energies

MDPI

Article

Smart Cities: Data-Driven Solutions to Understand Disruptive Problems in Transportation—The Lisbon Case Study

Vitória Albuquerque [1], Ana Oliveira [2], Jorge Lourenço Barbosa [2], Rui Simão Rodrigues [2], Francisco Andrade [2], Miguel Sales Dias [1,2] and João Carlos Ferreira [2,3,*]

[1] NOVA Information Management School (NOVA IMS), Campus de Campolide, Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal; d20190115@novaims.unl.pt
[2] ISTAR-IUL, Instituto Universitário de Lisboa (ISCTE-IUL), 1649-026 Lisbon, Portugal; Ana_Melanda@iscte-iul.pt (A.O.); Jorge_Lourenco_Barbosa@iscte-iul.pt (J.L.B.); Rui_Simao_Rodrigues@iscte-iul.pt (R.S.R.); Francisco_Antonio_Andrade@iscte-iul.pt (F.A.); miguel.dias@iscte-iul.pt (M.S.D.)
[3] Inov Inesc Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal
* Correspondence: joao.carlos.ferreira@iscte-iul.pt

Abstract: Transportation data in a smart city environment is increasingly becoming available. This data availability allows building smart solutions that are viewed as meaningful by both city residents and city management authorities. Our research work was based on Lisbon mobility data available through the local municipality, where we integrated and cleaned different data sources and applied a CRISP-DM approach using Python. We focused on mobility problems and interdependence and cascading-effect solutions for the city of Lisbon. We developed data-driven approaches using artificial intelligence and visualization methods to understand traffic and accident problems, providing a big picture to competent authorities and supporting the city in being more prepared, adaptable, and responsive, and better able to recover from such events.

Keywords: transportation; traffic; accidents; data-driven; data visualization; smart cities

Citation: Albuquerque, V.; Oliveira, A.; Barbosa, J.C.; Rodrigues, R.S.; Andrade, F.; Dias, M.S.; Ferreira, J.C. Smart Cities: Data-Driven Solutions to Understand Disruptive Problems in Transportation—The Lisbon Case Study. *Energies* **2021**, *14*, 3044. https://doi.org/10.3390/en14113044

Academic Editor: Luis Hernández-Callejo

Received: 12 April 2021
Accepted: 18 May 2021
Published: 24 May 2021

1. Introduction

Transportation infrastructures (TIs) have a pivotal role in ensuring citizens' liveability, safety, security, and health in urban settings. In the environment, modern critical infrastructures have become more competent in the way they operate, function, and interact with citizens, customers, and between each other, leading to the birth of smart cities (SCs). The smart city can now be defined as a complex network of technologically advanced critical infrastructures connected in a digital environment, characterized by a massive and increasing presence of IoT objects and underlying technologies [1]. A TI becoming smarter involves regular operation and use, making it more adaptive, more intelligent, and more connected. Moreover, in the case of excessive optimization, it could also make TIs more vulnerable and subject to cascading effects, and therefore less resilient. Following recent disruptive human-made and natural events, including the COVID-19 pandemic, it has become clear that the smart city itself is not sufficient to protect citizen life, and shall move to an upgraded and more secure version of itself: the resilient city [1].

The population inhabiting metropolitan areas around the world is increasing at an alarming rate. In 2008, 50% of the world's population lived in urban areas, and was growing exponentially. By 2050 [2], it is expected that 70% of the world's population will live in metropolitan regions. Due to this rapid population growth, cities will face new challenges [3], such as increased waste, pollution, traffic congestion, and road accidents.

Traffic management is a topic the research community has been tackling for more than 40 years [4]. Cities' policymakers approach this problem by integrating new

technologies in their solutions, such as sensors, video images, microwave radar, infrared sensors, laser sensors, audio sensors, and Global positioning System (GPS) sensors equipped in smartphones and vehicles, which provide large quantities of data. With appropriate data science methodologies, scientists can now predict real-time traffic trajectories and patterns based on historical data [5].

The availability of such data sources with open access has led to new data-driven approaches, cost-effective mobile services, and applications that stand as the cornerstone of intelligent transportation systems (ITSs). Road traffic congestion is a problem that leads to delays, energy consumption, and environmental pollution [6]. ITSs can improve citizens' sustainable urban mobility and quality of life, providing solutions to pollution and traffic congestion, promoting reduction of road accidents [7] and energy consumption. ITSs can contribute to key public policy and systematic management, reducing traffic congestion and energy consumption in urban areas.

The World Health Organization (WHO) estimates road accidents are the ninth leading cause of death globally across all age groups, and are the main cause of death among people aged between 15 and 29 years. Road accidents are a cause of life losses, and bring health and socioeconomic costs [8].

All over the globe, countries are adopting measures to decrease road fatalities. Speed management, infrastructure design and improvement, enforcement of traffic laws, leadership on road safety, vehicle safety standards, and post-crash survival are some of the currently ongoing initiatives that aim to mitigate this hazard [8]. A large number of road accidents occur due to various factors that directly or indirectly affect conditions on the road for drivers, passengers, and pedestrians [9]. Factors such as gender and age, or environmental factors such as low brightness (dawn or dusk) and adverse weather conditions [10], influence the world's global number of road accidents.

In Portugal, drivers spend a daily average of 42 min in urban traffic and an average of 160 h each year in traffic jams [11]. Traffic congestion is one of the causes of road accidents. This is considered one of the most serious problems in today's Portuguese society and a public health issue. According to Autoridade Nacional de Segurança Rodoviária (ANSR) [12], compared to 2018, the number of accidents with victims increased by 4% in 2019, with a reduction of 9% in the number of fatalities. Despite this fall in fatalities, serious injuries increased by 9%. The geographic distribution of road traffic victims in 2019 shows that Lisbon and Porto have approximately 40% of the country's total number of victims.

Our research aim was to analyze and visualize data of two traffic phenomena: traffic congestion and road accidents, with Lisbon as a case study. Lisbon is the capital of Portugal, with a population of 508,368, and 2 million people (20% of Portugal's population) living in the urban metropolitan area commute to Lisbon. Around 370,000 vehicles enter Lisbon every day, adding to the 200,000 vehicles that already circulate in the city. We investigated and constructed an overview of traffic congestion and road accidents in Lisbon by analyzing 2019 data. Our main research goal was to correlate traffic and road accidents, identifying how people move in a city and how weather conditions affect traffic congestion and road accidents, with a multivariable analysis and visualization. Furthermore, our research goal aimed to provide data-driven guidelines and knowledge about traffic and road accidents in Lisbon to the city authorities and policymakers in the framework of a traffic management and visualization tool to help them mitigate such phenomena.

Our study addresses the following research questions:

- RQ1: How can we characterize Lisbon's road traffic patterns?
- RQ2: How can we characterize road accident patterns in Lisbon, and what are the external factors that contribute the most to this phenomenon?
- RQ3: With a better understanding of Lisbon's traffic and road accident patterns, how can we identify how citizens move in Lisbon with the help of data visualization?

The paper is structured as follows: Section 2 presents our literature review. In Section 3, Data-Driven Solution for Decision Support, we introduce the adopted cross-industry standard process for data mining (CRISP-DM) methodology and the application to our Lisbon case study, particularly the business and data understanding, data preparation and fusion, and data visualization phases. In Section 4, we provide a comparative analysis of traffic and road accidents, discuss our findings and compare them to the literature review, and analyze our approach to addressing the research questions. We also identify research gaps and the limitations of our research. Finally, in Section 5, we present conclusions and draw lines for further research.

2. Literature Review

Our literature survey covered traffic congestion and road accidents in urban areas. We found just a few articles in these fields, with a focus on data analysis and visualization of traffic congestion and road accident. On the other hand, we found that most articles used prediction models to address how external factors influenced traffic and road accidents in cities.

Studies [7,8,11,13] on road accidents in Spain, India, and the United States of America (Washington) showed different approaches and results. Palazón-Bru [10] studied Spain road accident data from 2015, and considered variables associated with the accident, the vehicle, and individuals. This information could be used by both police authorities and health services to make predictions and determine where to undertake possible interventions to reduce death risk. Factors like not using a seat belt; unfavorable lighting; interurban roads where greater speed is reached; and small vehicles driving alongside buses or trucks, which, in the event of a small vehicle rollover, could cause its occupants' death, were the more prevalent. The authors concluded that individuals younger than 60 years had lower mortality rates.

In India, road accidents were analyzed [9], and the authors concluded that the period between 3 p.m. and 6 p.m. corresponded to the peak traffic hours, but this changed between states in India. The study [9] showed that two-wheeled vehicles are involved in accidents more, and over 80% that occurred were the driver's fault. These authors also concluded that accident severity was growing due to the increase in the number of vehicles. Another interesting result was that 61% of accidents occurred on the weekend (Friday and Saturday), which correlated well with alcohol consumption.

Combining data analytics with data visualization allows effective communication of research insights and provides visualization tools to policymakers and public authorities. According to Chen [14], a pipeline of traffic visualization is an adequate tool to assess traffic data properties and discover hidden patterns in the data. Chen [14] proposed an analysis based on four factors: temporal, spatial, spatio-temporal, and multivariable.

Traffic congestion analysis and visualization combined with factors such as car trajectories in peak hours [15] has proven to be an interesting tool to understand mixed traffic conditions. A multivariable analysis of vehicle types and car flow models improves mixed traffic-flow trajectories.

Multivariable analysis has shown effectiveness in communicating data insights by providing visualizations of road accidents within a time range and city areas [16] where they most occur. Moreover, it can be combined with data regarding type of vehicle, lighting conditions, weather conditions, month of the year, and day of the week.

Other factors can be analyzed in a multivariable analysis and visualization regarding road conditions [17] and how they affect traffic congestion and road accidents. Tools such as RetinaNet enhance results by training images and providing better performance metrics of road conditions.

The impacts of traffic congestion on road accidents are a common study field, as congestion can trigger road accidents. Studies [18,19] showed the application of Poisson–Gamma models to modeling road accidents. Wang [19] specifically looked at this correlation to measure congestion using Poisson nonspatial and spatial models. Aguero-

Valverde [20] looked at road-accident frequency with spatial models using a Bayesian hierarchical approach to identify spatial correlations.

Visual analytics provides the ability to analyze several tasks, such as traffic congestion detection, accident monitoring, and flow-pattern recognition. These tools are essential to what constitutes an intelligent transportation system (ITS). According to Zhang [21], having access to large quantities of data obtained through multiple sensor sources facilitates the development of a data-driven ITS based on vision, multisource, and machine algorithms. Such a visual-analytics approach is particularly efficient at predicting and managing traffic flows in urban areas.

Using visual analytics to aid in the interpretation of forecasting models is becoming the norm. Andrienko [22] used visual-analysis techniques to study the relationship between traffic intensities and speeds practiced on the roads by using mathematical models. These models could be used to predict everyday traffic situations, as well as to simulate future events. According to this author, visualization, as a tool for prediction, should be developed evolutionarily and iteratively through a cycle of data analysis, model development, and analysis of predictions.

As mentioned at the beginning of this section, most of the analyzed traffic and road accident papers presented their analysis and prediction methodologies combined. For this reason, we will also discuss how the scientific community has addressed traffic and road accident prediction in their studies.

The research community in traffic frequently uses two specific machine-learning techniques for predicting traffic flow. Those are the long short-term memory (LSTM) and recurrent neural networks (RNNs).

ITS systems are considered essential tools for mitigating automobile traffic, and because of that, providing a good forecast of automobile flow is essential for this system's effectiveness. Several authors have proposed hybrid models for better traffic prediction in the city due to conventional machine-learning models' weak adaptability.

A specific traffic-flow-prediction framework was proposed by Xiangxue [23] by combining LSTM and RNN models to evaluate two urban networks. The results showed that this approach brought better quality than other machine-learning models.

Current prediction models have problems such as poor stability, significant data needs, and poor adaptability. Liu [24] defined a model derived from the RNN that combined long short-term memory (SDLSTM) with auto-regressive integrated moving average (ARIMA). The end result presented excellent adaptability and higher precision than common machine-learning methods. This hybrid model integrated computer vision and machine learning in a cloud-computing environment.

Zheng [25] used another LSTM model and compared it with the conventional machine-learning models such as ARIMA and BPNN (back-propagation neural network), and obtained substantially superior results.

Regarding road accidents, Norros [26] found that factors such as weather, traffic, and location could be included as variables when creating models to predict road accidents. Driver characteristics do not depend on the time of day, traffic, or weather, but weather and other external conditions affect traffic intensity and constitution.

Tang [13] used the Washington Incident Tracking System and applied a combined analysis using eight methods: four statistical methods (accelerated failure time (AFT), finite mixture (FM), random parameters hazard-based duration (RPHD), and quantile regression (QR); and four machine-learning methods (K-nearest neighbor (KNN), support vector machine (SVM), BPNN, and random forest (RF)) used in traffic-incident clearance-time analysis. All showed that temporal factors like day of the week, time of day, and month of year influence accidents.

Table 1 summarizes our literature review analysis, with a focus on the papers' dimensions and applications.

Table 1. Literature review summary.

Reference	Dimension	Application
[7]	Multivariable analysis in three levels: sensor networks formed by vehicles, cognitive management functionality placed inside the vehicles, and cognitive management functionality in the overall transportation infrastructure	Provide knowledge to vehicles, and manage traffic and safety
[9]	Multivariable analysis of city names, the type of accident, condition of light, severity, speed zone, consumption of alcohol, time and day of the accident, etc.	Identify patterns to take appropriate measures to reduce the risk of loss of lives and occurrence of accidents on roads
[10]	Multivariable analysis with variables associated with accidents (weekend, time, number of vehicles, road, brightness, and weather), vehicle (type and age of vehicle, and other types of vehicles in the accident), and individuals (gender, age, seat belt, and position in the vehicle)	Predictive model to help determine the probability of mortality
[13]	Eight methods for predicting incident clearance time, including four statistical models: accelerated failure time (AFT), quantile regression (QR), finite mixture (FM), and random parameters hazard-based duration (RPHD); and four machine-learning models: K-nearest neighbor (KNN), support vector machine (SVM), back propagation neural network (BPNN), and random forest (RF)	"Heterogeneity" models performed better than statistical models
[14]	Multivariable data-processing techniques depicting the temporal, spatial, numerical, and categorical properties of traffic data through visualization	Identify transport mobility patterns with traffic data visualizations
[15]	Multivariable analysis: six vehicle categories using various stability criteria, evaluated for mixed traffic conditions	Identify traffic-flow behavior
[16]	Multivariable analysis and visualization of geospatial data (Tableau and GeoCharts)	Identify the effects of traffic public policies using data-visualization methods
[17]	Road-conditions assessment with RetinaNet image processing	Identify road conditions with imaging training
[18]	Poisson–Gamma models for modeling motor vehicle crashes: a Bayesian perspective	Model road accidents

[19]	Poisson-based nonspatial (such as Poisson–lognormal and Poisson–Gamma) and spatial (Poisson–lognormal with conditional autoregressive priors) models	Identify spatial correlation in traffic congestion and road accidents
[20]	Bayesian hierarchical approach was used with conditional autoregressive effects for the spatial correlation terms	Identify effect of spatial correlation in models of road crash frequency at the segment level
[21]	Multivariable analysis of functionality and deployment	Data-driven intelligent transport systems performance optimization
[22]	Traffic simulations based on spatially abstracted transportation networks using dependency models derived from real traffic data	Identify correlations between the traffic intensity and movement speed on links of a spatially abstracted transportation network
[23]	Data-driven short-term data processing and LSTM–RNN	Forecast urban road network traffic
[24]	Traffic flow combination forecasting method based on improved LSTM and ARIMA	Forecast traffic flow
[25]	Traffic flow forecast through time series analysis based on deep learning	Forecast traffic flow
[26]	Palm distribution application to analyze road accident risk assessment	Identify correlations between traffic, road, and weather conditions in road accidents

3. Data-Driven Solution for Decision Support

The Cross-Industry Standard Process for Data Mining (CRISP-DM), which was adopted in our research, is a methodology that aims to create a standard approach to data-mining projects to reduce costs and increase reliability, repeatability, and manageability, making the data-mining process more efficient [27]. CRISP-DM [28] is composed of six phases, and we changed the process toward specific data visualization, taking into account data fusion from different sources (Figure 1). The first and second phases, business understanding and data understanding, are when the initial data are collected, described, explored, and verified. In the third phase, data preparation, data are selected and cleaned, exploring and verifying data quality, integration, and format. In the fourth phase, data fusion, we selected the data sources, like traffic, accidents, weather, city infrastructure, pollution, and data warehouse. In the fifth phase (data visualization), we defined visualization templates to automatically visualize temporal and spatial data to define goals based on municipalities' needs. Finally, the last-step decision and the big picture were provided to competent decision-making authorities.

Figure 1. Our data-driven methodology.

Applying CRISP-DM involved business understanding by looking at the aim of the challenges proposed by the Lisbon City Hall (CML) in the framework of Lisboa Inteligente—LxDataLab [29]. It also required the definition of our strategy to address the research questions. In the data-understanding phase, traffic and road accident data sets were described and categorized in features. This was followed by the data-preprocessing phase, in which the collected data were cleaned and normalized, generating new data sets to be used in the modeling phase comprising analysis and visualization.

3.1. Business Understanding

Our study addressed the challenges of "General traffic index and indexes for the main Lisbon city entrances" [30] and "Identification of road accidents patterns and correlation with external factors" [31], both launched by Lisboa Inteligente—LxDataLab [29] for academia. Our study's objective was to investigate and identify traffic and accident patterns in Lisbon's metropolitan area, define when and where traffic congestion and accidents road occur and how they relate, and how external factors such as weather and pandemics affect such phenomena.

3.2. Data Understanding

The data-understanding phase aimed to collect, describe, explore, and verify the data's quality. This step was structured in three subphases: describe, explore, and verify data quality.

Traffic congestion and road accident data sets were provided and collected from Lisboa Inteligente—LxDataLab [29] on the scope of the challenges launched for academia, as mentioned previously.

3.2.1. Traffic Data Understanding

The data-understanding phase aimed to collect, describe, explore, and verify the quality of the data available. This step contained three subphases (describe, explore, and verify data quality).

The traffic-congestion data set included Waze data [30], preprocessed by Lisboa Inteligente—LxDataLab [29], with data from traffic jams in the Lisbon metropolitan area, in the period of 1 January 2019 to 30 July 2020. The provided data set had a table structure with 25 columns and 12,619,459 rows in the comma-separated values (CSV) file format.

The data set provided had already been preprocessed before extraction. Some columns needed to be removed, and minor alterations and improvements in the data quality were made. Some duplicated columns with no values (Endnode, pubmillis, Roadtype, turnType, turntype, Typeentity) and some columns (Bbox, entity_location, entity_type, fiware_service, fiware_servicepath, and pubMillis) deemed irrelevant by the CML were removed. Additionally, the Country column was removed because all data were from Lisbon (Portugal).

Table 2 shows the Waze data schema, and each column corresponds to the column name, description, the removed columns (excluded columns), and the reasons for removal.

Table 2. Waze data schema/provided by LxDataLab from Waze.

ID	Column Name	Description	Excluded Columns	Reason
1	Bbox	Position start and position end		
2	City	Information from the city and the region		
3	Country	Country information	Excluded	All data were from Portugal (PO)
4	Delay	Delay of jam compared to free-flow speed, in seconds ("−1" in case of a block)		
5	Endnode	Nearest junction, street, city to jam end (supplied when available)		
6	Endnode	Nearest junction, street, city to jam end (supplied when available)	Excluded	Duplicated column
7	Entity_id	No description	Excluded	
8	Entity_location	No description	Excluded	Unused column (client)
9	Entity_ts	Date of the occurrence (UNIX time—milliseconds since epoch)		
10	Entity_type	No description	Excluded	Unused column (client)
11	Fiware_service	No description	Excluded	Unused column (client)
12	Fiware_service path	No description	Excluded	Unused column (client)
13	Length	Jam length in meters		
14	Level	Traffic congestion level (0 = free flow 5 = blocked).		
15	Position	Jam geographical reference in Geojson format		

16	pubMillis	Publication date (UNIX time—milliseconds since epoch) (excluded)	Excluded	Unused column (client)
17	pubmillis	Publication date (UNIX time—milliseconds since epoch) (excluded)	Excluded	Unused column (client)
18	RoadType	Type of road (18 different types of roads)		
19	Roadtype	Type of road (18 different types of roads)	Excluded	Duplicated column
20	Street	Street name (as is written in the database, no canonical form, may be null)		
21	turnType	What kind of turn is it—left, right, exit R or L, continue straight or NONE (no info) (supplied when available)	Excluded	No data
22	turntype	What kind of turn is it—left, right, exit R or L, continue straight or NONE (no info) (supplied when available)	Excluded	No data
23	Typeentity	The field typeEntity (corresponds to the field Type (irregularities))		
24	Typeentity	The field typeEntity (corresponds to the field Type (irregularities))	Excluded	Duplicated column
25	Validity_ts	Date of LxDataLab archive (UNIX time—milliseconds since epoch)		

After removing the columns, an assessment of the data quality was executed on the remaining columns, and although the data was consistent and no significant problems were found, some adjustments had to be made in the next step of the CRISP-DM methodology.

3.2.2. Road Accident Data Understanding

In the scope of the Lisboa Inteligente—LxDataLab [29] challenge 51, the provided data set was a single Excel file with aggregated data from Agência Nacional para a Segurança Rodoviária (ANSR), Polícia de Segurança Pública (PSP), and Guarda Nacional Republicana (GNR), with a list of road accidents (RSB) events in Lisbon involving vehicles and motorcycles from January to December 2019. The Excel file included four sheets with the following structure: accidents (Table 3) with 37 columns and 2768 rows; vehicles involved and their driver (Table 4), with 18 columns and 4834 rows; passengers (Table 5), with eight columns and 631 rows; and pedestrians (Table 6), with seven columns and 700 rows.

Table 3. Accident data schema.

Characteristics	Description
IdAcidente	Accident ID
Datahora	Date and hour
Dia da semana	Day of the week
Sentidos	Upward and downward
Latitude GPS	Latitude
Longitude GPS	Longitude
Via Trânsito	Left, right, or central transitway

Localizações	Inside or outside localities
Freguesia	Parish
Pov. Proxima	Nearby village
Tipo natureza	Runover, collision, or screen
Natureza	Type of runover, collision, or screen
Traçado 1	Straight or curved
Traçado 2	With slope, in level or bump
Traçado 3	With or without roadside
Traçado 4	Place where the accident occurred
Estado de conservação	State of road
Características Tecnicas	Highway or other
Reg Circulação	One or both ways
Marca Via	Marks on the road
Obstáculos	Obstacles
Sinais	Signals
Sinais Luminosos	Light signals
Tipo Piso	Pavement type
Intersecção Vias	Road intersection
Factores Atmosféricos	Weather conditions
Luminosidade	Luminosity
Cond Aderência	Adhesion conditions
VM	No description
FG	No description
FL	No description
Tipos Vias	Type of road
Via	Lane
Num arruamento	Street number
Km	No description
Nome arruamento	Street name
Localização 2	GPS signal

Table 4. Vehicles involved data schema.

Characteristics	Description
IdAcidente	Accident ID
Datahora	Date and hour
Id. Veiculo	Vehicle ID
Categoria Veículos	Vehicle category
Idade	Age
Sexo	Sex
Lesões a 30 dias	Type of injury
Acessórios Condutores	Driver accessories
Acções Condutores	Driver actions
Inf. Comp. a Acções e Manobras	No description
Licença Condução	Driver license
Tempo Condução Continuada	Driving time
Teste Alcool	Alcohol test
Carga Lotação	Freight
Certificado Adr	No description
Inspecção Periódica	Periodic inspection
Seguros	Insurance

Table 5. Passengers' data schema.

Characteristics	Description
IdAcidente	Accident ID
Datahora	Date and hour
Id. Veículo	Vehicle ID
Id. Passageiro	Pedestrian ID
Idade	Age
Sexo	Sex
Lesões a 30 dias	Type of injury in 30 days
Acessórios Passageiro	Passenger accessories

Table 6. Pedestrians' data schema.

Characteristics	Description
IdAcidente	Accident ID
Datahora	Date and hour
Id. Peao	Pedestrian ID
Idade	Age
Sexo	Sex
Lesões a 30 dias	Type of injury in 30 days
Acções Peão	Pedestrian actions

3.3. Data Preparation and Fusion

3.3.1. Traffic Data Preparation and Fusion

This phase of the CRISP-DM methodology was subdivided into four subphases: data selection, data cleaning, feature selection, and data integration. All columns were selected except for Country, Typeentity, and Validity_ts. The Country column had no value, since the data all were from Lisbon, Portugal. The Typeentity column had only five levels: NONE (12,531,806 entries), Small (46,417 entries), Medium (25,794 entries), Large (15,313 entries), and Huge (129 entries), with NONE being the value in almost all of the rows; for that reason, this column was removed. The Validity_ts column also was removed, given that this work's main focus was on the occurrence of each event and not the archive date, so the Entity_ts column was used to the detriment of Validity_ts (date of LxDataLab archive—Table 1).

After the final column selection, all the columns were carefully inspected [32,33], and we made a few improvements in the data quality, as depicted in Table 7.

Table 7. Waze dataset—details and data transformation.

ID	Column	Chosen	Type	Defects Detected	Corrections Applied
1	City	Yes	object	1,963,605 rows were empty 10,137 rows had the value "Null."	All the missing values and nan were replaced with "Lisboa" because all the data were from the Lisbon metropolitan area
2	Country	No	object		Not considered for analysis
3	Delay	Yes	float64	8,779,033 rows had the value "−1" when the level of traffic was 5	−1 was treated as a missing value and was replaced with a value prediction using the Pearson correlation and linear regression

4	Endnode	Yes	object	3,471,012 rows had the value "Null."	All the null values were replaced with "Unknown" for dashboard display improvement
5	Entity_ts	Yes	float64		
6	Length	Yes	float64		
7	Level	Yes	float64		
8	Position	Yes	object		
9	RoadType	Yes	float64		
10	Street	Yes	object	1,620,642 rows had the value "Null."	All the null values were replaced with "Unknown" for dashboard display improvement
11	Typeentity	No	object		Not considered for analysis
12	Validity_ts	No	float64		Not considered for analysis

When the traffic level was 5, the Delay column had the value "−1", and this value could not be used for analysis. To surpass this problem without impacting the column values, "−1" was treated like missing data and replaced with a value prediction using a Pearson correlation and linear regression [24,25]. Table 8 shows the Pearson correlation between all variables. The variables Length, Level, and Road Type were selected to find the missing values because of the high correlation with the variable Delay.

Table 8. Pearson correlation table.

Characteristics	Delay	Entity_ts	Length	Level	Road Type	Validity_ts
delay	1.00	0.04	0.37	0.40	0.15	0.04
entity_ts	0.04	1.00	−0.18	0.30	0.14	1.00
length	0.37	−0.18	1.00	−0.53	−0.02	−0.18
level	0.40	0.30	−0.53	1.00	0.11	0.30
roadtype	0.15	0.14	−0.02	0.11	1.00	0.14
validity_ts	0.04	1.00	−0.18	0.30	0.14	1.00

When applying the values predicted by the linear regression algorithm to the original data set, we could observe an increase in the mean and median, and a decrease in the standard deviation, which could be explained by the increase in traffic occurrences with a high delay (level 5 traffic level). Table 9 shows the impact of the application of predicted values to the original dataset:

Table 9. Original vs. postprediction delay results.

Data Description	Original Delay	Postprediction Delay
Mean	46	235
Median	−1	267
Standard deviation	107	96
Minimum Value	−1	0
Maximum Value	6133	6133

According to the terms of the proposed challenge by the Lisbon City Hall, which stated that the aim was to "evaluate the traffic, the main entry points, and the main roads within the capital that are freeways", we filtered the main entry points of the city to include only road types 2 and 3 (primary streets and freeways, respectively), reducing the total number of rows to 5,123,746, or 4,982,407 primary streets and 141,339 freeways.

With the reduced dimensionality [32,33], we created new features to improve the display of information:

- entity_Date: Conversion of the entity_ts from UNIX time to standard date format (year-month-day hour:minute:second).
- Traffic Level: Level of the traffic according to the variable Traffic:
  - 1 = Low traffic
  - 2 = Low to medium traffic
  - 3 = Medium traffic
  - 4 = High traffic
  - 5 = Traffic jam
- length_KM: Length of the traffic in kilometers
- delay_M: Delay in minutes
- delay_H: Delay in hours
- Date_key: Date identification in a format yyyymmdd

3.3.2. Road Accident Data Preparation and Fusion

We performed data cleaning and preprocessing in Python, using the Spyder platform [34] and Python libraries, such as Numpy [35], Pandas [36], Matplotlib [37], and Seaborn [38].

From the initial data set comprising 4 Excel sheets, four corresponding new data sets were generated: 'acidentes', referring to accidents; 'veíc-cond', referring to vehicles involved; 'passageiros', corresponding to passengers; and 'peões', meaning pedestrians.

In the next step, data cleaning, we replaced some strings with null values since they did not add value. For example, 'NÃO DEFINIDO' was replaced by 'nan'. Some columns were deleted (acidentes: IdAcidente, Sentidos, Pov Próxima, Km; veíc-cond: IdAcidente, id_veiculo, certificado_adr; passageiros: IdAcidente, id_veiculo, id_passageiro; peões: IdAcidente, id_peao), while others were renamed and then converted for the string type. In addition, in the data cleaning, due to the occurrence of a large number of 'nan' in the categoria_veiculo column, regarding the veíc-cond data frame, 'nan' values were replaced, taking into account the values in the tipo_veiculo column. The data set was divided by type of accident: mislead (despiste), collision (colisão), and runover (atropelamento). The Datahora column was divided in date and hour to show on what day of week and what hour of the day more accidents occurred, or even in what season of the year (spring, summer, autumn, or winter). From the Datahora, we therefore created hora (hour) and mes (month), as well as new columns: nome_estacao (season) and momentoDia (moment of the day), to analyze the moment of the day and the moment of the year when accidents were more likely to happen. In the veíc-cond, passageiros, and peões data sets, age was aggregated to create an average column. A new data set was created with latitude and longitude columns to visualize the Lisbon map's geographic coordinates.

Tables 10–13 show each data set's first row after the transformations mentioned above, namely the new added columns.

Table 10. Accident data schema after data preparation.

Column	Data of 1st Row
Date and time	2019-01-02 15:10:00
Day of the week	Wednesday
Latitude	38.7683669
Longitude	−9.1728989
Traffic side	Right
Location	In the city
Parish	Lumiar

Accident type	Runover
Type	Pedestrian runover
Track_1	Straight
Track_2	Plateau
Track_3	Without road berm
Track_4	In parking area
Conservation status	Good
Technical characteristics	Road without separator
Road lane direction	One direction
Road tracing	nan
Obstacles	Nonexistent
Signals	nan
Light signals	nan
Road lane type	Concrete
Road lane intersection	nan
Weather conditions	nan
Luminosity	Daylight
Adherence conditions	Dry and clean
VM	0
FG	0
FL	1
Type of road	Road
Road	0
Street number	0
Street name	Azinhaga Ulmeiros
Location_2	Without GPS—estimated
Date	2019-01-02 00:00:00
Month	1
Hour	15
Season	1
Season name	Winter
Daytime	12 h–15 h

Table 11. Vehicles involved data schema after data preparation.

Column	Data of 1st Row
Date and time	2019-06-10 11:30:00
Vehicle category	Car
Vehicle type	Passengers
Age	32
Gender	Female
Injuries in 30 days	With no injuries
Driver's accessories	Seat belt
Driver's actions	Sudden braking
Infractions	Exceed speed to existing road characteristics
Driver's license	With driving license/driving license according to vehicle
Driving time	Less than 1 h
Alcohol test	Not submitted
Load capacity	Without load
Regular inspection	Valid

Insurance	With insurance
Age range	30–39
Date	2019-06-10 00:00:00
Month	6
Hour	11
Season	3
season name	Summer
Daytime	9 h–12 h

Table 12. Passengers involved data schema after data preparation.

Column	Data of 1st Row
AccidentID	20201823347
Date and time	2019-12-18 13:00:00
Age	22
Gender	Female
Injuries in 30 days	Bruised
Passenger's accessories	With helmet/seat belt
Age range	18–29
Date	2019-12-18 00:00:00
Month	12
Hour	13
Season	1
Season name	Winter
Daytime	12 h–15 h

Table 13. Pedestrians involved data schema after data preparation.

Column	Data of 1st Row
Date and time	2019-01-17 16:12:00
Age	59
Gender	Male
Injuries in 30 days	Bruised
Pedestrian accessories	Construction works running
Age range	50–59
Date	2019-01-17 00:00:00
Month	1
Hour	16
Season	1
Season name	Winter
Daytime	15 h–18 h

3.4. Data Visualization

3.4.1. Traffic Model

Our modeling results showed that in 2019, 96.90% of the traffic congestion occurred on primary streets, and 3.10% of the traffic congestion occurred on freeways, as shown in Figure 2.

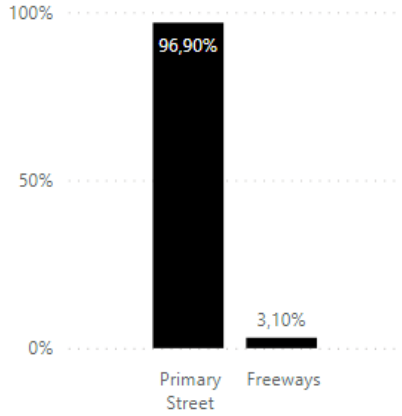

Figure 2. Traffic distribution by road type in Lisbon in 2019.

Figure 3 displays the distribution of traffic levels presented for both freeways and primary streets: 64.76% of the traffic corresponded to a level 5 occurrence (traffic jam), 17.78% to level 3 (medium traffic), 12.78% to level 4 (high traffic), 4.48% to level 2 (low to medium traffic), and only 0.20% to level 1 (low traffic).

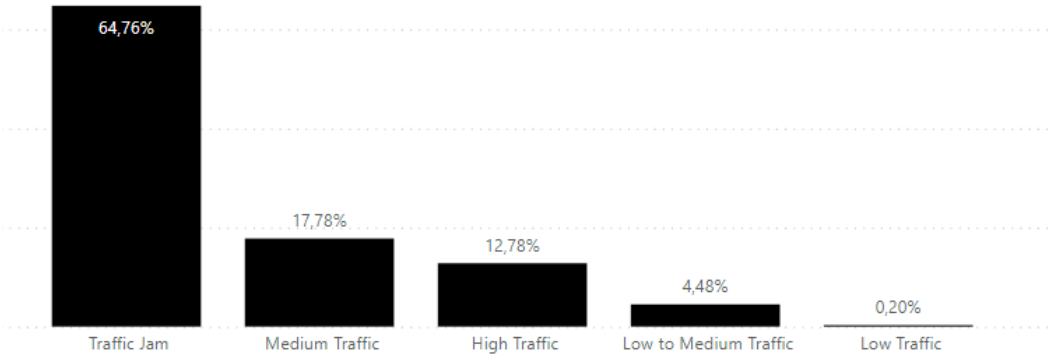

Figure 3. Traffic distribution by severity level in Lisbon in 2019.

Our data analysis showed that 5.39% of level 5 traffic occurrences happened on freeways versus 66.66% on primary streets, and 34.30% of level 4 traffic occurred on freeways versus 12.09% on primary streets. The data also showed that traffic level 3 traffic was more common on freeways, with 44.28% vs. 16.93% on primary streets, and level 2 traffic was more common on freeways, with 14.62% compared to primary streets with 4.16%. The model also showed that level 1 traffic was the least-common traffic on both freeways (1.41%) and primary streets (0.16%).

The average traffic length on freeways was 1.71 km, and the average delay was 5.27 min. In comparison, primary streets had an average traffic length of 0.19 km and an average delay of 3.86 min. Table 14 displays the average traffic delay and length distributed by road type and traffic level.

Table 14. Average delay and traffic length by road type.

Road Type	Average Delay (min)	Average Traffic Length (km)
Freeway	5.27	1.71
Low traffic	1.26	2.35
Low to medium traffic	2.19	2.25
Medium traffic	3.67	1.89
High traffic	8.87	1.46

Traffic jam	4.90	0.25
Primary Street	3.86	0.19
Low traffic	1.10	0.61
Low to medium traffic	1.29	0.54
Medium traffic	1.87	0.41
High traffic	3.65	0.34
Traffic jam	4.57	0.09
Total	3.90	0.24

According to our analysis, rush hours in Lisbon occur between 8 a.m. and 9 a.m. in the morning and between 5 p.m. and 6 p.m. in the evening, and make up 24.77% of daily traffic occurrences (11.34% in the morning and 13.43% in the afternoon).

Lisbon boroughs Paço do Lumiar, Sacavém, Lumiar, Ajuda, and Belém together displayed 80.93% of the Lisbon metropolitan area's traffic occurrences. In 2019, the months with higher traffic were October (16.61%), November (16.91%), and December (15.95%), and the months with lower traffic were April (1.83%), February (2.76%), and January (2.88%).

Our analysis also showed that the freeway IC17 and CRIL had a higher percentage of traffic occurrences (26.61%) and an average delay of 4.84 min, followed by A5 with 22.16% and an average delay of 4.53 min. The freeway with the highest delay was IP7/Eixo N-S, with an average delay of 8.07 min, followed by IC-19, with an average delay of 5 min. Figure 4 shows the distribution of traffic on Lisbon freeways.

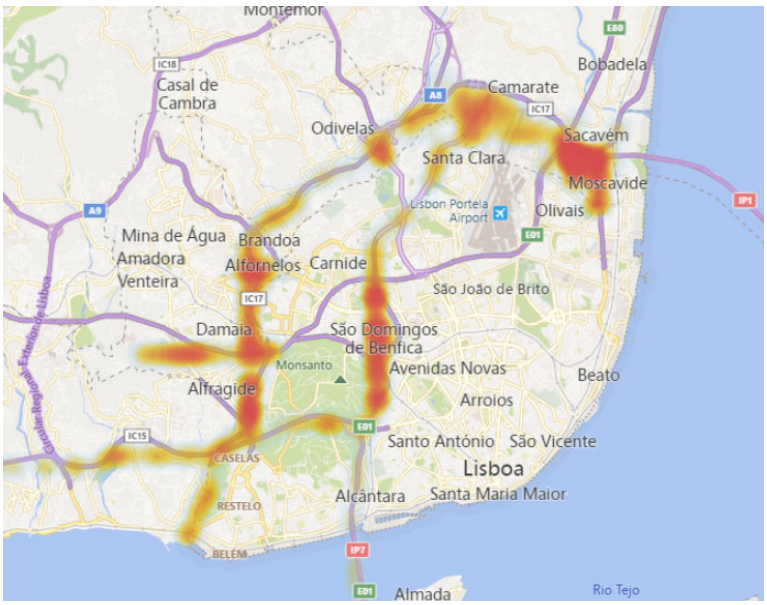

Figure 4. Distribution of traffic on freeways in Lisbon in 2019. Red corresponds to locations where more traffic congestion occurred on freeways; orange and yellow are locations where traffic congestion was less common on freeways.

Analysis of primary roads showed that Rua Direita in Lumiar, Rua Auta da Palma Carlos in Sacavém, and Calçada da Ajuda in Ajuda combined to make up 54.75% of the traffic occurrences on Lisbon's primary streets. According to the data, these streets were also the ones with the highest delays, where drivers experienced an average delay of 5 min. Figure 5 shows the distribution of traffic on primary streets.

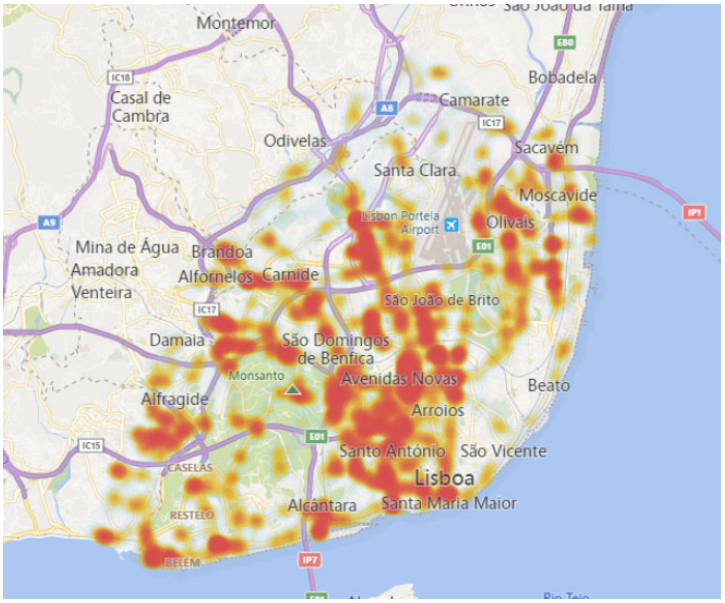

Figure 5. Distribution of traffic on primary streets in Lisbon in 2019. Red corresponds to locations where more traffic congestion occurred on primary streets; orange and yellow are locations where traffic congestion was less common on primary streets.

3.4.2. Road Accident Model

Modeling, which involved analysis and visualization, was performed in Python with the Spyder platform. Python libraries such as Numpy [35], Pandas [36], Matplotlib [37], and Seaborn [38] were used for statistical analysis, and Folium [39] and Geopandas [40] were used for spatial analysis visualization.

In the entire universe of accidents, we could gauge that 56.8% of the accidents were collisions, 24.3% were runovers, and 18.8% were misleads. Accidents that occurred during the day represented 66%, mainly between 3 p.m. and 6 p.m. It was also identified that 52% of accidents occurred in autumn and spring.

Figure 6a–c shows the road accident distribution by type, hour of the day, and season of the year.

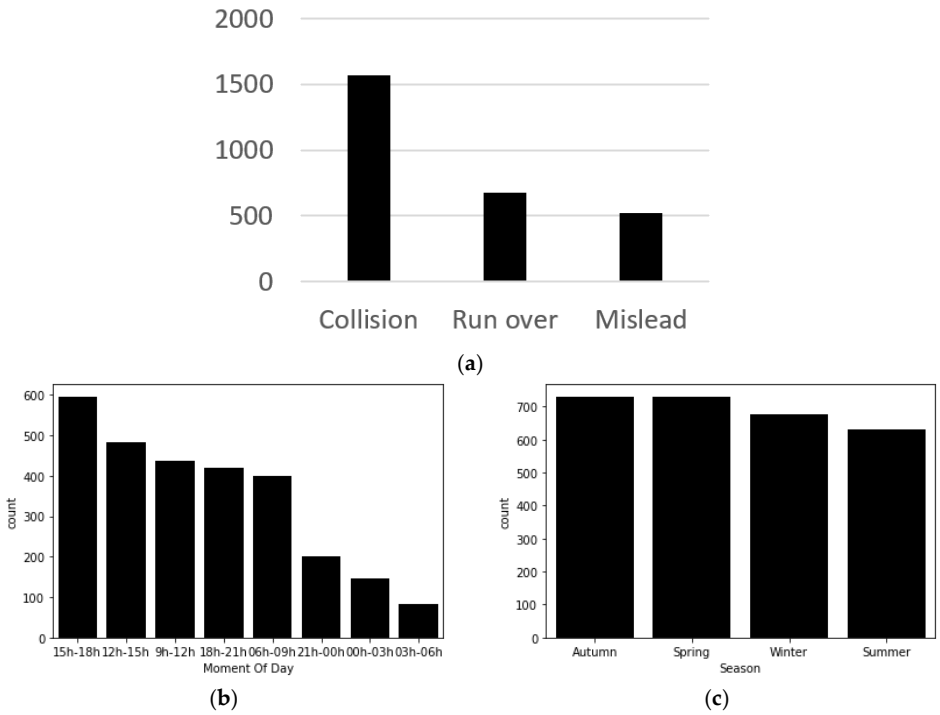

Figure 6. (**a**) Distribution of road accidents by type—collision, runover, and misleading; (**b**) road accident count by time of the day; (**c**) road accident count by season.

October and November were the months where 30% of misleads happened; of these, 54% were in broad daylight, and 39% were at night but with illumination. On Thursday and Saturday, there was a higher prevalence of misleads; this was lesser on Sunday.

It is clear that runovers and collisions (Figures 7 and 8) occurred more at Avenidas Novas, and misleads (Figure 9) at São Domingos de Benfica. On the other hand, runovers were more likely to occur in Lisbon's more touristic areas, and collisions mostly in the Lisbon city center.

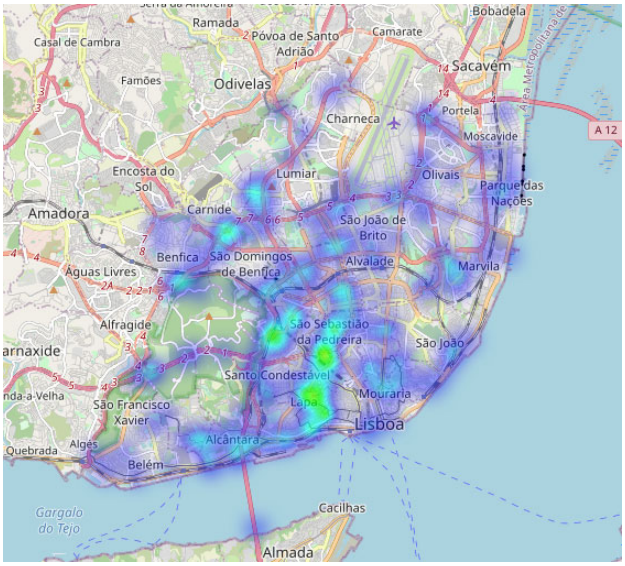

Figure 7. Mislead distribution in Lisbon in 2019. Orange and yellow represent streets where more misleads occurred; green and blue represent less common places for mislead occurrences.

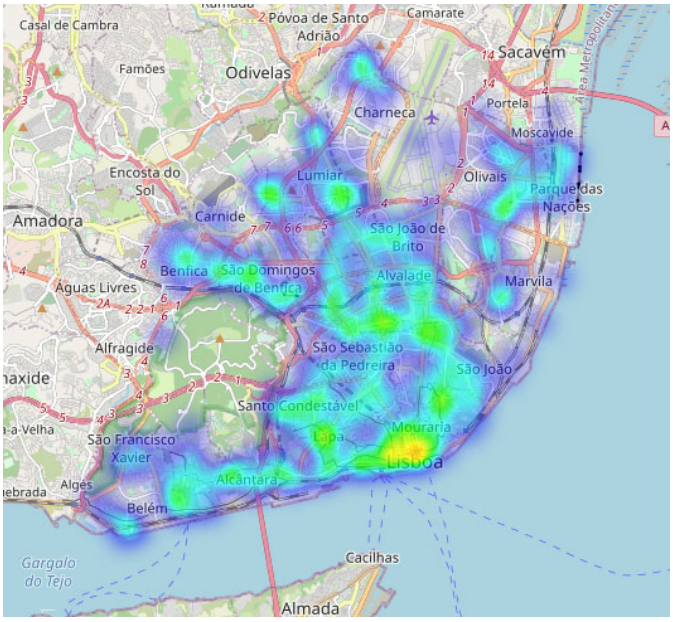

Figure 8. Runover distribution in Lisbon in 2019. Orange and yellow represent streets where more runovers occurred; green and blue represent less common places for runover occurrences.

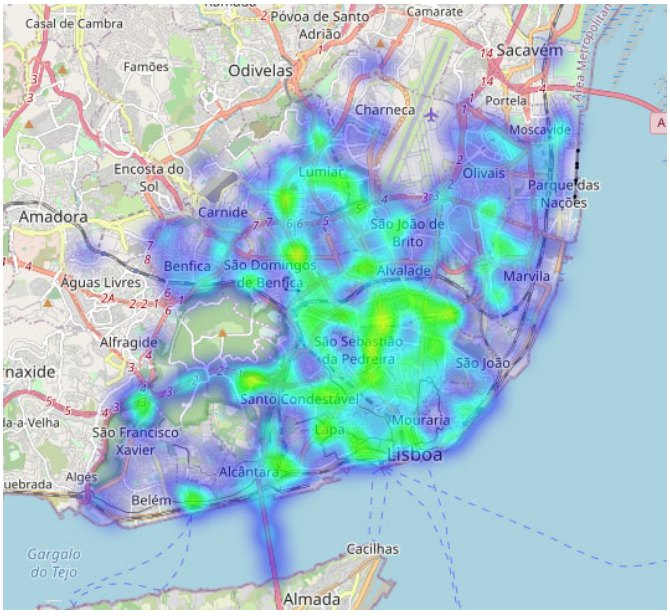

Figure 9. Collision distribution in Lisbon in 2019. Orange and yellow represent streets where more collisions occurred; green and blue represent less common places for collision occurrences.

Data visualizations of the incidence of accidents in Lisbon are shown as heatmaps in Figures 7–9. Areas represented in orange and yellow were the most active, representing streets where more accidents occurred; green and blue represent areas with lower occurrences of accidents.

Misleads occurred more in São Domingos de Benfica, Benfica, Carnide, and Lumiar (Figure 7), corresponding to entrances and exits to Lisbon's outskirts. The central city axis from Campo Grande, Avenida da República, passing by Saldanha, Marquês do Pombal, to Avenida Infante Santo had a strong incidence of mislead occurrences.

Approximately 95% of runovers involved pedestrians in broad daylight, and most happened between 3 p.m. and 6 p.m. on Tuesday and Friday. Runover incidents were

scattered throughout the city (Figure 8). There was a strong incidence in the downtown area, Terreiro do Paço, that could be associated with distracted tourists strolling. Other areas include Parque das Nações, São Domingos de Benfica, and Alcântara, corresponding to entrances to and exits from Lisbon. From Campo Grande to Avenida da República, passing by Saldanha, Marquês do Pombal, Avenida Infante Santo to Avenida 24 de Julho are the streets where most runovers took place. Moreover, the same pattern was also observed in the analysis and visualization of misleads (Figure 7).

Finally, regarding collisions, 65% of them occurred during daylight, and 21% between 3 p.m. and 6 p.m. on Thursdays and Fridays. Side collisions with another moving vehicle represented 35% of the total, while the more expressed others were rear collision with another moving vehicle, and collision with other situations involved.

Collision occurrence was scattered throughout the city, with focuses in Marvila, Alcântara (entrance and exit to Ponte 25 de Abril) and downtown, Terreiro do Paço, and stronger occurrences in Alvalade, Areeiro, Avenidas Novas, Campo de Ourique, and Estrela boroughs.

Generally speaking, the vehicle analysis showed that 89% of accidents corresponded to a passenger vehicle with a driver 18–29 years of age (28%). In addition, 84% of accidents occurred on a dry and clean road, and 15% on a wet road. It was possible to see that of all the passengers, 64.8% were females, and 33.8% were between 18–29 years of age. Of these, 99% suffered minor injuries and were wearing a seat belt or a helmet.

Pedestrians involved in accidents were mainly females (55.9%), and 94% suffered minor injuries, of which 41% were crossing roads on a signalized zebra crossing. Pedestrians aged 18 to 29 years and 70+ were the main citizens involved.

Moreover, misleads and runovers had a high incidence at the entrances and exits of Lisbon, and along the central axis of the city from Campo Grande to Avenida Infante Santo. However, these were phenomena that occurred all over the city.

4. Discussion

The major finding, addressing our first study on Lisbon traffic patterns in 2019, was that the predominance of traffic occurrences on primary streets represented most of the traffic events in the Lisbon metropolitan area.

Other findings showed that the average driver was likely to spend between 3 to 5 min more on each road—primary street or freeway—when a traffic event occurred. These traffic occurrences tended to peak between 8 a.m. and 9 a.m. and between 5 p.m. and 6 p.m.

When observing the data patterns, it was possible to assess that Lisbon boroughs, at the city limits, had the highest concentration of traffic jams. Most of these boroughs are neighboring municipalities, part of Lisbon's network of entrances and exits.

According to the data analysis, freeway traffic tended to have a higher length and delay than primary street traffic. However, traffic occurrences on freeways tended to be less severe than on primary streets. However, some freeways had slower and lengthier traffic jams than others, such as IC 17, A5, and Eixo Norte-Sul. The primary streets that stood out with more traffic congestion were R. Direita, R. Alta da Palma Carlos, and Calçada da Ajuda.

Traffic congestion was susceptible to pendular variations in its intensity, with critical peaks at the city's main entrances, especially during the morning and evening peak hours. However, traffic congestion tended to normalize throughout the evening. Our findings showed that roads were especially prone to long traffic jams when public events occurred. When a traffic event occurred, it was very likely to be of the heavy type.

Road accident patterns in Lisbon strongly correlated with traffic congestion and external factors, but varied according to the type of occurrence. Our findings showed that most accidents occurred in bustling traffic areas, during the daytime (between 3 p.m. and 6 p.m.), corresponding to the afternoon traffic peak when traffic congestion peaked. Friday was the weekday with a higher prevalence of road accidents, mostly when

commuters were traveling back home and rushing for weekend leave. Seasonality also influenced road accidents, as most accidents occurred in autumn and spring.

Hence, external factors such as weather, location, luminosity, and time of the day were crucial to understanding Lisbon's road accident phenomena. Age also influenced accidents. The population between 18 and 29 years old were the most involved in road accidents. Similarities with the literature [7,8,11] were observed, especially regarding our study's external factors.

Road accidents in Lisbon were scattered in the city by type of occurrence. Misleads and runovers had a higher incidence at Lisbon's entrances and exits, and along the central axis of the city from Campo Grande to Avenida Infante Santo. These occurrences corresponded to traffic-congestion patterns observed on primary streets and freeways. São Domingos de Benfica, Carnide, Lumiar, and Alcântara were the zones where most misleads and runovers occurred, corresponding to the entrances and exits of freeways—IC 17, A5, and Eixo Norte-Sul—with high congestion levels. Collisions, on the other hand, were a phenomenon that spread all over the city.

Our literature survey [7,8,11] found few articles about data analysis and visualization of traffic congestion and road accidents. We addressed this by developing data analytics and visualization, providing effective communication of research insights and visualization tools to policymakers.

Following insights from previous studies [14,15], we developed data analyses and visualizations based on different and combined factors, such as temporal, spatial, spatio-temporal, and multivariable.

Multivariable analysis has shown effectiveness in communicating data insights [16], providing visualizations of data regarding traffic congestion and road accidents within a time range and city areas, type of vehicle, lighting conditions, weather conditions, month of the year, and day of the week.

The results of our traffic-congestion and road-accident analyses and visualizations showed similar findings to insights found in the literature review, as shown in Table 15.

Table 15. Literature review and results synthesis.

Reference	Literature Review	Results
[9]	Multivariable analysis of city names, type of accident, condition of light, severity, speed zone, consumption of alcohol, time and day of the accident, etc.	Multivariable analysis applied to Lisbon road accidents identified patterns by type of accident, lighting condition, vehicle speed, and day and time of the accident
[14]	Multivariable data-processing techniques depicting the temporal, spatial, numerical, and categorical properties of traffic data through visualization	Multivariable data analysis and visualizations with variables associated with Lisbon traffic congestion (day and time) and road accidents (weekend, time, number of vehicles, road, brightness, and weather); vehicle (type and age of vehicle, and other types of vehicles in the accident); and individuals (gender, seat belt, and position in the vehicle)
[15]	Multivariable analysis: six vehicle categories using various stability criteria, evaluated for mixed traffic conditions	Lisbon traffic congestion patterns according to street type, and road accident patterns according to vehicle type and street type
[16]	Multivariable analysis and visualization of geospatial data (Tableau and GeoCharts)	Lisbon geomapping heatmaps as a visualization method in traffic congestion and road accidents

According to our literature review (Table 1), further work is needed to better understand the correlation between traffic congestion and road accidents. Poisson-based nonspatial (such as Poisson–lognormal and Poisson–Gamma) and spatial (Poisson–lognormal with conditional autoregressive priors) models [18,19] should be developed further regarding spatial correlation in traffic congestion and road accidents, and should provide a broader and deeper understanding of this phenomenon.

Moreover, in order to provide a data-driven analytics and visualization tool that helps decision-makers (public authorities and stakeholders), which is one of our research goals, we need to develop prediction models. Although we did not address predictive models in our study, we discussed them in the literature review in Section 2. Andrienko [22] used a visualization technique developed iteratively with data analysis, model development, and analysis of predictions. Moreover, long short-term memory (LSTM) and recurrent neural networks (RNNs) are the most used traffic prediction models [23].

Our study's limitations were related to incomplete data features and a lack of information regarding external factors data—weather, air quality, events, sports, music, and COVID-19. Waze traffic data needs to be understood in the context of Lisbon traffic; in particular, the percentage of the overall Lisbon traffic it represents in reality. Additionally, more data are required; namely, the numbers of cars in the city and cars that commute to the city, car speed, accident occurrences, external factors such as public events—soccer games, music festivals—as well as COVID-19 pandemic data. These additional data sources are needed to provide broader analysis and visualization of traffic and road accident patterns in the city.

5. Conclusions

Lisbon traffic congestion and road accidents are ongoing issues in Lisbon's urban mobility, carbon-emission reduction, and road safety. In a city where 370,000 vehicles enter every day, adding to the 200,000 vehicles that already circulate in the city, traffic congestion and road accidents are key challenges that need to be tackled to improve citizens' quality of life.

This study accomplished our research aim and goals in characterizing traffic congestion and road accidents in Lisbon (RQ1), as well as characterizing road accident patterns in Lisbon, and the external factors that contribute the most to this phenomenon (RQ2).

The developed multivariable analyses and visualizations showed congestion and road accident patterns related to their features and external factors. Traffic congestion and road accidents correlated with highway street type, on Friday and during afternoon peak hours. Road accident characterization also featured number of vehicles involved, brightness, and weather; and variables associated with individuals, such as gender, seat belt, and position in the vehicle. This enriched the characterization of Lisbon road accidents and identified the external factors that contributed the most to this phenomenon.

Poisson nonspatial and spatial models [18,19] need to be constructed to better understand the correlation between traffic congestion and road accidents; this will be addressed in future work.

Although in our study there was a strong correlation between traffic congestion and road accidents, this relation was not necessarily linear, meaning one can happen without the other occurring, and this needs more to be studied in more detail.

The developed data analytics and visualization tool for traffic congestion and road accidents provides a traffic visualization pipeline to assess traffic data properties based on a multivariable analysis that finds urban mobility patterns.

Traffic congestion and road accident visualization provided knowledge and insights on how citizens move in Lisbon (RQ3), enabling a better understanding of road accident scenarios and related events. It also provided data-driven guidelines and knowledge about traffic congestion and road accidents in Lisbon to the city authorities and

policymakers in the framework of a traffic management and visualization tool to help them mitigate such phenomena.

Future work aims to better understand these results' implications, especially regarding external factors such as weather, air quality, events, crowd flow, and bike data. Studies [41,42] of the Lisbon bike-sharing system have shown the influence of external factors, such as weather and bike type, in bike use and ride frequency. Moreover, pedestrian walkability, cycleways, bike stations, and bike accident data are of interest to correlate with this study's findings and understand the overall Lisbon urban mobility scenario.

To this aim, an integrated urban mobility dashboard with data analysis and visualization can contribute to providing city management authorities and policymakers with an overall picture of the city's urban mobility, enabling implementation of smart solutions toward a more resilient city.

This urban mobility tool would allow city management authorities and decision-makers to explore and better understand the profile of Lisbon metropolitan area commuters by means of an interactive dashboard depicting georeferenced data with different borough, geographical, demographic, economic, social, planning and environmental information, generating combined visualizations.

Moreover, this analytical and visualization tool would provide a complete monitoring and management resource of the entire urban ecosystem that could be replicated in other cities. It could also be integrated into the Lisbon Intelligent Management Platform—Plataforma de Gestão Inteligente de Lisboa (PGIL) [43], an existing data platform used by the City Hall, further developing PGIL capacity to process and provide useful information on the operational and strategic management of the city to the various stakeholders.

Author Contributions: Methodology, A.O., J.L.B., R.S.R., F.A. and V.A.; data-mining processes for accidents, A.O., and for traffic, J.L.B.; state-of-the-art revision, V.A., R.S.R. and A.O.; visualization, A.O. and J.L.B.; system development, V.A.; writing—original draft preparation, A.O., J.L.B., V.A., and J.C.F.; supervision and contribution to the article's writing, M.S.D.; research coordination and contribution to the article's writing, J.C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation for Science and Technology (FCT) through ISTAR-IUL's projects UIDB/04466/2020 and UIDP/04466/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study used third-party data. Restrictions apply to the availability of these data. Data were obtained from Câmara Municipal de Lisboa (CML) and Lisboa Inteligente, and are available at https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/criacao-indicador-de-trafego-geral-e-indicadores-para-cada-uma-das-principais-vias-de-entrada-na-cidade (accessed on 29 March 2021) and https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/identificacao-de-pontos-de-incidencia-dos-acidentes-rodoviarios-e-da-sua-correlacao-com-outros-fatores (accessed on 29 March 2021) with the permission of Câmara Municipal de Lisboa (CML) and Lisboa Inteligente.

Acknowledgments. J.C.F. received support from the Portuguese National Funds through FITEC—Programa Interface, with reference CIT INOV—INESC INOVAÇÃO—Financiamento Base.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Allam, Z.; Newman, P. Redefining the Smart City: Culture, Metabolism and Governance. *Smart Cities* **2018**, *1*, 4–25, doi:10.3390/smartcities1010002.
2. *World Urbanization Prospects*; The 2007 Revision Report; UN: New York, NY, USA, 2008.
3. Albino, V.; Berardi, U.; Dangelico, R.M. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *J. Urban Technol.* **2015**, *22*, 3–21, doi:10.1080/10630732.2014.942092.

4. Lana, I.; Del Ser, J.; Velez, M.; Vlahogianni, E.I. Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 93–109, doi:10.1109/mits.2018.2806634.
5. Nagy, A.M.; Simon, V. Survey on traffic prediction in smart cities. *Pervasive Mob. Comput.* **2018**, *50*, 148–163, doi:10.1016/j.pmcj.2018.07.004.
6. Petrovska, N.; Stevanović, A. Traffic Congestion Analysis Visualisation Tool. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems (IEEE), Las Palmas de Gran Canaria, Spain, 15–18 September 2015; pp. 1489–1494.
7. Dimitrakopoulos, G.; Demestichas, P. Intelligent Transportation Systems. *IEEE Veh. Technol. Mag.* **2010**, *5*, 77–84, doi:10.1109/mvt.2009.935537.
8. World Health Organization. *Save Lives: A Road Safety Technical Package*; WHO: New Delhi, India, 2017.
9. Krishnan, P.V.; Sheel, V.C.; Viswanadh, M.V.S.; Shetty, C.; Seema, S. Data Analysis of Road Traffic Accidents to Minimize the rate of Accidents. In Proceedings of the 3rd International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 20–22 December 2018; pp. 247–253.
10. Palazón-Bru, A.; Prieto-Castelló, M.J.; La Rosa, D.M.F.-d.; Macanás-Martínez, A.; Mares-García, E.; Carbonell-Torregrosa, M.d.l.Á.; Gil-Guillén, V.F.; Cardona-Llorens, A.; Marhuenda-Amorós, D. Development, and Internal, and External Validation of a Scoring System to Predict 30-Day Mortality after Having a Traffic Accident Traveling by Private Car or Van: An Analysis of 164,790 Subjects and 79,664 Accidents. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9518, doi:10.3390/ijerph17249518.
11. Nunes, D. 42 Minutos de Fila por Dia: Lisboa é a Cidade Ibérica com Mais Trânsito, Diário de Notícias, June 2019. Available online: https://www.dn.pt/dinheiro/42-minutos-de-fila-por-dia-lisboa-e-a-cidade-iberica-com-mais-transito-10976480.html (accessed on 29 March 2021).
12. Relatório Anual de Segurança Rodoviária. Autoridade Nacional de Segurança Rodoviária, Portugal, 2019. Available online: http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2019/Relat%C3%B3rio%20Anual%20Sinistralidade%20Rodovi%C3%A1ria%202019.pdf (accessed on 29 March 2021).
13. Tang, J.; Zheng, L.; Han, C.; Yin, W.; Zhang, Y.; Zou, Y.; Huang, H. Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review. *Anal. Methods Accid. Res.* **2020**, *27*, 100123, doi:10.1016/j.amar.2020.100123.
14. Chen, W.; Guo, F.; Wang, F.-Y. A Survey of Traffic Data Visualization. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2970–2984, doi:10.1109/tits.2015.2436897.
15. Surya, H.R.; Raju, N.; Arkatkar, S.S. Stability Analysis of Mixed Traffic Flow using Car-Following Models on Trajectory Data. In Proceedings of the 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengularu, India, 5–9 January 2021; pp. 656–661.
16. Müller, L.; Moser, C.; Paris, G.; Freitas, L.; Oliveira, M.; Signoretti, W.; Manssour, I.H.; Silveira, M.S. Hit by the Data: A visual data analysis regarding the effects of traffic public policies. *arXiv* **2021**, arXiv:2102.07621.
17. Ochoa-Ruiz, G.; Angulo-Murillo, A.A.; Ochoa-Zezzatti, A.; Aguilar-Lobo, L.M.; Vega-Fernández, J.A.; Natraj, S. An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment. *Appl. Sci.* **2020**, *10*, 3974, doi:10.3390/app10113974.
18. Lord, D.; Miranda-Moreno, L.F. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. *Saf. Sci.* **2008**, *46*, 751–770, doi:10.1016/j.ssci.2007.03.005.
19. Wang, C.; Quddus, M.A.; Ison, S.G. Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. *Accid. Anal. Prev.* **2009**, *41*, 798–808, doi:10.1016/j.aap.2009.04.002.
20. Aguero-Valverde, J.; Jovanis, P.P. Analysis of Road Crash Frequency with Spatial Models. *Transp. Res. Rec.* **2008**, *2061*, 55–63, doi:10.3141/2061-07.
21. Zhang, J.; Wang, F.-Y.; Wang, K.; Lin, W.-H.; Xu, X.; Chen, C. Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639, doi:10.1109/tits.2011.2158001.
22. Andrienko, N.; Andrienko, G.; Rinzivillo, S. Experiences from Supporting Predictive Analytics of Vehicle Traffic. Available online: http://predictive-workshop.github.io/papers/vpa2014_3.pdf (accessed on 29 March 2021).
23. Xiangxue, W.; Lunhui, X.; Kaixun, C. Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN. *Arab. J. Sci. Eng.* **2019**, *44*, 3043–3060, doi:10.1007/s13369-018-3390-0.
24. Liu, B.; Tang, X.; Cheng, J.; Shi, P. Traffic flow combination forecasting method based on improved LSTM and ARIMA. *Int. J. Embed. Syst.* **2020**, *12*, 22–30, doi:10.1504/ijes.2020.105287.
25. Zheng, J.; Huang, M. Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning. *IEEE Access* **2020**, *8*, 82562–82570, doi:10.1109/access.2020.2990738.
26. Norros, I.; Kuusela, P.; Innamaa, S.; Pilli-Sihvola, E.; Rajamäki, R. The Palm distribution of traffic conditions and its application to accident risk assessment. *Anal. Methods Accid. Res.* **2016**, *12*, 48–65, doi:10.1016/j.amar.2016.10.002.
27. Wirth, R.; Hipp, J. CRISP-DM: Towards a Standard Process Model for Data Mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 11–13 April 2000; pp. 29–39.
28. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. *CRISP-DM 1.0: Step-by-Step Data Mining Guide*; SPSS Inc.: Chicago, IL, USA, 2000.

29. LxDataLab—Laboratório de Dados Urbanos de Lisboa. Available online: https://lisboainteligente.cm-lisboa.pt/lxdatalab/ (accessed on 29 March 2021).
30. C.M. Lisbon, Criação de Indicador de Trafego Geral e Indicadores para Cada Uma Das Principais Vias de Entrada na Cidade. Available online: https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/criacao-indicador-de-trafego-geral-e-indicadores-para-cada-uma-das-principais-vias-de-entrada-na-cidade/ (accessed on 29 March 2021).
31. Identificação de Pontos de Incidencia dos Acidentes Rodoviarios e da sua Correlação com Outros Fatores. Available online: https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/identificacao-de-pontos-de-incidencia-dos-acidentes-rodoviarios-e-da-sua-correlacao-com-outros-fatores/ (accessed on 29 March 2021).
32. Gorard, S. Handling missing data in numeric analyses. *Int. J. Soc. Res. Methodol.* **2020**, *23*, 651–660, doi:10.1080/13645579.2020.1729974.
33. Kumar, S. Predict Missing Values in the Dataset. 2020. Available online: https://towardsdatascience.com/predict-missing-values-in-the-dataset-897912a54b7b (accessed on 29 March 2021).
34. Spyder IDE. Available online: https://www.spyder-ide.org/ (accessed on 29 March 2021).
35. Nbsp; NumPy. Available online: https://numpy.org/ (accessed on 29 March 2021).
36. Nbsp; Pandas. Available online: https://pandas.pydata.org/ (accessed on 29 March 2021).
37. MatPlotLib. Available online: https://matplotlib.org/ (accessed on29 March 2021).
38. Seaborn. Available online: https://seaborn.pydata.org/ (accessed on 29 March 2021).
39. Folium. Available online: https://pypi.org/project/folium/ (accessed on 29 March 2021).
40. Geopandas. Available online: https://geopandas.org/ (accessed on 29 March 2021).
41. Albuquerque, V.; Andrade, F.; Ferreira, J.C.; Dias, M.S. Understanding Spatiotemporal Station and Trip Activity Patterns in the Lisbon Bike-Sharing System. In *Intelligent Transport Systems, From Research and Development to the Market Uptake*; Martins, A.L., Ferreira, J.C., Kocian, A., Costa, V., Eds.; Springer: Cham, Switzerland, 2021; pp. 16–34.
42. Albuquerque, V.; Andrade, F.; Ferreira, J.; Dias, M.; Bacao, F. Bike-sharing mobility patterns: A data-driven analysis for the city of Lisbon. *EAI Endorsed Trans. Smart Cities* **2021**, 169580, doi:10.4108/eai.4-5-2021.169580.
43. Plataforma de Gestão Inteligente de Lisboa. Available online: https://lisboainteligente.cm-lisboa.pt/lxi-iniciativas/plataforma-de-gestao-inteligente-de-lisboa/ (accessed on 6 April 2021).