INSTITUTO
UNIVERSITÁRIO
DE LISBOA

A data-driven approach to road accidents in the municipality of Lisbon

João Carlos Pereira Sousa Macedo Mesquitela

Master in Computer Engineering

Supervisor:
Prof. Doctor Luís Miguel Martins Nunes, Associated Professor
ISCTE-IUL

Co-Supervisor:
Prof. Doctor João Carlos Amaro Ferreira, Assistant Professor, with habilitation
ISCTE-IUL

July, 2021

Page left intentionally blank

iscte
TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

A data-driven approach to road accidents in the municipality of Lisbon

João Carlos Pereira Sousa Macedo Mesquitela

Master in Computer Engineering

Supervisor:
Prof. Doctor Luís Miguel Martins Nunes, Associated Professor
ISCTE-IUL

Co-Supervisor:
Prof. Doctor João Carlos Amaro Ferreira, Assistant Professor, with habilitation
ISCTE-IUL

July, 2021

Page left intentionally blank

To my wife Ines and my daughter Joana for their love, unconditional support, encouragement, patience, and help overcome the obstacles that arose along the way. This work, as well as my life, is dedicated to you!

Page left intentionally blank

Acknowledgments

This document on data-driven analysis of road accidents in Lisbon municipality derives from the author's work. However, it was only possible thanks to the contribution of several people and entities to whom we express our sincere thanks:

To ISCTE - Instituto Universitário de Lisboa for the opportunity granted;

To Professors Luis Nunes and João Fernandes, co-supervisors of this work, I thank their support, knowledge sharing, and valuable contributions to this dissertation. Above all, for following me throughout this journey and stimulating my interest in knowledge and academic life. As well as for the friendship and availability;

To the teachers of ISCTE - Instituto Universitário de Lisboa, for the relevant contribution to my scientific training;

To the Autoridade Nacional de Segurança Rodoviária, in the person of Dr. Ana Maria Coroado for monitoring and providing the needed data;

To all friends and colleagues who, directly or indirectly, contributed or helped prepare this work for the patience, attention, and strength they provided in less easy moments.

To all, my sincere and profound Thank You!

Page left intentionally blank

Resumo

Os acidentes de trânsito em áreas urbanas conduzem à redução da qualidade de vida e à desigualdade social nas cidades, especialmente nos países em desenvolvimento. O crescimento da malha urbana, assim como, a densidade populacional raramente é acompanhada pelo desenvolvimento ou dimensionamento da infraestrutura rodoviária.

É um facto que o número e a gravidade dos acidentes rodoviários em Portugal têm vindo a diminuir ao longo dos últimos trinta anos, o que permitiu aproximarmos da média Europeia, apesar destes factos a situação continua a ser preocupante.

Apesar da adoção de programas como o Programa de Segurança Rodoviária da Comissão Europeia ou, numa base nacional, o PENSE 2020 - Plano Estratégico Nacional para a Segurança Rodoviária os números de acidentes de viação com vítimas no distrito de Lisboa continuam a ser mais elevados do que a média europeia.

Desta forma e para efeitos deste trabalho realizamos uma análise de dados exploratória (AED) aos dados dos incidentes de transito registados no sistema de gestão de ocorrências do Regimento de Sapadores Bombeiros de Lisboa e a os dados de acidentes rodoviários reportados à ANSR pelas forças de segurança (GNR e PSP) através do Boletim Estatístico de Acidentes de Viação (BEAV) e ocorridos no concelho de Lisboa entre 2010 e 2020 por forma a identificar a existência de Pontos Negros nas vias de Lisboa e quais os fatores mais significantes e contribuintes que permitam explicar a sua existência.

Os dados relativos aos acidentes rodoviários foram também georreferenciados para capitalizar a sua existência espacial e, consequentemente, compreender melhor os padrões espaciais existentes e os fatores de risco.

Posteriormente através do recurso ArcGIS Pro aplicaram-se os algoritmos das ferramentas Densidade de Kernel e Hot Spot Analysis (Getis-Ord Gi*), identificando a existência dos pontos negros, e que fatores humanos, ambientais e circunstanciais têm influência na gravidade dos acidentes e que algumas variáveis de exposição foram consideradas importantes na explicação da ocorrência dos mesmos, sendo a validade do conteúdo garantida através de uma comissão de especialistas.

Pretende-se, assim, contribuir para a identificação das zonas de concentração de acidentes da cidade de Lisboa (hotspots), tendo em conta as suas condições influenciadoras. Potenciando a segurança rodoviária no município.

Palavras-chave: Acidentes Rodoviários, Pontos Negros, GIS, Analise de dados

Page left intentionally blank

Abstract

Traffic accidents in urban areas lead to reduced quality of life and social inequality in cities, specially in third world countries. The growth of the urban mesh and the population density is seldom accompanied by the development or sizing of the road infrastructure.

It is a fact that the number and severity of road accidents in Portugal have been decreasing over the last thirty years, bringing us closer to the European average. However, despite these facts, the situation remains worrying.

Despite the adoption of programs such as the European Commission Road Safety Program and the recent EU Road Safety Policy Framework 2021-2030 or, on a national basis, the PENSE 2020 - National Strategic Plan for Road Safety the number of road accidents with victims in the district of Lisbon is still higher than the European average.

Thus, and for this dissertation, we conducted an exploratory data analysis (EDA) on the combined data of traffic incidents recorded in the occurrence management system of the Lisbon Fire Brigade Regiment (RSB) and the road accidents reported to ANSR by the security forces (GNR and PSP) through the Statistical Bulletin of Traffic Accidents (BEAV). Furthermore, with data from occurrences in the Municipality of Lisbon between 2010 and 2020, to identify the existence of Black Spots in Lisbon's roads and which are the most significant and contributing factors to explain their existence.

The data on road accidents were also georeferenced to capitalize their spatial existence and, consequently, better understand the existing spatial patterns and risk factors.

Subsequently, through the use of the ArcGIS Pro we apply the algorithms of the Kernel Density and Hot Spot Analysis (Getis-Ord Gi*) tools, identifying the existence of the black spots, and that human, environmental and circumstantial factors have an influence on the severity of accidents, being the content validity guaranteed through an expert committee.

This way, our research goal is to contribute to identify accident concentration areas in the city of Lisbon (hotspots), considering their influencing conditions.

Keywords: Road Accidents, Black Spots, GIS, Data Analysis

Page left intentionally blank

Glossary

Accident with victims - An occurrence on the public highway involving at least one vehicle registered by the official entities, resulting in victims or material damage. Accidents with victims can be fatal/deadly, severe, or light depending on the victims' injuries' severity.

Categorical variable - A statistical variable measured on a nominal scale.

Dead or Fatal Victim (30 days) - Victim with injuries resulting from the accident whose death occurs within 30 days of the accident.

Death or fatality (local) - A victim of an accident who dies either at the accident scene or on the way to the Hospital.

Driver - Person in charge of a vehicle on the public highway.

Fatal accident - Accident with at least one fatal victim regardless of the injury severity of any other involved persons.

Minor Accident - An accident resulting in at least one minor injury and in which no fatalities, deaths, or serious injuries are registered.

Passenger - A person assigned to a vehicle on the public highway which is not a driver.

Pedestrian - Person on foot. A person pushing or holding a bicycle.

Road victim - A human being who suffers bodily harm as a result of a road accident.

Serious accident - An accident resulting in at least one seriously injured person, with no death.

Seriously Injured - Victim of an accident whose body injuries result in a period of hospitalization of more than 24 hours.

Slightly Injured - Victim of an accident who is not considered seriously injured.

Some of the definitions in this glossary have been adapted from [1][2][3]

Page left intentionally blank

Index

List of Figures

List of Tables

Acronyms

AADT - Annual Average Daily Traffic

ANSR - National Road Safety Authority

BEAV - Road Accidents Statistics Bulletin

CRISP-DM (Cross Industry Standard Process for Data Mining)

CVI- Content Validity Index

CVR - Content Validity Ratio

DGV - General Direction of Traffic

ENSR - National Road Safety Strategy

GIS - Geographic Information System

GNR - National Republican Guard

GPS - Global Positioning System

IMTT - Mobility and Land Transport Institute

ISCTE-IUL - University Institute of Lisbon

I-CVI - Item Content Validity Index

PSP - Public Security Police

WHO – World Health Organization

ZAA - Accident Accumulation Zones

Page left intentionally blank

CHAPTER 1

Introduction

1.1. Motivation and Framework

Recent decades have witnessed a significant increase in the motorization rate worldwide (image 1). Associated with urban expansion caused by population growth and migration to large urban centres in recent decades, this increase has led to increased Road Traffic Accidents (RTA) in urban areas. This urban growth requires continuous improvement in the capacity and dimensioning of road networks, which, as a rule, hardly ever happens. This situation puts the safety of the most vulnerable users (pedestrians, cyclists, and drivers of two-wheeled motor vehicles) at risk. Consequently, RTA is the leading cause of death and /or disability worldwide.

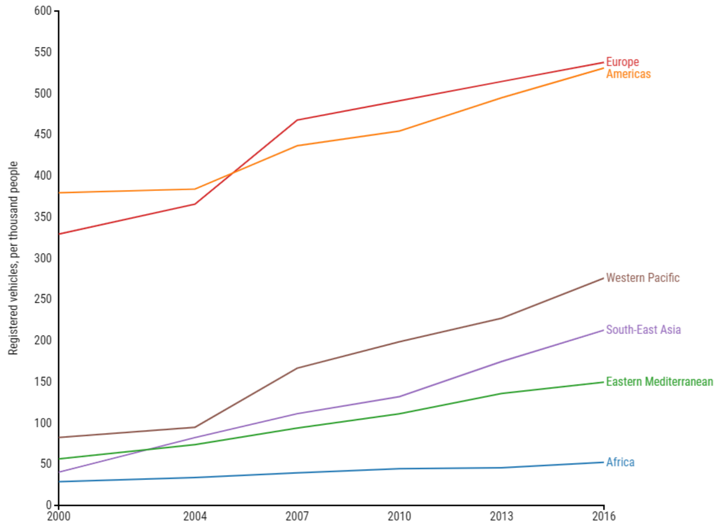

Image 1 - Registered vehicles in the world, per capita [4]

According to the World Health Organization's (WHO) global report on the state of road safety [5], there are more than 1.35 million traffic fatalities annually due to RTAs. For every fatality in a road accident, more than five people, approximately, end up seriously injured (20-50 million) with irreversible consequences for their lives. Severely injured people are often costlier to society because of the need for rehabilitation and long-term care.

The same document states that road accidents were, in 2018, the eighth leading cause of death for all age groups. In the 5 to 29 age group (children and young adults), they are the leading cause of death. Moreover, the same study predicted that this will be the fifth leading cause of death in 2030 (image 2).

2004

Rank	Disease or injury
1	Ischaemic heart disease
2	Cerebrovascular disease
3	Lower respiratory infections
4	Chronic obstructive pulmonary disease
5	Diarrhoeal diseases
6	HIV/AIDS
7	Tuberculosis
8	Trachea, bronchus, lung cancers
9	Road traffic injuries
10	Prematurity & low-birth weight

2030

Rank	Disease or injury
1	Ischaemic heart disease
2	Cerebrovascular disease
3	Chronic obstructive pulmonary disease
4	Lower respiratory infections
5	Road traffic injuries
6	Trachea, bronchus, lung cancers
7	Diabetes mellitus
8	Hypertensive heart disease
9	Stomach cancer
10	HIV/AIDS

Image 2 - Top 10 of the Leading causes of, 2004 and 2030 compared [6]

The implementation of several concerted intervention strategies at the Level of the 27 EU Member States (Local, National, and European) made it possible to reduce the number of road fatalities by 43% between 2001 and 2010 and 21% between 2010 and 2018. However, despite the efforts of the authorities, about 22.800 deaths were recorded in RTAs in 2019 [6],[7]. Although the underlying trend is downward (image 3), progress has slowed in most countries since 2013, and the EU target of halving the number of road fatalities by 2020 (compared to the 2010 baseline scenario) may not be achieved. According to data from the European Commission, 2020 was an exception in that the number of fatalities decrease significantly due to the measures taken to combat the pandemic (lockdown), but even that significant decrease was not be enough to reach the target [8].

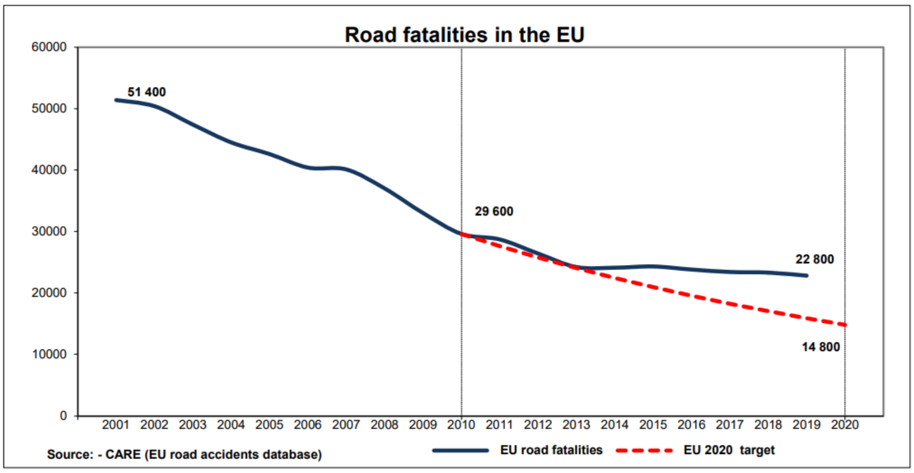

Image 3 - Downward trend in the number of road traffic fatalities in the EU [8]

The human and social cost caused by mobility is high and unsustainable. In financial terms, the annual burden of RTAs in the EU is €280 billion, or about 2% of the GDP, according to a recent report [9].

At a national level, measures have been implemented to minimize this problem and reduce the number of fatalities. These measures have had positive results. Between 2000 and 2018, Portugal reduced the number of deaths per million inhabitants by 58%, ranking 11th on the list of countries that have reduced the number of deaths the most, although four percentage spots above the EU28 average (-26%).

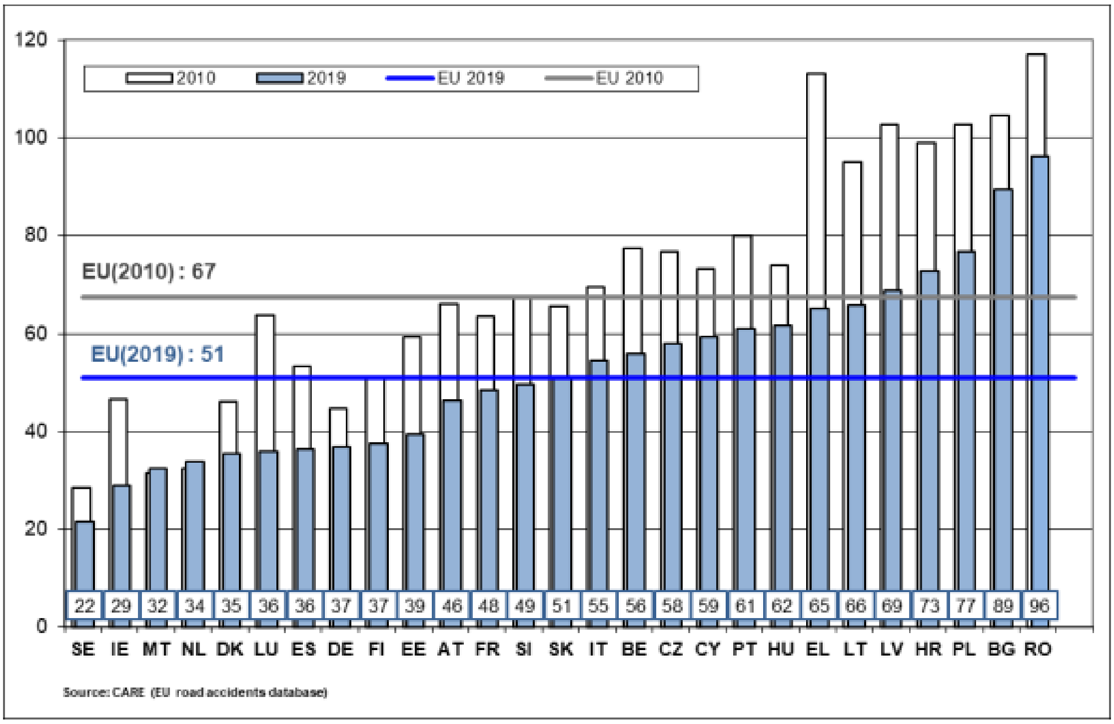

Image 4 - Trend in road fatality numbers per million inhabitants by country [8]

Data published by the Portuguese National Road Safety Authority (ANSR) [10], for the period from 1995 to 2019, showed that investments in infrastructure and safer vehicles, associated with integrated road safety policies, were decisive in saving 26.000 lives and avoiding 187.000 serious injuries. In addition to the lives saved, the economic savings were 4.5 times greater than what would have been spent if the investment had not taken place.

In this context, one might think that fatalities are gradually becoming a less serious problem worldwide. However, despite the decrease in the number of casualties and accidents, RTAs are still a worrisome problem due to the increase in the number of vehicles. According to accident data published by the Portuguese National Road Safety Authority, the number of road accidents inside cities is higher than the country, but the severity index is higher in the latter.

Recognizing this problem, both European and national organisations responsible for road prevention and safety have committed to developing and implementing prevention and driver awareness programs. These activities are part of the creation and implementation of programs such as the European Commission's Road Safety Programme (Towards a European road safety area: policy orientations on road safety 2011-2020) [11] and the recent EU Road Safety Policy Framework 2021-2030 [12]. At the national level, they include the PENSE 2020 - National Strategic Plan for Road Safety [13] and the still-developing National Road Safety Strategy 2021-2030 – Vision Zero 2030.

According to the final document, this "intends to continue to address, during its term/validity period, national indicators of the European average and, in the medium term, of the countries with the best historical performance in these indicators" based on the following strategic objectives: i. Improve road safety management; ii. Increase insurance safety; iii. To have safer infrastructures; iv. To have safer vehicles; v. To improve assistance and support to victims.

As mentioned above, Portugal has had great success in reducing road fatalities by implementing road prevention policies at a national level. As the number of incidents decreased, new and ambitious goals were set that required the strict participation of all municipalities, which allowed for greater precision in the implementation of national policies. In this process, guidelines were made to prepare Municipal Road Safety Plans under the National Safety Strategy (NRSS) so that each municipality could identify its role in achieving the national goals.

The new road safety policy is essentially focused on a specific territorial region, the municipality. Therefore, the research for this dissertation focused on the need identified by the Mobility Division of the Municipality of Lisbon with the goal of identifying the road accidents incidence spots and their correlation with other factors. Referred to in the literature as “hotspots”, these spots were identified at the municipal level, and a proposal for their correction was developed.

The proposed measures are part of the calming traffic methodology, which reduces accident indicators by lowering traffic speeds and controlling traffic flows. One of the significant difficulties of the bodies responsible for road safety in Portugal and the rest of the world is knowing where and how to implement appropriate road prevention measures to reduce accidents. In this case, computer tools, GIS, and spatial analysis are particularly relevant for the present work. This research addresses a current, critical topic for emergency managers and provides a technical asset for mobility managers in Portugal.

1.2. Goals and Research Questions

The main research goal of this dissertation is to identify the road traffic accident-prone zones (hotspots) in the Lisbon municipality and correlate them with their influencing factors. This investigation responded to challenge number 51 from the Lisboa Inteligente—LxDataLab project[1] [14].

This challenge from Lisboa Inteligente aims to give the municipality authorities the necessary data to improve the city's traffic planning and management by implementing steps to limit the number of accidents and victims in these locations.

With this motivation, it is established as main objectives for the present study:

- To perform a spatial-temporal statistical analysis of road accidents with victims in the Municipality of Lisbon;
- To identify the roads with a higher occurrence of road accidents with victims.
- Characterization of the roads and areas identified as problematic.
- Temporal Analysis (Biennial and Seasonal) of the identified hotspots
- Identification of the RTAs Hotspots and the pinpointing of their causal factors.
- Creation of digital artifacts (images and shapefiles) with the Hotspots locations in the municipality.

To achieve this purpose, a spatial-temporal multivariable analysis of the road traffic accident (with victims) between 2010 and 2019 in the Lisbon council municipality was conducted. Geographic Information Systems (GIS) combined with spatial and statistical analysis algorithms and tools were used to identify the presence and location of hotspots and their relation with diverse factors (human, environmental and vehicular) that contribute to their existence or activation.

[1] A collaboration between Lisbon City Council (CML) and several universities and research institutions, including ISCTE-IUL, whose mission is to create analysis and visualization solutions supported by Big Data to improve the planning, management, and identification of Lisbon´s emergencies.

Geographic Information Systems (GIS) allows researchers to use various computational methods and tools to combine geographic data, statistical data, and mapping data to identify the spatial characteristics of a hotspot [15]. According to Filipe Ferreira Esteves [16], "GIS is computational tools that allow the integration and manipulation of different types of information, especially suitable for spatial variables of global, regional or local nature. They are a decision support system that involves integrating georeferenced data in a problem-oriented environment, particularly those where the spatial component is strongly present". Identifying these critical spots through GIS is especially relevant, as it makes it possible to understand the causes and factors associated with accidents, allowing consistent decisions by the city council's mobility department.

For that, it is essential to identify the existence of a pattern of spatial clusters in the accident data, using the following two spatial analysis algorithms:

Kernel density estimation (KDE), which is a technique for interpolating and analysing spatial patterns of spots. From a set of known spots, it identifies the intensity with which a given variable occurs in space. It is ideal for formulating explanations and illustrating conclusions and is an easily understandable statistical method for non-mathematicians. With GIS, it is possible to visualize the concentration of processes and describe process changes at the local level.

Hot spot analysis in a neighborhood context, which is based on the Getis-Ord Gi* statistical calculation, presents the high (hotspot) and low (cold spot) clustering values, resulting in the z-score (the number of standard deviations from the mean of an information spot) and p-value (the probability of obtaining the observed results of a test, assuming the null hypothesis is correct). An area with a high p-value is not necessarily a statistically significant hotspot. A spot must have a high value and be "surrounded" by other spots with equally high values to be a statistically significant hotspot. The local sum of a location and its neighbours is compared proportionally to the sum of all the spots. When the local sum is quite different from the expected local sum, and if this difference is too large, a statistically significant z-value occurs.

1.3. Document Structure

After defining the objectives, this dissertation consists of six chapters, each of which reflects a different step of the research process.

The present chapter provides an overview of the research theme, aims, methodology applied, and a summary of the organization.

Chapter 2 – Literature Review: This chapter explores the theoretical framework necessary for the development of this dissertation.

Chapter 3 – Business Understanding: This chapter introduces the CRISP-DM Methodology. It explains RTAs, the city of Lisbon's emergency system, and the territory characteristics based on the author's 20 years of professional experience in the National Authority for Emergency and Civil Protection.

Chapter 4 – Data Understanding and Data Preparation: This chapter is devoted to the Data Understanding and Data Preparation Phases of the CRISP-DM methodology. It describes the data gathering and treatment procedures.

Chapter 5 – Data Visualization: This chapter analyses the data presented by the data visualisation artifacts that have been created to provide the necessary support to CML decision-makers. It also contains the content validation analysis provided by the Expert Committee.

Chapter 6 – Conclusions: The study's conclusions, as well as recommendations, research limitations, and a proposal for further work are presented in the sixth and final chapter.

1.4. Methodology

1.4.1. Project development methodology

Data analysis is the process of using tools, techniques, and methods to examine in-depth data to identify meaningful information that can be applied to make decisions or solve problems. According to a simplified view, preparation, preprocessing, analysis, and post-processing are the four main phases of a data analysis project.

We tried to follow best practices during the dissertation development to increase the chances of success. There are various implementation methodologies in a data extraction project (information). One of the most widely used methodologies used in this dissertation is the framework considered by Colin Shearer [16] as a standard for implementing data mining projects, i.e., CRISP-DM (Cross-Industry Standard Process for Data Mining).

The life cycle of this framework has six continuous and interactive, non-sequential phases. Although this is a data mining cycle, CRISP-DM allows feedback on any phase as new data is added to the process.

To adapt this methodology to the dissertation, we redesigned the CRISP-DM life cycle (image 6) from the one proposed by Ana Oliveira et al. [17]. In addition, we shifted the approach toward a more targeted data visualization, considering all the attention applied to the data, the data fusion, and the Visualization of the results.

In that sense, the new adapted framework phases are:

Image 5 - Phases of our altered CRISP-DM Process Model, adapted from [17] *and [21]*

Phases of CRISP-DM

1. Business understanding – This phase consists of identifying the aim and requirements of the challenge proposed by challenge number 51 from the Lisboa Inteligente—LxDataLab project [14]. After this assessment, the entire literature review process (theoretical framework) on the universe of road accidents was initiated. At this stage, after the proper framework on the

subject, the strategy for approaching the research questions defined by the challenge under analysis is also defined.

2. Data understanding – Phase for acquiring the necessary data, describe and explore its quality.

3. Data preparation – Phase of data preparation and cleaning (Data Featuring; Inconsistencies; Duplicates; Outliers; Missing Data, etc.).

4. Data Fusion – In this phase, the ETL is created, all the data sources (accidents, weather, traffic) are aggregated in a unique dataset.

5. Data Visualization – Phase where all the digital artifacts are created to depict the temporal and spatial data, as well the artifacts with identification of the hotspot's areas are produced to support the CML requirements.

6. Decision Support – The authorities receive the statistical analysis of road accidents with victims in the municipality and the digital artifacts produced with the pinpoint locations of the city's hotspots and the identification of the underlying factors that induce or influence their existence. This way, they will have data contributing to the decision-making process regarding the safety measures to be applied to mobility in Lisbon.

Image 6 - Generic tasks and outputs of the CRISP-DM reference model, adapted from[17] *and [21]*

Page left intentionally blank

CHAPTER 2

Literature Review

2.1 The Approach Followed for the Literature Review

To achieve the proposed objectives and answer the research objectives mentioned earlier, the author opted for a hybrid approach and combined two methodologies to conduct the systematic literature review. The PRISMA *Statement - Preferred Reporting Items for Systematic Reviews and Meta-Analyses*, whose objective is to ensure that systematic reviews (SR) and meta-analyses are done in a thorough, clear, and precise manner; and, complementarily, the *snowball sampling* methodology (non-probabilistic sampling using reference chains), a technique that searches for new sources of information based on the references used by the documents that scientifically support this dissertation.

Therefore, the systematic search was done in three academic search engines: Scopus[2], Biblioteca do Conhecimento Online[3] , and Web of Science[4]. The first is the primary source of the articles consulted.

The time interval considered for these articles was the publication date between 2010 and 2020.

The location of the articles needed for this dissertation was made through different combinations of groups of keywords. Through them, the existing contents in the respective databases were extracted and filtered:

1. ("Road Traffic Accidents" AND "analysis" AND Portugal AND spatial pattern);
2. (RTA AND accident AND mapping AND spatiotemporal analysis);
3. (Traffic accident AND GIS, AND Kernel density estimator AND KDE+ OR "Getis-Ord Gi*");
4. (“Lisbon” AND “RTA” AND "GIS";
5. ("KDE" AND "KDE+") AND (hot spot analysis);
6. (Lisboa AND Acidentes de transito);
7. (human factors AND road accidents).

The results obtained were filtered according to a rationale applied to the *Covidence*[5] systematic literature review tool, consisting of three steps:

[2] https://www.scopus.com/
[3] https://www.b-on.pt/
[4] https://www.webofknowledge.com/
[5] https://app.covidence.org/

1. Eligibility Criteria – at this stage, the results are approved or rejected according to the following criteria:
    i) Language (only articles written in English or Portuguese are accepted);
    ii) Date of publication of the article between 2010 and 2021;
    iii) Scientific literature exclusively from scientific articles or reviews, apart from grey literature from the Lisbon City Council, the National Road Safety Authority, and the European Commission. These, for being intervention plans, technical documentation (e.g., Municipal Emergency Plan of the Civil Protection of Lisbon) and reports, cannot be published in the scientific community; and the article The CRISP-DM Model: The New Blueprint for Data Mining, for not meeting the publication date criterion.
    iv) The topic is relevant to this dissertation.
2. Reading the article abstract;
3. Reading the full article.

After analyzing and filtering the results, the following values (image 7) were obtained:

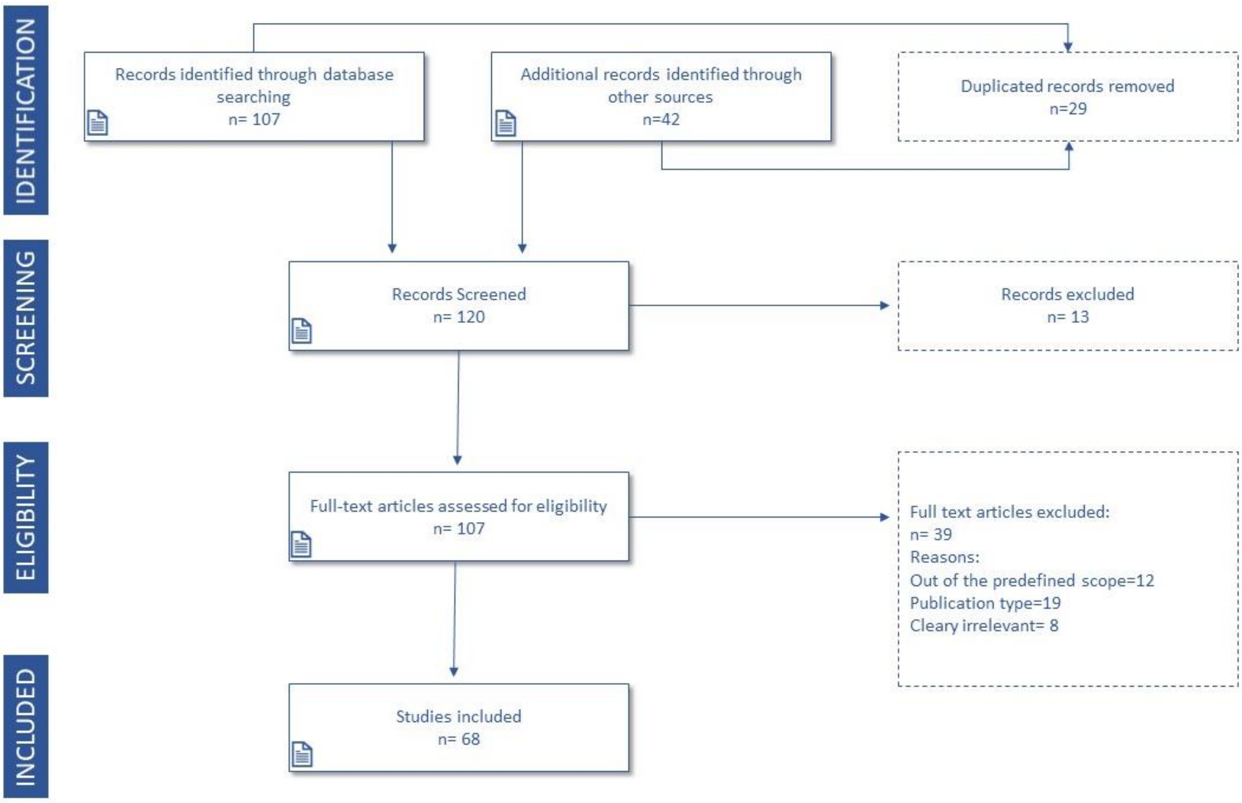

Image 7 - PRISMA Diagram of this Dissertation

Then, VOSviewer[6] was selected to create visualizations of the keyword groups (Image 8) used by the authors of the studies chosen for this dissertation.

The map visibly highlights the term "*GIS*" as the central concept of the links and relationships established, followed by the term "*Road Safety*", traffic accidents, and Hotspots.

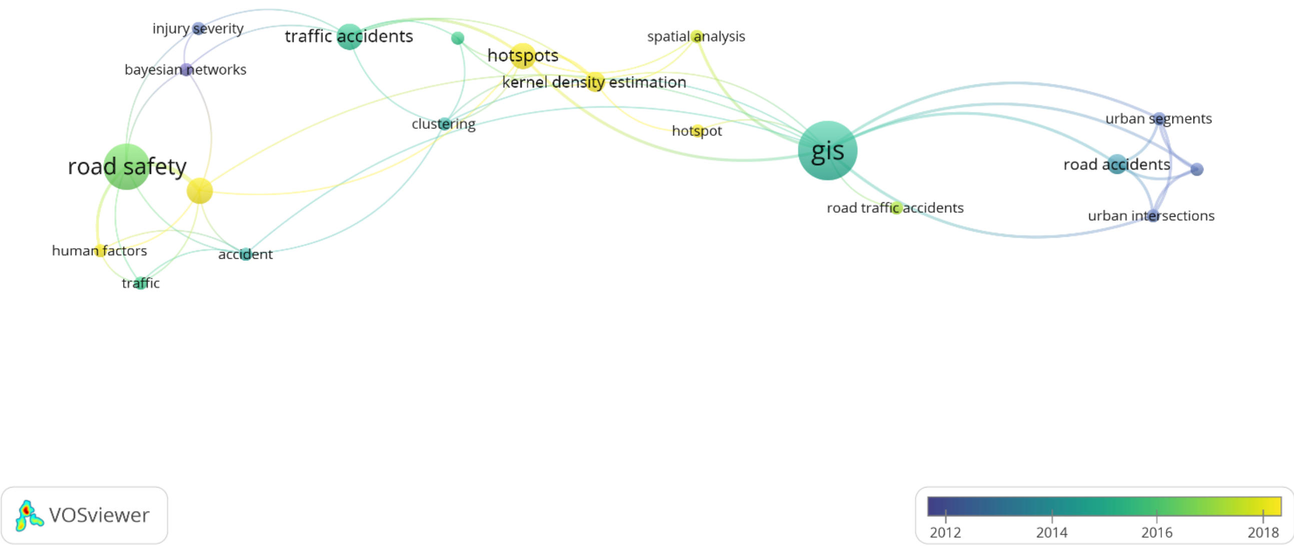

Image 8 - VOS Viewer keywords Visualization

A critical component of reducing road traffic accidents is analysing the locations of the incidents. This is founded on the recognition that a road traffic accident by definition has a geographical component, as it occurs in a specific area. Additionally, road traffic accidents do not occur at random. They are prone to congregate in particular locations for reasons that can be explained by a variety of circumstances referred to as "influencing factors." These clusters, or concentrations of RTAs, which are usually referred to in the literature as "hotspots", are described as "every spot that has a higher number of accidents than other similar spots due to local risk factors"[18].

This notion refers to the concept that hotspots are prone areas where the geometry and the traffic design (e.g., congested intersections, sharp curves, inefficient vertical signalization, etc.) play a significant role in accidents, which can be reduced if the influencing factors are identified and corrected by the authorities.

[6] https://www.vosviewer.com/

There is no universally accepted and precise definition of accident hotspots [19] as definitions differ among and are adapted to suit each country's features and governmental goals. Rune Elvik compiled hotspot definitions in European countries [20] in order to demonstrate the differences.

According to the literature review and the comparison of approaches for defining hotspots (Table 1), no European country has fully adopted an identical methodology for determining hotspots. Applied methods of identification, each with distinct strengths and weaknesses in their field of application, are beneficial in locating hotspots. In view of the above, this dissertation uses the hotspot definition provided by the Portuguese National Road Safety Authority (ANSR): "A stretch of road with a maximum length of 200 meters, in which there were at least five accidents with victims in the year under review" [10].

Table 1 - Summary of Hotspot definitions found in literature, adapted from [20] [21],[22]

Country	RTA Hotspot Definition
Austria	Three or more similar accidents with injuries within three years period and a risk coefficient of at least 0.8
Belgium	Three or more severe accidents within 100 meters stretch of road over the period of 1 year
Czech Republic	Three or more accidents with injuries within 250 meters stretch of road over the period of 1 year, or, 3 similar accidents in the period of 3 years, or, 5 similar accidents in the period of 1 year.
England	More than Twelve accidents within 300 meters stretch of road over the period of 3 years
Germany	Three accidents within 300 meters stretch of road over the period of 1 year, or, 5 accidents over the period of 3 year
Hungary	Three or more accidents within 100 meters stretch of road over the period of 3 year
Netherlands	Five accidents on a single location over the period of 3 year, or, 10 accidents over the period of 5 year
Norway	Four or more severe accidents within 100 meters stretch of road over the period of 1 year
Portugal	Five or more severe accidents within 200 meters stretch of road over the period of 1 year and indicators with severity up to 20 in one year of the analysis.
Spain	The country lacks consistency in its definition of the concept. The method for calculating Hotspots differs per administrative division, e.g.: Andalucia - Five or more severe accidents within 1000 meters stretch of road over the period of 1 year, or, three accidents with injured victims (at least 1 deceased) in the last three years; Valencia – Three or more accidents within 200 meters stretch of road over the period of 3 year

The academic world has used different statistical models to try to classify hotspots. This topic gained momentum in the last quarter of the 20th century, specifically in the late 1970s. Gaussian regression, Multivariate Poisson regression, and negative binomial models were applied in several studies, taking into account the randomness of events in space and time [23][24]. The results of the mentioned models were incorrect because they considered the spatial characteristics of an RTA as constant for a given period, which S. Lakshmi [15] described as occurring "when the assumption that mean and variance must be equal is violated. When the accident data are overdispersed (variance exceeds mean) or under-dispersed (mean exceeds variance), it will lead to erroneous inference with the parameters which determine the crash frequency." The identification of a hotspot requires a much more detailed analysis of the causes and factors —e.g., the severity of the crash, the road conditions, the vehicle conditions, and the weather factors present/existing when the incident occurred [19][25][26]. This issue is developed in detail in chapter 3, "Business Understanding".

Nowadays, researchers are combining GIS and spatial analysis techniques to examine the spatial distribution and spatial dependence of RTA events in a two-dimensional (2D) planar space, utilising global indices such as Global Moran's I (spatial autocorrelation), Getis-Ord G and Getis-Ord G*, which enable investigators to identify statistically significant hot spots, cold spots and spatial outliers. Moreover, local indices such as KDE and Local Anselin Moran can be used to identify the spatial locations of RTA clusters.

Geographic Information Systems (GIS) have established themselves as a critical and powerful tool for spatial data analysis over time due to their multiple features and uses. These systems enable the storage, modelling, analysis, and visualisation of georeferenced data (image 9). Spatial data analysis is the study of spatially referenced phenomena using approaches that attempt to describe and explain why specific events occur in particular locations.

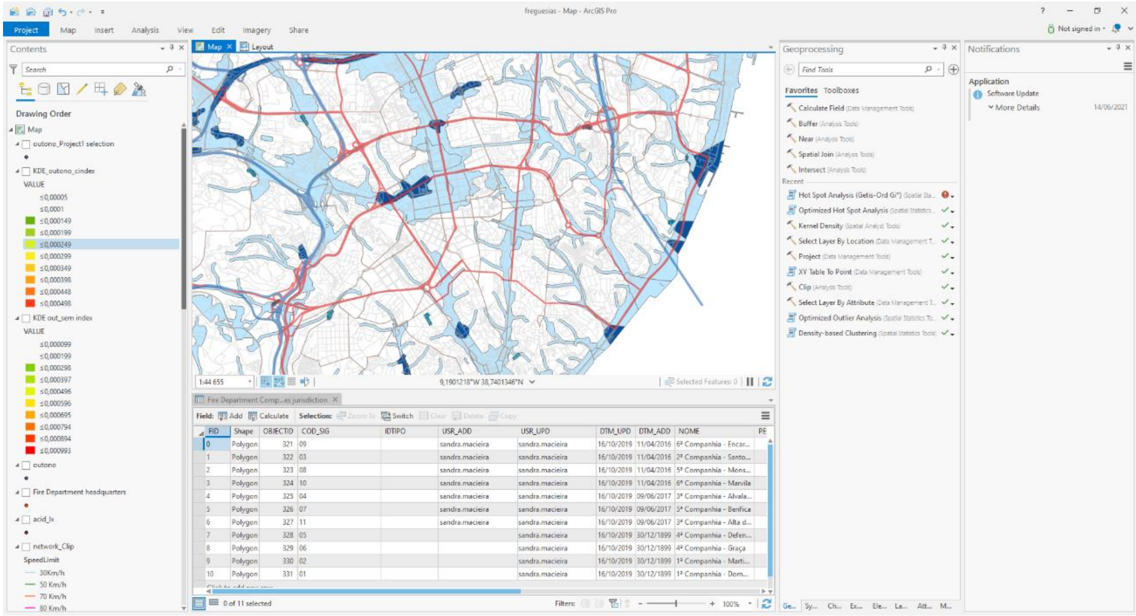

Image 9 - ArcGIS Pro Interface

Numerous studies and scientific discoveries have resulted from the application of GIS technologies, including crime studies [27], Cancer Epidemiology [28]–[31], Zika virus incidence [32], and COVID-19 [33]–[37]. In the case of the current study, GIS has been applied to explore the locations characterized by a high occurrence of road traffic accidents [38]–[42], know as hotspots [43].

These studies could only be carried out properly due to the massification of GPS in cars and mobile devices (smartphones). The access to accurate geolocation data and incident reports appropriately formatted for statistical and spatial analysis provided researchers with previously unattainable reliability levels in survey results [44]. The classic method for RTA geolocation was the mileage pile, which is the combination of the road name and the hectometric demarcation (EU) or Highway location marker (US) to reference the location of the accident. However, this method was highly fallible due to road construction processes, subcontracting, overlapping roads, and changes in nomenclature or classification of the infrastructure [45], which often led to incorrect location identification, introducing systematic and random errors into the analysis.

These conclusions are supported by geographic analysis of existing data and research, such as those shown below. These studies are primarily dependent on the sort of data used and the study's objective.

In the Portuguese city of Vila Real, Inês Franco [21] evaluated three different methods for identifying hazardous road locations (HRL): the NNH clustering algorithm, the KDE algorithm, and the Point Density algorithm.

Luís Ramos [46] recurred to a Visual Analytics approach for the dynamic identification of black spots of road traffic accidents in Portugal

Khanh Giang Le, Pei Liu & Liang-Tay Lin [38] used a combination of KDE and COMAP techniques (comprehensive mitigation assessment process) to identify RTA hotspots. Both analyses determined relatively similar hotspots, but the ranking of some hotspots was quite different due to the integration of the severity index. In this evaluation, Khanh Giang Le proves that in addition to accident frequency, accident severity level is also essential because it helps to highlight the accidents that involve significant damage.

Kaygsz et al. [47] conducted spatiotemporal analysis of traffic accidents on Turkish motorways. Due to data quality, only the South Anatolian Motorway was numerically investigated. Both the KDE and Network KDE techniques were used to identify hotspots, which mainly occurred during the summer. The most common accident types in the study region were rear-ended collisions, stationary object collisions, and run-offs.

Using accident data from Pennsylvania between 1996 and 2000, Aguero-Valverde and Paul P. Jovanis [48] approximated the annual frequency of accidents in the United States at the county level according to hierarchical models. Sociodemographic data, weather conditions, transportation infrastructure, and journey time were among the elements considered. They discovered that parameters such as low poverty, age range (specifically, 0 to 24 and over 64 years of age), road length and traffic flow density contribute to an increased likelihood of road accidents.

Ayodeji E. Iyanda [49] used a smoothing method to predict two years of road accident events based on previous historic RTA in all 36 states and the federal territory of Nigeria. This study employed Moran's I, a spatial autocorrelation statistic, to determine the degree of randomness of RTA severity among the 36 states. It also used Anselin's local indicator of spatial association (LISA), which provides a statistic for each location with an assessment of significance and establishes a proportional relationship between the sum of the local statistics and a corresponding global statistic. According to the findings, the northern portion of Nigeria has the highest RTA severity compared to the southern states. Local research suggests that these clusters in the northern part of Nigeria did not happen by chance; the geographic patterns are considerably clustered.

Kernel density estimation is a widely used and well-established spatial technique for estimating crash intensity and identifying hotspots [19],[50],[51],[52],[53]. This method converts a collection of point events into a continuous surface that indicates their density and enables researchers to examine the variance in the mean value of the event under study over the study area — i.e., how events are distributed in space.

Saffet Erdogan et al. [51] investigated accident hotspots and identified unsafe zones on highways in the Turkish city of Afyon. KDE and repeatability analysis were applied to analyse the data. Both assessments identified nearly identical locations as hot spots. The majority of them were around junction points, crossroads and access routes to villages and towns. They also discovered that the number of accidents increases in the summer and winter, particularly in August and December. Weekends had higher frequencies as well. Furthermore, they found that the majority of deadly accidents occurred after midnight.

Tessa Anderson [19] recommends utilizing GIS and KDE to examine regional patterns, such as injury crash hotspots, and then merging the data with clustering techniques.

Prasannakumar et al. [54] studied 1468 accidents that were reported to the police in Thiruvananthapuram, India, in 2008. Unlike many other work reports, these provided each event's coordinates. To conduct the investigation, researchers compared the spatial autocorrelation methods of Moran's I, Getis-Ord Gi** statistical index and KDE (Image 10).

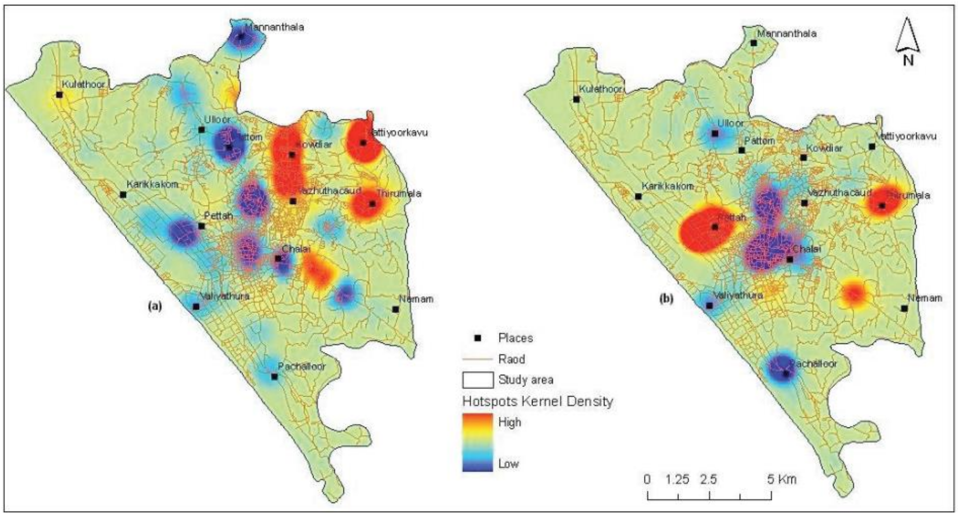

Image 10 - Spatial characteristics of accident hotspots (a). monsoon and b). non-monsoon period) [54]

When the Getis-Ord Gi* and KDEs were applied, the Moran I revealed that the data tend to cluster in specific time zones [54]. Furthermore, using methods such as the Moran I, hotspot analysis and KDE, Aghajani et al. [55] discovered a significant relationship between accident density, topography and rainfall maps — i.e., more mountainous regions and locations with higher rainfall have a higher accident density. Of the studies described above, few suggest accident prevention strategies. Instead, they focus only on the process of discovering hotspots and providing a descriptive analysis.

In this dissertation, Global Moran's I was used to evaluate the data for spatial autocorrelation. Then we used the Kernel Density Estimator to identify accident hotspots because it is a widely used method that is simple to apply and implement and produces visually appealing results [56] similar to those obtained by the authors mentioned above. Moreover, when combined with hot spot analysis (Getis-Ord Gi*), it allows for identifying problematic hotspots in Lisbon's road network.

2.1.1 Global Moran´s I

Global Moran's I is a widely used spatial autocorrelation indicator. The initial measure on spatial autocorrelation for Road traffic accidents in this study was Global Moran's I. We used Global Moran's I to identify whether the spatial output pattern is clustered, scattered, or random and concentration levels [54]. The null hypothesis states that the feature values are distributed randomly throughout the research region. Moran's Index values vary from -1 to 1, with a value near +1.0 indicating clustering, a value near -1.0 indicating dispersion [57], and a value of "0" indicating perfect geographic randomness [58].

The p-value is a probability that indicates whether the observed values were generated by random processes or by other geographical activities. When the essential z-score values are 1.655 at a 90% confidence level and the p-value is more than 0.10 (> 0.10), the null hypothesis can be accepted. The null hypothesis can be rejected when the crucial z-score values are more significant than 1.65 at a 90% confidence level, and the p-value is less than 0.10 (0.10). This is because significant clustering exists beyond this region [59]. In the tails of the normal distribution, very high (+2.58) or very low (-2.58) z-scores are discovered, which are associated with very small p-values (0.01) and represent clustering or dispersion at a 99 percent confidence level (Image 11). Substantial hotspots (shades of red in Image 3.2) are indicated by high z-scores linked with small p-values, whereas significant cold spots (shades of blue in Image 11) are indicated by low z-scores coupled with small p-values [59].

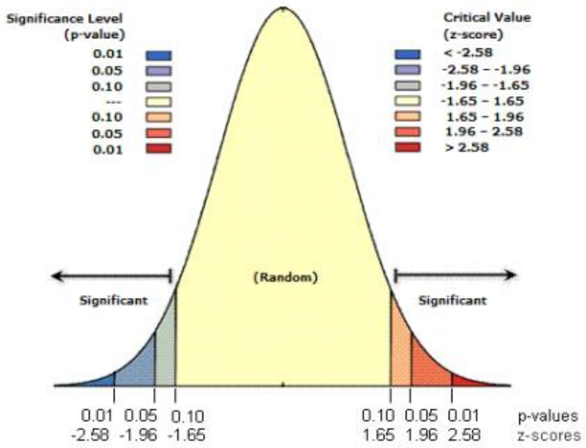

Image 11 - Normal Distribution of Global Moran's I [59]

The Spatial Autocorrelation tool was performed for each iteration with distance criteria defined by the ArcGIS Incremental Spatial Autocorrelation. The direct distance between two points was utilized to determine the locational proximity of data events, whereas the inverse distance weighting approach was employed to determine the locational proximity of neighboring points. Because each data point is analyzed in terms of its neighbors, determined by a distance threshold, it is important to pick a distance threshold that maximizes spatial autocorrelation[54]. The Global Moran I-Index is calculated using the following equation:

(1)

$$I = \frac{N \sum_{i=l}^{n} \sum_{j=l}^{n} W_{ij}(x_i - \bar{x})(X_j - \bar{x})}{\left(\sum_{i=l}^{n} \sum_{j=l}^{n} w_{ij}\right) \sum_{i=l}^{n}(x_i - \bar{x})^2}$$

Equation 1 - Global Moran's index equation

Where I is the Moran's I statistics, N is the number of observations, x_i is the value of variable at location i, x_j is the value of a variable at location j, and wij is the weight that determines the relationship between i and j.

2.1.2 Kernel density estimator for road accident analysis

One of the most popular techniques for identifying hotspots is spatial point pattern analysis, and there are several methods for doing so [19]. These can be divided into two categories [60]: first-order methods, which measure the variation of the process's mean value, such as quadrant counting or KDE, and second-order methods, which examine the spatial dependence of points and use functions to explain this level of spatial dependence.

As previously stated, KDE is the most widely used non-parametric statistical method for smoothing data because it is simple to understand and implement. It uses a kernel, K(x), centred in the estimation location (Xi), to calculate the density value of a set of point events (n) in a given area, interpolating them and generating a continuous, more or less smooth surface. The kernel is a weighted function that can be Gaussian, quadratic or conic in nature. In other words, the function counts the number of points within a region of influence, weighting them according to their distance from the location of interest; the bandwidth h determines the region of influence.

Equation 1 expresses the Kernel density function $\hat{f}$

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right) \quad (2)$$

Equation 2 - Kernel function

The chosen values for the kernel and the bandwidth are two parameters that can influence the final result. However, according to several authors, including Wand and Jones [61], the final result is relatively unaffected by the type of kernel used. The bandwidth is the most crucial parameter and has the most significant impact on the final result. It determines the smoothness of the surface produced; bands that are too small produce non-smooth results, while bands that are too wide produce the opposite effect. Because it varies with the study area and data and sometimes requires a subjective judgment when applied, there is no universal formula for determining this parameter.

This dissertation applied the KDE method in ArcGIS PRO 2.8 [59] with the Spatial Analyst extension. After testing several bandwidths, a bandwidth of 500 meters was chosen. The hotspots covered a large portion of the study area with larger bands, making it impossible to pinpoint hotspot concentration locations on specific roads. In comparison, only a few hotspots were identified with smaller bands, resulting in an uneven surface that required multiple attempts to achieve [62].

2.1.3 Hotspot Analysis (Getis-Ord Gi*)

The Getis-Ord Gi* are a group of statistics with several properties that make them appealing for quantifying dependence in a geographically distributed variable, particularly when combined with Anseli Moran's I [63].

The Getis-Ord Gi* analyze evidence of spatial patterns, according to is creators, Getis and Ord [64], "*deepen the knowledge of the process that give rise to spatial dependence and enhance detection of local pockets of dependence that may not be detected when using the global statistic*".

The Getis-Ord Gi* statistical computation is used in this dissertation to determine where the spatial clusters (hotspots) of road accidents with victims occur in the study area. In addition to identifying these clusters, the algorithm uses the z-score and p-value values to determine where the highest/lowest values for spatial clusters occur.

The Getis-Ord Gi* spatial cluster analysis analyses the characteristics of each occurrence of road accidents with victims, accessing each record in a neighborhood context and looking for the distance that ensures that each record (accident) has at least one neighbor. A high-value occurrence does not necessarily imply a hot zone.

To be statistically significant, it must have a high value and be surrounded by additional high-value occurrences.

The sum of a given characteristic, [incidence], with its neighbors is proportionally compared to the sum of all attributes. When the local sum differs significantly from what is predicted, and this difference is not a random outcome, a z-score with statistical significance is produced.

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{x}_{\Sigma_j^n=1} w_{i,j}}{S\sqrt{\frac{\left[n\Sigma_j^n=1 w_{i,j}^2 - \left(\Sigma_j^n=1 w_{i,j}\right)^2\right]}{n-1}}}, \quad (3)$$

$$\bar{x} = \frac{\sum_{j=1}^{n} XX_j}{n}$$

$$S = \sqrt{\frac{\Sigma_j^n = 1x_j^2}{n} - (\bar{x})^2}$$

*Equation 3 - Getis-Ord Gi**

where: Gi* is the spatial autocorrelation statistic of an event i over n events. The term, xj is the attribute value for feature j, wi,j is the spatial weight between feature i and j, n is equal to the total number of features.

Once calculated, t the determined value is a z-score, and the greater the score, the stronger the clustering of high values (hot spot). When the z-score number is negative (low-value), more cold spot values are clustered.

Page left intentionally blank

CHAPTER 3

Business Understanding

As mentioned in the introductory chapter, the first phase of the CRISP-DM focuses on identifying and assimilating the problem that needs to be solved. In order to do so, it is necessary to understand the whole universe related to the phenomenon under study — in this case, Road Traffic Accidents (RTAs). Consequently, in this chapter, the physical characteristics of Lisbon, the road network and the regulation of emergency management in the municipality of Lisbon are described.

This explanatory approach to the city of Lisbon's characteristics and its emergency system is is the author's contribution to knowledge based on 20 years of accumulated professional experience in the National Authority for Emergency and Civil Protection Service. It creates the basis for comprehending the issues and identifying the most significant factors in the municipality that contribute to the RTAs that will be used later in the "Data" chapter. This survey is of utmost importance since it might influence or even modify the final results.

Before addressing the characteristics of Lisbon, it is crucial to understand the phenomenon of Road Traffic Accidents and their causes.

3.1 The Road Traffic Accident Definition

The Portuguese National Road Safety Authority (ANSR) defines a road accident as "an occurrence on or from the public highway involving at least one vehicle, known to the supervising entities (GNR and PSP), and resulting in victims and/or material damage" [10].

As a result, an accident is defined as an unforeseeable or unplanned event that causes moral, property or personal harm. Depending on the type of damage, the accident may also be classified as an "accident with victims", in which at least one person is injured; a "fatal accident", in which at least one person is killed (on the spot or within 30 days); and a "seriously injured accident", in which at least one person is seriously injured (individual bodily injury requiring more than 24 hours of hospitalization, but where no death occurs).

3.1.1 The Causes of a Road Traffic Accident

A traffic accident rarely occurs due to a single common cause (factor) [20]. Pedro Oliveira, in a study conducted for the Associação de Cidadãos Auto-Mobilizados (ACA-M) [65], stated that " Contrary to what is being advocated, speed is not a cause of accidents, but rather an accident-causing factor, like rain, the bends in a road, alcohol, or disrespect for the Highway Code. Although road safety campaigns traditionally focus on the slogan "speed kills", the truth is that speed promotes conceivable mistakes on the part of the driver or on the other dynamic elements that make up the road infrastructure, such as third party vehicles, pedestrians, and various objects. Despite initial difficulties due to a lack of credible statistical data, which inevitably led to methodological errors and bias in the results, the causal factors that contribute to a traffic accident have been studied by several academics over the past few decades.

The advent of the practice of security agents registering RTAs allowed researchers to detect the relationship between one or more factors that implicitly or explicitly may contribute to an RTA to a greater or lesser extent [21]. In this way, accidents started to be understood as multicausal, meaning that they tend to occur when the conjunction of one or more variables are present that can act independently or in conjunction with one another (driver, road and road environment, and vehicle) and can be evident (traffic, weather conditions, infrastructure properties and vehicle properties) or latent (human perception, decision making, psychological and cognitive state) [26][66][67][68].

Thus, an RTA is a series of events that result in a system failure or breakdown, and it is rarely the result of a single cause. An RTA is an unusual or out-of-the-ordinary occurrence for the average road user. If appropriate and timely evasive actions disrupt the causal chain, the incident can be avoided.

3.1.2 Factors Contributing to a Road Traffic Accident

Driving entails operating a vehicle in a highly dynamic road environment and interacting closely with other road users. As previously stated, RTAs are typically caused by the convergence of factors closely related to the three elements of the road system: the road user, the vehicle and the road environment. For the purposes of this dissertation, these factors will be referred to as human, vehicular, and environmental factors, in that order.

Human factors are cited in the Highway Safety Manual (HSM), published by the American Association of State Highway Transportation Officials (AASHTO), as probable causes in more than 90% of RTAs.

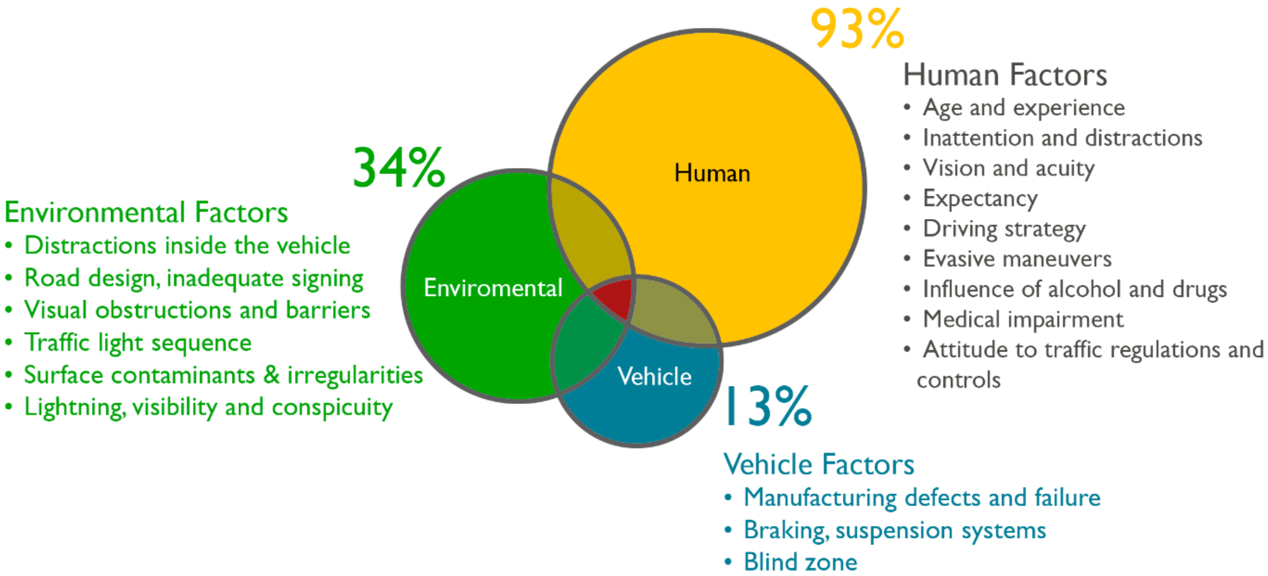

Image 12 - Contributing factors to Vehicle Crashes

3.1.3 Human Factors: Personality and Behavior

Human factors, including vehicle drivers and pedestrians, are often cited as the primary cause of road accidents. Many reasons can be identified for attributing RTA blame to humans.

The term "man/driver" includes a variety of variables that may contribute to the incidence of an accident, including the following:

- Perception and information processing;
- Driver perception;
- Time required for response;
- Individual distinctions;
- Mobile Phone Use;
- Exhaustion;
- Alcohol and substance abuse;
- Neuropsychological, medical, and psychiatric disorders.
- Age and experience

- Inattention and distractions
- Driving strategy
- Attitude to traffic regulations and controls

According to scientific research, "The normal inclination is for the driver to match his driving speed to the conditions dictated by the route's geometry. Drivers adjust their speed to the road's geometry when they are unaware of such risk-prone features in specific locations"[69]. In this manner, drivers retain the authority to choose the actions they exhibit on public roads, even though they are mindful of legal requirements.

3.1.4 Vehicular Factors

All vehicle-related characteristics, deficiencies and defects, as well as related motorist interactions and responses that cause an RTA or worsen the severity of a vehicle impact, are considered vehicular factors.

The term "vehicular factors" includes a variety of variables that may contribute to the incidence of an accident, including the following:

- Manufacturing defects and failure
- Braking, suspension systems
- Mechanical Defects
- Loss of Vehicle Control
- Vehicle blind zones

3.1.5 Environmental Factors

Environmental factors include elements found both inside and outside the vehicle and the roadway and its surroundings. Distractions in the vehicle, such as navigational aids, internal radio, air conditioning, window and mirror adjustments and conversations or arguments with passengers and small children can all affect a driver's response. According to studies [70], loud music can impair a driver's reaction time and performance in the same way that noise does. It is also common to drive while distracted by a cell phone. The roadway environment encompasses both the roadway and the surrounding environment, including weather and darkness.

The term "environmental factors" includes a variety of variables that may contribute to the incidence of an accident, including the following:

- Distractions inside the vehicle
- Road design, inadequate signing
- Visual obstructions and barriers
- Traffic light sequence
- Surface contaminants & irregularities
- Lightning, visibility, and conspicuity

Dynamic conditions that change quickly and suddenly are common in road traffic systems. Human factors, vehicular factors and environmental factors all play a role in traffic incidents. In a traffic crash, we assess the road user's skills and shortcomings, experiences with other road users, cars and the environment. Human vision, awareness, judgment and efficiency will all remain essential aspects of driving before self-driving cars become the standard on our roads.

3.2 Emergency Management and Safety in the City of Lisbon

Moored at the mouth of the Tagus, the city of Lisbon is one of the oldest cities in the Western world. Its geographical position in a Euro-Atlantic context has made Lisbon a space of geostrategic relevance, defining the contours, from very early on, for the city to become the capital of Portugal. As a city, Lisbon is the Tagus, the Atlantic, the climate, the relief and its historical past. Lisbon is also its industry, commerce, demography, urbanism, technology, knowledge, culture, transport and communication routes. Thus, the city is a complete combination of natural and anthropic factors [71], which, together, potentiate a constellation of dangers associated with existing vulnerabilities, transforming them into natural focuses of risk [71].

Our inability to act directly on all potentially dangerous processes (some of which are very unpredictable and, for this very reason, more dangerous) encourages us to reduce vulnerability. The most recent conceptual model concerning risk identified three determining factors for the reduction of vulnerability: i) exposure of the population, ii) awareness of the population, and iii) the response of society [72]. Given these factors and the scope of this dissertation, a city is also safety and, consequently, prevention, protection and rescue. The relief component of the response capacity depends on a list of exceedingly complex factors. The technical and technological aspects, which are substantially different from season to season, necessarily intersect with the nature and dynamics of a city, which are also volatile from season to season.

In Portugal, the government is responsible for the protection and assistance of its citizens. The Portuguese legal system regulates this obligation through Law no. 27/2006, later rectified by the Organic Law 1/2011 and Law 80/2015, which ratifies the Civil Protection Basic Law (LBPC). According to article 1 no. 1 of the LBPC, "civil protection" is defined as "the activity undertaken by the state, autonomous regions, and local governments, citizens, and all public and private entities with the goal of preventing collective risks associated with a severe accident or catastrophe situations, mitigating their effects, and protecting and assisting people and property in danger."

According to Law 80/2015, which modifies and upgrades Law no. 27/2006, Portugal's civil protection system comprises four tiers: national, regional, district and municipal. Due to the objective and purposes of this dissertation, we will only discuss the last one (the municipal level). The municipal level's importance in the current structure of the civil protection system is reflected in the principle of subsidiarity, as stated in Article 5 paragraph d) of Law no. 27/2006, which establishes that the higher level of the civil protection subsystem shall intervene only if and to the extent that the civil protection objectives cannot be met by the civil protection subsystem immediately below, given the magnitude and seriousness of the occurrences. According to Manuel João Ribeiro [28], Municipal Civil Protection Services (SMPC) serve as the foundation and support for the national civil protection system. He argues that the responsibility for the initial response in the event of significant accidents or disasters belongs to the municipal level, thereby adhering to the subsidiarity principle. In this context, SMPCs play a critical role in the current national system, as their existence and effectiveness are critical for the conduct of all civil protection activities, regardless of whether their level of operation is municipal, district, regional or national. Concerning the Lisbon SMPC’s mission, it is clearly stated in article 10 of Law nº 65/2007 that these structures have the authority to ensure the proper operation of all municipal civil protection organisms, as well as to centralise, treat and disseminate all information received regarding municipal civil protection. The competencies are defined in three dimensions: planning and operations, prevention and safety, and public information.

The Lisbon Municipal Civil Protection Service was established in 1985 and recently converted into a department by order no. 5347/2015, published in the Diario da Republica (Official Gazette), second series, No. 98 - May 21, 2015. It is one of the Municipal Services of the Municipality of Lisbon, alongside the Fire Brigade Regiment and the Municipal Police, that ensures the safety and security of people and property in the city of Lisbon. In terms of dependency, Article 9 of Law 65/2007 of November 12 clearly states that the mayor of the city is the political authority and directs the SMPC. Additionally, this legal diploma defines the figure of the Municipal Operational Commander (COM), which previously lacked a counterpart in the Civil Protection system, as the operational command structure of the National Authority for Emergency and Civil Protection (ANEPC) contemplated only the national operational commander and district operational commanders. Consequently, the COM, which is hierarchically and functionally subordinate to the mayor, was created. In the case of the Lisbon municipality and others with a professional Fire Brigade Regiment (RSB), the Commander in Chief is nominated to the position by the aggregation.

3.2.1 The Lisbon Fire Brigade Regiment (RSB)

According to the Civil Protection Fundamental Law, firefighters are the first responders in emergencies and other incidents posing a danger to life, property or the environment. Technically and operationally, this duty falls to the Fire Brigade Regiment (RSB) in Lisbon. Therefore, it is important to emphasise the RSB's mission as an organization under examination. The RSB's primary goal is to ensure the protection of people and property in the city of Lisbon through a continuous program of prevention, preparation and rescue in the event of a major accident or disaster. Moreover, the RSB relies on interoperability with Lisbon's Municipal Civil Protection Services, which engages in certain measures and monitors their resolution to mitigate consequences, to carry out its work.

3.2.2 Operations

As previously mentioned, all operational activities are focused on protecting and assisting citizens within the context of the RSB’s mission. If necessary, it’s possible that its mission can be extended to include other areas in order to support other emergency agencies (or by the request of the National Authority for Emergency and Civil Protection).

The regiment's organizational structure consists of approximately 867 professionals (the staff establishment plan can accommodate up to 1112 operational personnel) divided into 11 intervention zones based on the geographical area of operation, the population density, and the ability of the fire brigade vehicles to arrive at the theatre of operations (incident area) in less than six minutes. This rationale is due to the need for firefighters to arrive at the site of a fire before the occurrence of a flashover, which is the form of fire that occurs rapidly and is characterized by an increase in the rate of fire growth and development. This is typically a transition from a fire at a single source to a fire that consumes all other combustible materials in the area without coming into direct contact with the flames. During and after a flashover, the primary form of heat transfer is radiation. The regiment is divided into six companies, each with two firehouses (HQ and Station), for a total of 11 firehouses (3 battalions, six operational companies, one special intervention company, one company for command and services) and a fire academy.

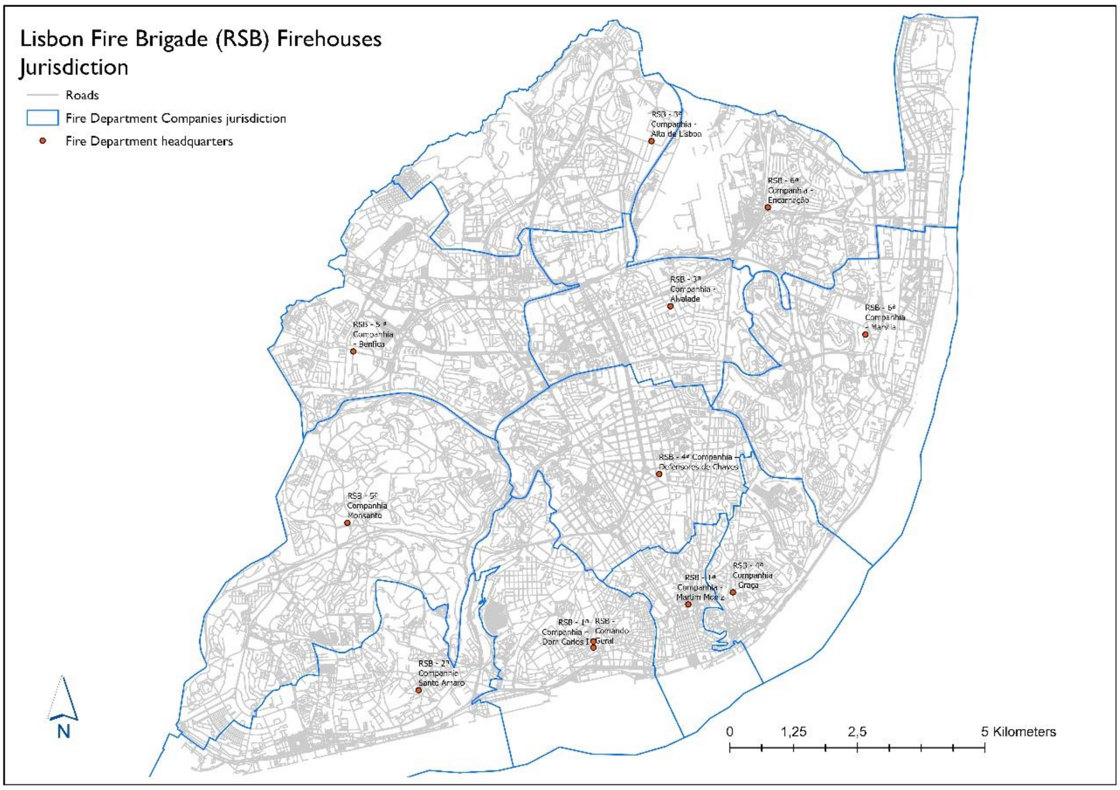

Image 13 - Lisbon Fire Brigade Firehouses Jurisdiction

3.2.3 Operational Response to Road Traffic Accidents

Each of the companies mentioned above is responsible for ensuring the relief of the area assigned to it, thus enabling greater effectiveness in combating occurrences and greater proximity to citizens. All of these firehouses are duly equipped with various combat means. In the field of accidents, the vehicles designed for specific operations include the Rescue and Tactical Assistance Vehicles (VSAT) and the

Urban Fire Fighting Vehicle (VUCI). The two existing VSAT are strategically positioned (on each extreme of the city) to arrive as soon as possible to an occurrence and hold the specific equipment that supports (the most complicated) extrication operations. The VUCI is the primary vehicle of the RSB for intervention in traffic accidents. In addition to all of the equipment necessary to extinguish an ignition, this vehicle is equipped with the proper material for extrication operations. The crew of this vehicle includes 6 members, one of whom is a certified Ambulance Rescue Team Leader (medical emergency technician) (TAS) and Automatic External Defibrillation (AED), who can perform the first assessment of the victims and, if necessary, intervention until the arrival of vehicles (and specialized teams) from the National Institute of Medical Emergency (INEM).

3.3 The City of Lisbon

Lisbon is a city vulnerable to several risk situations due to "its geographical and social, economic and political characteristics, for its importance as a capital" [73]. Thus, in order to ensure a proper theoretical framework within the scope of this work, it is important to outline Lisbon's relevant characteristics. This will allow us to establish if any relationships exist between these characteristics and the road traffic accidents that happen daily.

The Municipal Emergency Plan for the city of Lisbon is a benchmark of excellence regarding civil protection activity that, ultimately, all civil protection agents consult for information on their specific missions, which supports better coordination [74]. This strategic document identifies a variety of risk situations, including the one that is the focus of this dissertation: serious traffic accidents.

Thus, and always taking into consideration the objectives we propose to achieve with this dissertation, we have chosen to highlight the following as they impact RTAs:

- Altimetry and Hypsometry;
- Climate;
- Road infrastructure

The Municipality of Lisbon is central to the Metropolitan Area of Lisbon (AML) and is the capital of Portugal. Geographically, it is located on the right bank of the Tagus River, near its mouth, at 38º42' 30,5" North latitude and 9º Greenwich longitude 0º. It is bounded on the south and east by the Tagus River. To the west, it borders the Municipality of Oeiras, and to the north, the municipalities of Loures, Amadora, and Odivelas.

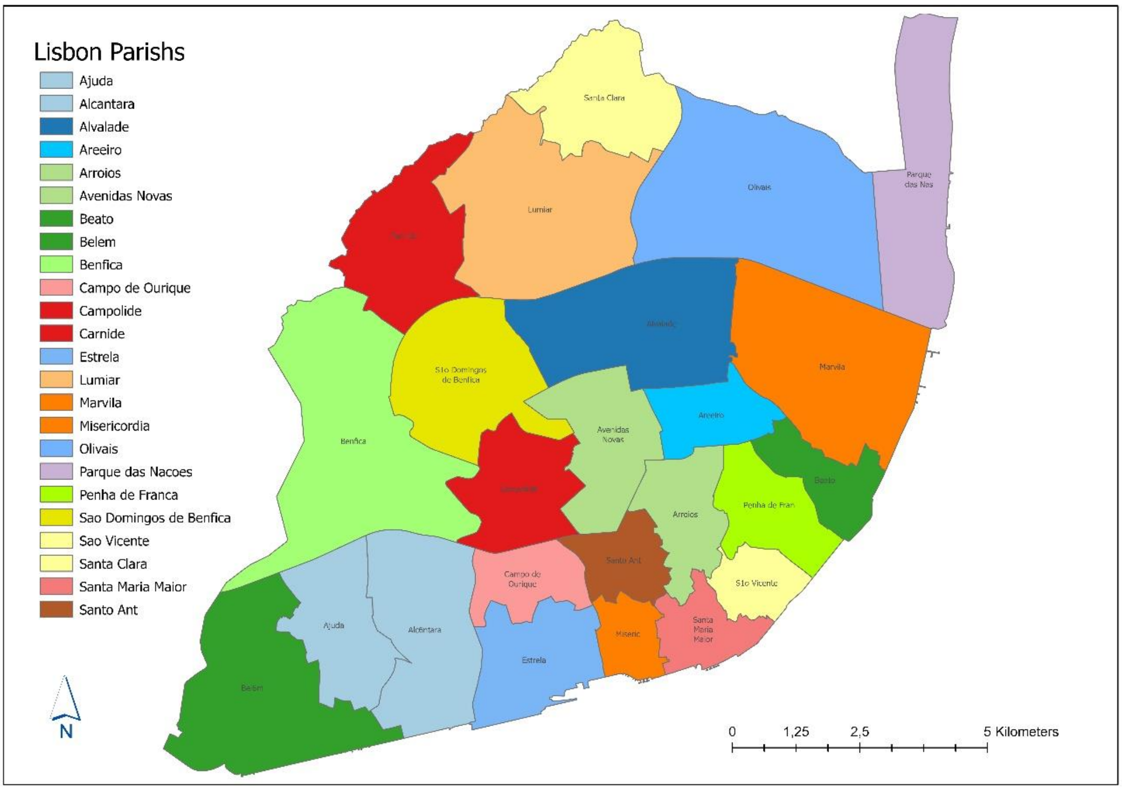

Image 14 - Lisbon Civil Parishes

The Municipality of Lisbon has a total area of 84 Km2 and is currently divided (image 14) into 24 administrative divisions (civil parishes).

As regards the Territorial Units for Statistics (NUTS), the Municipality of Lisbon is part of the Lisbon and Tagus Valley Region (NUTS II) and the sub-region of Greater Lisbon [75].

The territory of the city of Lisbon is fundamentally characterised by areas with an altitude below 100m; the major exception is the Serra de Monsanto with an altitude of 230m (at its highest point). The difference between altimetric quotas in Lisbon occurs between sea level and the quota of 230m, as is visible on the altimetric map (image 15). The areas that are part of the entire riverside zone and lower valleys have heights, in general, lower than 30m, while the slopes have heights of approximately 70m. The central zone of the Serra de Monsanto reaches heights above 150m; however, it is mostly characterised by areas in the order of 70 to 100m [33]. In the northern area of the county, we highlight the area of Ameixoeira and Aeroporto, where, as in the Serra de Monsanto, sporadically, it could reaches heights of 150 to 220m. It is also worth mentioning Telheiras and Carnide because they are mostly composed of areas between 100 and 150m, returning to register identical values only in the area of Campolide [76].

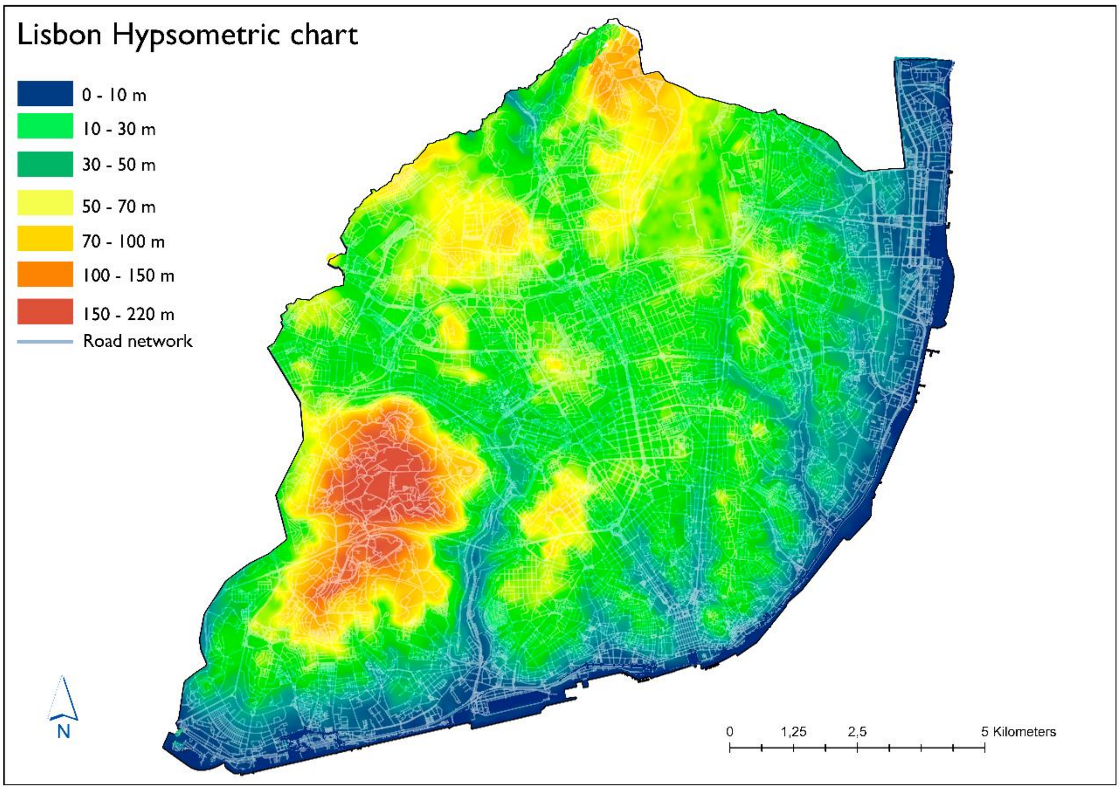

Image 15 - Lisbon Hipsometric Chart

The altimetric characteristics of the city of Lisbon are directly related to its particular circumstances, the Tagus Valley, its riverside margin, the Monsanto mountain range, and the plateau area. The first two deserve special mention, as they have lower elevations, which favors the circulation and accumulation of water [73].

The slopes are steeper in the areas next to the main water lines of the city, in particular Vale de Alcântara and Vale de Chelas, to the south and east, respectively. Of the valleys mentioned, Vale de Alcântara is the deepest, with the largest area of steep slopes (between 30% and 40%), which also includes the southeast slope of the Serra de Monsanto. The hillside areas usually have slopes of 25%.

The main valleys of Lisbon flow into the Tagus estuary. They are in Alcântara, Chelas, Avenida da Liberdade, Arroios, Marvila and Olivais. To the east, along the bank, valleys define hills with relatively steep and varied slopes. The Serra de Monsanto, as Lisbon's volcanic complex, stands out to the west and dominates the landscape. To the north, extensive plateau areas, where most of the terrain is steep, predominate [76]. The possibility of slope movements is related to several factors: the geological nature of the formations, geomorphology, and the presence or circulation of water. [74]

For emergency planning purposes, the following elements are considered to be possible climatic triggers: concentrated precipitation, strong winds, cold and heat waves [73]. In general, the city of Lisbon has a Mediterranean climate with a hot and dry summer. The average temperature is 16ºC. The lowest figures are registered in December, January and February (average values of 10ºC), and the highest occur July through September (average values between 20ºC and 25ºC). Precipitation occurs mainly between October and April, with annual average annual precipitation values between 650mm and 760 mm. The highest figures happen from November to February (with 160mm), and the lowest are in July and August (with values between 3 and 7 mm) [73].

Precipitation is of special importance because, according to a study of the European Environment Agency, Lisbon is the "fourth most waterproof city in Europe", and it favours the occurrence of rapid floods. The most susceptible areas "are the coastal areas, with valleys, low elevations, slight slopes and/or wetlands that in the past have been covered by main watercourses"[73], as can be seen in the flood risk map for the city of Lisbon (Image 16). Flooding "can occur when there is intense, sudden, or prolonged precipitation. This situation is aggravated when it coincides with the peak of the high tide and, even more, with the influence of the Storm surge phenomenon"[73].

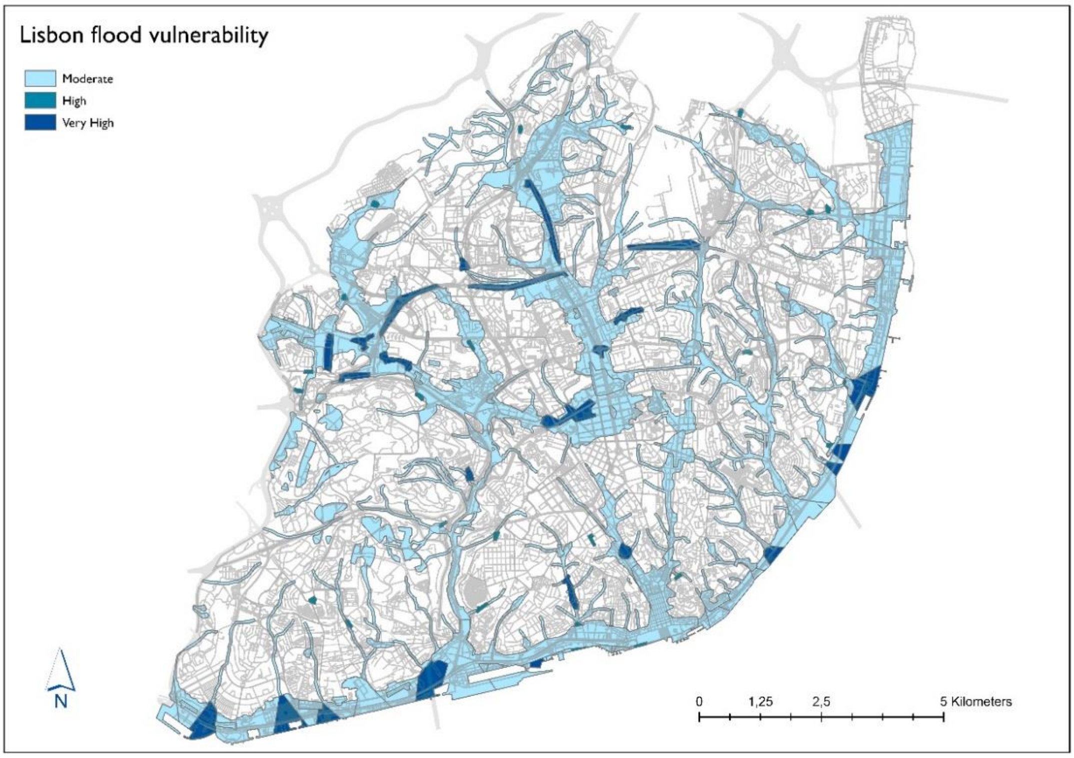

Image 16 - Lisbon flood vulnerability map

Beyond this generic characterization, Lisbon "is sometimes influenced by unpredictable weather, which causes exceptional conditions (...)". Among them, it is important to highlight the following due to their impact: "(...) Extreme figures of minimum (...) or maximum temperature (...) values above 40º C (...) High values of rainfall in short periods (...) Strong and very strong wind and gusts with very high speeds (...)" [73] These are all contributing factors to the occurrence of road accidents that will be addressed later.

Taking the road network and its characteristics into consideration is fundamental. After all, the city of Lisbon, with its high road traffic flows, especially during rush hours, registers a high number of traffic accidents. They are largely the consequence of roads that allow high speeds, even though they are poorly maintained or designed, in addition to unfavourable weather conditions [73].

In the national context, Lisbon is an urban center where daily population flows in and out of the 'region' are more intense. This flow is known as commuter traffic (traffic caused by commuters to and from work). According to the 2011 Census [77], the car is the most used means of transportation between home and work or home and school. In fact, "about 62% of the population, who commutes daily to carry out their activities, choose to use the car, as a driver or passenger"[77].

In a survey conducted by the National Institute of Statistics (2017) on the mobility and functionality of the territory of the Lisbon Metropolitan Area (LMA), we can see that in the first decades of the twenty-first century, there was an apparent increase in the metropolitan population, currently with 2.8 million residents, and an emptying of the urban centers, which led to a dispersion of residents throughout the territory of the LMA.

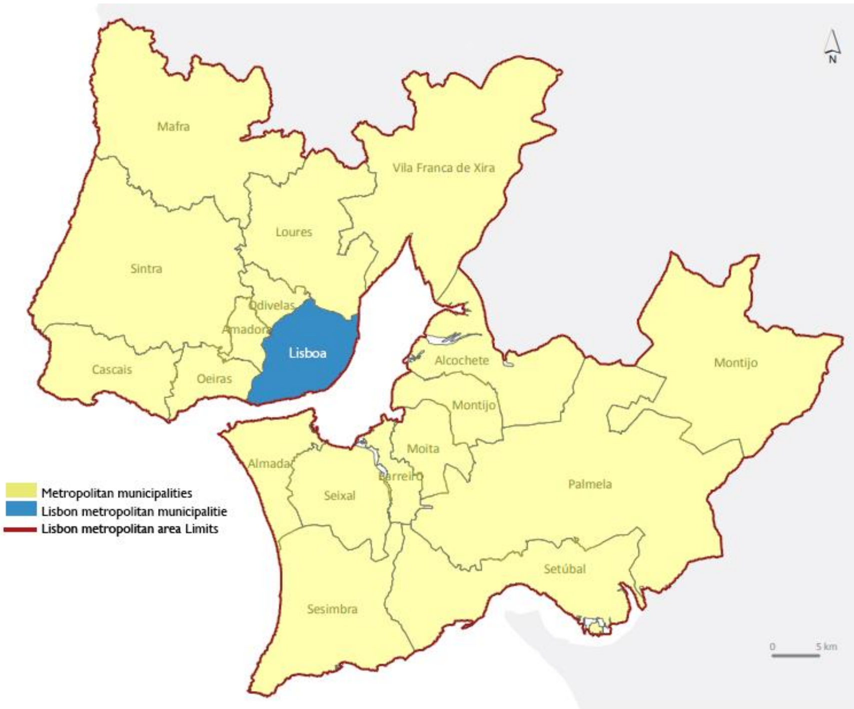

Image 17 - Lisbon Metropolitan Area

This phenomenon has led to a significant increase in commuter traffic, which is corroborated by the values of the average annual daily traffic (AADT)[7] from the Institute for Mobility and Transport (IMT), showing the commuter traffic flows in the Lisbon Metropolitan Area (Image 18) and demonstrating the strong relationship between the municipality of Lisbon and the other 17 metropolitan municipalities that make up the Greater Lisbon area (Alcochete, Almada, Amadora, Barreiro, Cascais, Loures, Mafra, Moita, Montijo, Odivelas, Oeiras, Palmela, Seixal, Sesimbra, Setúbal, Sintra and Vila Franca de Xira).

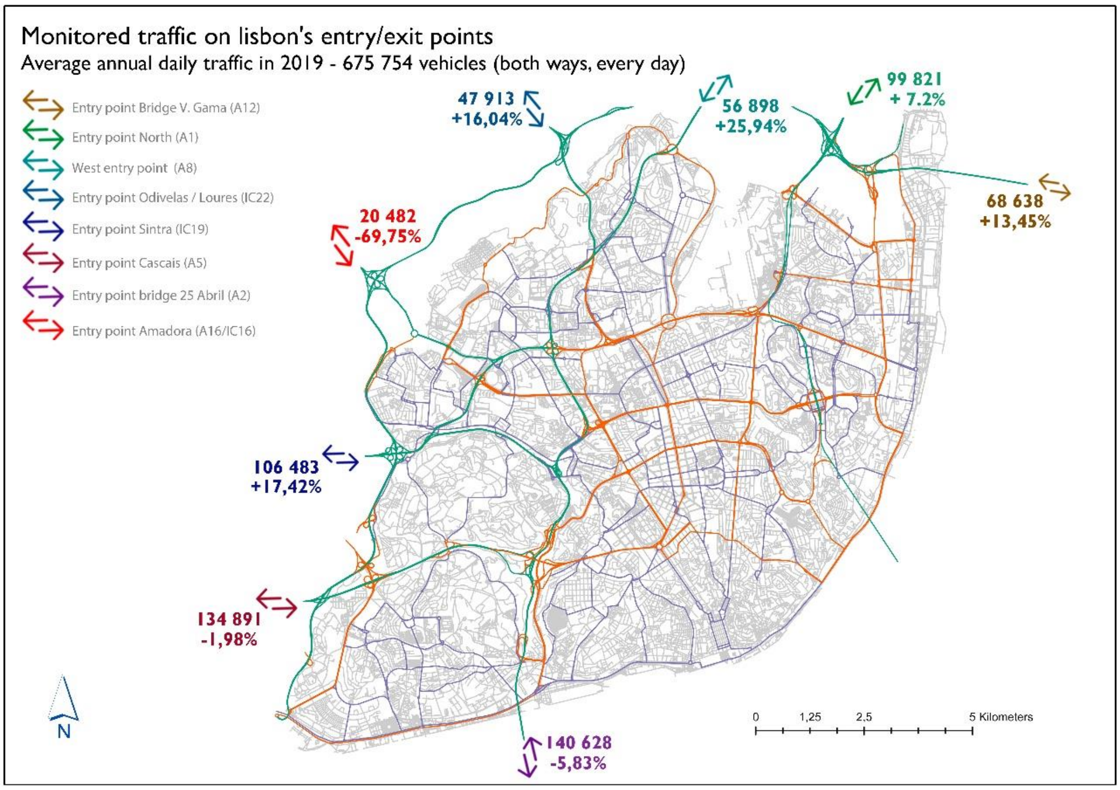

Image 18 - Commuter traffic flows on Lisbon entry / exit points

Every working day (weekdays), an average of 301,000 vehicles (table 1) enter the main accesses of the City of Lisbon, in addition to the 367,000 vehicles that reside in the municipality (Table 2).

[7] AADT - the entire volume of traffic that a road receives in both directions over the course of a year, divided by the number of days in the year.

Table 2 - Average Annual Daily Traffic of the entry points of Lisbon (both ways)

Entry Point	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	%
A1 - North	93 099	91 564	89 784	90 007	93 315	95 294	97 159	97 168	99 821	93 099	7,22%
A12 - Vasco da Gama Bridge	60 341	54 967	53 443	54 306	55 628	59 308	62 304	65 318	68 638	60 341	13,75%
A16/IC16 - Amadora	34 768	14 564	12 214	11 510	12 223	13 863	15 697	17 351	19 234	20 482	- 69,75%
A2 – 25 Abril Bridge	143 406	138 937	136 263	135 514	138 475	140 730	143 481	143 109	140 628	143 406	- 5,83%
A5 - Cascais	129 532	121 726	125 947	128 746	131 521	116 291	133 683	134 329	134 891	129 532	- 1,98%
A8 - West	45 178	44 002	48 094	48 683	50 891	52 571	54 859	55 962	56 898	45 178	25,94%
IC19 - Sintra	90 683	87 915	88 745	88 345	105 055	73 562	64 752	84 719	106 483	90 683	17,42%
IC22 – Odivelas / Loures	41 291	36 790	45 267	41 042	39 761	31 533	46 191	47 175	47 913	41 291	16,04%
Total	**649 099**	**618 094**	**533 148**	**599 053**	**469 479**	**628 509**	**584 986**	**619 780**	**647 014**	**675 754**	**4,11%**

Table 3 - Number of license plates with valid insurance registered in the Lisbon municipality - Portuguese Insurance and Pension Funds Supervisory Authority 2021

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Moped	4 660	4 449	4 344	4279	4 511	4 400	4 258	3314	3 587	3 486
Light	313 277	305 019	301 203	301467	291 709	323 741	333 019	323875	329 059	331 956
Motorbikes	13 008	13 111	13 369	13014	13 409	15 766	17 046	21777	23 975	25 945
Heavy	3 180	1 860	1 693	1 520	1 567	11 581	6 642	5156	6 083	5 575
total	**337 305**	**326 299**	**322 302**	**321 800**	**312 763**	**355 488**	**360 965**	**354 122**	**362 704**	**366 962**

Between 06:00 and 20:00, there are theoretically 631,000 vehicles circulating on the 1790 km of roads in the city, which inevitably increases the number of road accidents. Based on this evidence, the emergency plan for the city of Lisbon states that the road network is of decisive importance to reduce the number of accidents. The different characteristics of the city of Lisbon's road network allow it to be classified hierarchically as a municipal and supralocal network (image 19), enabling daily access to Lisbon in different ways.

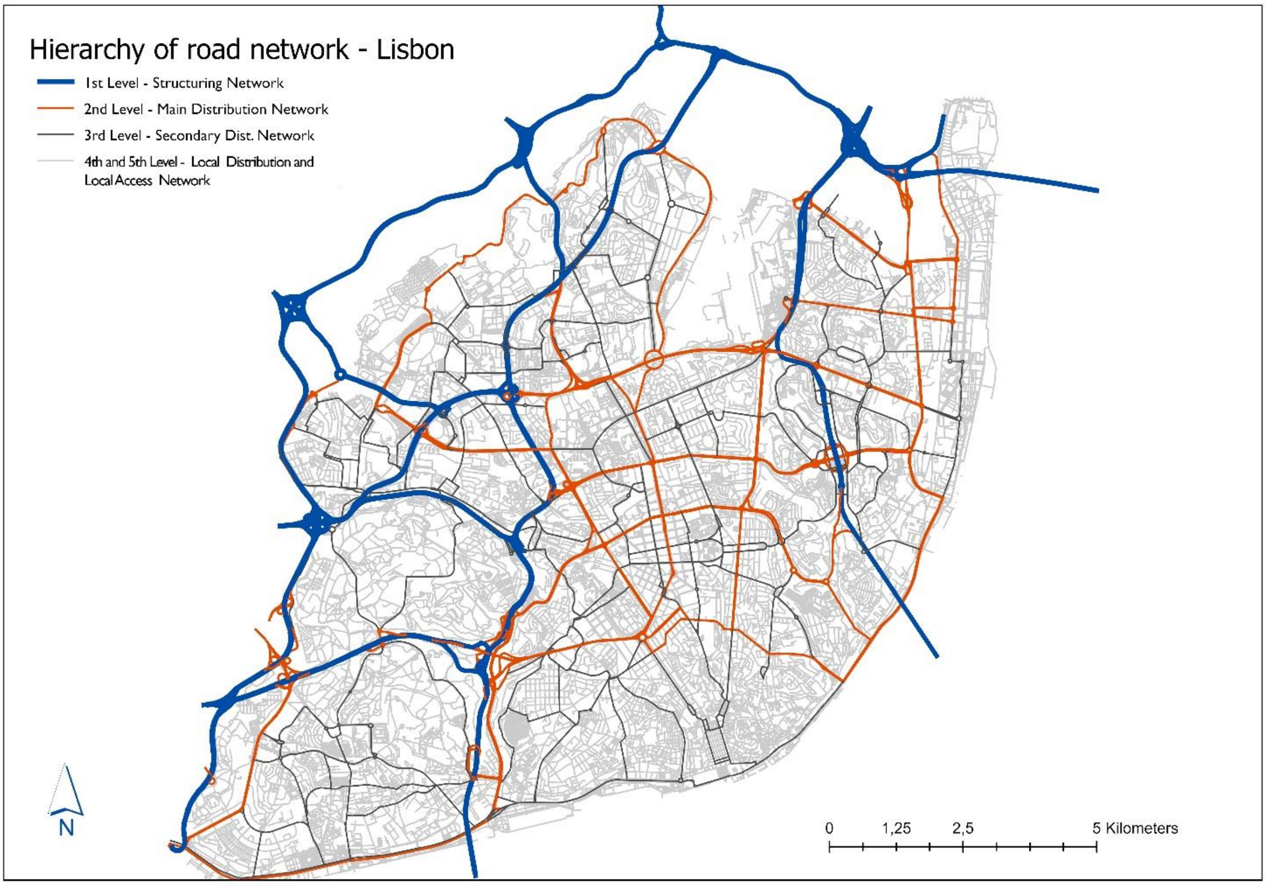

Image 19 - Lisbon road network hierarchy

According to the PME [73], "the road network should be ordered according to the functions and characteristics of the roads in:

- 1st Level - Structuring Network – ensures the connections within the municipalities and their transposition, in addition to long-distance travel within the city of Lisbon;
- 2nd Level - Main Distribution Network – ensures the distribution of the largest internal traffic flows within the Municipality, as well as the medium routes and the access to the structuring network;
- 3rd Level - Secondary Distribution Network – comprised of internal routes, ensuring the proximity distribution, besides guiding the traffic to the superior routes;
- 4th Level - Local Distribution Network (proximity network) – the structuring roads of the neighbourhoods, with some flow capacity, but where the main element is the pedestrian;
- 5th Level - Local Access Network (neighbourhood network) – ensures road access to buildings, with privileged conditions for pedestrian circulation (...)". Between the various axes [28, pp. 96] "(...) that serve the city, the highlight is given to the main access corridors, namely:
    - - Cascais (IC15 and EN6/ Avenida Marginal);
    - - Sintra/ Amadora (IC19 and EN117);

- - Amadora/ Loures (IC22/ Radial de Odivelas);
- - West (IC1/ A8);
- - North (IP1/A1, EN10 and IC2/ Variant to EN10);
- - Vasco da Gama Bridge (IC13/ A12);
- - 25 de Abril Bridge (IP2/ A2)".

The municipality of Lisbon's road network consists of 24,057 road segments totaling 1790 km in length, divided between 1st and 2nd levels (340Km, 32%) and 3rd, 4th and 5th levels (730 Km, 68%).

By analysing the distribution of these roads across the 24 civil parishes of Lisbon, we can determine their density and assess those that may be at a higher risk of accidents due to higher road density.

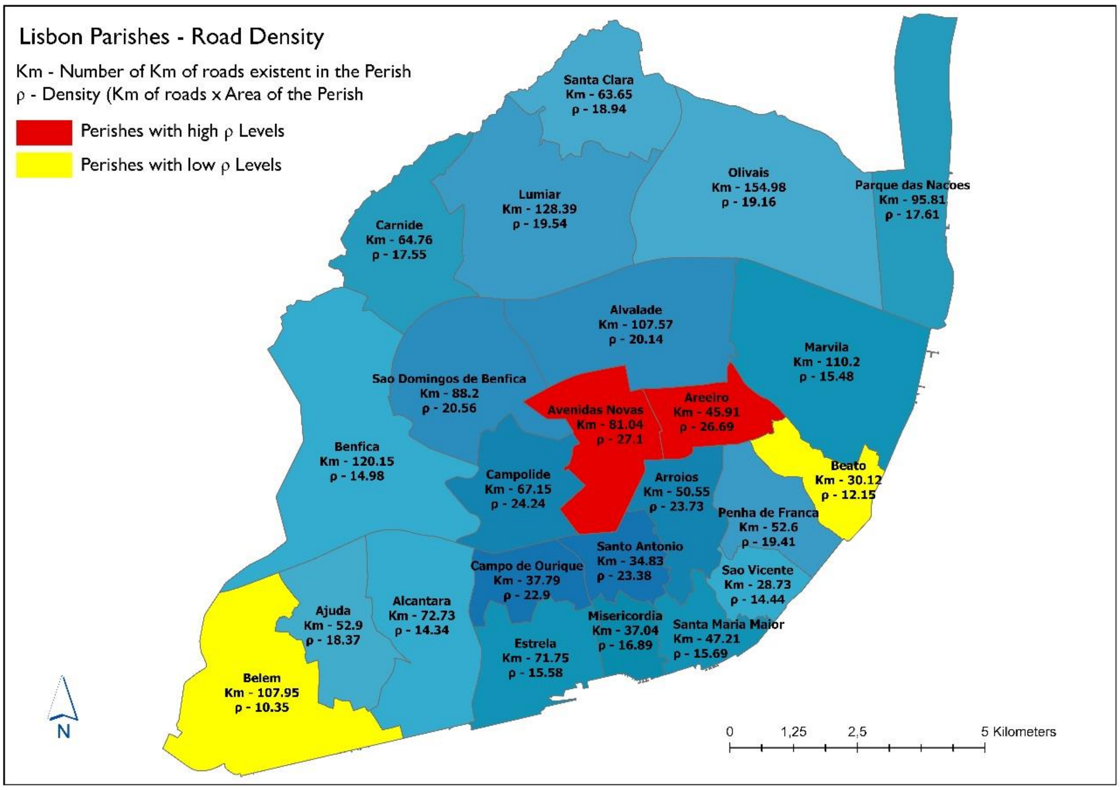

Image 20 - Lisbon civil pariches road density

After analysing the results expressed in image 20, it is clear that the civil parishes of Avenidas Novas and Areeiro have the highest road density. On the other end of the spectrum, Beato and Belém have the lowest density.

Three essential elements define a traffic stream: flow, speed. and density [78] A road network segment is said to be overloaded when it receives more traffic than it can accommodate. A saturation level of more than 90/100%, for example, implies a restricted and erratic traffic flow — i.e., the volume exceeds the artery's capacity, resulting in queues and waves of stop-and-go traffic. The images below (Image 21 and Image 22) represent the city of Lisbon´s saturation level during rush hours (morning 06:00-10:00 and afternoon 17:00-20:00). The sections with saturation levels greater than 90% (sections exceeding their theoretical capacity) are shaded in black.

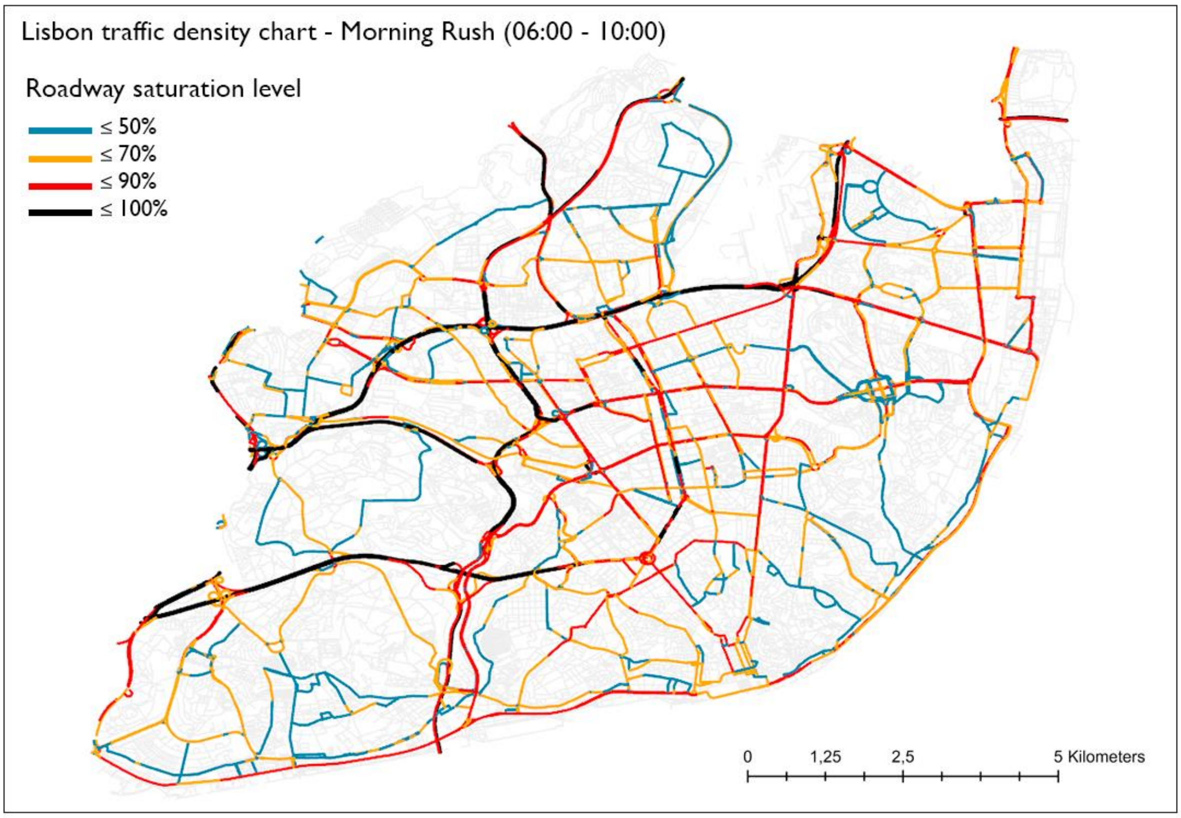

Image 21 - Lisbon traffic density chart - Morning

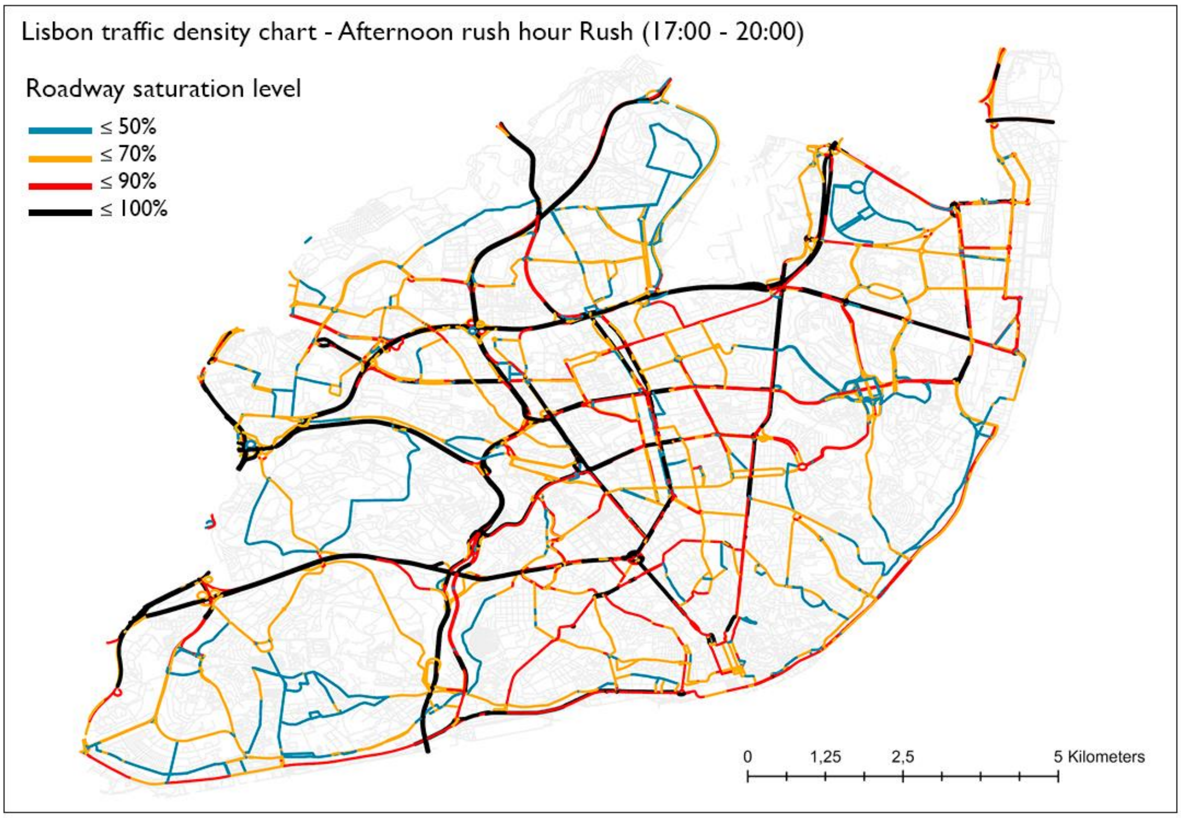

Image 22 - Lisbon traffic density chart - Afternoon

These results show that during morning rush hour, when traffic is most congested, there are a greater number of saturated roadways, which coincides substantially with radioconcentric penetration and distribution movements in the city center. The most congested routes in the Lisbon network are readily visible in both of the peak travel windows studied. In terms of the traffic load on Lisbon's road network during these time periods, the busiest axes with the highest demand has values of up to 1,650 PCU[8] per hour and per direction.

The zones identified by their high level of congestion reveal that the congestion produced by the density of vehicles is more representative at the times under consideration and prevalent mainly on three stretches of the Segunda Circular (near the Airport, on the Campo Grande Viaduct stretch and near the exit of the Benfica radial), in the city centre on the Avenida das Forças Armadas, Avenida dos Estados unidos da América, Avenida Calouste Gulbenkian, Avenida da Republica, Avenida da Liberdade, Avenida de Berna, Avenida da India and Avenida 24 de Julho.

[8] To deal with the heterogeneity in a mixed traffic environment, the Passenger Car Unit (PCU) is a relative weightage factor provided to the traffic volume of individual vehicle categories.

Page left intentionally blank

CHAPTER 4

Data Understanding and Data Preparation

4.1 Data Understanding

The CRISP-DM methodology's [79] data understanding phase is vital for achieving the goal and anticipating the emergence of unforeseen challenges in the following steps. It is composed of four distinct tasks: "Collect Initial Data," "Describe Data," "Explore Data," and "Verify Data Quality"[80].

This process entails gathering data, exploring it using visualization and analysis techniques, determining its quality, and finding relevant features to answer the research questions.

The ultimate goal of road accident analysis is to reduce the number and severity of accidents and casualties. To achieve this goal, high-quality, reliable, and accurate data must exist to understand which factors directly influence road accidents and the injuries that result from them.

4.1.1 Collect Initial Data

This subchapter addresses the collecting initial data task of the data understanding phase of the CRISP-DM methodology. The data acquisition process allows researchers to obtain the necessary information to achieve the objective defined in the research/process.

For this dissertation, the data provided by CML in the scope of challenge number 51 - "Identification of road accidents incidence points and their correlation with other factors" - did not fulfill the requirements for a correct statistical analysis. The records only included the year 2018, which is insufficient to identify hotspots accurately. The academy recommends that the ideal period for data collection when conducting a study on road accidents should be between 5 and 10 consecutive years since this interval represents a compromise between the number of years required for the sample to be considered statistically significant and the number of years necessary to ensure that the situation under study is not outdated.

Given the above situation, and after due coordination and consent from the supervisors of this dissertation, the necessary contacts to acquire the data required for this dissertation were made. These contacts included the Commander of the Lisbon Fire Brigade, Lieutenant Colonel Tiago Lopes, who received an official data request for 10 years of data through email from ISCTE-IUL. After the request was approved, the RSB asked Dr. Ulisses Pinto of the Information Systems Department of the Planning and Strategic Projects Division of CML to provide the dataset, and the file was delivered in Microsoft Excel Open XML Spreadsheet (.xlsx) format. The dataset provided was full of critical operational information but lacked the necessary fields for correct analysis — e.g., pavement conditions data and atmospheric data.

The next step was to request the data about meteorological conditions with an emphasis on the precipitation for the interval being analysed from Dr. Nuno Miguel Moreira of IPMA's Numerical Weather Forecast, Weather Watch and Earth Observation Division (DivMV). The data were delivered in comma-separated values (.CSV) format after a favourable order from the Institute, and they contained the atmospheric conditions data for 2 points of Lisbon city: The Main Automatic Station (nº 535 - EMA I), located at the Instituto Geofísico D. Luís (Lat. 38,71907778 Lon. -9,14972222 Elevation 77 m), and the Main Automatic Station (nº 579 - EMA I) Gago Coutinho, located near the Lisbon Airport (Lat. 38,76620278 Lon. -9,12749444 Elevation 104 m).

In order to have a complete dataset and obtain the missing data, we asked the owner of Windguru website (windguru.cz), Mr. Vaclav Hornik, for the historical wind information and acquired the historical hourly precipitation observations from the Open Weather Map History API [81], both files were delivered in comma-separated values (.CSV) format.

Having obtained the atmospheric data, it was necessary to complete the missing information to clearly understand why and when accidents occur in Lisbon. For that purpose, we contacted Dr. Ana Coroado, from the Safety Data Collection and Analysis Enforcement Unit of the ANSR, in order to obtain data from the Reports of the Statistical Bulletin of Traffic Accidents, the BEAV (Appendix A), which "is a statistical notation instrument filled in by the supervising entities (GNR and PSP) whenever they are aware of the occurrence of a traffic accident, to collect elements that allow depicting it as accurately as possible" [82]. To this end, ISCTE-IUL signed a protocol with ANSR that allowed and will allow future access to data from this National Authority. The files were delivered in Microsoft Excel Open XML Spreadsheet (.xlsx) format.

Although we had obtained all this data, we had not yet answered several questions that are essential to developing a full understanding of the traffic accidents in Lisbon: What is the volume of traffic on Lisbon's roads? How many vehicles are there in Lisbon, and what is the weight of commuter traffic? For that purpose, we asked Eng. Isabel Botelho from the Information Systems Services Department of the Institute for Mobility and Transport (IMT) for the average annual daily traffic (AADT) data. After approval, the dataset was submitted in comma-separated values (.CSV) format.

Since it was not possible to obtain the traffic volume data from the Gertrude traffic regulation system (a system of 1600 traffic sensors installed under the pavements of Lisbon's roads that allows the CML to manage the traffic light network and enable real-time analysis[9]), we needed to search for an alternative. The answer was TomTom Historical Traffic Stats [83], a website that provides anonymously collected data from TomTom's in-car navigation devices, smartphones and commercial fleets, allowing us to analyze the historical location data and traffic insights on road network speed, trip duration, and sample size on the road network.

Image 23 intends to demonstrate the several data sources utilized in this dissertation

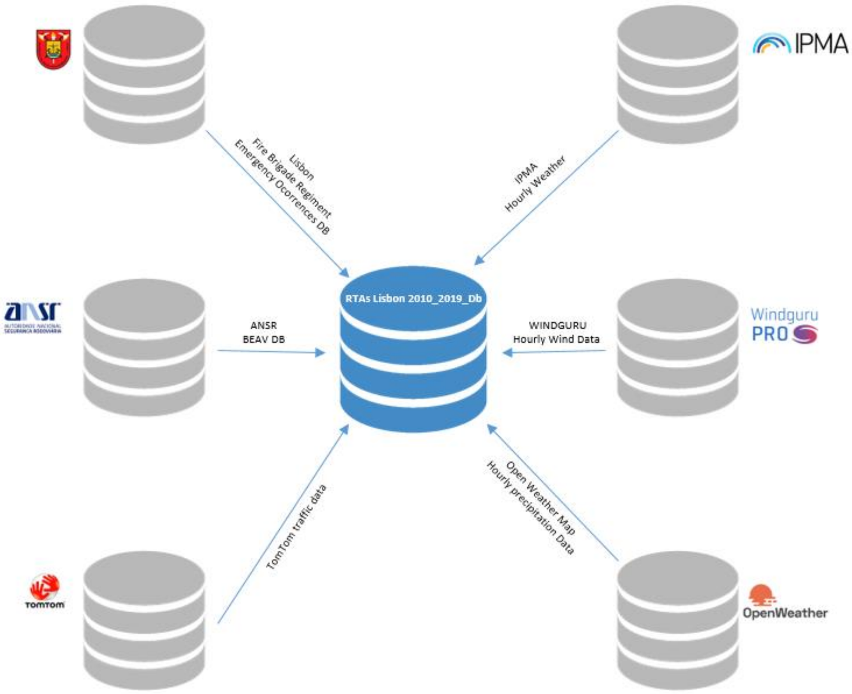

Image 23 - Data Sources

4.2 Describe Data

The CRISP-DM methodology identifies descriptive data analysis as one of the tasks required for data comprehension. The purpose of this step is to find problems in the quality and relevancy of the data. This is a critical stage in determining whether the acquired data contains all of the necessary information to solve the identified problem and deciding if the process can move forward to the information extraction step. This task includes identifying the data collection method, examining the

---

[9] This system is being replaced by the new Lisbon Intelligent Mobility System (SIM.Lx)

data format and size (number of observations and attributes) and conducting a preliminary analysis of the data using descriptive statistics through the use of IBM SPSS software to understand its distribution. This approach to the data allowed us to define a strategy for cleaning and the steps necessary to merge the data from the various sources. It was also possible to identify the existing problems in the datasets, which will be described separately below.

4.2.1 Lisbon Fire Brigade Regiment Incident Command System (GO)

The dataset included 157,214 emergency occurrences from all classifications and families performed by the RSB during the period under analysis. For this initial approach, a query with the SQL subcommand (GET DATA command) in IBM SPSS was created in order to select just the occurrences related to the three natures associated with an RTA — namely, accidents involving vehicles, accidents involving extrication operations and accidents involving pedestrians, which returned a total of 10,097 records. The first problem identified was related to an unintentional database update in 2019 that created a new nomenclature for the accident classification, which resulted in six typologies instead of the three correct ones (image 24).

1 943 025 112 018	25/11/2018 15:34:00	-9,16396947651384	38,717195733242	Rua Ferreira Borges	2101 - Acidentes - Rodoviários - Atropelamento
478 727 042 019	27/04/2019 15:11:00	-9,1611831679875 7	38,772737323243	Alameda das Linhas de Torres	2101 -> Acidentes -> Rodoviários -> Atropelamento
3 802 012 019	02/01/2019 17:52:00	-9,11295117174941	38,7567506385503	Avenida Dr. Augusto de Castro (antiga Avenida Central da Malha I de Chelas)	2102 - Acidentes - Rodoviários - Com viaturas
156 309 022 019	09/02/2019 22:18:00	-9,14471905971368	38,7078568103783	Rua de São Paulo	2102 -> Acidentes -> Rodoviários -> Com viaturas
110 323 012 019	23/01/2019 02:06:00	-9,11725272312503	38,7482624927774	Avenida do Santo Condestável (antiga Avenida Principal de Chelas)	2103 - Acidentes - Rodoviários - C/ Encarcerados
127 201 022 019	01/02/2019 12:35:00	-9,17963286925816	38,7235103528874	Estrada da Pimenteira	2103 -> Acidentes -> Rodoviários -> C/ Encarcerados

Image 24 - Duplicated typologies on RSBs incident command system

It was also found that the data relating to occurrences had less information available for analysis, since georeferencing of events was only possible after September 2012[10] due to lack of latitude and longitude coordinates before that time. Also, the street names in the database records were quite disparate — e.g., the same street could be described as "Avenida", "Av." or "Av", and some records included the street and police number.

4.2.2 ANSR - Data of Accidents Registered in the Statistical Bulletin of Traffic Accidents

The dataset was composed of 42,691 accident records. As previously defined, this dissertation is only researching accidents that produced victims (Deaths, Serious Injured, Light Injured). Thus, the first task applied to the data was, again, a query with the SQL subcommand (GET DATA command) in IBM SPSS to select just those accidents that corresponded to this criterion. As a result, a total of 22,725 records were obtained. As with the RSB data, the years 2010 and 2011 (6,323 records) did not contain

[10] Year in which the RSB emergency vehicles were equipped with the georeferencing module

geographic information to allow for their georeferencing, and the same problems related to street names were present.

4.2.3 Weather Data from The Portuguese Institute for Sea and Atmosphere

The records contained in the dataset showed several inconsistencies, particularly concerning the recording of wind speed and direction (absences) and precipitation. Several precipitation records had a value of "-990" (value expressed in mm/hour), which was identified initially as a natural phenomenon known as negative rainfall[11], but was later clarified by IPMA specialists as a weather station code error (sensor not working).

4.2.4 Data from WindGuru

The records included in the dataset contained 87,829 lines of hourly wind information speed and direction. Very few missing data were encountered.

4.2.5 Data from TomTom

As a commercial service, TomTom's data is only available through a contract. It was not possible to obtain data for the whole period under analysis due to the financial burden that this acquisition represented, so it was decided to limit the acquisition to a period of one year (2019; organized in monthly datasets subdivided into weekdays and weekend data). Tables 4 and 5 contain the values obtained through the TomTom probes (GPS, smartphones and commercial fleets) for the seven daily periods defined for this dissertation. Due to the cost to obtain the data for a 10-year span, it was assumed for the purposes of this dissertation that the values obtained are sufficiently representative of the dynamics of Lisbon traffic. So, the data results were applied to the full extent of the dataset.

Table 4 - Time intervals for TomTom data at weekdays

	Weekdays						
Time Sets	Mon 06:00-10:00	Mon 10:00-12:00	Mon 12:00-14:00	Mon 14:00-17:00	Mon 17:00-20:00	Mon 20:00-00:00	Mon 00:00-06:00
	Tue 06:00-10:00	Tue 10:00-12:00	Tue 12:00-14:00	Tue 14:00-17:00	Tue 17:00-20:00	Tue 20:00-00:00	Tue 00:00-06:00
	Wed 06:00-10:00	Wed 10:00-12:00	Wed 12:00-14:00	Wed 14:00-17:00	Wed 17:00-20:00	Wed 20:00-00:00	Wed 00:00-06:00
	Thu 06:00-10:00	Thu 10:00-12:00	Thu 12:00-14:00	Thu 14:00-17:00	Thu 17:00-20:00	Thu 20:00-00:00	Thu 00:00-06:00
	Fri 06:00-10:00	Fri 10:00-12:00	Fri 12:00-14:00	Fri 14:00-17:00	Fri 17:00-20:00	Fri 20:00-00:00	Fri 00:00-06:00

[11] Negative rainfall – "occurs when a particular area experiences more atmospheric evaporation than water downpour over a period of time. The phenomenon usually occurs when temperatures are unusually high, and precipitation is unusually low for the season"[99].

Table 5 - Time intervals for TomTom data at weekends

Weekends							
Time Sets	Sat 06:00-10:00	Sat 10:00-12:00	Sat 12:00-14:00	Sat 14:00-17:00	Sat 17:00-20:00	Sat 20:00-00:00	Sat 00:00-06:00
	Sun 06:00-10:00	Sun 10:00-12:00	Sun 12:00-14:00	Sun 14:00-17:00	Sun 17:00-20:00	Sun 20:00-00:00	Sun 00:00-06:00

4.2.6 Data from the Open Weather Map

The records included in the dataset contained 87,829 lines of hourly weather information. Very few missing data were encountered.

4.3 Verify Data Quality

This procedure's data quality analysis findings indicated substantial flaws that required a significant amount of effort to resolve. Therefore, in this dissertation, the tasks of exploring the data and verifying the data's quality were completed virtually concurrently with the data preparation step, which is entirely acceptable according to the CRISP-DM methodology, a "highly iterative, creative process with many parallel activities" [80].

4.4 Data Preparation

The data preparation phase consists of five distinct processes (select data, clean data, construct data, integrate data, and format data) that culminate in the creation of the final dataset(s) that will be utilised for modeling or the project's critical analysis activity [79]. This phase is also the most time-demanding data mining process, accounting for up to 50% or 70% of overall project time [84].

The first step in building the ETL that will enable the final dataset to be created is to define which tables and columns are necessary (previously identified in the preceding phase) and which may be discarded during the data preparation process. Three different criteria were utilised to pick the data for the dissertation final dataset, as indicated by Peter Chapman [80]: relevance to the study purpose, quality and technical constraints. This way, several columns of the different datasets were eliminated to obtain only the necessary data and avoid wasting time in the data cleaning phase. After this step, we started the correction phase of the problems previously identified in the datasets. As we applied practically the same processes to both files and shared characteristics between them (RSB and ANSR), they will be described together.

4.4.1 Lisbon Fire Brigade Regiment and ANSR Datasets

The first step performed on the RSB data was to normalise the duplication of categories. To this end, the six pre-existent categories were standardised by applying the codes of the occurrences' classification table established by the Permanent Operational Standard 3101/2009 of ANEPC (2001;2002;2103).

AGRUPAMENTOS_POR_CATEGORIAS	OCB_NAT	OCO_CAT	OCO_SUBCAT_1	OCO_SUBCAT_2
Acidentes Rodoviarios	2101	Acidentes	Rodoviários	Atropelamento
Acidentes Rodoviarios	2102	Acidentes	Rodoviários	Com viaturas
Acidentes Rodoviarios	2103	Acidentes	Rodoviários	C/ Encarcerados

Image 25 – RTA classification update

Once this nonconformity was overcome, the focus was centred on correcting the lack of georeferencing data in both files; the RSB file contained only georeferenced information after September 2012[12] and the ANSR file only from 1 January 2012. To perform the address geocoding process (obtain latitude and longitude coordinates through the address), it was necessary to normalise the addresses contained in the RSB and ANSR files since the street names in the databases' records were quite disparate — e.g., the same street prefix could be described as "Avenida", "Av." or "Av", and some records included the street and police number in the same column.

To this end, the street numbers were extracted, and a comparison of those addresses with the addresses present in the CML road network database was performed. This process was achieved through the Excel plug-in, Fuzzy Look Up, which allows for the comparison of strings (Table 6) by assigning a percentage of similarity between the strings. Thus, each record needed to be checked to determine whether the correspondence was correct or not. It was necessary to make corrections manually in some cases.

[12] Year in which the RSB emergency vehicles were equipped with the georeferencing module.

Table 6 - Fuzzy LookUp Results

OCO_ID	OCO_SIG	Street name CML	Similarity
153018394	Rua Alfandega	Rua Alfandega	100%
15502076	Rua Azedo Gneco	Rua Azedo Gneco	100%
652917	Rua da Rosa	Rua da Rosa	100%
18353255	Rua Jacinto Nunes	Rua Jacinto Nunes	100%
1153564	Rua Morais Soares	Rua Morais Soares	100%
2354135	Calcada Carriche	Calcada de Carriche	95%
24554196	Rua Amoreiras	Rua das Amoreiras	93%
3304342	Av Eusébio da Silva Ferreira	Avenida Eusébio da Silva Ferreira	86%
10204462	Av. General Norton de Matos	Avenida General Norton de Matos	86%
22354707	Avª Defensores Chaves	Avenida Defensores de Chaves	62%
2304711	Rua de S. Caetano	Rua de São Caetano	62%
24464722	R. Conselheiro Vaz	Rua Conselheiro Lopo Vaz	56%

Once the nomenclature was standardised, it was possible to proceed with the georeferencing of the missing data in the datasets using the online application Geocode (https://geocode.localfocus.nl/), which uses a key API from Google that allows for the conversion of addresses into coordinates. The address entered had to correspond with an existing address in Google's database, and the format of the address entered had to include the name of the street, parish, district and country. As a result of this process, the missing georeferences in the dataset were obtained (372 registries in the RSB file and 183 in the ANSR). Where it was not possible to identify the police number, the centre of the road was assumed as the point of the accident. A final correction was made to adjust the incident time to the nearest hour (pandas.Series.dt.round) in order to normalise the file and facilitate its future joining with other files.

4.4.2 Weather Data from The Portuguese Institute for Sea and Atmosphere

The atmospheric data provided by IPMA were composed of 28 columns and 87,829 lines with the hourly record (00:00 - 23:00) of the atmospheric conditions of 2 points in Lisbon: the Main Automatic Station (nº 535 - EMA I), located in the Instituto Geofísico D. Luís (Lat. 38,71907778 Lon. -9,14972222 Elevation 77 m), and the Main Automatic Station (nº 579 - EMA I) Gago Coutinho, located near the Lisbon Airport (Lat. 38,76620278 Lon. -9,12749444 Elevation 104 m). To correct the (-990) error and have a complete meteorological data set, we replaced the missing data in the IPMA file with the existing data from the donor tables.

4.5 Clean Data

Data cleaning is time-consuming and labour-intensive, especially when numerous data cleaning processes are required to remove large amounts of data from various data sources, as in this particular case. For this analysis, Python was chosen to clean the data, and the following libraries were used: “pandas [85]”; “NumPy [86]” and “datacleaner [87]” [13] The intended result of this process was to build a high-quality dataset. As a result, we carried out the following data cleaning operations (DCO) in order to improve the quality of the dataset: detect and fill missing values, standardize values from the fields, detect the use of special characters, clean the duplicate records and remove irrelevant data (columns) and white spaces.

4.6 Construct Data

The feature engineering phase of the CRISP-DM framework is addressed in this task; new characteristics are established or built for the dataset depending on the needs stated in the "describe data" task.

The final dataset contained 78 attributes — eight of which were the following derived attributes (new data created by aggregating or calculating fields from existing columns): Year, Month, Day, Period_Day, Weekday, Weekdays, Season, Group_Age — and two generated records: the first calculated through the ANSR equation (4) for the Gravity Indicator (GI), and the other through the ANSR Severity Index (5).

$$GI=100 \times F + 10 \times SI^a + 3 \times SI^b \tag{4}$$

GI = Gravity Indicator; F = Fatalities at 30 days; SI^a = Serious injuries and SI^b = Slight injuries.

$$SIx = \frac{NF}{NAV} x100 \tag{5}$$

Six =Severity Index for the year n; NF = number of fatalities for the year n; and NAV = Number of Accidents with Victims for the year n.

[13] The pandas and NumPy libraries were used for the data cleaning process, and the datacleaner library was used to address the encoding of the categorical variables of the datasets.

Table 7 - Calculation of the gravity Indicator and annual Severity Index values

Year	Fatalities 24h	Fatalities 30 Days	Serious injuries	Slight injuries	GI	Six	Rounded Six	NAV
2010	16	27	107	2919	12527	0,99	1	2734
2011	18	29	89	2579	11527	1,29	1,3	2256
2012	19	19	73	2599	10427	0,85	0,9	2233
2013	10	18	91	2663	10699	0,78	0,8	2296
2014	9	9	91	2754	10072	0,38	0,4	2370
2015	12	12	94	2963	11029	0,47	0,5	2544
2016	9	12	59	3150	11240	0,44	0,4	2697
2017	16	18	76	3276	12388	0,64	0,6	2813
2018	10	30	61	3327	13591	1,07	1,1	2813
2019	2	8	31	3326	11088	0,84	0,8	950

4.7 Integrate Data

In this phase of CRISP-DM, we proceed to the ETL construction; all data sources are combined in a single database, enabling analysis. To make the treatments applied to the data in the data understanding, data preparation and integrate data phases explicit, a graphic workflow (image 26) was created to promote quick reading and comprehension.

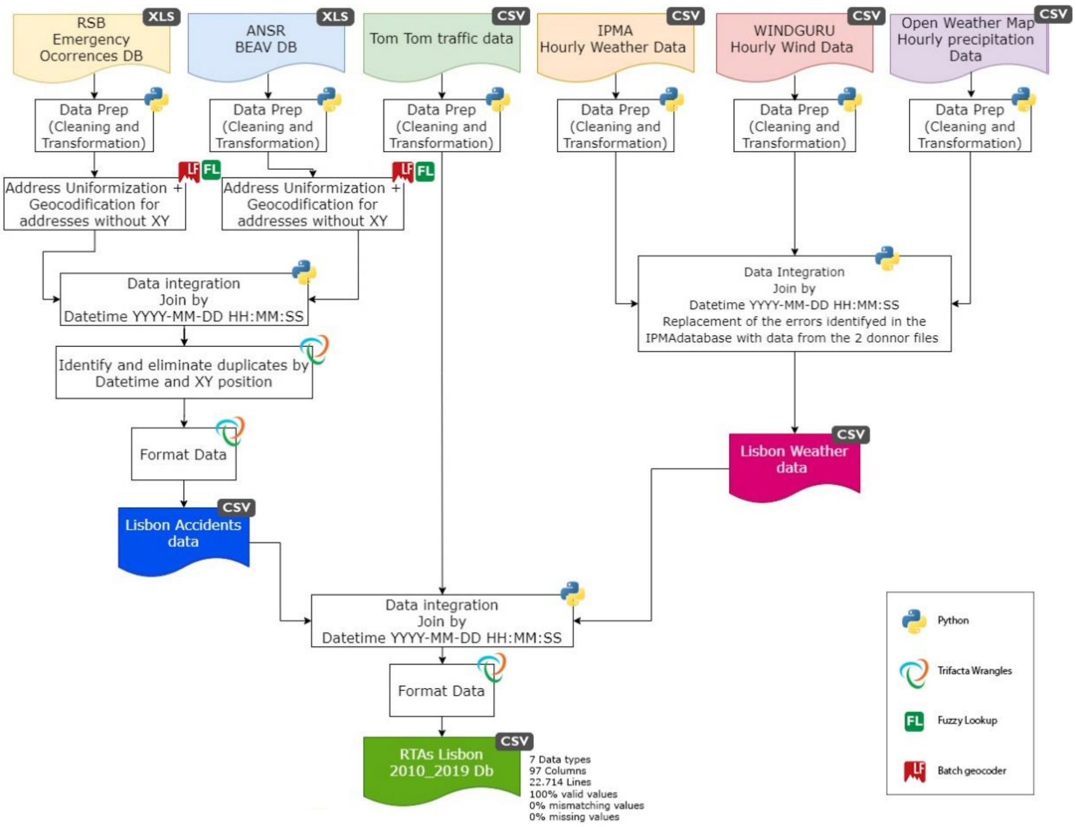

Image 26 - CRISP-DM flow applied to the data

The integration process was addressed in three stages:

1. A single dataset (Lisbon Accidents data.csv) was created in Python from the aggregation (joined by the Datetime field) of the two files containing the accident data (RSB and ANSR). A search for duplicates was executed in the Trifacta software [88] with the following criteria: XY location, Datetime and number of victims. The duplicate entries were eliminated.
2. A single dataset (Lisbon Weather data.csv) was created in Python from the aggregation (joined by the Datetime field) of the three files containing the weather data (IPMA, WINDGURU and Open weather).
3. Several datasets (Lisbon Accidents data, Lisbon weather data and TomTom Traffic Data) were merged according to the Datetime field, resulting in a single data file containing all of the relevant information for the dissertation. It is composed of 97 columns and 22,714 lines (each corresponding to a single accident that produced injuries in those involved), and it is structured as follows:

Table 8 - RTAs Lisbon 2010_2019_Db data schema

Column Name	Data Type	Description
Accident_ID	Original	The accident id
Date_Time	Featured	Date and hour of the accident
Year	Featured	Year of accident occurrence
Month	Featured	Month of accident occurrence
Day	Featured	Day of accident occurrence
Day_Section	Featured	Time of day when the accident occurred (Early morning; Morning Rush Hour; Morning; Lunch Rush Hour; Afternoon; Afternoon Rush Hour; Night)
Weekday	Featured	Day of the week when the accident occurred
Weekdays	Featured	Weekday or Weekend
Season	Featured	Season in which the accident occurred
Street	Original	Name of the street on which the accident occurred
Latitude	Original	Latitude where the accident occurred
Longitude	Original	Longitude where the accident occurred
Roadway_Type_ID	Original	ID for the type of Roadway (IP, IC, AE, EN)
Roadway_Type	Original	Roadway type (IP, IC, AE, EN, Street)
Parish	Original	Civil parish where the accident occurred
Location	Original	Identification of whether the accident occurred inside or outside the localities

Nature	Original	Nature of the Accident: - Hit-and-run accident, Animal hit-and-run accident - Pedestrian collision - Collision - Hit-and-run collision - Collision with other situations - Collision with vehicle or obstacle in the lane - Head-on collision - Side collision with another moving vehicle - Rear-end collision with another moving vehicle - Crash with a rollover - Collision with immobilized vehicle or obstacle - Crash with a restraint device - Crash with runaway - Crash with side restraint in place - Crash without restraint device - Simple crash
Injuries	Original	Severity of injuries (Slight injuries, Serious injuries, Death)
Sex	Original	Gender of the driver (Male-Female)
Driver_Age	Original	Driver age
Group_Age	Featured	Driver group age (<20; 20-29; 30-39; 40-49; 50-59; 60-64; 65-69; 70-74; 74-79; 80-84; 85-90; 90+)
Number_serious_injured	Original	Number of persons suffering severe injuries as a result of the RTA
Number_slightly_injured	Original	Number of persons suffering slight injuries as a result of the RTA
Number_Dead	Original	Number of persons deceased until 30 days after the accident as a result of the RTA[14]
GI	Featured	Calculated Gravity Indicator
SI	Featured	Calculated Severity Index
At_Factors	Original	Atmospheric conditions present at the time of the accident
Luminosity	Original	Existing light conditions present at the time of the accident: • Daylight • At night with street lights • At night without street lights • Dawn or twilight • Sun glare
Surface_cond	Original	Contains the adherence conditions at the time of the RTA: • With water accumulated in the lane • Wet pavement • Dry and clean pavement • Ice • With gravel or sand • With mud • With oil • On a wet road surface

[14] Mortal victim at 30 days is any person who dies at the place of the traffic accident or as a consequence of it up to 30 days after the occurrence — e.g., a seriously injured person who dies 5 days after the accident. This concept is used for road accident statistics purposes. In Portugal, it is regulated by law through Order 27808/2009.

Driver_Action	Original	Manoeuvre performed in response to the RTA
Cause_of_accident	Original	Identified cause of the accident
Vehicle_Category	Original	Class of the vehicle
Vehicle_Age	Original	Age of the vehicle
Roadway_specs	Original	Technical characteristics of the roadway
Roadway_st_of_cons	Original	Conservation state of the roadway
Roadway_Int	Original	Type of intersection where the RTA occurred: • At Junction • At crossroads • NOT DEFINED • On a deceleration lane • Outside of intersection • On connecting branch - driveway • On connecting branch - exit • At traffic circle • On acceleration lane • On deceleration lane • Outside the intersection • At a roundabout
Obstacles	Original	If there were any obstacles on the roadway
Directions	Original	The direction of the road in which the accident occurred: • On the increasing direction of mileage • On the decreasing direction of mileage
Traffic_signs	Original	If there were traffic signs on location and is type
Traffic_lights	Original	If there were traffic lights at the accident location and, if so, the condition of the traffic lights
Pavement_type	Original	Constitution of the roadway pavement
Roadway_geometry1	Original	Geometry of the section of the roadway where the RTS occurred: • Curve • Strait • N/A
Roadway_geometry2	Original	• At level • With inclination • On a slope
Roadway_geometry3	Original	• Paved roadside • Unpaved roadside • No roadside edge or impracticable
Roadway_geometry4	Original	• On parking lot • On the roadway • On the sidewalk • On the roadside
FRC	Original	Functional road class of this segment, ranging from the motorway (0) to destination road (8)
SpeedLimit	Original	Speed limit (kph)
AvgSp	Original	Average arithmetic speed for this time period (kph)
HvgSp	Original	Average harmonic speed for this time period (kph)
MedSp	Original	Median speed for this time period (kph)
Air_Temp	Original	Air temperature [ºC]
Thermal_sen	Original	Accounts for the human perception of temperature
T_min	Original	Minimum temperature at the moment
T_max	Original	Maximum temperature at the moment
WW_H	Featured	Weather warnings – High Temperature
WW_L	Featured	Weather warnings – Low Temperature

At_pressure	Original	Atmospheric pressure, hPa
Rel_hum	Original	Relative humidity [%] - The amount of moisture in the air in comparison to the amount of moisture the air can "hold" at that temperature
Wind speed	Original	Wind speed in Kph
Wind Classification	Original	IPMAs classification for wind conditions
Beaufort	Featured	Beaufort Scale Wind intensity - The Beaufort Scale is an empirical measure that relates wind speed to observed conditions at sea or on land [89]: 0. Calm 1. Light air 2. Light breeze 3. Gentle breeze 4. Moderate breeze 5. Fresh breeze 6. Strong breeze 7. Near gale 8. Gale 9. Strong gale3 10. Storm 11. Violent storm 12. Hurricane
Wind_Direction	Original	Wind direction in degrees
Wind_Direction_card	Featured	Wind direction, cardinal points
Max_WG	Original	Maximum value of wind gust in Kph
Precipitation1h	Original	Rain volume for the last hour, mm
Precipitation _Class_1	Featured	IPMAs classification for precipitation conditions at 1 hour
Precipitation3h	Original	Rain accumulated volume for the last 3 hours, mm
WW_P	Featured	Weather warnings – precipitation
Cloud cover	Original	Cloudiness, %
weather_main	Original	Group of weather parameters (Rain, Snow, Extreme, etc.)
weather_description	Original	Weather condition within the group

4.8 Format Data

All the data format processes were executed in Trifacta[15] and produced a final dataset named RTAs_Lisbon_2010_2019_db.csv

4.9 Other Data

The collaboration between ISCTE-IUL and the Lisbon municipality included the provision of additional geographic and cartographic data (Shape Files), such as the urban area limit; road network classification, limits and centerlines; land use characterization; civil parish limits; firehouse locations; road network speed limits; altimetry and flood vulnerability; road slope; horizontal and vertical signalling. However, not all of the information available was used.

[15] https://www.trifacta.com/

CHAPTER 5

Data Visualization

The fifth phase of our adaptation of the CRISP-DM framework is dedicated to analysing the data extracted from the dataset. First, the statistical data of the dataset was analysed using IBM's SPSS (Statistical Package for Social Sciences) software, which is the most widely used software for quantitative data analysis, and Tableau, which is the leading data visualization tool in the business intelligence industry. Second, existing hotspots in the municipality of Lisbon and the causal factors for their existence were identified.

5.1 Descriptive Analysis of Road Traffic Accidents in Lisbon Municipality

As mentioned earlier, the increasing number of traffic accidents is a challenge for transport infrastructures because of health issues, which are critical, and because of the financial burden, they pose to society. This descriptive study of road accidents and their causal factors is an essential contribution to the safety analysis of the organizations responsible for vehicular mobility, as it supports them in adopting appropriate preventive measures to reduce the number of accidents and the severity of their consequences. The present data analysis covers the years 2010 to 2019 and assesses the 22,725 road accident records that recorded injuries in the municipality of Lisbon.

5.1.1 Accidents in Lisbon

The analysis of the annual accident rate (image 27) shows a decrease in the year 2012 and a growing trend in the number of accidents until 2017, plateauing from this year until 2019. It would be standard practice to estimate that these values would hold steady and not experience significant growth in relation to annual trends. However, this prediction was complicated by the enormous decrease experienced in 2020 due to the COVID-19 pandemic.

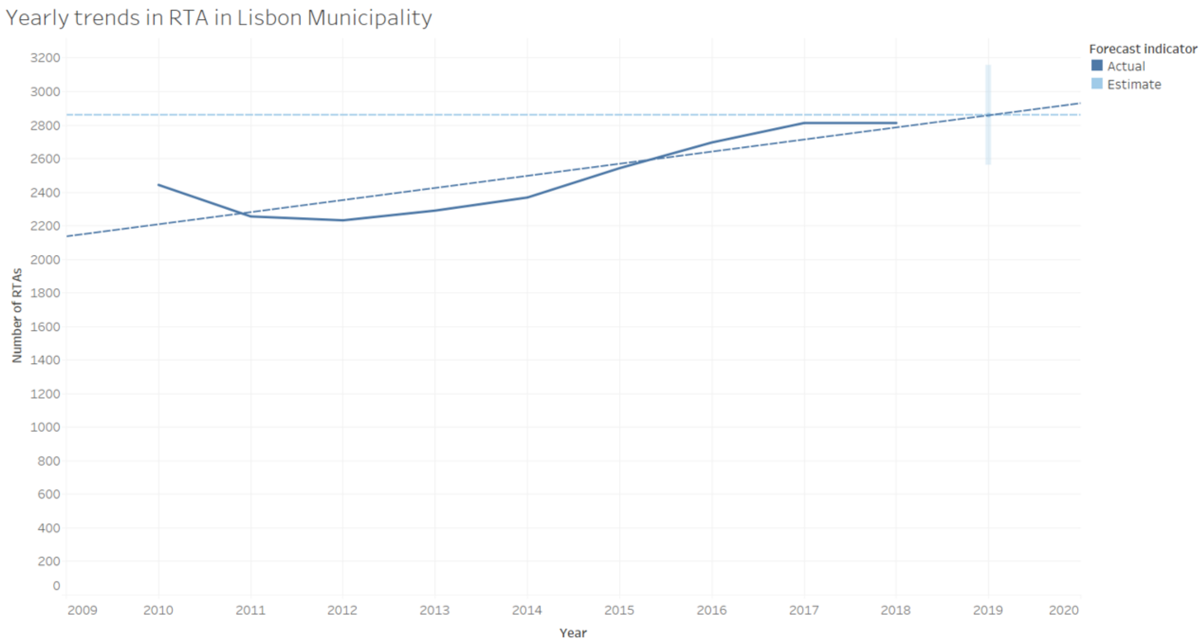

Image 27 – Yearly trends in RTA in Lisbon Municipality

As it is possible to observe, there has been a marked decrease in road fatalities in the last two years of our research. Image 28 shows two lines, a yellow one representing the number of Deaths/Fatal Victims at 24 hours, in other words, all the victims whose death occurs at the place of the accident or during the journey to the health unit, and a red line that represents the number of Dead/Fatal Victim at 30 days, which is all the victims whose death occurs at the place of the accident or during the period of 30 days after its occurrence.

The definition of 30-day fatality enables Portugal's mortality statistics to be compared to other countries, allowing for greater clarity when comparing mortality figures between countries. This statistical count of fatalities covers those who died on the scene or on their way to the hospital, as well as those who occurred within 30 days of the accident. As it is the official concept for accounting for victims, this is the value applied to all statistics and calculations made in this dissertation.

Image 28 - Trends in deaths resulting from RTA (24h and 30 days)

As with the mortality curve, the number of victims with serious injuries (Image 29) has shown a downward trend, which is thought to be due to increased enforcement measures.

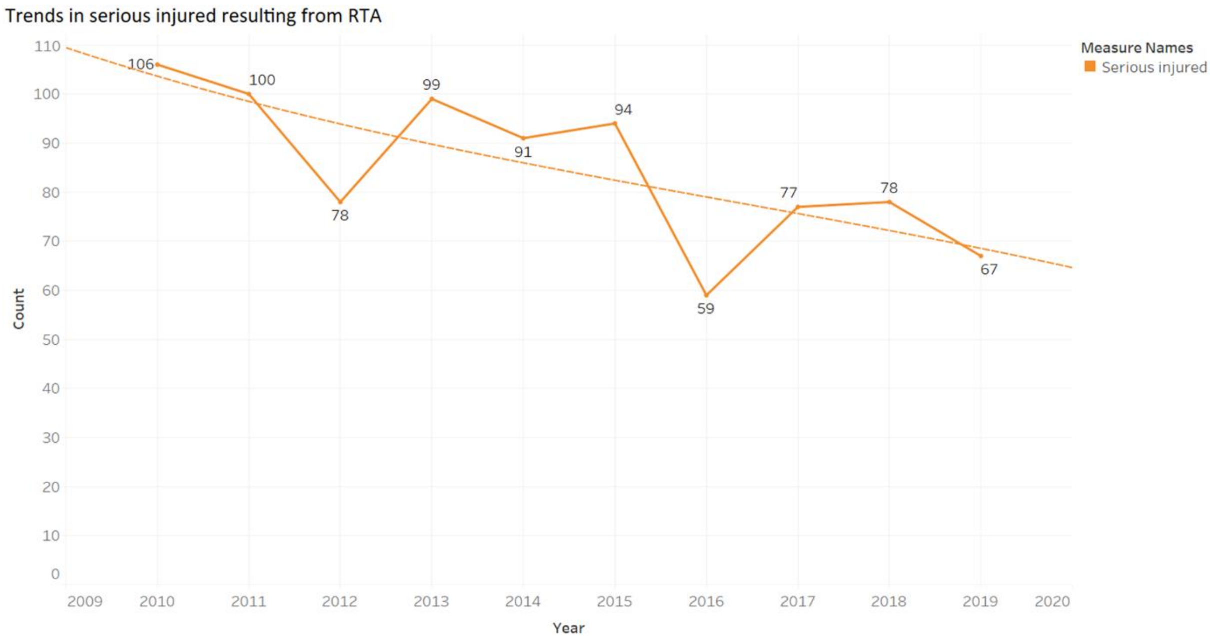

Image 29 - Trends in serious injured resulting from RTA

Regarding the monthly distribution (image 30), we can observe that, historically, the months of October and November present the highest number of accidents. In contrast, the month of August features an accentuated reduction (approximately 800 accidents), primarily due to the exodus of the population from Lisbon to take holidays (traditionally held at this time of year).

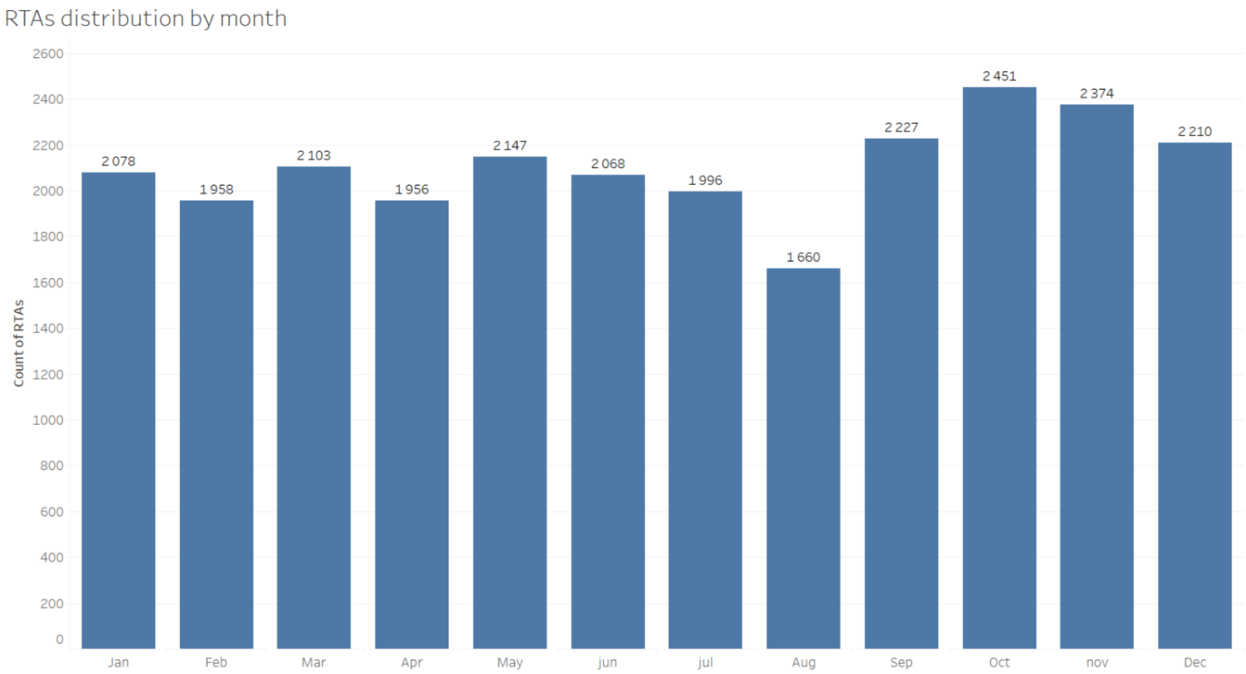

Image 30 - RTAs distribution by month

Regarding seasonal accidents (image 31), data analysis shows that most accidents occur in the autumn, which was already expected due to the weight of October and November and is thought to be related to the resumption of the working period after summer holidays. Similarly, because Lisboans are away from the city during the summer, when the end of the school year traditionally marks the beginning of the holiday periods, the naturally lower flow of cars in the city results in fewer road accidents.

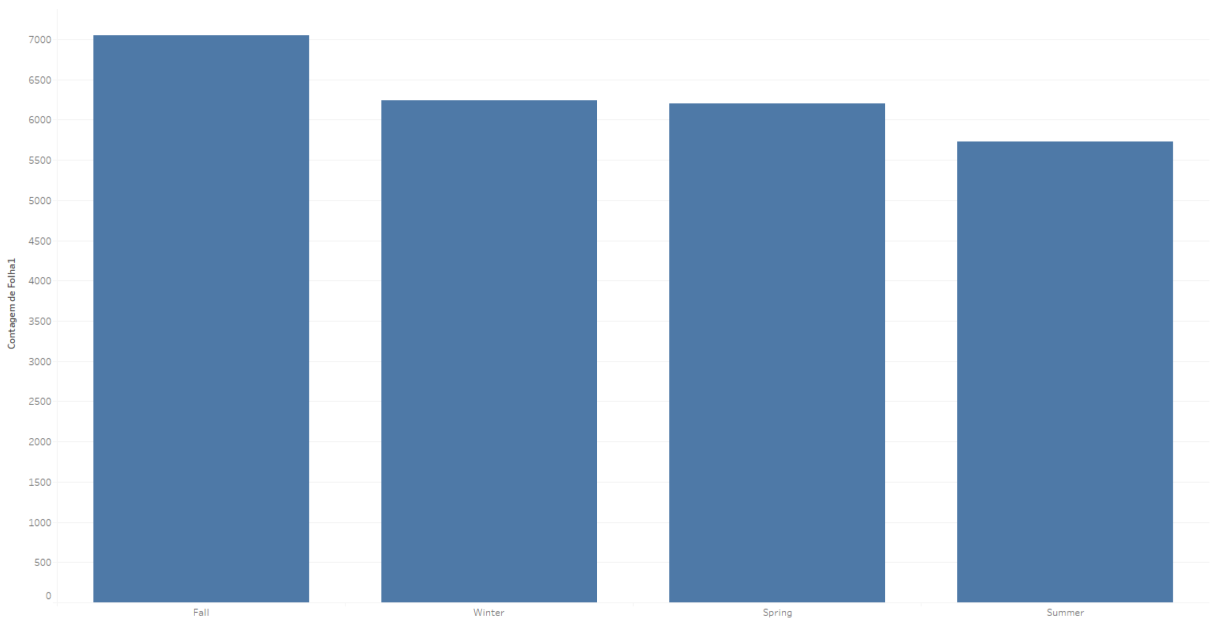

Image 31 - RTAs distribution by season

In terms of weekly accident distribution (image 32), we can observe that Fridays are particularly prone to traffic accidents involving victims; additionally, the number of serious injuries and fatalities reaches its peak on this day of the week, particularly during the night and early morning hours Friday and Saturday. When we segment the data by age group, we can see that the vast majority of victims are between the ages of 20 and 29 and that speeding was the primary cause of these accidents.

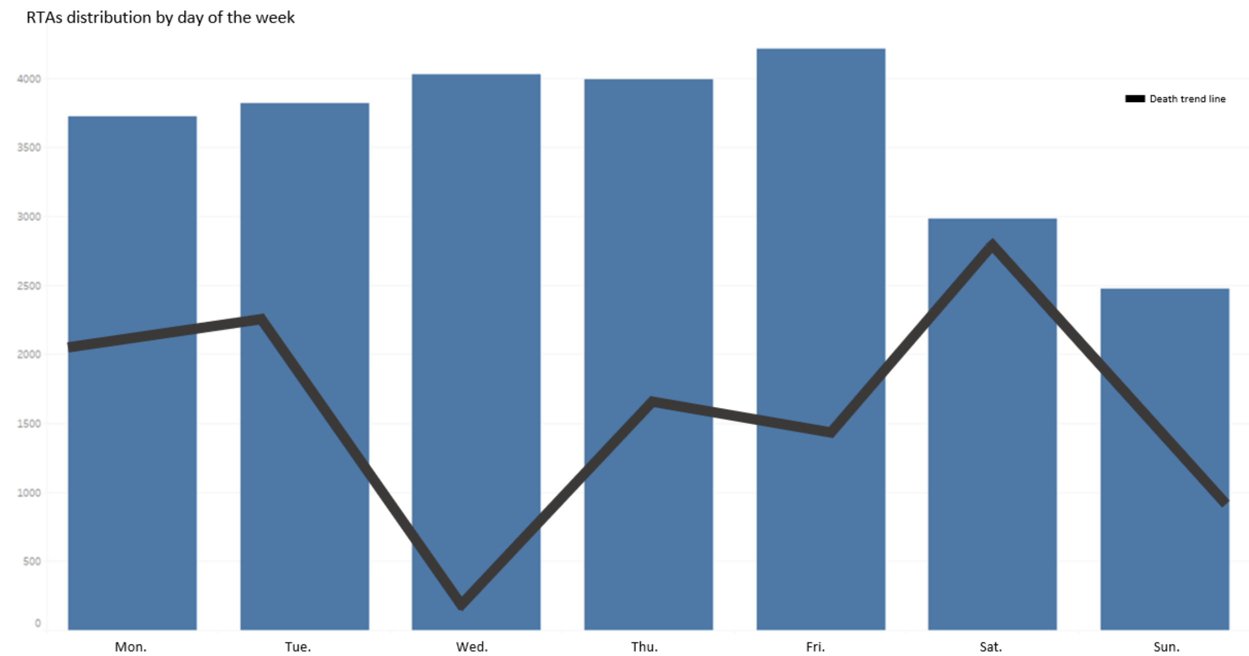

Image 32 - RTAs distribution by day of the week

As previously mentioned, in order to structure the research from a temporal point of view, the 24 hours of the day were divided into the following seven time segments (image 33):

1. Dawn - Covers the time from 00:00 to 06:00 and is characterised by little traffic density. During this period, there is an increase in average speeds of vehicles in circulation.
2. Morning Rush - Between 06:00 and 06:00, this period is characterised by a high traffic density due to the amount of commuter. This period ranks third in terms of daily accidents.
3. Morning - The time from 10:00 to 12:00 is characterised by an increase in average traffic speeds due to the decrease of traffic density.
4. Lunch Rush - Covers the time from 12:00 to 14:00 and is characterised by an increase in traffic density due to workers leaving their workplace to go to lunch. During this period, there is a decrease in traffic speeds, particularly in downtown Lisbon.
5. Afternoon - Covers the time from 14:00 to 17:00 and, as in the morning period, is characterized by an increase in average traffic speeds due to the decrease in traffic density. During this period, the second highest number of accidents is registered.
6. Afternoon Rush - The period from 17:00 to 20:00 is characterised by a high traffic density due to the return of commuter traffic to the satellite municipalities of the Lisbon Metropolitan Area. This is the daily period during which the highest number of accidents with victims are registered.
7. Night - Covers the time from 20:00 to 00:00 and is the third daily period during which an increase in average traffic speeds is again registered due to a decrease in traffic density.

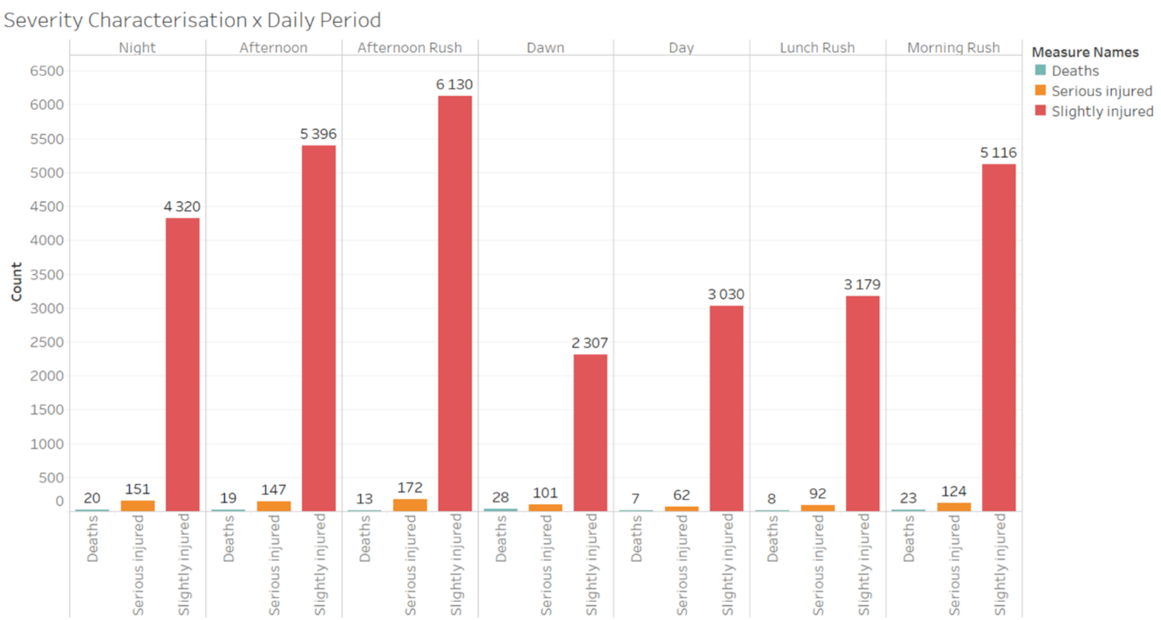

Image 33 - RTAs injury type separated by time segment

An analysis of the RTA by accident category (image 34) shows that pedestrian collisions are undoubtedly the most common type of road accident (with victims) in the municipality, which is to be expected because a collision between a vehicle and a pedestrian will inevitably cause extensive damage to the latter, Evidence of this is that pedestrian collisions are the category that generates the highest number of severe injuries and deaths on the roads in the city of Lisbon.

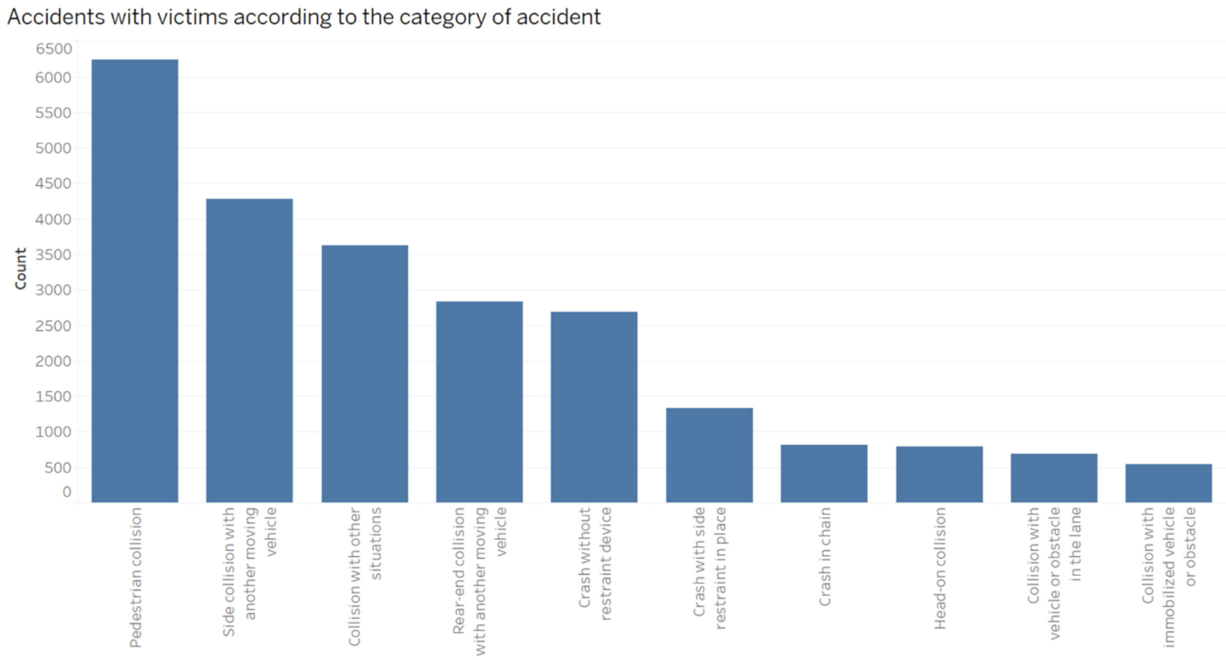

Image 34 - RTAs accordingly his typology

Avenida 24 de Julho is the one with the highest number of injuries to pedestrians in the three classes (image 35) under consideration (slight injuries, serious injuries, and deaths), followed by Avenida Almirante Reis, Avenida da República and Avenida Infante Dom Henrique. Regarding the second most frequent type of accident, collision, we can see that the Avenida General Norton de Matos (second ring road) has the highest number of occurrences, followed by the stretch of the A5 near the Viaduct Eduardo Pacheco, Avenida Infante Dom Henrique, Eixo Norte Sul, Avenida da República and Avenida Marechal Craveiro Lopes.

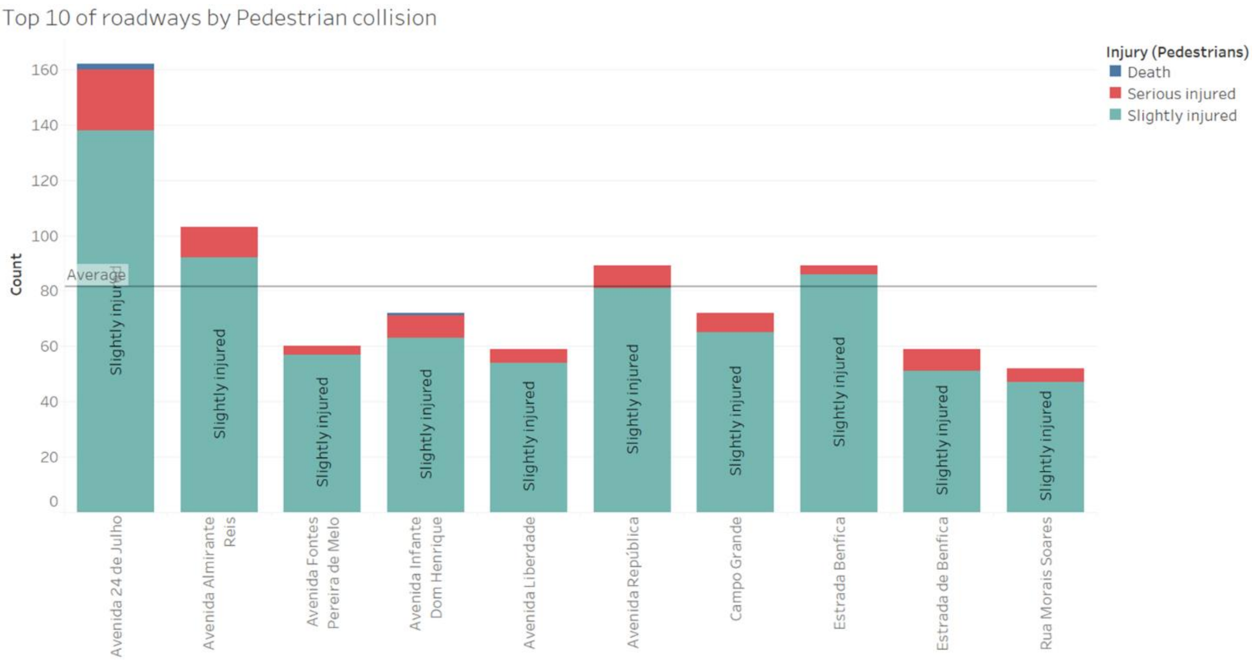

Image 35 - Top 10 roads - Pedestrian collisions by injure type

As this analysis focuses on Lisbon, we attempted to determine which civil parishes had the most significant number of accidents. According to Image 36, the parish with the highest incidence of accidents throughout the observation period was Santa Maria dos Olivais. Despite being one of the parishes with the highest density of roads, it has the most accidents. It is also possible to observe in the image 35 that Avenidas Novas and Areeiro are in eighth and twenty-second place in the number of casualties.

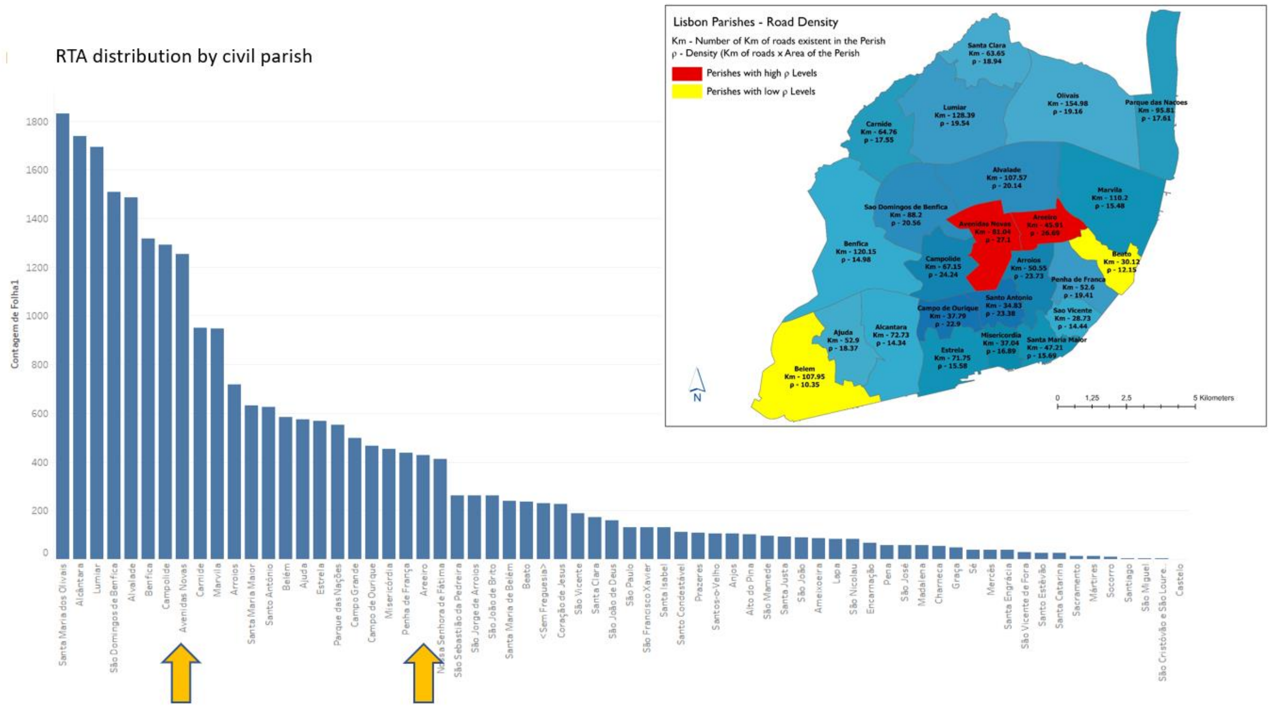

Image 36 - RTAs distribution by civil parishes

The study of the data identified that the vast majority of accidents are suffered/caused by light vehicles (image 37), and the second place belongs to motorbikes with a cylinder capacity not exceeding 125 cm3. This data reveals a trend that is gaining strength since 2009. On August 13, there was a legislative change that allowed holders of a driving license category B and over 25 years of age to drive motorbikes with cylinder capacity up to 125cc, This authorization produced a significant increase of this type of motorbike in circulation on the streets of Lisbon, which in turn caused an increase in accidents in this category, and it is expected that in the coming years this growth trend will be even more significant due to the appearance and growth of home delivery companies (125% compared to 2019, motivated by the pandemic and the restrictions imposed) that make use of this same means of transport.

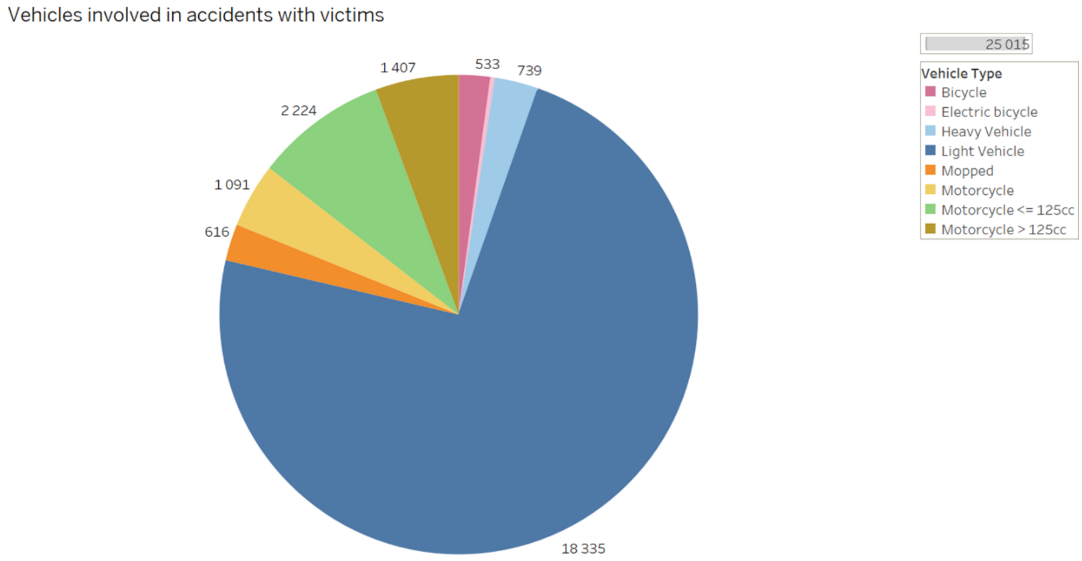

Image 37 - RTAs distribution by vehicle type

With regard to the age bracket of those involved in a road accident with victims (image 38) we see that the 30 to 39 age bracket is the most affected, followed by the 20 to 29 age bracket, proportionately male drivers are much more likely to be involved in an accident than female drivers, in our research 18054 males were involved in a RTA vs 6018 females.

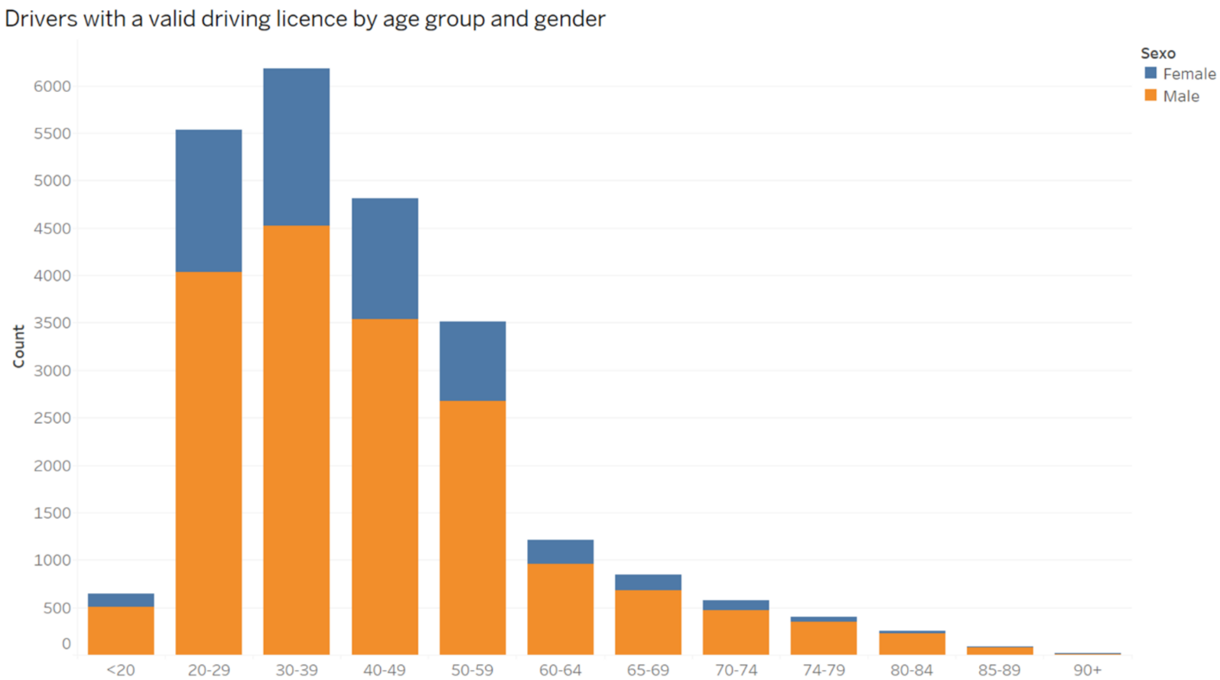

Image 38 - Drivers with valid driving license by sex and age group

When we examine the atmospheric factors (image 39), we can see that the vast majority of accidents (82,9%) occur in good weather, accidents with registered precipitation account for only 16,9 percent of our records, and the remaining categories of weather conditions, such as fog, wind, snow, and hail, have no expression (0,3%), so we chose to exclude them from this analysis.

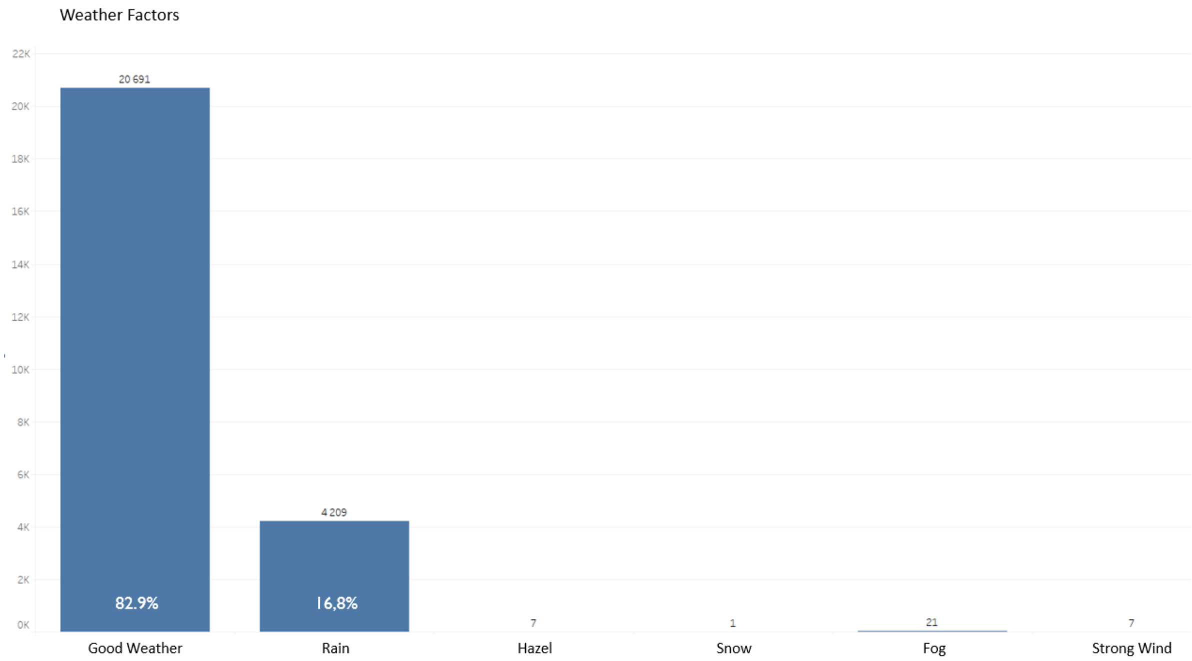

Image 39 - Distribution of RTAs by weather conditions

Correlating the RTAs with the two statistically significant atmospheric factors (image 40), good weather, and rain, with the road maintenance conditions, we can see that the general state of maintenance of the roads in Lisbon is in good condition, the number of RTAs occurring with pavement signs in poor condition corresponds to only 0.02% of the general total, so we can assume that the pavement conditions on the roads in Lisbon are generally good.

Relating the previous factors with the lighting factor, we can see that the vast majority of incidents (68%) occurred in broad daylight, 28% in night conditions with lighting, 3% during dawn and dusk periods, and 1% for the remaining lighting conditions (Night without lighting, and dawn). These values differ little when comparing RTAs occurring in good weather and those occurring with precipitation.

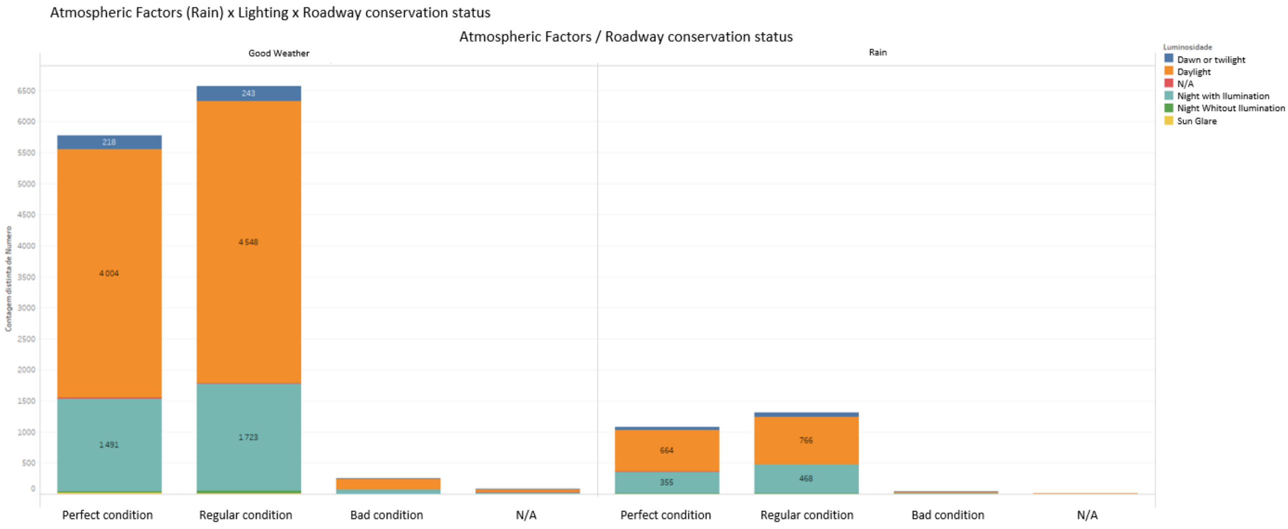

Image 40 - Distribution of RTAs by weather factors, lighting conditions and road surface

Making a selection of the accidents that occurred in situations of precipitation with the conditions of road surface adherence (image 41), we can observe that only thirty-one episodes of aquaplaning were registered in the city of Lisbon during the period under analysis, except for two incidents, the location of these RTAs coincides with the areas identified in Image 16 as extremely prone to flooding.

Regarding the accidents recorded as having occurred with wet soil conditions (when precipitation was not yet enough to wash the oils and mud from the floor), we found an average value of precipitation in the order of 0,6 mm/h. Another conclusion that we were able to identify is that the number of these occurrences is contrary to what would be expected much lower than the number of accidents that took place with precipitation values above 3.4 mm/h, which may indicate that drivers, knowing that the road is slippery, observe more carefully and pay more attention to the speed at which they move.

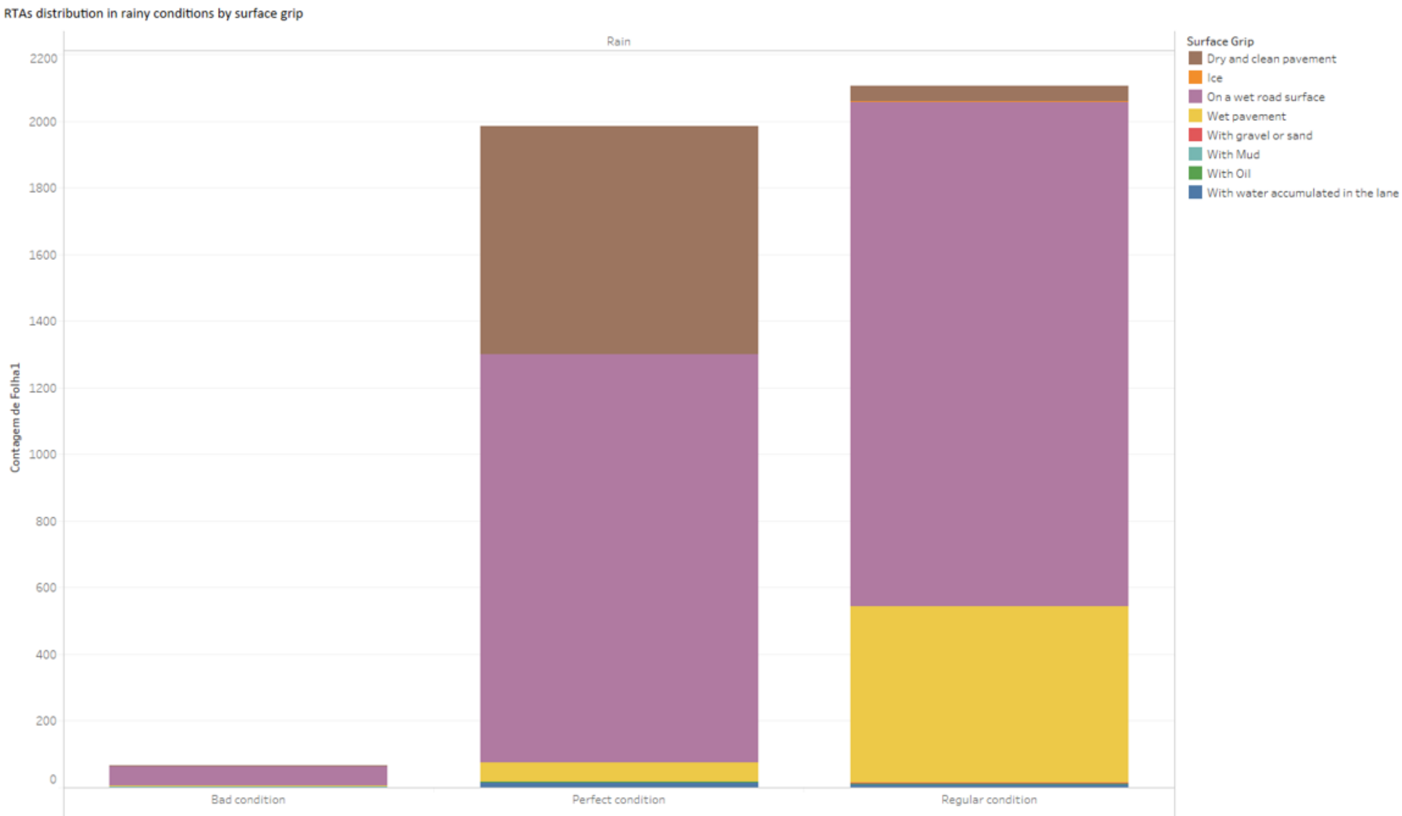

Image 41 - RTAs distribution in rainy conditions by surface grip

The analysis of the influence of the factor "wind" on the RTAs proved fruitless (image 42). The wind regime in Lisbon at the surface is dominated by a strong flow coming from the north "northerly" this regime is strongly related to the Azores anticyclone and is predominant during the summer months, in winter, or in situations of bad weather it is observed a rotation of the dominant winds to the south-southeast quadrants. This pattern is perfectly identifiable in our study, and no outliers or patterns were identified that could justify an increase of RTAs due to a specific value of gust or wind azimuth.

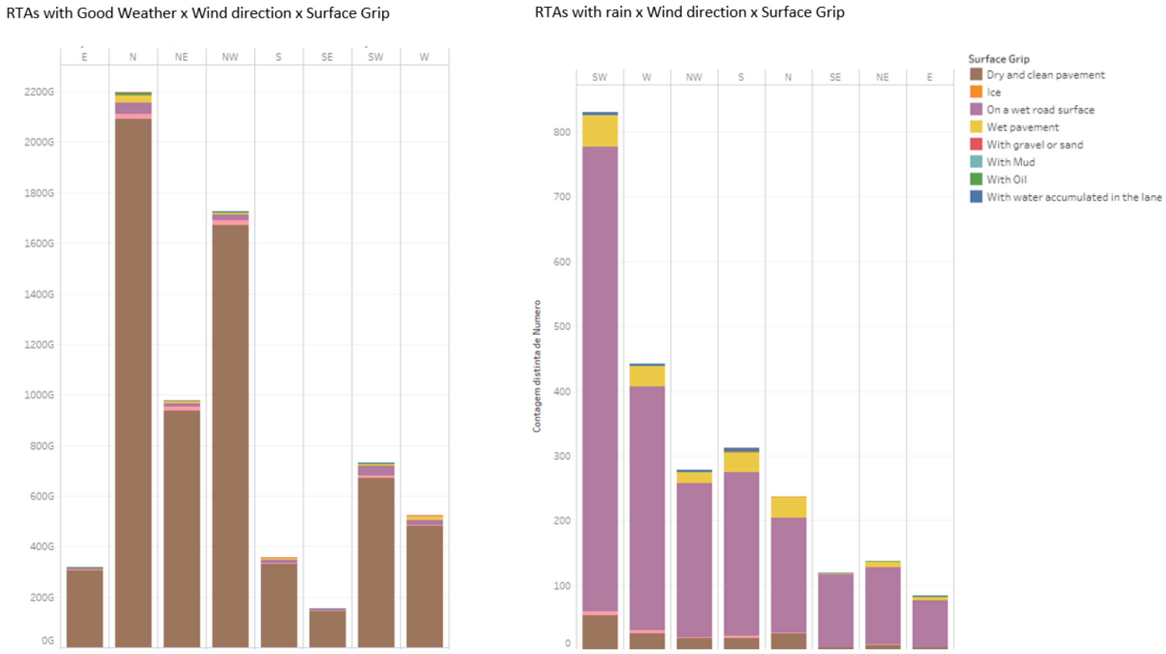

Image 42 - RTAs distribution by wind conditions

An examination of the records of the causes (contributing factors) attributed by enforcement authorities, excluding occurrences where no causal factor can be identified (image 43), reveals that the vast majority of accidents are caused by speeding, followed by irregular maneuvers, disregard for vertical signs, and disrespect for traffic lights. This analysis validates the previously defended thesis that human factors are the primary cause of road accidents.

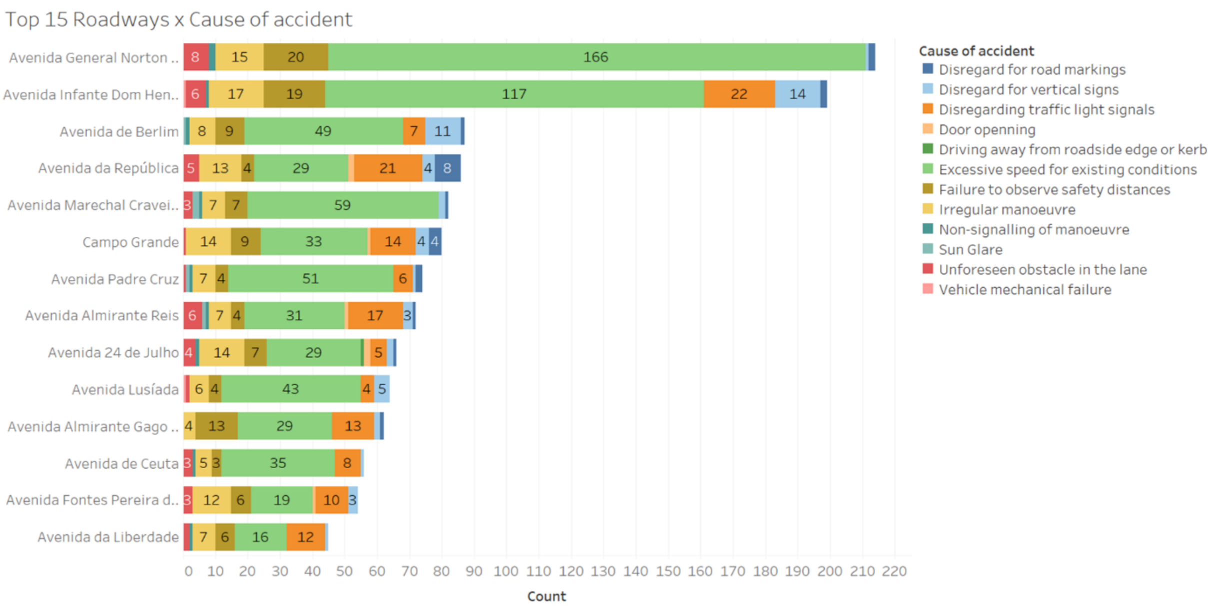

Image 43 - Analysis of causes of an RTA on the top 15 streets with more accidents

5.1.2 Distribution of RTAs by Fire Station Areas of Responsibility (AoRs)

Since this is a statistical analysis of road accidents in the city of Lisbon, which intends to contribute with data for the improvement of mobility and safety in the municipality, we could not fail to analyze the distribution of the three classes of road accidents that more means and personnel commit, namely RTAs, vehicle-pedestrian collisions, and accidents with trapped victims.

In this way, it is possible to observe that the RSB headquarters with more interventions of the road accidents typology (image 44) is the Benfica headquarters (5th company), followed by the Defensores de Chaves headquarters (4th company). The Defensores de Chaves headquarters, due to its proximity to the city center, is the one with more occurrences of this type of accident. Following and sharing the logic of proximity to the city center, we have the axis created by the headquarters of Benfica, Dom Carlos I, and Alvalade.

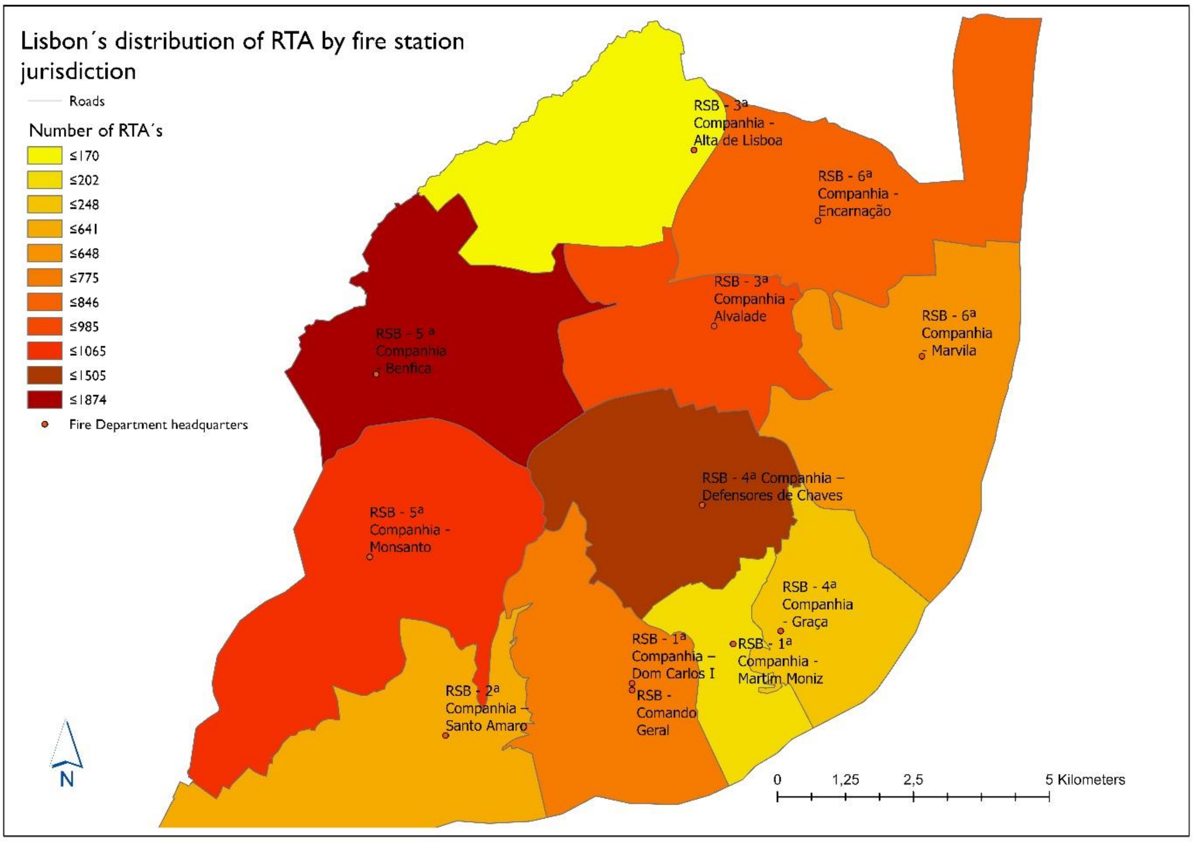

Image 44 - RTAs distribution by Fire Station Jurisdiction

Regarding incidents with incarcerated persons (image 45), the headquarters with the highest number of incidents is the Benfica barracks, mainly due to its location within the area of responsibility of the Eixo Norte-Sul (the road with the highest number of records of this type), Avenida General Norton de Matos and Calçada de Carriche. In the second place the Defensores de Chaves fire station responsible for the occurrences in Avenida João XXI, Avenida da Republica, and Avenida General Correia Barreto, then the Monsanto fire station with interventions at the end of the Eixo Norte-Sul, in the access of the Highway A5, Avenida das Descobertas and the entrance of Avenida da Ponte.

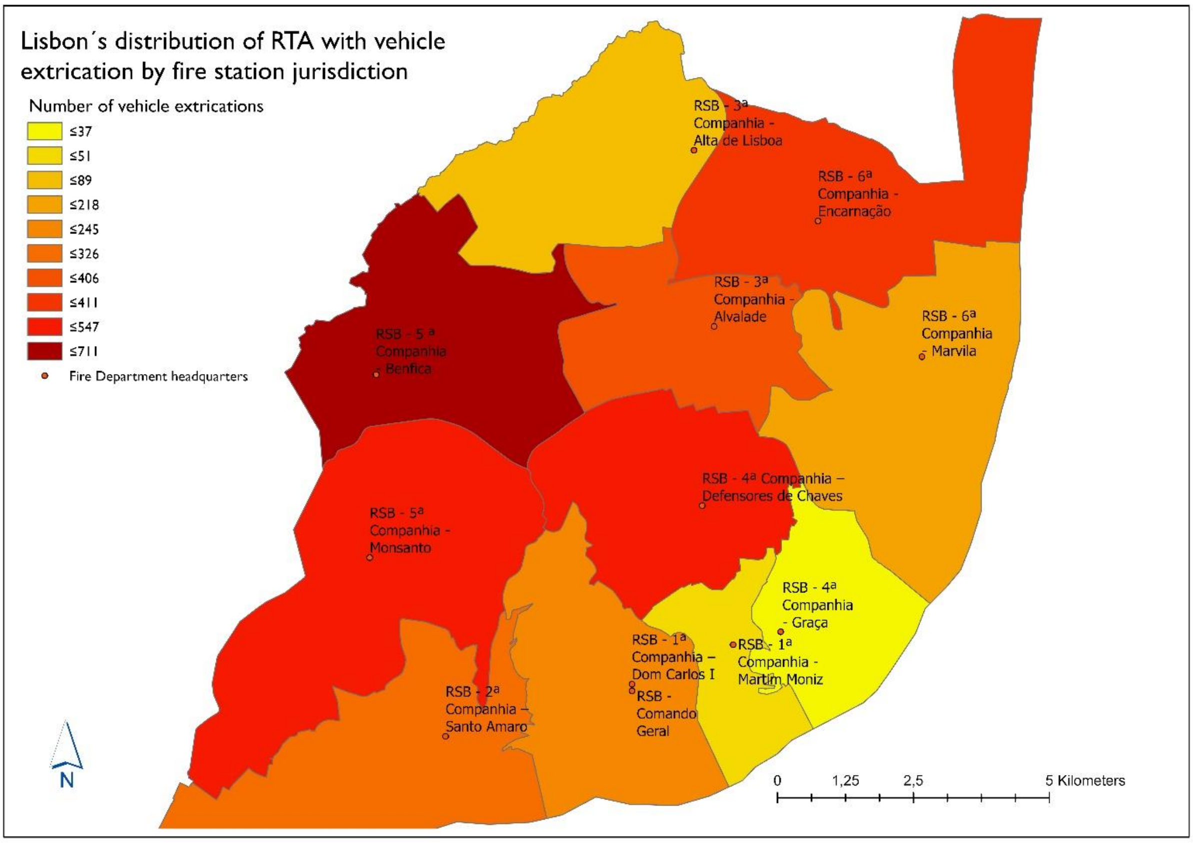

Image 45 - RTAs with extrication operations distribution by fire station jurisdiction

Regarding the class of road accidents that produce more victims - the vehicle-pedestrian collisions (image 46), we can see that the Defensores de Chaves headquarters is the one that rescue more victims due to its centrality and proximity to the streets and avenues that register more incidents, namely Avenida da República, Avenida Almirante Reis, Avenida Fontes Pereira de Melo, Rua da Escola Politécnica and Rua Melo Gouveia. Then we have the Headquarters of the 1st Company - Dom Carlos I with many occurrences on Avenida 24 de Julho and the Headquarters of the 5th Company, Benfica, with occurrences on Benfica Road, Avenida do Uruguai, Laranjeiras Road, and Avenida Lusíada.

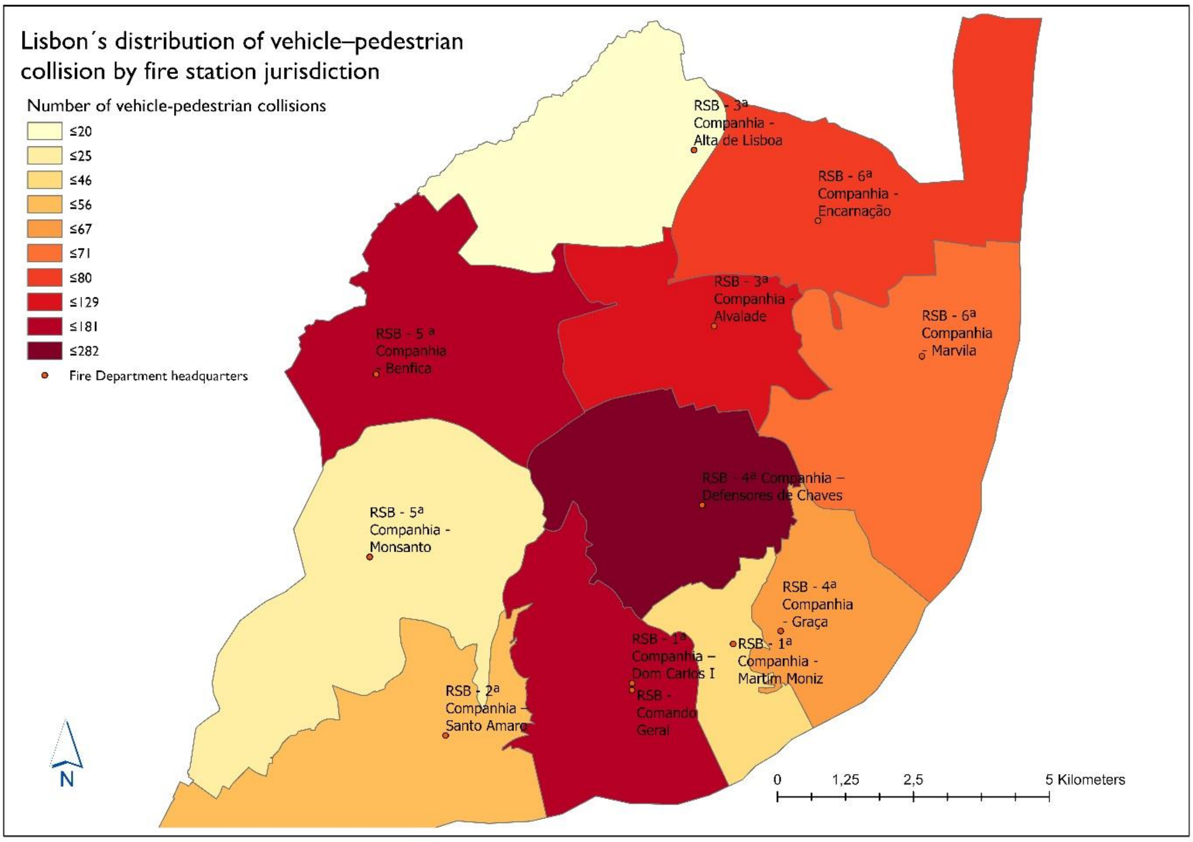

Image 46 - Vehicle-pedestrian collision distribution by fire station jurisdiction

5.2 Hotspot Analysis

Using a GIS-assisted technique (ArcGIS Pro) to identify accident hotspots within the Lisbon municipality area will aid in the identification of contributing causes and will assist the city council and its departments in implementing mitigation measures to make driving safe in these areas.

As stated in Chapter 3, Kernel Density Estimation (KDE) and Getis-Ord Gi* hotspot analysis were used to identify hotspots of road traffic accidents with victims in the Lisbon city area based on the location and characteristics of the road accident.

Using Kernel density analysis and Getis-Ord Gi* hotspots analysis, different densities and levels of confidence were determined for each year's data. The different maps produced by the Kernel density analysis and Getis-Ord Gi* hotspots analysis are displayed below for each biennium from 2010 to 2019, as weall for the decade. Several hotspot areas occurred in the study area consistently throughout the study period.

The data were checked for spatial autocorrelation using Global Moran's before performing KDE and Getis-Ord Gi* hotspot analysis to ensure that the data contained spatial clustering required for hotspot analysis. The ANSR definitions (5 victims, 200 meters) of a hotspot were aggregated and used to determine an appropriate search radius distance for KDE and Getis-Ord Gi*.

5.2.1 Biennial Analysis

In order to obtain the parameters required to proceed with the hotspot analysis, the data for the biennium were selected in ArcGis Pro 2.8 using a SQL Query that allowed us to select only the road accidents with victims for the years in review (2010-2011; 2012-2013; 2014-2015;2016-2017 and 2018-2019), after which the five biennia were subjected each one to the Moran I calculation to assess the existence of clusters of RTAs.

The inverse distance and the Euclidean distance method were introduced to conceptualize spatial relationships; no threshold was used. The results[16] obtained demonstrate a strong cluster Padron of accidents, with a positive Moran's Index of 0.99, a Z-score of 65, and a P-value of 0.0001, giving us less than 1% likelihood for the clustering pattern result of random chance. Thus, we can reject the null hypothesis.

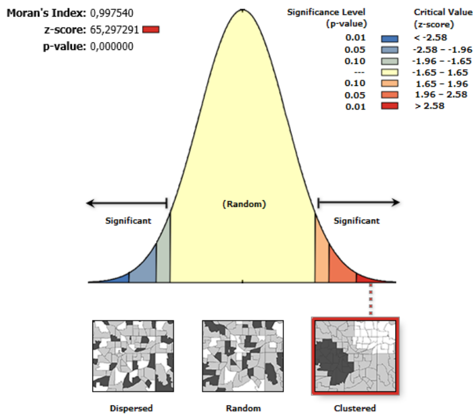

Image 47 - Spatial Autocorrelation Report Biennial analysis

[16] SInce the results of the five analisys give similar results - less than 1% likelihood for the clustering pattern result of random chance, we just discust in the text the first result (2010-2011).

The next step after identifying spatial clustering of road traffic accidents in 2010 and 2011 was to conduct a Kernel Density Estimation (KDE) in ArcGis 2.8 utilizing the tool spatial analyst tool. Tessa Anderson [19] identifies cell size and search radius (bandwidth) as the two most important elements influencing the KDE technique. According to several experts, the bandwidth is the most critical parameter for determining the best density surface [90]. As a result, the selection of bandwidth will have a considerable impact on the outcome of the hotspots. Specifically, the smaller the bandwidth, the smaller the hotspots.

The value of bandwidth will influence the smoothness of the density surface. The smoother the density surface, the larger the bandwidth. As a result, it is critical to select an ideal bandwidth.

To determine the desired bandwidth, we used Ito Fumiya's study [91], which suggests making incremental 50-meter jumps until the hotspot plot reaches the equilibrium, i.e., larger bands, hotspots cover a large portion of the study area, making it challenging to identify hotspot concentration points on specific roads correctly. In comparison, only a few hotspots are identified with smaller bands, resulting in a rough surface. In our case, we reached a balance in the bandwidth of 250 meters.

2010 – 2011

The Kernel Density Estimation (KDE) analysis presents for the biennium 2010-2011 (image 48) an accumulation of hotspots in Avenida de Berlin, Avenida Infante Dom Henrique, Avenida Marechal Craveiro Lopes, Avenida General Norton de Matos and Avenida Eusébio da Silva Ferreira.

The Eixo Norte-Sul, A5, IC19, Avenida da Ponte and Avenida de Ceuta also have concentrations of hotspots.

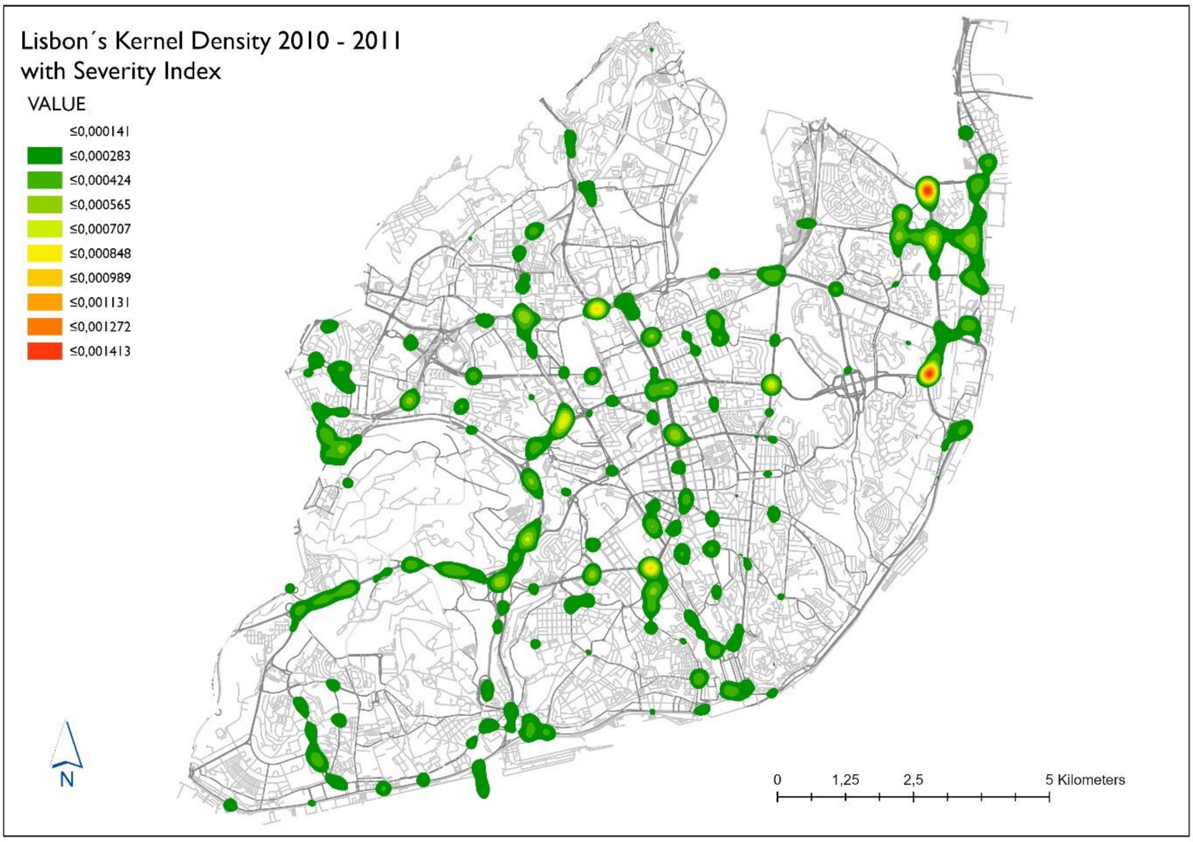

Image 48 - KDE analysis Biennial analysis 2010-2011

2012 – 2013

The 2012-2013 (image 49)period presents the first outlier of this analysis to the 2010-2019 decade, as it is possible to observe a marked decrease in the number of hotspots of accidents with victims. Not only are there fewer accidents, but traditionally prone areas are not active (with hotspots), as is the case of downtown Lisbon, the city's entry points and in the 1st and 2nd Level roads, with only a few small concentrations identified in the Eixo Norte-Sul, Avenida da Republica, Avenida Marechal Craveiro Lopes near Campo Grande, Avenida Marechal Gomes da Costa, Avenida de Berlim and Avenida Infante Dom Henrique. But what justifies this decrease? Situating ourselves in the year under consideration (2012) and analyzing the records contained in the INRIX report [84] [85], Portugal presented a drop in traffic hours of 50% in relation to 2011, and - 18% in 2013, all motivated by the economic crisis that began on 6 April 2011 with the official request for aid to the troika, which caused the colossal recession that we all experience, another detail that may justify this decrease is the unemployment rate that at the time was 17.5%, much of this experienced in the large capitals such as Lisbon and Porto.

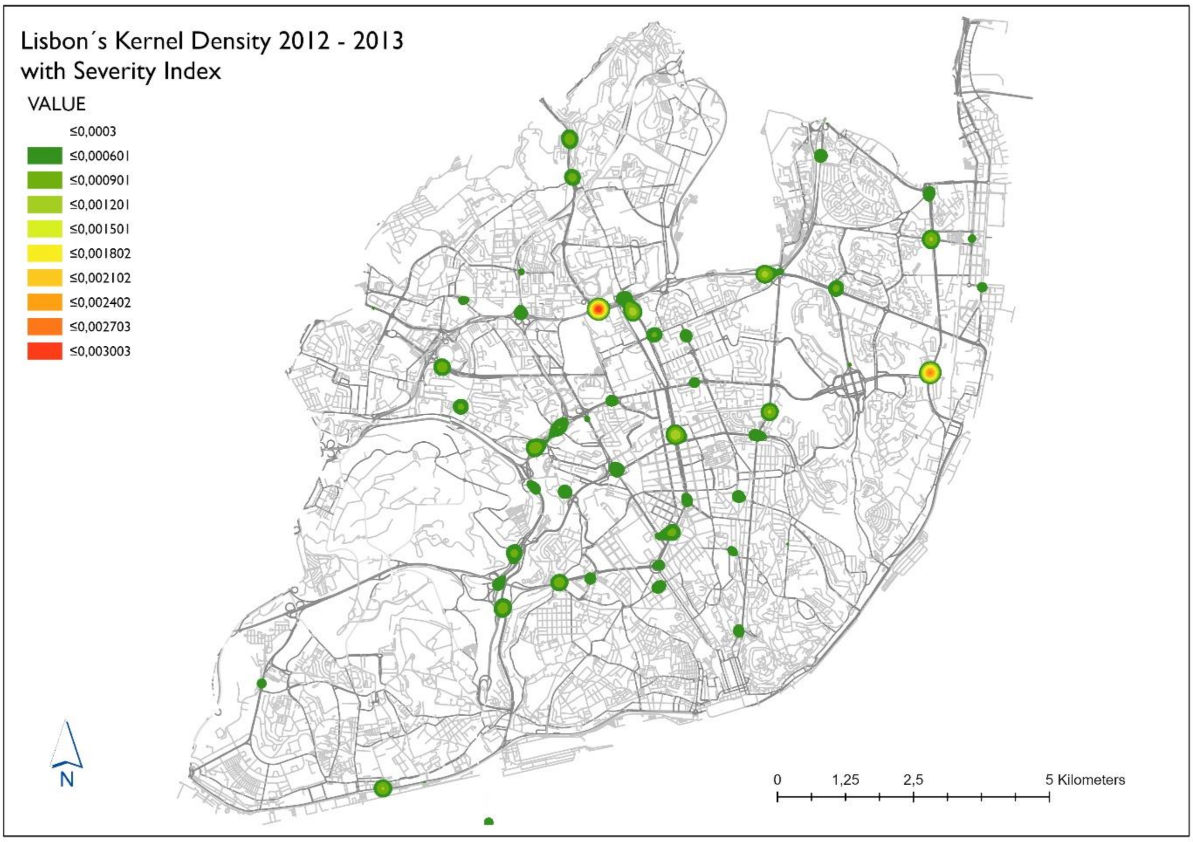

Image 49 - KDE analysis Biennial analysis 2011-2012

2014-2015

The years 2014 and 2015 (image 50) were years of economic recovery. In the second quarter of 2013, the economy emerged from a ten-quarter stretch of a technical recession and resumed growth. GDP rebounded by 2% in nine months, thanks partly to austerity relief brought about by the resumption of pensioners and government servant's allowances. The employment also began to respond favourably.

As observable data from this study proves this statement, the growth in the number of accidents with victims is consistent with the increase in the number of vehicles circulating in the city, observing the activation of hotspots in the city centre (Avenida da Republica, Praça Marques do Pombal, Avenida Fontes Pereira de Melo, Avenida da Liberdade, and downtown area), main access roads (Av. de Ceuta, Eixo Norte-Sul, Viaduct Eng. Eduardo Pacheco, IC19, A5, Avenida da Ponte, and Segunda Circular) and in the riverside area (Avenida Ribeira das Naus, Avenida da India, and 24 de Julho).

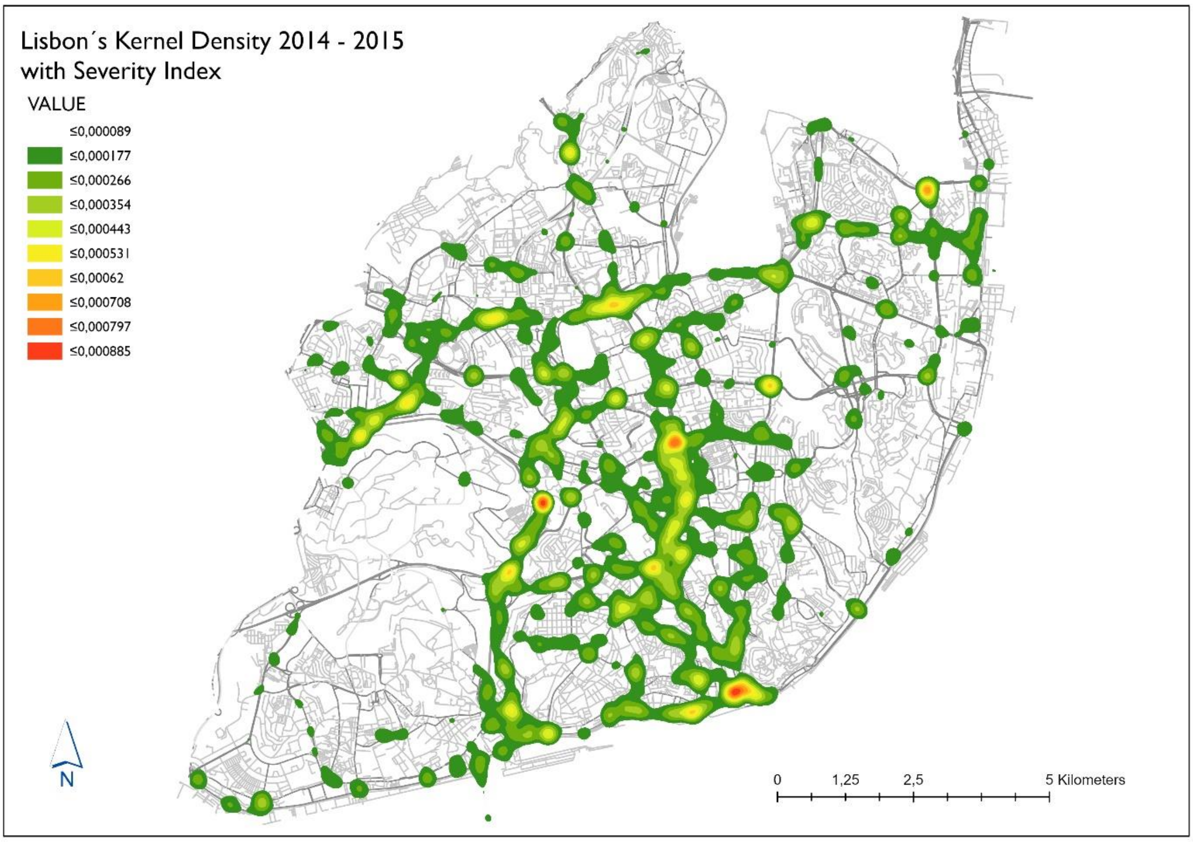

Image 50 - KDE analysis Biennial analysis 2014-2015

2016-2017

The 2016-2017 Biennium (image 51) was very similar to the 2014 and 2015 figures, with little variation in relation to the hotspots of traffic accidents with victims, and it is already possible to identify a pattern of location of the hotspots in the city.

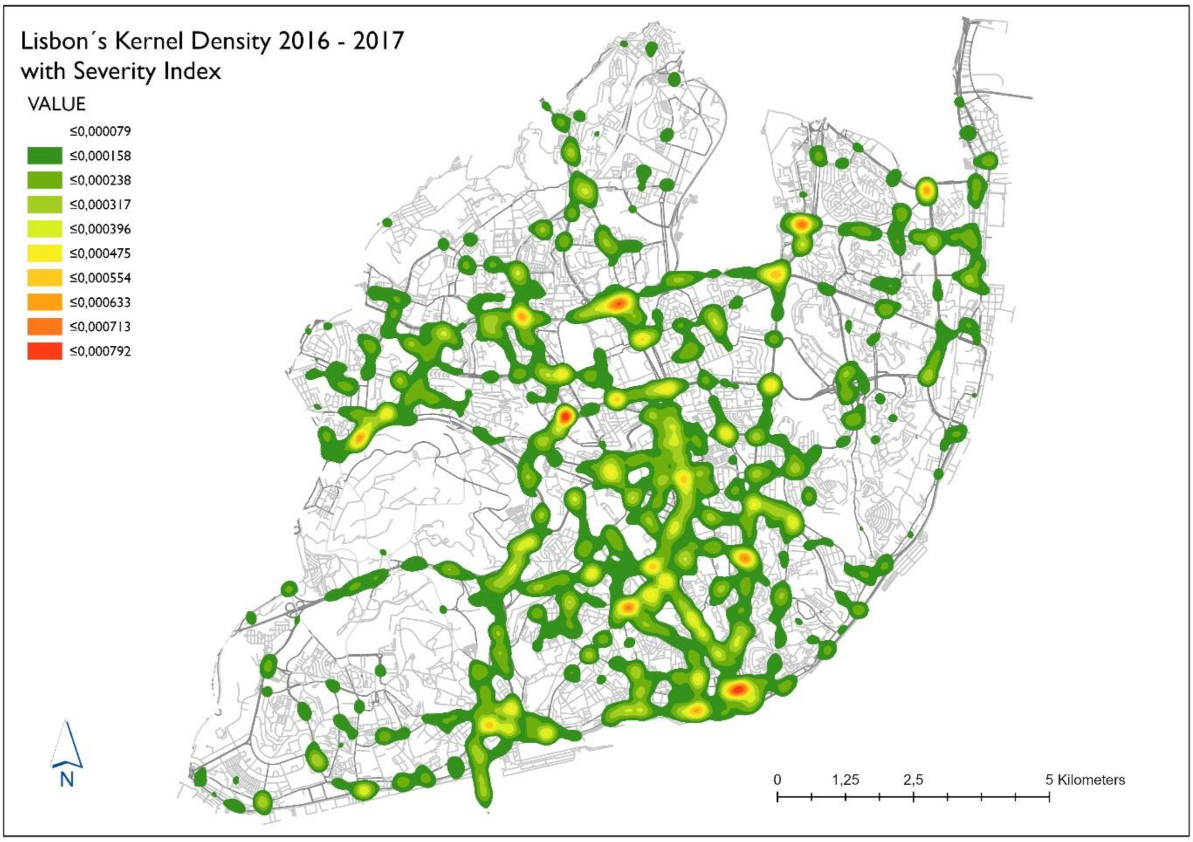

Image 51 - KDE analysis Biennial analysis 2016-2017

2018-2019

The 2018-2019 biennium (image 52) presents the second outlier in our data. There is a marked decrease in the number and location of hotspots, particularly in the city centre, the spots that were previously spread across the 3rd Level, 4th Level, and 5th Level roads are practically non-existent, If in 2012-2013 the economic crisis justified the reduction in traffic and consequently the reduction in accidents with victims, the current biennium cannot justify the same, Portugal is in total economic growth, according to the Tourism of Portugal [92] Lisbon is one of the European cities with the highest demand, registering numbers in the order of 7.5 million tourists only in 2018. To understand this phenomenon, we asked the Lisbon Municipal Civil Protection and the Lisbon Municipal Police what could justify this decrease? The conclusion we reached is both surprising and revealing. Once again, explaining the thesis defended in this dissertation that human factors are undoubtedly the most significant contributor to a road accident with victims.

In 2019, the Lisbon Municipal Police (PML), in addition to the 21 fixed speed cameras, started to control the speed with mobile speed cameras strategically placed in areas where the speed limit is 30 km/h (4th Level and 5th Level roads) such as residential areas or schools. The control of mobile speed control in the municipality allowed PML to record a record number of offenses, with a total of 61,540 severe infractions, ten times more than in 2015 (6842). This data and the study of the average speeds recorded by TomTom showed that drivers adjusted their speed to the limit imposed by law mainly on the roads mentioned above, resulting in a clear decrease in road accidents.

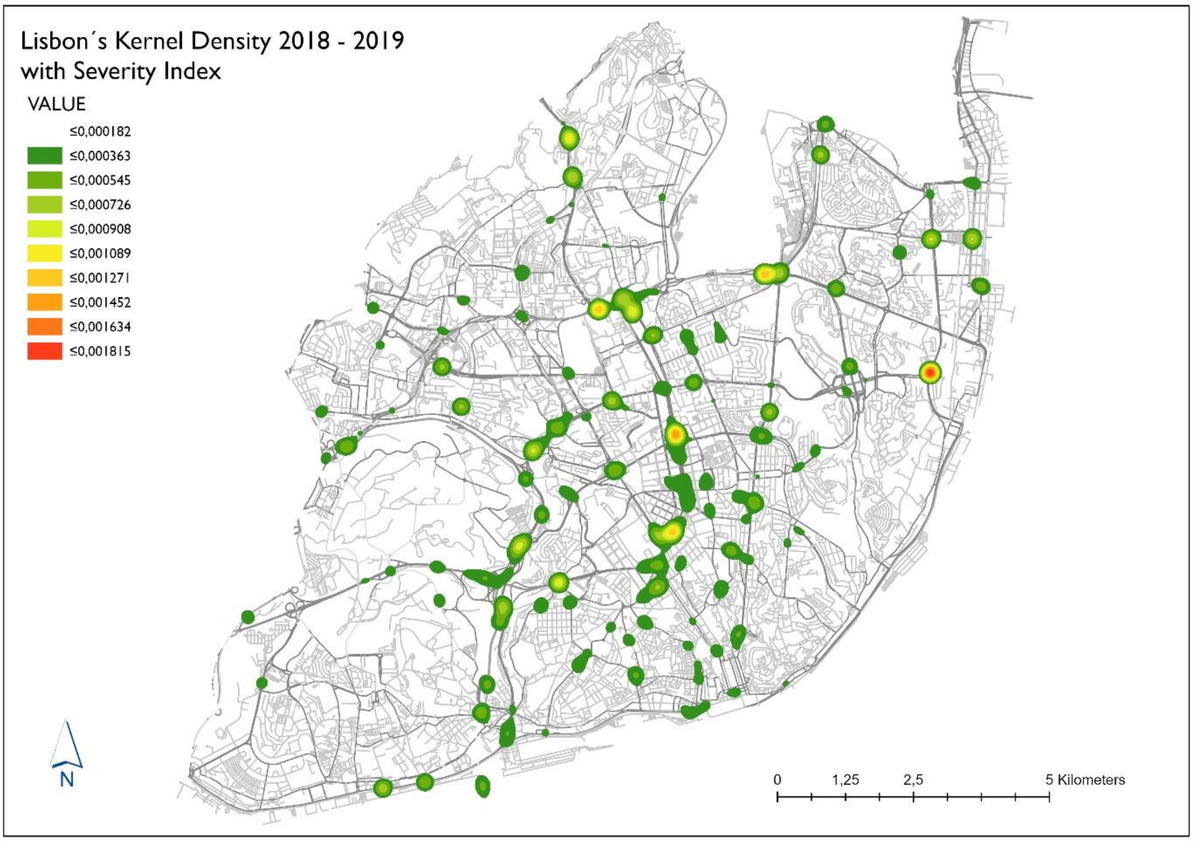

Image 52 - KDE analysis Biennial analysis 2018-2019

5.2.2 Seasonal Analysis

The temporal study of the kernel density for the seasons of the year of accidents with victims (image 53) allows us to identify that in spring, there is a lower concentration of hotspots in the outlying areas of the city in comparison with the other seasons of the year, being observable that it is the city centre, as well as downtown Pombal that concentrates the accidents, an example of this being Avenida da Republica, Avenida Fontes Pereira de Melo, and Avenida Liberdade. In summer, there is an increase in the incidence and frequency of incidents in the central area of the city, in the area of the Eixo Norte-Sul, accesses to the 25 de Abril, General Norton de Matos Avenue with the Benfica Road and at the end of the A5 towards the Duarte Pacheco viaduct at the exit to the Bridge. The aforementioned roads continue to be hotspots throughout the autumn and winter. However, there are a greater accident concentration on the outskirts of Lisbon, specifically along with Avenida General Norton de Matos and the Eixo Norte-Sul, with the highest concentration of accidents in the area that overlaps Avenida Engenheiro Duarte Pacheco.

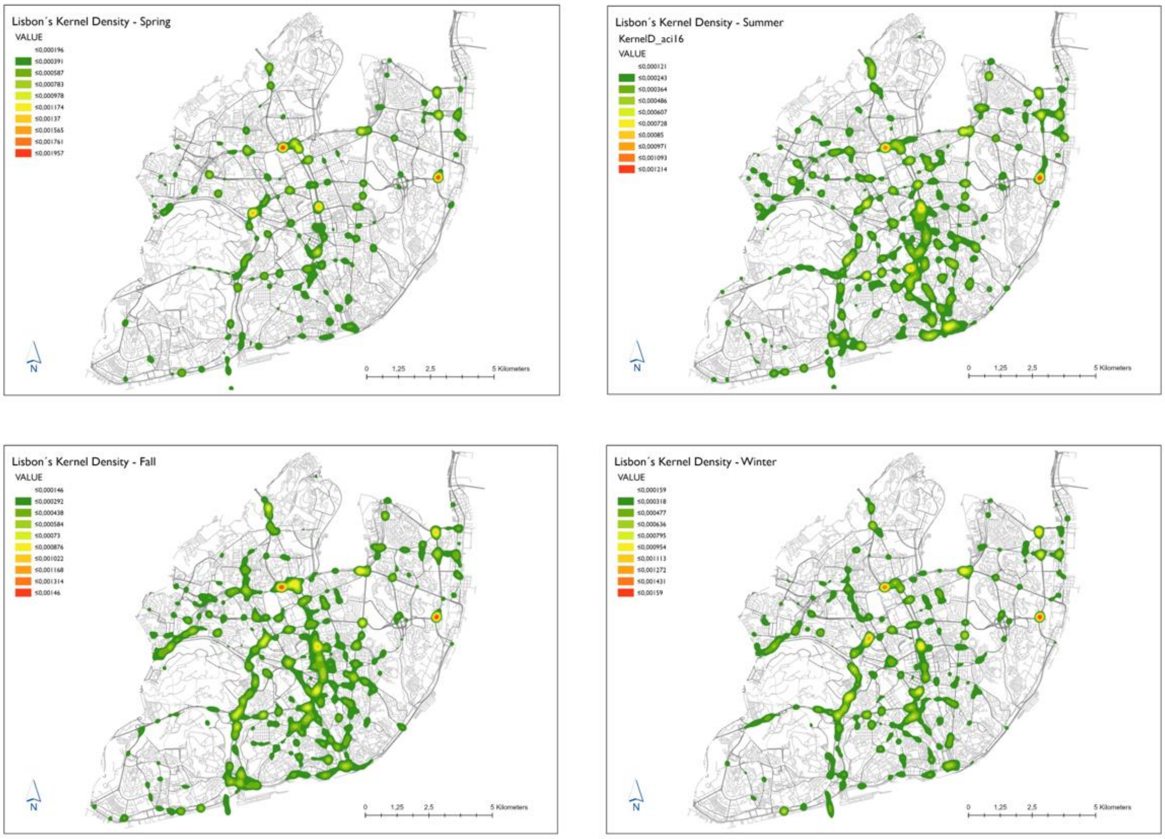

Image 53 - KDE analysis seasonal analysis 2010-2019

5.2.3 10-Year Analysis

In order to be able to perform an assessment and identification of the hotspots of traffic accidents with victims existing in the area of the municipality of Lisbon that reflects the real problems without suffering the influences or deviations that may exist in spatio-temporal analyses of lower temporal amplitude, we performed, as in the biannual analyses, the KDE analysis of all road accidents with victims that fulfilled the ANSR hotspot cluster requirement (5 victims, 200 meters). The results obtained demonstrate a strong cluster pattern of accidents, with a positive Moran's Index (image 54) of 0.99, a Z-score of 200, and a P-value of 0.0001, giving us less than 1% likelihood for the clustering pattern result of random chance. Thus, we can reject the null hypothesis.

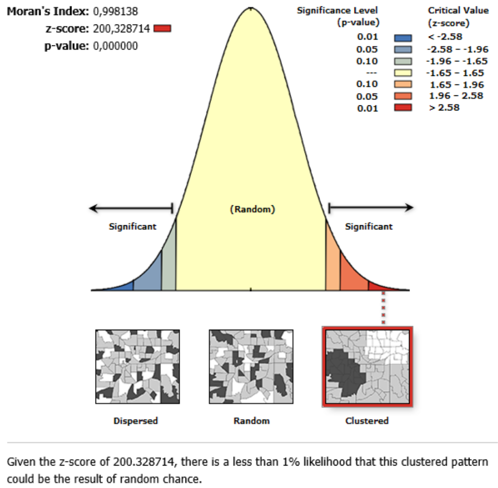

Image 54 - Spatial Autocorrelation Report - Decade analysis

As the hotspots locations (image 55) will be subject to confirmation of statistical significance through Hotspot Analysis (Getis-Ord Gi*) we will only identify the effective Lisbon hotspots locations in the next subchapter.

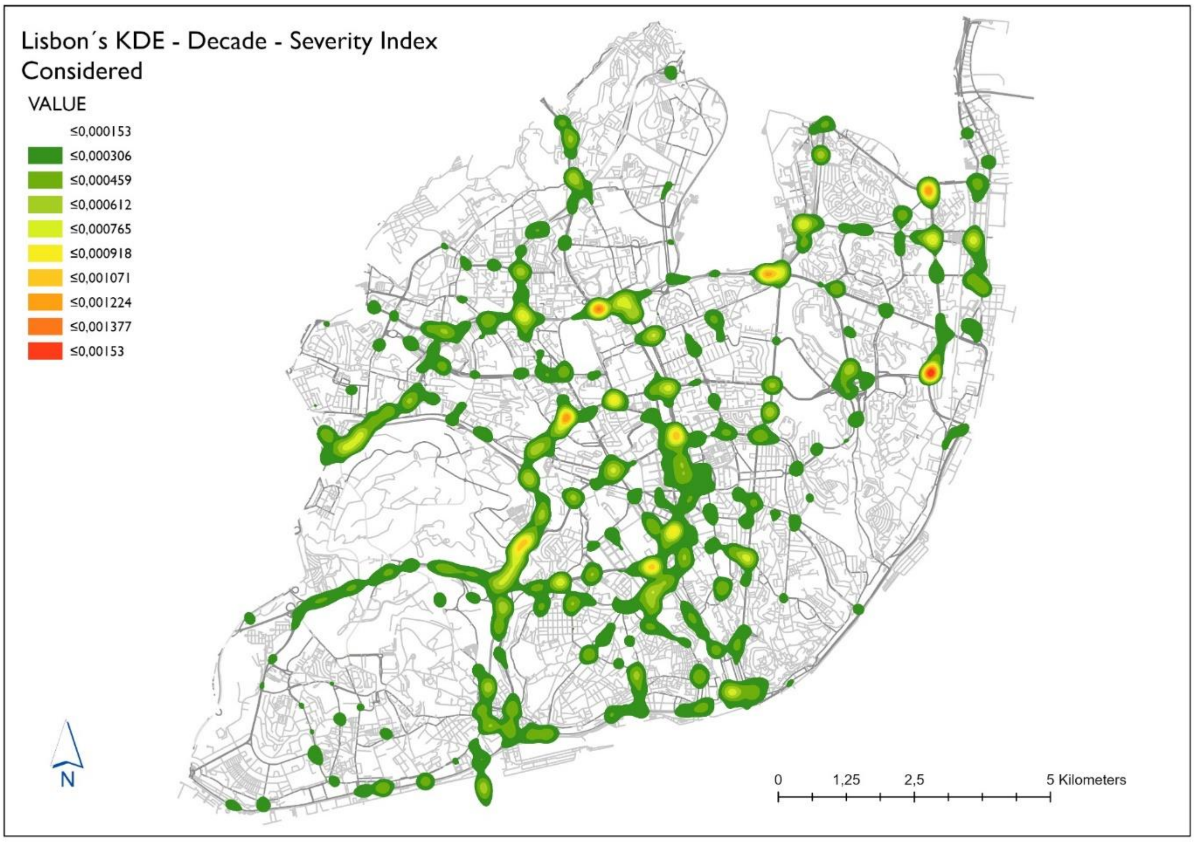

Image 55 - KDE analysis decade analysis 2010-2019

5.2.4 Getis Ord Gi* Analisys

The previous analysis (KDE) permitted the identification of the RTAs hotspots. However, as previously stated, the KDE technique lacks an evaluation of the statistical significance of the discovered hotspots. As a result, it's critical to look into the statistical significance of the hotspot and identify the most prone regions with statistical significance. According to many researchers, it is preferable if the statistical value of the discovered hotspot is determined objectively and proactively [51]. Our statistically significant evaluation procedure began with the null hypothesis: "accidents in a roadway segment occurred at random." As previously stated, we initially used spatial autocorrelation global Moran's I test to verify the statistical testing for the null hypothesis and determine whether RTAs points were organized in clusters of identical values. Accordingly to Prasannakumar [54], this is a required step before performing the Hotspot Analysis (Getis-Ord Gi*) on ArcGIS PRO.

In this way and for the identification of hotspots of road accidents with victims in the municipality of Lisbon, the Hotspot Analysis Getis-Ord Gi* (image 56) was applied to our database, presenting a clustering of high values (hot spots) with a confidence level of 99% for the following locations: Avenida Dom João II, Eixo Norte-Sul, Avenida General Norton de Matos, Avenida Marechal Craveiro Lopes, Avenida Eusébio da Silva Ferreira, Itinerário Complementar 17, Estrada de Monsanto, A5, Acesso Avenida da Ponte, Ponte 25 de Abril, Avenida Vinte e Quatro de Julho, Campo dos Mártires da Pátria, Rua Joaquim António de Aguiar, Tunel Marquês.

At a 95% confidence interval were identified as statistically hotspots: Avenida Fontes Pereira de Melo, Rua Sousa Lopes, Avenida da República, Avenida Álvaro Pais, Avenida General Roçadas, Avenida Mouzinho de Albuquerque, Avenida de Brasília, Avenida Infante Dom Henrique e Avenida da India.

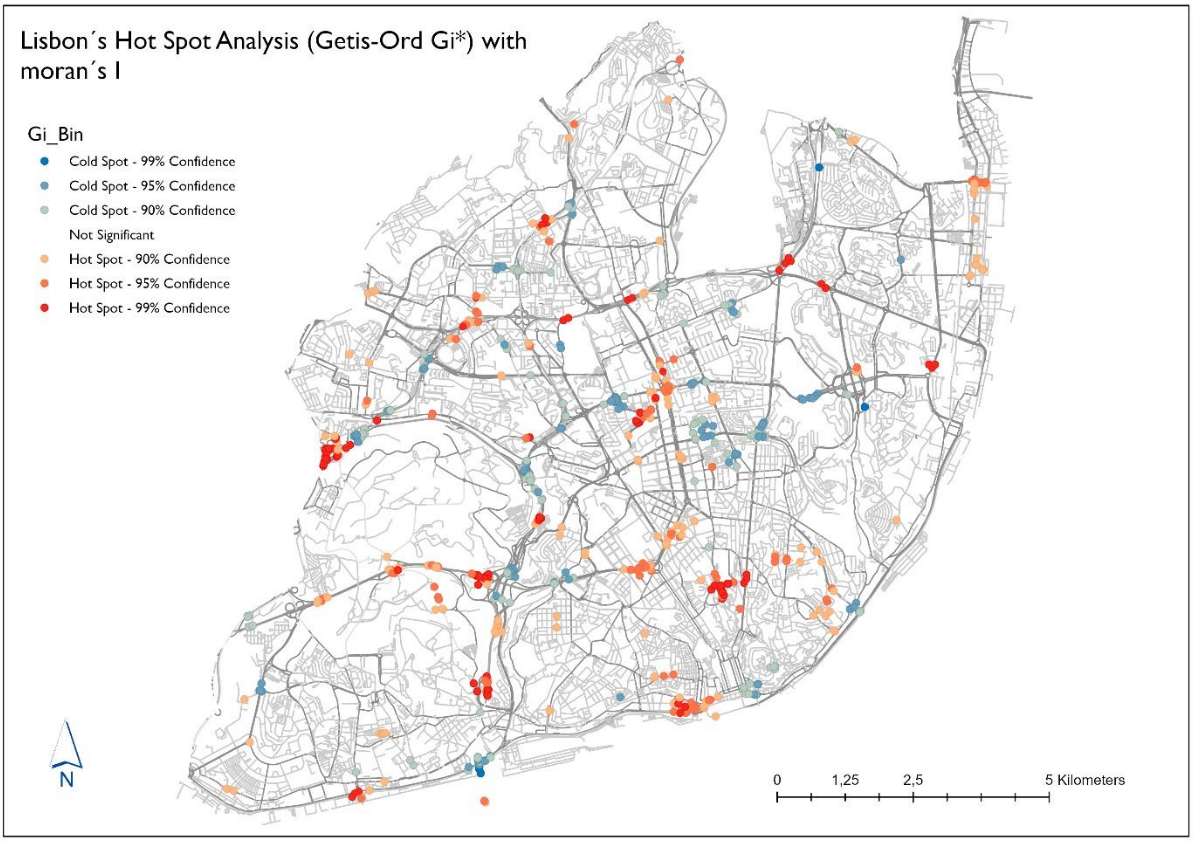

Image 56 - Hotspot analysis (Getis-Ord Gi) decade analysis 2010-2019*

5.2.5 Hotspot Identification and Characterization

After the identification of the location of statistically significant hotspots performed by the Hotspot Analysis Getis-Ord Gi* algorithm, the results were analyzed, and a selection of the most active hotspots was arranged by the number of accidents and severity of victims was made, as well the identification of the problematic roads (image 57). If Lisbon Municipality had already applied the Municipal Road Safety Index (ISRM) we could compare this classification with the one calculated by CML.

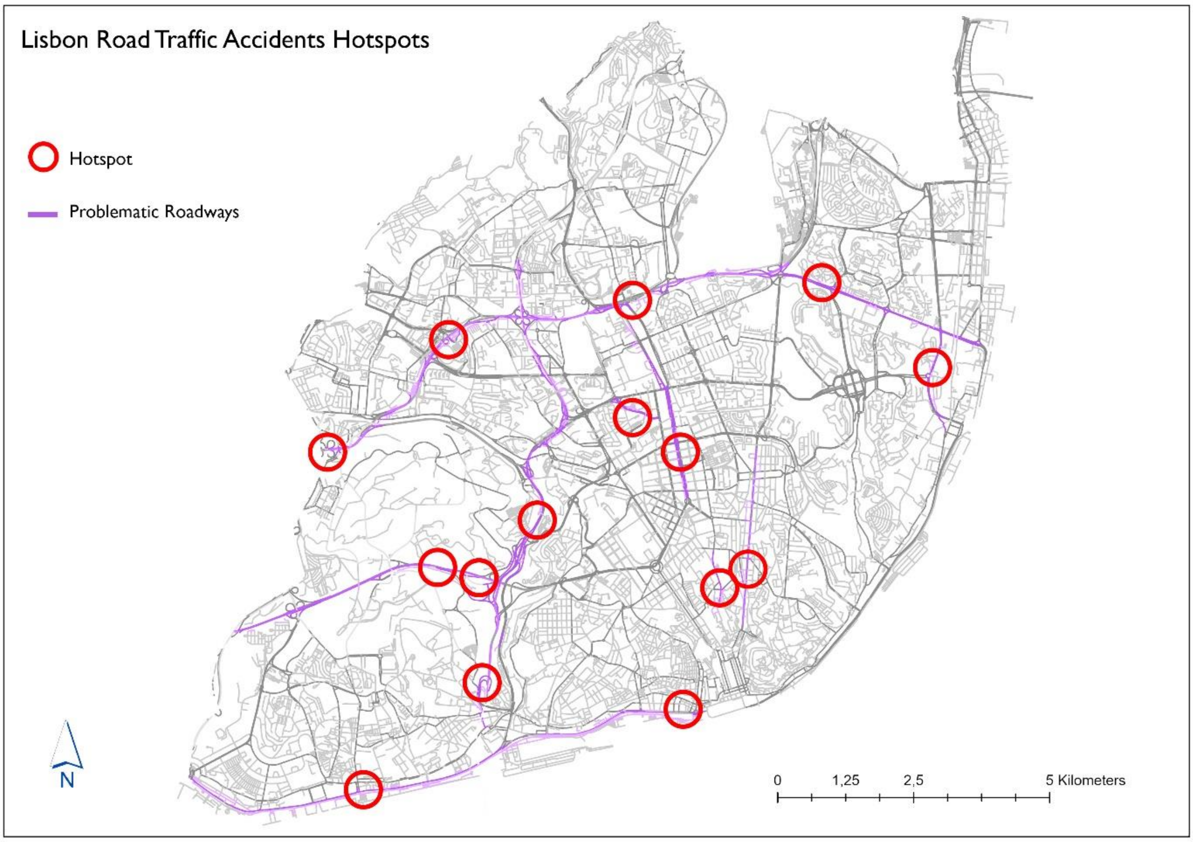

Image 57 - Lisbon Road traffic Accidents Hotspots and problematic roadways identification

Segunda Circular

The data shows that the Segunda Circular is the most problematic roadway in this study.

It comprises three avenues (Avenida General Norton de Matos, Avenida Marechal Craveiro Lopes, and Avenida Eusébio da Silva Ferreira, the late only exist from 2015, before that there were just two avenues), is 10 kilometers long and is the primary distributor of internal traffic in the city of Lisbon.

To better understand the analysis, the Segunda Circular was segmented by its constituent avenues, and the Hotspots were identified and characterized accordingly.

Av. Eusébio de Silva Ferreira

This section of the Segunda Circular between IC 19 and Estrada da Luz has 3,7 Km in extension. It is a 2nd Level Roadway (Main Distribution Network). The speed limit is 80Km/h and registered one fatality, one serious injury, and 156 minor injuries during the time frame of this study. Regarding pedestrian collisions, this stretch recorded 10 (ten) Pedestrian collisions and 199 road traffic accidents.

The identified hotspot (image 58) is located westbound at the final stretch of Av. Eusébio de Silva Ferreira (Segunda Circular) between the exit to benfica / buraca, just after the Escola Superior de Comunicação and the access to the IC19 / IC17 and A5.

The most common accident typology on the hotspot is skidding, followed by rear-end collision with another moving vehicle, and Crash with a rollover.

The visible rise in accidents (skidding) in this hotspot when there is an episode of precipitation is manifest. 41% of the records analyzed for this location occurred on rainy days. This data becomes especially significant when we consider that in Lisbon, during our analysis, there were on average only 100 days of precipitation per year and that these conditions (rain) only correspond to 17% of the RTAs cases in our study.

The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by disrespect for the safety distances.

Regarding the time distribution, we observe that most accidents occur at night, followed by lunchtime rush and morning rush in good weather conditions. With precipitation conditions, the most frequent period is during the morning rush and dawn.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly the human factors that, combined with atmospheric conditions, potentiate the occurrence of RTA in this location. As a suggestion and extending the concept to the rest of the Segunda Circular, we believe that the reduction of speed from 80km/h to 50 km/h combined with the existence of average speed cameras (located near the airport, Campo Grande viaduct, and after the exit to Benfica Buraca) would drastically reduce the number of accidents on this roadway.

Image 58 - Av. Eusébio de Silva Ferreira Hotspot localization

Av. General Norton de Matos

This section of the Segunda Circular between Estrada da Luz and Campo Grande viaduct has 1,4 Km in extension. It is a 2nd Level Roadway (Main Distribution Network). The speed limit is 80Km/h and registered five fatalities, 15 serious injuries, and 338 minor injuries during the time frame of this study. Regarding pedestrian collisions, this stretch recorded 12 Pedestrian collisions and 673 road traffic accidents (in fact, from the three avenues that compose the Segunda Circular, this stretch is the one with most cases of crashes, pedestrian collisions, and rear-end collisions).

The identified hotspot (image 59) is located just after the campo grande viaduct, near José de Alvalade Stadium.

The most common accident typology on the hotspot is rear-ended collision with another moving vehicle.

In terms of meteorológical influence, 74% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by disrespect for the safety distances.

Regarding the time distribution, we observe that most accidents occur during the morning rush, followed by the afternoon and afternoon rush.

The majority of the Pedestrian collisions occurred at night, dawn, and morning rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 59 - Av. General Norton de Matos Hotspot localization

Av. Marechal Craveiro Lopes

This section of the Segunda Circular between Campo Grande Viaduct and Prior Velho has 4,3Km in extension. It is a 2nd Level Roadway (Main Distribution Network). The speed limit is 80Km/h and registered seven fatalities, ten serious injuries, and 261 minor injuries during the time frame of this study. Regarding pedestrian collisions, this stretch recorded 2 Pedestrian collisions and 352 road traffic accidents.

The identified hotspot (image 60) is located just after the viaduct above the rotunda do relogio.

The most common accident typology on the hotspot is rear-ended collision with another moving vehicle, multiple vehicle collision, and Crash with a rollover.

In terms of meteorológical influence, 69% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by disrespect for the safety distances.

Regarding the time distribution, we observe that most accidents occur during the afternoon rush, followed by the afternoon and morning rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

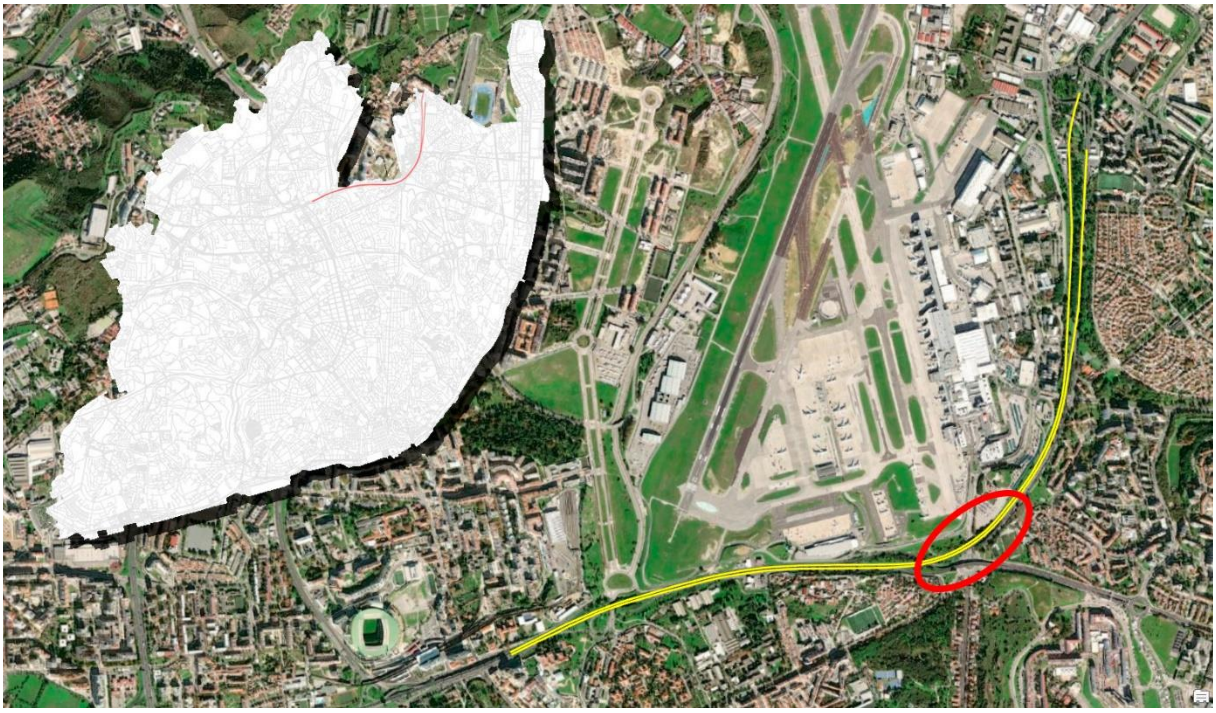

Image 60 - Av. Marechal Craveiro Lopes Hotspot localization

Avenida de Berlim

The Avenida de Berlim iniciates in the crossroads with Avenida Dom João II and Avenida do Pacífico and finishes near the airport has 3Km in extension. It is a 3rd Level Roadway (Secondary Distribution Network). The speed limit is 50Km/h and registered two severe injuries and 144 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 60 Pedestrian collisions and 302 road traffic accidents.

The identified hotspot (image 61) is located in the junction with Avenida Cidade do Porto.

The most common accident typology on the hotspot is a Pedestrian collision and a side collision with another moving vehicle.

In terms of meteorológical influence, 75% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by the disregard for vertical signs and disregarding for traffic lights signals.

Regarding the time distribution, we observe that most accidents occur during the afternoon rush, followed by the morning rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 61 - Av. de Berlim Hotspot localization

Avenida Almirante Reis

The Avenida Almirante Reis initiates in Palma street and finishes near the Francisco sá Carneiro square. It is a 3rd Level Roadway (Secondary Distribution Network). The speed limit is 50Km/h and registered six severe injuries and 92 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 104 pedestrian collisions and 275 road traffic accidents.

The identified hotspot (image 62) is located at the crossroads with Angola street.

The most common accident typology on the hotspot is a Pedestrian collision and a side collision with another moving vehicle.

In terms of meteorológical influence, 87% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by the disregard for vertical signs and disregarding for traffic lights signals.

Regarding the time distribution, we observe that most accidents occur during the night, followed by the afternoon rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 62 - Av. Almirante Reis Hotspot localization

Avenida da Republica

The Avenida da Republica initiates in Campo Grande and finishes near the Duque de Saldanha square. It is a 3rd Level Roadway (Secondary Distribution Network). The speed limit is 50Km/h, and as 1.5 Km in length, and registered two deaths, six severe injuries, and 139 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 113 pedestrian collisions and 392 road traffic accidents.

The identified hotspot (image 63) is located at the crossroads of Avenida Elias Garcia.

The most common accident typology on the hotspot is a Pedestrian collision and a side collision with another moving vehicle.

In terms of meteorological influence, 84% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by the disregard for vertical signs.

Regarding the time distribution, we observe that most accidents occur during the afternoon rush, followed by the morning rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 63 - Av. da República Hotspot localization

Eixo Norte Sul

Eixo Norte-Sul (IP7) is a highway that connects the north and the south of the city, initiates in the bridge 25 de Abril, and finishes near Tunnel of Grilo (between the connections with A8, A1, and A12). It is a 1st Level Roadway (Structuring Network). The speed limit is 90Km/h, but in some stretches is 70Km/h has 9.5 Km in length and registered three deaths, ten severe injuries, and 602 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway does not record any pedestrian collisions and registers 1013 road traffic accidents.

The identified hotspots (image 64) are located in 3 different places through the IP7: The first one is located near Campolide in the access to the IP7 from the radial de Benfica. The second one is in Duarte Pacheco viaduct, and the third one is near Lumiar.

Image 64 - Eixo Norte-Sul Hotspot localization

The most common accident typology is the crashes whit a retaining device and rear-ends collisions with another vehicle.

In terms of meteorológical influence, only 57% of the RTAs occurred in good weather conditions— the great majority of RTAs occurring in precipitation conditions (< 1.2mm/h) were located in the second hotspot, the area of Duarte Pacheco viaduct, just after the exit to Belem. Analyzing the flood vulnerability map (image 66), we can perceive why there are so many accidents in this area.

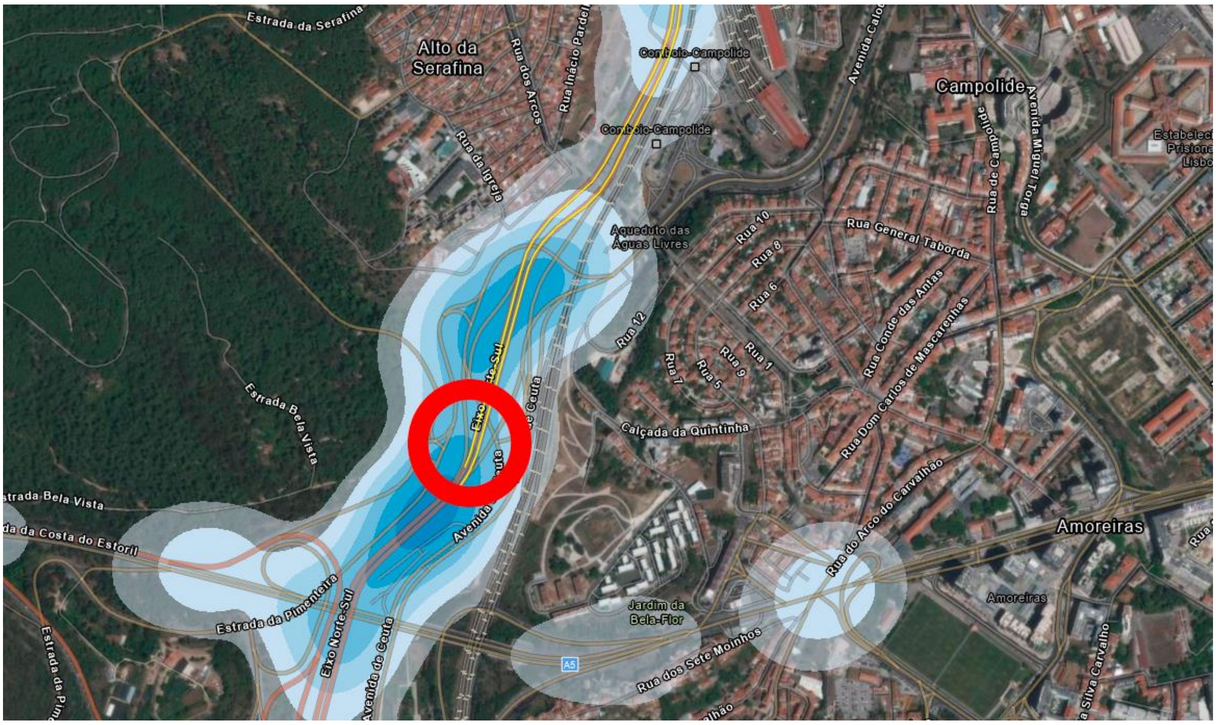

Image 65 - Hotspot located in a flood-prone area

The most common factor for the accidents on the three locations is the excessive speed for the existing conditions followed by the failure to observe safety distances.

Regarding the time distribution, we observe that most accidents occur during the morning rush, followed by the afternoon.

Given the above, we believe that the leading causes for the existence of these hotspots, in particular, the second one are a mix of human factors due to the speeding and environmental ones (roadway with poor drainage and weather conditions).

Avenida 24 de Julho

The Avenida 24 de Julho initiates in Cais do Sodré Train station and finishes near the Museu do Oriente. It is a 4th Level Local Distribution Network (proximity network). The speed limit is 50Km/h, as 3 Km in length, and registered three deaths, three severe injuries, and 109 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 165 pedestrian collisions and 363 road traffic accidents.

The identified hotspots (image 66) are located at the eastern part of the roadway near Cais do sodré train station, the second one in the junction near Jardim de Santos.

The most common accident typology on the hotspot is a Pedestrian collision and a collision with other situations.

In terms of meteorológical influence, 89% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by irregular maneuvering.

Regarding the time distribution, we observe that most accidents (including the Pedestrian collisions) occur during the afternoon rush, followed by dawn.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 66 - Av. 24 de Julho Hotspot localization

Avenida das Indias

The Avenida das Indias initiates near the Museu do Oriente and finishes near Alges. It is a 4th Level Local Distribution Network (proximity network). The speed limit is 50Km/h, as 5 Km in length, and registered eight deaths, twenty-seven severe injuries, and 126 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 21 pedestrian collisions and 267 road traffic accidents.

The identified hotspot (image 67) is located near Museu dos Coches.

The most common accident typology on the hotspot is rear-ended collision and Collision with other situations.

In terms of meteorológical influence, 85% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by disregarding traffic light signals.

Regarding the time distribution, we observe that most accidents occur during the morning rush, followed by the afternoon rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 67 - Av. da Índia Hotspot localization

Ponte 25 de Abril

The access to the 25 Abril bridge is located near Alvito. It is a 2nd Level Roadway (Main Distribution Network). The speed limit is 80Km/h, as 0.5 Km in length, and registered one death, three severe injuries, and 107 minor injuries during the time frame of this study in a total of 234 road traffic accidents.

The identified hotspot (image 68) is located near Alvito.

The most common accident typology on the hotspot is Crash with side restraint in place and Collision with other situations.

In terms of meteorológical influence, 71% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by irregular maneuver.

Regarding the time distribution, we observe that most accidents occur during the afternoon rush, followed by the morning rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

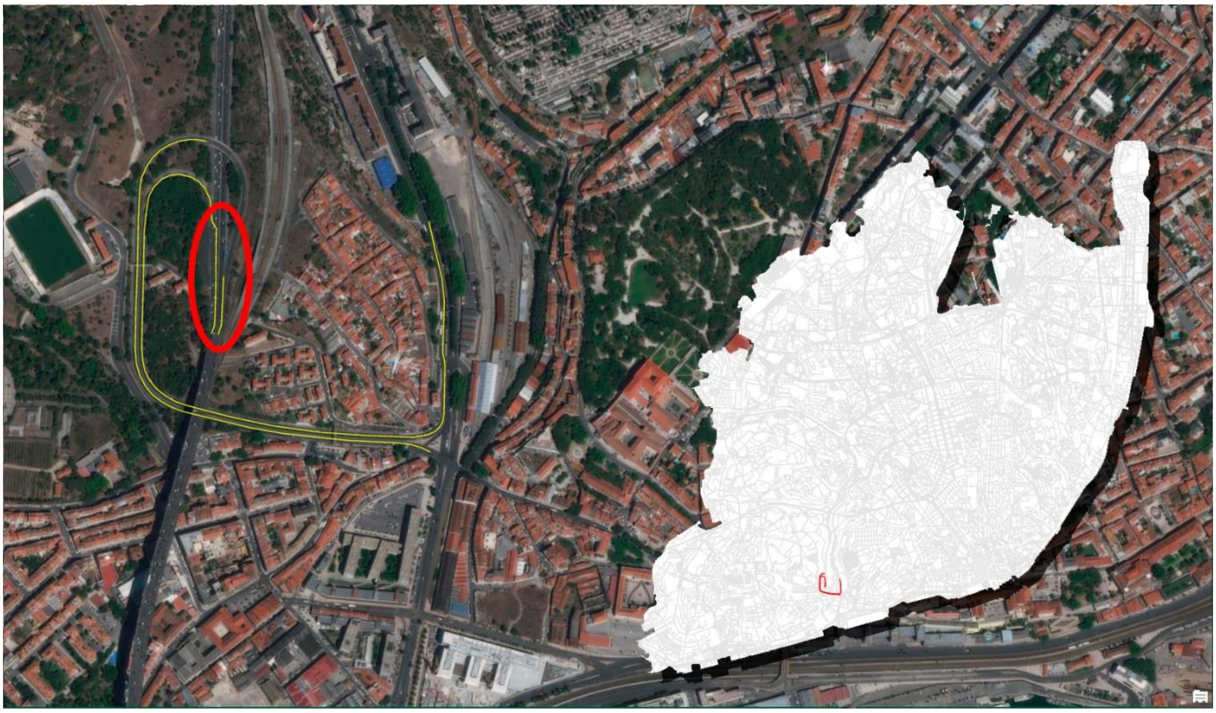

Image 68 - 25 de Abril Bridge Access Hotspot localization

Avenida Infante Dom Henrique

The Avenida Infante Dom Henrique initiates in José Queiros square and finishes near the Praça do Comercio square. It is a 2nd Level Roadway (Main Distribution Network). The speed limit is 50Km/h, with 12Km in length is the biggest avenue of Lisbon and registered seven deaths, 73 severe injuries, and 812 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway recorded 75 pedestrian collisions and 642 road traffic accidents.

The identified hotspot (image 69) is located at the crossroads of Avenida de Berlim.

The most common accident typology on the hotspot is a side collision with another moving vehicle.

In terms of meteorológical influence, 81% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by the disregard for traffic light signals.

Regarding the time distribution, we observe that most accidents occur during the morning rush, followed by the afternoon rush.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

Image 69 - Av. Infante Dom Henrique Hotspot localization

A5

Highway A5 (A5) is a highway that connects Cascais to Lisbon, initiates in Duarte Pacheco Viaduct, and finishes near Cascais. It is a 1st Level Roadway (Structuring Network). The speed limit is 120Km/h, has 25 Km in length, seven of which are inside the Lisbon Municipality, and registered six deaths, ten severe injuries, and 426 minor injuries during the time frame of this study. Regarding pedestrian collisions, this roadway records two pedestrian collisions and registers 446 road traffic accidents.

The identified hotspots (image 70) are located in 3 different places through the A5: The first one is located near Duarte Pacheco viaduct. The second one is at the exit to A2, and the third one is near the Monsanto exit.

The most common accident typology on the hotspots is rear-ended collision and simple crash.

In terms of meteorológical influence, 74% of the RTAs occurred with good weather. The most common factor for the accidents in this location is the excessive speed for the existing conditions followed by irregular maneuvering.

Regarding the time distribution, we observe that most accidents occur during the afternoon rush, followed by night.

Given the above, we believe that the leading causes for the existence of this hotspot are undoubtedly human factors.

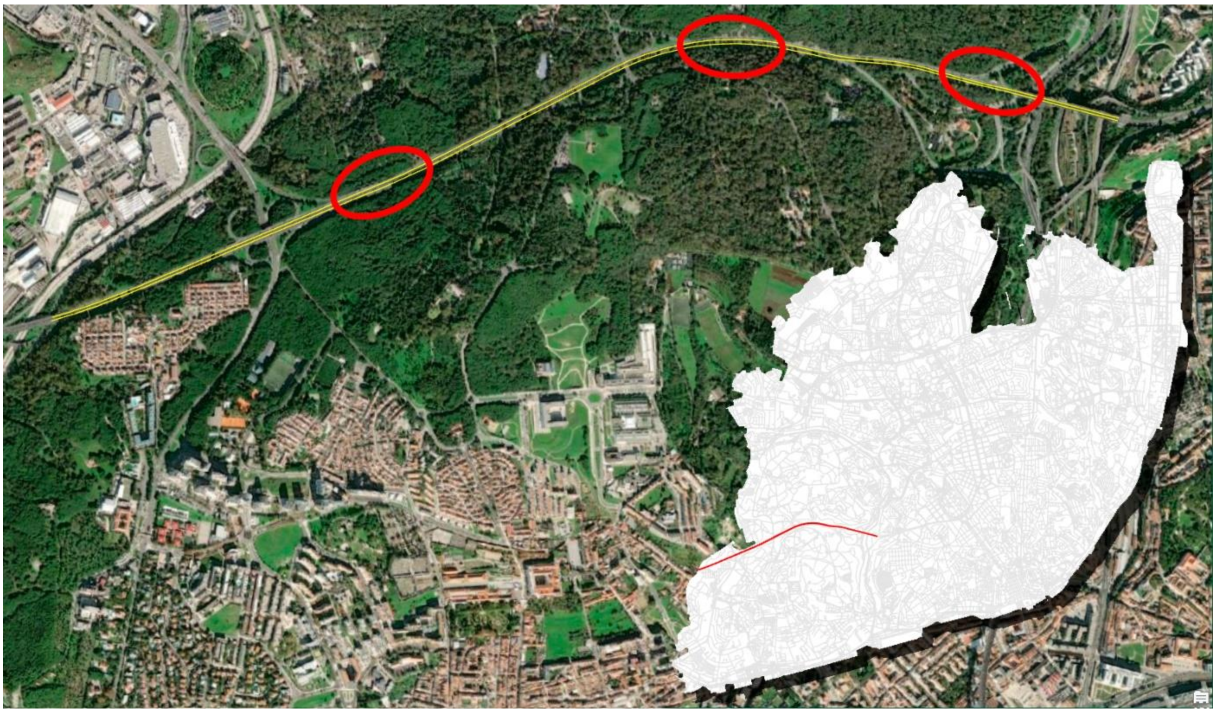

Image 70 - A5 Hotspot localization

5.3 The Validation of the Results by Experts

Although expert validation is subject to biases, it is nevertheless a valid method of verification. Given that content validity is critical to ensuring an assessment's overall validity, a systematic approach to content validation should be taken based on evidence and best practices, in that way for the confirmation of the artifacts created in this dissertation. The approach adopted was adapted from the framework proposed by Muhamad Yusoff [93] to calculate the Content Validity Ratio (CVR) and Content Validity Index (CVI) of the expert's opinion.

CVI: The Item-CVI method can be used to calculate CVI, which is the most extensively documented strategy for content validity in instrument development (I-CVI) [94].

I-CVI is calculated by dividing the number of experts who gave each item a rating of "extremely relevant" by the total number of experts. Values vary from 0 to 1; if I-CVI is greater than 0.79, the item is important; between 0.70 and 0.79, the item requires modifications; and less than 0.70, the item is not approved. [94].

I-CVI: has been determined as the number of “very relevant” (or 3-4) expert ratings for each question divided by the total number of experts on the panel.

(6)

$$I - CVI = \frac{Number\ of\ raters\ giving\ a\ rating\ of\ 3\ our\ 4}{Total\ number\ of\ raters}$$

CVR: was calculated for each question's importance, with a higher value (between 1 and -1) indicating greater agreement among expert panel members.

(7)

$$CVR = \frac{\left[n - \left(\frac{N}{2}\right)\right]}{\frac{N}{2}}$$

Where n=number of experts who indicated “essential” for a question, and N=total number of raters in the calculation.

5.3.1 Expert panel

Experts in the expert panel were defined as individuals who worked in civil protection, GIS experts, and Road Safety Experts.

Table 9 - Characteristics of the expert panel

Characteristics of the expert panel, N = 4	
Gender	2 Female, 2 Males
Academic Disciplines (some with more than one degree)	1 Geography and Science, 1 Civil Protection, 1 Education; 1 Science
Position	1 National Focal Point of Copernicus Emergency Management Service – GIS specialist - ANEPC; 1 Deputy National Operations – ANEPC; 1 Road Safety Observatory Division - technical officer – ANSR; 1 Senior IT specialist - ANEPC
General work experience	On average, 35 years
Assessment instruments/process documentation	4 experts
Theoretical knowledge about:	
GIS	3 experts
Spatial Analysis	2 experts
Road traffic accidents	2 experts
Statistics	3 experts
Civil Protection	3 experts
Road safety	2 experts
Local and regional authorities	2 experts

5.3.2 Expert content evaluation process

The expert evaluation process was carried out in three interactions/phases:

The first interaction consisted of one-on-one discussions with specialists, during which the problem and objectives were identified, the dissertation guidelines were presented, and discussed the best approaches to the theme in question. This iteration's evaluation was carried out to confirm that our artefacts were suitable for their intended use, as described by the objective statements presented in Chapter 1, Table 10 shows the results of the evaluations.

Table 10 - Experts first interaction results

Criteria	Objective Statement	Content Validity Ratio
Efficiency	The Identification of the location of RTAs hotspots is efficient to assist the process of decision support of CML's civil protection and mobility officers	0,5
Consistency with the municipality's needs	The research responds totally to the needs identified in challenge 51	1
Detail level	The proposed solution provides the level of detail required for the decision support process	1
Methodologies for determining hotspots	The solution complies with the latest methods in the literature for hotspot determination	1
Utility	The solution is an added value for the reduction of traffic accidents with victims in the city of Lisbon	1
Applicability	Does the solution have practical applicability in the field of road safety and civil protection?	0,5
Consistency	Is the solution consistent throughout its analysis?	1
Transferability	The results of the study can be applied to other contexts or areas?	1
Dependability	The methodology and processes employed in the study allow external researchers to follow the same method used by the researcher?	1
Content Validity Index		0.88

The second interaction consisted of several one-on-one discussions with three specialists, the civil protection expert, the Road safety expert and the GIS specialist, and were focused on resolving all of the issues raised in the previous iteration's review, and to validate the options taken.

Table 11 - Experts second interaction results

Criteria	Objective Statement	Content Validity Ratio
Efficiency	The Identification of the location of RTAs hotspots is efficient to assist the process of decision support of CML's civil protection and mobility officers	0,5
Consistency with the municipality's needs	The research responds totally to the needs identified in challenge 51	1
Detail level	The proposed solution provides the level of detail required for the decision support process	1
Methodologies for determining hotspots	The solution complies with the latest methods in the literature for hotspot determination	1
Utility	The solution is an added value for the reduction of traffic accidents with victims in the city of Lisbon	1
Applicability	Does the solution have practical applicability in the field of road safety and civil protection?	1
Consistency	Is the solution consistent throughout its analysis?	1
Transferability	The results of the study can be applied to other contexts or areas?	1
Dependability	The methodology and processes employed in the study allow external researchers to follow the same method used by the researcher?	1
Content Validity Index		0,94

The third phase consisted of a form sent to the experts so that, after analyzing the findings of this dissertation, they could assess various factors concerning it. Once the experts have completed providing scores on all items, they submitted online their responses to the researcher (a form template is presented in Appendix C).

Table 12 - Experts third interaction results

Criteria	Objective Statement	Content Validity Ratio
Efficiency	The Identification of the location of RTAs hotspots is efficient to assist the process of decision support of CML's civil protection and mobility officers	1
Consistency with the municipality's needs	The research responds totally to the needs identified in challenge 51	1
Detail level	The proposed solution provides the level of detail required for the decision support process	1
Methodologies for determining hotspots	The solution complies with the latest methods in the literature for hotspot determination	1
Utility	The solution is an added value for the reduction of traffic accidents with victims in the city of Lisbon	1
Applicability	Does the solution have practical applicability in the field of road safety and civil protection?	1
Consistency	Is the solution consistent throughout its analysis?	1
Transferability	The results of the study can be applied to other contexts or areas?	1
Dependability	The methodology and processes employed in the study allow external researchers to follow the same method used by the researcher?	1
Content Validity Index		1

CHAPTER 6

Conclusions & Future Work

6.1 Conclusions

The object of this dissertation was derived from a real need of a municipality taking into account the need of knowledge extraction from available data of the Lisbon City Council Mobility Office, translated into a project (number 51) of LXDataLab. This initiative involves the municipality and the academy searching for solutions through data analysis for real city problems. This challenge required the identification of areas prone to traffic accidents (hotspots) in the municipality of Lisbon and their causes.

The Lisboa Inteligente challenge aims to give the municipal authorities responsible for mobility and the management of Protection and Rescue. The necessary information to improve the planning and management of the city's traffic, allowing the identification and implementation of mitigating measures that limit the number of accidents and victims in these accident hotspots.

We had no doubts in accepting the challenge since, by profession and training, Civil Protection has been the fruit of my dedication for more than 20 years.

Thus, and to achieve the objective now materialized, a multivariable space-time analysis of the road accidents that occurred between 2010 and 2019 was carried out to identify the location of its hotspots—the path took until the presented result was not easy. Several methodologies were tested and discarded, several meetings with experts and advisors were necessary to define the most appropriate methodology for this dissertation. Several months of time and dedication were invested in providing the municipality with an artifact that could help correct failures and ultimately save lives.

The work began with a collection of data provided by CML to ISCTE-IUL, which included a database dump of Fire Brigade Regiment (RSB) occurrences for the year 2019 and some shapefiles with geographical data of the municipality, namely the slopes values traffic lights, intersections, and altimetry. It seemed from the beginning and after analyzing the data provided that it fell short of the real needs for an investigation of this nature. Thus began a meticulous process to gather the necessary data to allow a robust and efficient analysis.

We identified in the literature that an ideal period for data collection when conducting a study on road accidents should be between 5 and 10 consecutive years since this interval represents a compromise between the number of years required to have a reasonable sample of seasonal patterns and still be in the processable range.

So the first task was to request from the Fire Brigade Regiment the data necessary for the analysis (10 years of data), and why the RSB? Because the Regiment is responsible for responding to all road accidents in the municipality, and ultimately will be the potential beneficiary of this study.

After receiving this data, we realized that essential fields for our investigation were still missing, namely the causes of the accident, the drivers' profile, and the weather conditions at the time of the incident.

In this way, we started contacts with ANSR and IPMA to obtain the necessary data. From these contacts was born a cooperation protocol between ISCTE and ANSR which we are very proud of because it will allow other students to access a precious source of information.

After a new analysis, we realized that it was still not possible to measure the volume of traffic on the Lisbon roads at the time of the incident. IMTT and TomTom were essential to get this answer.

Once the necessary data was gathered, we began to prepare it, a Herculean task, cleaning and preparing data from 6 different sources with georeferenced information and in different formats, carrying out the whole process of feature engineering for the transformation and construction of derived data, etc. This phase is guaranteed to occupy 60% to 70% of the time invested by the student in this study.

After this step and having a new dataset that gave us the necessary data for the study, we began the task of statistical data analysis, identifying patterns and deviations, and the search for a justification for some of the results that proved challenging.

Once this phase was finished, we proceeded to all the processing and geographic analysis work using ArcGIS PRO version 2.8 and the applicability of spatial autocorrelation algorithms, namely the Kernel density estimation, the Moran I index and the Hotspot analysis Getis-Ord Gi*.

The results allowed us to identify twelve streets considered problematic and the location of fourteen hotspots of traffic accidents in the municipality. Its analysis allows us to state that the major contributing factor to traffic accidents in the city of Lisbon is the human factor. The vast majority, if not the totality of the causes of RTAs in the places under consideration, are due to excessive speed or disrespect of traffic signs and traffic signals.

We have seen that environmental factors (weather) have little influence on road traffic accidents in the study area, with the exception of two cases identified, namely in an area prone to flooding located in the North-South axis, near the Duarte Pacheco viaduct and in the Segunda Circular at the end of Avenida Eusébio de Silva Ferreira, near the exit to the IC19. With the exception of these factors, we were unable to identify any other factors that could potentiate a traffic accident.

We identified that the vast majority of accidents were Pedestian collisions, a truly worrying fact that caused nearly 7000 victims in the period under review. We learned that the most serious accidents occur on Friday night and continue into Saturday morning. The age group most prone to an RTA is between 30 and 39, but the age group between 20 and 29 is the one with the most serious accidents, and that, statistically, men have more accidents than women.

As the product is a digital artifact, an output, and not a framework, it had to be validated, so the whole process was gradually accompanied by a panel of 4 experts from various branches related to the project's scope. There were several iterations made, as well as changes/corrections suggested.

To formalize this support, three iterations of formal evaluation were carried out to ensure the validity and usefulness of the results.

The first interaction consisted of one-on-one discussions with specialists, during which the problem and objectives were identified, the dissertation guidelines were presented, and discussed the best approaches to the theme in question. This iteration's evaluation was carried out to confirm that our artifacts were suitable for their intended use, some specialists appointed the question of the applicability of this analysis in the real world i.e. what the real impact this analysis will have, changes will be made? And the efficiency of the chosen method (algorithm used for the analysis).

The second interaction consisted of several one-on-one discussions with three specialists, the civil protection expert, the Road safety expert and the GIS specialist, and were focused on resolving all of the issues raised in the previous iteration's review, and to validate the options taken.

In the third and final interaction the experts analyzed all the results produced as well the digital artifacts and assess the various factors concerning it. Once the experts have completed providing scores on all items, they submitted the final evaluation online to the researcher.

Road accidents are a complex subject, with several variables that must be linked for the occurrence of a road traffic accident. This was undoubtedly my greatest difficulty, having to search for alternatives and sources of data as despair as TomTom or the data of the number of cars with insurance in the municipality of Lisbon. In the same way, this study is also difficult because the inexistence of an integrated database where a researcher can access all the data necessary for his investigation makes the whole process slow and costly.

6.2 Future Work

As this dissertation has shown, most road accidents with victims occur due to human error, distraction, use of mobile phones, overconfidence, alcohol, tiredness, etc. Of these, we note that speeding is undoubtedly the most significant contributor to road accidents.

Driving is a complex task that necessitates more than just knowledge of and compliance to the Highway Code. It necessarily requires special attention to the vehicle being driven and what is happening outside of it, i.e., to the road infrastructure (the road and its signposting, third-party vehicles, pedestrians, and weather conditions). Driving is thus an exercise including the dynamics of the vehicle itself in constant engagement with the dynamics of the others, each of which is permeable to the effect of the various factors.

Thus, we suggest a holistic approach that addresses all of the risk factors associated with RTAs, the proposed approach consists of 5 vectors:

1. Municipal Road Safety Plan: The elaboration of a Municipal Road Safety Plan (PMSR) adapted to the reality of the Municipality of Lisbon that aims at reducing fatalities and reducing the severity of their consequences and that promotes effective measures to encourage drivers and pedestrians to comply with safety rules, to reduce their exposure to the risk of accidents, with particular focus on the following areas as stipulated by ANSR [95]: Civic, school and professional education; Driving education and exams; Driver behavior; Vehicle safety; Driver and vehicle supervision; Improvement of infrastructures; Improvement in the rescue and support to victims; Safety studies; Cooperation and coordination of entities and Communication. At the time of writing, the Municipality of Lisbon still did not have a PMSR, there being only a working group proposal dated 2018 [96].

2. Effective Enforcement - More punitive policies for speeding. Reinforcement of the number of mobile radars of the Lisbon Municipal Police which significantly reduced the number of accidents with victims during 2019, as previously analyzed in chapter 5.2.1; The introduction of variable speed radars on the Lisbon roads most prone to severe accidents, as is the case of the Segunda Circular. The increase in enforcement actions (alcohol) by the Municipal Police and Public Security Police near nightlife areas and main access roads, particularly on Friday nights and Saturday mornings (dawn).

3. A safer pedestrian policy - The improvement of safety conditions for pedestrians and other vulnerable users. Lisbon has been improving these conditions a lot in this area. However, there is still a long way to go. We must continue along this path by continuously improving the pedestrian infrastructure, increasing visibility, and building pedestrian bridges over the busiest roads and those where the most pedestrian collisions are registered.

4. Education for risk - a commitment to improving the education and training of our drivers, the improvement and expansion of programs similar to the road safety education program - Safety starts with you [97] that encourages young people to adopt civic behavior as pedestrians, passengers, and drivers.

5. Data Support - Reliable data to identify the problems and help decision-makers regarding the policy to be applied.

Concerning this dissertation and in order to correct the problems faced during the data survey process and in order to have a consistent database capable of satisfying the needs of point 5 of the holistic proposal presented, it is essential to develop a set of recommendations to fill the information gaps identified during the said data survey process:

1. Changes to the way of processing the statistical bulletin of road accidents, which are being adopted, will allow for the densification of data and a greater understanding of them, which until now did not exist, i.e., the existence of integration between IPMA and ANSR, so that when the accident is entered in ANSR's database, the meteorological values for that date, time and location are automatically added to the BEAV.

2. The integration between ANSR and the georeferenced data of the recent e-segurnet[17] application from the Portuguese Insurers Association [98] will allow the coordinates of the accident to be compared with those recorded by the security forces and registered in the BEAV.

3. The causes of the accident could also be analyzed, i.e., densification of the data identifying the cause of the crash: were the drivers under the influence of alcohol? Was the reason for the distraction of the accident because they were on their mobile phones? Why was the pedestrian hit by a car? Was he or she wearing headphones or on a mobile phone? These and other questions would allow researchers, in the future, to develop more detailed studies according to the reality of road accidents with victims in the municipality of Lisbon.

[17] The e-SEGURNET application facilitates the participation and communication of a motor accident to your insurer. Created by the insurers, this application is an alternative to the traditional friendly declaration (DAAA). It also allows you to keep your data and that of your vehicles easily accessible. The geolocation, use of photos and help in filling out the form make it easier to report when the accident happens.

References

[1] EUROPEAN COMMISSION, “ERSO - Annual Accident Report 2016,” 2016. Accessed: Apr. 19, 2021. [Online]. Available: https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/statistics/dacota/asr2016.pdf

[2] WHO, “WORLD REPORT ON ROAD TRAFFIC INJURY PREVENTION - Glossary of terms,” 2016. Accessed: Apr. 19, 2021. [Online]. Available: http://www.unece.org/trans/main/wp6/

[3] Autoridade Nacional de Segurança Rodoviária, “Relatório Anual da Sinistralidade Rodoviária 2020,” Lisboa, 2020. Accessed: Apr. 19, 2021. [Online]. Available: http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2020/RelatórioSinistralidadeAno2020.pdf

[4] “WHO | Death on the roads.” https://extranet.who.int/roadsafety/death-on-the-roads/#trends/vehicles (accessed Apr. 29, 2021).

[5] “Global status report on road safety 2018.” https://www.who.int/publications/i/item/9789241565684 (accessed Mar. 31, 2021).

[6] “Road safety: Data show improvements in 2018 but further concrete and swift actions are needed.” https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1951 (accessed Apr. 13, 2021).

[7] “Estatísticas da segurança rodoviária relativas a 2019: O que está por detrás dos números?” https://ec.europa.eu/commission/presscorner/detail/pt/qanda_20_1004 (accessed Apr. 29, 2021).

[8] E. Comission, “2020 road safety statistics: what is behind the figures? | Mobility and Transport,” 2020. https://ec.europa.eu/transport/modes/road/news/2021-04-20-road-safety-statistics-2020_en (accessed Aug. 24, 2021).

[9] D.-G. for M. and T. (European Commission), “Handbook on the external costs of transport - Publications Office of the EU,” 2019. Accessed: Apr. 13, 2021. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/9781f65f-8448-11ea-bf12-01aa75ed71a1

[10] Autoridade Nacional de Segurança Rodoviária, “Relatório Anual de Segurança Rodoviária,” Lisbon, 2019. Accessed: Mar. 31, 2021. [Online]. Available: http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2019/Relatório Anual Sinistralidade Rodoviária 2019.pdf

[11] E. COMMISSION, *Towards a European road safety area: policy orientations on road safety 2011-2020*. 2010.

[12] EUROPEAN COMMISSION, “EU Road Safety Policy Framework 2021-2030 - Next steps towards ‘Vision Zero’ EN,” 2019. Accessed: Apr. 13, 2021. [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

[13] P. do C. de Ministros, *Resolução do Conselho de Ministros n.º 85/2017 - Plano Estratégico Nacional de Segurança Rodoviária - PENSE 2020*.

[14] CML, “Iniciativas - Lisboa Inteligente,” 2021. https://lisboainteligente.cm-lisboa.pt/iniciativas/ (accessed Jun. 28, 2021).

[15] S. Lakshmi, I. Srikanth, and M. Arockiasamy, “Identification of Traffic Accident Hotspots using Geographical Information System (GIS),” *Int. J. Eng. Adv. Technol.*, no. 2, pp. 2249–8958, 2019, doi: 10.35940/ijeat.B3848.129219.

[16] “FLUP - Tese de Mestrado em Sistemas de Informação Geográfica e Ordenamento de Território.” https://sigarra.up.pt/flup/pt/teses.tese?p_aluno_id=103675&p_processo=17587&p_lang=1 (accessed May 03, 2021).

[17] V. Albuquerque, A. Oliveira, J. Lourenço Barbosa, R. S. Rodrigues, F. Andrade, M. S. Dias, J. C. Ferreira, and L. Hernández-Callejo, “Smart Cities: Data-Driven Solutions to Understand Disruptive Problems in Transportation-The Lisbon Case Study,” 2021, doi: 10.3390/en14113044.

[18] R. Elvik, “Speed limits, enforcement, and health consequences,” *Annual Review of Public Health*, vol. 33. pp. 225–238, Apr. 21, 2012. doi: 10.1146/annurev-publhealth-031811-124634.

[19] T. K. T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots,” *Accid. Anal. Prev.*, vol. 41, no. 3, pp. 359–364, May 2009, doi: 10.1016/j.aap.2008.12.014.

[20] R. Elvik, “A survey of operational definitions of hazardous road locations in some European countries,” *Accid. Anal. Prev.*, vol. 40, no. 6, pp. 1830–1835, Nov. 2008, doi: 10.1016/j.aap.2008.08.001.

[21] I. Moreira, “1 Identification of Hazardous Road Locations in Urban Areas through Spatial Analysis : Application to Vila Real,” 2011.

[22] Ž. BEng and N. magingtraff, “COMPARISON OF METHODS FOR DETERMINING CRASH HOTSPOTS IN THE ROAD TRAFFIC.”

[23] A. Ceder and M. Livneh, “Relationships between road accidents and hourly traffic flow-I. Analyses and interpretation,” *Accid. Anal. Prev.*, vol. 14, no. 1, pp. 19–34, Feb. 1982, doi: 10.1016/0001-4575(82)90004-5.

[24] S. Oppe, “The use of multiplicative models for analysis of road safety data,” *Accid. Anal. Prev.*, vol. 11, no. 2, pp. 101–115, Jun. 1979, doi: 10.1016/0001-4575(79)90018-6.

[25] J. Otilia da Costa Elisabete Freitas Paulo António Alves Pereira and M. Alice Prudêncio Jacques, “ACIDENTES RODOVIÁRIOS DAS ESTRADAS NACIONAIS DE PORTUGAL: ESTUDO DA ASSOCIAÇÃO ENTRE AS VARIÁVEIS RECOLHIDAS,” 2011. Accessed: Apr. 14, 2021. [Online]. Available: http://repositorium.sdum.uminho.pt/handle/1822/15483

[26] K. Bucsuházy, E. Matuchová, R. Zůvala, P. Moravcová, M. Kostíková, and R. Mikulec, “Human factors contributing to the road traffic accident occurrence,” in *Transportation Research Procedia*, Jan. 2020, vol. 45, pp. 555–561. doi: 10.1016/j.trpro.2020.03.057.

[27] P. L. Brito, E. G. V. Jesus, R. M. S. Sant’Ana, C. Martins, J. P. M. Delgado, and V. O. Fernandes, “Official crime data versus collaborative crime mapping at a Brazilian City,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2014, vol. 40, no. 2, pp. 137–144. doi: 10.5194/isprsarchives-XL-2-137-2014.

[28] N. Rybnikova, R. G. Stevens, D. I. Gregorio, H. Samociuk, and B. A. Portnov, “Kernel density analysis reveals a halo pattern of breast cancer incidence in Connecticut,” *Spat. Spatiotemporal. Epidemiol.*, vol. 26, pp. 143–151, 2018, doi: 10.1016/j.sste.2018.06.003.

[29] P. Goovaerts, “Combining area-based and individual-level data in the geostatistical mapping of late-stage cancer incidence,” *Spat. Spatiotemporal. Epidemiol.*, vol. 1, no. 1, pp. 61–71, 2009, doi: 10.1016/j.sste.2009.07.001.

[30] J. A. P. Góes, D. G. Souza, L. A. Andrade, J. Cunha, S. Kameo, M. A. O. Góes, A. F. de Menezes, M. A. P. Nunes, K. C. G. M. Araújo, and A. D. Santos, “Trend and spatial analysis of prostate cancer mortality in the state of Sergipe, Brazil,” *Geospat. Health*, vol. 13, no. 2, pp. 353–358, 2018, doi: 10.4081/gh.2018.732.

[31] T. B. Sasongko, O. Arifin, and H. Al Fatta, “Optimization of hyper parameter bandwidth on naïve Bayes kernel density estimation for the breast cancer classification,” in *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, 2019, pp. 226–231. doi: 10.1109/ICOIACT46704.2019.8938497.

[32] N. Méndez, M. Oviedo-Pastrana, S. Mattar, I. Caicedo-Castro, and G. Arrieta, “Zika virus disease, microcephaly and Guillain-Barré syndrome in Colombia: Epidemiological situation during 21 months of the Zika virus outbreak, 2015-2017,” *Arch. Public Heal.*, vol. 75, no. 1, 2017, doi: 10.1186/s13690-017-0233-5.

[33] D. E. Cuartas, D. Arango-Londoño, G. Guzmán-Escarria, E. Muñoz, D. Caicedo, D. Ortega, A. Fandiño-Losada, J. Mena, M. Torres, L. Barrera, L. Barrera, and F. Méndez, “Sars-cov-2 spatio temporal analysis in Cali, Colombia | Análisis espacio-temporal del sars-cov-2 en Cali, Colombia,” *Rev. Salud Publica*, vol. 22, no. 2, pp. 1–6, 2020, doi: 10.15446/rsap.V22n2.86431.

[34] F. E. Rex, C. A. S. Borges, and P. S. Käfer, "Spatial analysis of the COVID-19 distribution pattern in São Paulo State, Brazil | Análise espacial do padrão de distribuição do COVID-19 no Estado de São Paulo, Brasil," *Cienc. e Saude Coletiva*, vol. 25, no. 9, pp. 3377–3384, 2020, doi: 10.1590/1413-81232020259.17082020.

[35] B. Niu, R. Liang, S. Zhang, H. Zhang, X. Qu, Q. Su, L. Zheng, and Q. Chen, "Epidemic analysis of COVID-19 in Italy based on spatiotemporal geographic information and Google Trends," *Transbound. Emerg. Dis.*, 2020, doi: 10.1111/tbed.13902.

[36] R. Elson, T. M. Davies, I. R. Lake, R. Vivancos, P. B. Blomquist, A. Charlett, and G. Dabrera, "The spatio-temporal distribution of COVID-19 infection in England between January and June 2020," *Epidemiol. Infect.*, vol. 149, p. e73, 2021, doi: 10.1017/S0950268821000534.

[37] W. Shi, C. Tong, A. Zhang, B. Wang, Z. Shi, Y. Yao, and P. Jia, "An extended Weight Kernel Density Estimation model forecasts COVID-19 onset risk and identifies spatiotemporal variations of lockdown effects in China," *Commun. Biol.*, vol. 4, no. 1, 2021, doi: 10.1038/s42003-021-01677-2.

[38] K. G. K. G. Le, P. Liu, and L.-T. L.-T. Lin, "Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam," vol. 23, no. 2, pp. 153–164, Apr. 2020, doi: 10.1080/10095020.2019.1683437.

[39] K. M. S. Newaz, S. Hasanat-E-Rabbi, and S. Miaji, "Spatio-temporal study of road traffic crash on a national highway of Bangladesh," in *2017 4th International Conference on Transportation Information and Safety, ICTIS 2017 - Proceedings*, 2017, pp. 60–66. doi: 10.1109/ICTIS.2017.8047743.

[40] A. E. Hie Tan, "Planar and network spatial analyses of road traffic accidents - A review of methods and tools," in *IET Conference Publications*, 2018, vol. 2018, no. CP750.

[41] H. B. T. P. Jayathilaka and V. P. I. S. Wijeratne, "GIS-based space and time analysis of road accidents in an urban environment: A case study of the Colombo MC area," 2020.

[42] C. Toval-Ruiz, C. Rojas-Roque, and A. Hernández-Vásquez, "Predictors and geographic analysis of road traffic accidents in leon, nicaragua | Preditores e análise geográfica de acidentes de trânsito em león, nicarágua," *Rev. Bras. Epidemiol.*, vol. 24, pp. 1–13, 2021, doi: 10.1590/1980-549720210003.

[43] M. A. Dereli and S. Erdogan, "A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods," *Transp. Res. Part A Policy Pract.*, vol. 103, pp. 106–117, 2017, doi: 10.1016/j.tra.2017.05.031.

[44] G. Yalcin and H. S. Duzgun, "Spatial analysis of two-wheeled vehicles traffic crashes: Osmaniye in Turkey," *KSCE J. Civ. Eng.*, vol. 19, no. 7, pp. 2225–2232, 2015, doi: 10.1007/s12205-015-0661-0.

[45] A. Fang, X. Peng, J. Zhou, and L. Tang, “Research on the Map-matching and Spatial-temporal Visualization of Expressway Traffic Accident Information,” in *2018 3rd IEEE International Conference on Intelligent Transportation Engineering, ICITE 2018*, Oct. 2018, pp. 23–27. doi: 10.1109/ICITE.2018.8492572.

[46] L. Ramos, L. Silva, M. Y. Santos, and J. M. Pires, “Detection of Road Accident Accumulation Zones with a Visual Analytics Approach,” *Procedia Comput. Sci.*, vol. 64, pp. 969–976, Jan. 2015, doi: 10.1016/J.PROCS.2015.08.615.

[47] Ö. Kaygisiz, Ş. Düzgün, A. Yildiz, and M. Senbil, “Spatio-temporal accident analysis for accident prevention in relation to behavioral factors in driving: The case of South Anatolian Motorway,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 33, pp. 128–140, Aug. 2015, doi: 10.1016/j.trf.2015.07.002.

[48] J. Aguero-Valverde and P. P. Jovanis, “Spatial analysis of fatal and injury crashes in Pennsylvania,” *Accid. Anal. Prev.*, vol. 38, no. 3, pp. 618–625, May 2006, doi: 10.1016/j.aap.2005.12.006.

[49] A. E. Iyanda, “Geographic analysis of road accident severity index in Nigeria,” *Int. J. Inj. Contr. Saf. Promot.*, vol. 26, no. 1, pp. 72–81, Jan. 2019, doi: 10.1080/17457300.2018.1476387.

[50] S. S. Pulugurtha, V. K. Krishnakumar, and S. S. Nambisan, “New methods to identify and rank high pedestrian crash zones: An illustration,” *Accid. Anal. Prev.*, vol. 39, no. 4, pp. 800–811, 2007, doi: 10.1016/j.aap.2006.12.001.

[51] S. Erdogan, I. Yilmaz, T. Baybura, and M. Gullu, “Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar,” *Accid. Anal. Prev.*, vol. 40, no. 1, pp. 174–181, Jan. 2008, doi: 10.1016/j.aap.2007.05.004.

[52] L. Thakali, T. J. Kwon, and L. Fu, “Identification of crash hotspots using kernel density estimation and kriging methods: a comparison,” *J. Mod. Transp.*, vol. 23, no. 2, pp. 93–106, 2015, doi: 10.1007/s40534-015-0068-0.

[53] F. Keskin, F. Yenilmez, M. Çolak, I. Yavuzer, and H. Ş. Düzgün, “Analysis of traffic incidents in METU campus,” in *Procedia - Social and Behavioral Sciences*, 2011, vol. 19, pp. 61–70. doi: 10.1016/j.sbspro.2011.05.108.

[54] V. Prasannakumar, H. Vijith, R. Charutha, and N. Geetha, “Spatio-temporal clustering of road accidents: GIS based analysis and assessment,” in *Procedia - Social and Behavioral Sciences*, Jan. 2011, vol. 21, pp. 317–325. doi: 10.1016/j.sbspro.2011.07.020.

[55] M. A. Aghajani, R. S. Dezfoulian, A. R. Arjroody, and M. Rezaei, “Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (case study: Ilam Province, Iran),” in *Transportation Research Procedia*, 2017, vol. 25, pp. 2126–2138. doi: 10.1016/j.trpro.2017.05.409.

[56] A. J. Brimicombe, “Detecting clusters in spatially repetitive point event data sets,” *CyberGeo*, vol. 2007, pp. 1–13, 2007.

[57] G. O. Korter, O. E. Olubusoye, and A. A. Salisu, “Spatial analysis of road traffic crashes in Oyo state of Nigeria,” *J. Sustain. Dev.*, vol. 7, no. 4, 2014, doi: 10.5539/jsd.v7n4p151.

[58] W. J. Fu, P. K. Jiang, G. M. Zhou, and K. L. Zhao, “Using Moran’s i and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China,” *Biogeosciences*, vol. 11, no. 8, 2014, doi: 10.5194/bg-11-2401-2014.

[59] “Spatial Autocorrelation (Global Moran’s I) (Spatial Statistics)—ArcGIS Pro | Documentation.” https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm (accessed Jul. 15, 2021).

[60] D. O’sullivan and D. W. S. Wong, “A surface-based approach to measuring spatial segregation,” *Geogr. Anal.*, vol. 39, no. 2, 2007, doi: 10.1111/j.1538-4632.2007.00699.x.

[61] M. P. Wand and M. C. Jones, *Kernel Smoothing (Monographs on Statistics and Applied Probability)*, vol. 53, no. 9. 1995.

[62] C. Plug, J. Xia, and C. Caulfield, “Spatial and temporal visualisation techniques for crash analysis,” *Accid. Anal. Prev.*, vol. 43, no. 6, pp. 1937–1946, Nov. 2011, doi: 10.1016/j.aap.2011.05.007.

[63] L. T. Truong and S. V. C. Somenahalli, “Using GIS to identify pedestrian- vehicle crash hot spots and unsafe bus stops,” *J. Public Transp.*, vol. 14, no. 1, 2011, doi: 10.5038/2375-0901.14.1.6.

[64] A. Getis and J. K. Ord, “The Analysis of Spatial Association by Use of Distance Statistics,” *Geogr. Anal.*, vol. 24, no. 3, 1992, doi: 10.1111/j.1538-4632.1992.tb00261.x.

[65] P. M. Oliveira, “Os Factores Potenciadores da Sinistralidade Rodoviária,” Lisbon, 2007.

[66] A. Borsos, S. Birth, and H. J. Vollpracht, “The role of human factors in road design,” in *6th IEEE Conference on Cognitive Infocommunications, CogInfoCom 2015 - Proceedings*, Jan. 2016, pp. 363–367. doi: 10.1109/CogInfoCom.2015.7390620.

[67] M. Vlkovský, M. Šmerek, and J. Michálek, “Cargo Securing during Transport Depending on the Type of a Road,” in *IOP Conference Series: Materials Science and Engineering*, Nov. 2017, vol. 245, no. 4, p. 042001. doi: 10.1088/1757-899X/245/4/042001.

[68] A. Gregoriades and K. C. Mouskos, “Black spots identification through a Bayesian Networks quantification of accident risk index,” *Transp. Res. Part C Emerg. Technol.*, vol. 28, pp. 28–43, Mar. 2013, doi: 10.1016/j.trc.2012.12.008.

[69] A. M. P. Moreira, “Avaliação de traçados rodoviários,” 2000. Accessed: Apr. 16, 2021. [Online]. Available: https://repositorio-aberto.up.pt/handle/10216/12670

[70] C. A. Catalina, S. García-Herrero, E. Cabrerizo, S. Herrera, S. García-Pineda, F. Mohamadi, and M. A. Mariscal, “Music Distraction among Young Drivers: Analysis by Gender and Experience,”

2020, doi: 10.1155/2020/6039762.

[71] A. PEDROSA, António; PEREIRA, "'Cidades e dinâmicas urbanas: a gestão de relações complexas em territórios de risco,'" in *Riscos naturais, antrópicos e mistos. Homenagem ao Professor Doutor Fernando Rebelo*, Coimbra: Departamento de Geografia, Faculdade de Letras, Universidade de Coimbra, 2013, pp. 599–617.

[72] L. Lourenço and F. Tedim, *Realidades e desafios na gestão dos riscos: diálogo entre ciência e utilizadores*, Imprensa d. Coimbra: Imprensa da Universidade de Coimbra Núcleo de Investigação Científica de Incêndios Florestais, 2014. doi: http://dx.doi.org/10.14195/978-972-8330-23-1.

[73] SMPC, *PLANO MUNICIPAL de EMERGÊNCIA de PROTEÇÃO CIVIL de LISBOA*. 2017.

[74] T. Alexandra, A. Lopes, and C. Da Fonseca, "O PARADIGMA DO PLANEAMENTO DE EMERGÊNCIA DE PROTECÇÃO CIVIL EM PORTUGAL," 2010. Accessed: Jan. 09, 2021. [Online]. Available: http://hdl.handle.net/10316/14572

[75] M. munhos, C. Pinto, S. Efigénio, and J. Vicente, "Lx_Risk - Caracterização Sócio-Urbanística," *Departamento de Informação Geográfica e Cadastro – Câmara Municipal de Lisboa*. http://lxrisk.cm-lisboa.pt/caract_socio_urb.html (accessed Jan. 10, 2021).

[76] CML, "Relatório síntese de Caracterização Biofísica de Lisboa no âmbito da Revisão do Plano Director Municipal de Lisboa," 2010. Accessed: Jan. 09, 2021. [Online]. Available: https://www.lisboa.pt/fileadmin/cidade_temas/urbanismo/pdm/EstCarat_RSinteseCaracterizacaoBiofisica.pdf

[77] INE, "Instituto Nacional de Estatistica, Censos 2011." https://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=ine_censos_publicacao_det&contexto=pu&PUBLICACOESpub_boui=156651739&PUBLICACOESmodo=2&selTab=tab1&pcensos=61969554 (accessed Jan. 09, 2021).

[78] M. A. . A. S. Chowdhury, *Fundamentals of Intelligent Transportation Systems Planning - Mashrur A*. 2003.

[79] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *J. Data Warehous.*, vol. 5, no. 4, 2000.

[80] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide," 2000.

[81] "OpenWeatherMap® API: Get Historical & Current Weather Data | RapidAPI." https://rapidapi.com/blog/lp/openweathermap/?utm_source=google&utm_medium=cpc&utm_campaign=Alpha&utm_term=openweathermap_e&gclid=CjwKCAjwruSHBhAtEiwA_qCppnYNPM1L2DN_O3ZYEHcCHrXEMCC-S86LRPAm0C3_ipNDj1WcO-yVpRoCEYcQAvD_BwE (accessed Nov. 23, 2020).

[82] ANSR, “Manual de Preenchimento do BEAV.” 2015. Accessed: Jun. 12, 2021. [Online]. Available: http://www.ansr.pt/Estatisticas/BEAV/Documents/MANUAL PREENCHIMENTO BEAV.pdf

[83] T. NV, “TomTom Move,” *TomTom NV*, 2020. https://move.tomtom.com/login (accessed Mar. 11, 2021).

[84] J. Pérez, E. Iturbide, V. Olivares, M. Hidalgo, A. Martínez, and N. Almanza, “A Data Preparation Methodology in Data Mining Applied to Mortality Population Databases,” *J. Med. Syst.*, vol. 39, no. 11, Nov. 2015, doi: 10.1007/s10916-015-0312-5.

[85] “pandas - Python Data Analysis Library.” https://pandas.pydata.org/ (accessed Jun. 30, 2021).

[86] “NumPy.” https://numpy.org/ (accessed Jun. 30, 2021).

[87] “GitHub - rhiever/datacleaner: A Python tool that automatically cleans data sets and readies them for analysis.” https://github.com/rhiever/datacleaner (accessed Jun. 30, 2021).

[88] “Data Wrangling Software and Tools | Trifacta.” https://www.trifacta.com/ (accessed Jun. 30, 2021).

[89] “The Beaufort Scale | Royal Meteorological Society.” https://www.rmets.org/resource/beaufort-scale (accessed Jun. 18, 2021).

[90] B. A. MIKELBANK, “Quantitative Geography: Perspectives on Spatial Data Analysis, by A. S. Fotheringham, C. Brunsdon, and M. Charlton,” *Geogr. Anal.*, vol. 33, no. 4, 2010, doi: 10.1111/j.1538-4632.2001.tb00453.x.

[91] “Crime Occurrence Factors from a Microscale Environmental Characteristics Point of View,” *J. Soc. Saf. Sci.*, vol. 13, 2010, doi: 10.11314/jisss.13.109.

[92] Turismo de Portugal, “Turismo em pt.” https://travelbi.turismodeportugal.pt/pt-pt/Paginas/turismo-em-numeros-2018.aspx (accessed Jul. 18, 2021).

[93] M. S. B. Yusoff, “ABC of Content Validation and Content Validity Index Calculation,” *Educ. Med. J.*, vol. 11, no. 2, pp. 49–54, Jun. 2019, doi: 10.21315/EIMJ2019.11.2.6.

[94] I. B. Rodrigues, J. D. Adachi, K. A. Beattie, and J. C. MacDermid, “Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis,” *BMC Musculoskelet. Disord. 2017 181*, vol. 18, no. 1, pp. 1–9, Dec. 2017, doi: 10.1186/S12891-017-1914-5.

[95] ANSR, “Guia para elaboração de planos municipais de segurança rodoviária,” 2015. http://www.ansr.pt/SegurancaRodoviaria/PlanosdeSegurancaRodoviaria/Documents/Guia_Planos_Municiapais_Seguranca_Rodoviaria.pdf (accessed Aug. 29, 2021).

[96] CML, “2.º SUPLEMENTO AO BOLETIM MUNICIPAL N.º 1392,” *Boletim Municipal da CML*, Oct. 22, 2020. https://bmpesquisa.cm-lisboa.pt/pls/OKUL/app_bm.download_my_file?p_file=3484 (accessed Aug. 29, 2021).

[97] CML, “Câmara Municipal de Lisboa, Educação, Programas,” 2020. https://www.lisboa.pt/cidade/educacao/programas (accessed Aug. 29, 2021).

[98] Portuguese Insurers Association, “APS e-SEGURNET. Participação eletrónica de sinistros.” https://e-segurnet.pt/ (accessed Aug. 19, 2021).

[99] “What is Negative Rainfall? - Geography Realm.” https://www.geographyrealm.com/what-is-negative-rainfall/ (accessed Jun. 23, 2021).

Page left intentionally blank

Appendices

Page left intentionally blank

Appendix A – Boletim Estatístico de Acidentes de Viação (BEAV)

Instrumento de notação registado no I.N.E., sob o n.º 10111, válido até 31/12/2014

Nº Boletim

Entidade Fiscalizadora

ANSR
AUTORIDADE NACIONAL SEGURANÇA RODOVIÁRIA
Ministério da Administração Interna

BOLETIM ESTATÍSTICO DE ACIDENTES DE VIAÇÃO

(Reservado ao carimbo da unidade)

A - a preencher em todos os acidentes **B e seguintes - a preencher apenas em acidentes com vítimas**

A - IDENTIFICAÇÃO DO ACIDENTE

A1 **DATA/HORA**

Ano Mês Dia Hora Min.

A2 **LOCALIZAÇÃO**

1. ☐ Fora das localidades
☐ Dentro das localidades

2. Distrito

Concelho

Freguesia

Povoação (ou a mais próxima)

Coordenadas GPS
Latitude +
Longitude -

3. Designação de via

Km

Arruamento

n.º

4. Se houver separador central indique em que sentido
1 ☐ Crescente
2 ☐ Decrescente

A3 **TIPO DE ACIDENTE**
1 ☐ Acidente só com danos materiais
2 ☐ Acidente com vítimas
Mortos
Feridos graves
Feridos leves

A4 **NATUREZA DO ACIDENTE**
1 ☐ Despiste
2 ☐ Colisão
3 ☐ Atropelamento

A5 **NÚMERO DE VEÍCULOS INTERVENIENTES**
Ciclomotor e motociclo
Veículo ligeiro
Veículo pesado
Outros

A6 **CONDUTORES INTERVENIENTES**

1. SEXO
A B C
1 ☐ ☐ ☐ Masculino
2 ☐ ☐ ☐ Feminino

2. DATA DE NASCIMENTO
A Ano Mês Dia
B Ano Mês Dia
C Ano Mês Dia

B - CIRCUNSTÂNCIAS EXTERNAS

B1 **CARACTERÍSTICAS TÉCNICAS DA VIA**

1. ESTRADA COM SEPARADOR
1 Autoestrada – nº de vias de trânsito no sentido
2 Outra via – nº de vias de trânsito no sentido

2. ESTRADA SEM SEPARADOR – nº de vias no sentido

3. VIA DE TRÂNSITO
1 ☐ Esquerda
2 ☐ Direita
3 ☐ Central

B2 **TRAÇADO DA VIA**

1. EM PLANTA
1 ☐ Reta
2 ☐ Curva

2. EM PERFIL
1 ☐ Em patamar
2 ☐ Com inclinação
3 ☐ Em lomba

3. 1 ☐ Sem berma ou impraticável
2 ☐ Berma não pavimentada
3 ☐ Berma pavimentada

4. SITUAÇÃO DO ACIDENTE
1 ☐ Em plena via
2 ☐ Na berma
3 ☐ No passeio
4 ☐ Em via ou pista reservada
5 ☐ Em parque de estacionamento

5. INTERSEÇÃO DE VIAS
1 ☐ **Fora da interseção**
Em interseção de nível
2 ☐ Em cruzamento
3 ☐ Em entroncamento
4 ☐ Em rotunda
5 ☐ Em passagem de nível
Em interseção desnivelada
6 ☐ Em via de aceleração
7 ☐ Em via de desaceleração
8 ☐ Em ramo de ligação – entrada
9 ☐ Em ramo de ligação – saída

6. ACIDENTE EM OBRAS DE ARTE
1 ☐ Túnel
2 ☐ Viaduto/Ponte
3 ☐ Passagem estreita

B3 **REGIME DE CIRCULAÇÃO**

1. FAIXA DE RODAGEM COM
1 ☐ Sentido único
2 ☐ Dois sentidos
3 ☐ Reversível

2. VELOCIDADE PERMITIDA NO LANÇO
Limite geral ____ Km/h
Limite local ____ Km/h

B4 **PAVIMENTO**

1. TIPO DE PISO
1 ☐ Terra batida
2 ☐ Betuminoso
3 ☐ Betão de cimento
4 ☐ Calçada

2. ESTADO DE CONSERVAÇÃO
1 ☐ Em bom estado
2 ☐ Em estado regular
3 ☐ Em mau estado

3. OBSTÁCULOS OU OBRAS
1 ☐ Inexistentes
2 ☐ Não sinalizados
3 ☐ Insuficientemente sinalizados
4 ☐ Corretamente sinalizados

4. CONDIÇÕES DE ADERÊNCIA
1 ☐ Seco e limpo
2 ☐ Húmido
3 ☐ Molhado
4 ☐ Com água acumulada na faixa de rodagem
5 ☐ Com gelo, geada ou neve
6 ☐ Com lama
7 ☐ Com gravilha ou areia
8 ☐ Com óleo

B5 **SINALIZAÇÃO**

1. MARCAS NO PAVIMENTO
1 ☐ Sem marcas rodoviárias ou pouco visíveis
2 ☐ Com marcas – separadoras de sentido de trânsito
3 ☐ Com marcas – separadoras de sentido e de vias de trânsito

2. SINALIZAÇÃO LUMINOSA
1 ☐ Inexistente
2 ☐ A funcionar normalmente
3 ☐ Intermitente
4 ☐ Desligada

3. SINAIS
1 ☐ Stop
2 ☐ Cedência de passagem
3 ☐ Proibição de ultrapassagem
4 ☐ Passagem de peões
5 ☐ Outros

B6 **LUMINOSIDADE**
1 ☐ Em pleno dia
2 ☐ Sol encandeante
3 ☐ Aurora ou crepúsculo
4 ☐ Noite, sem iluminação
5 ☐ Noite, com iluminação

B7 **FATORES ATMOSFÉRICOS**
1 ☐ Bom tempo
2 ☐ Chuva
3 ☐ Vento forte
4 ☐ Nevoeiro
5 ☐ Neve
6 ☐ Nuvem de fumo
7 ☐ Granizo

C - NATUREZA DO ACIDENTE

DESPISTE
1 ☐ Despiste simples
Com transposição do separador central
2 ☐ Com dispositivo de retenção
3 ☐ Sem dispositivo de retenção
4 ☐ Com transposição do dispositivo de retenção lateral
5 ☐ Com capotamento
6 ☐ Com colisão com veículo imobilizado ou obstáculo
7 ☐ Com fuga

COLISÃO
8 ☐ Frontal
9 ☐ Traseira com outro veículo em movimento
10 ☐ Lateral com outro veículo em movimento
11 ☐ Com veículo ou obstáculo na faixa de rodagem
12 ☐ Choque em cadeia
13 ☐ Com fuga
14 ☐ Outras situações

ATROPELAMENTO
15 ☐ De peões
16 ☐ De animais
17 ☐ Com fuga

Incêndio posterior. A B C ☐ ☐ ☐ A preencher no caso de se verificar

D - VEÍCULOS INTERVENIENTES

D1 **CATEGORIA/CLASSE**

1. VEÍCULOS A, B e C
A B C
1 ☐ ☐ ☐ Velocípede
2 ☐ ☐ ☐ Velocípede c/motor
3 ☐ ☐ ☐ Ciclomotor
4 ☐ ☐ ☐ Triciclo
5 ☐ ☐ ☐ Motociclo cilindrada ≤ 125cc
6 ☐ ☐ ☐ Motociclo cilindrada > 125cc
7 ☐ ☐ ☐ Automóvel ligeiro
8 ☐ ☐ ☐ Automóvel pesado
9 ☐ ☐ ☐ Veículo agrícola
10 ☐ ☐ ☐ Máquina industrial
11 ☐ ☐ ☐ Veículo sobre carris
12 ☐ ☐ ☐ Veículo de tração animal
13 ☐ ☐ ☐ Quadriciclo
14 ☐ ☐ ☐ Desconhecido

2. Se for automóvel ligeiro ou pesado, indicar o tipo:
A B C
1 ☐ ☐ ☐ Passageiros
2 ☐ ☐ ☐ Mercadorias
3 ☐ ☐ ☐ Misto
4 ☐ ☐ ☐ Trator
5 ☐ ☐ ☐ Veículo especial. Qual?

3. A B C

1 ☐ ☐ ☐ Sem semireboque/reboque
2 ☐ ☐ ☐ Com semireboque/reboque

D2 TIPO DE SERVIÇO

A B C

1 ☐ ☐ ☐ Particular
2 ☐ ☐ ☐ Público

D3 ANO DE MATRÍCULA

A ☐☐☐☐ B ☐☐☐☐ C ☐☐☐☐

D4 INSPEÇÃO PERIÓDICA

A B C

1 ☐ ☐ ☐ Não obrigatória
2 ☐ ☐ ☐ Válida
3 ☐ ☐ ☐ Sem validade

D5 CERTIFICADO ADR

1. Preencher apenas no caso de transporte de mercadorias perigosas

A B C

1 ☐ ☐ ☐ Válido
2 ☐ ☐ ☐ Sem validade
3 ☐ ☐ ☐ Inexistente

2. MATÉRIA/OBJETO PERIGOSO TRANSPORTADO ☐☐☐☐

D6 CARGA/LOTAÇÃO/PNEUS

1. CARGA/LOTAÇÃO

A B C

1 ☐ ☐ ☐ Sem carga
2 ☐ ☐ ☐ Com excesso de carga
3 ☐ ☐ ☐ Carga bem acondicionada
4 ☐ ☐ ☐ Carga mal acondicionada
5 ☐ ☐ ☐ Com lotação excedida

2. PNEUS

A B C

1 ☐ ☐ ☐ Sem deficiência
2 ☐ ☐ ☐ Com deficiência

3. TACÓGRAFO

A B C

1 ☐ ☐ ☐ Sem tacógrafo ou desativado
2 ☐ ☐ ☐ Com tacógrafo

D7 SEGURO

A B C

1 ☐ ☐ ☐ Com seguro
2 ☐ ☐ ☐ Sem seguro
3 ☐ ☐ ☐ Isento

E - CONDUTORES INTERVENIENTES

E1 CARACTERÍSTICAS DA HABILITAÇÃO DE CONDUÇÃO

1. LICENÇA/CARTA DE CONDUÇÃO

A B C

1 ☐ ☐ ☐ Com licença/carta adequada ao veículo
2 ☐ ☐ ☐ Com licença/carta não adequada ao veículo
3 ☐ ☐ ☐ Em situação de instrução/exame
4 ☐ ☐ ☐ Caducada/suspensa
5 ☐ ☐ ☐ Sem licença/carta
6 ☐ ☐ ☐ Não necessária ao veículo que conduz

2. PAÍS DE EMISSÃO

A B C

1 ☐ ☐ ☐ Portugal
2 ☐ ☐ ☐ Outro(s) A ☐☐☐ B ☐☐☐ C ☐☐☐

3. ANO DA HABILITAÇÃO

Relativamente ao veículo que conduzia

A ☐☐☐☐ B ☐☐☐☐ C ☐☐☐☐

4. CERTIFICADO ADR

A B C

1 ☐ ☐ ☐ Válido
2 ☐ ☐ ☐ Sem validade
3 ☐ ☐ ☐ Inexistente

E2 CONDIÇÕES PSÍCO/FÍSICAS

1. CONTROLO DO NÍVEL DE ALCOOLEMIA

A B C

1 ☐ ☐ ☐ **Submetido ao teste de alcoolemia**

Não submetido por

2 ☐ ☐ ☐ Doença
3 ☐ ☐ ☐ Lesão ou morte decorrente do acidente
4 ☐ ☐ ☐ Condutor não contactado na altura do acidente
5 ☐ ☐ ☐ Fuga
6 ☐ ☐ ☐ Recusa
7 ☐ ☐ ☐ Outra

2. TAXA DE ALCOOLEMIA

A ☐,☐☐ B ☐,☐☐ C ☐,☐☐

3. OUTROS FATORES

A B C

1 ☐ ☐ ☐ Normal
2 ☐ ☐ ☐ Droga por despistagem
3 ☐ ☐ ☐ Sono/sonolência
4 ☐ ☐ ☐ Distração
5 ☐ ☐ ☐ Doença súbita
6 ☐ ☐ ☐ Fadiga

4. TEMPO DE CONDUÇÃO CONTINUADA

A B C

1 ☐ ☐ ☐ Menos de 1 hora
2 ☐ ☐ ☐ De 1 a 3 horas
3 ☐ ☐ ☐ De 3 a 5 horas
4 ☐ ☐ ☐ Mais de 5 horas
5 ☐ ☐ ☐ Ignorada

E3 AÇÕES E MANOBRAS ANTES DO ACIDENTE

1. A B C

1 ☐ ☐ ☐ Início de marcha
2 ☐ ☐ ☐ Saída de parqueamento ou rua particular
3 ☐ ☐ ☐ Em marcha normal
4 ☐ ☐ ☐ Ultrapassagem pela esquerda
5 ☐ ☐ ☐ Ultrapassagem pela direita
6 ☐ ☐ ☐ Mudança de direção para a esquerda
7 ☐ ☐ ☐ Mudança de direção para a direita
8 ☐ ☐ ☐ Marcha atrás
9 ☐ ☐ ☐ Circulação em sentido oposto ao estabelecido
10 ☐ ☐ ☐ Travagem brusca
11 ☐ ☐ ☐ Parado ou estacionado
12 ☐ ☐ ☐ Inversão do sentido de marcha
13 ☐ ☐ ☐ Trânsito em filas paralelas
14 ☐ ☐ ☐ Mudança de via de trânsito para a esquerda
15 ☐ ☐ ☐ Mudança de via de trânsito para a direita
16 ☐ ☐ ☐ Desvio brusco/saída de fila de trânsito
17 ☐ ☐ ☐ Atravessando a via

2. ESQUEMA ☐☐ (Ver esquema em anexo)

E4 INFORMAÇÃO COMPLEMENTAR A AÇÕES E MANOBRAS

A B C

1 ☐ ☐ ☐ Desrespeito da sinalização vertical
2 ☐ ☐ ☐ Desrespeito das marcas rodoviárias
3 ☐ ☐ ☐ Desrespeito da sinalização semafórica
4 ☐ ☐ ☐ Manobra irregular
5 ☐ ☐ ☐ Velocidade excessiva para as condições existentes
6 ☐ ☐ ☐ Não sinalização da manobra
7 ☐ ☐ ☐ Desrespeito das distâncias de segurança
8 ☐ ☐ ☐ Circulação afastada da berma ou passeio
9 ☐ ☐ ☐ Rebentamento pneumático
10 ☐ ☐ ☐ Queda de carga ou objeto
11 ☐ ☐ ☐ Falha mecânica do veículo
12 ☐ ☐ ☐ Ausência de luzes quando obrigatórias
13 ☐ ☐ ☐ Obstáculo imprevisto na faixa de rodagem
14 ☐ ☐ ☐ Abertura de porta
15 ☐ ☐ ☐ Encandeamento
16 ☐ ☐ ☐ Não identificada

E5 ACESSÓRIOS DE SEGURANÇA

A B C

1 ☐ ☐ ☐ Capacete
2 ☐ ☐ ☐ Cinto de segurança
3 ☐ ☐ ☐ Sem uso de cinto/capacete
4 ☐ ☐ ☐ Isento

F - CONSEQUÊNCIAS DO ACIDENTE

F1 CONDUTORES VÍTIMAS

1. GRAU DE GRAVIDADE DAS LESÕES

A B C

1 ☐ ☐ ☐ Morto
2 ☐ ☐ ☐ Ferido grave
3 ☐ ☐ ☐ Ferido leve

F2 PASSAGEIROS VÍTIMAS

Veículo A Veículo B Veículo C

1. SEXO

a b c d | i j l m | r s t u

1 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Masculino
2 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Feminino

2. IDADE

a ☐☐ b ☐☐ | i ☐☐ j ☐☐ | r ☐☐ s ☐☐

c ☐☐ d ☐☐ | l ☐☐ m ☐☐ | t ☐☐ u ☐☐

3. POSIÇÃO NO VEÍCULO

a b c d | i j l m | r s t u

1 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ À frente
2 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ À retaguarda
3 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Desconhecido

4. USO DE ACESSÓRIOS DE SEGURANÇA

a b c d | i j l m | r s t u

1 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ C/ capacete/cinto segurança
2 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ C/ sistema retenção de crianças
3 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ S/ uso capacete/cinto segurança
4 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ S/ sistema retenção de crianças

5. GRAU DE GRAVIDADE DAS LESÕES

a b c d | i j l m | r s t u

1 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Morto
2 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Ferido grave
3 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Ferido leve
4 ☐☐☐☐ | ☐☐☐☐ | ☐☐☐☐ Ileso

F3 PEÕES VÍTIMAS

1. SEXO

a b c d

1 ☐☐☐☐ Masculino
2 ☐☐☐☐ Feminino

2. a b c d

1 ☐☐☐☐ Peão isolado
2 ☐☐☐☐ Peões em grupo
3 ☐☐☐☐ Conduzindo à mão velocípedes, carros de crianças ou de deficientes físicos
4 ☐☐☐☐ Deslocando-se sobre patins, trotinetes ou outros

3. IDADE

a ☐☐ b ☐☐ c ☐☐ d ☐☐

4. CONDIÇÕES PSÍCO-FÍSICAS

a b c d

1 ☐☐☐☐ Sem restrições
2 ☐☐☐☐ Com visão deficiente
3 ☐☐☐☐ Com audição deficiente
4 ☐☐☐☐ Com deficiência motora

Influenciada pelo álcool

5 a ☐,☐☐ b ☐,☐☐ c ☐,☐☐ d ☐,☐☐

5. AÇÕES

a b c d

1 ☐☐☐☐ A sair ou entrar num veículo
2 ☐☐☐☐ Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo
3 ☐☐☐☐ Em plena faixa de rodagem
4 ☐☐☐☐ Em trabalhos na via
5 ☐☐☐☐ Atravessando fora da passagem de peões, a menos de 50 m de uma passagem
6 ☐☐☐☐ Atravessando fora da passagem de peões a mais de 50 m de uma passagem ou quando não exista passagem
7 ☐☐☐☐ Atravessando em passagem sinalizada
8 ☐☐☐☐ Atravessando em passagem sinalizada com desrespeito da sinalização semafórica
9 ☐☐☐☐ Em ilhéu ou refúgio na via
10 ☐☐☐☐ Transitando pela direita da faixa de rodagem
11 ☐☐☐☐ Transitando pela esquerda da faixa de rodagem
12 ☐☐☐☐ Transitando pela berma ou passeio

6. UTILIZAÇÃO DE MATERIAL REFLETOR

a b c d

1 ☐☐☐☐ Sim
2 ☐☐☐☐ Não

7. GRAVIDADE DAS LESÕES

a b c d

1 ☐☐☐☐ Morto
2 ☐☐☐☐ Ferido grave
3 ☐☐☐☐ Ferido leve

DATA ____/____/____

Número de boletins utilizados neste acidente ☐☐

Nome ______________________________

(Posto) ______________________________

Appendix B – Database Sample

Accident_ID	Date_Time	Year	Month	Day	Day_Section	Weekday	Weekdays	Season	Street	Latitude	Longitude	Roadway_Type_ID	Roadway_Type
20191651938	02/01/2019 18:00	2019	Jan	2	Afternoon Rush Hour	Wednesday	Weekdays	Winter	Avenida Marechal Craveiro Lopes	38,759862	-9,153388	2	Street
20191649654	03/01/2019 18:00	2019	Jan	3	Afternoon Rush Hour	Thursday	Weekdays	Winter	Avenida da Liberdade	38,718295	-9,144548	2	Street
20191656727	04/01/2019 10:00	2019	Jan	4	Morning	Friday	Weekdays	Winter	Avenida General Norton de Matos	38,759576	-9,154869	2	Street
20191654467	04/01/2019 20:00	2019	Jan	4	Night	Friday	Weekdays	Winter	Avenida General Correia Barreto	38,73683	-9,168621	2	Street
20191794374	06/01/2019 11:00	2019	Jan	6	Morning	Sunday	Weekend	Winter	Avenida da Índia	38,703009	-9,173179	2	Street
20191651891	07/01/2019 17:00	2019	Jan	7	Afternoon Rush Hour	Monday	Weekdays	Winter	Avenida Almirante Gago Coutinho	38,744264	-9,132014	2	Street
20191654379	09/01/2019 01:00	2019	Jan	9	Early morning	Wednesday	Weekdays	Winter	Avenida da República	38,74071	-9,146273	2	Street
20191657679	11/01/2019 05:00	2019	Jan	11	Early morning	Friday	Weekdays	Winter	Avenida Almirante Reis	38,735304	-9,133985	2	Street
20191655925	11/01/2019 16:00	2019	Jan	11	Afternoon	Friday	Weekdays	Winter	Avenida de Ceuta	38,718678	-9,174791	2	Street
20191654477	12/01/2019 15:00	2019	Jan	12	Afternoon	Saturday	Weekend	Winter	Avenida da República	38,74071	-9,146273	2	Street
20191656499	14/01/2019 22:00	2019	Jan	14	Night	Monday	Weekdays	Winter	Avenida Marechal Craveiro Lopes	38,759862	-9,153388	2	Street
20191656769	16/01/2019 19:00	2019	Jan	16	Afternoon Rush Hour	Wednesday	Weekdays	Winter	Avenida da República	38,74071	-9,146273	2	Street
20191656785	18/01/2019 08:00	2019	Jan	18	Morning Rush Hour	Friday	Weekdays	Winter	Avenida da Índia	38,703009	-9,173179	2	Street
20191657868	18/01/2019 08:00	2019	Jan	18	Morning Rush Hour	Friday	Weekdays	Winter	Avenida General Correia Barreto	38,73683	-9,168621	2	Street
20191657775	19/01/2019 10:00	2019	Jan	19	Morning	Saturday	Weekend	Winter	Avenida Marechal Craveiro Lopes	38,759862	-9,153388	2	Street
20191659447	19/01/2019 23:00	2019	Jan	19	Night	Saturday	Weekend	Winter	Avenida da República	38,74071	-9,146273	2	Street
20191657239	21/01/2019 09:00	2019	Jan	21	Morning Rush Hour	Monday	Weekdays	Winter	Avenida Marechal Craveiro Lopes	38,759862	-9,153388	2	Street
20191658438	21/01/2019 17:00	2019	Jan	21	Afternoon Rush Hour	Monday	Weekdays	Winter	Avenida Marechal Craveiro Lopes	38,759862	-9,153388	2	Street
20191658427	23/01/2019 16:00	2019	Jan	23	Afternoon	Wednesday	Weekdays	Winter	Avenida da Liberdade	38,719087	-9,145237	2	Street
20191658344	27/01/2019 13:00	2019	Jan	27	Lunch Rush Hour	Sunday	Weekend	Winter	Avenida Almirante Reis	38,735304	-9,133985	2	Street
20191659427	28/01/2019 11:00	2019	Jan	28	Morning	Monday	Weekdays	Winter	Avenida Almirante Reis	38,735304	-9,133985	2	Street
20191659429	28/01/2019 14:00	2019	Jan	28	Afternoon	Monday	Weekdays	Winter	Avenida Marechal Francisco da Costa Gomes	38,737368	-9,126179	2	Street
20191659425	29/01/2019 18:00	2019	Jan	29	Afternoon Rush Hour	Tuesday	Weekdays	Winter	Avenida da República	38,74071	-9,146273	2	Street
20191659710	30/01/2019 12:00	2019	Jan	30	Lunch Rush Hour	Wednesday	Weekdays	Winter	Avenida Almirante Reis	38,735304	-9,133985	2	Street
20191659558	31/01/2019 00:00	2019	Jan	31	Early morning	Thursday	Weekdays	Winter	Avenida Almirante Reis	38,735304	-9,133985	2	Street

Parish	Location	Nature	Injuries	Sex	Driver_Age	Group_Age	Number_serious_injured	Number_slightly_injured	Number_Dead	GI	SI
Alvalade	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Female	53	50-59	0	1	0	11088	3,28
Santo António	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Female	38	30-39	0	1	0	11088	1,17
Lumiar	Within Locations	Side collision with another moving vehicle	Slightly injured	Male	53	50-59	0	1	0	11088	2,79
Benfica	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Male	59	50-59	0	1	0	11088	2,79
Belém	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Male	48	40-49	0	1	0	11088	1,38
Alvalade	Within Locations	Side collision with another moving vehicle	Slightly injured	Male	33	30-39	0	1	0	11088	0,78
Arroios	Within Locations	Side collision with another moving vehicle	Slightly injured	Male	45	40-49	0	5	0	11088	1,3
Arroios	Within Locations	Side collision with another moving vehicle	Slightly injured	Male	74	70-74	0	1	0	11088	1,39
Alcântara	Within Locations	Crash in chain	Slightly injured	Female	63	60-64	0	2	0	11088	1,87
Avenidas Novas	Within Locations	Pedestrian collision	Slightly injured	Female	44	40-49	0	1	0	11088	1,3
Lumiar	Within Locations	Crash in chain	Slightly injured	Male	30	30-39	0	1	0	11088	3,28
Avenidas Novas	Within Locations	Collision with other situations	Slightly injured	Male	38	30-39	0	1	0	11088	1,3
Belém	Within Locations	Crash in chain	Slightly injured	Female	30	30-39	0	1	0	11088	1,38
Benfica	Within Locations	Crash with a restraint device	Slightly injured	Male	27	20-29	0	1	0	11088	0,66
Lumiar	Within Locations	Side collision with another moving vehicle	Slightly injured	Female	47	40-49	0	2	0	11088	3,28
Avenidas Novas	Within Locations	Collision with other situations	Slightly injured	Male	42	40-49	0	7	0	11088	1,3
Alvalade	Within Locations	Crash in chain	Slightly injured	Male	76	74-79	0	1	0	11088	3,28
Alvalade	Within Locations	Crash with a restraint device	Slightly injured	Male	68	65-69	0	1	0	11088	3,28
Santo António	Within Locations	Collision with other situations	Slightly injured	Male	26	20-29	0	1	0	11088	1,17
Arroios	Within Locations	Collision with vehicle or obstacle in the lane	Slightly injured	Male	34	30-39	0	1	0	11088	1,39
Arroios	Within Locations	Side collision with another moving vehicle	Slightly injured	Male	44	40-49	0	1	0	11088	1,39
Penha de França	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Female	38	30-39	0	1	0	11088	0,7
Avenidas Novas	Within Locations	Collision with other situations	Slightly injured	Male	86	85-89	0	1	0	11088	1,3
Arroios	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Male	70	70-74	0	2	0	11088	1,39
Arroios	Within Locations	Rear-end collision with another moving vehicle	Slightly injured	Male	27	20-29	0	1	0	11088	1,39

At_Factors	Luminosity	Surface_cond	Driver_Action	Cause_of_accident	Vehicle_Category	Vehicle_Age	Roadway_specs
Good Weather	Night, with illumination	Dry and clean pavement	In Normal Operation	Unidentified	Light Vehicle	14	Other lane
Good Weather	Night, with illumination	Dry and clean pavement	Normal driving	Excessive speed for the existing conditions	Light Vehicle	19	Road without separator
Good weather	In broad daylight	Dry and clean pavement	Sharp detour/leaving a traffic queue	Non-respect of safety distances	Light Vehicle	7	Road without separator
Good weather	Night, with illumination	Dry and clean pavement	Normal operation	Not identified	Light Vehicle	14	Other lane
Good weather	In broad daylight	Dry and clean pavement	Normal operation	Not Identified	Light Vehicle	19	Road without separator
Good Weather	Night, with illumination	Dry and clean pavement	Travel Start	Not identified	Motorcycle <= 125cc	11	Road without separator
Good Weather	Night, with illumination	Dry and clean pavement	Normal Operation	Not identified	Light Vehicle	3	Road without separator
Good Weather	Night, with illumination	Dry and clean pavement	Normal Operation	Non-respect of traffic signalization	Light Vehicle	2	Other lane
Good Weather	In broad daylight	Dry and clean pavement	Stopped or parked	Not identified	Light Vehicle	9	Other lane
Good Weather	Broad daylight	Dry and clean pavement	Normal operation	Traffic signalization disrespect	Light Vehicle	19	Other lane
Good Weather	Night, with illumination	Dry and clean pavement	Normal operation	Not identified	Light Vehicle	4	Other track
Good Weather	Night, with illumination	Dry and clean pavement	Normal operation	Not identified	Motorcycle > 125cc	3	Other track
Good Weather	In broad daylight	Dry and clean pavement	Normal operation	Not identified	Light Vehicle	5	Road without separator
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Excessive speed for the existing conditions	Motorcycle <= 125cc	2	Other lane
Rain	Broad daylight	Wet pavement	Normal operation	Not identified	Light Vehicle	18	Other track
Good weather	Night, with illumination	Dry and clean pavement	Changing traffic lane to the right	Not identified	Light Vehicle	4	Other track
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Not Identified	Light Vehicle	3	Other track
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Not Identified	Light Vehicle	9	Other track
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Not Identified	Light Vehicle	21	Road without separator
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Excessive speed for the existing conditions	Light Vehicle	11	Other lane
Good Weather	Broad daylight	Dry and clean pavement	Exiting a parking lot or private road	Irregular maneuver	Light Vehicle	9	Other track
Good weather	Broad daylight	Dry and clean pavement	Normal operation	Not Identified	Bicycle	NULL	Other track
Rain	Night, with lighting	Wet pavement	Normal operation	Not Identified	Light Vehicle	13	Other track
Rain	Broad daylight	Wet pavement	Normal operation	Excessive speed for the existing conditions	Light Vehicle	17	Other track
Rain	Night, with illumination	Wet pavement	Normal operation	Not identified	Mopped	NULL	Other route

Roadway_st_of_cons	Roadway_Int	Obstacles	Directions	Traffic_signs	Traffic_lights	Pavement_type	Roadway_geometry1	Roadway_geometry2
In good condition	Outside the intersection	Non-existent	In decreasing order of mileage	Other	Nonexistent	Bituminous	Straight	On a slope
In good condition	At junction	Non-existent	NOT SET	NOT DEFINED	Functioning normally	Bituminous	Straight	With inclination
In fair condition	At junction	Non-existent	NOT SET	Others	Functioning normally	Bituminous	Straight	At level
In fair condition	Outside the intersection	Non-existent	Mileage increasing	Others	Not available	Bituminous	Straight	Break-even
In fair condition	Outside the intersection	Nonexistent	NOT SET	Others	None	Bituminous	Straight	Break-even
In fair condition	NOT DEFINED	Nonexistent	NOT SET	Others	Working normally	Bituminous	Curve	Break-even
In fair condition	At a junction	Nonexistent	NOT SET	NOT DEFINED	Working normally	Bituminous	Straight	At break-even
In fair condition	At a junction	Nonexistent	NOT SET	Others	Working normally	Bituminous	Straight	With steep inclines
In fair condition	Outside Intersection	Nonexistent	Mileage decreasing	Others	None	Bituminous	Straight	At break-even
In fair condition	Outside Intersection	Nonexistent	Mileage increasing	Others	Working normally	Bituminous	Straight	break-even
In fair condition	Outside Intersection	Correctly signposted	Mileage decreasing	Others	None	Bituminous	Straight	Break-even
In fair condition	Intersection	Non-existent	NOT SET	Others	Working normally	Bituminous	Straight	Break-even
In fair condition	Outside Intersection	Non-existent	NOT SET	Other	Not available	Bituminous	Straight	Break-even
In fair condition	Outside Intersection	Non-existent	Mileage increasing	Other	Non-existent	Bituminous	Straight	Lows
In fair condition	Outside Intersection	Non-existent	Mileage increasing	Other	None	Bituminous	Straight	Lows
In good condition	At intersection	Non-existent	Mileage decreasing	Other	Functioning normally	Bituminous	Straight	Lows
In good condition	At Junction	Nonexistent	Mileage increasing	Others	Nonexistent	Bituminous	Straight	Lows
In fair condition	Outside of the intersection	Nonexistent	Mileage decreasing	NOT DEFINED	Nonexistent	Bituminous	Straight	break-even
In fair condition	Outside the intersection	Nonexistent	NOT SET	Others	Normally functioning	Bituminous	Straight	With steep inclines
In fair condition	At intersection	Correctly signposted	Mileage decreasing	Crosswalk	Working normally	Bituminous	Straight	At levels
In fair condition	Outside of the intersection	Non-existent	In the downward direction of the mileage	NOT DEFINED	Working normally	Bituminous	Straight	At tilt
In fair condition	Outside the intersection	Non-existent	Mileage decreasing	Other	None	Bituminous	Straight	At tilt
In fair condition	Intersection	Non-existent	Mileage increasing	Pedestrian crossing	Working normally	Bituminous	Straight	At level
In fair condition	Intersection	Non-existent	Mileage decreasing	Other	Working normally	Bituminous	Straight	At tilt
In fair condition	Outside Intersection	Nonexistent	Increasing the mileage	Others	Working normally	Bituminous	Straight	At level

Roadway_geometry3	Roadway_geometry4	FRC	SpeedLimit	AvgSp	HvgSp	MedSp	Air_Temp	Thermal_sen	T_min	T_max	WW_H	WW_L	At_pressure	Rel_hum	Wind speed
Paved berm	On the roadway	2	80	67,79	65,69	63,51	15	13	13	18	NULL	NULL	1020	29	22
No shoulder or impassable	On the roadway	4	50	29,22	19,55	23,74	16	15	15	17	NULL	NULL	1017	28	22
Paved berm	On the roadway	5	50	28,03	29,09	27,43	15	14	13	17	NULL	NULL	1017	28	20
Unpaved berm	On the roadway	4	80	78,7	79,8	71,29	13	13	11	15	NULL	NULL	1015	28	19
No berm or impassable	On the roadway	3	50	64,01	64,88	61,7	12	12	11	14	NULL	NULL	1018	27	19
Paved berm	On the roadway	4	50	47,13	37,91	39,32	11	10	10	13	NULL	NULL	1018	26	19
Paved berm	On the roadway	4	50	57,03	53,07	45,6	11	9	9	12	NULL	NULL	1022	26	18
No berm or impassable	On the roadway	4	50	51,39	42,44	47,3	10	8	9	11	NULL	NULL	1022	26	17
Paved berm	On the roadway	4	50	47,49	44,28	41,5	14	11	12	16	NULL	NULL	1026	26	20
No berm or impassable	On the roadway	4	50	35,63	28,29	26,19	11	8	10	13	NULL	NULL	1027	15	9
Paved roadside	On the roadway	4	80	78,2	79,27	76,64	9	6	8	11	NULL	NULL	1026	14	9
No berm or impassable	On the roadway	4	50	13,66	8,69	9,43	9	6	8	11	NULL	NULL	1025	14	9
No berm or impassible	On the roadway	3	50	66,3	66,74	63,94	12	8	10	14	NULL	NULL	1024	14	9
Paved berm	On the roadway	4	80	83,06	86,16	84,13	13	11	12	15	NULL	NULL	1032	11	9
Paved berm	On the roadway	2	80	68,5	68,02	60,74	12	10	11	13	NULL	NULL	1031	10	8
Paved berm	On the roadway	5	50	43,1	43,57	38,84	11	9	10	14	NULL	NULL	1031	10	7
Paved berm	On the roadway	2	80	59,1	58,88	47,21	10	7	7	13	NULL	NULL	1022	10	8
Paved berm	On the roadway	2	80	60,09	58,22	57,81	13	11	12	16	NULL	NULL	1011	96	9
No berm or impassable	On the roadway	4	50	25,43	16,2	17,43	9	6	7	12	NULL	NULL	1029	97	10
Paved roadside	On the roadway	4	50	31,9	32,18	29,4	9	6	8	12	NULL	NULL	1029	98	11
No berm or impassable	On the roadway	4	50	32,9	32,83	30,47	11	8	9	13	NULL	NULL	1021	75	11
Paved roadside	On the roadway	4	50	37,34	26,56	22,4	11	7	8	13	NULL	NULL	1021	77	5
No berm or impassable	On the roadway	4	50	16,65	10,86	10,87	14	9	13	15	NULL	NULL	1015	96	5
Paved berm	On the roadway	4	50	39,5	37,29	28,09	13	13	13	15	NULL	NULL	1015	98	11
Paved berm	On the roadway	4	50	48,99	36,68	47,9	13	9	12	14	NULL	NULL	1011	96	9

Wind Classification	Beaufort	Wind_Direction	Wind_Direction_card	Max_WG	Precipitation1h	Precipitation _Class_1	Precipitation3h	WW_P	Cloud cover	weather_main	weather_description
Moderate wind	Gentle breeze	15	N	51	NULL	NULL	NULL	NULL	1	the sky is clear	scattered clouds
Moderate wind	Gentle breeze	360	N	43	NULL	NULL	NULL	NULL	0	mist	scattered clouds
Moderate wind	Gentle breeze	360	N	43	NULL	NULL	NULL	NULL	0	fog	sky is clear
Fair wind	Light breeze	360	N	43	NULL	NULL	NULL	NULL	0	mist	sky is clear
Moderate wind	Gentle breeze	15	N	41	NULL	NULL	NULL	NULL	0	the sky is clear	scattered clouds
Moderate wind	Gentle breeze	80	E	39	NULL	NULL	NULL	NULL	0	mist	scattered clouds
Moderate wind	Gentle breeze	80	E	39	NULL	NULL	NULL	NULL	0	mist	sky is clear
Moderate wind	Gentle breeze	10	N	38	NULL	NULL	NULL	NULL	2	the sky is clear	broken clouds
Moderate wind	Gentle breeze	10	N	38	NULL	NULL	NULL	NULL	1	the sky is clear	scattered clouds
Light wind	Light breeze	359	N	29	NULL	NULL	NULL	NULL	2	the sky is clear	scattered clouds
Light wind	Light breeze	3	N	29	NULL	NULL	NULL	NULL	0	the sky is clear	sky is clear
Light wind	Light breeze	3	N	29	NULL	NULL	NULL	NULL	1	the sky is clear	sky is clear
Light wind	Light breeze	3	N	29	NULL	NULL	NULL	NULL	1	the sky is clear	sky is clear
Weak wind	Light air	310	NW	28	NULL	NULL	NULL	NULL	0	the sky is clear	sky is clear
Weak wind	Light air	1	N	28	NULL	NULL	NULL	NULL	0	the sky is clear	moderate rain
Weak wind	Light air	354	N	26	NULL	NULL	NULL	NULL	0	the sky is clear	sky is clear
Weak wind	Light air	1	N	23	NULL	NULL	NULL	NULL	0	the sky is clear	sky is clear
Moderate wind	Gentle breeze	80	E	39	NULL	NULL	NULL	NULL	0	sky is clear	sky is clear
Moderate wind	Gentle breeze	125	SE	19	NULL	NULL	NULL	NULL	38	Cloudy	scattered clouds
Moderate wind	Gentle breeze	131	SE	20	NULL	NULL	NULL	NULL	5	clear	sky is clear
Light wind	Light breeze	169	S	12	NULL	NULL	NULL	NULL	99	cloud	overcast clouds
Light wind	Light breeze	168	S	11	NULL	NULL	NULL	NULL	100	Cloudiness	overcast clouds
Light wind	Light breeze	272	W	12	3.5	Moderate	NULL	NULL	100	Rain	moderate rain
Light wind	Light breeze	250	W	13	2.68	Moderate	NULL	NULL	40	Rain	light rain
Weak wind	Light air	311	NW	11	0.6	Moderate	NULL	NULL	100	Rain	light rain

Appendix C – Experts examination form

Experts first interaction

*Obrigatório

1. Is this research is efficient? *

Marcar apenas uma oval.

- Yes
- No

2. The Identification of the location of RTAs hotspots is efficient to assist the process of decision support of CML's civil protection and mobility officers *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

3. The research respond to the CML needs? *

Marcar apenas uma oval.

- Yes
- No

4. The research respond totaly to the needs identified in challenge 51 *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

5. The Detail level of the research is suitable *

Marcar apenas uma oval.

- Yes
- No

6. The proposed solution provides the level of detail required for the decision support process *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

7. The applied methodologies for determining the hotspots are the correct ones? *

Marcar apenas uma oval.

- Yes
- No

8. The solution complies with the latest methods in the literature for hotspot determination *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

9. The analysis carried out is useful (Utility)? *

Marcar apenas uma oval.

- Yes
- No

10. The solution is an added value for the reduction of traffic accidents with victims in the city of Lisbon *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

11. The solution has applicability? *

Marcar apenas uma oval.

- Yes
- No

12. The solution has practical applicability in the field on road safety and civil protection? *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

13. The solution is consistent? *

Marcar apenas uma oval.

- Yes
- No

14. The solution is consistent throughout its analysis? *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definiely

15. The solution is transferable (Transferability)? *

Marcar apenas uma oval.

- Yes
- No

16. The results of the study can be applied to other contexts or areas? *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely

17. The solution is reproducible in academia? *

Marcar apenas uma oval.

- Yes
- No

18. The methodology and processes employed in the study allow external researchers to follow the same method used by the researcher? *

Marcar apenas uma oval.

	1	2	3	4	5	
Not						Definitely