UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Urban mobility safety in Barcelona: an evaluation of risk factors and techniques to achieve Vision Zero

Ahmad Aiash

ADVERTIMENT La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del repositori institucional UPCommons (http://upcommons.upc.edu/tesis) i el repositori cooperatiu TDX (http://www.tdx.cat/) ha estat autoritzada pels titulars dels drets de propietat intel·lectual **únicament per a usos privats** emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei UPCommons o TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a UPCommons (*framing*). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del repositorio institucional UPCommons (http://upcommons.upc.edu/tesis) y el repositorio cooperativo TDR (http://www.tdx.cat/?locale-attribute=es) ha sido autorizada por los titulares de los derechos de propiedad intelectual **únicamente para usos privados enmarcados** en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio UPCommons No se autoriza la presentación de su contenido en una ventana o marco ajeno a UPCommons (*framing*). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the institutional repository UPCommons (http://upcommons.upc.edu/tesis) and the cooperative repository TDX (http://www.tdx.cat/?locale- attribute=en) has been authorized by the titular of the intellectual property rights **only for private uses** placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading nor availability from a site foreign to the UPCommons service. Introducing its content in a window or frame foreign to the UPCommons service is not authorized (framing). These rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Urban Mobility Safety in Barcelona: An evaluation of risk factors and techniques to achieve Vision Zero

PhD thesis

Ahmad Aiash

Supervisor:

Prof. Francesc Robusté

PhD Program in Civil Engineering
Barcelona Innovative Transportation (BIT) Research group
Department of Civil and Environmental Engineering
Barcelona School of Civil Engineering
Universitat Politècnica de Catalunya – UPC BarcelonaTech

Barcelona, 2023

Acknowledgement

First of all, I would like to thank my supervisor Prof. Francesc Robusté for his support and guidance that has been provided and given to me along my entire PhD journey. Of course, without the support of my parents and sisters, I will not be able to achieve what I have achieved, and I will not be able to have this chance and have the PhD degree. And thanks to my friends who are always besides me and supporting me through my good and difficult times. I have to admit that it was a unique journey that I had here in Barcelona where I learned a lot from it, and I enjoyed my time. I would also like to thank the Doctoral Training Network (DTN) – European Institute of Innovation and Technology (EIT) Urban Mobility for giving me the chance to be part of it to develop my different skills that can help me to further my career and expand my network in Europe where I have met amazing people.

Abstract

In order to converge towards *Vision Zero* concept in road safety, traffic crashes need great study and analysis. This thesis is aimed at this goal, applied to the urban and metropolitan environment of Barcelona. The thesis is organized as a compendium of eight papers, each dealing with a specific safety risk factor.

We use different techniques according to the objectives and data: traditional statistical models, time series analysis, and machine learning techniques. Random Utility Models such as Binary Probit Model, Chi-square Automatic Interaction Detector, Bayesian networks such as Tree Augmented Naïve tree, and Kohonen clustering. The methodology is generic and can be extrapolated to other cities, regions/countries, and datasets.

When searching for risk factors of severe or fatal injuries, with the data of 47,153 traffic accidents that occurred in Barcelona between 2016 and 2019, we found the worst injured odds for males and 65 years (and older) persons, pedestrians and drivers (higher probabilities compared to passengers), weekends, afternoon, and night timings. Working hours timing and season (summer) have also influence in the number of severe crashes.

During the COVID-19 pandemic, all levels of injuries that resulted from traffic crashes have shown approximately 39% (slight), 30% (severe), and 36% (fatal) injuries reduction. City areas surrounded by higher noise levels, may lead to more traffic crashes as a potential distraction factor, but further research is needed (noise and traffic levels are correlated).

Barcelona had a total population of 1,636,762 inhabitants in 2019, distributed on 10 districts. Eixample (literally, "broadening" of the city, planned by civil engineer Ildefons Cerdà in 1856), which is the district that has the highest population, highest usage of private transport mode, and highest density of passenger cars/km^2 compared to all other districts, has the highest risk of having all types of injuries resulting from traffic crashes.

Elderly pedestrians are more likely to have severe or fatal injuries, especially during the morning period. When evaluating bicycle lanes and stations safety, the most vulnerable age category for cycling is 26 to 50 years old people during morning and evening timing. Of course, bicycle lanes and bicycle stations in zones that have a restricted speed limit of 30 km/h, showed a lower crash injury ratio.

Finally, motorcycle crash injuries have worrying high values in Barcelona, a motorcycle-city. Alcoholism and pavement in poor condition can indeed increase the probability of having different levels of injuries. Young motorcyclists (25-40 years old) have higher odds for crash injuries, unlike elderly users who are less likely to be involved in motorcycle crashes.

After analyzing all these factors in detail with the proper methodologies and available data, this document concludes with recommendations and future research lines.

Keywords: Traffic crash, accidents, risk factors, severities, injuries, data analysis, time-series, probit, CHAID, Bayesian network, TAN, Kohonen clustering, working hours, COVID-19, noise, city districts, pedestrians, bicycles, motorcycles, Barcelona.

Resumen

Para converger hacia el concepto *Visión Cero* en seguridad vial, conviene estudiar los accidentes de tráfico en detalle. Esta tesis se enfoca a este objetivo, aplicado al entorno urbano y metropolitano de Barcelona. La tesis está organizada como un compendio de ocho artículos, cada uno de los cuales trata un factor de riesgo de seguridad específico.

Utilizamos diferentes técnicas según los objetivos y datos: modelos estadísticos tradicionales, análisis de series temporales y técnicas de aprendizaje automático. Modelos de utilidad aleatoria, como el modelo *probit* binario, el detector de interacción automática de chi-cuadrado, las redes bayesianas, y el agrupamiento de Kohonen. La metodología es genérica y puede extrapolarse a otras ciudades, regiones/países y conjuntos de datos.

Al buscar factores de riesgo de lesiones graves o mortales, con los datos de 47.153 accidentes de tráfico ocurridos en Barcelona entre 2016 y 2019, encontramos las peores probabilidades de lesionarse para hombres y personas mayores de 65 años, peatones y conductores, fines de semana, y horarios de tarde y noche. El horario de trabajo y la temporada (verano) también influyen en el número de accidentes graves.

Durante la pandemia de COVID-19, todos los niveles de lesiones que resultaron de accidentes de tráfico han mostrado una reducción de aproximadamente 39 % (leve), 30 % (grave) y 36 % (mortal) para las lesiones. Los barrios con niveles de ruido más altos pueden dar lugar a más accidentes de tráfico como posible factor de distracción, pero se necesita más investigación al respecto (los niveles de ruido y tráfico están correlacionados).

Barcelona tenía una población total de 1.636.762 habitantes en 2019, repartidos en 10 distritos. Eixample (proyectado por el ingeniero de caminos Ildefons Cerdà en 1856), que es el distrito con mayor población, mayor uso del transporte privado y mayor densidad de turismos/km^2, tiene el mayor riesgo de tener todo tipo de accidentes de tráfico.

Los peatones de edad avanzada tienen más probabilidades de sufrir lesiones graves o fatales, especialmente durante el período matutino. Al evaluar la seguridad de las ciclovías y estaciones, los ciclistas más vulnerables son los de 26 a 50 años durante el horario de la mañana y la tarde. Por supuesto, en las zonas que tienen un límite de velocidad restringido a 30 km/h, se encontró una tasa de lesiones por choque más baja.

Las lesiones por accidentes de moto tienen valores preocupantemente elevados en Barcelona, una ciudad de motos. El alcoholismo y el pavimento en malas condiciones pueden aumentar la probabilidad de tener diferentes niveles de lesiones. Los motociclistas jóvenes (25-40 años) tienen mayores probabilidades de sufrir lesiones por accidentes, a diferencia de los usuarios de edad avanzada que tienen menos probabilidades.

Tras analizar todos estos factores en detalle con las metodologías adecuadas y los datos disponibles, el documento concluye con recomendaciones y futuras líneas de investigación.

Palabras clave: accidente de tráfico, factores de riesgo, gravedad, lesiones, análisis de datos, series temporales, probit, CHAID, red bayesiana, TAN, agrupación de Kohonen, horas de trabajo, COVID-19, ruido, barrios de la ciudad, peatones, bicicletas, motocicletas, Barcelona.

Resum

Per a convergir cap al concepte *Visió Zero* en seguretat viària, convé estudiar els accidents de trànsit detalladament. Aquesta tesi està enfocada a aquest objectiu, aplicat a l'entorn urbà i metropolità de Barcelona. La tesi està organitzada com un compendi de vuit articles, cadascun dels quals tracta un factor de risc de seguretat específic.

Utilitzem diferents tècniques segons els objectius i les dades: models estadístics tradicionals, anàlisi de sèries temporals i tècniques d'aprenentatge automàtic. Models d'utilitat aleatòria, el detector d'interacció automàtica de txi-quadrat, les xarxes bayesianes i l'agrupament de Kohonen. La metodologia és genèrica extrapolable a d'altres ciutats, regions/països i dades.

Quant a factors de risc de lesions greus o mortals, amb les dades de 47.153 accidents de trànsit ocorreguts a Barcelona entre el 2016 i el 2019, trobem les pitjors probabilitats de lesionar-se per a homes i persones majors de 65 anys, vianants i conductors, caps de setmana, i horaris de tarda i nit. L'horari de treball i la temporada (estiu) també influeixen en el nombre d'accidents greus.

Durant la pandèmia de COVID-19, tots els nivells de lesions que van resultar d'accidents de trànsit han mostrat una reducció d'aproximadament 39% (lleu), 30% (greu) i 36% (mortal) per a les lesions. Els barris amb nivells de soroll més alts poden donar lloc a més accidents de trànsit com a possible factor de distracció, però cal més investigació sobre això (els nivells de soroll i trànsit estan correlacionats).

Barcelona tenia una població total de 1.636.762 habitants el 2019, repartits en 10 districtes. Eixample (projectat per l'enginyer de camins Ildefons Cerdà el 1856), que és el districte amb més població, més ús del transport privat i més densitat de turismes/km^2, té el risc més gran de tenir tot tipus d'accidents de trànsit.

Els vianants d'edat avançada tenen més probabilitats de patir lesions greus o fatals, especialment durant el període matutí. En avaluar la seguretat dels carrils bici i les estacions, els ciclistes més vulnerables són els de 26 a 50 anys durant l'horari del matí i la tarda. Per descomptat, a les zones que tenen un límit de velocitat restringit a 30 km/h, es va trobar una taxa de lesions per xoc més baixa.

Les lesions per accidents de moto tenen valors elevats a Barcelona, ciutat de motos. L'alcoholisme i el paviment en condicions dolentes poden augmentar la probabilitat de tenir diferents nivells de lesions. Els motociclistes joves (25-40 anys) tenen més probabilitats de patir lesions per accidents, a diferència dels usuaris d'edat avançada.

Després d'analitzar tots aquests factors detalladament amb les metodologies adequades i les dades disponibles, el document conclou amb recomanacions i futures línies de recerca.

Paraules clau: accident de trànsit, factors de risc, gravetat, lesions, anàlisi de dades, sèries temporals, probit, CHAID, xarxa bayesiana, TAN, agrupació de Kohonen, hores de treball, COVID-19, soroll, barris de la ciutat, vianants, bicicletes, motocicletes, Barcelona.

Publications and Conferences

This thesis is a paper-based thesis that is comprised of eight papers that have been published in different journals, conferences, book chapter and/or still under-review:

1- Aiash, A. and F. Robusté. (2021). Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree. *International Journal of Injury Control and Safety Promotion*. 29(2): 256-264.

2- Aiash, A. and F. Robusté. (2021). Working hours and traffic accident injuries case study in Barcelona. *Transportation Research Procedia*. 58: 678-682.

3- Aiash, A. and F. Robusté. (2022). Traffic crash injuries occurrence varieties across Barcelona districts. *Revista Internacional de Desastres Naturales, Accidentes e Infraestructura Civil.* 22, 6.

4- Aiash, A. and F. Robusté. (2023). Total noise level's passive impact on traffic crashes: a potential distraction indicator. *International Journal of Environmental Studies.*

5- Aiash, A. and F. Robusté. (2023). Chapter 4 - COVID-19 pandemic effects on traffic crash patterns and injuries in Barcelona, Spain: An interpretable approach. In *XA2ITS: Explainable AI for Intelligent Transportation Systems.* 72- 89. Taylor & Francis – Routledge.

6- Aiash, A. and F. Robusté. (2023). Analyzing pedestrians' crash injury risk factors in Barcelona. *Transportation Research Procedia.* (Accepted)

7- Aiash, A. and F. Robusté. (2023). Evaluating bicycle lanes and stations safety in Barcelona: a Bayesian network approach to analyze injury severity. (Under review)

8- Aiash, A. and F. Robusté. (2023). Supervised and unsupervised techniques to analyze risk factors associated with motorcycle crash injuries. (Under review)

Table of Contents

List of tables

List of figures

1. Introduction and research objective

1.1. Introduction

Traffic crashes are still a wide cause of fatalities and severe injuries globally. Based on World Health Organization (WHO), more than 1.3 million fatalities are caused by traffic crashes besides the disability and injuries that are caused due to them. Additionally, traffic crashes have significant economic costs, including both medical expenses and lost productivity. This cost is estimated to be 3 % of the gross domestic product (GDP) for most of the countries. There are enormous risk factors that can impact the severity of traffic crashes and their occurrences. This situation is raising the need to examine potential risk factors to understand their impact. In some cases, a combination of certain risk factors can work as a group to affect these possible consequences and increase their severity. Evidence-based strategies can be then conducted, as a result, after analyzing and identifying the potential significant risk factors. For instance, for the policies that can be imposed through data-driven insights and analysis, new policies can be prepared and imposed to establish speed enforcement programs and campaigns, which in return can have a significant impact on traffic crash rates. Speed cameras or pedestrian crosswalks introducing can also be a result of the data analysis results. Overall, policymakers and researchers can conduct certain strategies that target the significant risk factors in order to lessen their impact after better understanding their impact.

In current times, data possesses vast shares of the daily routine of our existing life. This is coming from our dependency on new technology that is collecting the data naturally with categorizing the preferred choices based on each individual. These data is coming from as basically the way we communicate with each other to conducting business. Valuable insights and help us to make better decisions can be the potential advantages that can be obtained from these vast amounts of data can have. Therefore, grasping the importance of data has become increasingly essential. Extracting meaningful information, patterns, and specific details can all be obtained with the help of advanced analytics and machine learning techniques. Analyzing data can also help in predicting several factors in a variety of specializations, such as predicting disease outbreaks. A traffic crash is one of the research paths that data can play a critical role in understanding its causes and consequences. Eventually, policymakers and the responsible authorities can develop targeted measures to lessen these crashes and their severity. High-risk road users, such as drivers with a history of crashes and/or traffic violations can also be identified with the help of data collection and analysis. Targeted education and awareness campaigns to improve driver behavior and reduce the incidence of crashes can be developed. New laws can also be imposed to reduce these incidents by the increasing the amount of certain traffic fines. The effectiveness of traffic safety policies and interventions can be evaluated. This can be achieved by observing trends over time to identify the potential changes in the frequency and severity of crashes. For instance, policymakers and experts can compare crash rates before and after the implementation of a new policy or measure so that its effectiveness can be determined. Eventually, this type of extracted and concluded information can be utilized to refine policies in order to improve their impact.

An enormous number of risk factors can impact the incidence and severity of road crashes. The age of the person involved in the traffic crash is one of these risk variables. The influence of age on traffic crash abilities has been extensively researched, with several elements of traffic crashes included. According to some of the studies explained later in this thesis in the form of research articles, young and elderly road users are more likely to be involved in collisions than middle-aged road users. This might be owing to their impaired cognitive and physical capacities. For drivers, for instance, the danger of being involved in a traffic crash is greater within the first few months after receiving a driver's license, since new drivers tend to overestimate their skills while underestimating the risks connected with driving. This condition might be the amplifier for them engaging in unsafe activities such as speeding, tailgating, and texting. Peer pressure is another factor that might contribute to young drivers engaging in parlous driving. Drug and alcohol usage can also increase the likelihood of getting involved in a traffic crash for young drivers.

Vision, hearing, and response time tend to decline in senior drivers and as individuals age in general. This can have an impact on their decision-making ability, whether they are crossing the street as pedestrians or driving. Chronic medical disorders that might impair driving safety, such as arthritis, dementia, and Parkinson's disease, are more common among older persons. When having comparable medical issues, vision, memory, and motor abilities can all be compromised, making it harder for the elderly to react rapidly to various road conditions. Furthermore, due to their prospective medical condition, older road users are more prone to be under the influence of more drugs. Certain medical drugs might impair judgment and response time because they have side effects such sleepiness, dizziness, and impaired vision.

As previously demonstrated, the risk associated with age can be reduced. In many nations, for example, drivers must complete specific conditions before acquiring or renewing a license, as outlined in rules and legislation. To renew a driver's license, several nations require the elderly to pass a vision exam or attend a driving course. Several nations have enacted regulations requiring young drivers to complete a certain number of hours of supervised driving before receiving a full driving license. Driver education programs can help young drivers obtain the knowledge they need to drive safely. These education programs may comprise a course of teaching, practical training, and simulation exercises to imitate real-life driving conditions. Older drivers might benefit from such training to keep their knowledge and skills up to date, as well as to identify any concerns that may influence their driving ability. Finally, age might be a substantial risk factor in road accidents. As a result, this risk factor must be investigated in research articles pertaining to various forms of traffic accidents.

Gender is one of several additional characteristics that might influence the incidence and severity of traffic crashes. Gender is commonly seen as an indirect risk factor that might contribute to traffic crashes. Men accounted for more traffic collision deaths and severities than women, as will be examined further in the next portions of this thesis by offering various references and demonstrating prior processes. This tendency has been seen for a long time, owing to a variety of variables such as variances in driving habits and decision-making, vehicle selection, and risk-taking attitudes in general. Furthermore, studies have found that males are more prone than females to speed, drive while under the influence of drugs or alcohol, and participate in aggressive driving behaviors such as tailgating and weaving in and out of traffic. These actions can increase the likelihood of an crash and contribute to men's greater mortality rates. Females, on the other hand, are more careful while driving, which might result in less unsafe driving habits. These practices are associated with decreased crash risk and so lead to lower mortality rates. However, it is crucial to remember that gender is not the only element that determines driving behavior, since traffic collision occurrence and severity are generally influenced by a mix of variables such as age, experience, and culture. Finally, gender is a complicated risk factor that can impact the likelihood and severity of traffic crashes. As a result, safe driving practices in particular, and safe road usage in general, for both males and females, are critical considerations.

Another possible feature that might impact the incidence and severity of traffic crashes is a temporal risk factor. The temporal risk factor in road collisions refers to the impact of time-related variables on the frequency and severity of a crash. These factors might include the time of day, the day of the week, and the season. Understanding how these characteristics, along with the variables associated with them, contribute to the likelihood of a crash can be a critical step in devising successful methods for lowering the number of crashes and better understanding traffic crash patterns. Congestion is typically higher during rush hour in the morning and evening than during the night, which might contribute to an increase in collision occurrence. While, in terms of severity, the danger of a fatal collision is often higher at night than during the day due to various other variables, such as alcohol intake and over-speeding, as will be illustrated later in the sections. Another temporal component that influences the likelihood of a traffic crash is the day of the week. Weekends, holidays, and the night timing are linked with a greater risk of a crash because of the increased alcohol intake, exhaustion, and distraction, and individuals tend to mingle and stay out longer on weekends and holidays. Seasonal conditions can also influence the likelihood of a traffic crash. The level of traffic on the highways often increases during the summer months as more people take road trips. As a result, investigating temporal risk variables can play an essential role in creating effective solutions to reduce the frequency or severity of collisions. Increased law enforcement presence at high traffic periods, education programs to promote awareness of the hazards, and infrastructural upgrades are examples of such approaches.

Another important risk factor in traffic collision occurrence and severity is location or spatial characteristics. Because of the higher speed limitations that these routes often have, crashes on highways or other high-speed roads are frequently more serious than those on smaller roads. Intersections are another possible high-risk area owing to potential uncertainty or difficulties negotiating these sorts of intersections, which can lead to a higher chance of a traffic collision. Another factor to consider is the physical characteristics of the road itself, which might influence collision risk. Poorly built, inadequately lit, or in disrepair roads might increase the danger of a collision. A steep curve or blind spot on a particular route, for example, may make it difficult for drivers to notice incoming vehicles. Inadequate illumination on a certain route may also be a risk aspect for road users owing to a loss or impediment of vision. Another issue to examine while investigating the geographical variables of traffic collisions is the emergency response at a certain road or site. When a crash occurs in a distant or rural region, emergency services may take longer to arrive. Similarly, places with high traffic volumes or congestion may make emergency services more difficult to provide. Overall, recognizing these spatial elements can aid in the development of successful initiatives such as better road design, more traffic safety enforcement, and teaching drivers about the hazards associated with various types of roads and locations.

Certain individuals are more likely to suffer the consequences of traffic crashes than other users. According to the World Health Organization, vulnerable road users such as pedestrians, cyclists, and motorcyclists account for over half of the 1.3 million fatalities resulting from traffic crashes worldwide. Various factors can increase the severity of injuries resulting from these crashes, including characteristics of the vulnerable road users, such as their gender and age, the timing of the crash, and the location where the crashes occurred, as stated previously. Therefore, this thesis attempts to understand the impact of different risk factors on traffic crashes, in general, and the injuries resulting from these crashes. However, the focus of this study is not only considering specific types of injuries or crashes that are occurring in Barcelona. As a result, both machine learning techniques and traditional statistical models are employed in order to achieve the established target of this thesis. The findings of this thesis should provide holistic insights and results regarding urban mobility safety in Barcelona. Aspects that are included in this thesis are the overall traffic crashes status analysis with including several risk factors that may impact the level of injury as stated previously, user characteristics and spatiotemporal factors, besides focusing on vulnerable road users that may suffer more from traffic crashes compared to other road users including cyclists, pedestrians, and motorcyclists. To be more specific, this thesis is comprised of eight chapters, as each chapter represents a research paper.

The first paper focuses on the overall traffic crashes in Barcelona, including several risk factors that can contribute to traffic crashes. In this paper, a binary probit and a decision tree are employed to examine and identify the significant risk factors and determine their potential consequential impact on traffic crash severity. For the temporal factors, two papers are conducted to examine the impact of time on traffic crash injuries and occurrences. The first paper focuses on the COVID-19 impact on traffic crash injuries and patterns in Barcelona. For this study, a time series analysis is carried out to extract the impact of the pandemic events on traffic crashes and to provide better insights into it. The impact of certain risk factors is compared before and during the pandemic to investigate the differences that certain risk factors can affect the level of injury when a traffic crash occurs. In this paper, a prediction model is developed based on the previous traffic crash data that occurred before the pandemic in order to estimate the changes that occurred during the pandemic and compare the prediction with the actual number of traffic crashes. Additionally, and as stated previously, a comparison between the impact of certain risk factors before and during the pandemic is conducted by employing a decision tree model. The second paper investigates the impact of the working hour timing period and seasons of the year on traffic crashes focusing on the provided care for slightly injured persons. The conditional probabilities that are estimated for grasping the impact of the risk factors are calculated by a Bayesian network model by choosing the tree-augmented naive Bayes (TAN) technique.

Another aspect is the spatial factor that can influence traffic crashes. Therefore, a paper is conducted that considers traffic crash occurrence locations in Barcelona. This paper focuses on the frequency and severity of traffic crashes that may differ significantly from one district to another. As a result, this spatial influence on traffic crashes may also affect the level of injuries that is resulted from traffic crashes. There are ten distinct districts across Barcelona, each with a varied population and size. Eventually, this study aims to categorize these districts based on traffic crashes and the injuries that are caused by them, whether they are minor, severe, or fatal injuries. A Bayesian network represented by the TAN method is, then, employed to examine this impact.

Another different factor that may impact traffic crashes is the total noise levels. This study examines the trace of total noise levels on traffic crashes to determine its potential passive impact by investigating the correlations between the total noise levels in the city and drivers and pedestrians traffic crashes. The main reason for considering this factor is due to the lack of similar studies, up to our knowledge, that are focusing on this type of potential risk on road users that can eventually lead to traffic crashes. The trace of total noise levels is then measured by utilizing a multinomial logit regression. The road users that are included are drivers, pedestrians and passengers.

Three possible vulnerable road users may suffer more from traffic crashes compared to other road users, as stated previously. Pedestrians are part of those vulnerable road users who need more attention in order to identify and undermine the factors that can cause severe and fatal injuries in traffic crashes. Eventually, the possible risk factors that can cause traffic crash injuries are examined. Then, BN models are employed to determine the conditional probabilities between these chosen risk factors and the dependent variable, which is the level of injury resulting from traffic crashes.

Another vulnerable road user is the cyclist that can also be involved in a high-risk traffic crash. Therefore, this thesis considered this road user in order to provide a holistic insight related to vulnerable road users' risk factors in Barcelona. Part of the evaluation steps that should be followed is evaluating bicycle infrastructure safety due to the spike in utilizing them in order to provide sustainable transport modes. The spike in bicycle usage is captured besides the appropriate infrastructures increase. More than 496 stations and 400 km are designated and distributed in Barcelona. Geographic Information System (GIS) is utilized to extract bicycle lanes, stations, and zones and then merge it with bicycle crashes injuries data in 2019 in Barcelona, considering other risk factors, including gender, age, and the time of the day. For the analysis, a TAN model is employed to identify these potential risk factors and grasp their impact by determining their conditional probabilities.

Motorcyclists are one of the three vulnerable road users, and motorcycle is one of the highly used modes of transport in Barcelona. This situation can be attributed to consequences such as traffic crashes and severe and fatal injuries. Similar to the rest of traffic crash injuries, several factors can impact the traffic crash injury level, which needs to be investigated and examined carefully. In this study, gender, age, temporal factors, alcoholism, drugs, or medicines, conditions of the road, metrological factors, objects or animals on the road, and the crash type. Eventually, this study attempts to examine the impact of the aforementioned risk factors. Supervised and unsupervised approaches are utilized to examine these potential risk factors related to motorcycle crash injuries, including a binary probit and a Kohonen technique.

1.2. Research objective

It is not a surprise that the traffic crash research topic is one of the hot topics that is focused on by an enormous number of researchers due to the substantial consequences that can be caused on different levels: lives lost, disabilities, infrastructures, monetary losses, and time loss. However, each research considered a certain amount of risk factors that can lead to traffic crashes in the first place, and other researchers considered the level of injury that resulted from them. Imagining traffic crashes as a tree that is rooted deep down with its enormous number of roots and going up with its trunk and branches, the research topic that can be studied is somehow similar to it. The only solution to entirely stop traffic crashes is to cut all vehicles off, but this is a fantasy solution to be implemented in my own view in the current time. The thing is that vehicles are for sure considered as the mainstream for this massive issue, but for sure, it is not the only cause for it. There are different varieties of types of transport modes that can be utilized, from bicycles to scooters, even with skateboards, besides other transport modes such as trains and airplanes.

Imagine someone who coming back from his/her job in his/her private car while, on the other side, another person preferred to come back from his/her job using a bicycle and then suddenly had a traffic crash between both of them. By looking into the big picture, the speed of the car was slightly higher than the speed limit on the road and then hit the bicycle at the intersection. By giving a closer look into this crash, the site of the crash has not had enough light, no protected lanes for cyclists, and no traffic lights; the crash happened during the night timing, the person was 26 years old and was using the bicycle while the person who was using the car was 67 years old. All these factors can be fused together to form a case that led to this specific traffic crash. Therefore, analyzing an enormous number of traffic crash risk factors can provide meticulous insight that can provide a profile for a certain area or region, which in return can spot the light on these potentially hazardous features and help in targeting them by proactively acting to prevent this type of factors or alleviate their consequences impact or simply acting on the current situation if there is any risk that should be reduced or any measure that should be improved.

Back to the tree of traffic crashes, the formation of this tree is formed by these factors. Treating this issue is the way how the solution can be found. Out of these treatments is the process of carefully analyzing the risk factors, which in the following chapters, the different strategies of the current governments and local authorities to reduce traffic crash occurrences and severities and the initiatives that are related to this reduction are going to be unfolded. However, covering all possible sick branches of this tree is not possible through this thesis, as each possible topic can stand on its own as a whole thesis, but the focus of this thesis is to provide an overall insight into different risk factors that are endangering road users in Barcelona. Some uncommon factors are also covered as a part of this thesis to be prepared for further studies that can be conducted in the future for researchers. Counting all the branches that are not covered in this thesis is deemed to be a strenuous task to be carried out, but some of the thesis topics can be enumerated. Mobile phone use is one of the thesis's possible perilous factors that can amplify the risk of traffic crashes, which is not considered in this thesis. The reason for this is that it requires certain equipment and specific labs that can measure this risk besides involving participants that should test different scenarios to determine this risk. Traffic violations that are committed by road users can also be linked to traffic crashes. However, this type of data was not available, so it cannot be included in this thesis. Overall, the main obstacles that are hindering expanding and covering a wider range of topics related to traffic crashes in this thesis are mainly data availability and the existence and availability of relevant equipment that is required to conduct the research.

So why in this thesis and through most of the papers that are conducted in this study are stating a crash, not an accident? The answer is very simple, which is a crash can be avoided if carefully examined, while an accident can just happen without any preapprehensions. Two papers stated traffic accidents in this thesis, not traffic crashes, as both were conducted during the early stages of developing this thesis, and through understanding and analyzing traffic crashes, more insights were spotted and grasped, which led to changing the term accident to crash.

As a result, several datasets are utilized besides different techniques to analyze, identify, and predict risk factors. The reason for this is to lessen the number of these types of traffic crashes by detecting potential risk factors and their impact. Afterward, road authorities can legislate the required laws to undermine their influence on traffic crash severities or even can promote self-awareness recommendations for road users regarding these factors. Additionally, it can help the authorities to keep in mind these factors while constructing new roads to proactively act to prevent them. Eventually, this study can be considered as a path for the Vision Zero goal in Barcelona, which is aiming to have zero severe and fatal injuries resulting from traffic crashes.

Intentionally blank page

2. Traffic accident severities analysis in Barcelona using a binary probit and CHAID tree

2.1. Abstract

Traffic accidents are still wide causation for fatalities in the globe. The set of alarm for this cause of deaths is still on, due to the fact that the number of fatalities is still representing an enormous issue and a challenge for most governments. In Barcelona, similar to the rest of the world, traffic accidents are threatening lives and raising the need to lessen the number of both fatalities and severities. This study is conducted to grasp the correlations between different classification factors with accident severities and fatalities. A total of 47,153 traffic accident cases that occurred between 2016 and 2019 are utilized. Then, a binary probit model is exploited to grasp the impact of several risk factors. The results confirmed that males and 65 years and older injured persons are more exposed to severe or fatal injuries compared to other categories. Pedestrians and drivers are found to have higher probabilities compared to passengers in being involved in severe or fatal injuries. Weekends, afternoon, night timings all have higher odds of having severe or fatal traffic accidents. Chi-square automatic interaction detector is, then, applied with depicting similar results with more details regarding the categories of the independent variable. The findings of this study can help road authorities in targeting these risk factors that can increase fatalities or severities of traffic accidents in order to mitigate their impact or lessen the number of this type of accidents to achieve Vision Zero.

***Keywords*:** Barcelona, Accidents, Severities, Probit model, CHAID

2.2. Introduction

"Every fatal or serious crash on our roads is a tragedy. It is our moral obligation – our shared responsibility – to take road safety seriously." Violeta Bulc. A traffic accident is still a major issue in Europe and the rest of the world. About 25 thousand lives were claimed, while more than a million persons were injured as a cause of traffic accidents in 2016 in Europe (European Road Safety Observatory, 2018). A report (Carson et al., 2020) was prepared to provide a performance index for 32 European countries regarding road safety. The report showed that there was a 3% decrease in the number of accidents in 2019 compared to 2018 for the total number of accidents that happened in the EU27. 7,023 road fatalities less in 2019 than in 2010 for the total number of EU27. Greece has the highest decreasing percentage with -44.4%, while

Malta had the highest increasing percentage of 6.7%. For Spain, the percentage of road fatalities was 30.4% by comparing the two years 2010 and 2019. However, the medium-term target was halving the number of deaths, but as shown, the progress of lessening the number in some countries is stagnated.

In 2017, the EU transport ministers adopted the Valletta declaration to reduce the number of severe injuries, which can be considered as a consistent plan with the EU Commission (Council of the European Union, 2017). The European Commission has also set a new target to lessen the number of serious injuries to half between 2020 and 2030. Most European countries, therefore, have set this time period for the new national road safety strategy that includes a target for death and severe injuries reduction. Key Performance Indicators (KPIs) were introduced to assess the performance of meeting the targets during the defined period (European Transport Safety Council, 2019). Those indicators are speed compliance, utilizing the seat belt beside the child restraint system, protective equipment using, alcohol influence during driving, distractions while driving that include handheld devices, new cars safety, the safety of the infrastructure, and lastly, the care for post-crash. The reason for setting such indicators is acting in a proactive manner for identifying the problems due to their ability in providing a full picture of the status of the level of road safety (European Transport Safety Council, 2018). Spain has set a strategy between 2011-2020 (Traffic General Directorate, 2011) to achieve safe and sustainable mobility. This strategy is mainly comprised of five main elements. The ecological aspect is one of those elements that is consisted of reducing air and noise pollution and improving energy efficiency besides other targets. The second aspect is the competitive aspect that aims to improve journeys quality and goods distribution alongside guaranteeing the regularity of journeys. For the safe aspect, this element is setting the plan to reduce the number of traffic accidents and severe injuries in parallel with enhancing the care that is given to road accident victims. Moreover, this aspect is aiming to decrease the rate of accidents among vulnerable road users. The healthy element is consisting of promoting cycling and walking and enhancing the mental and physical health of people. Lastly, the universal aspect that composed of securing an acceptable cost for all social sectors who use public transport. Additionally, the accessibility for all modes of transports should be improved for all people. The order for elements is negligible as all the aspects are important and part of the plan. Some aspects have more categories. However, the most relevant categories are mentioned. The vision for the Spanish strategy is “The citizens have the right to a Safe Mobility

System in which everyone, citizens and agents involved, has a responsibility". This vision is consisted of five different values shape this vision including shared rights and duties, sustainable mobility, safe users, roads, environments, and vehicle.

Catalonia is, indeed, on the same path of having a strategic plan related to traffic accidents. The first strategic plan was established in 1999, and now Catalonia has a total of six strategic plans including the last published plan (Servei Català de Trànsit, 2014). The target of reducing the number of traffic accident severities and fatalities was met in the last three prepared plans, as all of them exceeded the planned reduction percentages. Based on the last report (Servei Català de Trànsit, 2014), the number of fatalities is one of the main challenges that still needs improvement in both urban and inter-urban areas. The priority for vulnerable road users should be maintained. Campaigns can also be conducted to target risky road users. Therefore, pillars were set to pledge the success of the implementation of the strategic plan. Law enforcement, search for efficiency and focus results, education and training, aspiring to leadership in Europe, shared responsibility in road safety, and lastly, the use of technologies associated with road safety are those main pillars. The overall target for the last strategic plan is to reach 50 % less in the number of fatalities in 2020 compared to 2010, while the overall target for 2050 is to reach zero fatalities and severe injuries with lifelong consequences under the Vision Zero.

Barcelona has also set a plan for the period 2019-2022 regarding road safety (Servei catala de Transit 2019). A 20 % reduction in fatalities is set to be a goal to be accomplished in 2022 compared to the number of fatalities in 2018. For severe injuries, a 16 % reduction should be reached in 2022 compared also to 2018. Therefore, a number of aspects should be considered in order to achieve those goals. The accidents that can lead to severe injuries should have the priority to prevent the occurring of similar accidents. Vulnerable users should also have the priority in identifying certain measures to protect them. The reasons for causing a high number of accidents including accident types and specific road users should be resolved. The causes for fatalities related to traffic accidents are also considered so that the fatalities numbers can be decreased. Four main action domains are conducted to achieve the proposed goals. Education is one of the four spheres that composed of prevention, training, communication, and research action and projects related to road safety. Engineering is the second sphere that is considered to sustain the proposed targets by involving in the technical measures related to vehicles, roads, and other related aspects. The third sphere is enforcement that means the correction of conduct which includes the implication of different regulations and the operations planning. Lastly, support and patronage measurements for the road accidents victims are under the last sphere which is emergency. Identifying the reasons for traffic accident severities and deaths, as mentioned earlier, are one of the spheres that are laid to undermine their impacts and reducing the number of fatalities and severities. Several studies were conducted to comprehend the effect of different potential factors so that the risk factors can be targeted and neutralized.

A study (García-Altés and Pérez, 2007) revealed the economic cost of traffic accidents in Barcelona in 2003. A 367 million is estimated for the total accident cost of accidents with 329 million for the direct cost. The gender differences in being involved in a risk injury due to traffic accidents by age, mode of transport, and severity were studied by research (Santamariña-Rubio et al., 2014). The data comprised of the time spent by people, who are residents and over three years old, in traveling in Catalonia in the period between 2004 and 2008. A Poisson regression model was exploited to delve into those differences. A significant impact in traffic accident injuries for both gender and age was distinguished. Males were found to have a higher risk of having severe and slight injuries amid the group of child pedestrians and young drivers. In contrast, females had a higher risk of injuries in older groups. Furthermore, the fatalities rate of males was higher than females. Most of the pedestrian fatality who are hit by a motorized two-wheeled vehicle is found to be because of Traumatic Brain Injury (Rebollo-Soria et al. 2016). Another study (Ferrandoet et al., 1998) examined the gender, age, and type of user impact on injuries in Barcelona. The data represented the traffic injuries cumulative incident rate of disability in 1993. Males were more exposed to the incidence compared to females when considering motorcycle and car users. Female pedestrians were found to be more involved in incidents compared to males. Consistently, additional research (Hidalgo-Fuentes and Sospedra-Baeza, 2019) showed that males drivers were more engaged in a higher level of severities due to motorcycle/moped accidents compared to females. For age categories, research (AlKheder et al., 2020) concluded that drivers who are older than 60 years were found to be more exposed to severe or fatal injuries compared to other age categories.

Volume to capacity ratio (v/c) influence on accident rates was explored by a study (Zhou and Sisiopiku, 1997) with including different other risk factors. The study exposed that there is an inverse relationship between the number of the v/c ratio and the number of traffic accidents. In the case of weekdays and weekends comparison related to traffic accidents and when the v/c ratio is high, weekdays had a higher number of traffic accidents. Contradictory, a recent study

(Kufera et al., 2020) revealed that crashes had higher odds of happening on weekends compared to weekdays when the impact of a new casino was examined in suburban Another study (Ramírez and Valencia, 2021) presented the correlations between the spatiotemporal factors and injuries and fatalities related to traffic accidents in Columbia. The findings of this study showed that there was a higher risk of being involved in traffic accidents from Wednesday to Saturday compared to Sunday. Additionally, dry and sunny weather was found to have an impact on decreasing the odds of having traffic accidents similar to the impact of a larger number of traffic lights. In Barcelona, a study (Robuste et al., 2003) found that speed limits were not followed by drivers as the average measured speed limits were higher than legal speeds.

Therefore, the main objective of this study is to analyze different classification variables that are associated with traffic accidents' severities. The reason for this is to lessen the number of these types of traffic accidents by detecting the potential risk factors. Afterward, road authorities can legislate the required laws to undermine their influence on traffic accidents severities. Additionally, it can help the authorities to keep in mind these factors while constructing new roads to proactively act to prevent them. Eventually, this study can be considered as a path for the Vision Zero goal in Barcelona, Spain. Two approaches are utilized parametric and non-parametric models including the probit model and Chi-square automatic interaction detector model. This study exploited a large number of traffic accidents that occurred in Barcelona including 47,153 cases with focusing on five potential risk factors. Till now, no similar study is conducted in Barcelona with this amount of data and applying similar approaches. This study should highlight those potential independent variables alongside validating the models that are exploited for the data provided.

2.3. Methodology

2.3.1. Traffic accidents data in Barcelona

Barcelona province is providing a public access database for different data types which is called "Ajuntament de Barcelona's open data service". The dataset themes include administration, city and services, economy and business, population, and territory theme. Traffic accidents are classified under the city and services theme (Barcelona's City Hall Open Data Service, 2020). The data that is exploited in this study is comprised of 4 years duration from 2016-2019. The data is, also, consisting of different categories for the response variable that is the level of injury that consisted of different levels of slight injuries, severe injuries, and fatal injuries. In more detail, each category is also divided into other subcategories. However, slight injury cases are all merged together to facilitate the analysis procedure in this study. For different severe and fatal injuries, both are integrated together. Therefore, a total of only two categories are remained, so that that the response variable is, now, a binary variable. The reason for this is, both severe and fatal injuries are focused on to determine the factors that impact the occurrence for both classes and because of the number of fatal injuries cases to have a better statistic estimate. Moreover, the objective of this study is help road authorities in achieving the Vision Zero target for both fatalities and severe injuries results from traffic accidents. Therefore, both severe and fatal injuries are focused on in this study. Alongside combining response variable classes, blank and undefined data are eliminated from the final data set as a part of the data preparation. Moreover, some classification variables categories are combined, as well, based on the data and to have a better insight into the dataset. Therefore, the process of classifying the variables is carried out. In general, some statisticians can describe the variable classification as a process to classify a variable as a categorical or continuous variable. In more details, categorical variables are classified into subgroups including nominal, ordinal, and dichotomous variables.

Table 1, as shown, is presenting the classification variables frequencies with respect to injuries in the city of Barcelona during 2016-2019. The accidents reported by the Guàrdia Urbana in the city of Barcelona. A total of 47,153 report cases are utilized after the data preparation has been achieved. A total of five different classification variables are fetched for this study in order to apply a model that can exhibit the correlations and how each factor can impact the dependent variable. It is worth mentioning that the accidents occurred in the urban area as all the reported accidents happened inside the city region. The difference between rural and urban accidents can be noticed as higher number of pre-hospitals deaths were found at rural areas compared to urban area (Li et al., 2008). The study found that the lower use of seatbelt, helmets or other safety precautions were the reasons for the deaths in the rural region. Similarly, the accident injury level at a rural area was found to be higher compared to an accident at an urban area (Hao and Kamga, 2017). Weekends and weekdays are classified under the week which is an independent variable. Daytime is another independent variable that is also considered that includes morning, afternoon, and night timings. In addition, gender impact is considered. Four different age categories are considered in this study including injured who are less than 15 years old, between 16 and 24 years old, between 25 and 64 years, and lastly, injured who are 65 years and older. Three different user types are implicated including pedestrian, passenger, and drivers.

Table1. Classification variables frequencies.

		Injury level		
		Severe or fatal injury	Slight injury	Total
Week:				
	Weekend	246	9359	9,605
	Weekday	693	36,855	37,548
Daytime:				
	Afternoon	488	23274	23762
	Night	160	5215	5375
	Morning	291	17725	18016
Gender:				
	Female	305	18156	18461
	Male	634	28058	28692
Age:				
	0 - 15	25	1513	1538
	16-24	121	6893	7014
	25-64	655	34537	35192
	65 +	138	3271	3409
User:				
	Pedestrian	269	4574	4843
	Passenger	86	8989	9075
	Driver	584	32651	33235
Total		939	46214	47153

2.3.2. Binary probit model

In this study, a binary probit model is a traditional statistical method that lays under the parametric approaches that are employed to grasp the correlations between the classification variables and the response variable by following previous proceedings (Yu and Abdel-Aty, 2014) (Garrido et al., 2014). SAS University Edition is exploited for employing the binary probit model (SAS, 2013). The reason for choosing the probit model is having the level of injury as a dependent variable and establish, later on, a comparison between this traditional model and the data mining technique in the overall results and the concluded correlations. The employed data is comprised of a binary response variable which consists of the first category which is slight injury, and the second category is severe or fatal injury. Assume for each injured i, U_i is the net utility of injury level. The total number of exogenous variables (X_i) is I. U_i is related to X_i.

$$U_i = X_i\beta + \varepsilon_i \tag{1}$$

Where the parameter estimates and explanatory variables vectors are β and X_i, respectively. The error is ε_i and it is assumed to follow a standard normal distribution.

The latent model is

$$y_i = X_i\beta + \varepsilon_i \tag{2}$$

When the relationship between U_i and the observable injury level y_i satisfies:

$$y_i = \begin{cases} 1 & if\ \ U_i > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

$\Pr\,(y_t = 1|x_t)$ is the conditional probability. The change in this conditional probability is measured by j^{th} which is the coefficients vector element $\beta, \beta_j(j\epsilon\{1,2,\dots,I\})$. This change is measured when there is a unit change x_i^j. j^{th} is an element in vector X_i. The conditional probability is also assumed to have a normal distribution. Then, the standard normal cumulative distribution function is $\phi(.)$. The conditional probability function is:

$$\Pr\,(y_i = 1|x_i) = \phi(X_i\beta) \tag{4}$$

2.3.3. CHAID tree

Decision trees (DTs) models, in general, are considered as a classifier method that uses values of attributes to make a discrete prediction. DT is normally a non-parametric technique that does not have any dependency on any functional form and no prior probabilistic knowledge requirement (Abellán et al., 2013). Chi-square automatic interaction detector that is abbreviated as CHAID tree is part of DT family. This model is developed by (Kass 1980) as an efficient approach for tree growing. (Dobra, 2018). To compare the results that are extracted from the binary probit model with a machine learning model, CHAID tree is chosen due to its ability in determining the correlations between the independent variables and the dependent variable (Berry and Linoff, 2004). Several reasons are behind choosing CHAID tree to analyze the data. Firstly, it can develop two levels and more for any tree level. Therefore, this technique is not only a binary model as in this study several independent variables have more than two categories. This technique can also handle all types of variables and provide a good classification for the chosen variables by developing a simple chart that depicts the correlations for the interpretation process that follows the deployed model as the interpretation process can be a hideous process with other data mining techniques. It has to be mentioned that several other techniques were explored before choosing the CHAID tree as the data mining technique in this study. Neural network, XGBoost tree, and Linear Support Vector Machine have all provided similar accumulated accuracy with 98.059 % accuracy similar to the employed CHAID tree. Additionally, CHAID exploits statistical significance test as criterion and evaluate all values of independent variables. Similar values are merged that are related to the dependent variable and other values that are not homogeneous are maintained. The first branch of the constructed tree is formed by the best independent variable. Then, each child node is developed from a group of the similar (homogeneous) values. Continually, this process is maintained till the tree is developed. Due to the dependent variable that is used in this paper is categorical variable, chi-square test is used as a measurement level of this variable.

The type of the dependent variable determines the process of calculating the p-values. In this study, the target variable is a categorical variable. Therefore, Pearson chi-square statistic and its corresponding p-values are calculated as shown in equation 5 and 6, respectively. Where a chi-square distribution $d = (J-1)(I-1)$ degrees of freedom is followed by X_d^2 . Where X is the predictor and Y is the target variable. The observed cell frequency is n_{ij}, while the expected cell frequency $(x_n = i \, , \, y_n = j)$ is $\widehat{m}_{ij}$. Categories are compared during the merge process to consider only the records that belong to the comparison categories. All the current nodes are utilized during the split step as all the categories are considered in p-value calculations. Based on the observed and expected frequencies, the likelihood ratio chi-square and its p-value is calculated in equation 7 and 8, respectively. For the expected frequencies it is calculated based on two cases. The one with no case weights which is calculated as shown in equation 10 and 11. For the one with specified weights, the expected frequencies are calculated in equation 12 and 13. α_i and β_i are estimated parameters. This model is exploited using IBM Watson with using SPSS modeler set. More details and steps for this algorithm can be fround at guide (IBM 2016).

$$n_{ij} = \sum\nolimits_n f_n I(x_n = i \wedge y_n = j) \tag{5}$$

$$X^2 = 2 \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}} \tag{6}$$

$$p = \Pr(X_d^2 > X^2) \tag{7}$$

$$G^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij} \ln(\frac{n_{ij}}{\widehat{m}_{ij}}) \quad (8)$$

$$p = \Pr(X_d^2 > G^2) \quad (9)$$

$$\widehat{m}_{ij} = \frac{n_{i.}n_{.j}}{n_{..}} \quad (10)$$

Where

$$n_{i.} = \sum_{j=1}^{J} n_{ij}, \quad n_{.j} = \sum_{i=1}^{I} n_{ij},\ n_{..} = \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}. \quad (11)$$

$$\widehat{m}_{ij} = \overline{w}_{ij}^{-1}\alpha_i\beta_i \quad (12)$$

Where

$$\overline{w}_{ij} = \frac{w_{ij}}{n_{ij}}, \quad w_{ij} = \sum_{n\in D} w_n f_n I\,(x_n = i \wedge y_n = j) \quad (13)$$

2.4. Results and discussion

Table 1 depicts how the model is significant. All likelihood ratio, score, and Wald are statistically significant with p-values less than 0.0001 for the applied probit model. Table 2 presents the significance of the classification variables that are used in the model. This table is showing the hypothesis tests for each variable individually. Weekday, the daytime, gender, age, and user are significantly improving the model-fit based on their Chi-square test and their p-values.

Table 1 Testing Global Null Hypothesis

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.0512	9	<.0001
Score	519.1789	9	<.0001
Wald	431.3261	9	<.0001

Table 2 Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Week	1	20.4102	<.0001
Daytime	2	48.5499	<.0001
Gender	1	21.2514	<.0001
Age	3	41.1812	<.0001
User	2	279.5212	<.0001

Table 3 portrays the coefficient estimates for all the classification variables alongside their standard errors, Chi-square test statistics, and their p-values. Accidents on weekends compared to accidents on weekdays increases the odds of having severe or fatal injuries by 0.1474. In Madrid, the probabilities of having traffic accidents, in general, on weekends were found to be higher compared to weekdays (Kanaan et al., 2009). Weekends and public holidays were also found to have higher odds of severe injuries that are resulted from a traffic accident (Zhang et al., 2013). Accidents that happened in the afternoon and night period compared to accidents that happened in the morning period rise the log-odds of having severe or fatal injuries by 0.1263 and 0.3010, respectively. In contrast, a study found that the early morning period alongside other factors can lead to higher odds of having severe injuries (Quddus et al., 2002). Females injured decrease the odds by 0.1433 compared to males injured. This finding is consistent with a previously conducted study that found that males were more likely to have severe or fatal injuries compared to females (Vorko-Jović et al., 2006). For age categories, injured who was 15 years old and younger, between 16 and 24, and between 25 and 64, all decrease odds compared to injured who were older than 65 years old with 0.4948, 0.283, and 0.2282, respectively. To compare this age category, a study concluded that drivers who were between the age group of 46-55 were at higher risk compared to older drivers (Kadilar, 2016). Pedestrians have an increased odd of having fatal or severe injuries with 0.5336 compared to drivers, while passengers have reduced the odds of having severe or fatal injuries by 0.2259 compared to drivers. A study that was conducted in France confirmed that pedestrians are at a higher risk compared to car occupants (Bouaoun et al., 2015). Besides the chosen and concluded risk factors in this study, there are other potential risk factors that can influence traffic accident severity such as the vehicle type. For instance, bus accidents were found to increase the odds of having fatal injuries when the bus crushed with non-motor vehicles (Samerei et al., 2021). Another conducted study (Samerei et al., 2021) focused on the factors that impact bus-pedestrian accidents and found that intersections, darkness, pedestrians walking on carriageway with traffic, high speed zone alongside other factors can lead to a higher risk of pedestrian fatality.

Table 3 Analysis of the classification variables[1]

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.9796	0.0515	1477.5193	<.0001
Week	Weekend	1	0.1474	0.0326	20.4102	<.0001
	Weekday (Reference)	0	0	.	.	.
Daytime	Afternoon	1	0.1263	0.0308	16.7993	<.0001
	Night	1	0.301	0.0438	47.2094	<.0001
	Morning (Reference)	0	0	.	.	.
Gender	Female	1	-0.1433	0.0311	21.2514	<.0001
	Male (Reference)	0	0	.	.	.
Age	0 -15	1	-0.4948	0.0938	27.8364	<.0001
	16-24	1	-0.2833	0.058	23.8553	<.0001
	25-64	1	-0.2282	0.0475	23.0581	<.0001
	65+ (Reference)	0	0	.	.	.
User	Pedestrian	1	0.5336	0.0387	190.038	<.0001
	Passenger	1	-0.2259	0.0473	22.8035	<.0001
	Driver (Reference)	0	0	.	.	.

Figure 1 is depicting the developed tree for the applied model including all previously mentioned independent variables. The tree is presented for the 30 % testing set after the model 70 % training set. The thickness of the branch is presenting the number of accident injuries. The thicker branch means that this branch has a higher number of traffic accident injuries. Each node has the number of injuries that it represents including both slight injuries number and the severe or fatal injuries number. The categories are only mentioned that are branches of the node which represent a different independent variable. The most important variable is the one that the model started with which is the user that includes driver, passenger, and the pedestrian. In general, and through scanning the developed tree, males have a higher number of severe or fatal injuries compared to females during certain conditions for drivers. Older pedestrians are also at great risk compared to other age groups. Passengers have higher odds of having severe or fatal injuries during the evening or night period compared to the morning

[1] The reference for the response variable is severe or fatal injuries category.

period during weekends. In contrast, drivers are more likely to have severe or fatal injuries resulting from traffic accidents during the morning period.

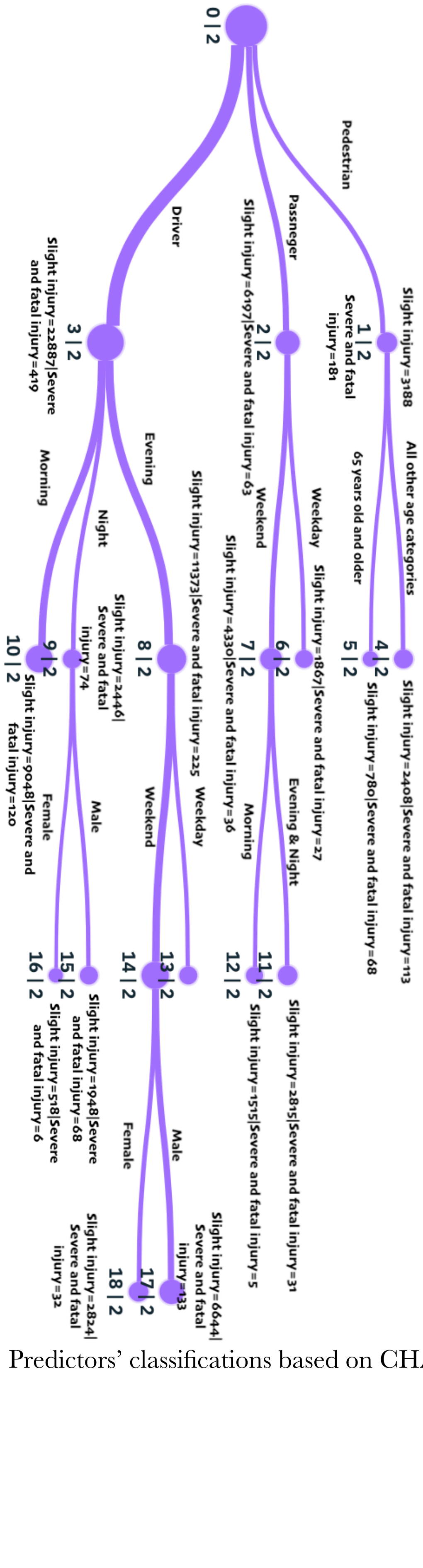


Figure 1 Predictors' classifications based on CHAID tree

2.5. Conclusions

Traffic accidents are still wide causation for fatalities around the globe. The set of alarm for this cause of deaths is still on due to the fact that the number of fatalities is still representing an enormous issue and a challenge for most governments. In Barcelona, similar to the rest of the world, traffic accidents are threatening lives and raising the need to lessen the number of both fatalities and severities. Different strategies and plans are set to meet this target; however, the number of deaths is exemplifying a formidable issue as each life counts, alongside the impact of traffic accidents in causing severe injuries and disabilities. Eventually, this study is conducted to attempt to specify the correlations between certain chosen classification variables and locate the areas of high occurrence for traffic accident severities and fatalities. The results of this study can be considered as an initiation for future studies that deem traffic accident research in Barcelona. A binary probit model is exploited for this reason to determine the correlation between the classification variables and the response variable. In this study, five different classification variables are considered including temporal factors represented by two risk factors, age of the injured, gender, and lastly, user type including the driver, pedestrian, and passenger.

The impact of the selected factors is studied by the binary probit model. The results of the model show that females injured have a low probability of being involved in severe or fatal injuries compared to males. For the temporal factors, accidents that happen on weekends, afternoon, or night have higher odds of having severe or fatal injuries for the involved parties. Injured who are 65 years or older are more engaged in severe or fatal accidents compared to other age categories. Pedestrians are also more implicated in severe or fatal injuries. Following the results of the applied binary probit model, the results of the applied CHAID tree are consistent with the previous results. Driver, passenger, pedestrian are the most important predictors based on the applied CHAID. Older pedestrian, male drivers during the evening and night period are found to have a higher risk of fatal or severe injuries compared to other categories. These findings can help road planners and road safety auditors to detect these factors that can lead to a higher risk of having severe or fatal injuries. Then, legislate rules that can alleviate their impact or lessen the number of this type of accidents.

2.6. References

Abellán, J., G. López, and J. de Oña. 2013. "Analysis of traffic accident severity using Decision Rules via Decision Trees." *Expert Systems with Applications* 40 (15): 6047-6054.

AlKheder, Sharaf, Fahad AlRukaibi, and Ahmad Aiash. 2020. "Risk analysis of traffic accidents' severities: An application of three data mining models." *ISA Transactions* 106: 213-220 .

Barcelona's City Hall Open Data Service. n.d. *People involved in accidents managed by the Police in the city of Barcelona.* https://opendata-ajuntament.barcelona.cat/data/en/dataset/accidents-persones-gu-bcn.

Berry, M. J. A., and G. S. Linoff. 2004. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* Wiley.

Bouaoun, L., M. M. Haddak, and E. Amoros. 2015. "Road crash fatality rates in France: A comparison of road user types, taking account of travel practices." *Accident Analysis & Prevention* 75: 217-225.

Carson, J., D. Adminaité-Fodor, and G. Jost. 2020. *14th Road Safety Performance Index Report.* Brussels: ETSC.

Council of the European Union. 2017. "Council conclusions on road safety - endorsing the Valletta Declaration of March 2017." Brussels.

Dobra, A. 2018. "Decision Tree Classification." *Encyclopedia of Database Systems.* Springer, New York, NY.

European Road Safety Observatory. 2018. *Annual Accident Report 2018.* European Commission.

European Transport Safety Council. 2018. *Briefing: 5th EU Road Safety Action Programme 2020-2030.* ETSC.

European Transport Safety Council. 2019. *Briefing | EU Strategic Action Plan on Road Safety.* ETSC.

Ferrando, J., A. Plasència, E. MacKenzie, M. Orós, P. Arribas, and C. Borrell. 1998. "Disabilities resulting from traffic injuries in barcelona, spain: 1-year incidence by age, gender and type of user." *Accident Analysis & Prevention* 30 (6): 723–730.

García-Altés, A, and K Pérez. 2007. "The economic cost of road traffic crashes in an urban setting." *Injury Prevention* 13: 65-68.

Garrido, R., A. Bastos , A. de Almeida , and J. P. Elvas . 2014. "Prediction of road accident severity using the ordered probit model." *Transportation Research Procedia* 3: 214 – 223.

Hao, W., and C. Kamga. 2017. "Difference in rural and urban driver-injury severities in highway–rail grade crossing accidents." *International Journal of Injury Control and Safety Promotion* 24 (2): 174-182.

Hidalgo-Fuentes, Sergio, and María J. Sospedra-Baeza. 2019. "Factores asociados a los accidentes de motocicleta en Barcelona, España Contributing factors of motorcycle crashes in Barcelona, Spain." *Ciencias Psicológicas* 13 (2): 265-274.

IBM. 2016. *IBM SPSS Modeler 18.0 Algorithms Guide.* IBM Corporation.

Kadilar, G. O. 2016. "Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach." *International Journal of Injury Control and Safety Promotion* 23 (2): 135-144.

Kanaan, A., P. Huertas, A. Santiago, J. A. Sánchez, and P. Martínez. 2009. "Incidence of different health factors and their influence on traffic accidents in the province of Madrid, Spain." *Legal Medicine* 11 (1): S333-S336.

Kass, G. V. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29 (2): 119-127.

Kufera, Joseph A., Ahmad Al-Hadidi, Daniel G. Knopp, Zachary D.W. Dezman, Timothy J. Kerns, Olasunmbo Eva Okedele, Geoffrey L. Rosenthal, and J. Kathleen Tracy. 2020. "The impact of a new casino on the motor vehicle crash patterns in suburban Maryland." *Accident Analysis and Prevention* 142.

Li , M.–D., J.–L. Doong , K.–K. Chang , T.–H. Lu , and M.–C. Jeng. 2008. "Differences in urban and rural accident characteristics and medical service utilization for traffic fatalities in less-motorized societies." *Journal of Safety Research* 39 (6): 623-630.

Quddus, M. A., R. B. Noland, and H. C. Chin . 2002. "An analysis of motorcycle injury and vehicle damage severity using ordered probit models." *Journal of Safety Research* 33 (4): 445-462 .

Ramírez, Andrés Felipe, and Carlos Valencia. 2021. "Spatiotemporal correlation study of traffic accidents with fatalities and injuries in Bogota (Colombia)." *Accident Analysis & Prevention* 149.

Rebollo-Soria, M. C., C. Arregui-Dalmases , D. Sánchez-Molina , J. Velázquez-Ameijide, and I. Galtés. 2016. "Injury pattern in lethal motorbikes-pedestrian collisions, in the area of Barcelona, Spain." *Journal of Forensic and Legal Medicine* 43: 80-84.

Robuste, F., A. Lopez-Pita, and M. Velez. 2003. "Optimal social speed limits in a highway: Suitability of the 120 km/h speed limit in Barcelona, Spain." *Actes de la European Transport Conference.* Estrasburg.

Samerei, S. A., K. Aghabayk, A. Mohammadi, and N. Shiwakoti. 2021. "Data mining approach to model bus crash severity in Australia." *Journal of Safety Research* 76: 73-82.

Samerei, S. A., K. Aghabayk, N. Shiwakoti, and S. Karimi. 2021. "Modelling bus-pedestrian crash severity in the state of Victoria, Australia." *International Journal of Injury Control and Safety Promotion* 28 (2): 233-242.

Santamariña-Rubio, E., K. Pérez, M. Olabarria, and A. M. Novoa. 2014. "Gender differences in road traffic injury rate using time travelled as ameasure of exposure." *Accident Analysis and Prevention* 65: 1-7.

SAS. 2013. "Chapter 79 The PROBIT Procedure." In *SAS/STAT® 13.1 User's Guide*. North Carolina: SAS.

Servei Català de Trànsit. 2014. *Strategic Road Safety Plan for Catalonia 2014-2020.* Generalitat de Catalunya.

Servei catala de Transit. 2019. *Pla Local de Seguretat Vıària 2019-2022.* Barcelona: Ajuntament de Barcelona.

Traffic General Directorate. 2011. *Spanish Road Safety Strategy 2011-2020. Executive Summary.* Traffic General Directorate.

Vorko-Jović, A., J. Kern, and Z. Biloglav. 2006. "Risk factors in urban road traffic accidents." *Journal of Safety Research* 37 (1): 93-98.

Yu, R., and M. Abdel-Aty. 2014. "Using hierarchical Bayesian binary probit models to analyze crashinjury severity on high speed facilities with real-time traffic data." *Accident Analysis & Prevention* 62: 161–167.

Zhang, G., K. K.W. Yau, and G. Chen. 2013. "Risk factors associated with traffic violations and accident severity in China." *Accident Analysis & Prevention* 59: 18-25.

Zhou, Min, and Virginia P. Sisiopiku. 1997. "Relationship Between Volume-to-Capacity Ratios and Accident Rates." *Journal of the Transportation Research Board* 1581 (1): 47-52.

3. Working hours and traffic accident injuries: Case study in Barcelona

3.1. Abstract

Working hours in Spain is generally last for 8 hours per day with a maximum of 40 hours per week. Working hours, therefore, can impact the traffic flow and characteristics due to the intensity of road usage by different road users. On the other side, traffic accidents can also be impacted by several temporal factors that may lead to a higher number of traffic accidents or may increase the level of injury alongside other risk factors. In this study, these timings are examined to comprehend the influence of the temporal factors represented by the working hours scheme on traffic accident injuries in Barcelona. Another temporal factor which is the season of the year is included to provide a wider and clearer image for the conducted results and the current situation. The data is collected from the open data service provided by the City Hall of Barcelona. Data preparation and segregation to include the categories of working hours that may lead to a different level of injury resulted from a traffic accident is, firstly, carried. Then, a machine learning model is applied to classify the correlations for both the temporal factors and traffic accidents. Eventually, a Tree Augmented Naïve Bayes model is applied. The results showed that both working hours timing and summer season have higher probabilities of having traffic accidents slight injuries with different medical care assistance provided compared to other timings.

Keywords: Working hours; Traffic accidents; Bayesian network; Barcelona

3.2. Introduction

Spain has set a comprehensive labor law to organize and regulate the whole working employees' and employers' relationships with covering other aspects including health and safety at work, Social Security, the procedural law, and special relationships conditions for employment de Vivero (2019). As part of these regulations and rules, the maximum number of hours per week has been set to be 40 hours which is estimated based on an average over an annual period. Generally, the working time periods each day is set to start at 9:00 a.m. and end at 7:00 p.m. with having two hours break during the working period. This break period usually starts at 2:00 p.m. and ends at 4:00 p.m. for most employees. These timings, indeed, can be a reason for traffic congestion or flow intensity on different roads due to the fact that several groups of people are heading to different destinations at the same time throughout the day.

Real-time traffic data can be affected by different temporal factors similar to the working hours timings and other temporal factors besides the spatial factors. Several studies have examined the effect of the temporal factors on traffic real-time data. In some European countries, holidays' periods shown lower traffic compared to other seasons Stathopoulos and Karlaftis (2001). Peak hours have shown to be another influential factor related to traffic flow. Based on a study that is conducted to examine traffic flow patterns in Shanghai, China Yang et al. (2019), the results showed that peak hours had worse traffic congestion compared to non-peak hours.

Traffic accidents, on the other hand, can also be correlated with different temporal factors that can increase the occurrence of them or affect the level of injury. A study Kashani and Zandi (2020) shown that weekends can be correlated with a higher number of traffic accidents compared to weekdays. The same study also revealed that hot timing had more accidents during weekdays. These timings were found to be 8 a.m. and 2 p.m. While for weekends, these timings differ from weekdays' hot timings. 1 p.m., 8 p.m., and 10 p.m. were the hot timings for accidents during weekends. Slight injuries accidents were found to be more likely to happen when the traffic flow is high (Quddus et al., 2010). Similar to traffic flow, traffic accidents were found to vary according to the season of the year (Harirforoush, 2017) (Le et al., 2019). Morning, afternoon, or late-night timings were found to have lower probabilities of having fatal injuries compared to early morning timing in Cartagena, Colombia (Cantillo et al. 2020). Consistently, time was found to be an important independent variable related to traffic accidents severities (Li et al., 2020). As mentioned earlier, traffic congestion can be highly affected by the timing of the day. Therefore, a study (Hyodo and Todoroki, 2018) showed congested and the mixed flow state can increase the risk of having traffic accidents different types including property damages and slight injury accidents.

The objective of conducting this study is to examine the influence of certain temporal factors represented by the working hours and the season of the year in Barcelona, Spain. Data collected and preparation process is carried. Then, a Bayesian network is applied to the exploited data to extract the results.

3.3. Methodology

3.3.1. Data description

The data the is exploited in this study is collected from the Barcelona open data service which is called Barcelona's City Hall Open Data Service (Ajuntament de Barcelona's open data service, 2019). The data that is gathered consists of different traffic injuries. In this study, slight injury levels with different categories are selected for this purpose that occurred in Barcelona in 2019. Two temporal factors are selected for the objective of this study including the working hours timing and the season of the year. Beginning with the season potential temporal risk factor, the four seasons are included in the examined period and categorized based on their climate The main climate data of Barcelona (2020). The summer season is considered to start in May and end in August. Followed by the autumn season that begins in September and ends in November. The winter season is considered to start in December and end in February. Spring, therefore, is considered to start in March and end in April. For the working hours, two categories are included: during work and the other rest time. Working hours are considered to start at 9 a.m. and end at 7 p.m. with having two hours break. These two hours are considered to start at 2 p.m. and end at 4 p.m. with considering them with the other rest category when the initiation of the analysis part is established. However, during the data analysis, this time-period is considered to start from 2:01 p.m. to 4 p.m.

These previously mentioned categorizations are for the independent variables, while for the dependent variable, three categories are included. The first category is represented by the person who had a slight injury with medical assistance. The second category is represented by the person who had a slight injury but rejected health care. The third and last category is represented by the person who had a slight injury with having a hospitalization up to 24 hours. Table 4 is displaying the two different independent variables' general statistics. The total number for both predictors is 11620 accidents that occurred in 2019 in Barcelona and belonged to the slight injuries that required different levels of medical care. As mentioned earlier, the working hours variable consists of two categories, while the season has three categories. Both predictors are considered as categorical type variables alongside the dependent variable when the analysis part is carried.

Table 4 Main independent variables statistics

	Count	Mean	Min	Max	Range	Variance	Standard Deviation	Standard Error of Mean
Working hours	11620	1.55	1	2	1	0.247	0.497	0.005
Season	11620	2.68	1	4	3	1.405	1.185	0.011

3.3.2. Bayesian network

Bayesian network is a member of probabilistic graphical models (GM)s that provides laconic descriptions of the distribution of joint probability for the given random variables. Two methods can be exploited when applying Bayesian network models when using IBM Watson Studio software platform with utilizing SPSS modeler. Tree Augmented Naïve Bayes and Markov Blanket estimation are both methods that can be utilized. In this study, Tree Augmented Naïve Bayes is implemented to classify the correlation between the two predictors and the level of medical care that is provided for persons that are involved in traffic accidents. The reason for choosing Tree Augmented Naïve Bayes is for its simplicity as the number of variables is only two and the aim of this study is only to examine the correlation with considering the prediction part. Moreover, the two independent variables are classified under the same category which is the temporal variables category. The classifier of Tree Augmented Naïve Bayes that is exploited through the IBM software platform is based on this book (Friedman et al., 1997). The conditional probabilities are calculated, in general, by SPSS modeler as follows:

$$Pr\left(Y_i|X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j\right) = \frac{Pr(Y_i)\left(X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j|Y_i\right)}{Pr\left(X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j\right)}$$

$$\propto \Pr(Y_i) \prod_{k=1}^{n} Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right) \tag{1}$$

For $d_j = \left(x_1^j, x_2^j, \ldots, x_n^j\right)$ where d is the data set. d_j is the case that is classified to which it belongs to i^{th} target category, which in this study, the target Y_i is the medical care three levels provided for slightly injured persons. x is the predictor. n is the number of predictors which is two in this study. K is the number of non-redundant parameters. π_k is the parent set of the independent variable alongside the dependent variable, it maybe empty for Tree Augmented Naïve Bayes. The conditional probability is $Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right)$ that is associated with each node, which in this study, there are two nodes for predictors.

3.4. Results and discussion

The structure of the applied Tree Augmented Naïve Bayes structure consists of three nodes including medical care level, season, and time which is the working hours as shown in figure 2. The parent node for the season independent variable is only the medical care level. For the time node, this node is linked to two nodes including medical care level and the season. This makes sense since the fact that working hours are already part of the season of the year. Time has a higher importance value compared to season based on the applied model.

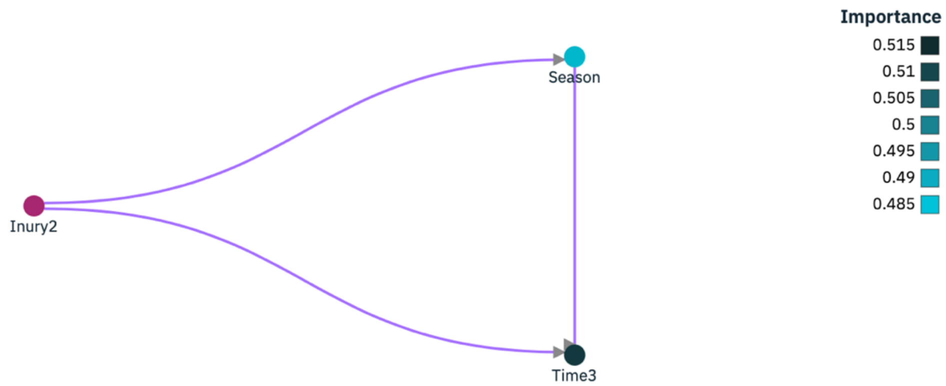


Figure 2 Tree Augmented Naïve Bayes structure

Table 5 and table 6 are both displaying the estimated conditional probabilities based on the applied Tree Augmented Naïve Bayes for season and time, respectively. For the season, summer has a higher conditional probability compared to all other seasons of the year, followed by autumn, winter, and lastly, the spring season. For the time variable, the working hours category has the highest conditional probabilities during all seasons for all medical care categories that are provided for the slightly injured person except spring season when a person got a slight injury, but this injured person rejected medical assistance.

Table 5 Season conditional probability

Parents		Season Probability			
Medical care level	Autumn	Winter	Spring	Summer	Total
With medical assistance	0.245	0.233	0.180	0.341	2,263
Rejected health care	0.243	0.231	0.188	0.338	420
Hospitalization up to 24 hours	0.255	0.235	0.164	0.346	5,428

Table 6 Time conditional probability

Parents		Time Probability		
Season	Medical care level	Working hours	Rest of the day	Total
Autumn	With medical assistance	0.514	0.486	555
Autumn	Rejected health care	0.520	0.480	102
Autumn	Hospitalization up to 24 hours	0.525	0.475	1,384
Spring	With medical assistance	0.576	0.424	408
Spring	Rejected health care	0.468	0.532	79
Spring	Hospitalization up to 24 hours	0.552	0.448	888
Winter	With medical assistance	0.547	0.453	528
Winter	Rejected health care	0.619	0.381	97
Winter	Hospitalization up to 24 hours	0.558	0.442	1,277
Summer	With medical assistance	0.585	0.415	772
Summer	Rejected health care	0.549	0.451	142
Summer	Hospitalization up to 24 hours	0.559	0.441	1,879

3.5. Conclusions

The fact that there are enormous factors that influence traffic accident occurrences and severities is leading to conducting several studies to examine these potential risk factors. Temporal factors are part of these factors that can impact traffic accidents. Working hours and the season of the year are part of these temporal factors that may lead to traffic accidents. Therefore, this study has exploited traffic accident injuries by focusing on the level of medical care that is provided for the slightly injured person in Barcelona in the year 2019.

Tree Augmented Naïve Bayes based on Bayesian network is employed to classify the correlations between the two predictors and the dependent variable. Four seasons are included to understand its temporal impact on the three levels of medical care. Two timings are included the working hours which starts at 9 a.m. and ends at 7 p.m. with having 2 hours break that is not included in this category which is from 2 p.m. to 4 p.m. period. The results show that the summer season has higher conditional probabilities of having slight injuries with including different medical care assistance compared to other seasons. For the timing, the working hours period, similar to the summer season, has the highest conditional probabilities for traffic accidents with being involved in slight injuries with different medical care assistance. For the future work and based on the concluded results, data from delivery operating firms is needed to grasp the impact of these different timings on traffic accidents occurring while the operation is maintained. Then, similar data analysis can be carried out to detect the impact of working hours on this category of employees as they may have a higher risk of being involved in traffic accidents compared to other categories during these timings. Other levels of injury resulted from traffic accidents may also be considered in the future work.

3.6. References

Ajuntament de Barcelona's open data service. (2019). Retrieved from https://opendata-ajuntament.barcelona.cat/en/

Cantillo, V., Márquez, L., and Díaz, C. (2020). An exploratory analysis of factors associated with traffic crashes severity in Cartagena, Colombia. *Accident Analysis & Prevention, 146*.

de Vivero, S. (2019). *Employment law overview Spain 2019-2020.* Barcelona: L&E Global.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning, 29*, 131–163.

Harirforoush, H. (2017). *Thesis: An Integrated GIS-Based and Spatiotemporal Analysis of Traffic Accidents: A Case Study in Sherbrooke.* Université de Sherbrooke.

Hyodo, S., and Todoroki, T. (2018). An Analysis of Risk Factors for Rear-Ender Accident on Urban Expressway Considering Accident Severity. *Transportation Research Procedia, 34*, 203-210.

Kashani, A., and Zandi , K. (2020). Influence of Traffic Parameters on the Temporal Distribution of Crashes. *Transportation Engineering, 24*, 954–961.

Le, K., Liu, P., and Lin, L.-T. (2019). Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. *Geo-spatial Information Science, 23*(2), 153-164.

Li, L., Prato, C., and Wang, Y. (2020). Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: A sequential approach of multivariate imputation by chained equations and random forest classifier. *Accident Analysis & Prevention, 146*.

Quddus, M., Wang, C., and Ison, S. (2010). Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. *Journal of Transportation Engineering, 136*(5), 424-435.

Stathopoulos, A., and Karlaftis, M. (2001). Temporal and Spatial Variations of Real-Time Traffic Data in Urban Areas. *Transportation Research Record: Journal of the Transportation Research Board, 1768*(1), 135-140.

The main climate data of Barcelona. (2020). Retrieved from https://www.barcelona.de/en/barcelona-climate-travel-season.html

Yang, S., Wu, J., Xu, Y., and Yang, T. (2019). Revealing heterogeneous spatiotemporal traffic flow patterns of urban road network via tensor decomposition-based clustering approach. *Physica A: Statistical Mechanics and its Applications, 526*,120688.

Intentionally blank page

4. COVID-19 pandemic effects on traffic crash patterns and injuries in Barcelona, Spain: An interpretable approach

4.1. Abstract

The COVID-19 pandemic has certainly affected our lives in many different aspects. Traffic, in general, and traffic crashes, in particular, have both been affected by the pandemic. This study, therefore, is conducted to examine the influence of the pandemic on traffic crashes different levels of injuries in Barcelona, Spain. Time-series analysis and forecast are both exploited to achieve the objective of conducting this study. The results show a dramatic drop in traffic crashes, especially during the state of alarm duration that was imposed in March 2020. All levels of injuries that resulted from traffic crashes have shown approximately 39 %, 30%, and 36% reduction for slight, severe, and fatal injuries, respectively. Additionally, the impact of certain chosen risk factors is found to be the same after employing two exhaustive chi-square automatic interaction (CHAID) models. Male injured persons and the time of the day represented by evening and night timing are found to have higher probabilities of having severe or fatal injuries both before and during the pandemic periods. The two applied exhaustive CHAID models have proven their ability to provide interpretable results.

***Keywords*:** COVID-19, Traffic crashes, Injuries, Time-series, Forecast, CHAID, Barcelona, Interpretability

4.2. Introduction

Corona Virus Disease (COVID-19) which is linked to severe acute respiratory syndrome (SARS) and some types of the common cold is a new strain of coronavirus that is spread in late 2019 (UNICEF; WHO; IFRC 2020). This virus can be transmitted from one person to another when respiratory droplets of an infected person are disseminated or being contacted by another person. Eventually, the COVID-19 outbreak has changed into a COVID-19 pandemic to change the implemented strategies that deal with the virus (The Lancet Infectious Diseases, 2020). One of the measures that have been implemented is imposing a curfew in different areas and regions to contain the number of infected cases (Koh, 2020). Until now, the unprecedented number of confirmed cases is exceeding 548 million cases globally, while the number of deaths is exceeding 6 million deaths (World Health Organization, 2022). Based on the European Centre for Disease Prevention and Control (ECDC, 2022), more than 230 million cases are confirmed in Europe, with more than 2 million deaths.

This pandemic has, for sure, influenced our life from different perspectives alongside the different measures that were imposed and implemented that may also have changed our way of living. A research study (Zhang et al., 2020) found that 25 % of people who answered a conducted survey stopped working. This category had shown worse mental and physical health compared to other categories. Consistently, a 31.7 % increase in suicide was captured in Hong Kong among people who were 65 years and older during the pandemic (Solomon 2020). Conforming to the previously mentioned situation, a study (de Pedraza and Guzi, 2020) found that the increasing number of COVID-19 cases had increased the levels of dissatisfaction and anxiety.

Moving from the mental and physical health aspect, traffic has also been influenced by the pandemic globally due to the measures that were imposed and the cautions from people to avoid unnecessary social connections. The impact of COVID-19 on traffic was found in South Korea related to the number of vehicles that were utilized during the pandemic. The study (Lee et al., 2020) found that the number of vehicles was 9.7 % lower than in 2019. However, the results also showed that the number of vehicles started to increase again as the number of patients decreased. For the environmental aspect related to traffic, a notable decrease was recognized for seven harmful gases during the pandemic in a Northwestern United States city (Xiang et al., 2020). In Kazakhstan, benzene and toluene concentrations were found to be 2 to 3 times higher during the same seasons in 2015-2019 compared to the COVID-19 pandemic period as an impact of the traffic-free urban condition (Kerimray et al., 2020). Another COVID-19 restriction effect is that vehicle miles and driving days were both reduced among adolescents (Stavrinos et al., 2020).

Traffic crashes are part of the other aspects of the other categories that have been intensively and passively affected by the pandemic due to the series of different occurred events worldwide. Curfew is one of these events that has been implemented as a measure for restricting unnecessary movement and lessening the number of confirmed cases. Considering the bright side of the pandemic, 17,600 injuries and 200 fatalities as a result of traffic crashes were avoided in Turkey when people stayed at their homes during past March and April in 2020 compared to the previous same periods (Oguzoglu, 2020). In Tarragona, Spain, the number of traffic crashes drastically decreased during the period between March and April during the pandemic (Saladié et al., 2020). Another study (Nuñez et al., 2020) examined the Emergency

Traumatology Service data at a tertiary hospital within the Spanish National Health System. The results revealed that there was a reduction in the number of traffic crash case visits besides other types of crashes during the pandemic. A huge drop in the number of traffic crash fatalities was noticed in Peru as the number of fatalities decreased by 12.22 million men per month (Calderon-Anyosa and Kaufman, 2021). In Louisiana, the United States, a study (Barnes et al. 2020) found a large reduction in the number of traffic crashes, injuries resulting from traffic crashes, and distracted drivers. The findings also found that the age category for persons who were involved in incidents during the pandemic was between 25 to 64. During the lockdown in Telangana, India, a drop of 77.9% in the cases of trauma that is related to traffic crashes was noticed (Maryada et al., 2020). However, there was no detected change in fragility fracture incidence cases for the elderly during the lockdown. A reduction in traffic crashes by 68% was found for the category that represented people who were older than 60 years old (Shen et al. 2021).

Despite the reduction in crashes (Sakelliadis et al., 2020) and traffic volumes during the pandemic in Greece, there was an increase in speeding, harsh acceleration and braking, and using the mobile phone (Katrakazas et al., 2020). Another research (Hu et al. 2021) detected a minor impact of stay-at-home restrictions on human mobility, whereas the same research found an increase in mobility after re-opening. Similarly, drivers were found to have a higher tendency to commit speed traffic violations that were involved in fatal crashes during the lockdown second month (Inada et al., 2021). Following the same trend, the number of traffic crash injuries, motor vehicle fatalities, and speeding violations increased during the pandemic (Meyer 2020). A similar case was found in Malaysia as the number of traffic crashes was high despite the decrease in the volume of traffic (Prasetijo et al., 2021). In Missouri, United States, the number of severe or fatal injuries that resulted from traffic crashes had no significant change (Qureshi et al., 2020). In contradiction, a research study (Muley et al., 2021) found a decrease in traffic violations during complete movement restriction in Qatar with more than 70%. In addition, the study showed there was a reduction in traffic crashes that accompanied the imposed preventive measures.

The main objective of this study is to examine the impact of the COVID-19 pandemic period on traffic crash injuries and patterns in Barcelona, Spain. For this reason, different time intervals are exploited alongside different techniques. Time series analysis is carried out to reveal traffic crash injuries during the selected time interval including the pandemic and then compare it with previous traffic crash injury occurrences. Additionally, a forecast model represented by a simple seasonal exponential smoothing technique is employed to compare the actual injuries with predictions besides detecting abnormal cases. Lastly, two identical supervised learning techniques represented by two exhaustive chi-square automatic interaction detector (CHAID) models are employed to compare the impact of chosen risk factors on traffic crash injuries during the pandemic and previous years. Using models with inherent interpretability is a crucial factor to gain experts' trust in such algorithms. The results can help road authorities in detecting and observing injuries during the pandemic in order to decide whether new rules should be implemented to confront unusual traffic crash cases besides determining the correlations between several risk factors and traffic crash injuries. The remainder of this chapter is organized as follows; the methodology section is set to be section two. The results are presented in the third section. The conclusions are presented in the last section of this chapter.

4.3. Methodology

4.3.1. Data description

Injuries that are resulted from traffic crashes are, indeed, representing a challenge worldwide. During the COVID-19 pandemic, several changes are noticed regarding traffic crashes, in general, and the injuries that are resulted from traffic crashes, as explained before. In Barcelona, Spain, similar to the rest of the world, this phenomenon, is still under examination. The Open Data Service of Barcelona's City Hall (Barcelona's City Hall Open Data Service, 2020) is providing the service of public access to the data related to traffic crashes alongside other types of data. Eventually, the spotted data that is required for carrying out the analysis to conduct this study is gathered from this open-access database. Mainly the data utilized in this study is the same for all different types of analysis that are carried out, however, the differences in choosing which factors are studied.

As shown in table 7, the gender predictor has two categories male and female. The age variable has five different classifications including 0-15 years old group, 16-24 years old group, 25-40 years old group, 41-64 years old group, and lastly 65 years old and older injured persons. The time of the day has three categories including morning (6 a.m. to 1 p.m.), evening (2 p.m. to 9 p.m.), and night timing (9 p.m. to 5 a.m.). This time classification is based on the collected data when the crash occurred. The injured person variable has pedestrian, passenger, and driver categories. The reasons for choosing these factors are based on previously conducted studies related to traffic crashes as these risk factors have already shown a significant impact on traffic crash injuries (Aiash and Robusté, 2021). For the first exhaustive CHAID tree, data from 2016-2019 is utilized. 2020 data is used to represent the COVID-19 pandemic. All blank categories are eliminated from both data sets. The data shown in table 7 is only utilized for the CHAID tree models. For the simple seasonal exponential smoothing, all data sets without eliminating blank categories for four years that occurred before 2020 including the years 2016, 2017, 2018, and 2019 are utilized as this part of the paper is only focusing on the date with only focusing on severe injury cases. Eventually, 870 severe injury cases are included for this purpose.

Table 7 Injuries during different years related to chosen risk factors

Risk factor	Category	2020 Severe or fatal injury	2020 Slight injury	2016-2019 Severe or fatal injury	2016-2019 Slight injury
Time of the day	Morning	27	482	488	23274
	Evening	36	1616	160	5215
	Night	94	2319	291	17725
Gender	Male	111	4270	634	28058
	Female	46	147	305	18156
Age	0-15	1	121	25	1513
	16-24	18	644	121	6893
	25-40	53	1806	296	19160
	41-64	65	1609	359	15377
	65+	20	237	138	3271
Injured	Driver	37	299	269	4573
	Passenger	15	415	86	8989
	Pedestrian	105	3703	584	32652

4.3.2. Time series forecasting

In this study, simple seasonal exponential smoothing is exploited to predict the pandemic period with considering the traffic crashes severe injuries. Exponential smoothing is considered one of the most popular modules used in time series. This popularity can be referred to its simplicity, efficiency, and reasonable accuracy (Montgomery et al., 1990). Due to the nature of traffic crashes data, the seasonal type of exponential smoothing is selected. For the observation and time interval, four years are included to conduct the prediction estimations for 2020 including the years 2016, 2017, 2018, and 2019. IBM Watson Studio Cloud is exploited to apply the time series model that provides the aimed forecast model. While applying the forecast model, several settings are set to achieve accurate results. Firstly, the time interval is estimated per day assuming the weekday first day is Monday with an increment value of 1. For the missing value, the linear interpolation replacement method is chosen with the lowest data quality of 5 %. The forecast option is utilized to extend the prediction estimation for the 2020 period based on the previously occurred severe injuries during the four years period. The number of extended future days is chosen as 365 days to cover the entire 2020 year and conduct the aimed comparison between the predictions and the actual number of severe injuries that occurred during the 2020 year. The IBM time series algorithm is designed by IBM with the help of Ruey Tsay at The University of Chicago. More details can be found regarding time series algorithms (Alfons et al., 2013) (Box et al., 1994) (Brockwell and Davis, 1991) (Gardner, 1985) (Harvey, 1989) (Makridakis et al., 1997) (Melard, 1984) (Pena et al., 2001).

After the previous setups are chosen, the model is conducted, and results are displayed and determined. A simple seasonal exponential smoothing method is applied. This method has level and season parameters. $Y(t = 1,2, \dots, n)$ which is the univariate time series under investigation. The total number of observations is n. $L(t)$ is the level parameters for the applied model. The seasonal length is S. $S(t)$ is the season parameter. $\hat{Y}_t(k)$ is the model estimated k-step ahead forecast at time t for Y series. α is the weight of level smoothing. δ is the trend smoothing weight. Formulas 1, 2, and 3 are presented to grasp the overall simple seasonal exponential smoothing algorithm estimations.

$$\mathrm{L(t)} = \begin{cases} \big(\mathrm{Y(t)} - \mathrm{S(t-s)}\big) + (1-\alpha)\mathrm{L(t-1)}, & \text{if Y(t) is not missing} \\ \mathrm{L(t-1)}, & \text{else} \end{cases} \tag{1}$$

$$\mathrm{S(t)} = \begin{cases} \delta\big(\mathrm{Y(t)} - \mathrm{L(t)}\big) + (1-\delta)\mathrm{S(t-s)}, & \text{if Y(t) not missing} \\ \mathrm{S(t-s)}, & \text{else} \end{cases} \tag{2}$$

$$\hat{\mathrm{Y}}_{\mathrm{t}}(\mathrm{k}) = \mathrm{L(t)} + \mathrm{S(t+k-s)} \tag{3}$$

Twelve measures are exploited and depicted to present the forecast model goodness of fit for severe injury category including equations from 4 to 15. Mean squared error (MSE), root mean squared deviation (RMSE), root mean squared prediction error (RMSPE), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), Akaike information criteria (AIC), Bayesian information criteria (BIC), R-squared(R^2), stationary R-squared ($\mathrm{R}_{\mathrm{s}}^2$), and Ljung-Box (Q) are all determined to estimate the goodness of fit for the applied time-series forecast model. $Z(t)$ is $Y(t)$ or transformed $Y(t)$. $\overline{\Delta Z}$ is the simple mean model for the differenced transformed series.

$$MSE = \frac{\sum(Y(t)-\hat{Y}(t))^2}{n-k} \tag{4}$$

$$RMSE = \sqrt{MSE} \tag{5}$$

$$MAE = \frac{1}{n}\sum\left|Y(t) - \hat{Y}(t)\right| \tag{6}$$

$$MAPE = \frac{1}{n}\sum\left|\frac{(Y(t)-\hat{Y}(t))}{Y(t)}\right| \tag{7}$$

$$RMSPE = \sqrt{\frac{100}{n}\sum\left(\left|\frac{(Y(t)-\hat{Y}(t))}{Y(t)}\right|\right)^2} \tag{8}$$

$$MaxAE = \max\left(\left|Y(t) - \hat{Y}(t)\right|\right) \tag{9}$$

$$MaxAPE = 100 \max\left(\left|\frac{Y(t)-\hat{Y}(t)}{Y(t)}\right|\right) \tag{10}$$

$$AIC = n \times \ln\left(\frac{\sum(Y(t)-\hat{Y}(t))^2}{n}\right) + 2k \tag{11}$$

$$BIC = n \times \ln\left(\frac{\sum(Y(t)-\hat{Y}(t))^2}{n}\right) + k \times \ln(n) \tag{12}$$

$$R^2 = 1 - \frac{\sum(Y(t)-\hat{Y}(t))^2}{\sum(Y(t)-\bar{Y})^2} \tag{13}$$

$$R_s^2 = 1 - \frac{\sum_t(Z(t)-\hat{Z}(t))^2}{\sum_t(Z(t)-\overline{\Delta Z})^2} \tag{14}$$

$$Q(K) = n(n+2)\sum_{k=1}^{K}\frac{r_k^2}{(n-k)} \tag{15}$$

4.3.3. Risk factor analysis

Chi-square automatic interaction detector (CHAID) which lies under supervised learning techniques that belong to the decision tree (DT) technique that is normally a non-parametric technique is exploited in this study. This technique does not have any dependency on any functional form and no prior probabilistic knowledge requirement (Abellán et al., 2013) and is developed by (Kass, 1980). One of the main reasons for applying the CHAID tree is its ability to develop a simple chart to detect the correlations between the independent variables and the dependent variable compared to other techniques that cannot provide this easiness in interpretation. It simply acts as a decision tree with the exception of using chi-squared based criterion instead of using the information gain or gain ratio criteria. Additionally, it is considered an efficient statistical method for segmentation or tree growing. The outputs of the CHAID tree are highly meaningful and easy to interpret. This process keeps repeating till the fully grown tree is formed. Part of the advantages of applying the CHAID tree is its ability in producing more than two categories as the CHAID tree is not a binary tree technique. Therefore, it can work with different types of variables and can create a wider tree than binary tree techniques. However, in this study, exhaustive CHAID is employed as it is identical to normal CHAID in the statistical tests but much safer in finding useful splits as it can continue to merge categories of independent variables until two super categories are left. Thus, the best split for each independent variable is found and then compare based on the p-values to choose which independent variable to split. Exhaustive CHAID is basically a modification of CHAID that is developed by (Biggs et al., 1991).

By following a previously conducted study that employed the CHAID tree (Aiash and Robusté 2021), exhaustive CHAID (AlKheder et al., 2020) and utilized IBM Watson with using SPSS modeler set, 70% is set for the training set and 30 % for is set for the testing set. Pearson chi-square statistic and its corresponding p-values are calculated because the injury level is a categorical variable as shown in equations 16 and 17, respectively. X_d^2 follows $d = (J-1)(I-1)$ a chi-square distribution degree of freedom. The independent variable is X, and the dependent variable is Y. n_{ij} is the observed cell frequency. $\widehat{m}_{ij}$ is the expected cell frequency $(x_n = i\ , y_n = j)$. During the merge process, categories are compared to consider only the records that belong to the comparison categories. During the split step, all the current nodes are used as the categories are considered in p-value calculations. The likelihood ratio chi-square and its p-value is calculated in equation 18, 19, and 20, respectively based on the observed and expected frequencies. The expected frequencies are calculated based on two cases. The expected frequencies with no case weights are calculated as shown in equations 21 and 22. The expected frequencies with specified weights are calculated in equations 23 and 24. The estimated parameters are α_i and β_i. Further details related to CHAID algorithms are mentioned in the IBM guide (IBM, 2016).

$$n_{ij} = \sum_{n} f_n I(x_n = i \wedge y_n = j) \tag{16}$$

$$X^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I}\frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \tag{17}$$

$$p = \Pr\left(X_d^2 > X^2\right) \tag{18}$$

$$G^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij} \ln\left(\frac{n_{ij}}{\hat{m}_{ij}}\right) \tag{19}$$

$$p = \Pr\left(X_d^2 > G^2\right) \tag{20}$$

$$\hat{m}_{ij} = \frac{n_{i.} n_{.j}}{n_{..}} \tag{21}$$

Where

$$n_{i.} = \sum_{j=1}^{J} n_{ij}, \quad n_{.j} = \sum_{i=1}^{I} n_{ij}, \; n_{..} = \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}. \tag{22}$$

$$\hat{m}_{ij} = \bar{w}_{ij}^{-1}\alpha_i\beta_i \tag{23}$$

Where

$$\bar{w}_{ij} = \frac{w_{ij}}{n_{ij}}, \qquad w_{ij} = \sum\nolimits_{n\in D} w_n f_n I\,(x_n = i \wedge y_n = j) \tag{24}$$

4.4. Results and discussion

4.4.1. Injury analysis

To understand the traffic crash injury occurrences during the pandemic in Barcelona, a radar chart is prepared to depict the variations that happened as shown in figure 3. As the pandemic happened in 2020, both years 2019 and 2020 are included in this chart for comparison. The first confirmed case related to COVID-19 was identified in Spain in late January 2020, followed by declaring the state of alarm on the 14th of March, and then ended on the 21st of June (Henríquez et al., 2020). This can be obviously noticed in figure 1. The total number of traffic crash injuries that consists of the three levels of injuries is included. During January and February, both months have similar total numbers for both years. Then, the number of traffic crash injuries is drastically dropped during the following months.

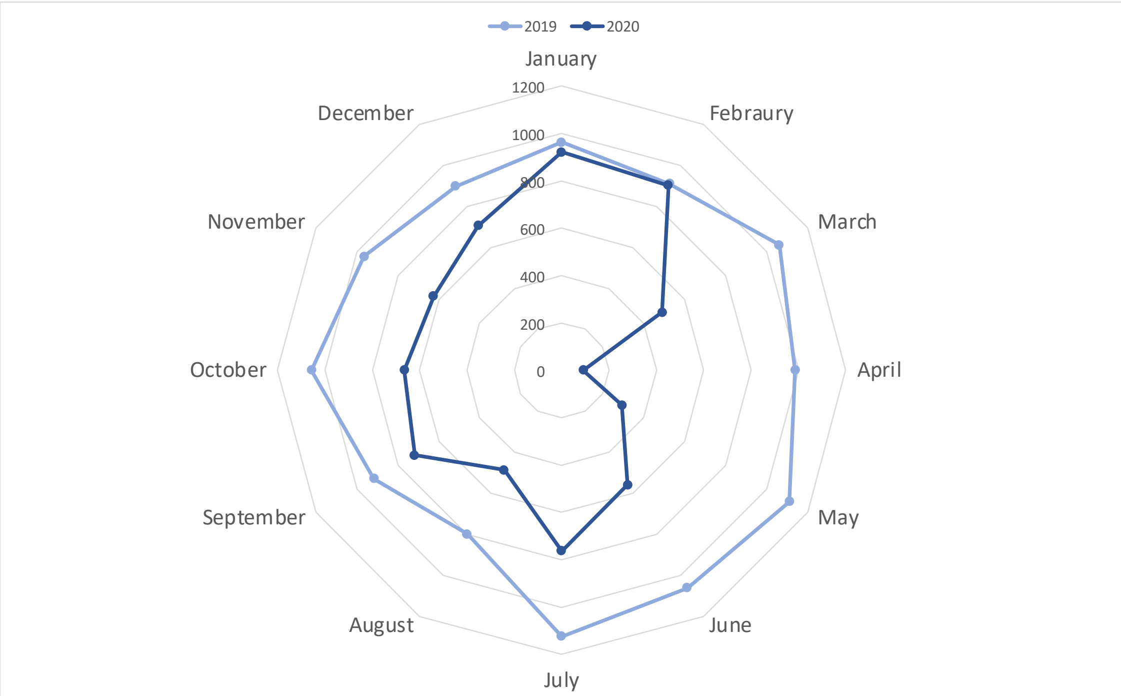


Figure 3 Traffic crashes injuries during 2019 and 2020

For time series, time decomposition analysis is exploited to examine the differences between the two periods before and after the pandemic. The times series analysis is separated based on the resulting level of injury. The slight injury is presented in figure 4. As shown, the number of traffic crashes is almost similar till the beginning of March 2020, when the numbers started to drastically drop as the state of alarm is set on the 14th of March 2020. Then, the numbers started to increase again in July of the same year. For the total slight injuries, 2019 had 11620 injuries, and 2020 had only 7059. For severe injuries that are shown in figure 5, the trend was different.

Indeed, fewer numbers of severe injuries can be noticeably detected during the same period when the state of alarm was set. The total number of severe injuries was 202 in 2019, while in 2020, severe injuries number was 147. However, in September 2020, there was an abnormal increase in severe injury cases per day compared to 2019. There were 4 severe injuries that occurred on the 30th of September 2020. This number was not found on any day in September 2019. This abnormal trend is also detected in October with more reported severe cases. Similarly, in December 2020, there were 5 severe injuries in only one day which was on the 12th of December 2020, which is higher than any day compared to 2019 for the same month. Figure 6 is depicting the number of fatal injuries that are reported during 2019 and 2020. One of the noticeable observations is that the number of fatalities was highly affected by the state of alarm and the imposed restrictions as these fatalities cases were almost zero during that entire period. The total number of fatal injuries in 2019 was 22 deaths, while in 2020, this number was 14.

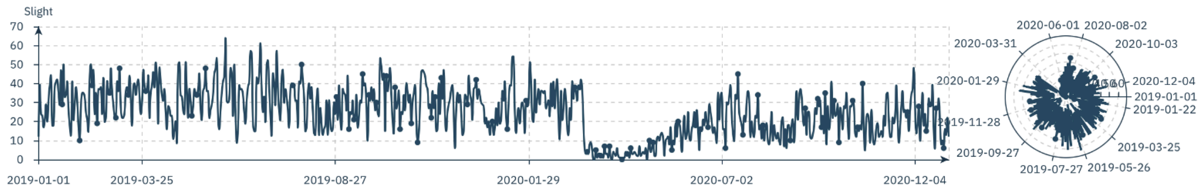


Figure 4 Slight injury time series analysis during 2019 and 2020

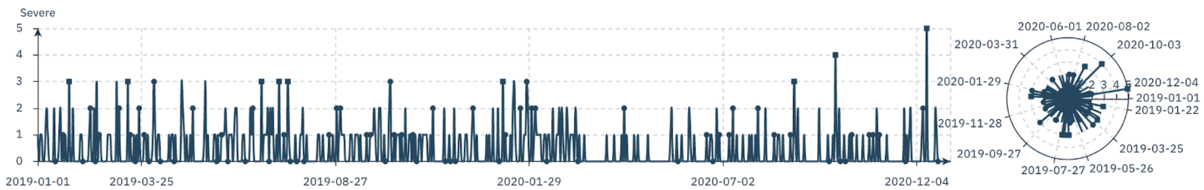


Figure 5 Severe injury time series analysis during 2019 and 2020

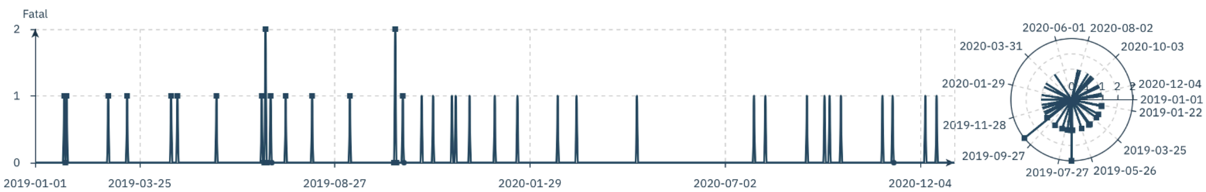


Figure 6 Fatal injury time series analysis during 2019 and 2020

The number of severe injuries that are found in unusual cases compared to previous years is recalling the examination of traffic volume during those specific durations to determine whether traffic volume can be the reason for these cases. Therefore, traffic volumes for both September and December are both included and depicted in figures 7 and 8. The data and figures are both gathered and brought from the TomTom data report for Barcelona (TomTom, 2020) based on data availability. The average traffic volume on working days for both 2019 and 2020 is shown. It can be concluded that the traffic volume has indeed decreased in 2020 compared to 2019 during both months. However, after the peak hours that have the highest traffic volume in both years, the average traffic volume gap between 2019 and 2020 is almost straightened.

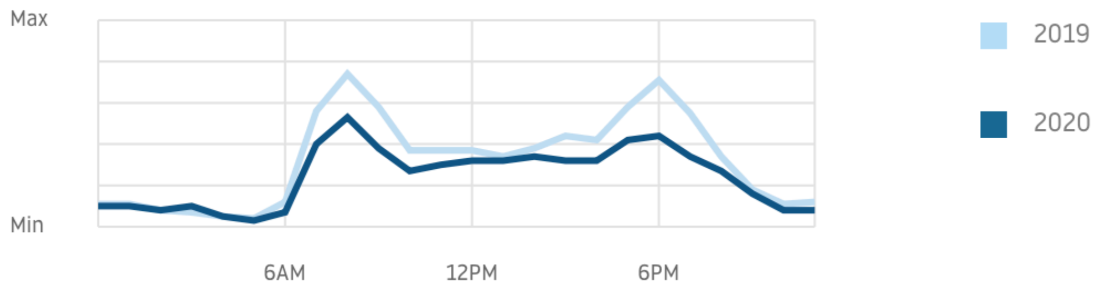


Figure 7 Average daily traffic during workdays in September based on (TomTom, 2020)

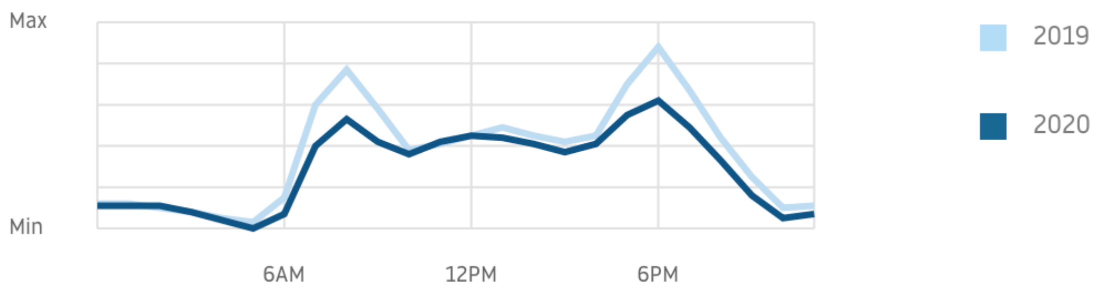


Figure 8 Average daily traffic during workdays in December based on (TomTom, 2020)

For the same reason of unusual severe injury cases, the overall statistics for several characteristics are presented to grasp the correlation between the pandemic and different aspects. The data for these aspects are gathered from the above-mentioned Open Data Service of Barcelona's City Hall. Part of these characteristics is the total number of registered vehicles. Figure 9 is depicting the total number of vehicles that are registered in the city of Barcelona. The total number of vehicles for the 2019 year is compared with the 2020 year. As shown, during the state of alarm the number of registered vehicles is, obviously, dramatically dropped especially in April. However, this trend has changed as the number started to increase again reaching a higher total number of vehicles compared to the previous year. Then, the spike decreased again to almost equalize the previous year's numbers. Another aspect that is examined is the reasons for traffic crash injuries for both 2019 and 2020 which are presented in figure 10. Alcohol, a road in poor condition, drugs or medicine, overspeeding, meteorological factors, and objects or animals on the road are all included and shown with their percentages that are determined based on the total number of reported reasons on the total number of crashes for both 2019 and 2020, individually. It can be shown that alcohol, a road in poor condition, overspeeding, and objects or animals on the road all have slightly increased in 2020 compared to 2019. It is worth mentioning that the low percentages for all included reasons are based on the reported reasons from the data source, as other factors are reported as "There is no mediated cause”. It can be concluded from those different overall statistics that the pandemic has influenced these aspects which may in turn affect the traffic crash injury cases and lead to this number of cases.

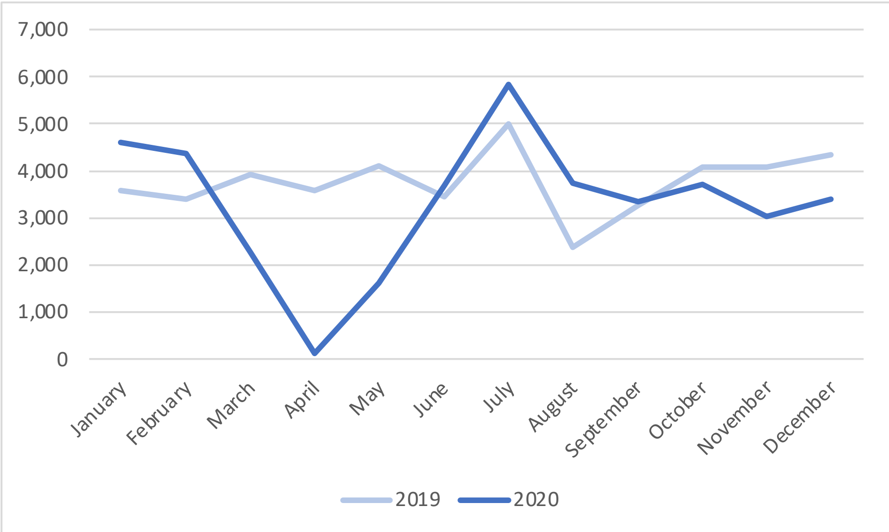


Figure 9 Number of vehicle registrations in Barcelona

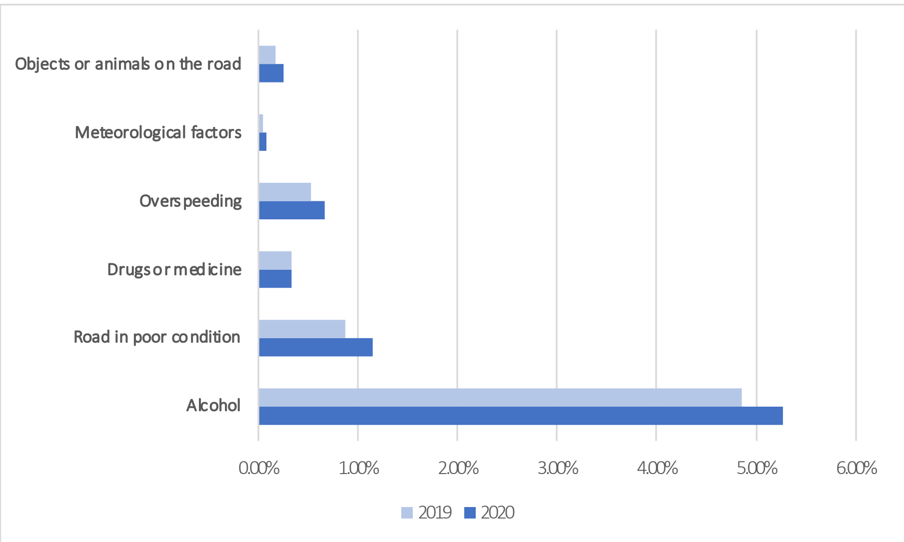


Figure 10 Percentages of several reasons for traffic crash injuries in Barcelona

4.4.2. Forecast model and goodness of fit

Table 8 is showing the results of the model goodness of fit measures. The results show an acceptable result for the carried measures regarding the goodness of fit for the applied model.

Table 9 is showing the results for the Ljung-Box. Df is the degree of freedom for the number of parameters that are free to vary when considering the severe injury target variable. Q(#) is the Ljung-Box statistics. The significance value is larger than 0.05 indicating that the model has random residual errors, which is a good indication for the model. In figure 11, the actual and upper forecasts are all presented alongside the actual crash reports that included only severe injuries. As observed previously, the number of severe injuries that occurred per day during September and December are also captured by the forecast as abnormal cases because their results exceeded the upper limits for the forecast.

Table 8 Forecast model fit for severe injury

MSE	0.729
RMSE	0.854
RMSPE	51.400
MAE	0.696
MAPE	48.338
MAXAE	5.377
MAXAPE	90.257
AIC	-459.223
BIC	-448.649
R-Squared	0.006
Stationary R-Squared	0.741

Table 9 Ljung-Box Q for the Forecast model of the severe injury

Q(#)	15.803
df	16.000
Significance	0.467

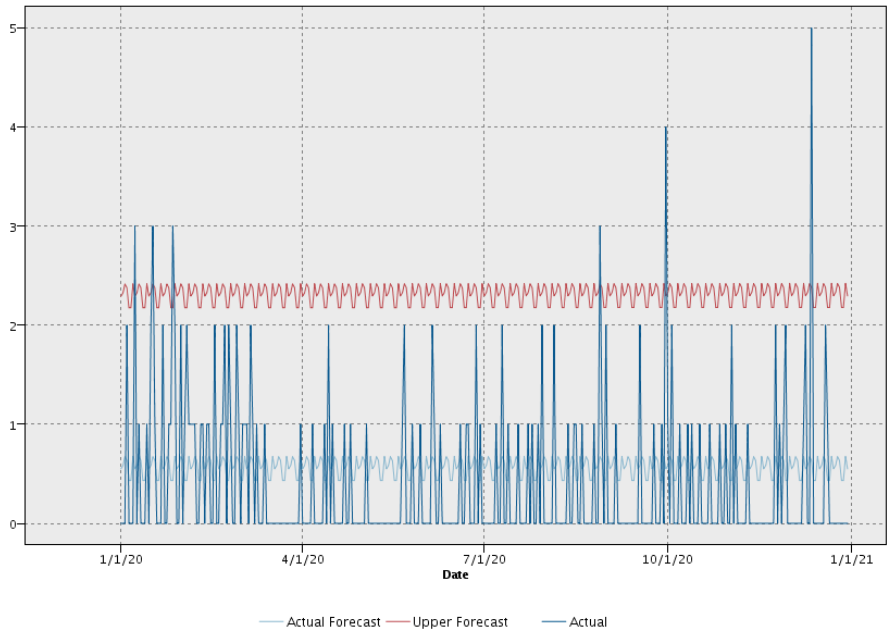


Figure 11 Forecast results vs actual observations for severe injuries in 2020

4.4.3. Risk factor analysis

To compare the correlations between four selected independent variables, two identical exhaustive CHAID models are employed. The four selected predictors are gender, age, time of the day, and injured person The training and testing prediction accuracy values for the first exhaustive CHAID are 98.02% and 97.98%, respectively. The training and testing prediction accuracy values for the second model are 96.67%and 96.34%, respectively. Figures 12 and 13 are depicting the results of both employed models. As shown, for traffic crashes that happened during 2016-2019, the injured person category is the most important variable, while for 2020, gender is the most important variable based on the applied exhaustive CHAID. For the first model, pedestrians and drivers are at a higher risk of having severe or fatal injuries compared to passengers. The branch of the pedestrian category is, then split based on the age category. Elderly pedestrians are shown to be more involved in severe or fatal injuries compared to other age categories. Driver branch is split based on the time of the day predictor. The evening period has the highest number of severe and fatal injuries compared to morning and night timing. Both evening and night timing are split based on gender, while the morning category is split based on age. Males in all branches have a higher number of severe or fatal injuries compared to females. For the second model, the male category is split based on the injured type, while the female category is split based on the age group. Drivers and passengers branch has a higher number of both slight, severe, and fatal injuries. Both morning and night timing have the same number of severe and fatal injuries which is less than the number of injuries during the evening timing. The females who are in age groups between 41 and older are showing the highest odds of having severe or fatal injuries compared to all other age categories in 2020.

Pedestrians in both years intervals are vulnerable road users when comparing the total severe and fatal injuries to total injuries that occurred for all injured persons. However, drivers have the highest number of severe or fatal injuries in total during the 2016–2019-time interval. On other branches of CHAID for 2016-2019, the age category has more complex variations compared to 2020 due to having different sets from different groups at the same branch. The time of the day risk factor shows that evening and night timing for both time intervals have a higher number of severe or fatal injuries compared to morning timing. Male injured persons are still considered a risk category in 2020 similar to the 2016–2019-time interval.

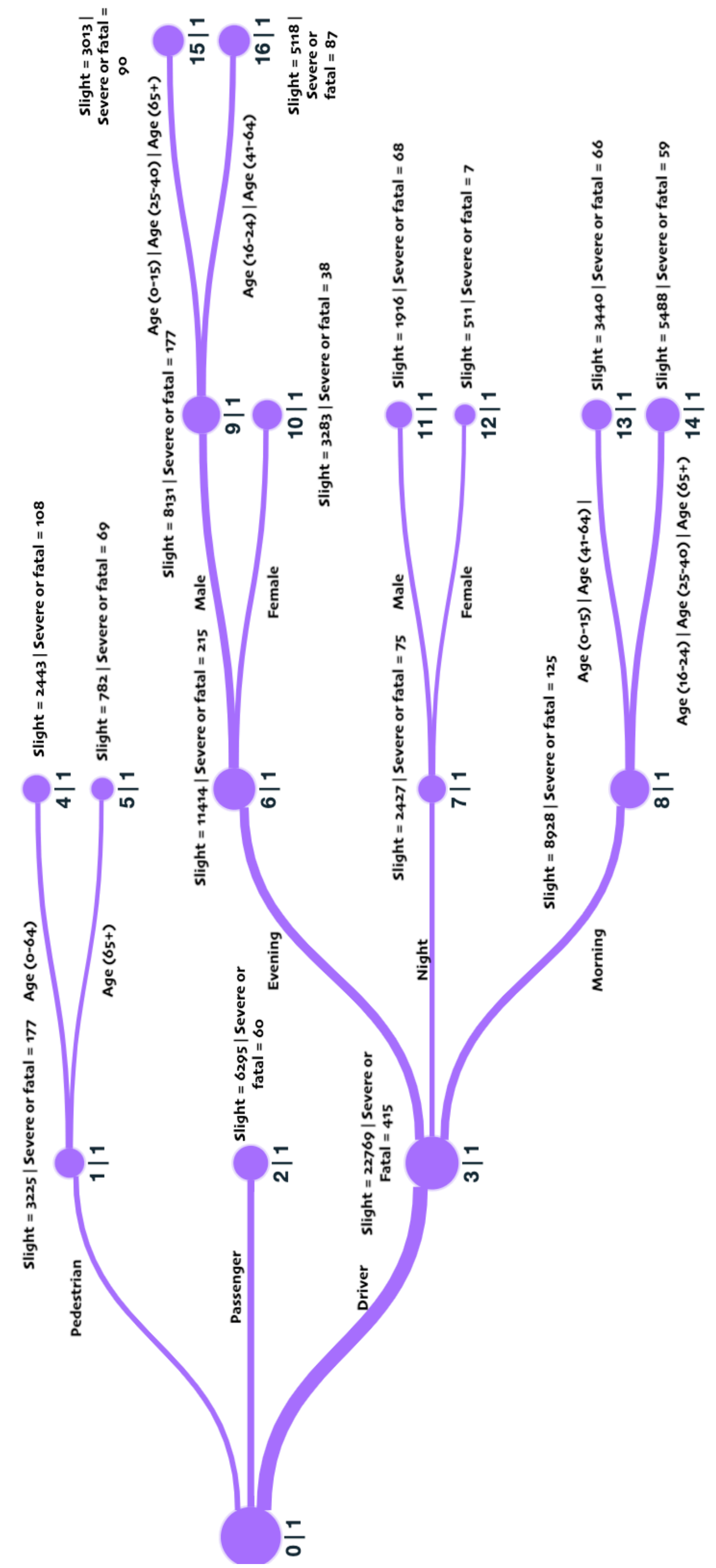


Figure 12 The structure of CHAID for traffic crash injuries between 2016-2019

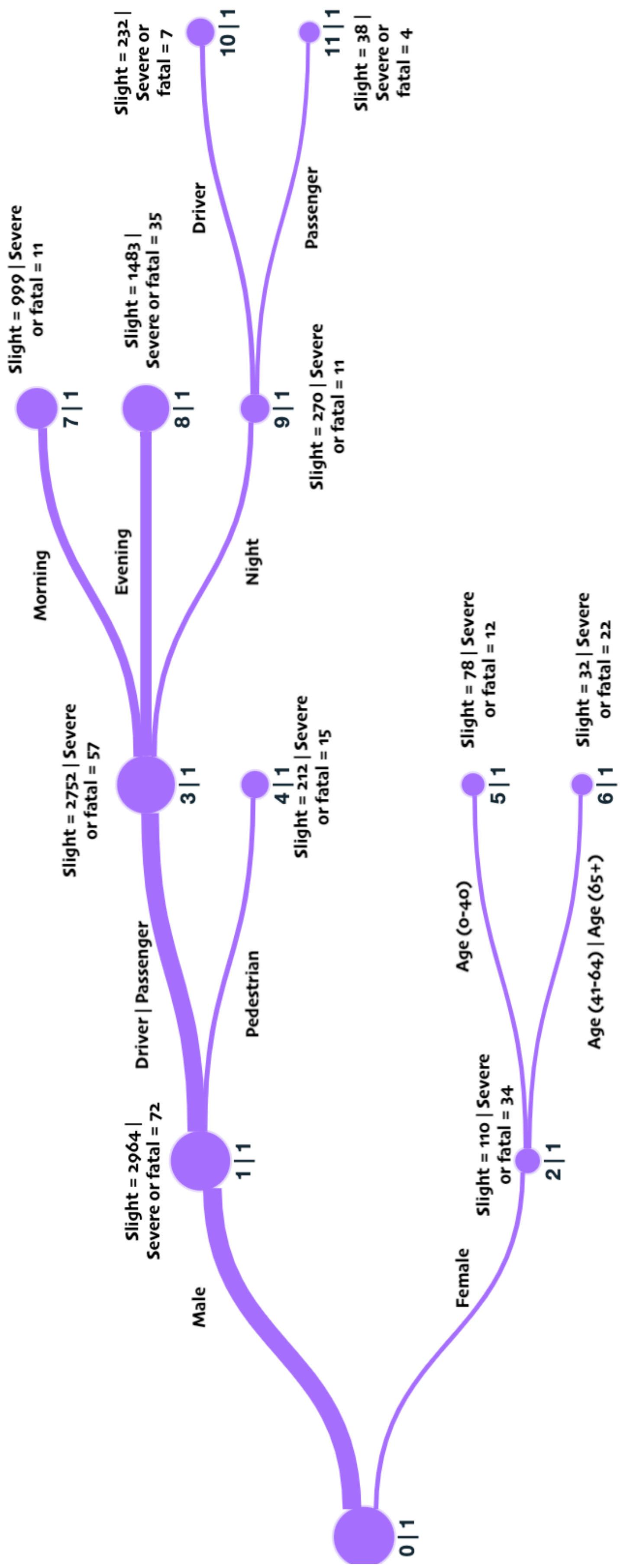


Figure 13 The structure of CHAID for traffic crash injuries in 2020

4.5. Conclusions

The COVID-19 pandemic is certainly affected our lives and many other different aspects globally. Traffic crashes are one of those aspects that are affected by this pandemic. However, the influence of COVID-19 varied in different regions related to traffic crashes, but generally, the number of traffic crashes decreased drastically worldwide due to the imposed restrictions to restrain the spread of this virus and to lessen the number of deaths and infected cases. Several studies in different countries have confirmed the impact of this pandemic. However, some studies have noticed an uncommon increase in certain traffic violations or traffic crashes after the restrictions are eased across some regions. In Barcelona, until now, no similar studies are conducted to examine the impact of COVID-19 on traffic crash injuries. Therefore, this study is conducted to provide the overall status before and during the pandemic regarding traffic crashes. The main analysis approach that is carried out in this study is time-series analysis and forecast to understand this potential impact on traffic crashes different levels of injuries that occurred in 2020 and then compare it with previous years. Exhaustive CHAID tree models are, then, employed to prepare a comparison between the pandemic period and four previous years.

The results have shown a dramatic drop in the number of traffic crashes during the pandemic, especially when the state of alarm started in March. This pattern continued till July as the number increased to reach a similar total number of the previous year. Slight injuries have shown a total percentage decrease of 39%. For the total decrease of severe injuries, a total percentage of 30% reduction during 2020 compared to 2019. Similarly, fatal injuries have dropped by 36%. However, despite the total reduction in all the levels of injuries that occurred in 2020, there was an abnormal increase during some days in severe injuries cases that happened in 2020 during September and December. Therefore, a forecast model is applied based on the four previous years including 2016, 2017, 2018, and 2019. These anomalous cases have also exceeded the upper limit of the forecast model to confirm this captured increase.

For the risk factors comparison, the results show that male injured person is still more involved in severe and fatal injuries compared to females before and during the pandemic. Evening and night timing has also a similar trend in 2020 when compared to 2016-2019 in having a higher number of severe and fatal injuries. Pedestrians during 2016-2019 have a much higher number of severe or fatal injuries compared to 2020. However, the number of severe or fatal injuries for pedestrians during 2020 is still high compared to other categories when focusing on the overall number of injuries. In 2020, elderly female injured persons are showing higher probabilities of having severe or fatal injuries compared to other age categories. Although the pandemic has affected the number of injuries, the impact of certain risk factors is found to be the same as before and during the pandemic. It can be shown that both CHAID models that are employed to examine the correlations have proven their ability to provide interpretable results.

Data availability

The datasets generated during and/or analyzed during the current study are available in the Barcelona's City Hall Open Data Service repository, https://opendata-ajuntament.barcelona.cat

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors reported there is no funding associated with the work featured in this article.

4.6. References

Abellán, J., López, G., and de Oña, J. (2013). Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications, 40*(15), 6047-6054.

Aiash, A., and Robusté, F. (2021). Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree. *International Journal of Injury Control and Safety Promotion, 29*(2), 256-264.

Alfons, A., Croux, C., and Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics, 7*(1), 226-248.

AlKheder, S., AlRukaibi, F., and Aiash, A. (2020). Risk analysis of traffic accidents' severities: An application of three data mining models. *ISA Transactions, 106*, 213-220 .

Barcelona's City Hall Open Data Service. (2020). *Open Data BCN.* (Barcelona's City Hall Open Data Service) Retrieved from https://opendata-ajuntament.barcelona.cat/en/

Barnes, S., Beland, L.-P., Huh, J., and Kim, D. (2020). *The Effect of COVID-19 Lockdown on Mobility and Traffic Accidents: Evidence from Louisiana.* Essen: Global Labor Organization (GLO).

Biggs, D., De Ville, B., and Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18*(1), 49-62.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (1994). *Time Series Analysis: Forecasting and Control, 3rd Edition.* Englewood Cliffs, N.J.: Prentice Hall.

Brockwell, P., and Davis, R. (1991). *Time Series: Theory and Methods 2nd edition.* Springer-Verlag.

Calderon-Anyosa, R., and Kaufman, J. (2021). Impact of COVID-19 lockdown policy on homicide, suicide, and motor vehicle deaths in Peru. *Preventive Medicine, 143*, 106331.

de Pedraza, P., and Guzi, M. (2020). *Life Dissatisfaction and Anxiety in COVID-19 pandemic.* Essen: Global Labor Organization (GLO).

ECDC. (2022). Retrieved from https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

Gardner, E. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting, 4*, 1-28.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge: Cambridge University Press.

Henríquez, J., Gonzalo-Almorox, E., García-Goñi, M., and Paolucci, F. (2020). The first months of the COVID-19 pandemic in Spain. *Health Policy Technol, 9*(4), 560–574.

Hu, S., Xiong, C., Yang, M., Younes , H., Luo, W., and Zhang, L. (2021). A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. *Transportation Research Part C: Emerging Technologies, 124*, 102955.

IBM. (2016). *IBM SPSS Modeler 18.0 Algorithms Guide.* IBM Corporation.

Inada, H., Ashraf, L., and Campbell, S. (2021). COVID-19 lockdown and fatal motor vehicle collisions due to speed-related traffic violations in Japan: a time-series study. *Injury Prevention, 27*, 98-100.

Kass, G. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 29*(2), 119-127.

Katrakazas, C., Michelaraki, E., Sekadakis, M., and Yannis, G. (2020). A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. *Transportation Research Interdisciplinary Perspectives, 7*, 100186.

Kerimray, A., Baimatova, N., Ibragimova, O., Bukenov, B., Kenessov, B., Plotitsyn, P., and Karaca, F. (2020). Assessing air quality changes in large cities during COVID-19 lockdowns: The impacts of traffic-free urban conditions in Almaty, Kazakhstan. *Science of The Total Environment, 730*, 139179.

Koh, D. (2020). COVID-19 lockdowns throughout the world. *Occupational Medicine, 70*(5), 322.

Lee, H., Park, S., Lee, G., Kim, J., Lee, J., Jung, Y., and Nam, E. (2020). The relationship between trends in COVID-19 prevalence and traffic levels in South Korea. *International Journal of Infectious Diseases, 96*, 399-407.

Makridakis, S., Wheelwright, S., and Hyndman, R. (1997). *Forecasting: Methods and applications 3rd edition.* New York: John Wiley and Sons.

Maryada, V., Mulpur, P., Guravareddy, A., Pedamallu , S., and Bhasker , B. (2020). Impact of COVID-19 Pandemic on Orthopaedic Trauma Volumes: a Multi-Centre Perspective From the State of Telangana. *Indian Journal of Orthopaedics, 54*, 368–373.

Melard, G. (1984). A fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics, 33*(1), 104–119.

Meyer, M. (2020). COVID Lockdowns, Social Distancing, and Fatal Car Crashes: More Deaths on Hobbesian Highways? *Cambridge Journal of Evidence-Based Policing, 4*, 238–259.

Montgomery, D., Johnson, L., and Gardiner, J. (1990). *Forecasting and Time Series Analysis.* McGraw-Hill.

Muley, D., Ghanim, M., Mohammad, A., and Kharbeche, M. (2021). Quantifying the impact of COVID–19 preventive measures on traffic in the State of Qatar. *Transport Policy, 103*, 45-59.

Nuñez, J., Sallent, A., Lakhani, K., Guerra-Farfan, E., Vidal, N., Ekhtiari, S., and Minguell, J. (2020). Impact of the COVID-19 Pandemic on an Emergency Traumatology Service: Experience at a Tertiary Trauma Centre in Spain. *Injury, 51*(7), 1414-1418.

Oguzoglu, U. (2020). COVID-19 Lockdowns and Decline in Traffic Related Deaths and Injuries. *IZA*, 13278.

Pena, D., Tiao, G., and Tsay, R. (2001). *A course in time series analysis.* New York: John Wiley and Sons.

Prasetijo, J., Zhang, G., Jawi, Z., Mahyeddin, M., Zainal, Z., Isradi, M., and Muthukrishnan, N. (2021). Crash model based on integrated design consistency with low traffic volumes (due to health disaster (COVID-19)/movement control order). *Innovative Infrastructure Solutions, 6*, 22.

Qureshi, A., Huang, W., Khan, S., Lobanova, I., Siddiq, F., Gomez, C., and Suri, M. (2020). Mandated societal lockdown and road traffic accidents. *Accident Analysis & Prevention, 146*, 105747.

Sakelliadis, E., Katsos, K., Zouzia, E., Spiliopoulou, C., and Tsiodras, S. (2020). Impact of Covid-19 lockdown on characteristics of autopsy cases in Greece. Comparison between 2019 and 2020. *Forensic Science International, 313*, 110365.

Saladié, Ò., Bustamante, E., and Gutiérrez, A. (2020). COVID-19 lockdown and reduction of traffic accidents in Tarragona province, Spain. *Transportation Research Interdisciplinary Perspectives, 8*, 100218.

Shen, J., Wang, C., Dong, C., Tang, Z., and Sun, H. (2021). Reductions in mortality resulting from COVID-19 quarantine measures in China. *Journal of Public Health*, 1–7.

Solomon, H. (2020). COVID-19 checklist: Mask, gloves, and video chatting with grandpa. *Psychiatry Research, 288*, 112986.

Stavrinos, D., McManus, B., Mrug, S., He, H., Gresham, B., Albright, M., Svancara, A. M., Whittington, C., Underhill, A., and White, D. (2020). Adolescent driving behavior before and during restrictions related to COVID-19. *Accident Analysis & Prevention, 144*, 105686 .

The Lancet Infectious Diseases. (2020). COVID-19, a pandemic or not? *The Lancet Infectious Diseases, 20*(4), 383.

TomTom. (2020). Barcelona traffic. *TomTom*. Retrieved fromhttps://www.tomtom.com/en_gb/traffic-index/barcelona-traffic/

UNICEF; WHO; IFRC. (2020). *Key Messages and Actions for COVID-19 Prevention and Control in Schools March 2020.* World Health Organization.

World Health Organization. (2022). Retrieved from https://covid19.who.int

Xiang, J., Austin, E., Gould, T., Larson, T., Shirai, J., Liu, Y., . . . Marshall, J., and Seto, E. (2020). Impacts of the COVID-19 responses on traffic-related air pollution in a Northwestern US city. *Science of The Total Environment, 747*, 141325.

Zhang, S., Wang, Y., Rauch, A., and Wei, F. (2020). Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak. *Psychiatry Research , 288*, 112958.

Intentionally blank page

5. Total noise level's passive impact on traffic crashes: a potential distraction indicator

5.1. Abstract

Noise has not been much studied in relation to factors causing traffic crashes. This study has selected two neighbourhoods located in Barcelona (Spain) in order to examine the potential impact of total noise levels recorded in the city on drivers' and pedestrians' concentration, thus leading to a traffic crash. A logistic regression model is then applied to determine whether or not the hypothesis is true. The results show that areas with higher total noise levels may lead to traffic crashes and further investigations are required.

***Keywords*:** Noise, Distraction, Traffic, Crash, Safety, Barcelona

5.2. Introduction

The number of risk factors that can be considered as an indication of a traffic crash includes many that can be hard to detect. Most published studies appear to consider the following: crash characteristics, road types, temporal (Aiash and Robusté, 202A) or spatial factors alongside traditional risk factors that may have a consequential impact on traffic crashes such as gender (Aiash and Robusté, 2021B), age (AlKheder et al., 2020), and social characteristics.

Noise is one of the undocumented factors such as drivers' inattention and drowsy driving (Sagberg, 2018) that is, still, not considered by studies that are examining traffic crashes despite its known mischievous and deleterious effects on human health. There are various definitions of noise. A Directive of the European Union (EU) defines noise as "unwanted or harmful outdoor sound created by human activities, including noise from the road, rail, airports, and from industrial sites." (Environmental Noise Directive, 2002). Mohamed et al. (Mohamed et al., 2021) categorize noise levels into several groups based on the sound level measured in decibels (dB). Noise with 0 dB is near total calm. 15 dB is designated for sounds similar to whisper, and 60 dB is for sounds similar to an ordinary discussion. A level of 90 dB is for sounds similar to a grass cutting machine, whereas a car siren is designated as 110 dB. Being exposed to sounds that exceed 85 dB is considered harmful for both adults and children, and immediate damage can happen when being exposed to sounds over 140 dB. Population growth, urbanization, industrialization and transport networks etc were found to be significant factors contributing to higher levels of noise according to (Sordello et al. 2019) and (Swaddle et al., 2015). Table 1 shows previous researchers' findings on the consequences of noise for human health.

Table 10 State of art for noise consequences on human health

Article Identifier	**Noise consequences**
(Farooqi et al., 2019)	Urban noise levels can lead to headaches, sleeplessness, and high blood pressure levels.
(Al-Harthy et al., 2021)	Sleep disturbance, insomnia, irritation, and fright can be some of the syndromes that were associated with regions near the airport.
(Gilani and Mir, 2022)	Traffic noise can cause insomnia.
(Wolfgang, 2008)	Road traffic noise levels can increase the risk of having cardiovascular health problems if the levels of noise were above 60 dB.
(Foraster et al., 2011)	Road traffic noise and nitrogen dioxide could have an effect on cardiovascular health.
(Mekel et al., 2012)	Road traffic noise can affect the sleep of children who were between 0-14 years old .
(Sørensen et al., 2012)	A 10 dB noise can increase the risk of having diabetes.
(Barceló et al., 2016)	The long-term impact of traffic noise can be correlated with myocardial infarction mortality.

Noise can have another consequential mental impact on human beings also. (Foraster et al., 2022) examined the potential influence of traffic noise on the cognition of children at schools who were over one year old in Barcelona. This type of noise had a negative impact; there was evidence of slower working memory development and attention. (Weyde et al., 2017) found that road traffic noise can cause the inattention of children. (Hygge and Knez, 2001), analysing the effects of noise, heat and indoor lighting on cognitive performance and self-reported affect upon 128 participants aged between 18 and 19 years, found that ventilation noise affected the participants' attention; the participants appeared to work faster but with less accuracy. Additionally, the study found that there was an interaction between heat and noise that affected the long-term recall memory related to a certain task. For noise and vibration to which bus drivers are exposed during their daily work, (Rahmani et al., 2021) found an increase in response time but a decrease in true response rate. The true response *rate* means the accuracy of the participant in choosing the correct answer. The response *time* means the time taken to finish the task. These two indicators represent the cognitive performance of the drivers who participated (Rahmani et al., 2021). (Rahmani et al., 2021) used a simple test to measure cognitive performance. This test evaluates the selection of a person for certain words with different colours to determine the attention, skills capacity, and reaction time of the participant. The frequency of very inaccurate answers by the participants for the conducted evaluation method that was based on detecting similar numbers under noise and quiet environment is also found as a part of the consequential effects of noise according to (Smith, 1988). Consistent with previous findings, (Domhnaill et al., 2021) found that the executive function of old adults was negatively affected by road traffic noise in case of long-term exposure.

Other researchers have investigated traffic volume and noise levels relationship, e.g., (Abdur-Rouf and Shaaban, 2022). (Escobar and Pérez, 2018) studied street classification when examining urban noise and average annual daily traffic (ADT) data to grasp this relationship. The results showed that there is a relationship between traffic flow and noise levels. (Ross et al., 2011) found that noise levels were temporally correlated with traffic and combustion pollutants. Since ADT can be linked to traffic crashes, many studies have investigated traffic crashes and their relationship with ADT and annual average daily traffic (AADT) worldwide. (Gwynn. 1967) found that traffic crashes with injury rates or higher totals happened in both low and high hourly traffic volume ranges. Consistently, (Ceder and Livneh, 1978) analysed traffic crash data on interurban road sections, including details related to ADT. The research found that higher ADT can lead to higher traffic crash occurrences.

The main research objective is to examine the potential passive effect on drivers and pedestrians of total noise levels on traffic crashes by examining the total noise levels that are recorded in the city of Barcelona to detect whether there is a correlation between drivers and pedestrians and the total noise levels as a potential risk factor that can lead to traffic crashes. Total noise levels are calculated by the city council of Barcelona according to the CNOSSOS-EU method, established in Directive 2002/49/EC S. (Kephalopoulos et al., 2012). The main sources of noise are the recorded total noise in an area in general, road traffic noise, rail and tramway traffic noise, and industrial noise according to (Barcelona City Council, 2017). To the best of our knowledge, there is a lack of similar studies in Spain and elsewhere focused on noise as a potential distraction risk factor for road users resulting in a traffic crash. Eventually, the total noise levels that are measured and recorded in l'Antiga Esquerra de l'Eixample and Sant Antoni neighbourhoods in Barcelona are investigated alongside traffic crashes data from the same year - 2017. Then, a logistic regression model is applied to determine whether or not the hypothesis can be considered true, viz., that total noise levels can be a potential risk factor related to traffic crashes and can have a negative effect on drivers and pedestrians as it can be a source of distraction for them. Therefore, the trace of total noise levels can be captured, and this investigation can set the groundwork for future studies related to noise as a potential risk factor for traffic crashes. The remainder of the paper is organized as follows: section two presents the details about the chosen areas' characteristics, the data and factors, and the applied model. Section three presents the results. Lastly, the conclusions are presented in section four.

5.3. Methodology

5.3.1. Case areas characteristics

The district of interest selected for study is Eixample. The reason for choosing this district is the intensity of traffic crash data availability compared to other districts. The two chosen neighbourhoods are L'Antiga Esquerra de l'Eixample and Sant Antoni. The reason for choosing these areas is that the number of crashes that occurred in the Eixample district is the highest compared to all other districts. This is crucial in the analysis part as a lower number of cases of traffic crashes can significantly impact the results of the applied model. Additionally, the contradiction between the two neighbourhoods regarding the recorded noise levels in the same district is not found in other districts with comparable characteristics that are shown later. This contradiction is related to having two neighbourhoods, with one that has high levels of noise and the other neighbourhood with lower noise levels but with comparable characteristics that can impact the occurrence of traffic crashes. The other reasons that may impact traffic crashes are not considered in this study, as the main objective of this study is the potential noise level impact on traffic crashes. L'Antiga Esquerra de l'Eixample had a population of 42,512 in 2017, with 66.9 % of the inhabitants being between 16 and 64 years according to (Ajuntament de Barcelona, 2017A). Sant Antoni had a population of 38,412 in 2017, and 66.5 % of the inhabitants were between 16 and 64 years old according to (Ajuntament de Barcelona 2017B). Males represented 53.3% of the inhabitants, and females represented 46.7% of the inhabitants in L'Antiga Esquerra de l'Eixample. For Sant Antoni, males represented 51.6% of the inhabitants, and females represented 48.4% of the inhabitants. Age and gender statistics for both areas are compared in this study by reference to the previous work of (Aiash and Robusté, 2021) correlating the age and gender of the injured person and traffic crashes. This work shows similarities between the two studied areas that are chosen to examine the impact of noise levels on traffic crashes to eliminate population factor differences between them. The economic activity differences between the two neighbourhoods are as follows; L'Antiga Esquerra had 16.7 % of its activities for industry or warehouse activities. In Sant Antoni, and for the same activity type, the area had 23.6%. Based on the same references, L'Antiga Esquerra had in the year 2017 a total of 308 vehicles per 1000 inhabitants, and Sant Antoni had 255. For the motorcycles, 148 and 196 motorcycles per 1000 inhabitants were in Sant Antoni and L'Antiga Esquerra, respectively. For the unemployment rate, Sant Antoni and L'Antiga Esquerra reported unemployment cases of 1,598 and 1,433 in October 2017, respectively.

The traffic status is then compared between both areas to eliminate the differences in that respect. The reason for this comparison is the possible impact of traffic flow on traffic crashes based on the findings of the studies that are mentioned previously. Therefore, this factor can be eliminated if both areas show similar traffic status. To provide a comparison of traffic status for vehicles entering or exiting the two selected neighbourhoods, 44 designated points are used for this purpose with QGIS software and based on the same data source, Barcelona's City Hall Open Data Service (figure 14). These points provide traffic actual status by combining the data of October, November, and December of 2017. This period is chosen based on the available data that are gathered in 2017.

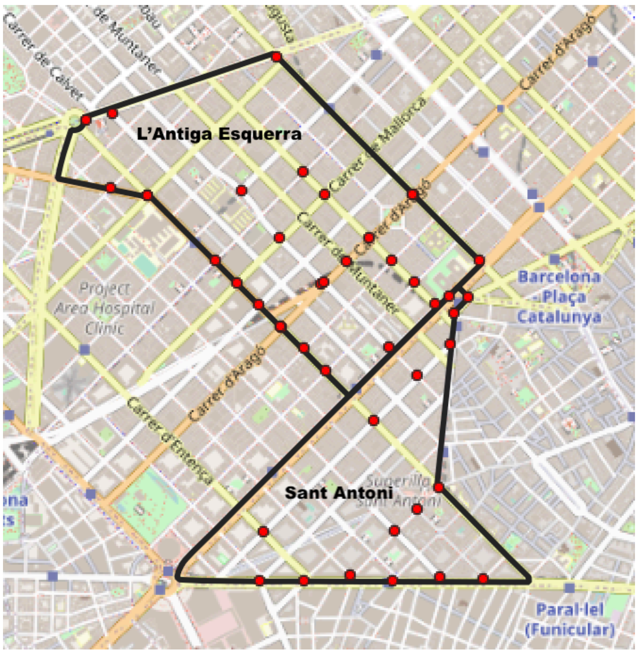


Figure 14 The designated points to measure traffic status

The data from these designated points are mainly collected from sensors, placed beneath the asphalt, which can track changes brought on by the passage of metal masses (vehicles) in the magnetic field. Other tools used to measure the status of the road at the designated points can include infrared sensors and image-processing cameras. Based on certain established thresholds for each station reported in the database source, the service level of the related section information is then translated in a qualitative style from each detection station (values without units represented by value zero, from value 1 to value 5). Qualitative style means that the responsible operators of the Mobility Management Centre in the city of Barcelona are transforming the data from just numbers (how many vehicles passed the censor or the detector), which are quantitative data, to words that are represented by the categories from (0 to 5), each representing the traffic status of a certain road. This means that the responsible department for this type of data is categorizing the data based on these categories to provide the overall status for different districts and road segments.

After the data are translated, the traffic status is described as follows; value zero is assigned to points with no available data based on the source where those data are collected to show if these detectors are not functioning well or there is no traffic where the detector is located. Values 1 and 2 are designated for very fluid and fluid points, respectively. Dense and very dense points data are described by 3 and 4 values, respectively. Value 5 is for congestion. Lastly, value 6 is for the cut road. These data are then used to depict the comparison between the neighbourhoods, as shown in figure 15. The examined neighbourhoods started at different points regarding no data availability, followed by a considerably close fluid frequency that is represented by value 1 (i.e., very fluid traffic status). When considering the only fluid status, the case is significantly different between the studied points in these different neighbourhoods. The slope draws near, then, when approaching the dense status that is represented by point 3. The values are becoming more similar when considering very dense, congested, and cut road traffic status. These findings provide an overall insight into the status of traffic in the selected neighbourhoods in order to expel the ADT impact on the following conducted results and to show several similarities between them regarding congested sections.

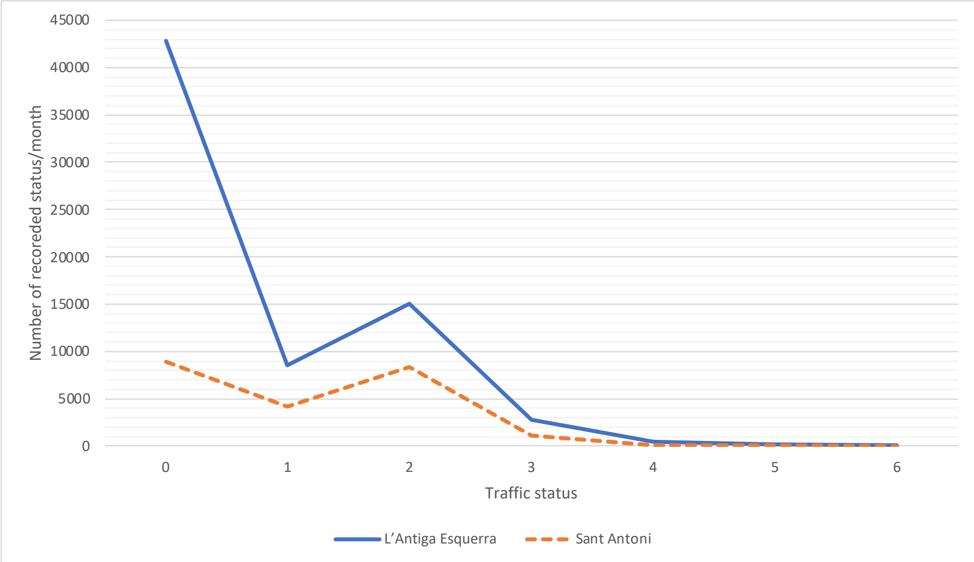


Figure 15 A comparison of traffic status between L'Antiga Esquerra and Sant Antoni

5.3.2. Noise levels data

The same database is used for collecting the total noise levels data in Barcelona - Barcelona's City Hall Open Data Service (Ajuntament de Barcelona, 2017C). The noise sources that Barcelona regulates include major road traffic, railway, tramway, and industry. These total noise levels are calculated, as stated earlier, according to the CNOSSOS-EU method, established in Directive 2002/49/EC (Kephalopoulos et al., 2012). Other sources of noise are excluded from regulation: leisure and crowding elements, pedestrian streets, parks, and inner courtyards. For the total noise levels data, there are different types of noise levels. The basic three periods of noise levels are L_{Day}, $L_{Evening}$, and L_{Night} which refers to equivalent continuous sound pressure levels during the day, evening, and night periods, respectively. Based on the environmental ordinance for the city of Barcelona, the time period for each noise category is 7 a.m. to 9 p.m. (day), 9 p.m. to 11 p.m. (evening), and 11 p.m. to 7 a.m. (night). Based on WHO report (WHO Regional Office for Europe, 2018) and European standards (European Union law, 2003), L_{Night} and L_{DEN} are the most used indicators to assess the exposure to noise in different health studies. L_{DEN} refers to day-evening-night weighted sound pressure level with penalizing evening period with 5 dB and night period with 10 dB. The formula to determine L_{DEN} (European Union law, 2003), is as follows:

$$L_{DEN} = 10\ \log\left[\frac{1}{24}\left(t_d \times 10^{\frac{L_{Day}}{10}} + t_e \times 10^{\frac{L_{Evening+5}}{10}} + t_n \times 10^{\frac{L_{Night+10}}{10}}\right)\right] \quad (1)$$

t_d is the resulting length of the period for the daytime, which in this study is 14 hours. t_e is the length of the evening period that should be between 2 and 4 hours. In this study, and based on the collected data, this period is equal to 2 hours. t_n is the resulting length of the night period timing which in this study is equivalent to 8 hours. The total number of studied hours is 24 hours which is explained by the dominator in the equation.

The data exploited in this study comprise different total levels percentages for different noise categories, as shown in table 11. Based on Barcelona's City Hall Open Data Service (Ajuntament de Barcelona, 2017C) database, noise data present the percentages of the population exposed to the different noise ranges according to the time period and the type of source etc. The data are available at different levels, including city, district and neighbourhood levels. These data are prepared every five years and serve as a management tool to fight noise pollution. The range of noise levels is categorized into ten categories based on the original data source. These categories range from 0 to 40 dB, 40 to 45 dB, 45 to 50 dB, 50 to 55 dB, 55 to 60 dB, 60 to 65 dB, 65 to 70 dB, 70 to 75 dB, 75 to 80 dB, and lastly, more than 80 dB category. The percentage is given based on the collected data for each category in each chosen neighbourhood. According to (Schlittmeier et al., 2015), cognitive performance is affected by road traffic noise levels that exceed 70 dB compared to 50 dB. Additionally, (Wang et al., 2021) and (Berglund at al., 2021) found that noise levels that exceed 70 dB can cause hearing impairment for a large majority of people. Therefore, noise levels that range between 70-75, 75-80 dB, and more than 80 dB are chosen in this study to identify the level of noise that might be a distraction indicator. The total noise in the neighbourhood and transit noise which records road traffic noise, have the same percentages based on the database. Thus, the total noise categories are presented to examine the differences between the neighbourhoods. Both categories 70-75 dB and 75-80 dB for both types of reported noise categories are higher in L'Antiga Esquerra de l'Eixample than in Sant Antoni.

Table 11 Noise levels categories percentages

Name of the neighbourhood	Neighbourhood code	Range of noise levels	Total L_{Day}	Total $L_{Evening}$	Total L_{Night}
l'Antiga Esquerra de l'Eixample	8	70-75 dB	**26.28%**	**13.50%**	**0.02%**
l'Antiga Esquerra de l'Eixample	8	75-80 dB	**1.16%**	**0.05%**	0.00%
l'Antiga Esquerra de l'Eixample	8	≥80 dB	0.00%	0.00%	0.00%
Sant Antoni	10	70-75 dB	12.26%	3.34%	0.00%
Sant Antoni	10	75-80 dB	0.08%	0.00%	0.00%
Sant Antoni	10	≥80 dB	0.00%	0.00%	0.00%

The percentages for L_{DEN} are already determined and ready to be used from the database directly. Then, the average values for total L_{DEN} noise are determined by following (Essen et al. 2005) and only considering ranges categories between 70-75 and 75-80 dB as the category that is higher than 80 dB has a zero value for both areas, as shown in table 12. Because the collected values are percentages, the weighted average of total L_{DEN} noise in a certain neighbourhood is determined and compared to the other neighbourhood using the following formula:

$$W = \frac{\sum_{i=1}^{2} b_i \times a_i}{\sum_{i=1}^{2} b_i} \tag{2}$$

b_i is the abbreviation of the average range noise levels, which 72.5 is for b_1 and 77.5 is for b_2, while a_i is the percentage value of total L_{DEN} for each category.

Table 12 Total L_DEN noise levels

Name of the neighbourhood	Range of noise levels	Total L_{DEN}	W	Abbreviation
l'Antiga Esquerra de l'Eixample	70-75 dB	35.54%		
l'Antiga Esquerra de l'Eixample	75-80 dB	3.96%	0.19223667	W1
Sant Antoni	70-75 dB	22.58%		
Sant Antoni	75-80 dB	2.23%	0.12065833	W2

5.3.3. Traffic crashes data

The traffic crashes data consist only of the road users who are making decisions while using the road, similar to drivers and pedestrians, alongside the location of the crash, which is the neighbourhood. Therefore, the passenger category is included when exploiting the data for the analysis step in order to distinguish the differences between this type of road user and other road users who are capable of making the decisions on the road. Both drivers and pedestrians are the ones who are making decisions on the road, unlike passengers, who cannot do so. Therefore, passenger existence in the analysis part represents the null hypothesis of claiming that noise does not have an influence on traffic crashes. This means if there are correlations between passengers, traffic crashes, and noise levels in different areas, this will exclude the hypothesis that there is a correlation between noise and traffic crashes as the crash happened because of decisions made by drivers and pedestrians that are included in this study. Tables 13 and 14 portray the overall crashes that occurred in 2017 in both L'Antiga Esquerra de l'Eixample and Sant Antoni. Driver, passenger and pedestrian who had traffic crash injuries in l'Antiga Esquerra de l'Eixample and Sant Antoni are included in these tables. The number of crashes that drivers are involved in is the highest among other categories, including passengers and pedestrians, besides showing that the neighbourhood that has the highest weighted average of total L_{DEN} has more crashes in general, viz., l'Antiga Esquerra de l'Eixample that is represented by W1.

Table 13 Crashes in l'Antiga Esquerra de l'Eixample and Sant Antoni

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Driver	715	74.25	715	74.25
Passenger	146	15.16	861	89.41
Pedestrian	102	10.59	963	100
Total crashes	963	100	963	100

Table 14 Road user crashes based on the total noise levels

	W1	W2
Driver	522	193
Passenger	96	50
Pedestrian	62	40

5.3.4. Noise impact analysis

A logistic regression model is applied in this study to understand the correlation between the noise levels in the Eixample region, including the two chosen neighbourhoods. In this study and in order to examine the impact of noise level as an element of distraction, three road users are involved in the model: the pedestrian, driver, and passenger. The main concern is to distinguish the impact of the trace of total noise levels that can hinder drivers or pedestrians from concentrating and eventually lead to traffic crashes. Then, each neighbourhood is then substituted by the different weighted average of the total L_{DEN} noise category that it represents when initiating the analysis step. For the logistic regression equation and by following (Alkheder et al., 2019), the dependent variable, which is the weighted average of total L_{DEN} noise in a certain neighbourhood W, is assigned to be y. W2 is assigned to be y=1, and W1 is assigned to be y=2 which is the reference level. Then y=1 is compared to y=2. The logistic regression function is as follows:

$$logit(p_1) = \ln\left(\frac{p(y=1|x)}{p(y=2|x)}\right) = \alpha_1 + \sum_{i=1}^{m} \beta_{1i}\, x_i \tag{3}$$

x_i is the value for the ith independent variable. α_1 is the intercept of the y=1 logistic function. The number of variables is m.The corresponding coefficient is β_1. The condition probability of kth outcome category is:

$$p(y = k|x) = \frac{\exp(\alpha_k + \sum_{i=1}^{m} \beta_{ki} x_i)}{1 + \sum_{k=1}^{K-1} (\alpha_k + \sum_{i=1}^{m} \beta_{ki} x_i)} \tag{4}$$

α_K is the intercept of the kth logit function. The corresponding coefficient of the kth logit function (4) is β_{ki}. K is the number of outcome category.

5.4. Results

Table 15 portrays the model's significance based on the p-value. The p-value for the model is less than 0.05. This indicates that the null hypothesis is rejected. This null hypothesis states that there is no relationship between the injured road user and the weighted average of total L_{DEN} noise levels. This means that the alternative hypothesis is statistically significant which supports the existence of having a relationship between the response variable represented by the total noise levels for each area after determining the weighted average (W) of total L_{DEN} noise for each chosen area and the classification variable represented by the road users including drivers, passengers, and pedestrians.

Table 15 Logistic regression model fitting

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Significance
Intercept Only	23.936	-	-	-
Final	16.026	7.91	2	0.0192

Each W including W1 and W2 as previously shown in table 12 represents the weighted average of total L_{DEN} noise decibels for each area and includes it in the analysis to grasp the correlation between the total noise levels and the road users. Table 16 depicts the estimates of the classification variable correlations. The positive value of the driver category for its calculated exponentiation of the B coefficient (odds ratio) with a significant p-value which is less than 0.05 indicates that drivers are more likely to have a crash at L'Antiga Esquerra de l'Eixample that has higher W than Sant Antoni. In general, the significant value indicates the possible correlation between drivers and W. The negative B (coefficient) value for the driver category indicates that pedestrians (the reference category) are less likely to have a traffic crash than are drivers in L'Antiga Esquerra de l'Eixample neighbourhood. But, the passenger category, on the other hand, has a value larger than 0.05, indicating that this is an insignificant category with respect to W.

Table 16 Weighted average of the total L_DEN estimation results

	Predictor classes	B	Standard error	Wald	df	Significance	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
W2	Intercept	0.417	0.239	3.038	1	0.081	-	-	-
	Driver	-0.597	0.259	5.308	1	0.021	0.550	0.331	0.915
	Passenger	-0.112	0.325	0.119	1	0.730	0.894	0.473	1.689
	Pedestrian (reference)	-	-	-	-	-	-	-	-

5.5. Conclusions

Noise has manifested its consequences on human health and the environment. But, noise, till now, was not investigated as a potential risk factor, not even considering its effect on the cognition of drivers who committed traffic crashes. Many cities and areas are recording noise levels that are transmitted into different neighbourhoods to evaluate the quality of life of their residents. This study, as a result, attempts to move against the influence of noise on traffic crashes that includes drivers or pedestrians. The impact of noise levels is examined by exploiting data sets that occurred in Barcelona in 2017. Two neighbourhoods are chosen to establish the comparison and the analysis. After the comparison demonstration between the two selected neighbourhoods, a logit regression model is applied to determine the hypothesis.

L'Antiga Esquerra de l'Eixample has a higher number of traffic crashes than Sant Antoni. After the comparison between L'Antiga Esquerra de l'Eixample and Sant Antoni, both areas show comparable characteristics. Then, both areas are examined and analysed for noise levels after grasping the overall characteristics. Regarding the traffic status, both areas show a similar reported status regarding congested and dense streets at specifically chosen points. A logit regression model is then applied to examine the impact of total noise levels on traffic crashes. The results show that the area with a higher level of total L_{DEN} may lead to traffic crashes that include drivers than the area that has lower levels.

Certain measures can be carried out to mitigate the impact of noise on traffic crashes. Firstly, areas with higher levels of noise should be monitored more in order to establish the sources of noise so that correct policies can target them in order to reduce their levels. Traffic noise can be reduced by imposing zones with a speed limit of 30 km/h according to (Rossi et al., 2020), which could lessen the impact of total noise levels on drivers.

The main limitation of the study is the collected data, as the most recent available data reported in 2017 in Barcelona has the total noise levels. This can hinder further analysis that can be carried out as the data for traffic crashes is available every year and can give better results with having a larger amount of data to be examined. For future work, large-scale analysis can be carried out to examine the impact of noise on road users in different regions. In addition, lab tests can also be conducted to determine the impact of noise more precisely. Then, multiple scenarios can be proposed to grasp the perception and reaction of the participants related to traffic crashes and noise. These multiple scenarios can include testing different noise levels on participants to examine the impact of each noise level on their concentration, their ability to make decisions, and probably their ability to avoid a traffic crash.

Data availability

The datasets generated during and/or analysed during the current study are available in the Barcelona's City Hall Open Data Service repository, https://opendata-ajuntament.barcelona.cat

Conflict of interest

The authors declare that there is no conflict of interest in any manner regarding the publication of this article.

Funding

The authors reported there is no funding associated with the work featured in this article.

5.6. References

Abdur-Rouf, K., and K. Shaaban. 2022. "A Before-and-After Traffic Noise Analysis of a Roundabout Converted to a Signalized Intersection." *Arabian Journal for Science and Engineering* .

Aiash, A., and F. Robusté. 2021A. "Working hours and traffic accident injuries: case study in Barcelona." *Transportation Research Procedia* 58: 678-682.

Aiash, A., and F. Robusté. 2021B. “Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree.” *International journal of injury control and safety promotion* 29 (2): 256-264.

Ajuntament de Barcelona. 2017a. *L’Antiga Esquerra de l’Eixample Eixample.* https://ajuntament.barcelona.cat/estadistica/catala/index.htm.

Ajuntament de Barcelona . 2017b. *Sant Antoni Eixample.* https://ajuntament.barcelona.cat/estadistica/catala/index.htm.

Al-Harthy, I., P. Amoatey, A. Al-Mamun, Z. Alabri, and M. S. Baawain. 2021. “Assessment of noise levels and induced annoyance in nearby residential areas of an airport region in Oman.” *Environmental Science and Pollution Research* 28: 45596–45608.

Alkheder, S., F. Alrukaibi, and A. Aiash. 2019. “Drivers’ response to variable message signs (VMS) in Kuwait.” *Cognition, Technology & Work* 21: 457–471.

AlKheder, S., F. AlRukaibi, and A. Aiash. 2020. “Risk analysis of traffic accidents’ severities: An application of three data mining models.” *ISA transactions* 106: 213-220.

Barceló, M. A., D. Varga, A. Tobias, J. Diaz, C. Linares, and M. Saez. 2016. “Long term effects of traffic noise on mortality in the city of Barcelona, 2004–2007.” *Environmental Research* 147: 193-206.

Barcelona City Council. 2017. *Environmental data maps.* Barcelona City Council. https://ajuntament.barcelona.cat/mapes-dades-ambientals/soroll/es/faqs.html.

Barcelona's City Hall Open Data Service. 2017. *Open Data BCN.* https://opendata-ajuntament.barcelona.cat/en/.

Berglund, B., T. Lindvall, and D. H. Schwela. 1999. *GUIDELINES FOR COMMUNITY NOISE.* Geneva: World Health Organization.

Ceder, A., and M. Livneh. 1978. “Further evaluation of the relationships between road accidents and average daily traffic.” *Accident Analysis & Prevention* 10 (2): 95-109.

Domhnaill, C. M., O. Douglas, S. Lyons, E. Murphy, and A. Nolan. 2021. “Road traffic noise and cognitive function in older adults: a cross-sectional investigation of The Irish Longitudinal Study on Ageing.” *BMC Public Health volume* 21: 1814 .

Environmental Noise Directive. 2002. “Directive 2002/49/EC of the European parliament and of the council of 25 June 2002 relating to the assessment and management of environmental noise.” *Official Journal of the European Communities* L189: 12-26.

Escobar, V. G., and C. J. Pérez. 2018. “An objective method of street classification for noise studies.” *Applied Acoustics* 141: 162-168.

Essen, H. P. , B. H. Boon, S. Mitchell, D. Yates, D. Greenwood, and N. Porter. 2005. *Sound Noise Limits Options for a uniform noise limiting scheme for EU airports.* Delft: CE.

European Union law. 2003. *Commission Recommendation of 6 August 2003 concerning the guidelines on the revised interim computation methods for industrial noise, aircraft noise, road traffic noise and railway noise, and related emission data.* Official Journal of the European Communities.

Farooqi, Z. U. R., M. Sabir, J. Latif, Z. Aslam, H. R. Ahmad, I. Ahmad, M. Imran, and P. Ilić. 2019. “Assessment of noise pollution and its effects on human health in industrial hub of Pakistan.” *Environmental Science and Pollution Research* 27: 2819–2828.

Foraster, M., A. Deltell, X. Basagaña, M. Medina-Ramón, I. Aguilera, L. Bouso, M. Grau, et al. 2011. “Local determinants of road traffic noise levels versus determinants of air pollution levels in a Mediterranean city.” *Environmental Research* 111 (1): 177-183.

Foraster, M., M. Esnaola, M. López-Vicente, I. Rivas, M. Álvarez-Pedrerol, C. Persavento, N. Sebastian-Galles, J. Pujol, P. Dadvand, and J. Sunyer. 2022. “Exposure to road traffic noise and cognitive development in schoolchildren in Barcelona, Spain: A population-based cohort study.” *PLoS Med* 19 (6): e1004001.

Gilani, T. A., and M. S. Mir. 2022. “A study on road traffic noise exposure and prevalence of insomnia.” *Environmental Science and Pollution Research* 29: 41065–41080.

Gwynn, D. W. 1967. "Relationship of accident rates and accident involvements with hourly volumes." *Eno Transportation Foundation* 21 (3): 407-418.

Hygge, S., and G. Knez. 2001. "Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect." *Journal of Environmental Psychology* 21: 291-299.

Kephalopoulos, S., M. Paviotti, and F. Anfosso-Lédée. 2012. *Common Noise Assessment Methods in Europe (CNOSSOS-EU), To be used by the EU Member States for strategic noise mapping following adoption as specified in the Environmental Noise Directive 2002/49/EC.* Ispra, Italy: European Commission Joint Research Centre Institute for Health and Consumer Protection.

Mekel, O. C. L., S. Sierig, and T. Claßen. 2012. "Feasibility study of HIA (Health Impact Assessment) on road traffic noise induced health effects on children." *Pollution Atmospherique* 216: 343-352.

Mohamed, A. O. , E. K. Paleologos, and F. M. Howari. 2021. "Chapter 19 - Noise pollution and its impact on human health and the environment." In *Pollution Assessment for Sustainable Practices in Applied Sciences and Engineering*, 975-1026. Butterworth-Heinemann - ELSEVIER.

Rahmani, R., M. Aliabadi, R. Golmohammadi, M. Babamiri, and M. Farhadian. 2021. "Evaluation of Cognitive Performance of City Bus Drivers with Respect to Noise and Vibration Exposure." *Acoustics Australia* 49: 529–539.

Ross, Z., I. Kheirbek, J. E. Clougherty, K. Ito, T. Matte, S. Markowitz, and H. Eisl. 2011. "Noise, air pollutants and traffic: Continuous measurement and correlation at a high-traffic location in New York City." *Environmental Research* 111 (8): 1054-1063.

Rossi, I. A., D. Vienneau, M. S. Ragettli, B. Flückiger, and M. Röösli. 2020. "Estimating the health benefits associated with a speed limit reduction to thirty kilometres per hour: A health impact assessment of noise and road traffic crashes for the Swiss city of Lausanne." *Environment International* 145: 106126.

Sagberg, F. 2018. "Characteristics of fatal road crashes involving unlicensed drivers or riders: implications for countermeasures." *Accident Analysis & Prevention* 117: 270-275.

Schlittmeier, S. J., A. Feil, A. Liebl, and J. Hellbrück. 2015. "The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges." *Noise & Health* 17 (76): 148–157.

Smith , A. P. 1988. "Acute effects of noise exposure: an experimental investigation of the effects of noise and task parameters on cognitive vigilance tasks." *International Archives of Occupational and Environmental Health* 60: 307–310.

Sordello, R., F. F. De Lachapelle, B. Livoreil , and S. Vanpeene . 2019. "Evidence of the environmental impact of noise pollution on biodiversity: a systematic map protocol." *Environmental Evidence* 8.

Sørensen, M., Z. J. Andersen, R. B. Nordsborg, T. Becker, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen. 2012. "Long-Term Exposure to Road Traffic Noise and Incident Diabetes: A Cohort Study." *Environ Health Perspect* 121 (2): 217–222.

Swaddle , J. P., C. D. Francis , J. R. Barber, C. B. Cooper, C. C. M. Kyba, D. M. Dominoni, G. Shannon, et al. 2015. "A framework to assess evolutionary responses to anthropogenic light and sound." *Trends in Ecology & Evolution* 30 (9): 550-560.

Wang, T.-C., T.-Y. Chang, R. S. Tyler, B.-F. Hwang, Y.-H. Chen, C.-M. Wu, C.-S. Liu, K.-C. Chen, C.-D. Lin , and M.-H. Tsai. 2021. "Association between exposure to road traffic noise and hearing impairment: a case-control study." *J Environ Health Sci Eng* 19 (2): 1483-1489.

Weyde, K. V., N. H. Krog, B. Oftedal, P. Magnus, S. Øverland, S. Stansfeld, M. J. Nieuwenhuijsen, M. Vrijheid, M. de C. Pascual , and G. M. Aasvang. 2017. "Road traffic noise and children's inattention." *Environmental Health* 16 (127).

WHO Regional Office for Europe. 2018. *Environmental Noise Guidelines for the European Region.* Copenhagen: World Health Organization.

Wolfgang , B. 2008. "Road traffic noise and cardiovascular risk." *Noise & Health* 10 (38): 27-33.

6. Traffic crash injuries occurrence varieties across Barcelona districts

6.1. Abstract

Barcelona (Spain) had a total population of 1,636,762 in 2019 that are distributed on 10 districts across the city. These districts have different characteristics in size and population density. As a result, this can lead to different traffic crashes injuries occurrences and numbers in these areas. Therefore, this study is attempting to determine the conditional probabilities for each district in order to identify the district that has highest number of injuries compared to other areas. A Bayesian network approach is utilized to analyze the dataset and identify the high-risk district alongside analyzing the varieties of traffic crashes during four-year intervals. The results have shown that Eixample, which is the district that has the highest population, highest usage of private transport mode, and highest density of passenger car per km^2 compared to all other districts, has the highest risk of having all types of injuries resulting from traffic crashes. For the temporal factor represented by the four years interval, traffic crashes occurrences varied from district to district based on the level of injury.

***Keywords*:** Barcelona, districts, injuries, traffic crashes, Bayesian network

6.2. Introduction

European countries have a plan that is updated from time to time in order to achieve less traffic crashes injuries with more focusing on fatalities and severe injuries. Figure 16 portrays the change of road fatalities in some European countries by comparing the data between 2010 and 2019 based on a *14th Road Safety Performance Index Report (2020)*. The negative percentages in the figure are referring to a decrease in road fatalities in the designated countries and vice versa. The report showed that the reduction difference for the total number of road deaths in EU27 between 2019 and 2010 is 7,023 deaths. Greece has the highest decreasing percentage with -44.4%, while Malta had the highest increasing percentage of 6.7%. For Spain, the percentage of road fatalities was 30.4% reduced by comparing the two years 2010 and 2019. However, the medium-term target was halving the number of deaths, but as shown, the progress of lessening the number in some countries is stagnated.

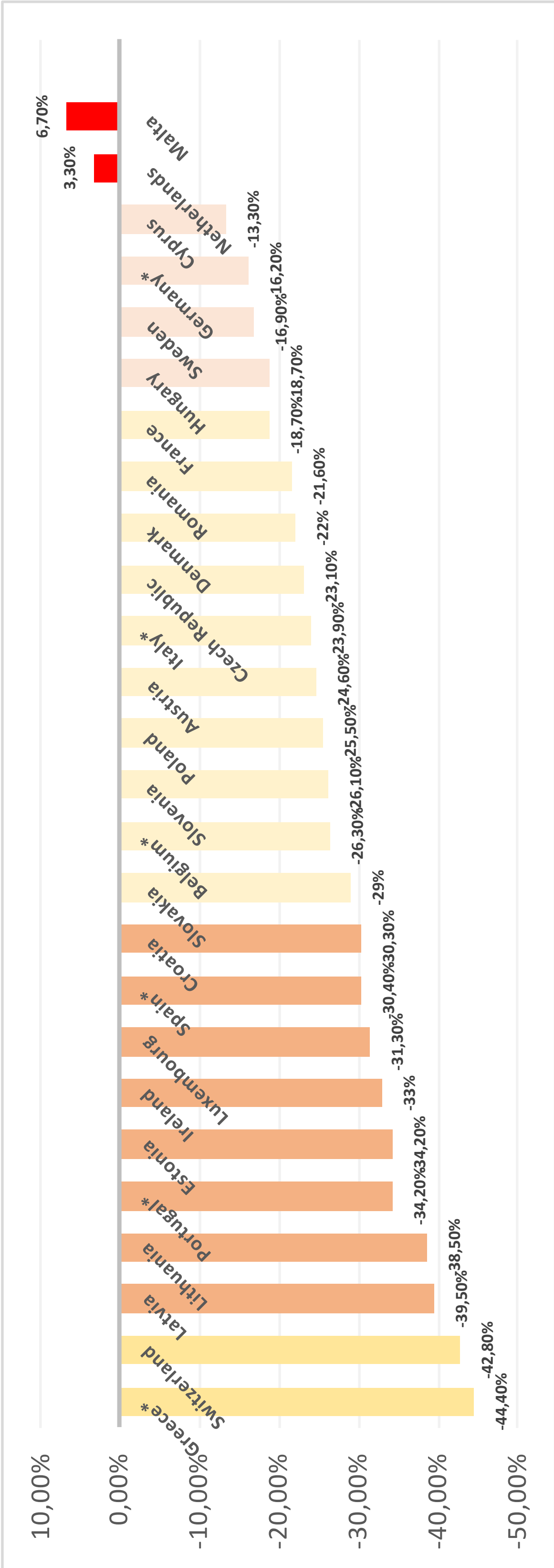


Figure 16 Relative change in road deaths 2010-2019 in Europe
(*Provisional data for 2019 for some indicated countries based on the provided report).

Therefore, many studies have attempted to examine traffic crashes from different perspectives to thoroughly detect and mitigate the risk factors that may impact the level of injury resulted from traffic crashes. In Barcelona, Spain, a recently published study (Aiash and Robusté, 2021A) has examined several risk factors. This study applied two models: a binary probit model and chi-square automatic interaction (CHAID) tree. The total traffic crashes that are examined are 47,153 crashes with including five potential risk factors. The findings showed that elderly and males injured-prone were more likely to have severe or fatal injuries compared to other categories. For the user type, pedestrians and drivers have shown similar trend in having severe of fatal injuries. For the temporal factors, weekends, afternoon, and night timing all are considered as a risk timing due to the likelihood of having severe or fatal injuries compared to other timings. Another recently published study (Aiash and Robusté, 2021B) have conducted an analysis for two other factors that can contribute traffic injuries. These two factors are both temporal factors including the working hours timing period and the season of the year. A Bayesian network is employed for this reason to conduct the analysis process and identify their potential impact of traffic injuries. The results showed that both summer season and working hours timing period were more likely to have higher number of reported injuries.

Bayesian network is one of the utilized approach models to analyze and predict traffic crash data. A study (Alkheder et al., 2020) has employed several data mining techniques to predict and detect the risk factors and their consequential impact. The study found that Bayesian network was accurately predicted the crash data better than CHAID tree and linear support vector machine. The results showed that males drivers, front seat passengers, and elderly drivers were all at higher risk of having severe or fatal injuries compare to other groups. The importance of using the seat belt in case of traffic crash was also asserted. The types of crashes and road classifications were found to have an impact on injuries resulted from traffic crashes. A side from the accurate prediction that Bayesian network can provide, a study (Zhao and Deng, 2015) has deployed the Bayesian network to make inference on crash types at urban intersections. The findings revealed that bicycle and electric bikes were more likely to be involved in frontal collision at urban intersections crashes than small cars and heavy vehicles, whereas, heavy vehicle was more likely to be involved in side collision compared to light vehicles.

Space-temporal factors can have a potential impact on traffic crash injuries. Several research studies are conducted to examine this impact. A study (Wang et al., 2013) have employed a

Bayesian network approach to analyze the spatiotemporal effect of traffic congestion. Fatal and severe injuries were found to be more associated with the increase in traffic congestion. Another research (Jia et al., 2018) found that residential density and bank and hospitals point of interests have an impact on traffic crashes. Consistently, mixed use development, urban residential, single-family residential, multi-family residential, business and, office district all are found to have strong correlations with traffic analysis zones level crashes (Pulugurtha et al., 2013).

6.3. Methodology

6.3.1. Barcelona districts

Barcelona has a total population of 1,636,762 in 2019 that are distributed on 10 districts across Barcelona. These districts are *Ciutat Vella, Eixample, Sants-Montjuïc, Les Corts, Sarrià-Sant Gervasi, Gràcia, Horta-Guinardó, Nou Barris, Sant Andreu,* and *Sant Martí,* as shown in Figure 17. Figure 18 is depicting the road network across Barcelona city with red lines based on (Ajuntament de Barcelona, 2022). Each of these districts is assigned to a different number based on the database source. Table 17 is presenting the detailed population in each district. The highest number of population and density per ha is found in *Eixample,* while the lowest population and density per ha is found in *Les Corts* and *Sarrià-Sant Gervasi,* respectively. This data is gathered from the City Hall of Barcelona database (Estadística i Difusio de Dades, 2019).

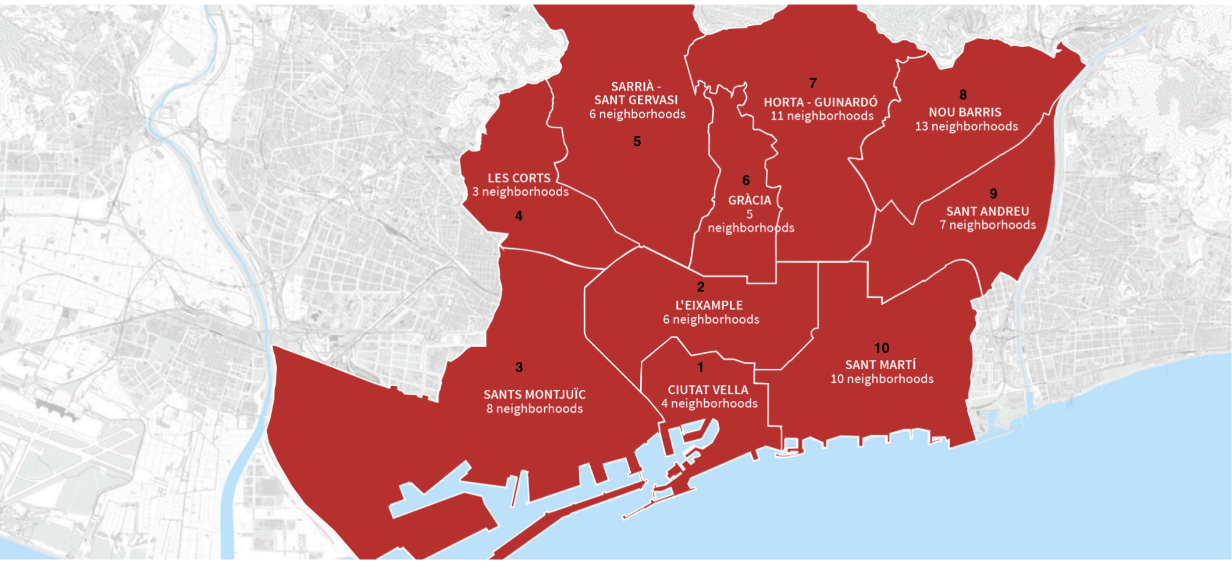


Figure 17 Barcelona districts' location and boundaries
https://www.barcelona.cat/en/living-in-bcn/living-neighbourhood

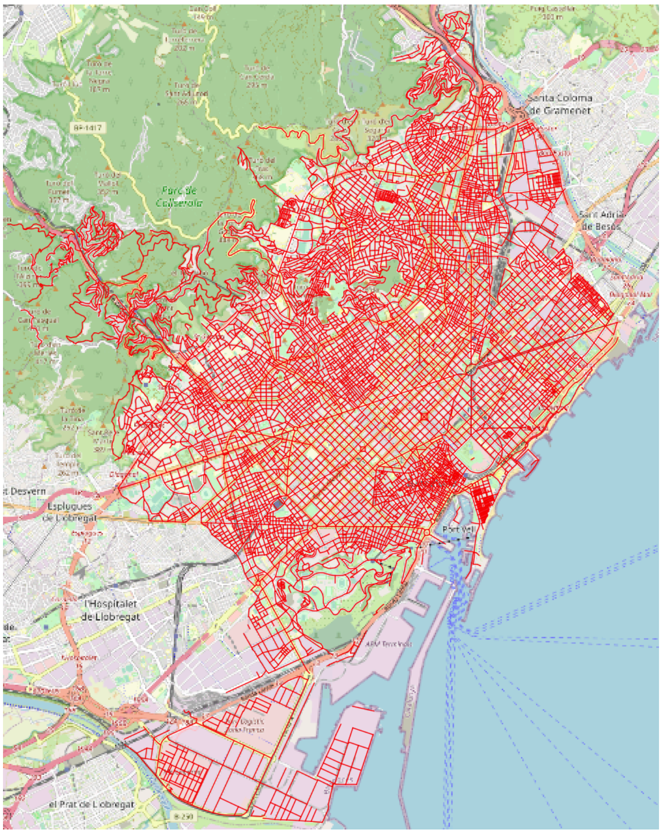


Figure 18 Road network in Barcelona city

Table 17 Barcelona Districts population and size

District	Population	Size in (ha)	Density (inhabitants/ha)
1. Ciutat Vella	103,429	412.1	251
2. Eixample	**265,910**	746.4	**356**
3. Sants-Montjuïc	184,091	2,267.6	81
4. Les Corts	81,974	601.1	136
5. Sarrià-Sant Gervasi	149,260	1,991.6	75
6. Gràcia	121,798	422.4	288
7. Horta-Guinardó	171,495	1,191.9	144
8. Nou Barris	170,669	805.1	212
9. Sant Andreu	149,821	658.9	227
10. Sant Martí	238,315	1,038.6	229

A recently published Statistical Yearbook of the City of Barcelona (Oficina Municipal de Dades, 2021) has included the indicators and vehicle stats across the different districts of Barcelona. Table 18 is presenting passenger car distributions across Barcelona districts. It can be noticed that *Eixample* has the highest number of passenger cars with having the highest density of passenger cars per km^2. Table 19 is showing the number of registered vehicles in Barcelona districts in 2020. Similar to previously noticed statistics, *Eixample* has the highest number of all types of registered vehicles except for registered mopeds and trucks. Consistently, table 20 is presenting the usage of different modes of transport in Barcelona districts in 2020.

It can be also noticed that *Eixample* has the highest usage of both on foot and bicycle mode of transport and private transport mode, while Sant Marti has the highest usage of public transport compared to other districts.

Table 18 Density Indicators of passenger cars in 2020

District	Passenger car	Surface (km^2)	Passenger car / km^2
Ciutat Vella	17.679	4.2	4.204
Eixample	**77.912**	7.5	**10.438**
Sants-Montju'ic	49.494	22.9	2.163
Les Carts	31.393	6.0	5.223
Sarria-Sant Gervasi	59.474	19.9	2.986
Gracia	33.709	4.2	7.980
Horta-Guinard6	49.751	11.9	4.174
Nou Barris	45.792	8.1	5.684
Sant Andreu	44.922	6.6	6.815
Sant Marti	68.992	10.4	6.610

Table 19 Vehicles registered by typology and district in 2020

District	Passenger cars	Motorcycles	Mopeds	Vans	Trucks	Other vehicles
Ciutat Vella	624	554	1.135	92	33	27
Eixample	**3.227**	**2.194**	1.171	**236**	52	**190**
Sants-Montju'ic	1.962	1.054	310	184	**53**	112
Les Carts	1.369	700	36	33	12	11
Sarria-Sant Gervasi	2.363	1.846	101	75	22	60
Gracia	1.278	1.002	103	53	10	17
Horta-Guinard 6	1.887	1.307	104	126	27	26
Nou Barris	1.645	826	78	114	25	34
Sant Andreu	2.018	822	56	142	25	34
Sant Marti	2.652	1.358	**2.583**	200	29	58

Table 20 Transport type usage by district in 2020

District	Total	By foot and Bicycle	Private transport	Public transport
Ciutat Vella	333.300	253.308	25.329	54.662
Eixample	**802.039**	**544.185**	**147.750**	110.104
Sants-Montju"ic	500.419	294.860	97.666	107.894
Les Corts	264.350	176.951	53.956	33.444
Sarria-Sant Gervasi	455.135	259.254	136.183	59.699
Gracia	380.898	250.154	68.424	62.320
Horta-Guinard6	490.973	271.132	108.942	110.899
Nou Barris	526.307	294.752	108.715	122.840
Sant Andreu	460.405	289.878	88.952	81.574
Sant Marti	713.044	435.258	137.993	**139.793**

6.3.2. Bayesian network

In this study, a Bayesian network is employed to analyze the utilized data that is gathered from *the City Hall of Barcelona* (Barcelona's City Hall Open Data Service, 2019). Tree Augmented Naïve Bayes technique is used through the IBM Watson Studio platform that is developed based on *Bayesian Network Classifiers* (Friedman et al., 1997). Two predictors are analyzed including all districts of Barcelona and a range of years that extended from 2016 to 2019. The dependent variable is the level of injury which consists of slight injury and severe or fatal injury. The total number of utilized crashes in this study is 47,081. The undefined areas or values are eliminated from the final utilized data. The conditional probabilities are determined as shown in the following equation and similar to a previously mentioned and conducted study (Aiash and Robusté, 2021). More details about the applied model can be found at IBM report (2016):

$$Pr\left(Y_i|X_1=x_1^j,X_2=x_2^j,\dots,X_n=x_n^j\right)=\frac{Pr(Y_i)\left(X_1=x_1^j,X_2=x_2^j,\dots,X_n=x_n^j|Y_i\right)}{Pr\left(X_1=x_1^j,X_2=x_2^j,\dots,X_n=x_n^j\right)}$$
$$\propto Pr\ (Y_i)\prod_{k=1}^{n}Pr\left(X_k=x_k^j\middle|\pi_k^j,Y_i\right) \tag{1}$$

d is the data set. the case that is classified to which it belongs to i^{th} target category is d_j. The target Y_i is level of injury including two categories: slight injury and severe or fatal injury. The independent variable is x. The number of independent variables which is two in this study is n. As a result, $d_j=\left(x_1^j,x_2^j,\dots,x_n^j\right)$. Non-redundant parameters number is K. The parent set of the independent variables alongside the dependent variable is π_k, it maybe empty for Tree Augmented Naïve Bayes. The conditional probability is $\Pr(X_k = x_k^j | \pi_k^j, Y_i)$ related to the developed two nodes.

6.4. Results and discussion

Figure 19 is showing the structure of the applied TAN model. It consists of three nodes starting with the injury level which is the dependent variable followed by district and year nodes. District node is including the ten different aforementioned districts in Barcelona, while the year node is including 4 years which they are 2016, 2017, 2018, and 2019. Based on the applied model, district as an independent variable is more important than year variable in predicting the utilized data. The prediction accuracy for the applied model is found to be for the training and testing is 98.01% and 98%, respectively.

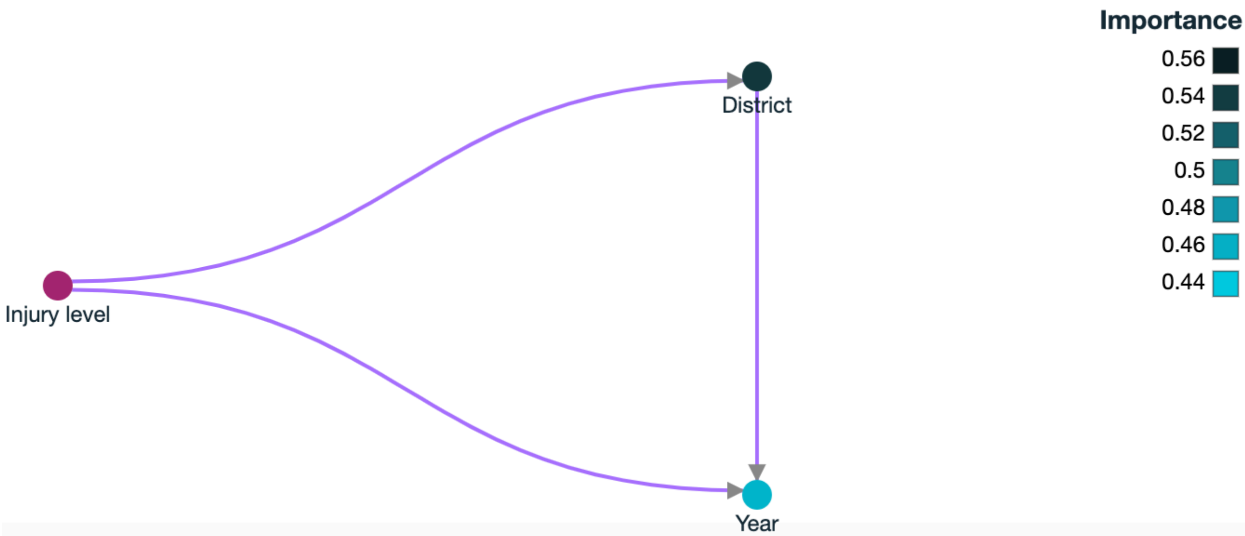


Figure 19 Structure of the applied TAN model

Table 21 and Table 22 are showing the determined conditional probabilities for both nodes of the independent variables. Table 21 is presenting the probabilities for the districts. As shown, *Eixample* has the highest probabilities compared to all other districts which indicate that both slight, severe, and fatal injuries are more likely to happen in this district. *Sant Martí* is following *Eixample* in having the highest probabilities, followed by *Sants-Montjuïc* and *Sarrià-Sant Gervasi.* It is worth mentioning that the differences between the last mentioned three districts is slightly low and training and testing sample should be considered as these results are representing 70 % of the utilized dataset, while the testing set represents 30%.

Table 21 Calculated conditional probabilities for districts

Parents	Probability										
Injury level	Ciutat Vella	Eixample	Gràcia	Sants-Montjuïc	Sarrià-Sant Gervasi	Horta-Guinardó	Les Corts	Sant Martí	Sant Andreu	Nou Barris	Total
Slight Injuries	0.051	**0.304**	0.046	0.111	0.107	0.07	0.073	0.125	0.062	0.051	32,482
Severe or fatal injuries	0.048	**0.273**	0.026	0.102	0.108	0.064	0.092	0.145	0.067	0.076	660

Table 22 is depicting the calculated conditional probabilities for different years with having two parents based on the applied model which they are the district and the dependent variable which is the level of injury. For *Ciutat Vella* as a parent node, all types of injuries have higher probabilities in 2018 compared to all other years. While for Eixample**,** slight injuries are higher in 2016, and severe or fatal injuries are higher in 2018 compared to other years. *Gràcia* has the highest probabilities in 2017 for all types of injuries. Slight injuries are higher in 2019 while severe and fatal injuries are higher in 2016 in Sants-Montjuïc district. Sarrià-Sant Gervasi has higher number of slight injuries in 2018, however, severe or fatal injuries are found to be high at three different years 2016, 2017 and 2018. For Horta-Guinardó, 2016 has the highest number of slight injuries and 2019 has the highest number of severe or fatal injuries. Les Corts has the highest number of slight injuries in 2016 and the highest number of severe or fatal injuries in 2017. In 2019, the highest number slight injuries is identified in Sant Martí, while in 2017, the highest number of severe or fatal injuries is identified at the same district. For Sant Andreu, 2019 has the highest probabilities for all types of injuries. Nou Barris has its highest number of slight injuries during 2018 and its highest number of severe or fatal injuries during 2016.

Table 22 Calculated conditional probabilities for years

Parents		Probability				
District	Injury level	2016	2017	2018	2019	Total
Ciutat Vella	Slight Injuries	0.225	0.253	**0.259**	0.262	1,655
Ciutat Vella	Severe or fatal injuries	0.25	0.188	**0.344**	0.219	32
Eixample	Slight Injuries	**0.262**	0.239	0.245	0.254	9,862
Eixample	Severe or fatal injuries	0.233	0.222	**0.306**	0.239	180
Gràcia	Slight Injuries	0.249	**0.257**	0.250	0.244	1,487
Gràcia	Severe or fatal injuries	0.118	**0.471**	0.118	0.294	17
Sants-Montjuïc	Slight Injuries	0.258	0.238	0.242	**0.262**	3,591
Sants-Montjuïc	Severe or fatal injuries	**0.328**	0.239	0.224	0.209	67
Sarrià-Sant Gervasi	Slight Injuries	0.241	0.246	**0.267**	0.247	3,479
Sarrià-Sant Gervasi	Severe or fatal injuries	**0.254**	**0.254**	**0.254**	0.239	71
Horta-Guinardó	Slight Injuries	**0.273**	0.240	0.256	0.231	2,283
Horta-Guinardó	Severe or fatal injuries	0.167	0.214	0.262	**0.357**	42
Les Corts	Slight Injuries	**0.256**	0.252	0.253	0.238	2,386
Les Corts	Severe or fatal injuries	0.295	**0.328**	0.197	0.180	61
Sant Martí	Slight Injuries	0.255	0.255	0.232	**0.258**	4,076
Sant Martí	Severe or fatal injuries	0.167	**0.365**	0.302	0.167	96
Sant Andreu	Slight Injuries	0.254	0.224	0.260	**0.262**	2,005
Sant Andreu	Severe or fatal injuries	0.205	0.182	0.250	**0.364**	44
Nou Barris	Slight Injuries	0.252	0.258	**0.262**	0.229	1,658
Nou Barris	Severe or fatal injuries	**0.360**	0.220	0.220	0.200	50

6.5. Conclusions

Traffic crashes intensity or occurrence can vary greatly from region to region or from district to district. Therefore, this spatial influence on traffic crashes can also impact the level of injury and its types. In Barcelona, 10 different districts with different populations and size that are distributed across the city. Consequently, this study is attempting to classify these districts according to traffic crashes and their resulted injuries whether it is a slight injury, severe, or fatal injury. A Bayesian network represented by TAN technique is employed in order to analyze the utilized data that consists of 47,081 traffic crashes that resulted into different level of injuries during four years interval including 2016, 2017, 2018, and 2019.

The results have shown that *Eixample* has the highest probabilities of having all types of injuries compared to all other districts. This could be a result of a combination of factors that can lead to this status. One of these factors is, the fact that, *Eixample* has the highest density per hectare and population compared to all other districts; street traffic is also high on the grid streets designed by Civil Engineer Ildefons Cerdà. Additionaly, *Eixample* has the highest figures related to owned passenger cars, motorcycles and vans, highest usage of private transport mode, and the highest density of passenger cars per km^2compared to all other districts. For the temporal factor represented by years, the number of injuries vary significantly from district to district during different years as there is no one single year that has the highest number of injuries for all districts. For the future work, an extensive analysis can be carried out related to traffic crashes in *Eixample* in order to identify risk factors. These risk factors can include spatiotemporal factors, average daily traffic, person characteristics who was involved in the crash, type of the crash, and the transport mode that led to the crash.

6.6. References

Aiash, A., and F. Robusté. 2021A. "Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree." *International Journal of Injury Control and Safety Promotion.*

Aiash, A., and F. Robusté. 2021B. "Working hours and traffic accident injuries: Case study in Barcelona." *Transportation Research Procedia* 58: 678-682.

Ajuntament de Barcelona. 2019. *Barcelona's City Hall Open Data Service.* https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=accident&sort=fecha_publicacion+desc.

Ajuntament de Barcelona. 2019. *Estadística i Difusió de Dades.* https://ajuntament.barcelona.cat/estadistica/castella/Territori/sup/a2019/S0202.htm.

Ajuntament de Barcelona. 2022. *Barcelona's City Hall Open Data Service.* https://opendata-ajuntament.barcelona.cat/data/en/dataset/mapa-graf-viari-carrers-wms.

AlKheder, S., F. AlRukaibi, and A. Aiash. 2020. "Risk analysis of traffic accidents' severities: An application of three data mining models." *ISA Transactions* 106: 213-220.

Carson, J., D. Adminaité-Fodor, and G. Jost. 2020. *14th Road Safety Performance Index Report.* Brussels: ETSC.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. "Bayesian Network Classifiers." *Machine Learning* 29: 131–163.

IBM. 2016. *IBM SPSS Modeler 18.0 Algorithms Guide.* IBM Corporation.

Jia, R., A. Khadka, and I. Kim. 2018. "Traffic crash analysis with point-of-interest spatial clustering." *Accident Analysis & Prevention* 121: 223-230.

Oficina Municipal de Dades. 2021. *Anuari Estadístic de la Ciutat de Barcelona 2021.* Barcelona: Ajuntament de Barcelona.

Pulugurtha, S. S., V. R. Duddu, and Y. Kotagiri. 2013. "Traffic analysis zone level crash estimation models based on land use characteristics." *Accident Analysis & Prevention* 50: 678-687.

Wang, C., M. Quddus, and S. Ison. 2013. "A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK." *Transportmetrica A: Transport Science* 9 (2): 124-148.

Zhao, J., and W. Deng. 2015. "The use of Bayesian network in analysis of urban intersection crashes in China." *Transport* 30: 411-420.

7. Analyzing pedestrians' crash injury risk factors in Barcelona

7.1. Abstract

Pedestrian crashes can be affected by different factors that can increase or decrease their possibilities of occurring and other consequences related to them, including the level of injury. In Barcelona, in particular, and in Spain, in general, several events happened during both the years 2020 and 2021. The first year had the pandemic and different events that affected the number of traffic crashes, similar to the following year, which had the introduction of applying the speed limit of 30 km/h inside the cities. Therefore, this study analyzes several risk factors related to pedestrian crashes during both the 2020 and 2021 years. A Bayesian network model is employed for this reason to determine the impact of those risk factors through these different years. This is done by calculating the conditional probabilities for the risk factors based on the applied model with determining the accuracy of the employed model in predicting the utilized data. The results show that elderly pedestrians are more likely to have severe or fatal injuries. The evening period shows higher probabilities of having pedestrian crash injuries compared to other periods of the day. Passage for pedestrians that are regulated by traffic lights is found to have a high probability of having severe and fatal injuries.

Keywords: Pedestrian, Risk factors, Injury, Bayesian network, Barcelona

7.2. Introduction

Due to the fact that traffic crashes are still representing a threat to the lives of road users (Aiash and Robusté, 2021A), certain road users suffer more from the consequences of these crashes. These vulnerable road users are more prone to traffic crash injuries compared to other road users. Based on (WHO, 2022), more than half of the 1.3 million fatalities that are resulted from traffic crashes around the world are for vulnerable road users, including pedestrians, cyclists, and motorcyclists. Several factors, on the other hand, can increase the possibility of having more severe injuries that are resulted from these crashes. Risk factors can include the characteristics of the vulnerable road users such as their gender and their age category, the temporal factors represented by the time when the crash occurred, and the spatial characteristics where the crash happened.

Therefore, several studies have focused on this category of road users in order to prevent or lessen the number of traffic crashes that are causing different levels of injuries so that the number of fatalities can be reduced. For age categories, children and elderly road users showed a higher number of fatalities compared to other groups (Chang, 2008) (Ponnaluri and Nagar, 2010). The correlation between the age of the pedestrian and the severity of the crash is also confirmed by another study (HU et al., 2020). For the gender of the injured pedestrian, males represented the majority of pedestrian fatalities compared to females in the U.S. (Clifton and Livi, 2005). Another study (Sun et al., 2019) examined several risk factors related to pedestrian crashes by employing the latent class clustering method and multinomial logit regression. The results have shown that pedestrians with involvement in drugs or alcohol were more likely to have severe and fatal injuries, similar to elderly pedestrians who were older than 65 years old and in adverse weather conditions. For other findings of the same study, pedestrian crash fatalities can be highly impacted when crossing a road far from an intersection. Speed limit (Munira et al., 2020) and the approach speed of the vehicle toward a pedestrian (Chakraborty et al., 2019) has an impact on increasing the possibility of fatal injuries for pedestrians. Additionally, periods of the day can have a different impact on injury severity (Pahukula et al., 2015). For socioeconomic status, the low socioeconomic status of the pedestrian was found to increase the odds of having injuries and fatalities (Cubbin and Smith, 2002) (Azetsop, 2010). The characteristics of the area and road can also impact the severity of pedestrian traffic crashes. A study (Zajac and Ivan, 2003) examined the impact of road features and the types of areas in a rural area. The results have shown that clear road width, village, downtown fringe, and low-density residential areas had a higher level of traffic crash severities.

This study attempts to analyze several risk factors that are related to traffic crashes that occurred for pedestrians in the city of Barcelona by utilizing traffic crash data that is provided by the open data service of the city during the years 2020 and 2021. Slight, severe, and fatal injuries are included to be analyzed. A Bayesian network model is then employed to determine the conditional probabilities for the risk factors in order to grasp their impact on the level of injuries that occurred for pedestrians. The remainder of the paper is organized as follows: the methodology section is set to be section two. Section three presents the results. Lastly, the conclusions are depicted in section four.

7.3. Methodology

7.3.1. Data description

Barcelona's City Hall Open Data Service is the database that is utilized in this study to gather data related to pedestrian crashes that occurred in Barcelona city (Ajuntament de Barcelona's Open Data Service, 2021). In figure 20, the total number of traffic crashes and different levels of injuries are depicted. The consistency of pedestrian crash numbers can be seen through the years from 2011 to 2019. Then, in 2020, the COVID-19 pandemic started with a variety of restrictions that limited the movement of road users that, resulted in a drastic decrease in the number of pedestrian crashes with different consequential levels of injuries. However, the decrement stopped in 2021 as the total number of these crashes started to increase again.

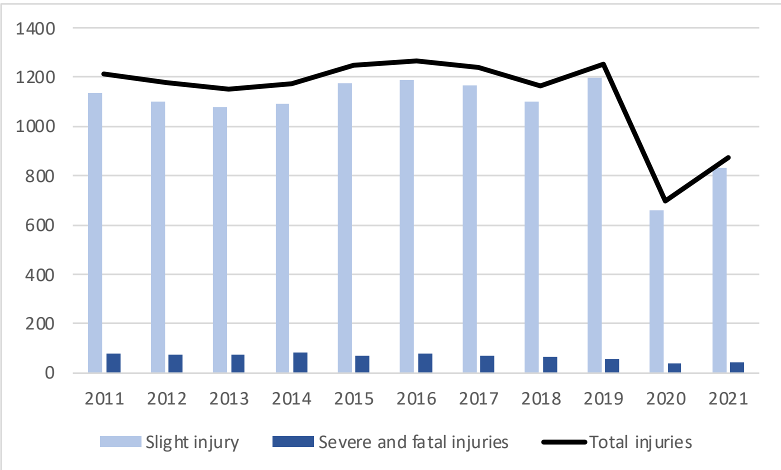


Figure 20 Pedestrian injuries across 10 years interval in Barcelona

The risk factors that are included to be analyzed with the employed model are shown in table 23 and table 24. It is worth mentioning that all traffic crashes with their risk factors happened during 2020 without merging 2020 year with 2021 due to the differences between the pandemic year as many events occurred in that year when the movement restrictions were imposed because of the pandemic. The daytime variable consists of three classes: morning, evening, and night timing, based on the database. The morning period is set to be from 6:00 to 13:00. The evening period is set to be from 14:00 to 21:00. The night period is set to be from 22:00 to 5:00. The gender includes male and female. The age category of the injured person includes the 0-15 group, 16-24 group, 25-40 group, 41- 64 group, and lastly, 65 and older age group for the pedestrian who got injuries resulting from the traffic crash. The last risk factor is the crash site characteristics, including six different sites: a crash on the sidewalk/platform, a crash on the passage that is regulated by the traffic light with having both horizontal and vertical signs, a crash on a passage with having horizontal signs but without having vertical signs, a crash on a passage without traffic lights but with having other vertical and horizontal signs, and last but not least, a crash outside the designated passage area for pedestrian.

Table 23 2020 pedestrian crash risk factors descriptive data during 2020

Risk factor	Risk factor class	Slight injury	Severe or fatal injury
Day time	Morning	255	8
	Evening	323	25
	Night	37	2
Gender	Male	344	13
	Female	271	22
Age	0-15	91	1
	16-24	56	2
	25-40	112	6
	41-64	210	14
	65+	146	12
Site of crash characteristics	On the sidewalk / Platform	78	1
	Passage regulated by traffic lights	271	18
	In a pedestrian area	28	0
	Passage without vertical signs	10	1
	Passage without traffic lights	86	4
	Outside the designated passage area	142	11

Table 24 2021 pedestrian crash risk factors descriptive data during 2020

		Slight injury	Severe or fatal injury
Day time	Morning	315	15
	Evening	411	22
	Night	27	4
Gender	Male	438	21
	Female	315	20
Age	0-15	98	9
	16-24	76	5
	25-40	161	8
	41-64	236	8
	65+	182	11
Site of crash characteristics	On the sidewalk / Platform	98	4
	Passage regulated by traffic lights	317	14
	In a pedestrian area	53	0
	Passage without vertical signs	7	1
	Passage without traffic lights	91	5
	Outside the designated passage area	187	17

7.3.2. Bayesian network

In this study, a Bayesian network (BN) is employed to determine the conditional probabilities between the risk factors and the dependent variable represented by the injury level of the pedestrian. For the given random variables, BN can provide laconic descriptions for the joint distribution of probability as it is considered a member of probabilistic graphical models. In this study and by following previous proceedings (Aiash and Robusté, 2021B) in employing BN, Tree Augmented Naïve Bayes is employed using the IBM Watson Studio software platform with utilizing SPSS modeler to classify the correlation between the four predictors and the level of injury that has happened to the pedestrians that are involved in traffic crashes. The training and testing set size partitions for the 2020 and 2021 years are set to be 70% and 30%, respectively. The classifier of Tree Augmented Naïve Bayes that is employed when utilizing IBM Watson Studio software with utilizing SPSS modeler is based on this book (Friedman et al., 1997). The conditional probabilities are determined as follows:

$$Pr\left(Y_i|X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j\right) = \frac{Pr(Y_i)\left(X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j|Y_i\right)}{Pr\left(X_1 = x_1^j, X_2 = x_2^j, \ldots, X_n = x_n^j\right)}$$

$$\propto \Pr\,(Y_i) \prod_{k=1}^{n} Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right) \tag{1}$$

For $d_j = \left(x_1^j, x_2^j, \ldots, x_n^j\right)$, the data set is d . The case is d_j. This case is classified to which it belongs to i^{th} dependent category. In this study, the target or dependent variable Y_i is the level of injury that consists of two categories including slight injury and severe or fatal injury. The predictor is x. The number of predictors is n. In this study, n is four. The number of non-redundant parameters is K. The parent set of the independent variable is π_k. The conditional probability for each node is represented by $Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right)$. In this study, there are four nodes for predictors.

7.4. Results and discussion

7.4.1. BN results for 2020 year

For the year 2020, the employed BN is shown in figure 21. The training and testing accuracy for the employed BN is 95.08% and 94.09%, respectively. All of the risk factors are included in the BN by providing their importance values. The most important risk factor based on this employed BN is the site of crash characteristics, while the least important risk factor is the age with its different included categories.

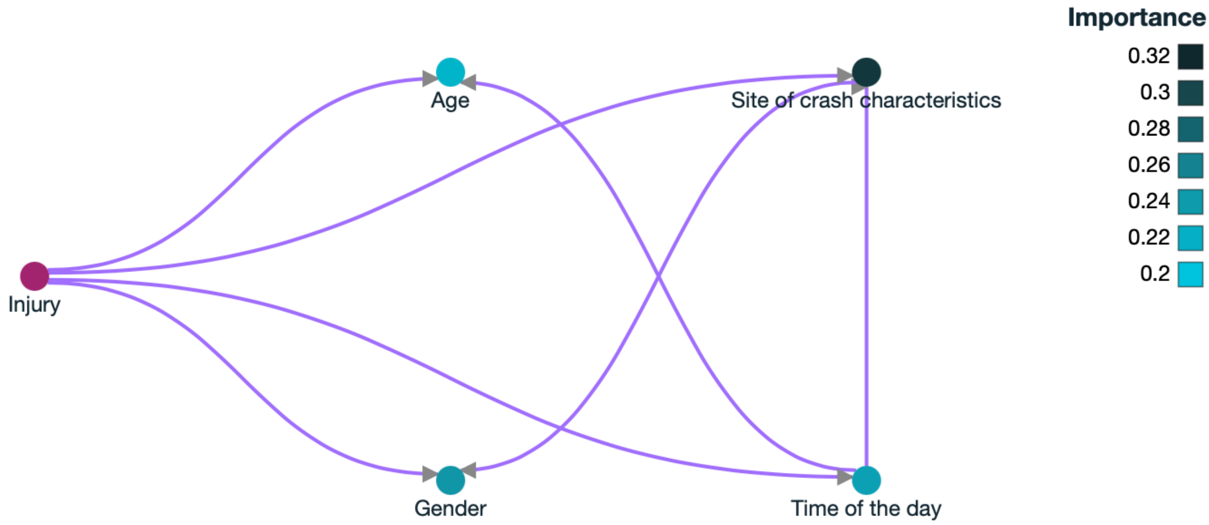


Figure 21 The structure of the applied BN for 2020

Tables 25, 26, 27, and 28 show the determined conditional probabilities for the chosen risk factors with respect to the dependent variable, that is, the level of injury. Table 25 shows the results for the age factor with two parents' nodes, including the time of the day and the injury level. For the morning period, severe and fatal injuries are more likely to occur to pedestrians who have an age between 41 years old and older compared to other categories. For slight injuries during the morning period, pedestrians who are 65 years old and older are more likely to have this type of injury compared to other groups. For the evening period, the 65 and older age group is found to have higher odds of having severe and fatal injuries. For slight injuries, the age group between 41 and 64 years old is found to have the highest probability. For the night period, younger age groups are more involved in pedestrian crashes. For severe and fatal injuries, the age group between 25 and 40 years old is more likely to have this type of injury compared to other groups. The age group 41 and 64 years old are more likely to have slight injuries during the night timing period.

Table 25 Age conditional probabilities based on the applied BN for 2020

Parents		Age				
Time of the day	Injury	0-15	16-24	25-40	41-64	65+
Morning	Severe or fatal injury	0.000	0.000	0.000	**0.500**	**0.500**
Morning	Slight injury	0.131	0.057	0.159	0.324	**0.330**
Evening	Severe or fatal injury	0.059	0.118	0.176	0.294	**0.353**
Evening	Slight injury	0.185	0.104	0.180	**0.324**	0.207
Night	Severe or fatal injury	0.000	0.000	**1.000**	0.000	0.000
Night	Slight injury	0.120	0.320	0.200	**0.360**	0.000

Table 26 shows the results for gender as a risk factor in 2020 year. The parents' nodes for this risk factor are the site crash characteristics and injury. Males are found to have higher probabilities of having all types of injuries in case of a traffic crash on the sidewalk compared to females. For the crashes that are happening on a passage regulated by traffic lights, females are more likely to have severe and fatal injuries there compared to males, while males are more likely to have slight injuries. Females are found to have higher odds of having slight injuries inside pedestrian areas. For passage without vertical signs, females are more likely to have severe or fatal injuries compared to males, who are more likely to have slight injuries. For passage without traffic lights, males are found to have higher conditional probabilities for all types of injuries. An interesting finding, females are found to be more involved in pedestrian traffic crashes outside the designated passage area.

Table 26 Gender conditional probabilities based on the applied BN for 2020

Parents		Gender	
Site of crash characteristics	Injury	Female	Male
On the sidewalk / Platform	Severe or fatal injury	0.000	**1.000**
On the sidewalk / Platform	Slight injury	0.283	**0.717**
Passage regulated by traffic lights	Severe or fatal injury	**0.667**	0.333
Passage regulated by traffic lights	Slight injury	0.400	**0.600**
In a pedestrian area	Slight injury	**0.524**	0.476
Passage without vertical signs	Severe or fatal injury	**1.000**	0.000
Passage without vertical signs	Slight injury	0.400	**0.600**
Passage without traffic lights	Severe or fatal injury	0.500	**0.500**
Passage without traffic lights	Slight injury	0.456	**0.544**
Outside the designated passage area	Severe or fatal injury	**0.750**	0.250
Outside the designated passage area	Slight injury	**0.569**	0.431

Table 27 presents the results for the site of crash characteristics risk factor with two parents' nodes: the time of the day and the injury level. Surprisingly, passages regulated by traffic lights show the highest odds for having all types of injuries that are resulted from traffic crashes during different day timings except for night timing as on the sidewalk has the highest odds for having severe or fatal injuries. Table 28 shows the time of day as a risk factor with only one parent node, that is, the injury level. The evening period has the highest odds of having all types of injuries from traffic crashes for pedestrians compared to other day timings.

Table 27 Site of crash characteristics conditional probabilities based on the applied BN for 2020

Parents		Site of crash characteristics					
Time of the day	Injury	On the sidewalk / Platform	Passage regulated by traffic lights	In a pedestrian area	Passage without vertical signs	Passage without traffic lights	Outside the designated passage area
Morning	Severe or fatal injury	0.000	**0.500**	0.000	0.000	0.167	0.333
Morning	Slight injury	0.085	**0.426**	0.051	0.023	0.131	0.284
Evening	Severe or fatal injury	0.000	**0.529**	0.000	0.059	0.059	0.353
Evening	Slight injury	0.167	**0.446**	0.054	0.005	0.122	0.207
Night	Severe or fatal injury	**1.000**	0.000	0.000	0.000	0.000	0.000
Night	Slight injury	0.040	**0.440**	0.000	0.000	0.280	0.240

Table 28 Time of the day conditional probabilities based on the applied BN for 2020

Parents	Time of the day		
Injury	Morning	Evening	Night
Severe or fatal injury	0.250	**0.708**	0.042
Slight injury	0.416	**0.525**	0.059

7.4.2. BN results for 2021

Similar to the 2020 year, the structure of BN for the 2021 year is depicted and shown in figure 22. The prediction accuracy for the training and testing sets are 95.32% and 93.70%, respectively. The structure is also based on the same risk factors as explained before. Eventually, four nodes represent those risk factors, and the fifth node is the injury level which is the dependent variable. Based on the applied BN, the most important risk factor is the age of the injured pedestrian, while the least important variable is gender.

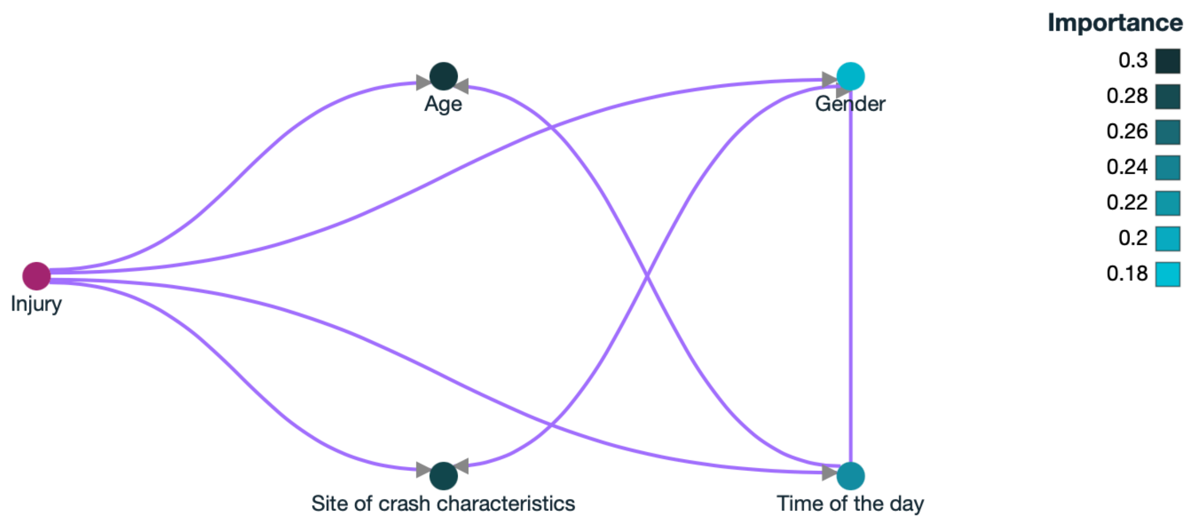


Figure 22 The structure of the applied BN for 2021 year

Tables 29, 30, 31, and 32 provide the results for the determined conditional probabilities for the employed BN for the year 2021. The results for the age category in table 29 show similar results compared to the year 2020 during the morning period, as 41 years old and older pedestrians are more likely to have severe or fatal injuries, 65 years old and older pedestrians are more likely to have slight injuries. While for the evening period, the results show that the age category from 0 to 15 years old is more involved in severe or fatal injuries, while the 41 to 64 years old age category is more involved in slight injuries during the same time of the day. The night timing period results show that 65 years old and older pedestrians can have higher odds of having severe or fatal injuries, while 41 to 64 years old age group pedestrians show higher probabilities of having slight injuries.

Table 29 Age conditional probabilities based on the applied BN for 2021 year

Parents		Age				
Time of the day	Injury	0-15	16-24	25-40	41-64	65+
Morning	Severe or fatal injury	0.100	0.100	0.200	**0.300**	**0.300**
Morning	Slight injury	0.077	0.047	0.189	0.326	**0.361**
Evening	Severe or fatal injury	**0.313**	0.063	0.250	0.188	0.188
Evening	Slight injury	0.175	0.124	0.240	**0.287**	0.175
Night	Severe or fatal injury	0.000	0.000	0.000	0.000	**1.000**
Night	Slight injury	0.000	0.238	0.238	**0.476**	0.048

Table 30 provides the conditional probabilities for the gender risk factor with having two parents' nodes: time of the day and injury. Males are showing higher possibilities of having slight, severe, and fatal injuries compared to females during all periods of the day except during the night, as females are more likely to have severe or fatal injuries compared to males.

Table 30 Gender conditional probabilities based on the applied BN for 2021 year

Parents		Gender	
Time of the day	Injury	Female	Male
Morning	Severe or fatal injury	0.400	**0.600**
Morning	Slight injury	0.365	**0.635**
Evening	Severe or fatal injury	0.438	**0.563**
Evening	Slight injury	0.440	**0.560**
Night	Severe or fatal injury	**1.000**	0.000
Night	Slight injury	0.333	**0.667**

Table 31 presents the conditional probabilities for the site of crash characteristics risk factor with two parents' nodes: gender and injury nodes. Passage regulated by traffic lights has the highest odds for passage regulated by traffic lights has the highest odds for females for all types of injuries. Table 32 depicts the results for the time of the day risk factor with only one parent node, that is, the injury. The evening period shows the highest probability of having all types of injuries for pedestrians compared to other periods of the day, which is similar to the year 2020.

Table 31 Site of crash characteristics conditional probabilities based on the applied BN for 2021 year

Parents		Site of crash characteristics					
Gender	Injury	On the sidewalk / Platform	Passage regulated by traffic lights	In a pedestrian area	Passage without vertical signs	Passage without traffic lights	Outside the designated passage area
Female	Severe or fatal injury	0.083	**0.500**	0.000	0.000	0.000	0.417
Female	Slight injury	0.141	**0.413**	0.066	0.014	0.108	0.258
Male	Severe or fatal injury	0.133	0.200	0.000	0.067	0.200	**0.400**
Male	Slight injury	0.120	**0.399**	0.066	0.009	0.158	0.247

Table 32 Time of the day conditional probabilities based on the applied BN for 2021 year

Parents	Time of the day		
Injury	Morning	Evening	Night
Severe or fatal injury	0.370	**0.593**	0.037
Slight injury	0.440	**0.520**	0.040

7.5. Conclusions

Vulnerable road users show a higher necessity to be protected compared to other road users due to the fact that they are more prone to traffic crash injuries. Part of the vulnerable road users are the pedestrians, that can be involved in traffic crashes with severe or fatal injuries. Therefore, this study examines different risk factors that can contribute to traffic crash injuries related to pedestrians. The data that is utilized is attributed to Barcelona city with implementing two years, including 2020 and 2021 years. Then, BN models are employed to determine the conditional probabilities between these chosen risk factors and the dependent variable, which is the level of injury resulting from traffic crashes, besides presenting the accuracy of the employed models in predicting the utilized data.

The results show similar findings for both years that are included in this study. Elderly pedestrians show higher probabilities of having severe or fatal injuries during the morning period. Passages regulated by traffic lights are showing a higher number of traffic crash injuries compared to other sites of pedestrian crashes. For the period of the day, the evening period is another risk factor for pedestrian traffic crash injuries. Both BN models that are employed for both years show a high prediction accuracy.

7.6. References

Aiash, A., and F. Robusté. 2021A. "Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree." *International Journal of Injury Control and Safety Promotion* 29 (2): 256-264.

Aiash, A., and F. Robusté. 2021B. "Working hours and traffic accident injuries: Case study in Barcelona."*Transportation Research Procedia* 58: 678-682.

Ajuntament de Barcelona. 2021. *Ajuntament de Barcelona's open data service.* https://opendata-ajuntament.barcelona.cat/en/.

Azetsop, J. 2010. "Social Justice Approach to Road Safety in Kenya: Addressing the Uneven Distribution of Road Traffic Injuries and Deaths across Population Groups." *Public Health Ethics* 3 (2): 115–127.

Chakraborty, A., Mukherjee, D., & Mitra, S. 2019. "Development of pedestrian crash prediction model for a developing country using artificial neural network." *International Journal of Injury Control and Safety Promotion, 26*(3), 283-293.

Chang, D. 2008. *National Pedestrian Crash Report (DOT HS 810 968).* Washington, DC: National Highway Traffic Safety Administration.

Clifton, K., and A. Livi. 2005. "Gender differences in walking behavior, attitudes about walking, and perceptions of the environment in three Maryland communities."*Research on Women's Issues in Transportation, Transportation Research Board.* Washington, DC.

Cubbin, C., and G. S. Smith. 2002. "Socioeconomic inequalities in injury: critical issues in design and analysis."*Annual Review of Public Health* 23: 349-75.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. "Bayesian Network Classifiers." *Machine Learning* 29: 131–163.

Pahukula, J., Hernandez, S., & Unnikrishnan, A. 2015. "A time of day analysis of crashes involving large trucks in urban areas." *Accident Analysis & Prevention, 75*, 155-163.

Ponnaluri, R., and F. Nagar. 2010. "Road crash risk among vulnerable population groups in Andhra Pradesh, India."*TRB 89th Annual Meeting Compendium of Papers DVD, Transportation Research Board.* Washington, DC.

Sun, M., Sun, X., & Shan, D. 2019. "Pedestrian crash analysis with latent class clustering method. " *Accident Analysis & Prevention, 124*, 50-57.

WHO. 2022. *Road traffic injuries.* https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

Zajac, S. S., and J. N. Ivan. 2003. "Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut."*Accident Analysis & Prevention* 35 (3): 369-379.

8. Evaluating bicycle lanes and stations safety in Barcelona: a Bayesian network approach to analyze injury severity

8.1. Abstract

Bicycle infrastructures are spreading day by day in different regions that are aiming to provide various sustainable transport modes. In Barcelona, Spain, the increase in using bicycles is captured alongside the increase in providing appropriate supporting infrastructures. One of the reasons that are leading to the detected spike in bicycle usage in Barcelona is the founded public bicycle sharing systems. More than 400 km of bicycle lanes are designated and provided with 496 stations that are distributed in different areas around Barcelona. In this study, Geographic Information System (GIS) is exploited to depict the bicycle lanes, stations, and zones. Then, the bicycle crashes data with three types of injuries that occurred in 2019 in Barcelona are combined with respect to the location aspect alongside other aspects including gender, age, and the time of the day. A Tree Augmented Naïve tree (TAN) which is part of the Bayesian network family is, then, applied to classify these potential factors. The results show that the most vulnerable age category that may have a slight or severe injury in most cases while doing cycling is 26 to 50 years old category during morning and evening timing. Additionally, bicycle different types of lanes and bicycle stations in zones that have a restricted speed limit of 30 km/h showed a fewer crash injury cases compared to other types. For TAN, the approach has shown a robust prediction performance when dealing with similar datasets.

***Keywords*:** Bicycle, Safety, Injury, Evaluation, TAN, Barcelona

8.2. Introduction

“Very quickly, we've moved from being a curiosity to a new urban transport mode. We invented the public-individual transport “Gilles Vesco. One of the preliminary initiations in Europe related to public bicycle sharing systems was introduced in 1965 called the “White Bike Plan” (Home, 1991) in Amsterdam, Netherland. Later on, a similar bike-sharing system in La Rochelle, France was introduced in 1974 called “Yellow Bikes” followed by a bike-sharing system that was called “Green Bike Schemes” in 1993 in Cambridge, United Kingdom (UK) (Shaheen et al., 2010). In 2007 and based on (Bührmannm, 2008), an 80 % increase in public and private bicycle usage was captured in Lyon, France. In Barcelona, and based on the previous reference, the bicycle users were 118 thousand for public bicycle sharing system (Bicing) when it was first established.

To grasp the concept of public bicycle sharing systems, this term should be defined first, it's a scheme that enables the user to collect the bicycle and rent it from a self-serve station and then return it to any other bicycle station that makes it ideal for point to point trip (Midgley, 2011) (New York City Department of City Planning, 2009). Several reasons are behind implementing public bicycle sharing systems in different countries. Part of these reasons for implementing public bicycle sharing schemes is the possibility of promoting the use of bicycles among the community, improving health, lessening travel cost and time, providing different transport modes for the users, and improving the travel experience (Ricci, 2015). Bicycle commuters were found to share a similar thought related to bicycles benefits based on a conducted survey that was analyzed and exploited in a previous study (Muñoz et al., 2013).

In Spain, a study (Anaya, 2010) examined the public bicycle systems with focusing on Barcelona. The first stage was recognized between 2002 and 2005 that have the first cycling projects. During this stage, third-generation systems incipient and promoting cycling projects with municipalities were established. Then, the growth of this type of project in more cities was reported. In 2007, the boom for cycling was captured with the announcing of the Bicing system. Consistently, and based on a more recent published paper (Pérez et al., 2017), there was a 72.5% increase in cycling in the year 2013 compared to 2009. The average annual economic benefit of cycling in Barcelona was estimated to be 4.7 million euros. Additionally, the study found that there was a drop in injury rates among cyclists despite the increasing cycling trips.

This global spike in bicycle usage is requiring studying the safety of the different aspects related to bicycles as bicycle crashes fatality rate (de Guerre et al., 2020) (Liu et al., 2015) similar to conducted studies (Aiash and Robusté, 2021) that are focusing on traffic crashes injuries, in general. Bicycle lanes and paths can play an important role in lessening cyclists' fatalities along busy roads with mixed traffic (Schepers et al., 2015). Similarly, a study has found that bicycle lanes can help in reducing the number of crashes and injuries when mixed traffic existed (Reynolds, et al. 2009). Marked bicycle lanes can also have a similar impact on lessening injury rates and crash occurrences (Kaplan, 1975) (Lott and Lott, 1976). However, bicycle lanes and paths are found to have an impact on attracting cyclists to distributor roads where these lanes are constructed (S. U. Jensen, 2006), that may lead to higher risk due to these distributor roads (Elvik et al., 2009). Another study (Smith and Walsh, 1988) found a negative influence of bicycle lanes for a section counter to on-road traffic flow on major roads, but, this negative effect did not remain over the long term.

In Toronto, a study (Aultman-Hall and Kaltenecker, 1999) found that the probability of having a crash for cyclists is higher on roads compared to off-road paths or on the sidewalks. In contradictory, sidewalks were found to have the highest probability of having major injuries and falls compared to the other two path types. Males' cyclists have higher incidents rate compared to females, and cyclists who are 50 years and older have a higher injury risk (Vanparijs et al., 2015). The expansion of bicycle infrastructures is found to have a positive impact on increasing the perceptions for cyclists related to the safety aspect (Branion-Calles et al., 2019). At intersections, bicycle crashes are found to be inversely proportioned to the width of pedestrian pavements (Kim et al., 2012). It was also found that bicycle crashes are more likely to happen at intersections with stop signs compared to no-stop signs intersections (Dong et al., 2021). Moreover, a study found that cyclists who were turning left at an intersection were more exposed to traffic crashes (Yuan and Chen, 2017).

On the other hand, crashes with large vehicles, lack of infrastructures for cyclists, steep terrain, and night timing are considered part of the risk factors that can lead to fatal injuries associated with bicycle crashes (Carvajal et al., 2020). Weather, windy condition, conditions with poor lighting, wet ground, loose sand, and dirt pavement are found to be risk factors related to bicycle crashes (Das et al., 2021). In a recently published study (Zhu, 2021), male cyclists have higher odds of having severe or fatal injuries similar to young drivers or cyclists, arterials roads, areas that have 100 km/h speed limit, and poor light conditions. Intersection, bicycle path, non-winter season besides other factors that are related to multiparty crashes are found to increase the chance of having severe injuries (Hosseinpour et al., 2021). Consistently, the zone that has 100 km/h, road classification, crash type, and other factors were also confirmed by another previously mentioned research (Zhu, 2021) as a potential risk that can increase the level of severity. Another risk factor is having a crash with a motor-vehicle which was found to be a significant factor related to severe and fatal injuries for cyclists (Hefny et al., 2012). For electric bicycles, safety helmets and other safety gear were found to be a necessary to protect cyclists from injuries (Wu et al., 2021). Similar studies (Otte, 1998) (Otte and Haasper, 2010) asserted the importance of wearing the helmet for cyclists for their safety, in general.

The objective of conducting this study is to evaluate the safety and the risk for all types of bicycle lanes and stations for the users in Barcelona with focusing on three types of injuries including slight, severe, and fatal injuries. Eventually, this study is prepared to measure this risk besides considering other potential risk factors such as age, gender, and the day. The results should provide a safety assessment for the bicycle different provided infrastructures in Barcelona and determine the risk probabilities for the different selected independent variables.

8.3. Bicycle in Barcelona

There are different types of bicycle lanes in Barcelona that have different characteristics. Firstly, these lanes bifurcate into two main types on the pavement and on the road. Both sidewalks/pavements and road types are divided into two types; one that has only one direction while the other has two directions. Pedestrians' areas are the areas where the pedestrian has priority over vehicles, bicycles, skates, and scooters. However, bicycles and the last-mentioned two types have priority over vehicles similar to pedestrians. These roads have speed limits of 10 and 20 km/h. For the bicycle roads, here, cyclists are cycling with the traffic flow with a similar previous priority of road users. These roads have a speed limit of 30 km/h. It is mandatory to cycle on these specifically designed roads for bicycles. Nevertheless, when these roads do not exist, cyclists can use the road or the pavement that has a 5-meter wide and 3 meters of free space with a speed limit of 20km/h and no signage prohibiting cycling. The lateral safety distance is set to be 1.5 meters for motor vehicles when sharing the road with cyclists. When driving behind a cyclist, a 3-meter distance should be kept between the motor vehicle and the bicycle. More traffic regulations can be found from the municipality of Barcelona.

As shown in figure 25, There are two types of bicycle lanes that are constructed in Barcelona using the QGIS software and based on the collected data from the Municipality of Barcelona (Barcelona's City Hall, 2019). The first type is bicycle lanes that are shown with red lines, while cycle lanes are shown in blue lines. Both are for cyclists; however, the type of road is different. Then, the length of all types of bicycle lanes is determined. For cycle lane type, the length is 265.977 km and 183.368 km is the length for the bicycle lane type. As the bicycle lane (or cycle way) is separated from the main road and sometimes shared with pedestrians on the pavement without designated lanes, while the cycle lane is on the same road, as the cyclists are sharing the same road with other vehicles or pedestrians with designated lanes. The yellow zones that are disseminated in different districts are the zones where the speed limit is set to be 30 km/h for all roads on those areas. These categorizations are collected and explained based on the information provided by the City Hall of Barcelona.

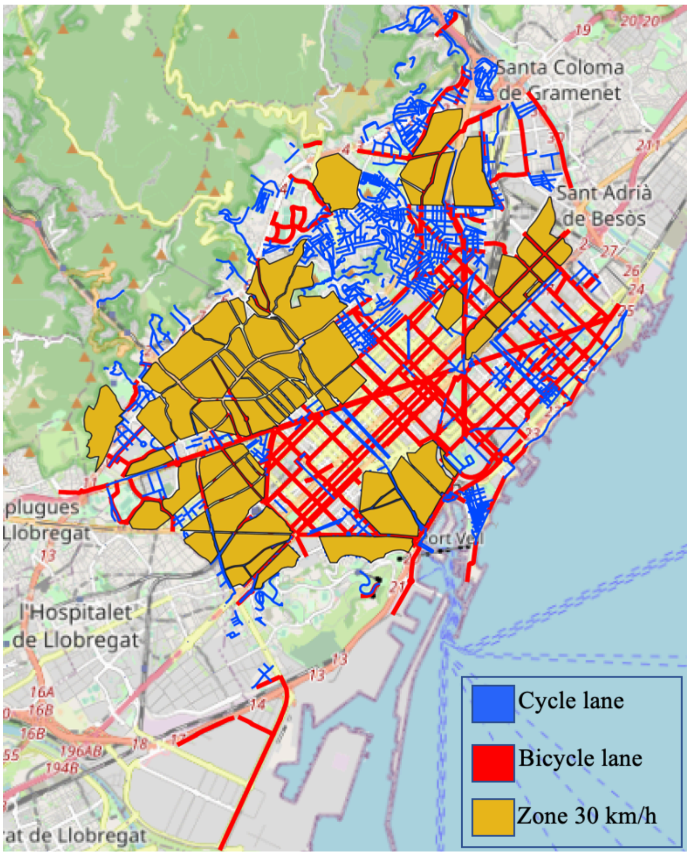


Figure 23 Barcelona bicycle paths alongside zone 30

For the lanes, and as mentioned earlier, there are different types of lanes that are designated in different areas in Barcelona. Firstly, there is the shared with other vehicles on the same road such as shared with buses and other vehicles as shown in figure 26. Secondly, there is the segregated type that is in parallel with the road but protected. Thirdly, there is the separated type which is on the same road but with a special lane designated only for bicycles. Fourthly, there is the type that is shared with pedestrians on the sidewalks or the pavement for the footpath.

Figure 24 Cycle lane paths on Pelai street in Barcelona

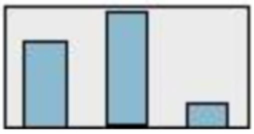
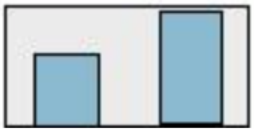
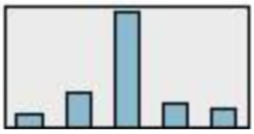
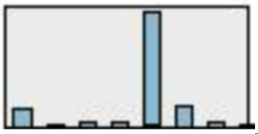
8.4. Methodology

8.4.1. The process of locating crashes

Although the bicycle crashes data is provided by the utilized database, the details regarding bicycle lanes are not provided with respect to traffic crashes. Therefore, the data cannot provide relative information regarding whether the crashes happened in or out of these lanes or within the areas that have a speed limit of 30 km/h or the areas that have bicycle stations. However, the coordinates are provided for each case. Thus, these cases are depicted using its coordinates by utilizing QGIS software. Similar to a previously conducted study (Soltani and Askari, Exploring spatial autocorrelation of traffic crashes based on severity 2017), localizing the crashes in order to measure the safety of these provided infrastructures related to injury severity is carried out. The location for bicycle lanes with its different types that are above mentioned are also depicted alongside the 30 km/h zones and the exact locations of bicycle stations. Then, a buffer is developed around both the lanes and stations. For lanes, the buffer is estimated to be 3 meters with assuming that the width of two lanes is 3 meters, while for the station, the buffer is estimated to be 10 meters after multiple trails with different buffers width that may be the best width that can detect crashes around these stations.

Zones with 30km/h do not need a developed buffer as their boundary is already defined. Then, crashes that are included in the analysis part are imported. Eventually, the number of crashes that occurred within these zones is counted with their coordinates. These details are merged with the original data alongside other given details including gender, age, and the day timing. Lastly, a total of 689 bicycle crashes that happened in Barcelona in 2019 are exploited. Table 1 is presenting the overall data the is used in this study. It has to be mentioned that the graph that is shown in table 1 is depicting the categories of the different variables with starting with the category that has 1 value in ascending order starting from the left side. Four different variables are included with different categories. For daytime variable, it consists of the three classes including the first class from the left which is morning, then evening period, lastly night period. The gender of cyclists who got the injury is also included with the first category which is female, and the second category is male as shown in the graph of table 33. The age category has five levels including the first age category from the left which is injured cyclists who have 16 years and younger, the second category is 17 to 25 years old cyclists, the third category is 26 to 50 years old cyclists, the fourth category is 51 to 64 years old cyclists, and the last category is 65 years old and older cyclists. For the type of the road, eight different types are represented including bicycle lane, bicycle lane in zone 30, cycle lane, cycle lane in zone 30, other type, other type in zone 30, other type near to bicycle station, and lastly, other type near to bicycle station in zone 30. Figure 27 shows the bicycle crashes location that is utilized in this study and occurred in Barcelona in 2019.

Table 33 The chosen independent variables related to bicycle injury crashes

Variable	Graph	Measurement	Valid
Day time		Nominal-categorical	689
Gender		Nominal-categorical	689
Age		Nominal-categorical	689
Type of the road		Nominal-categorical	689

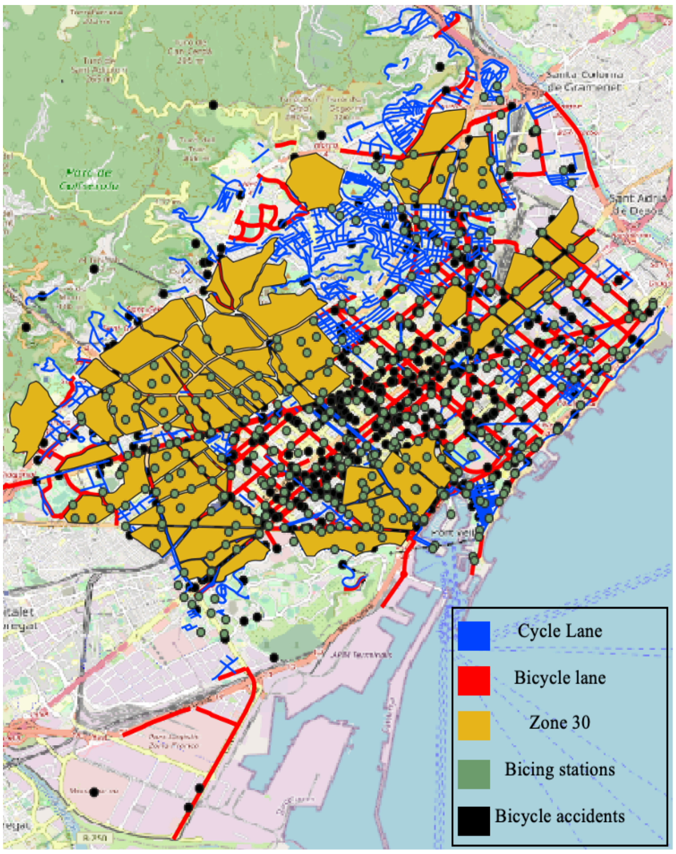


Figure 25 Bicycle crash location in Barcelona

8.4.2. Bayesian network

A Bayesian network classifier as Naïve Bayes is chosen to classify these risk factors due to its proven ability in predicting traffic crashes data (AlKheder et al., 2020) (Aldred et al., 2020) by exploiting the Tree Augmented Naïve Tree (TAN) approach. This technique is considered a simple approach (Kontkanen et al., 1998) (Langley et al., 1992) that can handle small and relatively small datasets very well. TAN is considered a natural extension to the Naïve Bayes classifier (Cerquides and de Mantaras, 2003) that is considered more coherent and best performing approach improvement compared to Naïve Bayes. A tree structure on the Naïve Bayes is imposed by TAN with restricting the interaction between the chosen variables to one. Class variables connect to all the variables by direct edges means (Padmanaban, 2014). Moreover, each variable in this developed structure of TAN can have two parents as each one of them can be connected to another one. Therefore, the complexity of Naïve Bayes is lessened in TAN but with sustaining the robustness and the ability to provide better accuracy (Friedman et al., 1997). The training sample is chosen to be 70 % while the rest is the testing sample. This model is applied by utilizing IBM Watson cloud service. The classifier of TAN that is exploited by following previous proceedings (Aiash and Robuste, 2021) through the IBM software platform (IBM, 2016) is based on this book (Friedman et al., 1997). The conditional probabilities are calculated, in general, by SPSS modeler as follows:

$$Pr\left(Y_i \middle| X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j\right) = \frac{Pr(Y_i)\left(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j \middle| Y_i\right)}{Pr\left(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j\right)}$$

$$\propto \Pr\,(Y_i) \prod_{k=1}^{n} Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right) \qquad (1)$$

d is the data set. d_j is the case as $d_j = \left(x_1^j, x_2^j, \dots, x_n^j\right)$. i^{th} target category that d_j belongs to. The dependent variable Y_i is the level of injury including slight, severe, and fatal injury category. The independent variable is x. The number of independent variables which are four this study is n. The number of non-redundant parameters is K. The parent set of the independent variable alongside the dependent variable is π_k, the value of it maybe empty for Tree Augmented Naïve Bayes. $Pr\left(X_k = x_k^j \middle| \pi_k^j, Y_i\right)$ is the conditional probability that is associated with each node. In this study, there are four nodes for independent variables. Followingly, the several steps are taken by the cloud to employ and construct the applied TAN. Firstly, the data D, predictors X, and the target Y are taken as a training data. Secondly, structure learning algorithm is used to learn a tree-like network over the predictor. Y the target variable is, then, added as a parent to each predictor X where $1 \leq i \leq n$. Then, the learning is carried out for the parameters. This structure learning algorithm is based on weight spanning tree MWST method (Chow 1968). Therefore, each edge related to common information between the two variables is given a weight by this method. MWST algorithm (Prim 1957), gives undirected tree after the weight matrix is developed. Defining the mutual information between two nodes such as X_i *and* X_j as stated in 2:

$$I(X_i, X_j) = \sum_{x_i, x_j} \Pr(x_i, x_j) \log(\frac{\Pr(x_i, x_j)}{\Pr(x_i) \Pr(x_j)}) \tag{2}$$

The conditional mutual information given the target replaces the previously mutual information between two predictors (Friedman et al., 1997) is shown in 3.

$$I(X_i, X_j|Y) = \sum_{x_i, x_j} \Pr(x_i, x_j, y_k) \log(\frac{\Pr(x_i, x_j|y_k)}{\Pr(x_i|y_k) \Pr(x_j|y_k)}) \tag{3}$$

$I(X_i, X_j|Y)$ is computed, $where\ i = 1, \dots, n\ and\ j = 1, \dots, n\ and\ i \neq j$ between each variables pair. A maximum weighted spanning tree with the weight of an edge is developed by using (Prim 1957). This edge connects $X_i\ to\ X_j$ by $I(X_i, X_j|Y)$. X_1 is, then, chosen as a root node to transform the undirected tree to directed tree. Lastly, all directions of all edges, are, then directed outward from it.

8.5. Results and discussion

The accuracy of the applied Tree Augmented Naïve Bayes for training and testing sets are 98.34 % and 97.09 %, respectively, as shown in table 34. The final structure of the applied model is shown in figure 28. The structure consists of all the independent variables including day-time variable, type of the road, age, and gender. The most important variable based on the applied TAN is day-time variable with 0.328 importance value followed by age and type of the road with 0.238 and 0.228, respectively. Then, lastly, the gender is the least important variable with 0.206 value based on the applied TAN.

Table 34 The prediction accuracy of the applied TAN

Partition	Training	Training Accuracy	Testing	Testing Accuracy
Correct	475	98.34 %	200	97.09 %
Wrong	8	1.66 %	6	2.91 %
Total	483		206	

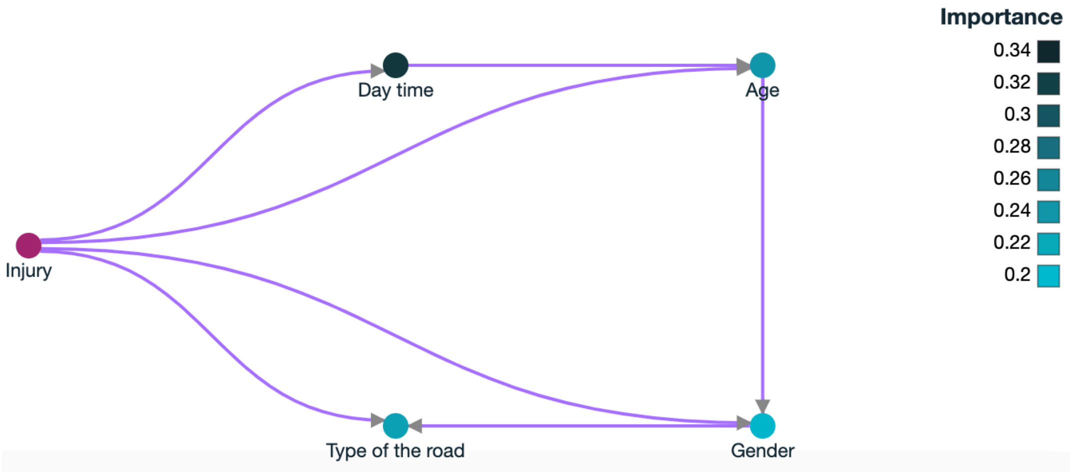


Figure 26 The structure of the applied TAN

Tables 35, 36, 37, and 38 are all showing the results of the calculated conditional probabilities for the selected variables with respect to the level of injuries. Table 35 is depicting the results for the day time variable. The results show that evenings have higher probabilities of having slight and severe injuries compared to other periods. Evening timing periods besides night timing periods were also found to have an association with crash severity (Hosseinpour et al., 2021). However, night timing is more likely to have fatal injuries compared to other timings. For the road variable, areas or roads that do not have the designated lanes for bicycles and/or located outside areas that have 30 km/h speed limits have the highest probabilities of having slight, severe, and fatal injuries in all different categories, as presented in table 36. These findings are consistent with previously conducted studies (Chen and Shen, 2019) that found that higher speed limits can lead to severe or fatal injuries and the importance of bicycle lanes in protecting cyclists (Nolan et al. 2021).

Table 35 Calculated conditional probabilities for day predictor

Parents		Number of Instances	
Injury	Morning	Evening	Night
Slight	0.394	**0.522**	0.084
Severe	0.143	**0.714**	0.143
Fatal	0.000	0.000	**1.000**

Table 36 Calculated conditional probabilities for road predictor

Parents		Number of Instances							
Gender	Injury	Bicycle lane	Bicycle lane in Zone 30	Cycle lane	Cycle lane in Zone 30	Other	Other in Zone 30	Other near to bicycle station	Other near to bicycle station in Zone 30
Male	Slight	0.108	0.000	0.020	0.007	**0.721**	0.128	0.013	0.003
Male	Severe	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
Male	Fatal	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
Female	Slight	0.118	0.006	0.028	0.011	**0.708**	0.129	0.000	0.000
Female	Severe	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000

Tables 37 and 38 are showing the results for the calculated conditional probabilities for age and gender variables, respectively. For age categories, the parent for this variable is the day variable besides the injury target variable. During the morning timing, the age category between 26 to 50 years old has higher odds of having slight injuries. For severe injury during morning timing, the age category that has 65 years or older is more likely to have severe injuries. The 26 to 50 years old age group and 51 to 64 years age group are both more likely to be involved in bicycle crashes during the evening timing with severe injuries. For night timing, slight injuries are more likely to occur in the age category between 26 to 50 years old, while for severe and fatal injuries, the age category the is between 17 to 25 years old has the higher probabilities compared to other categories. For gender variable, a male is more likely to have slight and severe injuries compared to females for most age categories and cases except for the age category 26 to 50 regarding only severe injuries and 65 years old and older category regarding having both slight and severe injuries. However, the probabilities for category 51 to 64 years old who had severe injuries are equally split between male and female categories.

Table 37 Calculated conditional probabilities for age predictor

Parents		Number of Instances				
Day time	Injury	16 -	17 to 25	26 to 50	51 to 64	65 +
Morning	Slight	0.048	0.144	**0.615**	0.096	0.096
Morning	Severe	0.000	0.000	0.000	0.000	**1.000**
Evening	Slight	0.069	0.161	**0.544**	0.141	0.085
Evening	Severe	0.000	0.200	**0.400**	**0.400**	0.000
Night	Slight	0.075	0.225	**0.625**	0.075	0.000
Night	Severe	0.000	**1.000**	0.000	0.000	0.000
Night	Fatal	0.000	**1.000**	0.000	0.000	0.000

Table 38 Calculated conditional probabilities for gender predictor

Parents		Number of Instances	
Age	Injury	Male	Female
16 -	Slight	**0.690**	0.310
17 to 25	Slight	**0.750**	0.250
17 to 25	Severe	**1.000**	0.000
17 to 25	Fatal	**1.000**	0.000
26 to 50	Slight	**0.615**	0.385
26 to 50	Severe	0.000	**1.000**
51 to 65	Slight	**0.661**	0.339
51 to 64	Severe	**0.500**	**0.500**
65 +	Slight	0.359	**0.641**
65 +	Severe	0.000	**1.000**

8.6. Conclusions

Cycling is increasing in many regions and expanding as a sustainable transport choice to commute in different regions. Therefore, bicycle stations and lanes need to be evaluated for their safety with respect to cyclists and traffic crash severity. Eventually, this study is conducted to provide a safety evaluation in Barcelona for these infrastructures alongside considering other related aspects including gender, age, and time of the day. Eventually, the results are expected to promote building similar infrastructures in order to protect cyclists and alleviate the risk of severe or fatal injuries due to traffic crashes. GIS is exploited to locate the occurrence of the cyclists' crashes and then to merge these details with other aforementioned risk factors. Subsequently, a Bayesian network model represented by the Tree Augmented Naïve Tree is utilized to analyze the data, provide the importance values for the factors, and classify the impact of these factors.

Eventually, the results have shown that the evening period is more likely to have slight and severe injuries, while the night period is more likely to have fatal injuries. For the type of the road variable, areas that do not lay within areas with a 30 km/h speed limit or that have bicycle or cycle lanes, or the bicycle stations have higher probabilities of having all types of injuries for both males and females injured parties. Male injured parties, however, are more likely to have slight, severe, and fatal injuries compared to females for most of the age groups. The most vulnerable age category that may have a slight or severe injury for most cases while doing cycling is 26 to 50 years old category during morning and evening timing, while for fatal injury, the age category that represents injured parties between 17 to 25 years has the highest odds during night timing. Lastly, TAN technique has proven its robust performance in dealing with this type of dataset.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

8.7. References

Aiash, A., and F. Robusté. 2021. "Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree." *International Journal of Injury Control and Safety Promotion.* 29(2): 256-264.

Aiash, Ahmad, and Francesc Robuste. 2021. "Working hours and traffic accident injuries: case study in Barcelona." *Transportation Research Procedia* 58: 678-682.

Aldred, R., S. García-Herrero, E. Anaya, S. Herrera, and M. Á. Mariscal. 2020. "Cyclist Injury Severity in Spain: A Bayesian Analysis of Police Road Injury Data Focusing on Involved Vehicles and Route Environment." *International Journal of Environmental Research and Public Health* 17 (1): 96.

AlKheder, S., F. AlRukaibi, and A. Aiash. 2020. "Risk analysis of traffic accidents' severities: An application of three data mining models." *ISA Transactions* 106: 213-220.

Anaya, E. 2010. "Are public bikes systems good for cycling? The view from Barcelona." *Bicycle Politics Workshop, 16th-17th September 2010, CeMoRe, Lancaster University, UK.* Lancashire, UK.

Aultman-Hall, L., and M. G. Kaltenecker. 1999. "Toronto bicycle commuter safety rates." *Accident Analysis & Prevention* 31 (6): 675-686.

Barcelona's City Hall. 2019. *Barcelona's City Hall Open Data Service.* https://opendata-ajuntament.barcelona.cat/en/.

Branion-Calles, M., T. Nelson, D. Fuller, L. Gauvin, and M. Winters. 2019. "Associations between individual characteristics, availability of bicycle infrastructure, and city-wide safety perceptions of bicycling: A cross-sectional survey of bicyclists in 6 Canadian and U.S. cities." *Transportation Research Part A: Policy and Practice* 123: 229-239.

Bührmann, S. 2008. "Bicycles as public-individual transport – European developments." *MEETBIKE – European Conference on Bicycle Transport and Networking.* Dresden.

Carvajal, G. A., O. L. Sarmiento, A. L. Medaglia, S. Cabrales, D. A. Rodríguez, D. A. Quistberg, and S. López. 2020. "Bicycle safety in Bogotá: A seven-year analysis of bicyclists' collisions and fatalities." *Accident Analysis & Prevention* 144 (105596).

Cerquides, J., and R. L. de Mantaras. 2003. *Tractable Bayesian Learning of Tree Augmented Naive Bayes Classifiers.* CSIC.

Chen, P., and Q. Shen. 2019. "Identifying high-risk built environments for severe bicycling injuries." *Journal of Safety Research* 68: 1-7.

Chow, C. K. 1968. "Approximating discrete probability distributions with dependence trees." *IEEE Transactions on Information Theory* 14 (3): 462-467.

Das, S., Z. Wei, X. "Jack" Kong, and X. Xiao. 2021. "Mining crowdsourced data on bicycle safety critical events." *Transportation Research Interdisciplinary Perspectives* 10 (100360).

de Guerre, L. E. V. M. , S. Sadiqi, L. P. H. Leenen, C. F. Oner , and S. M. Gaalen . 2020. "Injuries related to bicycle accidents: an epidemiological study in The Netherlands." *European Journal of Trauma and Emergency Surgery* 46: 413–418.

Dong, J., M. Hirota, T. Nomura, and J. Sato. 2021. "Analysis of Crossing Collision Accident Characteristics by Accident Party." *International Journal of Intelligent Transportation Systems Research* 19: 214–229.

Elvik, R., A. Høye, T. Vaa, and M. Sørensen. 2009. *The Handbook of Road Safety Measures.* Bingley: Emerald Group Publishing.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. "Bayesian Network Classifiers." *Machine Learning* 29: 131–163.

Hefny, A. F., H. O. Eid, M. Grivna, and F. M. Abu-Zidan. 2012. "Bicycle-related injuries requiring hospitalization in the United Arab Emirates." *Injury* 43 (9): 1547-1550.

Home, S. 1991. "Social Science." In *The Assault on Culture: Utopian Currents from Lettrisme to Class War*, 65-8. AK Press.

Hosseinpour, M., T. K. O. Madsen, A. V. Olesen, and H. Lahrmann. 2021. "An in-depth analysis of self-reported cycling injuries in single and multiparty bicycle crashes in Denmark." *Journal of Safety Research* 77: 114-124.

IBM. 2016. *IBM SPSS Modeler 18.0 Algorithms Guide.* IBM Corporation.

Kaplan, J. 1975. *Characteristics of the regular adult bicycle user.* University of Maryland, Civil Engineering Department.

Kim, M., E. Kim, J. Oh , and J. Jun . 2012. "Critical factors associated with bicycle accidents at 4-legged signalized urban intersections in South Korea." *KSCE Journal of Civil Engineering* 16: 627–632.

Kontkanen, P., P. Myllymaki, T. Silander, and H. Tirri. 1998. "Bayes Optimal Instance-Based Learning." *Machine Learning: ECML-98, Proceedings of the 10th European Conference* (Springer-Verlag) 1398: 77–88.

Langley, P., W. Iba, and K. Thompson. 1992. "An Analysis of Bayesian Classifiers." *In Proceedings of the Tenth National Conference on Artificial Intelligence* 223–228.

Liu, H.-T., C.-S. Rau, C.-C. Liang, S.-C. Wu, S.-Y. Hsu, H.-Y. Hsieh, and C.-H. Hsieh. 2015. "Bicycle-related hospitalizations at a Taiwanese level I Trauma Center." *BMC Public Health* 15 (722).

Lott, D. F., and D. Y. Lott. 1976. "Effect of bike lanes on ten classes of bicycle-automobile accidents in Davis, California." *Journal of Safety Research* 8 (4): 171–179.

Midgley, P. 2011. *Bicycle-sharing Schemes: Enhancing Sustainable Mobility in Urban Areas.* New York: United Nations- Department of Economic and Social Affairs.

Muñoz, B., A. Monzon, and D. Lois. 2013. "Cycling Habits and Other Psychological Variables Affecting Commuting by Bicycle in Madrid, Spain." *Transportation Research Record: Journal of the Transportation Research Board* 2382 (1): 1-9.

New York City Department of City Planning. 2009. "Bike-Share Opportunities in New York City." New York.

Nolan, J., J. Sinclair, and J. Savage. 2021. "Are bicycle lanes effective? The relationship between passing distance and road characteristics." *Accident Analysis & Prevention* 159 (106184).

Otte, D. 1998. "Benefit of in depth data for analysing injury mechanisms of accidents with bicyclists and motorcyclists." *International Journal of Crashworthiness* 3: 53-64.

Otte, D., and C. Haasper. 2010. "Effectiveness of the helmet for bicyclists on injury reduction in German road accident situations – state of affairs on GIDAS." *International Journal of Crashworthiness* 15 (2): 211-221.

Padmanaban, H. 2014. "Comparative Analysis of Naive Bayes and Tree Augmented Naive Bayes Models." *Master's Projects* (Master's Projects) 356.

Pérez, K., M. Olabarria, D. Rojas-Rueda, E. Santamariña-Rubio, C. Borrell, and M. Nieuwenhuijsen. 2017. "The health and economic benefits of active transport policies in Barcelona." *Journal of Transport & Health* 4: 316-324.

Prim, R. C. 1957. "Shortest Connection Networks And Some Generalizations." *Bell System Technical Journal* 36 (6): 1389-1401.

Reynolds, C. CO, M. A. Harris, K. Teschke, P. A. Cripton, and M. Winters. 2009. "literature, The impact of transportation infrastructure on bicycling injuries and crashes: a review of the." *Environmental Health* 8 (47).

Ricci, M. 2015. "Bike sharing: A review of evidence on impacts and processes of implementation and operation." *Research in Transportation Business & Management* 15: 28-38.

S. U. Jensen. 2006. *Effekter af sykelstier og cykelbaner-før-og-efter evaluering af trafiksikkerhed og trafikmængde ved anlæg af ensrettede sykelstier og cykelbaner i Københavns Kommune.* Lyngby: Trafitec - Forskerparken Scion-DTU.

Schepers, P., E. Fishman, R. Beelen, E. Heinen, W. Wijnen, and J. Parkin. 2015. "The mortality impact of bicycle paths and lanes related to physical activity, air pollution exposure and road safetyand road safety." *Journal of Transport & Health* 2 (4): 460-473.

Shaheen, S. A., S. Guzman, and H. Zhang. 2010. "Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future." *Transportation Research Record Journal of the Transportation Research Board* 2143 (1): 159-167.

Smith, R., and T. Walsh. 1988. "Safety impacts of bicycle lanes." *Transportation Research Record* 1168: 49-56.

Soltani, A., and S. Askari. 2017. "Exploring spatial autocorrelation of traffic crashes based on severity." *Injury* 48 (3): 637-647.

Vanparijs, J., L. I. Panis, R. Meeusen, and B. de Geus. 2015. "Exposure measurement in bicycle safety analysis: A review of the literature." *Accident Analysis & Prevention* 84: 9-19.

Wu, S., X. Li, F. Wei, X. Yan, and J. Qian. 2021. "A retrospective study of spine injuries in electric bicycles related collisions." *Injury.*

Yuan, Q., and H. Chen. 2017. "Factor comparison of passenger-vehicle to vulnerable road user crashes in Beijing, China." *International Journal of Crashworthiness* 22 (3): 260-270.

Zhu , S. 2021. "Analysis of the severity of vehicle-bicycle crashes with data mining techniques." *Journal of Safety Research* 76: 218-227.

9. Supervised and unsupervised techniques to analyze risk factors associated with motorcycle crash injuries

9.1. Abstract

Motorcycles are one of the highly used modes of transport in Barcelona, Spain, in particular, and in many different regions, in general. This situation is compromising safety on the road and may be attributed to a potential increase in traffic crashes. Therefore, this study examines several risk factors and their consequential impacts on the level of injury that is resulted in case of a traffic crash. Two approaches are employed to analyze the risk factors, including a supervised learning technique which is a binary probit model, and an unsupervised technique which is Kohonen clustering. The results for both models show that alcoholism and road in poor condition can indeed increase the probability of having different levels of injuries as reasons for the crash. Elderly users are less likely to be involved in motorcycle crash injuries compared to other age categories, especially the age group that ranges from 25-40 years old which has the highest odds. For both techniques, the performance in analyzing the utilized data shows that both approaches can be successfully utilized for this type of dataset.

Keywords: Motorcycle, Crash, Risk factors, Probit, Kohonen

9.2. Introduction

The estimation for the motorcycle segment is 770 million around the world (Konlan et al., 2020). The dramatic spike in using motorcycles is found in many European countries, especially those that lie in the south, including Spain, Italy, and Greece (Pinch and Reimer 2012), and this increase was found to be a multifactorial growth in Barcelona. However, several factors can affect this increase, such as gender and age. It was also found that employment status and education level can contribute to the increase in using the motorcycle (Marquet and Miralles-Guasch, 2016). A similar spike in using motorcycles is detected in Brazil. A study (de Oliveira et al., 2021) has examined several factors that can positively or negatively influence motorcycle usage. The results showed that there was an increase in the motorcycle fleet estimated by 869% between 1998 and 2018. Additionally, the study found that income per capita can positively increase the number of motorcycle fleets. This situation can be a risk sign for traffic crash occurrence. However, by examining motorcycles and their consequential crashes, some particular groups can suffer more from traffic crashes due to several risk factors that increase the severity of these crashes.

Indeed, motorcycle riders are found to have the highest risk of traffic crashes compared to other categories (Konkor, 2021). Several factors can influence the occurrence of traffic crashes and their severity. Helmets usage importance related to the safety of the motorcycle occupants is already affirmed due to its ability to mitigate or prevent severe injuries, as previously stated and noticed in Barcelona after applying the law. Similarly, other researchers have found that helmets were associated with a lower risk of having head and neck injuries (Rice et al., 2016) and severe and fatal injuries in general (Cunto and Ferreira, 2017). The helmet can benefit not only the riders but also the passengers by reducing the risk of fatality (Kashani et al. 2014). The economic aspect of helmet use for motorcyclists showed that non-helmeted injured riders had a much higher curing costs rate compared to riders who used it (Kim, et al. 2015). The number of years for the driving license issuance date when the rider had it can also impact the level of being involved in the crash (Moskal et al., 2012).

For fatal injuries, a study (Vajari et al., 2020) examined several risk factors that increase or lessen the probability of having fatal injuries. Part of these factors, motorcyclists who were over 59 years old, weekend crashes, morning rush hours crashes, multiple vehicles involved in the crash, t-intersections, roundabouts, and uncontrolled intersections, were all found to augment fatalities risk. The reason that age affects the severity can be due to the influence on the risk assessment ability of the rider in different age categories (Islam, 2021). However, a segregated motorcycle lane away from other traffic has shown a decrease in traffic crashes for motorcycle riders ((Sohadi et al., 2000). During nighttime, a study (Lemonakis et al., 2021) examined motorcyclists riding through different conditions and found that the traveling speed of riders was not reduced, which can implicate consequences. The type of motorcyclist can also have an important role in crash severity as super-sport motorcyclists were at four times more of higher risk compared to cruiser or regular motorcyclists (Teoh and Campbell, 2010). This type of motorcyclist was also found to travel more miles and longer trips (McCartt et al., 2011).

As stated previously, many factors can play an important role in increasing or decreasing the odds of impacting motorcycle crash occurrences and the consequences, including the resulting injuries that can happen in many different regions. This study, as a result, aims to analyze a wide variety of independent variables and their categories concerning motorcycle crashes. For this reason, 14,501 cases are examined in this study, with 19 classes grouped into five main predictors. Injury levels are the target variable consisting of two types: slight injury and severe or fatal injury class. For this reason, two approaches are exploited for analyzing the utilized dataset. One of the two approaches is a supervised technique which is a binary probit model, while the other technique is a non-supervised technique which is a Kohonen approach. The reason for having two different approaches is to provide different results for the same dataset to show if there are any differences between the two applied methods in identifying the risk factors. This can help identify an appropriate model that can analyze and examine this type of crash in Barcelona and other countries with a similar risk for motorcycle crashes. Additionally, the results can help show how different techniques can differ from each other in the interpretation process and in the way of determining the correlations between the dependent and independent variables.

9.3. Motorcycles in Barcelona

Back in time to 1991, motorcycle crashes accounted for half of the traffic injuries and fatalities cases in Barcelona, followed by car passenger and pedestrian (Plasència et al., 1995). Additionally, both the 20 to 24 and 15 to 20 age groups had the highest incidence of cases. Consistently, a study (Cirera et al., 2001) found that motorcycle and moped occupants had a higher risk of severe injuries, even when this study was conducted a couple of years after the previously stated study. The use of safety helmets then came into force for two-wheel motor vehicle occupants in Spain in 1992. A safety evaluation of their impact on the users was then preprepared to assess this law in Barcelona (Ferrando et al., 2000). The results of the evaluation showed that there was a reduction in fatalities among occupants after implementing this law. During a similar period, in Barcelona, the incident rate of disability for males was almost 1.5 higher than for females when considering motorcycle crashes (Ferrando et al., 1998). Years after, two-wheel-powered vehicles were still more involved in severe crashes in case of comparing them with other categories (Albalate and Fernández-Villadangos, 2009), emphasizing the impact of excess speed, gender, alcohol consumption, and road width on increasing the level of severity (Albalate and Fernández-Villadangos, 2010). In 2009, advance stop lines (ASL) were implemented. However, these stop lines were found to have a negative impact on motorcycle crashes, and sometimes they can increase injury possibilities (Pérez and Santamariña-Rubio, 2019). A variety of reasons can be behind the preference for motorcycles in Barcelona, in particular, and in different regions around the world, in general, such as parking convince and ability to maneuverability which can lead to saving time and money. This is leading to a risky mode of transport, as shown in table 39, that includes crashes that occurred during the years from 2016 to 2019 in Barcelona. More than 60 percent of road crash injuries are the consequence of riding a motorcycle compared to all other modes of transport, which indicate the perilous situation that a motorcycle exemplifies. This is consistent with the findings of the previously conducted study (Fagnant and Kockelman, 2015) in the U.S., that found that the motorcycle crash rate is 68% higher than other vehicles.

Table 39 Motorcycle injuries vs other modes of transport injuries

Type of the vehicle	Slight injury %	Severe or fatal injury %
Motorcycle	61,41	1,27
Other modes of transport	37,01	0,31

In this study, the data is exploited and extracted from the Barcelona city hall database service (Open Data BCN, 2021). Four years intervals are included and examined with respect to motorcycle crashes, including years from 2016 to 2019. However, data fusion is carried out between two databases' references to include a larger number of potential risk factors and their different categories, similar to a previously conducted study (Ospina-Mateus et al., 2021). Therefore, each case from each database is combined with the missing details from the other data file with the same case reference. Eventually, the final data set is prepared for the final data analysis process. A total of 14,501 cases with five independent variables are included to be analyzed in order to detect their potential impact on motorcycle crash injury severity. The total number of categories for the independent variables is 19 identified categories, as shown in table 40. The first predictor is the reason for the crashes, which includes seven classes. These classes are alcoholism, a road in poor condition, drugs or medicine, inadequate speeding and/or over-speeding, metrological factors, objects or animals on the road, and no specific reason for the crash category. Weekends and weekdays are classified under the week timing variable. Three-day timings are covered, day, evening, and night timing. Gender is also implicated within the independent variables. Four age groups commence with the first group, which consists of injured persons who are younger than or equal to 15 years old. The second group represents injured parties who are between 16 and 24, followed by the third group, which has people between 25 to 40 years old, followed by the 40 to 64 years old group. Lastly, the elderly group includes injured persons who are 65 years old and older. The dependent variable is the injury level which has two classes, slight injury and severe or fatal injury. All identified cases have higher slight injury percentages compared to severe or fatal injuries. However, some categories have higher percentages of severe or fatal injuries when compared to their total number of injuries, such as drugs or medicines, over or inadequate speeding, and objects or animals on the road.

Table 40 Slight, severe, and fatal injuries with respect to risk factors categories

Risk factors	Severe or fatal injury %	Slight injury %
Reason for crash		
Alcoholism	6.78	93.22
Road in poor condition	0.733	99.267
Drugs or medicines	22.857	77.143
Over Speeding or inadequate	29	71.000
Meteorological factors	8.333	91.667
Objects or animals on the road	29.167	70.833
Other	4.395	95.605
Weekday code		
Weekend	5.445	94.555
Weekday	4.375	95.625
Time of the day		
Morning	4.489	95.511
Evening	6.705	93.295
Night	4.118	95.882
Gender		
Male	4.488	95.512
Female	4.714	95.286
Age		
0-15	3.654	96.346
16-24	4.236	95.764
25-40	4.105	95.895
41-64	4.996	95.004
65+	8.576	91.424

9.4. Methodology

9.4.1. Binary probit

A binary probit model is chosen in this study to determine the correlations between the independent variables and the target variable by following previous proceedings (Yu and Abdel-Aty, 2014) (Garrido et al., 2014) (Aiash and Robusté, 2021). This model is considered a supervised learning technique as the utilized data have both inputs that refer to the risk factor and output, which refers to the target variable, in this case, the level of injury. Besides the above-mentioned reasons for choosing the binary probit, the probit model has proven its ability to reduce the bias parameter estimates (Aidoo et al., 2021). The target variable consists of two classes, including the slight injury and severe or fatal injury class, similar to previously conducted studies that have been conducted in different regions focusing on motorcycle crashes and different risk factors (Sivasankaran et al., 2021) (Rahman et al., 2021) (Salum et al., 2019) but with having the severe and fatal injury as one class. The net utility of injury level is U_i for each injured i. The total number of exogenous variables (X_i) is I. U_i is related to X_i.

$$U_i = X_i\beta + \varepsilon_i \tag{1}$$

Where the parameter estimates are represented by β and explanatory variables vectors are represented by X_i. ε_i is the error which is assumed to follow a standard normal distribution.

The latent model is

$$y_i = X_i\beta + \varepsilon_i \tag{2}$$

When the relationship between U_i and the observable injury level y_i satisfies:

$$y_i = \begin{cases} 1 & if\ \ U_i > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

The conditional probability is $\Pr(y_t = 1|x_t)$. j^{th} is the measurement the change in this conditional probability which is the coefficients vector element $\beta, \beta_j(j\epsilon\{1,2,\dots,I\})$. This change is measured when there is a unit change x_i^j. j^{th} is an element in vector X_i. The conditional probability is also assumed to have a normal distribution. Then, the standard normal cumulative distribution function is $\phi(.)$. The conditional probability function is:

$$\Pr(y_i = 1|x_i) = \phi(X_i\beta) \tag{4}$$

9.4.2. Clustering – Kohonen

Clustering is one of the methods that have been extensively utilized by previously conducted studies to examine traffic crash datasets globally (Weiss et al., 2016) (Soltani and Askari, 2017) (Sivasankaran and Balasubramanian, 2020) and for motorcycle crashes datasets as well (Francis et al., 2021). These techniques have shown their ability to reinforce the results of the applied traditional statistical models (Gazder et al., 2022). Kohonen is considered a special part of the neural network and is known as Knet or a self-organizing map that implements clustering. This approach is also considered an unsupervised learning technique due to its nature that does not use a target or a dependent variable similar to the previously applied probit model that has a target variable. Several reasons are behind choosing and applying the Kohonen technique in this study. Firstly, Kohonen has the ability to detect patterns in the variables of the utilized dataset. This technique can also be used to conduct a factor analysis or principal component analysis (PCA). Additionally, after comparing the performance of two other clustering techniques: k-means and two-step cluster, Kohonen has provided the best performance in clustering quality for the utilized dataset different classes, as this will be shown later in the results section.

The basic unit of the Kohonen model is a neuron. These neurons form two layers: the input layer and the output layer. The neurons that belong to the input layer are connected to all other neurons that belong to the output layer with having strengths or weights for these connections. Each record is developed based on the completion that is initiated by each unit with all other units to win this record. The parameters between input and output units or clusters are shown as wights or as cluster centers that are linked with each output unit. These cluster centers are updated similarly when building a k-means technique, but the difference is that clusters are sorted spatially in a two-dimensional grid. Moreover, each record affects clusters and the clusters within a neighborhood of the winning unit. So, the network is started with small random weights, and the input records are presented randomly to the network. The winning unit is detected by identifying the output unit that has the closest center to the output vector.

Then, an adjustment is carried out for the weight of the winning unit to get the cluster center closer to the input vector. η that, is the learning parameter is updated when each cycle ends. When the stopping criteria are reached, this process stops. It has to be mentioned that the winning unit of the output unit is the one that has the lowest activation, unlike the neural network approach. Two phases or stages are linked to the learning process: a gross structure phase that has a large η and neighborhood, unlike the second phase is the fine-tuning phase that has a smaller η neighborhood.

For the calculated distance in the Kohonen method, Eucludean distance is the type that is representing the distance between the vector and the cluster center for the output unit as shown in equation 5.

$$d_{ij} = \sqrt{\sum \left(x_{ik} - w_{jk}\right)^2} \tag{5}$$

The value of input value (kth) for the record (ith) is x_{ik}. The weight of kth on the output unit (jth) is w_{jk}.

The neighborhood function is as follows:

$$d_c(x, y) = \max_i |x_i - y_i| \tag{6}$$

The location of unit x is x_i on dimension i of the output grid, while the location of unit y is y_i on the same dimension. To consider an output unit o_j to be in another output unit o_i, $d_c(o_i, o_j)$ should be less than the neighborhood size n. It is worth mentioning that n usually has different sizes at different phases.

The weight change is calculated as follows:

$$\Delta W = \eta.(W - I) \tag{7}$$

The weight vector is W for the output unit that is being updated. The updating is carried out when the weights are adjusted by adding a difference portion between the current weight vector and the input weight, in case of the neighborhood is larger than zero for the winning output node. η is the learning parameter as mentioned before, and the input vector is I.

$$\Delta w = \eta.\left(w_j - i_j\right) \tag{8}$$

The weight related to input unit j for the output unit that is being updated is w_j. The jth input unit is i_j.

After each cycle, η is updated and its value is generally declining during the training cycles. In this study, exponential decay is chosen for η as it has given much better performance compared to linear decay:

$$\eta(t+1) = \eta(t).exp\left(\frac{log\left(\frac{\eta_{low}}{\eta(0)}\right)}{c}\right) \tag{9}$$

To prevent arithmetic errors, η_{low} has a minimum value of 0.0001 when taking the logarithm. More details related to the applied Kohonen technique can be found at (IBM 2016).

9.5. Results and discussion

9.5.1. Binary probit model

As shown in table 41, all variables with all their categories are included. The first risk factor is the reported reason for the crash, with seven different categories. No specific reason is a significant category with a positive B value that indicates that it can increase the odds of having severe or fatal injuries. Similarly, alcoholism and road in poor condition can both increase the probability of having severe or fatal injuries. For the week variable, weekend values show that weekends have a higher likelihood of having severe or fatal injuries compared to weekdays. For the day timing, the evening period shows higher significant odds of having severe or fatal injuries. Gender as a risk factor is found to be an insignificant variable related to motorcycle crashes based on the applied binary probit. For the age categories all age categories have higher odds of having severe or fatal injuries compared to people who are older than 65 years old.

Table 41 Binary probit estimations for the predictors categories

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	0.218	0.367	-0.936	0.501	0.352	1	0.553
No specific reason	**1.142**	**0.355**	**-1.838**	**-0.445**	**10.317**	**1**	**0.001**
Alcoholism	**1.866**	**0.370**	**-1.658**	**-0.208**	**6.367**	**1**	**0.012**
Road in poor condition	**1.957**	**0.496**	**-2.929**	**-0.984**	**15.541**	**1**	**<0.0001**
Drugs or medicines	0.230	0.450	-1.111	0.651	0.262	1	0.609
Over Speeding or inadequate	0.116	0.389	-0.647	0.879	0.089	1	0.766
Meteorological factors	0.370	0.706	-1.753	1.013	0.275	1	0.600
Objects or animals on the road	0						
Weekend	**0.104**	**0.044**	**0.018**	**0.189**	**5.630**	**1**	**0.018**
Weekday	0						
Morning	0.033	0.037	-0.040	0.105	0.785	1	0.376
Evening	**0.167**	**0.060**	**0.049**	**0.284**	**7.771**	**1**	**0.005**
Night	0						

Male	0.049	0.037	-0.120	0.023	1.747	1	0.186
Female	0						
0-15	**0.429**	**0.181**	**-0.784**	**-0.074**	**5.606**	**1**	**0.018**
16-24	**0.410**	**0.098**	**-0.603**	**-0.218**	**17.438**	**1**	**<0.0001**
25-40	**0.430**	**0.089**	**-0.605**	**-0.254**	**23.070**	**1**	**<0.0001**
41-64	**0.321**	**0.090**	**-0.497**	**-0.145**	**12.789**	**1**	**0.001**
65+	0						

9.5.2. Kohonen

9.5.2.1. Slight injury

To evaluate the quality of the developed cluster, the average Silhouette is calculated, and it is found to be 0.811, which refers to a good clustering quality and means that there is a strong structure for the prepared clusters. The Silhouette is a method that is used for the interpretation and validation process to evaluate the consistency within clusters of data that is proposed by Rousseeuw (Rousseeuw, 1987). Then, the importance values for all of the chosen risk factors are determined as well for the slight injury class. The results from figure 29 show that all risk factors have the same importance values based on the applied Kohonen.

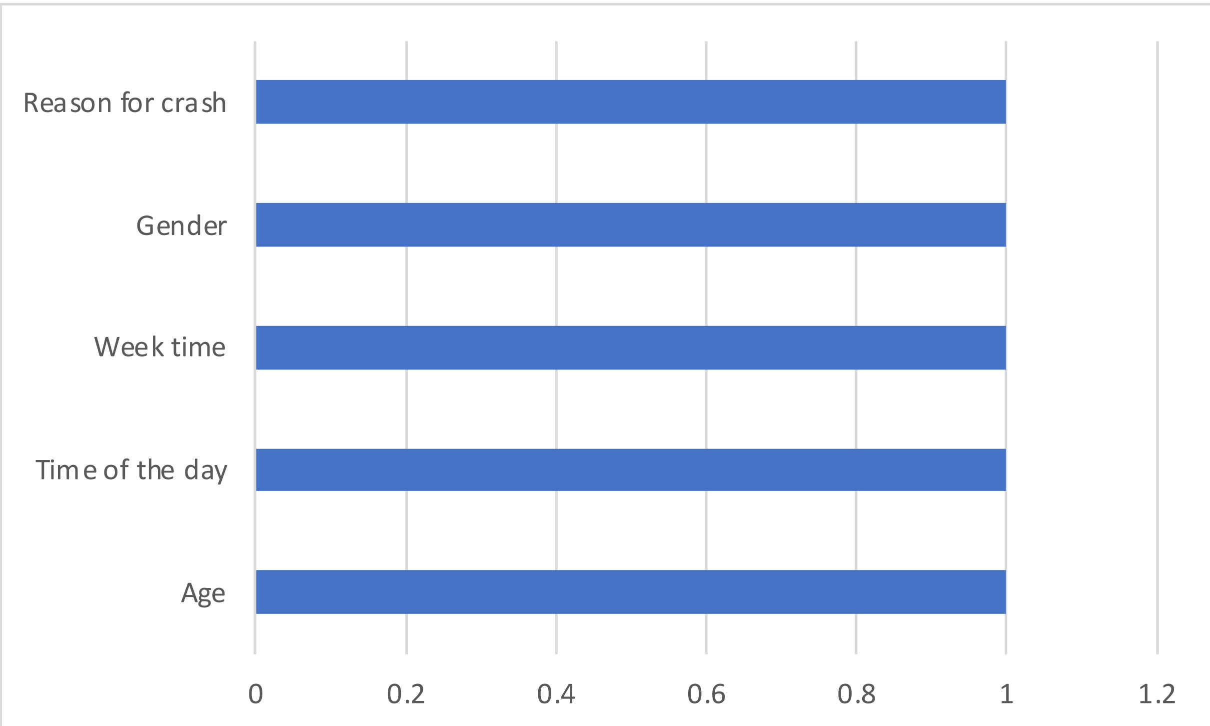


Figure 27 Importance values of the risk factors for slight injury class

Thirty-eight clusters are the total number of clusters that are developed for the slight injury class, as shown in figure 30. The cluster that has the highest percentage is found to be X=9, Y=04 that represents morning, 25-40 age group, no specific reason for the crash, a male injured person, and weekday, and has 13.13% size. Followingly, the group that has 11.47%, which represents the morning, 41-64 age group, no specific reason for the crash, male, and weekday is X=7, Y=0. Thirdly, group X=7, Y=6 has 9.93% that consists of night, no specific reason for the crash, 25-40 age group, male, and weekday. Fourthly, X=5, Y=6 has 9.64% that consists of night, no specific reason for the crash, 41-64, male, and weekday. Fifthly, X=9, Y=2 has 7.86% that consists of morning, no specific reason for the crash, 25-40 age group, female, and weekday. Then, group X=0, Y=6 has 5.13% and represents night, no specific reason for the crash, 25-40 age group, female, and weekday. The other group is X=9, Y=0 that has a 4.91% size and represents the morning, 41-64 age group, no specific reason for crash, female, and weekday. Lastly, group X=2, Y=0 has 4.08% and consists of morning, 16-24 age group, no specific reason for the crash, male, and weekday class.

Cluster X=0, Y=4 has 3.73% size and consists of night timing, 41-64 age group, no specific reason for the crash, female, and weekday timing. Followed by the group X=9, Y=6, this cluster has 2.89% and consists of morning, 25-40 age group, no specific reason, and weekend period. The other group is X=4, Y=4, which has 2.82% and consists of the evening, 25-40 age group, no specific reason, male, and weekday timing. Night, 16-24 age group, no specific reason for the crash, male, and weekday timing are the classes combination for group X=0, Y=4 that has 2.6% size. While for X=0, Y=0 has 2.55% size, the combination of classes consists of morning, 16-24 age group, no specific reason, female, and weekday timing. For X=5, Y=2, the cluster size is 2.29%, and the combination is evening, 16-24 age group, no specific reason, male, and weekday timing. Then, the cluster that has 1.79% is X=5, Y=0, which consists of morning, 41-64 age group, male, and weekday period. Lastly, X=6, Y=4 has 1.73% and consists of the evening, 25-40 age group, no specific reason, male, and weekday timing.

Cluster X=7, Y=2 has 1.55% and comprises the morning, 65+ age group, no specific reason for the crash, male, and weekday timing. The other cluster, X=3, Y=2, has 1.54% and consists of the evening, 16-24 age group, no specific reason for the crash, female, and weekday timing. X=8, Y=6 has 1.49% and comprises of night timing, 25-40 age group, no specific reason for the crash, female, and weekend timing. Night, 16-24 age group, no specific reason for the crash, male, and weekday timing are the combinations for the cluster X=1, Y=2 that has 1.43% size. The next cluster is X=2, Y=4, which has a 1.32% size and consists of night, 65+ age group, no specific reason, male, and weekday. X=3, Y=3 has 1.27% size and comprises of night, 41-64 age group, no specific reason for the crash, male, and weekend timing. The following cluster is

X=3, Y=0 that has 0.87% size and consists of morning, 16-24 age group, no specific reason for the crash, male, and weekend timing. The cluster that has 0.83% size is X=1, Y=6 and consists of night timing, 25-40 age group, no specific reason for the crash, female, and weekend period.

For X=7, Y=4, this cluster represents 0.56% of the total developed clusters and consists of the evening, 25-40 age group, alcoholism, male, and weekday timing. The other cluster is X=2, Y=6, which has 0.53% and comprises of night, 41-64 age group, no specific reason for the crash, female, and weekend timing. X=2, Y=2 has 0.36% size and consists of the evening, 16-24 age group, no specific reason, female, and weekday timing. For X=8, Y=4, this cluster combination has 0.34% size and morning, 25-40 age group, alcoholism, male, and weekday timing. X=7, Y=5 consists of night, 25-40 age group, a road in poor condition, male, and weekday timing and has 0.33%. X=9, Y=1 has 0.31% and consists of morning, 0-15 age group, no specific reason for the crash, female, and weekday timing. For the cluster that has 0.19%, X=6, Y=1 consists of morning, 41-64 age group, alcoholism, male, and weekday timing. X=1, Y=5 has 0.16% and consists of night timing, 0-15, no specific reason, female, and weekday timing. For X=4, Y=0, this cluster has 0.12% and consists of morning, 0-15 age group, no specific reason for the crash, male, and weekday period. The cluster X=3, Y=3 has 0.09% and comprises evening, +65, no specific reason, female, and weekday timing. X=2, Y=5 has 0.09%, similar to the previous cluster but consists of night, 0-15 age group, no specific reason for the crash, male, weekend timing. For clusters X=4, Y=3 and X=6, Y=3, both clusters have 0.03% cluster size. However, X=4, Y=3 consists of the evening, 0-15 age group, no specific reason for the crash, male, weekday period, while X=6, Y=3 cluster consists of the evening, +65 age group, no specific reason for the crash, male, and weekend timing. The last cluster, this cluster is X=3, Y=1 and has a 0.01% size cluster size and consists of morning, 16-24 age group, alcoholism, male, and weekday timing.

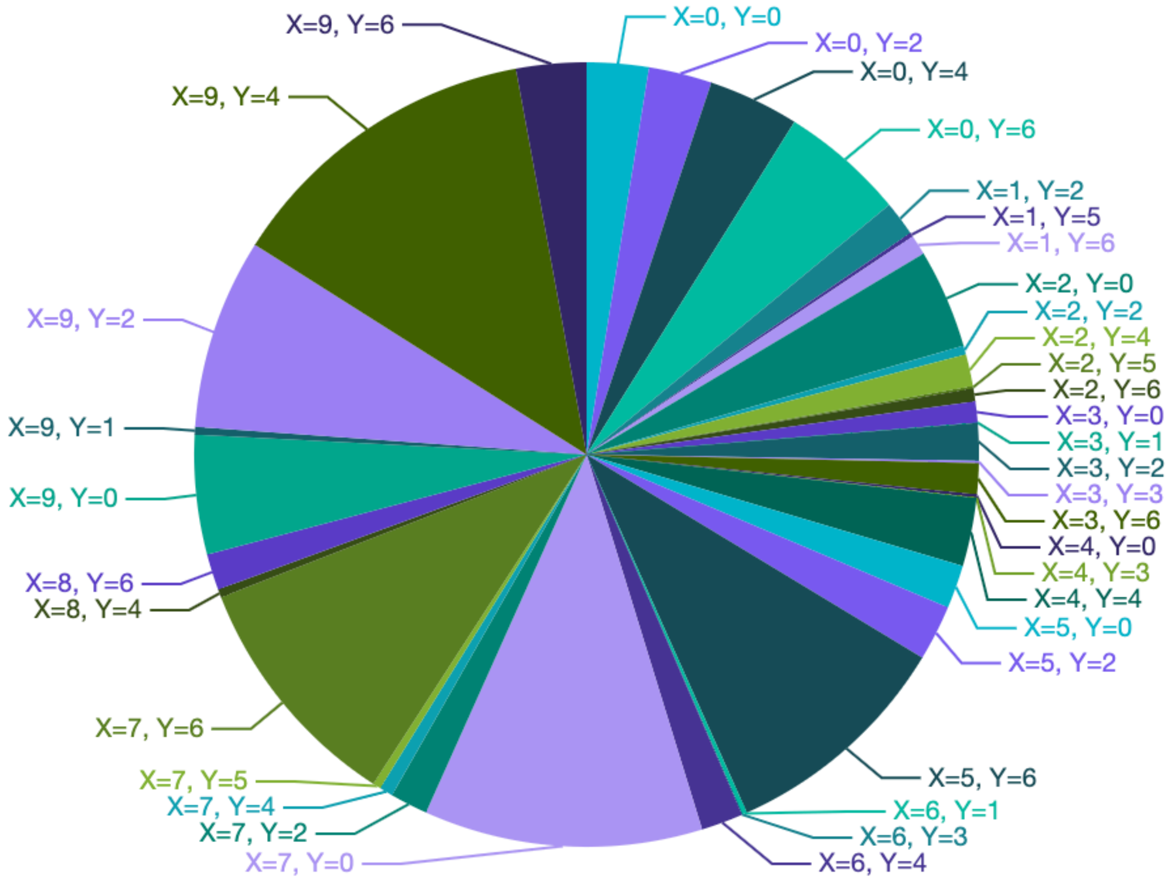


Figure 28 Clusters sizes and categories for slight injury

9.5.2.2. Severe and fatal injury

Similar to the slight injury class cluster, severe and fatal injury class cluster quality has also been evaluated by calculating the average Silhouette with a 0.725 determined value. The result, similar to the previous average Silhouette, indicates that there is a strong cluster structure, and the quality of clustering is good as well. While for the important values based on the applied Kohonen for the severe and fatal injury, age and time of the day variables have the highest value as their values are one, followed by the reason for the crash variable with the value of 0.84, as shown in figure 31. Then, the week time variable and gender variables have 0.49 and 0.46 values, respectively.

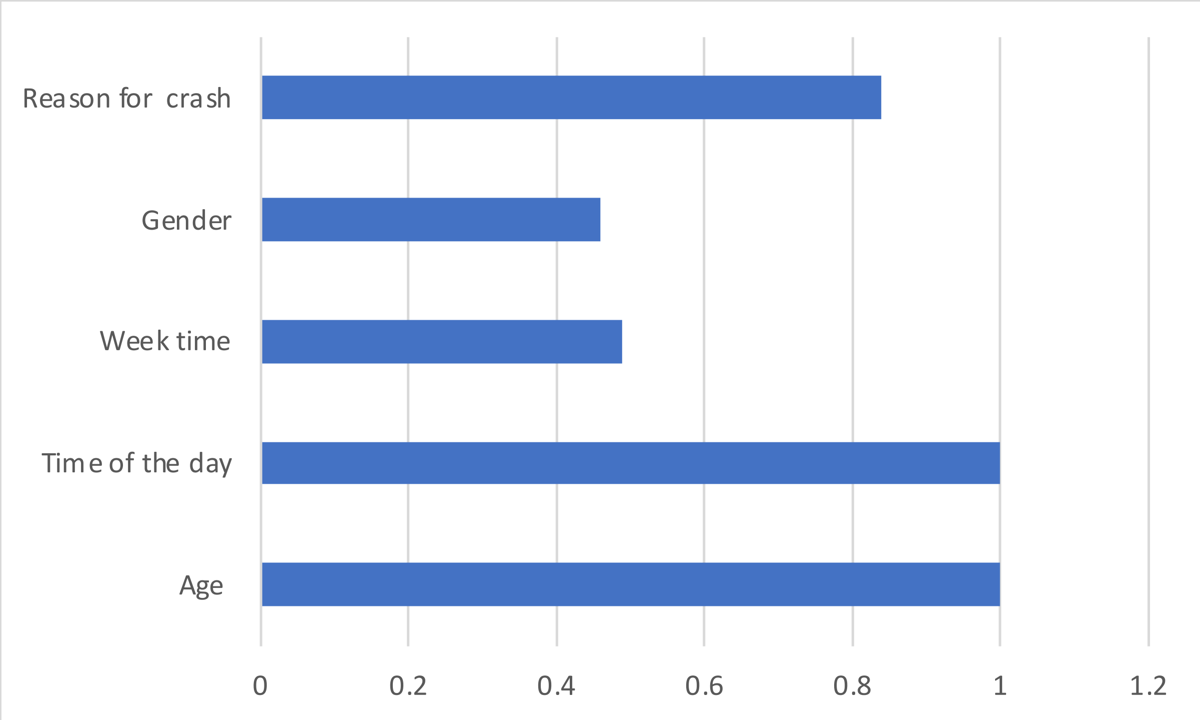


Figure 29 Importance values of the risk factors for severe and fatal injury class

The total number of developed clusters is similar to the previously developed clusters for the slight injury class, which is 38 clusters. The largest cluster in size is X=5, Y=6, which has 11.03% and consists of morning, 41-64 age group, weekday timing, male, and no specific reason for the crash, as shown in figure 32. After that, cluster X=2, Y=4 has 9.93% size and comprises the 25-40 age group, morning, weekday timing, male, and no specific reason for the crash. The next cluster, X=9, Y=2, has 9.1% and consists of 41-64, night, weekday timing, male, and no specific reason. Followingly, X=7, Y=0 has 7.63% size and consists of 25-40 age group, night, weekday, male, and no specific reason. X=0, Y=4 cluster represents 5.88% and has a combination of the 25-40 age group, morning, weekday timing, female, and no specific reason for the crash. For X=0, Y=0, this cluster has a 5.61% size and consists of the 25-40 age group, evening, weekday timing, male, and no specific reason. Cluster X=9, Y=6 has 5.06% size, and cluster inputs are 41-64 age group, morning timing, weekday timing, female, and no specific reason for the crash. For cluster X=9, Y=4, the cluster inputs are 41-64 size, night timing, weekday, female, and no specific reason for the crash, while cluster size is 4.87%. The cluster size for X=9, Y=0 is 4.32%, and the cluster inputs are 25-40 age group, night, weekday, female, and no specific reason for the crash.

Cluster X=2, Y=2 consists of the 16-24 age group, morning, weekday timing, male, and no specific reason for the crash, and has 3.68% size. X=7, Y=6 has 3.22% size and consists of the

40-64 age group, morning, weekday timing, male, and no specific reason for the crash. The cluster size of X=0, Y=6, and X=3, Y=6 is 2.76%. For X=0, Y=6, it consists of the 25-40 age group, morning, weekend, male, and no specific reason, while X=3, Y=0 consists of the 16-24 age group, evening, weekday, male, and no specific reason. For X=4, Y=2, this cluster has a 2.57% size and comprises a 65+ age group, night, weekday, male, and no specific reason for the crash. Then, X=5, Y=0 has 2.3% size and comprised of 25-40 age group, evening, weekend, male, and no specific reason. With 2.11% size, cluster X=0, Y=2 comes with having the combination of +65 age group, morning, weekday, male, and no specific reason for the crash. X=7, Y=4 has the same size as the previous cluster but with consisting of the 41-64 age group, evening, weekday, male, and no specific reason for the crash. X=3, Y=6 has 1.65 % size and consists of the 41-64 age group, morning timing, weekday, male, and over or inadequate speeding as a reason for the crash. Similar to the previous cluster size, X=4, Y=4 comprises of 16-24 age group, morning period, weekend, male, and no specific reason. X=1, Y=0 has 1.56% size and consists of 25-40 age group, evening, weekday period, male, and alcoholism. For X=6, Y=2, this cluster consists of the 25-40 age group, night timing, weekend period, male, and no specific reason, and has a 1.38% size. X=7, Y=2 has 1.29% size and comprises the 40-64 age group, night, weekend, male, and no specific reason for the crash. X=8, Y=6 consists of morning, 41-64 age group, no specific reason for the crash, weekend, and female and has the same cluster size as the previous cluster. For X=0, Y=5, this cluster has a 1.19% size and consists of morning, 25-40 age group, no specific reason for the crash, weekday, and female. For the last cluster that has a size larger than 1%, cluster X=6, Y=4 has 1.1% and comprises the evening, 41-64 age group, no specific reason for the crash, weekend, and male.

For clusters that have a size less than 1%, beginning with cluster X=5, Y=2 that has 0.74% size consists of night, 16-24 age group, no specific reason, weekend, and male. X=2, Y=0, and X=2, Y=6, which all have 0.55% size, but the first cluster consists of the 16-24 age group, evening, weekday, male, and alcoholism, while the second one consists of the 25-40 age group, morning timing, weekday male, and alcoholism. Cluster X=1, Y=6 has 0.46% size and consists of the 25-40 age group, morning timing, weekend, female, and alcoholism. The next cluster, X=4, Y=0, has a 0.37% size and comprises the 65+ age group, evening, weekend, male, and no specific reason. Cluster X=8, Y=2 has 0.28% size and consists of night, 41-64 age group, drugs or medicines as a reason for the crash, weekday timing, and male. Followed by four clusters that have the same size of 0.18%; these clusters are (X=0, Y=3), (X=1, Y=2), (X=5, Y=1), and (X=5, Y=4). The inputs of the first cluster are the 0-15 age group, morning, weekday, female, and no specific reason for the crash. For the inputs of the second cluster, the 0-15 age group, morning, weekday, male, and no specific reason for the crash are the combination of the inputs of this cluster. The third cluster has the 0-15 age group, night, weekday, male, and no specific reason for the crash. The inputs of the last cluster that has 0.18% size are morning, 0-15 age group, no specific reason for the crash, male, and weekend period. For the last three clusters of the developed 38 clusters, these three clusters are (X=2, Y=5), (X=4, Y=3), and (X=7, Y=1), and all of these last three clusters have 0.09% size. X=2, Y=5 consists of morning, 25-40 age group, objects or animals on the road as a reason for the crash, weekday, and male. For X=4, Y=3, the inputs are night, 16-24 age group, no specific reason, weekend, and female. The inputs for the last cluster X=7, Y=1, are night, 25-40 age group, meteorological factors as a reason for the crash, weekday, and male.

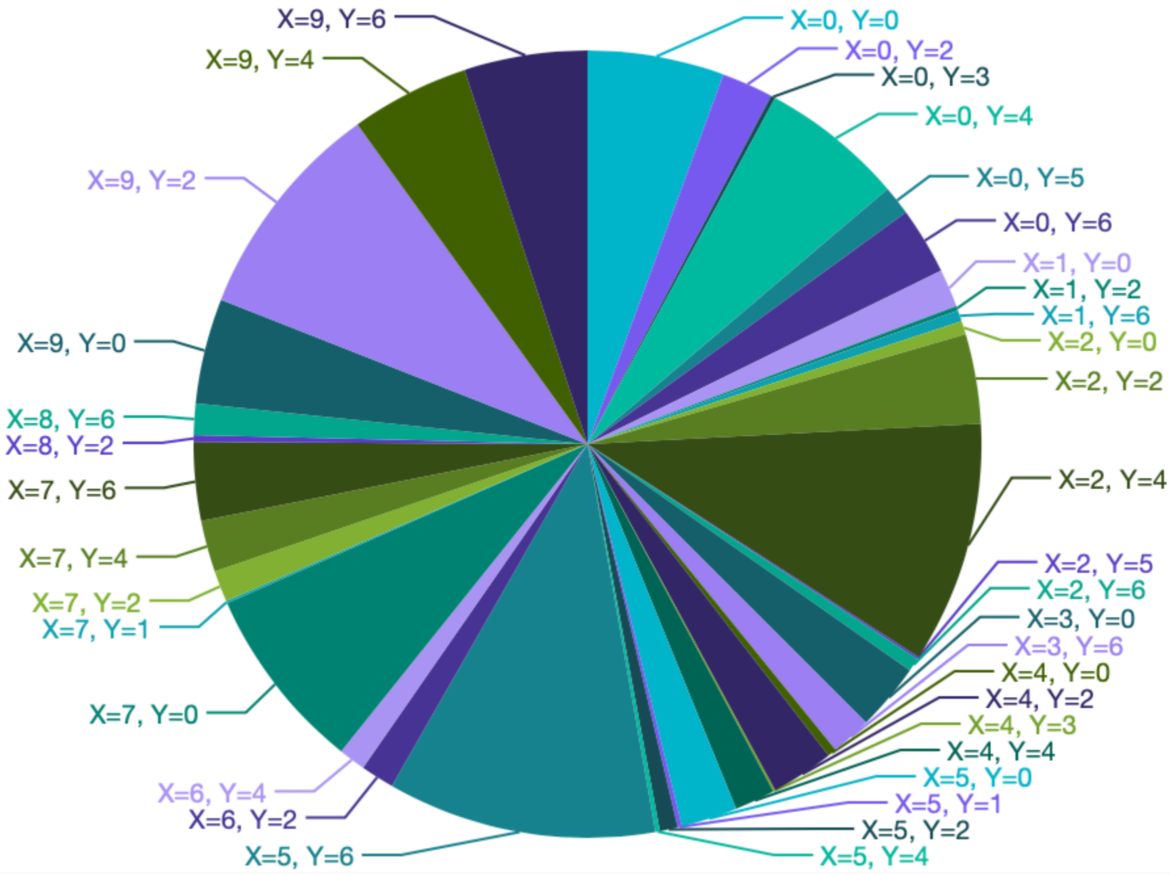


Figure 30 Cluster sizes and categories for severe and fatal injury class

9.6. Conclusions

Motorcycles are one of the most used modes of transport in many areas around the world, and Barcelona is considered one of these areas. Consequences can follow this mode of transport represented by traffic crashes as one of these consequences. Several factors can increase or decrease the injury level that is attributed to these crashes. Gender, age, temporal factors, and other types of factors such as injured status, whether the person is under the influence of alcohol, drugs, or medicines. Road conditions, metrological factors, objects or animals on the road, and the type of crash can all affect the severity of the injury. Therefore, this study attempts to analyze those different factors with respect to the level of injury by employing different techniques. Therefore, two approaches are employed to analyze these potential risk factors related to motorcycle crash injuries, including a supervised technique represented by the binary probit and an unsupervised technique represented by the Kohonen technique.

The results of the applied binary probit show that alcoholism and road in poor condition can play an important role in increasing the odds of having severe or fatal injuries. For different timing periods, evening and weekend timing periods are found to have higher probabilities of having severe or fatal injuries compared to other timing periods. Elderly motorcycle users are found to have the lowest odds of having severe or fatal injuries compared to all other age groups.

For the unsupervised technique, the results have shown that no specific reason for the crash is the most found category, which can be considered as an indication of the necessity to improve the details that are included in the crash report for both different injury categories resulting from the traffic crash. While for the documented reasons for the motorcycle crash, alcoholism is repeatedly found in several clusters. For roads in poor condition as a reason for the crash, this reason is only found in a cluster related to the slight injury category. For severe injuries, reasons for the crash, over or inadequate speeding, drugs or medicines, objects or animals on the road, and meteorological factors are found to be influential reasons. In contrast to the applied binary probit regarding gender significance, the results of Kohonen show that males are more likely to be involved in different levels of injuries compared to females' different cluster combinations total numbers as this category represents 63.2% of the clusters developed for the slight injury and 73.3% of the clusters developed for the severe and fatal injury. For the time of the crash as a predictor, morning, night, and weekday periods are all showing higher trends of having different types of injuries. For the age categories, the group that ranges from 25-40 years old has the highest number of clusters for both injury classes, while the group that is 65 years old and older than 65 years has the least number of developed clusters related to motorcycle crash injuries. For both techniques that are employed in this study, both have shown good performance in analyzing the utilized dataset but with the approach in conducting the outputs and identifying the risk factors with their different categories. As a result, this study can be considered a step forward in detecting these risk factors that can increase the odds of having different types of injuries related to motorcycle crashes and providing different appropriate techniques that can detect these factors and examine the correlations between them and the level of injury that can cause whether the approach is a supervised technique or unsupervised technique.

9.7. References

Aiash, A., and Robusté, F. (2021). Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree. *International Journal of Injury Control and Safety Promotion*, *29*(2), 256-264.

Aidoo, E., Amoh-Gyimah, R., and Ackaah, W. (2021). The relationship between driver and passenger's seatbelt use: a bivariate probit analysis. *International Journal of Injury Control and Safety Promotion*, *28*(2), 179-184.

Albalate, D., and Fernández-Villadangos, L. (2009). Exploring Determinants of Urban Motorcycle Accident Severity: The Case of Barcelona. *XREAP*, 02.

Albalate, D., Fernández-Villadangos, L. (2010). Motorcycle Injury Severity in Barcelona: The Role of Vehicle Type and Congestion. *Traffic Injury Prevention*, *11*(6), 623-631.

Cirera, E., Plasència, A., Ferrando, J., and Seguí-Gómez, M. (2001). Factors associated with severity and hospital admission of motor-vehicle injury cases in a southern European urban area. *European Journal of Epidemiology*, *17*, 201–208.

Cunto, F., and Ferreira, S. (2017). An analysis of the injury severity of motorcycle crashes in Brazil using mixed ordered response models. *Journal of Transportation Safety & Security*, *9*(1), 33-46.

de Oliveira , L., de Oliveira , I., Caliari, P., de Mello, C., and Maia, M. (2021). Influence of demographic and socioeconomic factors on motorcycle usage in Brazil. *Case Studies on Transport Policy*, *9*(4), 1757-1769.

Fagnant, D., and Kockelman, K. (2015). Motorcycle Use in the United States: Crash Experiences, Safety Perspectives, and Countermeasures. *Journal of Transportation Safety & Security*, *7*(1), 20-39.

Ferrando, J., Plasència, A., MacKenzie, E., Orós, M., Arribas, P., and Borrell, C. (1998). Disabilities resulting from traffic injuries in Barcelona, Spain: 1-year incidence by age, gender and type of user. *Accident Analysis & Prevention*, *30*(6), 723-730.

Ferrando, J., Plasència, A., Orós, M., Borrell, C., and Kraus, J. (2000). Impact of a helmet law on two wheel motor vehicle crash mortality in a southern European urban area. *Injury Prevention*, *6*(3), 184–188.

Francis, F., Moshiro, C., and Yngve, B. (2021). Investigation of road infrastructure and traffic density attributes at high-risk locations for motorcycle-related injuries using multiple correspondence and cluster analysis in urban Tanzania. *International Journal of Injury Control and Safety Promotion*, *28*(4), 428-438.

Garrido, R., Bastos , A., de Almeida , A., and Elvas , J. (2014). Prediction of road accident severity using the ordered probit model. *Transportation Research Procedia*, *3*, 214 – 223.

Gazder, U., Almalki, Y., Alam, M., and Arifuzzaman, M. (2022). The effect of different mobile uses on crash frequency among young drivers: application of statistical models and clustering analysis. *International Journal of Injury Control and Safety Promotion*.

IBM. (2016). IBM SPSS Modeler 18.0 Algorithms Guide. IBM Corporation.

Islam, M. (2021). The effect of motorcyclists' age on injury severities in single-motorcycle crashes with unobserved heterogeneity. *Journal of Safety Research*, *77*, 125-138.

Kashani, A., Rabieyan, R., and Besharati, M. (2014). A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers. *Journal of Safety Research*, *51*, 93-98.

Kim, C.-Y., Wiznia, D., Averbukh, L., Dai, F., and Leslie, M. (2015). The Economic Impact of Helmet Use on Motorcycle Accidents: A Systematic Review and Meta-analysis of the Literature from the Past 20 Years. *Traffic Injury Prevention*, *16*(7), 732-738.

Konkor, I. (2021). Examining the relationship between transportation mode and the experience of road traffic accident in the upper west region of Ghana. *Case Studies on Transport Policy*, *9*(2), 715-722.

Konlan, K., Doat, A., Mohammed, I., Amoah, R., Saah, J., Konlan, K., and Abdulai, J. (2020). Prevalence and Pattern of Road Traffic Accidents among Commercial Motorcyclists in the Central Tongu District, Ghana. *The Scientific World Journal*,10.

Lemonakis, P., Eliou, N., and Karakasidis, T. (2021). Investigation of speed and trajectory of motorcycle riders at curved road sections of two-lane rural roads under diverse lighting conditions. *Journal of Safety Research*, *78*, 138-145.

Marquet, O., and Miralles-Guasch, C. (2016). City of Motorcycles. On how objective and subjective factors are behind the rise of two-wheeled mobility in Barcelona. *Transport Policy*, *52*, 37-45.

McCartt, A., Blanar, L., Teoh, E., and Strouse, L. (2011). Overview of motorcycling in the United States: A national telephone survey. *Journal of Safety Research*, *42*(3), 177-184.

Moskal, A., Martin, J.-L., and Laumon, B. (2012). Risk factors for injury accidents among moped and motorcycle riders. *Accident Analysis & Prevention*, *49*, 5-11.

Open Data BCN. (2021). Ajuntament de Barcelona's open data service. Retrieved from https://opendata-ajuntament.barcelona.cat/en/

Ospina-Mateus, H., Jiménez, L., Lopez-Valdes, F., Garcia, S., Barrero, L., and Sana, S. (2021). Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 10051–10072.

Pérez, K., and Santamariña-Rubio, E. (2019). Do advanced stop lines for motorcycles improve road safety? *Journal of Transport & Health*, *15*, 100657.

Pinch, P., and Reimer , S. (2012). Moto-mobilities: Geographies of the Motorcycle and Motorcyclists. *Mobilities*, *7*(3), 439-457.

Plasència, A., Borrell, C., and Antó, J. (1995). Emergency department and hospital admissions and deaths from traffic injuries in Barcelona, Spain. A one-year population-based study. *Accident Analysis & Prevention*, *27*(4), 591-600.

Rahman, M., Zafri, N., Akter, T., and Pervaz, S. (2021). Identification of factors influencing severity of motorcycle crashes in Dhaka, Bangladesh using binary logistic regression model. *International Journal of Injury Control and Safety Promotion*, *28*(2), 141-152.

Rice, T., Troszak, L., Ouellet, J., Erhardt, T., Smith, G., and Tsai, B.-W. (2016). Motorcycle helmet use and the risk of head, neck, and fatal injury: Revisiting the Hurt Study. *Accident Analysis & Prevention*, *91*, 200-207.

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Salum, J., Kitali, A., Bwire, H., Sando, T., and Alluri, P. (2019). Severity of motorcycle crashes in Dar es Salaam, Tanzania. *Traffic Injury Prevention*, *20*(2), 189-195.

Sivasankaran, S., and Balasubramanian, V. (2020). Exploring the severity of bicycle–vehicle crashes using latent class clustering approach in India. *Journal of Safety Research*, *72*, 127-138.

Sivasankaran, S., Rangam, H., and Balasubramanian, V. (2021). Investigation of factors contributing to injury severity in single vehicle motorcycle crashes in India. *International Journal of Injury Control and Safety Promotion*, *28*(2), 243-254.

Sohadi , R., Mackay, M., and Hills, B. (2000). Multivariate Analysis of Motorcycle Accidents and the Effects of Exclusive Motorcycle Lanes in Malaysia. *Journal of Crash Prevention and Injury Control*, *2*(1), 11-17.

Soltani, A., and Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, *48*(3), 637-647.

Teoh, E., and Campbell, M. (2010). Role of motorcycle type in fatal motorcycle crashes. *Journal of Safety Research*, *41*(6), 507-512.

Vajari , M., Aghabayk, K., Sadeghian, M., and Shiwakoti, N. (2020). A multinomial logit model of motorcycle crash severity at Australian intersections. *Journal of Safety Research*, *73*, 17-24.

Weiss, H., Kaplan, S., and Prato, C. (2016). Fatal and serious road crashes involving young New Zealand drivers: a latent class clustering approach. *International Journal of Injury Control and Safety Promotion*, *23*(4), 427-443.

Yu, R., and Abdel-Aty, M. (2014). Using hierarchical Bayesian binary probit models to analyze crashinjury severity on high speed facilities with real-time traffic data. *Accident Analysis & Prevention*, *62*, 161–167.

10. Conclusions and future work

As shown and through the eight conducted papers in this thesis, several results were extracted, and different conclusions were conducted. For the overall traffic crash status in Barcelona, this study consists of 47,153 cases between the years 2016 and 2019. Two models were employed in this study to analyze the utilized data with the chosen risk factors, including the binary probit and CHAID tree. The results have shown that two road users are at higher risk, which are pedestrians and drivers. For the temporal factors, the results have shown that afternoon, weekend, and night timing periods are all having a high risk of severe and fatal injuries compared to other timing periods. A Bayesian network is employed in a study that focuses on the working hours timing and seasons of the year as risk factors related to traffic crashes. The findings have shown that both summer and working hours timing periods are showing higher risk to road users. The COVID-19 pandemic in Barcelona and its impact on traffic crashes has been examined through one of the conducted studies in this thesis. Several approaches were utilized to grasp this impact. The drop in traffic crash injuries was captured, showing a reduction of 39%, 30%, and 36% for slight, severe, and fatal injuries, respectively. Part of the findings of this study, males, evening, and night timing periods are showing similar impacts on traffic crash severity before and during the pandemic period after employing the CHAID tree model.

Another risk factor that has been examined is the total noise levels. Two neighborhoods' traffic crashes and noise data were compared and analyzed in order to determine the potential impact of total noise levels on road users. The findings have shown that the area with higher total noise levels may be attributed to a higher risk of having traffic crashes after applying a logistic regression model, and further studies should be conducted. The district's traffic crash intensity differences were analyzed and detected after employing a Bayesian network. The district with the highest population, private transport mode usage, and highest density of passenger cars per km^2 is found to have the highest risk of having severe traffic crashes and fatal injuries.

The results of the three studies that have considered the three vulnerable road users: pedestrians, cyclists, and motorcyclists are as follows. Passages regulated by traffic lights show higher probabilities of having severe and fatal injuries compared to other crash sites. The evening period is also found to be linked to a higher risk of severity and fatality among pedestrians. Elderly pedestrians are found to be more involved in severe and fatal injuries mostly during the morning period. While cyclists, the age group category between 25 to 50 years old during the morning and evening period, are at a higher risk in case of a traffic crash. Similar to the previous age category, motorcyclists who are between 25 to 40 have the highest probability of having traffic crash severities. Additionally, roads in poor condition and alcoholism can have a negative impact on motorcyclists, which eventually can lead to a traffic crash.

Part of the future work is collecting further datasets with additional details and risk factors that were still not considered in this thesis. Comparative analysis studies can also be part of future work that includes other cities around the world to examine the impact of several risk factors among these cities and how sociodemographic differences can play a role in traffic crash severities.

Several measures and recommendations can also be extracted based on the findings that are concluded. It has been shown that pedestrians are still at high risk compared to other road users when overall analyzing the traffic crash status in Barcelona in the first paper of this thesis. Therefore, another paper is conducted to meticulously grasp the factors that can lead to higher levels of injuries. The results have shown that passages regulated by traffic lights may represent a higher risk for pedestrians, which may be attributed to the high usage of this type of passage by pedestrians. Traffic calming at these designated passages is required to ensure speed reduction. Speed bumps, humps, and raised pavements can be part of the solution. Signals that can be activated by pedestrians are also recommended so that pedestrians can utilize them rather than getting used to ignoring traffic lights when the traffic volume is light. Some areas, such as the intersections that have parking nearby the passage regulated by a traffic light, are recommended to reduce or eliminate such parking from these places due to the risk that can present for pedestrians and other road users such as cyclists.

The timing period is another influential factor that impacts traffic crash severity. Evening and night timing periods have shown higher possibilities of having severe and fatal injuries. Usually, the illumination of the road can play a role in this case. Additionally, work time flexibility should be implemented to ensure that not all employees are going back home at similar times, especially during the evening period, due to many other factors that can impact road users, such as tiredness and traffic volume. For age as a risk factor, elderly road users are showing the highest risk compared to other age categories. Several measures can be followed by both the legislators and elderly road users to increase their safety and reduce the possibility of being involved in a traffic crash. The driving license should be renewed every period of time and more frequently to ensure that they can pass the sight test and other possible tests that can affect their abilities and impact their overall skills driving. A doctor consultation is also required to check if the medicines that are taken can impair their ability to drive. Another finding, male drivers have shown a higher risk of getting involved in severe or fatal injuries compared to females. Certain campaigns can target males to ensure that they grasp the risk of being involved in traffic crashes and consequential events that may occur in case of not cautious, especially while driving.

For the COVID-19 pandemic period, the risk factors that have previously been depicted have shown similar impacts before and during this period which indicates the importance of providing the appropriate solutions to lessen their impact. Additionally, the maintenance of the road during the pandemic period is affected negatively, which has negatively impacted the severity of the traffic crash. This shows the importance of maintaining the road in good condition for the road users to ensure their safety. Alcohol consumption has increased during this time period which can also increase the risk of traffic crashes. Tests for alcohol consumption is highly recommended to be implemented for drivers more frequently to ensure the enforcement of traffic rules.

Another important aspect that should be considered by traffic regulators is the total noise levels in different areas. Monitoring of highly noisy areas is required by the authorities to maintain low levels of total noise, especially noise levels higher than 70 dP. As shown in this thesis, this factor may represent a threat to drivers and eventually may cause a traffic crash. Additionally, this study confirms the role of this factor on traffic crashes and the necessity to conduct further studies related to this risk factor as up to our knowledge, this risk factor has still not been examined until the paper that has been conducted in this thesis. Therefore, part of the future work can be an extensive examination of this risk factor in order to accurately identify the impact of this risk factor. This can include lab experiment on certain participants under different noise level conditions and then measure the impact of distraction that can affect them.

For cyclists' safety, from the findings of this thesis, it can be seen that different types of cycle lanes are efficient in protecting cyclists from severe and fatal injuries. Therefore, these lanes are highly recommended to be widely disseminated through the city on a larger scale with more focus on the separated cycle lane as it has shown higher safety efficiency compared to the other types. Additionally, 30 km/h speed limit zones have also proven their efficiency in alleviating the risk for cyclists in being involved in severe or fatal injuries, which should also be considered by the legislators to include more areas that have these speed limits not only it can save cyclists but also can have an impact on the total noise levels that are generated from vehicles which can also decrease the impact of noise on road users in general and ensure the level of distraction by noise is low.

For motorcyclists, a road in poor condition can be a source of risk for them, which indicates maintaining the road in good condition is not only required for smooth riding but also for safer rides for all road users, not just for the motorcyclists. Again, for speeding, over-speeding appears to be an issue that danger the safety of motorcyclists, which requires more areas with a speed limit of 30 km/h besides ensuring that everyone is complying to the designated speed limit.