INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Patterns of Mobility in a Smart City.

Jorge Lourenço Valente Pinto Barbosa

Master in, Integrated Decision Support Systems

Supervisor:
PhD, João Carlos Amaro Ferreira, Auxiliar Professor with Habilitation,
ISCTE-IUL

Co-Supervisor
PhD, José Miguel de Oliveira Monteiro Sales Dias, Associated Professor with Habilitation,
ISCTE-IUL

October, 2021

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Patterns of Mobility in a Smart City.

Jorge Lourenço Valente Pinto Barbosa

Master in, Integrated Decision Support Systems

Supervisor:
PhD, João Carlos Amaro Ferreira, Auxiliar Professor with Habilitation,
ISCTE-IUL

Co-Supervisor
PhD, José Miguel de Oliveira Monteiro Sales Dias, Associated Professor with Habilitation,
ISCTE-IUL

October, 2021

Acknowledgments

Firstly, I would like to thank my supervisors, Professors João Carlos Amaro Ferreira and José Miguel de Oliveira Monteiro Sales Dias, for all the meetings, guidance, advice, and tips during the preparation and elaboration of this thesis. Furthermore, I want to thank all my colleagues in my study group Diogo Ribeiro, Pedro Mata, and Rui Rodrigues, whom I spent so many hours learning "how to fish" and working in these past two years. Most importantly, I want to thank my family, who supported me and cheered me in all my academic life, especially my mother, my father, and my brother. I also want to thank my girlfriend for all the patience she had while I was developing this thesis.

Finally, I want to thank all my friends that, in one way or another, encouraged, helped me, and did not let me stray too far from my goal. Also, I want to thank and apologize to my dogs for all the walks we did not do in the past two years.

Resumo

Os dados de transporte, no âmbito das cidades inteligentes, estão cada vez mais disponíveis. Estes dados permitem a construção de soluções inteligentes com impacto significativo na vida dos residentes e nos mecanismos das autoridades de gestão da cidade, os chamados Sistemas de Transporte Inteligentes. A nossa investigação incidiu sobre os dados de mobilidade urbana da cidade de Lisboa, disponibilizados pelo município. O principal objetivo da pesquisa foi abordar os problemas de mobilidade, interdependência e soluções de efeitos em cascata para a cidade de Lisboa. Para alcançar este objetivo foi desenvolvida uma metodologia baseada nos dados históricos do transito no centro urbano da cidade e principais acessos, com uma forte componente de visualização. Foi também aplicado um método baseado em series temporais para fazer a previsão das ocorrências de transito na cidade de Lisboa. Foi aplicada uma abordagem CRISP-DM, integrando diferentes fontes de dados, utilizando Python.

Esta tese tem como objetivo identificar padrões de mobilidade urbana com análise e visualização de dados, de forma a auxiliar as autoridades municipais no processo de tomada de decisão, nomeadamente estar mais preparada, adaptada e responsiva.

Palavras-Chave: Transportes; Previsão de tráfego; Padrões de tráfego; Data-driven; Visualização de dados; Cidades Inteligentes

Abstract

Transportation data in smart cities is becoming increasingly available. This data allows building meaningful, intelligent solutions for city residents and city management authorities, the so-called Intelligent Transportation Systems. Our research focused on Lisbon mobility data, provided by Lisbon municipality. The main research objective was to address mobility problems, interdependence, and cascading effects solutions for the city of Lisbon. We developed a data-driven approach based on historical data with a strong focus on visualization methods and dashboard creation. Also, we applied a method based on time series to do prediction based on the traffic congestion data provided. A CRISP-DM approach was applied, integrating different data sources, using Python. Hence, understand traffic patterns, and help the city authorities in the decision-making process, namely more preparedness, adaptability, responsiveness to events.

Keywords: Transportation; Traffic Forecast; Traffic Patterns; Data-driven; Data Visualization; Smart Cities

Index

Tables Index

Figures Index

List of Abbreviations

ANSR - Autoridade Nacional de Segurança Rodoviária

ARIMA - Auto-Regressive Integrated Moving Average

BPNN - Back-Propagation Neural Network

CANN - Cascaded artificial neural network

CFF - Curve Fit Forecast

CML - Lisbon City Hall

CNN - Convolutional Neural Networks

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

DNN - Deep Neural Network

DNN-BTF - Deep Neural Network Based Traffic Flow

ERNN - Elman Recurrent Neural Network

ESF - Exponential Smoothing Forecast

FBF - Forest-based Forecast

GC-LSTM - Graph Convolutional Long Short-Term Memory Neural Network

GC-LSTM - Traffic Graph Convolutional Long Short-Term Memory Neural Network

GLCM - Grey level of co-occurrence matrix

GMDH - Group method of data handling

GPS - Global Positioning System

HMM - Hidden Markov Model

IoT – Internet of Things

ITS – Intelligent Transportation Systems

k-NN - k-Nearest Neighbor

LR - Linear Regression

LSTM - Long Short-Term Memory

PCNN - Pulse Coupled Neural Network

PGIL - Lisbon Intelligent Management Platform - Plataforma de Gestão Inteligente de Lisboa

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PVD - Probe Vehicle Data

RNN - Recurrent Neural Networks

SC – Smart Cities

SLR – Systematic Literature Review

SMO - Sequential Minimal Optimisation

SSGRU - Selected Stacked Gated Recurrent Units model

TI – Transportation Infrastructures

VSR - Volume-Speed Rivers

Chapter 1 – Introduction

1.1. Topic Context

Transportation Infrastructure (TI) has a pivotal role in ensuring citizens' livability, safety, security, and health quality in urban areas. Critical Infrastructures have become more competent in the way they operate, function, and interact with citizens, and customers, as well as between each other, leading to Smart City (SC). The SC is defined as a complex network of technologically advanced Critical Infrastructures connected in a digital environment, characterized by a massive and increasing presence of the Internet of Things (IoT) underlying technologies [1]. Smarter TIs mean regular operation and use, making it more adaptive, intelligent, and connected. Nevertheless, in case of excessive optimization, it can make TIs more vulnerable, 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 SC itself is not sufficient to protect citizens' life and has to be developed in a 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, growing exponentially. By 2050 [2], it is expected that 70 % of the world population will live in metropolitan regions. Due to the rapid population growth, cities face new challenges [2], such as waste, pollution, traffic congestion, and increasing road accidents.

Traffic management is a topic the research community has been tackling for more than 40 years [3]. 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 [4].

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 (ITS). The European Union describes ITS as a vital factor to tackle the member states growing emissions and congestion problems, to support jobs and growth in transportation in metropolitan areas. Moreover, ITS development can create new services and improve daily commutes and operations of specific and combined modes of transportation. This led to the development and deployment of cooperative-ITS between the member states, leading to new ways for automation and effective data exchange through IoT technologies, making possible the connection between vehicles and road infrastructure and road users [5].

Road traffic congestion is a problem that leads to delays, energy consumption, and environmental pollution [6]. ITS can improve citizens' sustainable urban mobility and quality of life, providing solutions to pollution, traffic congestion, and energy consumption reduction [7]. Hence, ITS can contribute to critical public policy and systematic management, reducing urban areas' traffic congestion and energy consumption.

In Portugal, drivers spend a daily average of 42 minutes in urban traffic and an average of 160 hours each year in traffic jams [8]. During the recent worldwide pandemic of Covid-19, Lisbon went into lockdown, remote work was enforced, and the capital citizens spent more time at home. This created a drop of about 30% in traffic congestion [9], with a level of congestion within the city of 23% during the entirety of 2020. Even with this drop in traffic congestion, Lisbon occupies the position 139th position in the ranking, above larger and more populous cities like Shanghai (152th), Barcelona (164th), Toronto (168th), San Francisco (169th), or Madrid (316th) [10]. Traffic contributes to pollution that costs the Portuguese citizens an average of 1,159 euros every year in medical bills and causes an average of 500 premature deaths [11]. Traffic congestion is also one of the causes of road accidents, considered one of the most severe problems in today's Portuguese society and a public health issue according to Autoridade Nacional de Segurança Rodoviária (ANSR) [12].

1.2. Research Questions and Goals

Our research aims to look at the traffic congestion phenomena with Lisbon as a case study. Lisbon is the capital of Portugal with a population of 508,368, with 2 million people in the urban metropolitan area daily commuting to Lisbon. Due to the pendular movement of the population, 370,000 vehicles enter Lisbon daily, adding to the 200,000 vehicles that already circulate in the city, making almost a total of 600,000 vehicles. In this thesis, we investigate and present an overview of traffic congestion in Lisbon by analyzing data from 2019 and the first semester of 2020. We aim to identify traffic patterns and roads with different levels of congestion, identify how people move in the city in the rush hours, and how much time commuters spend on transportation. We also envisage to provide data-driven guidelines and knowledge about traffic congestion in Lisbon in the framework of a traffic management and visualization tool.

With this intense focus on the visualization component of the overall traffic within the city, we aim to help policymakers and stakeholders to make better-informed decisions on traffic behavior and help them mitigate such phenomena.

This thesis addresses the following research questions:

Research Question 1: How can we characterize Lisbon road traffic patterns?

Research Question 2: How can we predict traffic delay in Lisbon analysing continuous variables?

1.3. Methodologic Approach

Cross-Industry Standard Process for Data Mining (CRISP-DM), adopted in our research, is a methodology that aims to create a standard approach to data mining projects to reduce costs, increase reliability, repeatability, and manageability, making the data mining process more efficient [13]. CRISP-DM [14] is composed of 6 phases, and we adapted the process towards data visualization, considering data fusion from different sources (Figure 1). The first and second phases, business understanding and data understanding, are where the data is collected, described, explored, and verified. In the third phase, data preparation, data is selected and cleaned, exploring and verifying data quality, integrated, and formatted. In the fourth phase, data fusion, we select the data sources, like traffic, accidents, weather, city infrastructure, pollution, into a data warehouse. In the fifth phase (data visualization), we define visualization templates to automatically visualize temporal and spatial data to define goals based on the city authorities' needs. Finally, the last phase, the decision, and the overall picture are provided to competent decision-makers.

Applying CRISP-DM involves business understanding looking at the aim of the challenges proposed by the Lisbon City Hall (CML) in the framework of Lisboa Inteligente - LxDataLab [15]. It also requires the definition of our strategy to address the research questions. In the data understanding phase, traffic datasets are described and categorized in features. This is followed by the data preprocessing phase, where the collected data is cleaned and normalized, generating new datasets to be used in the modeling phase, comprising analysis and visualization.

Figure 1 - CRISP-DM Methodology

1.4. Structure and Organization of Dissertation

This dissertation was organized into four chapters. In Chapter 1, we introduce the theme of the dissertation, the topic relevance, questions, research goals, and the applied methodology. Chapter 2 presents the bibliometric analysis using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Then, we use the VOSviewer tool to visualize scientific landscapes in our literature dataset find the latest state-of-the-art methodologies applied to road traffic congestion. Chapter 3, we apply the CRISP-DM methodology to our traffic dataset using the Python [16] programming language and its libraries to present the finding using a dashboard tool called Power BI [17], then we apply a time-series-based method to create a prediction model. Finally, in Chapter 4, we discuss the results, present our conclusions and propose future work to be done.

Chapter 2 – Literature Review

2.1. Methods

2.1.1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses

A systematic literature review (SLR) was performed to answer the proposed research questions in (1.2) using the PRISMA methodology [18]. This methodology is composed of a flow chart and aims to identify the research community's studies that meet the eligibility criteria in a transparent, unbiased, and reproducible manner.

This method [19] provided a general guideline on what papers should be included from the title, abstract, and full-text screening and how to approach them when creating the literature dataset. The search was performed in February 2021 with a time constraint limited to the last five years, between 2017 and 2021, to focus mainly on recent literature and advances done by the scientific community on the matter.

2.1.2. Keywords Identification and Research Query

To begin our literature search, we identified a research query to generate the results. To achieve this, we conducted a search using the google scholar search engine [20] to understand which words were used by the community with common expressions related to traffic prediction, traffic visualization, traffic patterns, and intelligent transportation systems. The results were screened, and a survey to find the keywords used by the research community on the subject which ones could answer our research objective:

- What is the most recent literature on traffic patterns and visualization?

The final selection of keywords for the research query was as follows:

- "Traffic Congestion" and
- "Traffic Prediction" or
- "Traffic Forecasting" and
- "Traffic Patterns" and
- "Visualization"

2.1.3 Repositories and Bibliometric Analysis

Two repositories, Scopus and IEEE, were used to conduct our SLR. Studies showed multiple fields of knowledge ranging from Energy, Chemistry, Engineering, Mathematics, Computer Science, among other fields.

We used one free web and desktop reference management application in this research, Zotero [21]. This application enabled the quick removal of duplicated papers and structured the systematic literature review paper dataset. This application allowed us to organize and search all references from one library and save the metadata from different repositories.

2.1.4 Bibliometric Research Tool for Network Analysis - VOSviewer

VOSviewer is a tool for creating and visualizing bibliometric networks based on co-citation, co-authorship, and text mining capabilities enabling the extraction of co-occurrence networks of terms extracted from scientific literature [22]. This tool was chosen to conduct our bibliometric analysis. It conducts analysis with particular attention to graphical representation, and it can handle large sets of bibliometric data [23]. In addition, this tool creates graphical representations and maps authors, abstracts, titles, and keywords [23].

2.2 Literature Review

2.2.1. PRISMA Results

The PRISMA flow diagram has four primary steps: Identification, Screening, Eligibility, and Included. Each one increases the granularity of our SLR dataset.

In the “Identification” step, the research query was applied to two repositories, Scopus and IEEE, and 40 papers were identified (23 IEEE and 17 Scopus) and a total of 83 conference papers. Only papers in English published in peer-reviewed journals and conferences were considered. Although both repositories were used in this research, we did not find a high number of duplicated papers, so after the first step, three duplicated papers were identified and removed.

The screening step consists of two phases, title screening and abstract screening. In the first screening step, a total of 77 papers were removed.

From the screening phase, we found papers related to network traffic and air traffic. These papers were outside of the scope of our research and were removed. We also found papers related to smart cities and transportation that were out of the scope of our research, such as papers related to subway traffic control, public transportation, parking lot traffic management, pedestrian flow, and crowd movements. After the screening phase, we proceeded to the eligibility phase, and in this phase, full-text articles were assessed for eligibility, and none of the selected papers were removed.

The PRISMA flow diagram represented in Figure 2 illustrates the creation of the final literature dataset in the various steps of our PRISMA, where "n" is the number of papers.

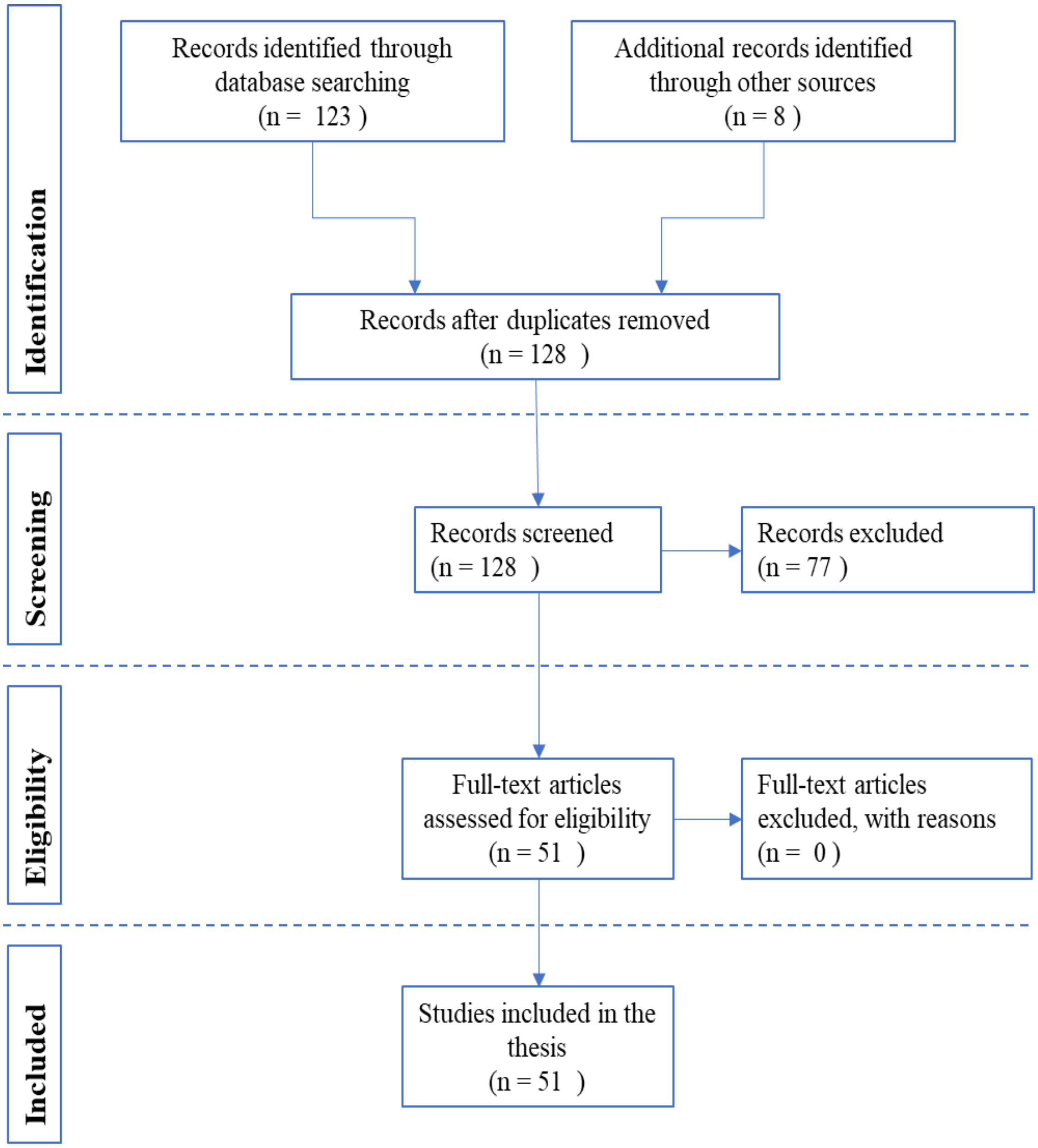

Figure 2 - PRISMA Flow Diagram

2.2.2. Identification of Main Journals

Our study resulted in a total of 32 journal papers that covered a wide variety of research fields, such as Computer Science, Engineering, Transportation, Automotive Engineering, and other fields, as shown in Table 1. The identified papers corresponded to a total of 17 different journals. The journal that showed more results was IEEE Access (9) followed by IEEE Transactions on Intelligent Transportation Systems (6), Transportation Research Part C:

Emerging Technologies (3), and IEEE Transactions on Visualization and Computer Graphics (2), followed by all the remaining journals with one result.

According to Scimago [24], most of the identified journals (24) are Q1 quartile ranked, which represents 75% of the journals, and the remaining are Q2 (4) and Q3 (5). The top publishing countries were the United States (18), followed by the United Kingdom (5), Netherlands (3), Switzerland (2), and then Germany (1), Romania (1), Egypt (1), and Indonesia (1).

Finally, the main publishers were Institute of Electrical and Electronics Engineers Inc. (16) followed by Elsevier Ltd. (4), IEEE Computer Society (2), Wiley-Blackwell Publishing Ltd (1), Hindawi Limited (1), MDPI Multidisciplinary Digital Publishing Institute (1), MDPI AG (1), IOS Press (1) Elsevier BV (1), Springer Berlin (1), Politechnica University of Bucharest (1), Inderscience Enterprises Ltd (1), Springer Berlin (1), Politechnica University of Bucharest (1), Science and Engineering Research Support Society (1), and Inderscience Enterprises Ltd (1)

Table 1 - Main Journals

Jornal	No.	Quartile Rank	Fields	Publisher	Publisher Country
IEEE Access	9	Q1	Computer Science; Engineering; Material Science	Institute of Electrical and Electronics Engineers Inc.	United States
IEEE Transactions on Intelligent Transportation Systems	6	Q1	Automotive Engineering; Computer Science; Mechanical Engineering	Institute of Electrical and Electronics Engineers Inc.	United States
Transportation Research Part C: Emerging Technologies	3	Q1	Automotive Engineering; Civil and Structural Engineering; Computer Science Applications; Management Science and Operations Research; Transportation	Elsevier Ltd.	United Kingdom

IEEE Transactions on Visualization and Computer Graphics	2	Q1	Computer Graphics and Computer-Aided Design; Computer Vision and Pattern Recognition; Signal Processing; Software	IEEE Computer Society	United States
IEEE Transactions on Vehicular Technology	1	Q1	Aerospace Engineering; Applied Mathematics; Automotive Engineering; Computer Networks and communications; Electrical and Electronic Engineering	Institute of Electrical and Electronics Engineers Inc.	United States
Computer-Aided Civil and Infrastructure Engineering	1	Q1	Civil and Structural Engineering; Computational Theory and Mathematics; Computer Graphics and Computer-Aided Design; Computer Science Applications	Wiley-Blackwell Publishing Ltd	United Kingdom
International Journal of Production Economics	1	Q1	Business Management and Accounting; Economics and Econometrics; Industrial and Manufacturing Engineering; Management Science and Operations Research	Elsevier Ltd.	Netherlands
Complexity	1	Q2	Computer Science; Multidisciplinary	Hindawi Limited	Egypt

Sensors (Switzerland)	1	Q2	Analytical Chemistry; Atomic and Molecular Physics and Optics; Biochemistry; Electrical and Electronic Engineering; Information Systems; Instrumentation; Medicine	MDPI Multidisciplinary Digital Publishing Institute	Switzerland
Arabian Journal for Science and Engineering	1	Q2	Multidisciplinary	Springer Berlin	Germany
ISPRS International Journal of Geo-Information	1	Q2	Earth and Planetary Sciences; Social Sciences	MDPI AG	Switzerland
International Journal of Advanced Science and Technology	1	Q3	Agricultural and Biological Sciences; Computer Science; Engineering	Science and Engineering Research Support Society	Indonesia
Journal of Intelligent and Fuzzy Systems	1	Q3	Artificial Intelligence; Engineering; Statistics and Probability	IOS Press	Netherlands
Journal of Traffic and Transportation Engineering (English Edition)	1	Q3	Civil and Structural Engineering; Transportation	Elsevier BV	Netherlands
International Journal of Embedded Systems	1	Q3	Computer Science	Inderscience Enterprises Ltd	United Kingdom
Applied Sciences	1	Q3	Chemical Engineering; Computer Science; Engineering; Materials	Politechnica University of Bucharest	Romania

	Science; Mathematics; Physics and Astronomy	

2.2.3. Identification of Main Conferences

Our research identified a total of 19 conference proceedings. All the conferences are displayed in Table 2. According to our SLR, all the conference papers were published in different conference proceedings. Our results demonstrated that only 5 conferences were from 2018, 5 from 2020, followed by 4 from 2019 and 3 from 2021, and finally 2 from 2017.

Table 2 - Main Conferences

Conference	No.
2021 International Conference on COMmunication Systems NETworkS (COMSNETS)	1
2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEECCOT)	1
2020 IEEE Region 10 Symposium (TENSYMP)	1
2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)	1
IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium	1
2019 20th IEEE International Conference on Mobile Data Management (MDM)	1
2020 6th International Conference on Signal Processing and Communication (ICSC)	1
2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)	1
2017 10th International Symposium on Computational Intelligence and Design (ISCID)	1
2017 IEEE Region 10 Symposium (TENSYMP)	1
2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)	1
2020 International Conference on Information Technology Systems and Innovation (ICITSI)	1
2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)	1
2019 6th International Conference on Systems and Informatics (ICSAI)	1
2019 IEEE Intelligent Transportation Systems Conference (ITSC)	1
2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)	1
2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)	1
2019 IEEE Symposium on Computers and Communications (ISCC)	1
2018 IEEE Global Communications Conference (GLOBECOM)	1

2.2.4. Keyword Occurrence Analysis

With the utilization of VOSviewer, a keyword occurrence analysis was performed on our literature dataset. This analysis was done with the full counting option, and the minimum number of occurrences of a keyword was set to 3. As a result, VOSviewer identified a total of 372 keywords in the dataset, from these 372, only 53 met the threshold, and from those, only 35 keywords were selected.

Most of the keywords were related to data models, traffic congestion, urban areas, machine learning, big data, visualization, and time series analysis. According to the results in Table 3, the top 5 identified keywords were roads (28 occurrences and link strength of 117), predictive models (23 occurrences and a link strength of 101), traffic congestion (18 occurrences and a link strength of 76), forecasting (14 occurrences and a link strength of 69) data models (11 occurrences and a link strength of 51).

Table 3 – Keywords Occurrences and Link Strength

Keywords	Occurrences	Total Link Strength
Roads	28	117
Predictive models	23	101
Traffic congestion	18	76
Forecasting	14	69
Data models	11	51
Urban areas	11	51
Deep learning	9	44
Neural networks	8	40
Traffic prediction	7	36
Roads and streets	8	31
Time series analysis	6	31
Machine learning	6	30
Visualization	6	30
Feature extraction	5	27
Trajectory	7	27
Learning systems	4	26
Traffic management	5	26
Big data	5	25
Prediction algorithms	5	24
Convolutional neural network	4	23
Data visualization	4	23
Intelligent transportation system	5	23
Prediction accuracy	4	23
Lstm	4	22

Training	4	19
Correlation	4	17
Hidden markov models	4	17
Long short-term memory	4	17
Road traffic	4	17
Motor transportation	5	16
Traffic control	4	16
Computational modeling	4	15
Prediction	3	15
Data mining	4	13
Visual analytics	4	12

VOSviewer identified 4 clusters, Figure 3 and Figure 4, with a total of 35 items and a total of 300 links in each cluster. We can highlight keywords related to data visualization (data visualization, visualization, visual analytics, and time-series analysis) in the color red, machine learning, data mining and traffic prediction in the color yellow, roads, urban areas predictive models in the color green, and traffic congestion big data and intelligent transport systems in the color blue. Also, in both Figure 3 and Figure 4, we can see a link between the different keywords in each node of the different clusters.

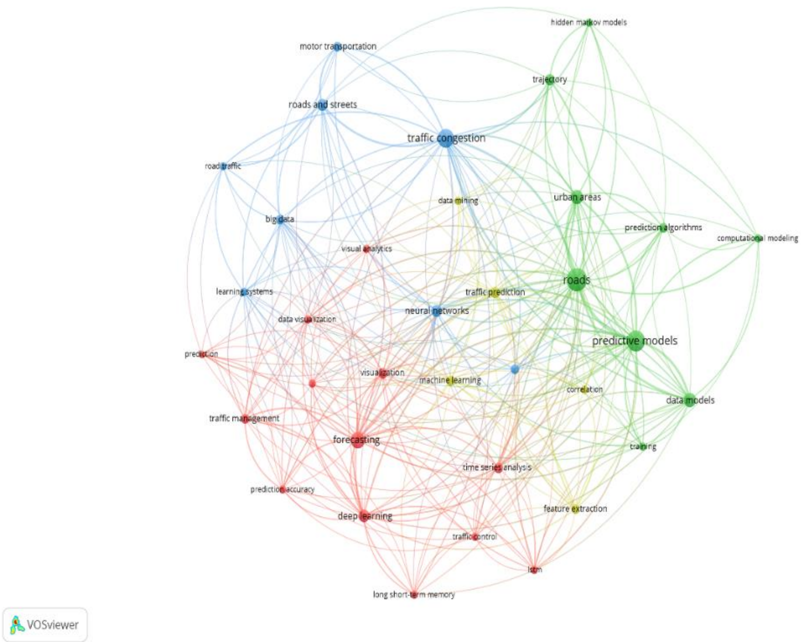

Figure 3 - Keyword Occurrence Visualization

When we analyzed Figure 4, we were also able to identify that the research community in the current year 2021 tends to use words like intelligent transportation systems, trajectories, and long short-term memory.

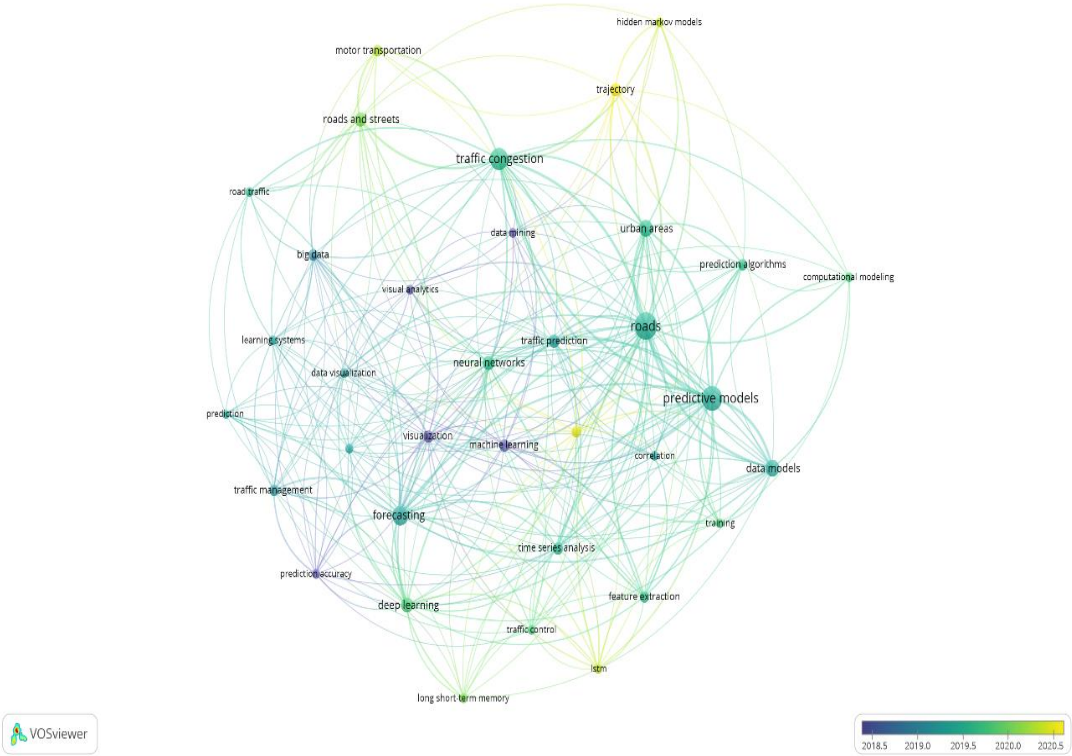

Figure 4 - Keyword Occurrence by Year Visualization

2.2.5. Title and Abstract Occurrence Analysis

VOSviewer has an option to analyze the titles and abstracts from our literature dataset. This analysis was performed with the full counting method, the minimum number of occurrences of a term set to 6, which resulted in the identification of 1698 terms from these, only 64 met the threshold. VOSviewer calculated a relevance score from those 64 terms, and based on that score, 54.6% of the most relevant terms were selected that corresponded to 35 terms represented in Table 4.

Table 4 – Title and Abstract Term Occurrences and Score

Term	Occurrences	Relevance Score
Car	6	4.338
Stability	7	4.3079
Busy traffic zone	6	3.4879
Application	10	1.5241
Point	8	1.4118
Traffic state	11	1.2783
Effectiveness	6	1.0518
Impact	9	0.9967
Addition	6	0.8895
Road congestion	8	0.8773

World	6	0.8544
Prediction accuracy	11	0.8216
Intelligent transportation system	15	0.8154
k-NN	6	0.7894
Performance	23	0.7368
Attention	6	0.7057
Person	8	0.702
Algorithm	31	0.6709
Historical data	7	0.6643
Traffic condition	17	0.6641
Road	48	0.6502
Work	9	0.588
Framework	16	0.5646
Number	9	0.5546
Intersection	9	0.5511
System	37	0.5508
Smart city	6	0.5117
Traffic management	10	0.4496
Urban road network	10	0.4333
Neural network	31	0.3553
Accuracy	18	0.3199
Experiment	9	0.313
Traffic jam	6	0.2874
Traffic congestion	53	0.2826
Accuracy	18	0.3199
Addition	6	0.8895
Algorithm	31	0.6709
Application	10	1.5241
Attention	6	0.7057
Busy traffic zone	6	3.4879
Car	6	4.338
Effectiveness	6	1.0518
Experiment	9	0.313
Framework	16	0.5646
Historical data	7	0.6643
Impact	9	0.9967
Intelligent transportation system	15	0.8154
Intersection	9	0.5511
k-NN	6	0.7894
Neural network	31	0.3553
Number	9	0.5546
Performance	23	0.7368
Person	8	0.702
Point	8	1.4118
Prediction accuracy	11	0.8216
Road	48	0.6502

Road congestion	8	0.8773
Smart city	6	0.5117
Stability	7	4.3079
System	37	0.5508
Traffic condition	17	0.6641
Traffic congestion	53	0.2826
Traffic jam	6	0.2874
Traffic management	10	0.4496
Traffic state	11	1.2783
Urban road network	10	0.4333
Work	9	0.588
World	6	0.8544

Figure 5 shows that this analysis performed by VOSviewer on the abstract and title created a network of 34 items and 311 links with a link strength in a total of 2135. This resulted in a total of five clusters (red, green, yellow, blue, and purple). The main nodes in each were traffic congestion (red), road (yellow), neural network (blue), urban road network (green), and stability (purple). Words related to traffic were the most identified in this analysis with traffic state, traffic congestion, traffic condition, traffic management, traffic jam. We also identified intelligent transportation systems and roads these terms also appeared in the keyword analysis.

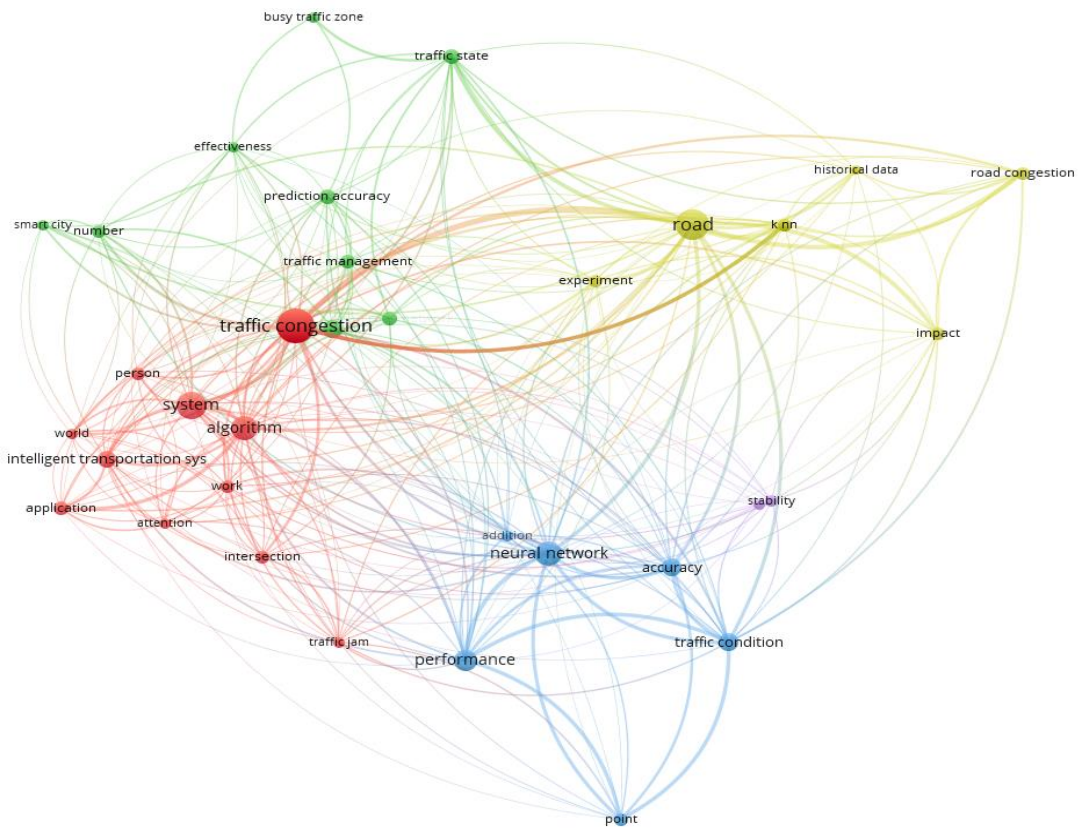

Figure 5 - Title and Abstract Network Visualization

Figure 6 shows that intelligent transportation systems, traffic congestion, and roads have a connection with almost all the other terms but were more used in the year 2018, 2019, and 2020 respectively.

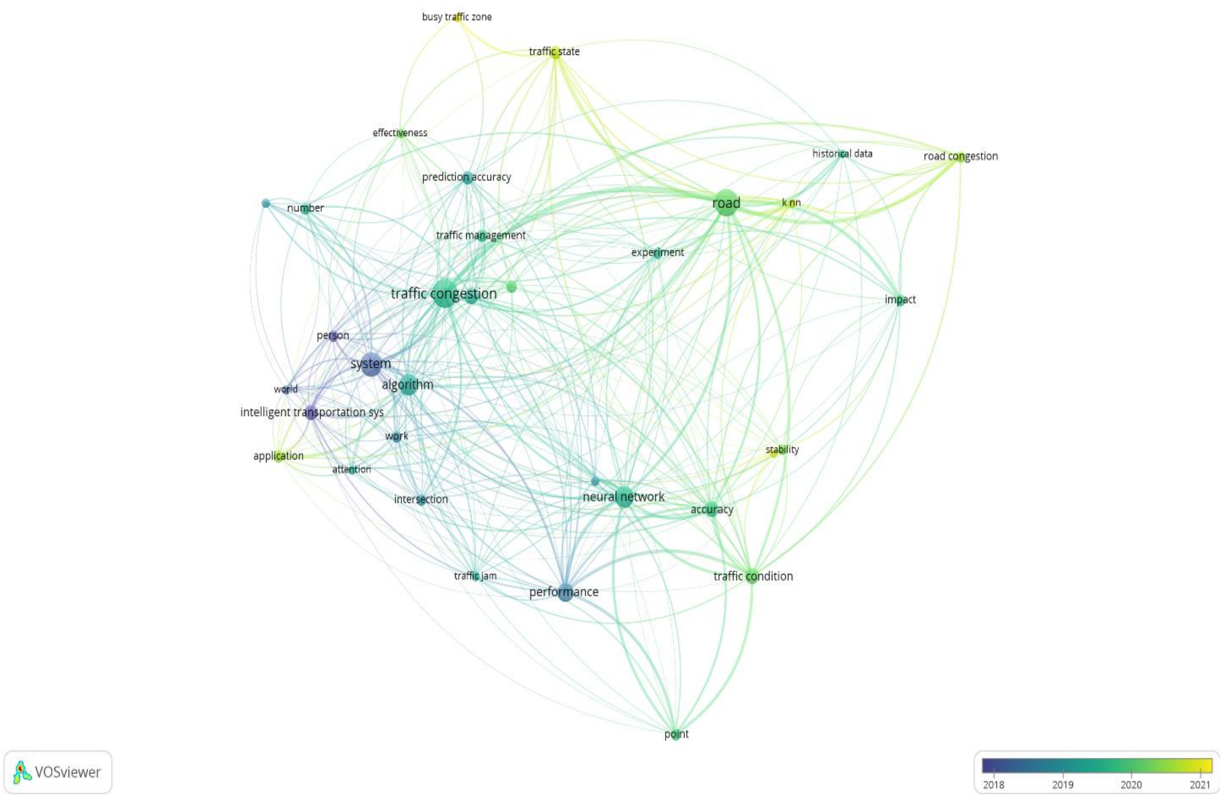

Figure 6 – Title and Abstract Text Network by Year

2.2.6. Author Co-Authorship Analysis

Analysis of the author and co-authorship was performed with the utilization of the same method (VOSviewer), and the counting method was set to full counting, and the minimum number of documents of an author was set to 1 this resulted in 215 authors out of 215 meet the threshold and from those 28 were selected for analysis (Table 5).

Table 5 - Author and Co-Authorship by Link Strength

Author	Documents	Total Link Strength
A. Drosou	1	7
A. Salamanis	1	7
A. Zamichos	1	7
D. D. Kehagias	1	7
D. Tzovaras	1	7
G. Margaritis	1	7
I. Kalamaras	1	7
S. Papadopoulos	1	7
C. Lee	1	6

D. Ebert	1	6
D. Kim	1	6
D. M. Farid	1	6
I. Alam	1	6
J. Ulisses	1	6
M. Alam	1	6
M. F. Ahmed	1	6
R. J. F. Rossetti	1	6
R. Maciejewski	1	6
S. Jin	1	6
S. Ko	1	6
S. Shatabda	1	6
Y. Kim	1	6
Wang, Fei Yue	2	1
Wang, Y.	2	1
Z. Kang	1	1
Z. Zhang	1	1
Zhang, Junping	1	1
Zhang, L.	1	1

The authors with the highest link strength were A. Drosou, A. Salamanis, A. Zamichos,D. D. Kehagias, D. Tzovaras, G. Margaritis, I. Kalamaras, S. Papadopoulos from the paper "An Interactive Visual Analytics Platform for Smart Intelligent Transportation Systems Management" [25] with a total link strength of 7 followed by C. Lee, D. Ebert, Y. Kim, S. Ko, S. Jin, D. Kim and R. Maciejewski from the paper "A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion" [26] and J. Ulisses, M. Alam, D. M. Farid, S. Shatabda, R. J. F. Rossetti, M. F. Ahmed, and I. Alam from the paper "Pattern mining from historical traffic big data" [27] with a link strength of 6. All the remaining authors had a link strength of 1, Wang, Y. and Zhang, I. [28], Junping Zhang and Fei Yue Wang [29], Z. Kang and Z. Zhang [30].

Figure 7 and Figure 8 show a total of 5 clusters with 72 links and 26 items and also that the scientific community tends to organize in groups of 7 authors.

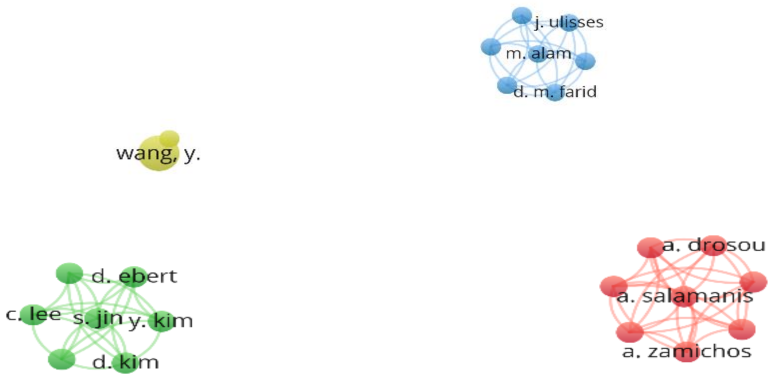

Figure 7 - Author and Co-Authorship Visualization

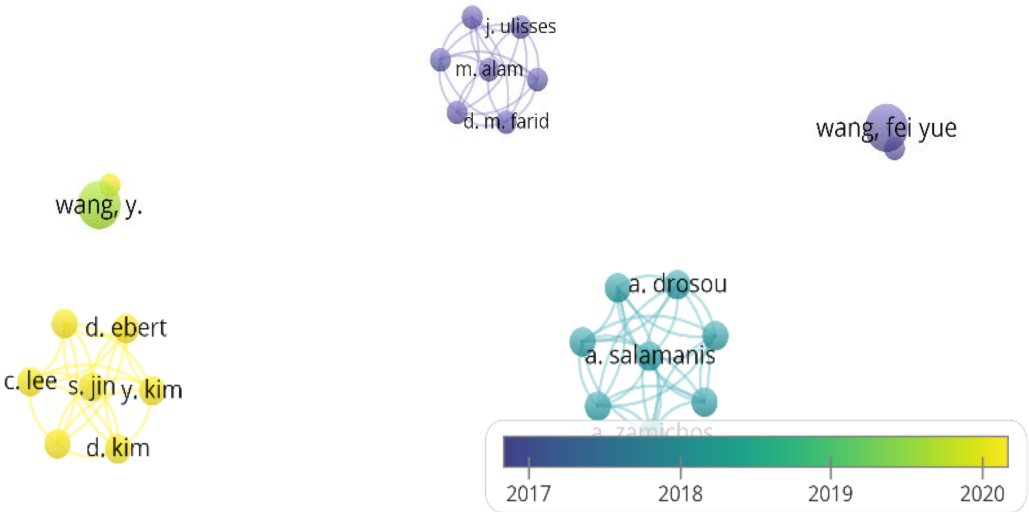

Figure 8 - Author and Co-Authorship by Year Visualization

2.2.7. Publication Ranked by Number of Citations

Identifying the most cited publications allowed us to have insights on what are the popular publications in the research community. This resulted in Table 6, where we can see the top 5 most cited papers, Junping Zhang, Fei-Yue Wang, (…), and Cheng Chen [29] with 747 citations, then the paper from Yuankai Wu, Huachun Tan, (…), and Zhuxi Jiang [31] with 261 citations, Wei Chen, Fangzhou Guo, Fei-Yue Wang [28] with 138 citations, followed by Zhiyong Cui, Kristian Henrickson, Ruimin Ke, and Yinhai Wang [32] with 72 citations, and finally Wang Xiangxue, Xu Lunhui, and Chen Kaixun [33] with a total of 33 citations, all the other authors had between 24 and 0 citations.

The publications can be consulted in Table 6 and are ranked by the number of citations these include the publications described above and all the others from our SRL dataset. Interestingly four of the most cited publications are from Q1-Ranked journals (IEEE

Transactions on Intelligent Transportation Systems and Transportation Research Part C: Emerging Technologies), also 3 of the top 5 most cited publications are from the same journal (IEEE Transactions on Intelligent Transportation Systems) and two of those are surveys. We also found out that IEEE Transactions on Intelligent Transportation Systems and Transportation Research Part C: Emerging Technologies are the most cited publications in our SLR e a total of 1299 total citations, making them the most cited journals of our SLR.

Table 6 - Publications Ranked by Number of Citations

Title	Authors	Year	Publication	Citations
Data-Driven Intelligent Transportation Systems: A Survey	Junping Zhang; Fei-Yue Wang; Kunfeng Wang; Wei-Hua Lin; Xin Xu; Cheng Chen	2011	IEEE Transactions on Intelligent Transportation Systems	747
A hybrid deep learning based traffic flow prediction method and its understanding	Yuankai Wu; HuachunTan; Lingqiao Qin; Bin Ran; Zhuxi Jiang	2018	Transportation Research Part C: Emerging Technologies	261
A Survey of Traffic Data Visualization	Wei Chen; Fangzhou Guo; Fei-Yue Wang	2015	IEEE Transactions on Intelligent Transportation Systems	138
Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting	Zhiyong Cui; Kristian Henrickson; Ruimin Ke; Yinhai Wang	2020	IEEE Transactions on Intelligent Transportation Systems	72
Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN	Wang Xiangxue ; Xu Lunhui ; Chen Kaixun	2019	Arabian Journal for Science and Engineering	33

2.2.8. Method and Application

With the advance of technology and networks, the volume and availability of traffic data resulted in the need for more data-driven approaches in ITS [34]. Our SLR covered traffic congestion in urban areas and the application of ITS. We found few articles in this field, focusing on data analysis and visualization of traffic congestion. Furthermore, we found that most articles use prediction models to address how external factors influence traffic in cities.

Data analytics combined with data visualization allows effective communication of research insights and provides visualization tools to policymakers and public authorities. According to Chen et al. [28], a pipeline of traffic visualization is an adequate tool to assess traffic data properties and discover hidden patterns in the data. Chen et al. [28] proposes an analysis based on four factors: temporal, spatial, Spatiotemporal, and multi-variable.

To reduce traffic congestion is necessary that the policymakers have intuitive traffic congestion platforms that help in the process of decision making [25], in this end, business intelligent visualization tools can provide great assistance, even more so if these platforms have real-time visualizations of the traffic state [25].

Traffic congestion analysis and visualization combined with car trajectories in peak hours [35] have proven to be an interesting tool for understanding mixed traffic conditions. In addition, a multi-variable analysis of vehicle types and car flow models improves mixed traffic flow trajectories.

Other factors can be analyzed in a multi-variable analysis and visualization regarding road conditions [35] and how they affect traffic congestion. In addition, tools such as RetinaNet enhance results training images and provide better performance metrics of road conditions.

Visual analytics provides the ability to analyze several tasks such as traffic congestion detection, accident monitoring, and flow patterns recognition. These tools are essential to what constitutes an ITS. According to Zhang et al. [29], having access to large quantities of data obtained through multiple sensor sources facilitate the development of a data-driven ITS based on vision, multisource, and machine algorithms. Such 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 et al. [36] used visual analysis techniques to study the relationship between traffic intensities and speeds practiced on the roads, using mathematical models. These models can 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. Lee et al. [26] created a visual analytics system designed to enable users to explore traffic congestion causes, propagation, and severity this system uses the Volume-Speed Rivers (VSRivers) visualization to present traffic volume and speed.

Traffic visualization is categorized into three groups: direct depiction, derived data visualization, and extracted patterns visualization [37]. Studies in the field proved multiple times the performance of visualization when we face a transportation problem [37].

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

The research community in traffic frequently uses neural networks machine learning techniques for predicting traffic flow. Those are the LSTM, Recurrent Neural Networks (RNN), Deep Neural Network (DNN), and other neural networks models.

ITS have different sets of applications, there are applications for traffic management (intersection monitoring, incident detection, and control, and vehicle classification) another set of applications can aid commuters on the road by providing maps and information in real-time [38], for these reasons ITS systems are considered essential tools for mitigating automobile traffic, and because of that, making 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 et al. [33] by combining LSTM and RNN models to evaluate two urban networks. The results showed that this approach brings better quality than other machine learning models.

Wu et al. [31] used DNN based traffic flow prediction that uses the complete periodicity, both weekly and daily, and the spatial-temporal characteristics of traffic flow and showed with visualization approach that this model surpassed other state-of-the-art models.

Modeling traffic conditions can be a difficult task with a tremendous negative impact on the citizens of a city. Chen et al. [39] observed that similar patterns occur in different neighboring times on consecutive days and that traffic has a multiscale property. With the utilization of a model based on deep CNN, Chen proposed a method called PCNN to predict short-term traffic conditions. This model outperformed other baseline methods in terms of short-term prediction [39]. Another example of the CNN method applied to a different type of dataset was proposed by Bui [40] the application of the method took into account a Sound dataset recorded in an urban road network, and this application showed promising results with accuracies of between 92% and 95%. Another example was proposed by Zhou [41], in this case, CNN was utilized to detect early congestion and operated over different time horizons, and results indicated that CNN-based methods had superior detection accuracy in that specific context.

Current prediction models have problems such as poor stability, significant data needs, and poor adaptability. Liu et al. [42] defined a model derived from the RNN, which combines Long Short-Term Memory (SDLSTM) with Auto-Regressive Integrated Moving Average (ARIMA). The result presents excellent adaptability and higher precision than standard machine learning methods. This hybrid model integrates computer vision, machine learning in a cloud computing environment.

Zheng et al. [43] used another LSTM model and compared it with the conventional machine learning models such as ARIMA and back-propagation neural network (BPNN), obtaining substantially superior results. Dong-Hoon Shin et al. [44] proposed the same method to make a prediction based on the correction of missing data, both temporal and spatial.

Wang et al. [45] defended that traffic congestion is not stationary and utilized the network theory, clustered the road points with traffic, and predicted the effects of traffic on the cluster using LSTM. Chu et al. [46] analyzed traffic on the lane level and utilized LSTM for traffic prediction, and the performance was improved, and the fitting error reduced when in comparison to other models.

According to our research, LSTM and its variations are fairly used when facing road traffic congestion problems. Zhiyong Cui et al. [32] proposed a Graph Convolutional Long Short-Term Memory Neural Network (GC-LSTM) to understand to forecast the entire traffic network and understand interactions between roadways, this method showed results that are 5% better compared to other models. Yauhui [47] used the same GC-LSTM method to create a traffic forecasting method and took into special account the temporal and spatial correlation of traffic flow. Another variation of the LSTM model was utilized by Di [48] for short-term prediction of congestion in different road segments, and that model outperformed 6 counterparts in the accuracy of its prediction.

Our research showed that CANN, Elman Recurrent Neural Network (ERNN) Shaokun Zhang et al. [30] used video surveillance data and CANN to predict traffic flow in certain positions and how weather, holidays, and temperature affected the results. Sadeghi-Niaraki A. et al. [49] used ERNN to predict traffic velocity and how weather and different temporal data affect traffic in Tehran.

k-NN, proposed by Lusiandro et al. [50] to visualize traffic intersection state and break down thickness, this model was also proposed by Priambodo B et al. [51] in this case, k-NN was utilized in combination with grey level co-occurrence matrix to predict traffic state and the high relationship between roads with surrounding roads.

Priambodo et al. [52] introduced a traffic prediction propagation models based on the HMM combined with GLCM based on sensors without any information about connected roads and studied spatial and temporal factors that influenced congestion in the area Tingting Sun et al. [53] proposed the same HMM and analyzed the correlation between external and internal traffic state and used traffic data from an area in Ningbo city, and the results in terms of accuracy reached 83.4%.

Traffic congestion prediction has been an ongoing problem since 1970 [54], studies have been done to mitigate this issue. Different problems have different approaches and different methods, Liu et al. [55] proposed the Random Forest algorithm due to its performance and robustness, Tao et al. [56] used the SSGRU to predict traffic flow.

Kundu et al. [57] investigated and compared the normal time-series and regression models with deep learning approaches and compared state-of-the-art models with more traditional approaches. Sinha [58] defined time series as an integral part of ITS. Zhou et al. [59] used time series forecasting models to predict traffic operation status, Chowdhury et al. [60] used the Circular Model of Traffic Forecasting (CMTF) to predict traffic in different nodes based on previous traffic data. According to He Z. et al. [61], identifying bottlenecks where traffic congestion occurs is essential to solving this problem a low-frequency PVD method was applied on intersections in a road network. Another example of time series utilization is using time series prediction based on group method data handling proposed by Song X. et al. [62], this approach displayed better results than alternative methods in daily traffic flow prediction.

Traffic has a multitude of temporal patterns, short-term (hourly, daily, and weekly) and long-term (monthly and yearly) [63]. Han et al. [65] proposed a framework based on deep clustering and used this method to supervise the representation learning in a visualized manner with unlabeled datasets to predict short-term traffic.

Husni [64] compared a multitude of different machine learning algorithms and concluded that Knowledge-Growing Bayes Classifier outperformed algorithms because it can adjust to changes in traffic conditions and predict as soon as the data is acquired and make decentralized predictions. Other interesting algorithms are used in traffic congestion prediction. Guo et al. [65] used ST-3DNet to capture spatial and temporal correlations in both local patterns and long-term patterns, ST-3DNet allowed to aggregate both in a final prediction outperforming other state-of-the-art algorithms. Sun et al. [66] proposed Fast Fourier transform combined with Wavelet analysis and applied these algorithms to three different datasets from England highways. Izhar et al. [67] considered diffident road-related factors and applied two different binary classifiers to avoid shortcomings with label generation when predicting traffic congestion. Alam et al. [27] analyzed historical traffic data and used a tool for traffic data observation and machine learning to find the annual average daily traffic and abnormal patterns in the historical data. Wang et al. [68] applied a matrix method based on wavelet decomposition for long-term prediction based on historical data, Zang L. et al. [69] used real-time traffic maps and did statistical analysis of the distance between vectors to find traffic patterns in certain periods, Ma X. et al. [70] used cluster analysis on matrices with different traffic patterns to explore temporal and spatial traffic congestion in large scale road networks, and [71] used video footage to calculate the percentage of congestion and a uses prediction via a clustering mechanism that makes its prediction based on images, Vighnesh et al. [72] used a similar method also based on image processing to calculate the number of vehicles on the road.

Table 7 illustrates the application and the method used in each paper:

Table 7 – Method and Application

Title	**Authors**	**Application**	**Method**
Data-Driven Intelligent Transportation Systems: A Survey	Junping Zhang; Fei-Yue Wang; Kunfeng Wang; Wei-Hua Lin; Xin Xu; Cheng Chen	Survey	discussing the functionality of key components and deployment of Data-driven ITS
A hybrid deep learning based traffic flow prediction method and its understanding	Yuankai Wu; HuachunTan; Lingqiao Qin; Bin Ran; Zhuxi Jiang	Use weekly and daily periodicity and spatial-temporal characteristics to predict traffic	Deep Neural Network Based Traffic Flow (DNN-BTF)
A Survey of Traffic Data Visualization	Wei Chen; Fangzhou Guo; Fei-Yue Wang	Survey	Traffic data visualization, overview of data processing techniques, summary of existing methods for depicting the temporal, spatial, numerical, and categorical properties of traffic data.
Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting	Zhiyong Cui; Kristian Henrickson; Ruimin Ke; Yinhai Wang	Propose a framework that can recognize the most influential road segments in a real-world dataset	Traffic Graph Convolutional Long Short-Term Memory Neural Network (GC-LSTM)

Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN	Wang Xiangxue ; Xu Lunhui ; Chen Kaixun	Data-driven framework to short-term traffic flow prediction for urban road networks	long short-term memory-recurrent neural network (LSTM-RNN)
Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data	He Z. ; Qi G. ; Lu L. ; Chen Y.	Identify the intersection traffic congestion in an urban road network	Probe Vehicle Data (PVD)-based method
Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis	Sepideh Kaffasha; An Truong Nguyena; Joe Zhu	Review and bibliometric analysis	Review of the application of ITS and the most recognized models with Big Data used in this context.
PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction	Meng Chen; Xiaohui Yu; Yang Liu	Short-term traffic congestion prediction and global traffic trend	Pulse Coupled Neural Network (PCNN)
A variational autoencoder solution for road traffic forecasting systems: Missing data imputation, dimension reduction, model selection, and anomaly detection	Guillem Boquet ; AntoniMorell ; JavierSerrano ; Jose Lopez Vicario	Learn how traffic data is generated and how to handle this data	Variational autoencoder (VAE)
Deep Spatial-Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting	Shengnan Guo; Youfang Lin; Shijie Li; Zhaoming Chen; Huaiyu Wan	Capture traffic correlations in spatial and temporal dimensions to find local and long-term patterns	ST-3DNet
A Match-Then-Predict Method for Daily Traffic Flow Forecasting Based on Group Method of Data Handling	Song X. ; Li W. ; Ma D ; Wang D ; Qu L ; Wang Y	Forecast daily traffic flow based on historical data	Contextual marching and time-series method for prediction based on group method of data handling (GMDH)
Prediction of Road Traffic Congestion	Yunxiang Liu; Hao Wu	Predict traffic congestion with	Random Forest

Based on Random Forest		variables like weather conditions, time period, special road conditions, and holidays	
An Interactive Visual Analytics Platform for Smart Intelligent Transportation Systems Management	Ilias Kalamaras ; Alexandros Zamichos ; Athanasios Salamanis ; Anastasios Drosou ; Dionysios D. Kehagias ; Georgios Margaritis ; Stavros Papadopoulos ; Dimitrios Tzovaras	Interactive visual analytics platform with historical and data prediction	Interactive interface
A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion	Chunggi Lee ; Yeonjun Kim; Seungmin Jin; Dongmin Kim; Ross Maciejewski; David Ebert; Sungahn Ko	Interactive visual analytics system that allows the user to see traffic conditions and forecasting.	Long Short-Term Memory (LSTM) and Volume-Speed Rivers (VSRivers)
A Fast Vehicular Traffic Flow Prediction Scheme Based on Fourier and Wavelet Analysis	Peng Sun; Noura AlJeri; Azzedine Boukerche	Application of two different thresholds values to decompose high-frequency noise and identify low-frequency traffic flow.	Fast Fourier transform and Wavelet analysis
Pattern mining from historical traffic big data	Ishteaque Alam ; Mohammad Fuad Ahmed ; Mohaiminul Alam ; João Ulisses ; Dewan Md. Farid; Swakkhar Shatabda; Rosaldo J. F. Rossetti	Analysis of historical big traffic data with the application of Java-based traffic data observation tool and machine learning to find the annual average daily traffic	Linear Regression, Sequential Minimal Optimisation (SMO) Regression, and M5 Base Regression Tree
Short-Term Traffic Prediction Based on DeepCluster in Large-Scale Road Networks	Lingyi Han; Kan Zheng; Long Zhao; Xianbin Wang; Xuemin Shen	Framework based on deep clustering	Deep clustering method to supervise the representation learning in a visualized manner

An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment	Gilberto Ochoa-Ruiz ; Alonso Andrés ; Angulo-Murillo ; Alberto Ochoa-Zezzatti	Multi-variable analysis and visualization regarding road conditions and how they affect traffic congestion	Detection of road conditions using RetinaNet
Traffic flow combination forecasting method based on improved LSTM and ARIMA	Boyi Liu; Xiangyan Tang; Jieren Cheng; Pengchao Shi	Comparison of traffic data singularity with the probability value in the dropout module and combining their unequal time intervals to achieve prediction	long short-term memory neural network and time-series autoregressive integrated moving average model (SDLSTM-ARIMA)
Traffic Flow Prediction Based on Cascaded Artificial Neural Network	Shaokun Zhang; Zejian Kang; Zhiyou Hong; Zhemin Zhang; Cheng Wang; Jonathan Li	Predict traffic flow at positions with the correlation of traffic data and road network distance	Cascaded artificial neural network (CANN)
Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning	Jianhu Zheng; Mingfang Huang	Traffic flow forecast through time series analysis.	Long short-term memory (LSTM) network
A Traffic Congestion Forecasting Model using CMTF and Machine Learning	Md. Mohiuddin Chowdhury; Mah mudul Hasan; Saimoom Safait; Dipankar Chaki; Jia Uddin	Traffic jam prediction model using pre-calculated density from node information table based on previous traffic data.	Time series analysis based on road nodes
Traffic Congestion Prediction by Spatiotemporal Propagation Patterns	PDF Xiaolei Di; Yu Xiao; Chao Zhu; Yang Deng; Qinpei Zhao; Weixiong Rao	A spatiotemporal model for short-term prediction of congestion level in each road segment.	long short-term memory neural network
Interactive, multiscale urban-traffic pattern exploration leveraging massive GPS trajectories	Wang Q.; Lu M.; Li Q.	Three-layer framework to recognize and visualize multiscale traffic patterns.	Visualization framework
Short-Term Traffic Flow Prediction Using the Modified	Sadeghi-Niaraki A.; Mirshafiei P.;	Short-term traffic flow prediction model with Modified	Elman Recurrent Neural Network

Elman Recurrent Neural Network Optimized through a Genetic Algorithm	Shakeri M.; Choi S.-M.	Elman Recurrent Neural Network model (GA-MENN)	
A Hybrid Stacked Traffic Volume Prediction Approach for a Sparse Road Network	Y. Tao; P. Sun; A. Boukerche	Predicting the traffic flow through a sparse road network	selected stacked gated recurrent units model (SSGRU)
A Novel Online Dynamic Temporal Context Neural Network Framework for the Prediction of Road Traffic Flow	Zoe Bartlett; Liangxiu Han; Trung Thanh Nguyen; Princy Johnson	Long and Short-term Traffic effect on traffic with the use of different temporal data segments	Prediction in an online framework with different temporal segments as inputs
Prediction of Traffic Congestion Based on LSTM Through Correction of Missing Temporal and Spatial Data	Dong-Hoon Shin; Kyungyong Chung; Roy C. Park	Traffic congestion prediction based on the correction of missing temporal and spatial values	Long short-term memory (LSTM)
Visual Cause Analytics for Traffic Congestion	Mingyu Pi; Hanbyul Yeon; Hyesook Son; Yun Jang	Analyze the cause of traffic congestion based on the traffic flow theory.	Visual Analytics system that analyzes the cause and influence of traffic congestion.
A review on traffic congestion detection methodologies and tools	Angayarkanni S.A. ; Sivakumar R. ; Ramana Rao Y.V.	Methodology review	Various traffic detection methods with vantages and disadvantages
Dynamic Traffic System Based on Real Time Detection of Traffic Congestion	Aditya Rao; Akshay Phadnis; Atul Patil; Tejal Rajput; Pravin Futane	Dynamic traffic system with footage that calculates the percentage of congestion	Prediction mechanism based on clustering
Early Identification of Recurrent Congestion in Heterogeneous Urban Traffic	Lin Zhu; Rajesh Krishnan; Fangce Guo; John W. Polak; Aruna Sivakumar	Automated congestion detection with different timeframes	Convolutional Neural Networks (CNN)
Foreseeing Congestion using LSTM on Urban Traffic Flow Clusters	Ziyue Wang; Parimala Thulasiraman	Predict traffic in a cluster road points affected by traffic using affinity propagation clustering	Long-short term memory neural network (LSTM)
Hybrid Feature Based Label Generation Approach for	Aamish Izhar; S. M. K. Quadri; S. A. M. Rizvi	Application of diffident road-related factors to avoid shortcomings with	Two different binary classifiers

Prediction of Traffic Congestion in Smart Cities		label generation when predicting traffic congestion.	
Implementation of the Advanced Traffic Management System using k-Nearest Neighbor Algorithm	Muchammad Arfan Lusiandro; Surya Michrandi Nasution; Casi Setianingsih	Simulation of Urban Mobility (SUMO) software to simulate traffic in an intersection	k-Nearest Neighbor (k-NN)
Multi-Step Prediction of Traffic Flow Based on Wavelet Decomposition Correlation Matrix	Zhumei Wang; Liang Zhang; Zhiming Ding	Analyzing the hidden patterns in the historical data of traffic flow	Correlation matrix sequence description method based on wavelet decomposition
Research on City Traffic Flow Forecast Based on Graph Convolutional Neural Network	Yaohui Hu	Using temporal and spatial correlation of traffic also provides a city traffic flow prediction method	Graph convolutional neural network
Sustainable Time Series Model for Vehicular Traffic Trends Prediction in Metropolitan Network	Adwitiya Sinha; Ratik Puri; Udit Balyan; Ritik Gupta; Ayush Verma	Analysis of traffic patterns for predicting transport trends	Utilization of sensor data in a specific location
Traffic Flow Prediction Model Based on LSTM with Finnish Dataset	Qingling Chu; Guangze Li; Ruijie Zhou; Zhengdong Ping	Traffic Flow Prediction Model by lane	long short-term memory (LSTM)
Traffic Forecasting with Deep Learning	Shounak Kundu; Maunendra Sankar Desarkar; P. K. Srijith	Deep learning models compared to standard time series and regression models	Multiple deep learning models
Traffic Prediction Using a Supervised Learning Approach	Vighnesh; Sanjana D	Adress traffic congestions caused by inefficient handling of traffic	Image processing
Stability Analysis of Mixed Traffic Flow using Car-Following Models on Trajectory Data	H R Surya; Narayana Raju; Shriniwas S Arkatkar	Analysis of Mixed Traffic Flow with different vehicles categories	Car-Following Models equations and trajectory data
Congestion Pattern Prediction for a Busy Traffic Zone Based on the Hidden Markov Model	Tingting Sun; Zhengfeng Huang; Hongdong Zhu; Yanhao Huang; Pengjun Zheng	Congestion pattern prediction model for a busy traffic zone	hidden Markov model (HMM)

Predicting Traffic Conditions Using Knowledge-Growing Bayes Classifier	Emir Husni; Surya Michrandi Nasution; Kuspriyanto; Rahadian Yusuf	Compare different algorithms to find the more reliable one for predicting traffic patterns	Knowledge-Growing Bayes Classifier
Predicting Traffic Flow Propagation Based on Congestion at Neighbouring Roads Using Hidden Markov Model	Bagus Priambodo; Azlina Ahmad; Rabiah Abdul Kadir	investigate different roads in a neighboring area based on the similarity of traffic conditions.	Grey level of co-occurrence matrix (GLCM) combined with hidden Markov model
Research on Traffic Situation Analysis for Urban Road Network Through Spatiotemporal Data Mining: A Case Study of Xi'an, China	Ruiyu Zhou; Hong Chen; Hengrui Chen; Enze Liu; Shangjing Jiang	predict the traffic operation status and explore spatiotemporal traffic congestion patterns.	Time-series forecasting models: Curve Fit Forecast (CFF), Exponential Smoothing Forecast (ESF), Forest-based Forecast (FBF).
Traffic Density Classification Using Sound Datasets: An Empirical Study on Traffic Flow at Asymmetric Roads	Khac-Hoai Nam Bui; Hyeonjeong Oh; Hongsuk Yi	Traffic density classification based on the road sound datasets.	Convolutional neural network (CNN)
How to Identify Patterns of Citywide Dynamic Traffic at a Low Cost? An In-Depth Neural Network Approach with Digital Maps	Zhang L; Gong K; Xu M.; Li A.; Dong Y.; Wang Y.	Analyze the distances between vectors and extract different traffic patterns.	Stacked convolutional autoencoder-based method
Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative matrix factorization algorithm	Ma X.; Li Y.; Chen P.	Cluster analysis on matrices with different traffic patterns to explore temporal and spatial traffic congestion	Matrix factorization algorithm
Spatio-temporal K-NN prediction of traffic state based on statistical features in neighbouring roads	Priambodo B. ; Ahmad A. ; Kadir R.A.	Predict traffic by determining the high relationship of roads within neighbouring roads.	k-nearest neighbour (K-NN)
Experiences from Supporting Predictive Analytics of Vehicle Traffic	Natalia Andrienko ; Gennady Andrienko ;	Traffic simulation based on spatially abstracted transportation	Interactive visual interfaces

	Salvatore Rinzivillo	networks using dependency models of real traffic data	

Chapter 3 – Data Analysis and Modeling

Our method followed the CRISP-DM approach and followed all the designated steps business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Our data understanding and data preparation phases resulted in various interactions with the Lisbon City Hall that also evaluated the model.

3.1. Business Understanding

Our study addresses challenges “General traffic index and indexes for the main Lisbon city entrances” [73], launched by Lisboa Inteligente - LxDataLab [15] to the academia. Our study's objective is to investigate and identify traffic congestion patterns in Lisbon’s metropolitan area, define when and where traffic congestionoccurs, how they relate, and how external factors such as weather and pandemics affect such phenomena.

3.2. Data Understanding

The data understanding phase aims to collect, describe, explore, and verify the data's quality. This step is structured in three subphases: Describe, explore and verify data quality.

Traffic congestion datasets were provided and collected from Lisboa Inteligente – LxDataLab [15] on the scope of the challenges launched to the academia, as mentioned previously.

The traffic congestion dataset included Waze data [73], pre-processed by Lisboa Inteligente - LxDataLab [15], with data from traffic jams in the Lisbon metropolitan area, in a period ranging between 1st January 2019 to 30th July 2020. The provided dataset had a table structure with 25 columns and 12,619,459 rows in comma-separated values (CSV) file format.

The dataset provided had already been pre-processed 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 the columns (Bbox, entity_location, entity_type, fiware_service, fiware_servicepath, and pubMillis), deemed irrelevant by the CML, were removed. Additionally, the column Country was removed because all data was from Lisbon (Portugal).

Table 8 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 8 - 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_servicepath	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.

3.3. Data Preparation

3.3.1. Visualizations Dataset

This phase of the CRISP-DM methodology is subdivided into four sub-phases: data selection, data cleaning, feature selection, and data integration. All columns were selected except for Country, Typeentity, and Validity_ts. The column Country has no value since the data is from Lisbon, Portugal. The column Typeentity had only five levels NONE (12,531, 806 entries), Small (46,417 entries), Medium (25,794 entries), Large (15,313 entries) Huge (129 entries), with NONE being the value in almost all of the rows and for that reason, this column was removed. The column Validity_ts was also removed, given that this thesis's primary focus is on the occurrence of each event and not the archive date, so the column Entity_ts was used to the detriment of Validity_ts (date of LxDataLab archive – Table 8).

After the final column selection, all the columns were carefully inspected, and we made a few improvements in the data quality, as depicted in Table 9:

Table 9 - Waze Dataset - Details and Data Transformation

ID	Column	Chosen	Type	Defects Detected	Corrections Applied
1	City	Yes	object	1963605 Rows were empty 10137 Rows had the value "Null."	All the missing values and nan were replaced with "Lisboa" Because all the data is from Lisbon metropolitan area
2	Country	No	object		Not considered for analysis
3	Delay	Yes	float64	8779033 Value "-1" when the level of traffic is 5	-1 treated as a missing value using and replaced with a value prediction using the Pearson Correlation and Linear Regression
4	Endnode	Yes	object	3471012 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	1620642 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 column delay 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 Pearson correlation and Linear Regression (LR) [42], [43]. Diagonal (Table 10) 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 10 - Pearson Correlation Table.

	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

Applying the values predicted by the LR algorithm to the original dataset, we can observe an increase in the mean and median and a decrease in the standard deviation, which can be explained by the increase in traffic occurrences with a high delay (level 5 traffic level). Table 11 shows the impact of the application of predicted values to the original dataset:

Table 11 - Original vs Post-prediction Delay Results.

	Original Delay	Post-prediction 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, 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. Time Series Application (Prophet Method)

Prophet uses time series, and we chose this method because it uses non-linear trends that fit with different timeframes it can leverage yearly, weekly, and daily seasonality, plus holiday effects, and is good at handling missing data and outliers to make its predictions [74]. This robust method was chosen in our thesis to create a prediction based on the metropolitan traffic within Lisbon.

We used the visualizations dataset to create the new dataset that was used for the application of the Prophet method. The application of this method receives 2 variables, ds and y, so we had to adhere to that. For the temporal variable (ds), we loaded the field year_month_day and changed the name to ds, and for the values (y), we have chosen the field "delay" and changed the name to y also added the length to this new data frame and then calculated the mean for the y and length and then grouped by ds resulting in a dataframe with 531 lines and 3 columns (Table 12).

Table 12 – Chunk of the Transformed Dataframe for Prophet

Line Number	ds	y	Length (m)
1	2019-01-01	183.40	310.12
2	2019-01-02	184.62	433.19
3	2019-01-03	157.97	399.68
…	…	…	…
531	2020-06-30	240.35	214.04

3.4. Traffic Modelling

3.4.1. Traffic Modelling Visualizations

Our main objective was to create the main traffic indicator with general traffic information. Figure 9 shows this indicator filtered by freeways according to our data, the average traffic delay for the entire city (Freeways and Primary streets) is 3.90 minutes, and the average delay for freeways is 5.27 minutes, also in freeways, the average traffic length is 1.71 kilometers, and the average traffic velocity is 19.52 kilometers per hour.

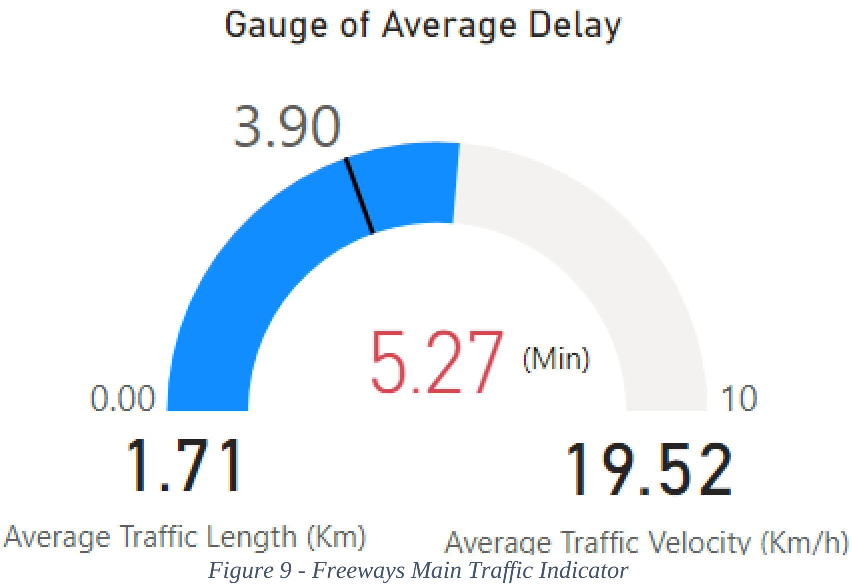

Figure 9 - Freeways Main Traffic Indicator

Figure 10 shows this indicator filtered by primary streets our data demonstrates again the average traffic in the city (3.90 minutes), primary streets and freeways included, and the average delay for primary streets is 3.86 minutes, and the average traffic length is 0.19 kilometers, and the average traffic velocity is 3.01 kilometers per hour,

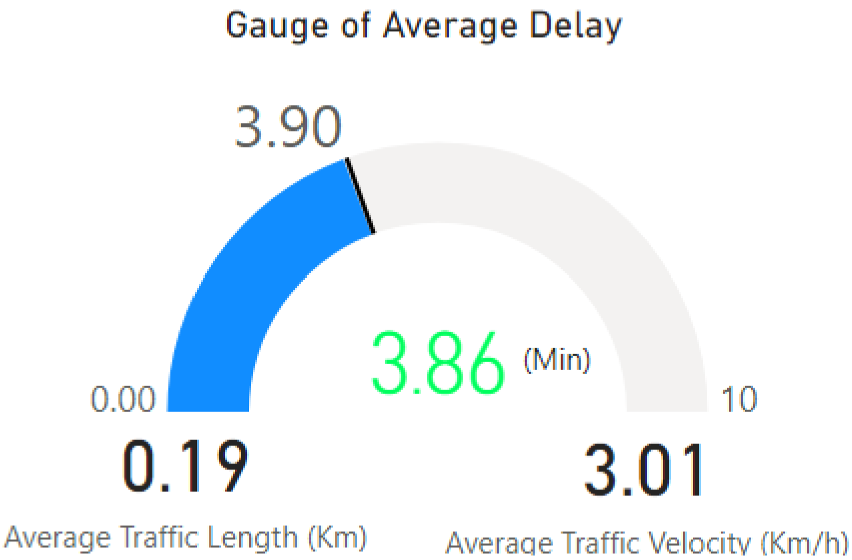

Figure 10 - Primary Streets Main Traffic Indicator

Our modeling results show that in 2019, 96.90% of the traffic congestion occurred in Primary streets, and 3.10% of the traffic congestion occurred in freeways, as shown in Figure 11.

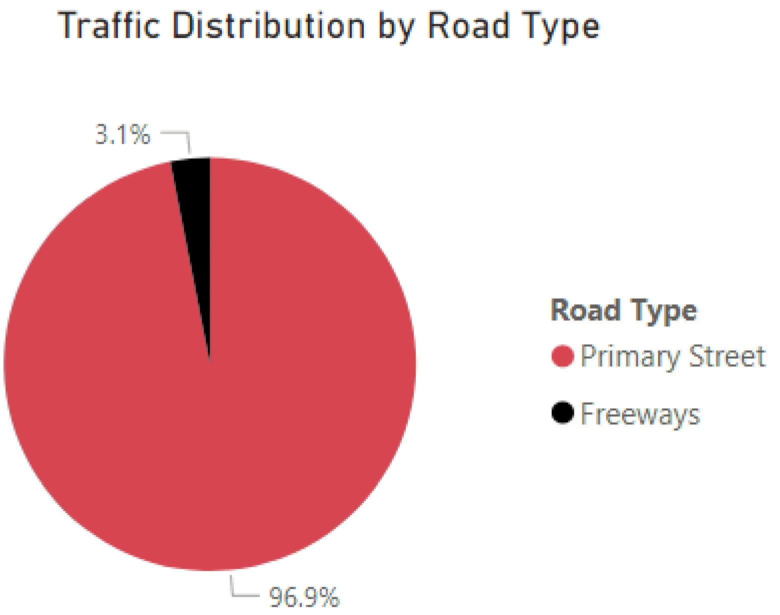

Figure 11 - Traffic Distribution by Road Type in Lisbon in 2019.

Figure 12 displays the distribution of traffic levels presented for both freeways and primary streets. 64.76% of the traffic corresponds 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).

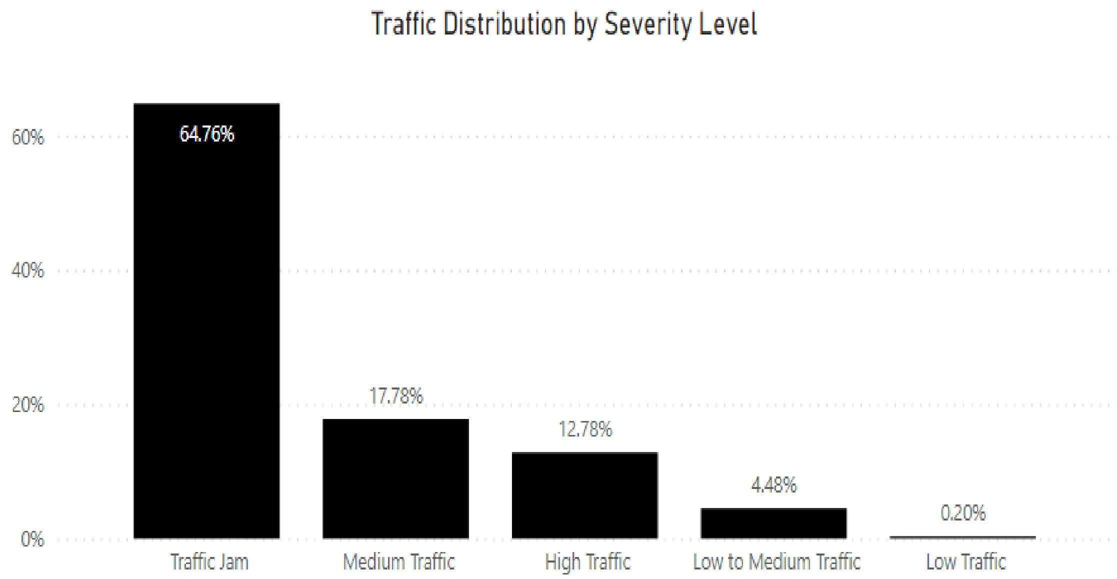

Figure 12 - Traffic Distribution by Severity Level in Lisbon in 2019.

Our data analysis (Figure 13, Figure 14) shows that level 5 traffic occurrences happened 5.39% in freeways versus 66.66% in primary streets, and level 4 traffic occurred 34.30% in freeways instead of 12.09% in primary streets. The data also showed that level 3 traffic is more common in freeways, with 44.28% vs 16.93% in primary streets, and level 2 traffic is more common in freeways, with 14.62% compared to primary streets with 4.16%. The model also shows that level 1 traffic is the less common in both freeways (1.41%) and primary streets (0.16%).

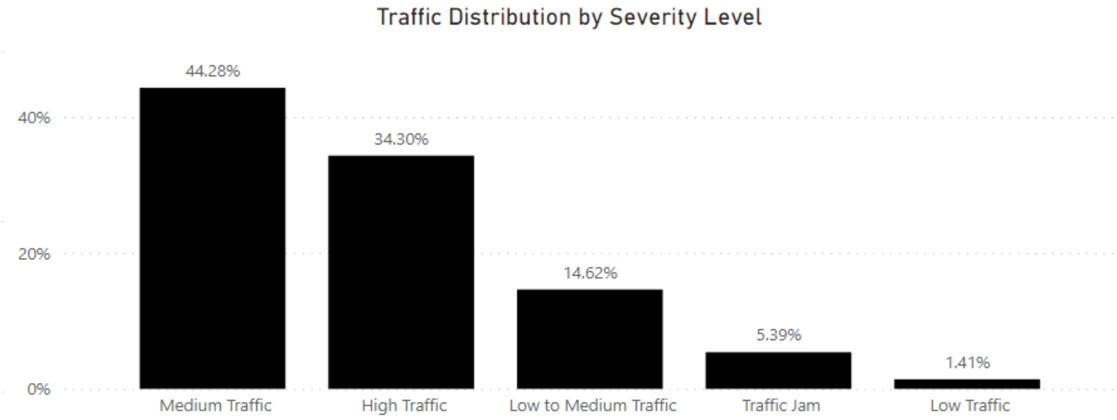

Figure 13 - Traffic Distribution in Freeways

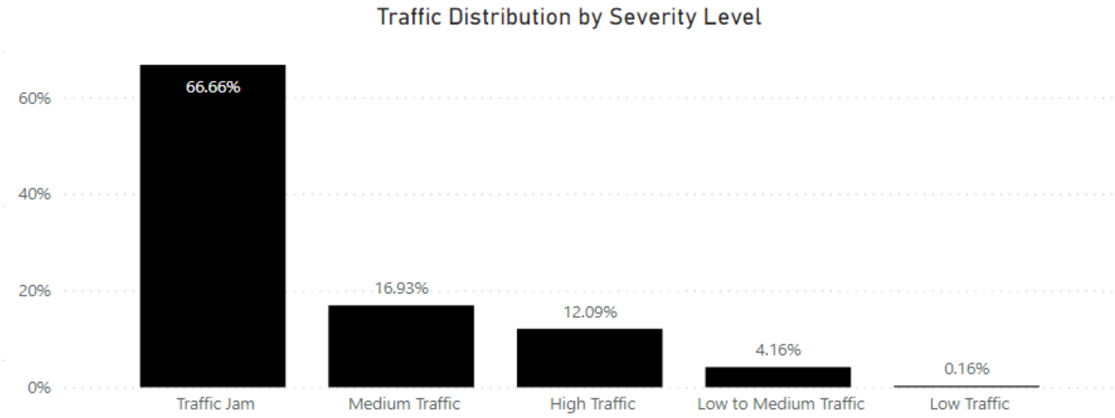

Figure 14 - Traffic Distribution in Primary Streets

The average traffic length in freeways is 1.71 kilometers, and the average delay is 5.27 minutes. In comparison, primary streets had an average traffic length of 0.19 kilometers and an average delay of 3.86 minutes. Table 13 displays the average traffic delay and length distributed by road type and traffic level.

Table 13 - Average Delay and Traffic Length by Road Type.

Road Type	Traffic Occurrence	Average Delay (minutes)	Average Traffic Length (kilometers)
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 am-9 am and between 5 pm-6 pm and make 24.77% of daily traffic occurrences (11.34% in the morning and 13.43% in the afternoon).

Lisbon boroughs Paço do Lumiar (23.53%), Sacavém (22.95%), Lumiar (20.33%), Ajuda (12.27%), and Belém (6.45%) together, display 85.53% of the Lisbon metropolitan area's traffic occurrences (Figure 15).

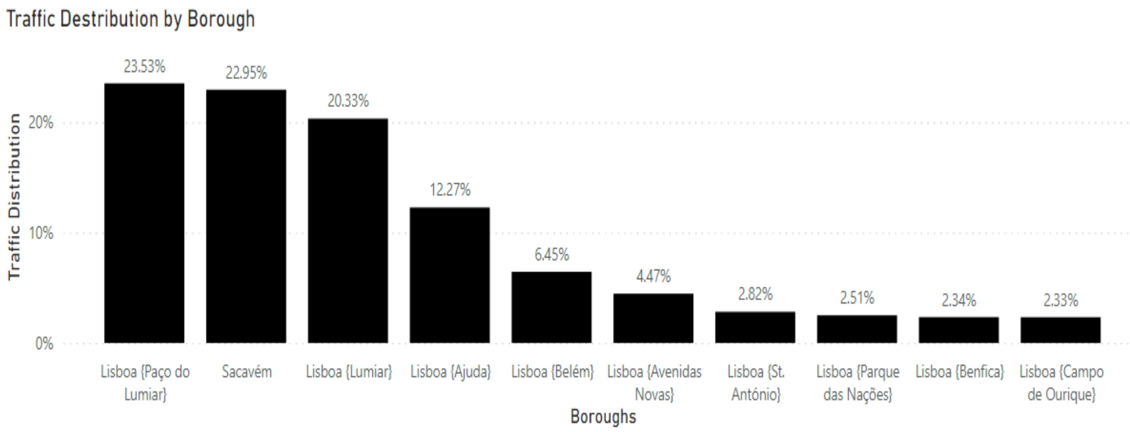

Figure 15 - Lisbon Traffic by Borough

In 2019 (Figure 16), the months with higher traffic were October (16.61%), November (16.91%), and December (15.95%), and the months with the lower traffic were April (1.83%), followed by February (2.76%) and January (2.88%).

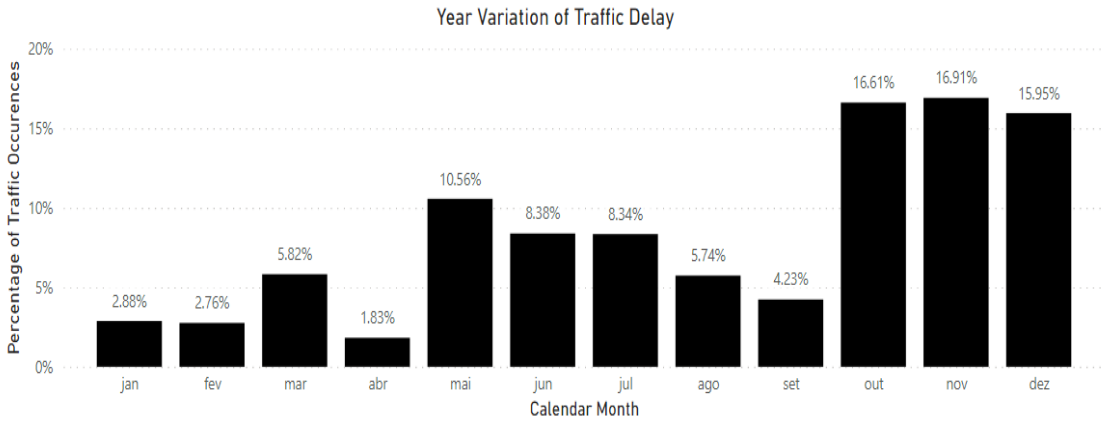

Figure 16 - Traffic Occurrences by Month

Our analysis (Figure 17) also shows that the freeway IC17 and CRIL have a higher percentage of traffic occurrences (26.61%) and have an average delay of 4.84 minutes, followed by A5 with 22.16% and an average delay of 4.53 minutes.

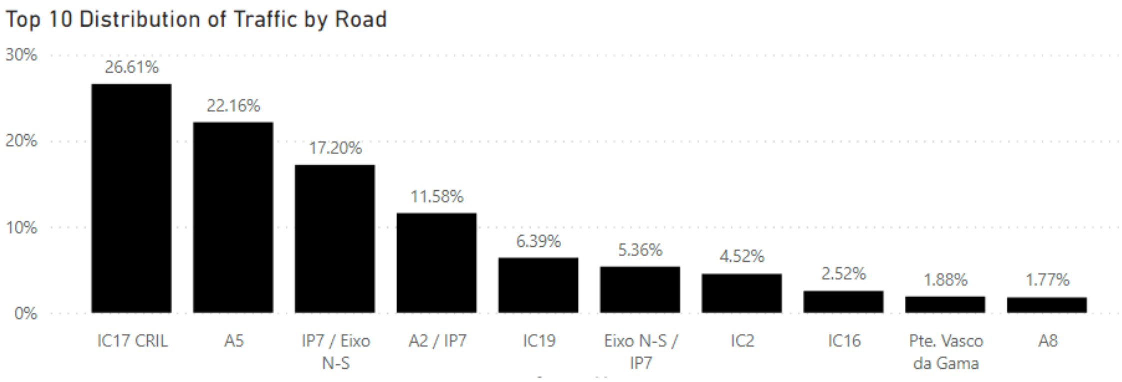

Figure 17 - Traffic Occurrences Distribution in Freeways

The freeway with the highest delay is IP7 / Eixo N-S, with an average delay of 8.07 minutes, followed by IC-19 with an average delay of 5 minutes. Figure 18 shows the distribution of traffic in Lisbon freeways.

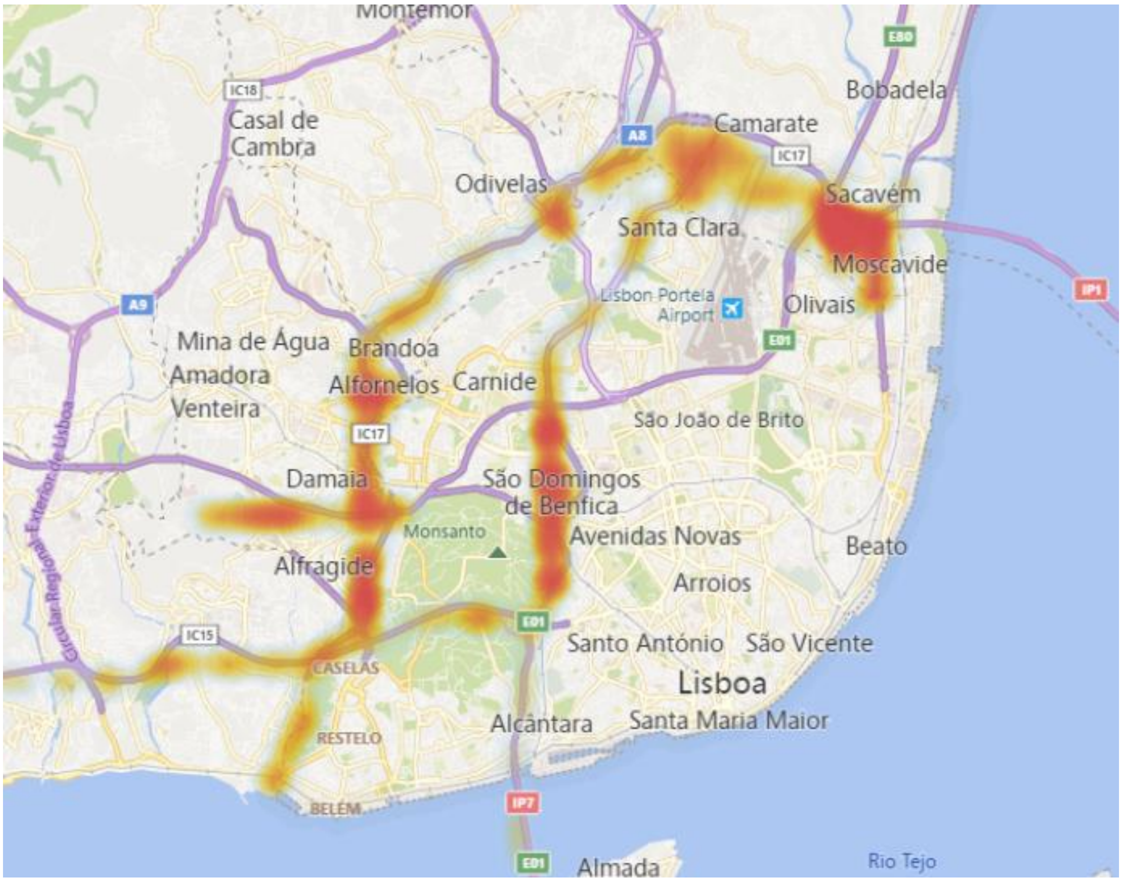

Figure 18 - Distribution of Traffic on Freeways in Lisbon in 2019. Red Corresponds to Locations Where More Traffic Congestion Occurred on Freeways; Orange and Yellow are Locations Where Traffic Congestion Was Less Common on Freeways.

Primary roads (Figure 19) analysis shows that Rua Direita in Lumiar, Rua Auta da Palma Carlos in Sacavém, and Calçada da Ajuda in Ajuda combined, make up to 54.75% of the traffic occurrences in Lisbon primary streets.

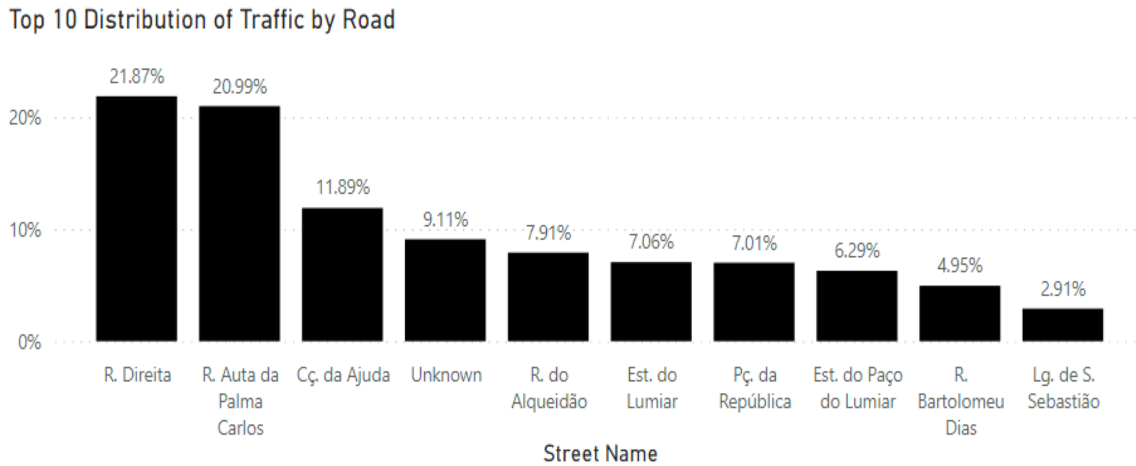

Figure 19 - Traffic Distribution in Primary Streets

According to the data, these streets are also the ones with the highest delays, where drivers experience an average delay of 5 minutes.

Figure 20 shows the distribution of traffic on primary streets.

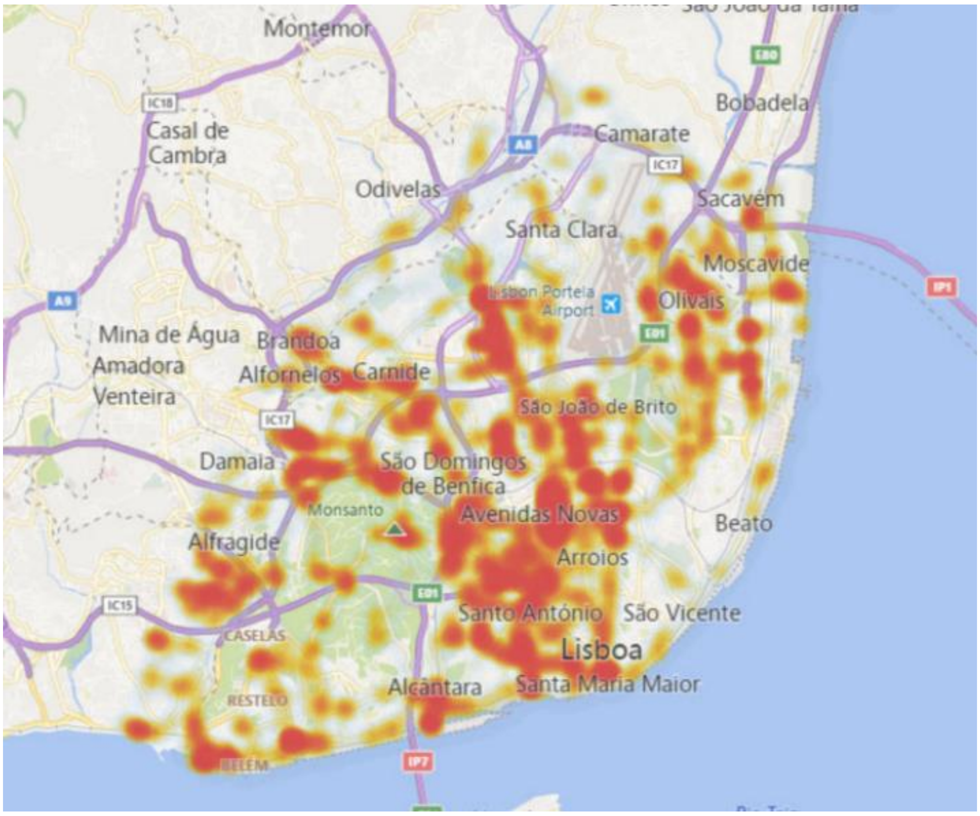

Figure 20 - Distribution of Traffic on Primary Streets in Lisbon in 2019. Red Corresponds to Locations Where More Traffic Congestion Occurred on Primary Streets; Orange and Yellow are Locations Where Traffic Congestion was Less Common on Primary Streets.

According to our data, we can see that there was a decrease in the time the drivers spent on traffic (Figure 21). This decrease is expressly observable in May 2020 until the end of our dataset in July 2020. According to our data, the total number of experienced delays in hours is around 4415.00 hours in 2020 and 213,710.00 hours in 2019, and the total traffic length is 9676.00 kilometers in 2020 and 590,840.00 kilometers in 2019.

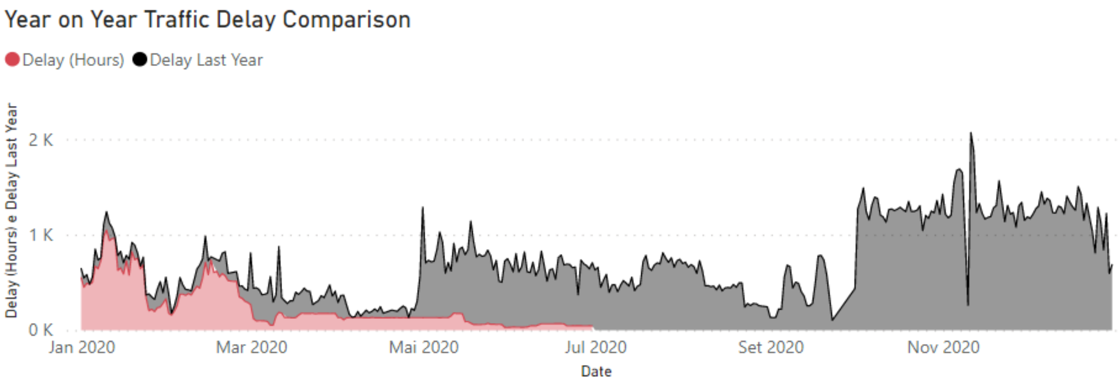

Figure 21 - Comparison Between Traffic Delay from 2020 (Red) and 2019 (Grey)

We can also see a decrease in the traffic length (Figure 22). For example, when comparing both the years 2019 and 2020, there is a decrease in traffic from March 2020 until July 2020.

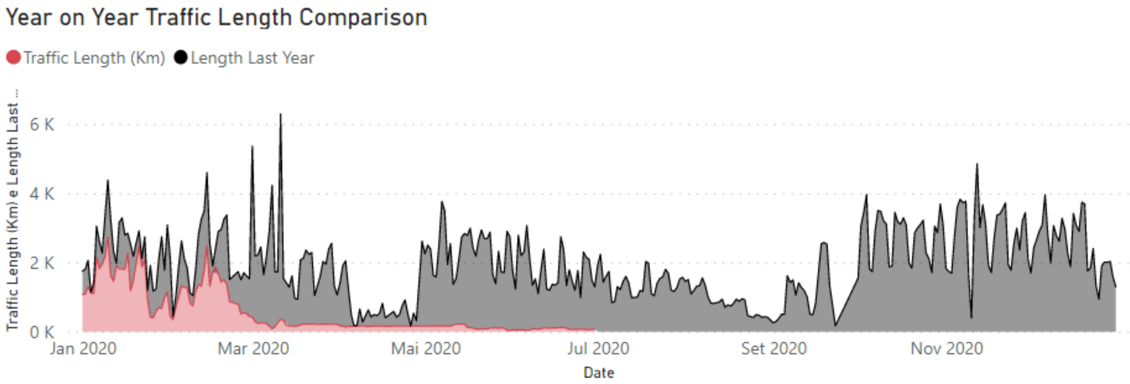

Figure 22 - - Comparison Between Traffic Length from 2020 (Red) and 2019 (Grey)

Looking at the traffic data from 2019 (Figure 23), we discovered that the traffic occurrence that happened more often are occurrences of medium traffic, followed by traffic jams and then high traffic. Also, the occurrence that tends to happen less is low traffic and then low to medium traffic for both years 2019 and 2020.

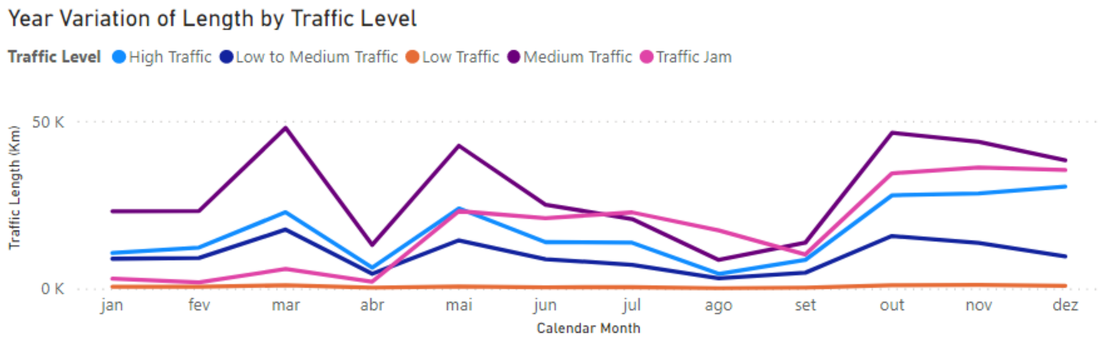

Figure 23 - Traffic Level Variation in 2019

Looking at the data from the first semester of 2020 (Figure 24), we found that traffic jam is the type of traffic occurrence that tends to happen more, followed by medium traffic and high traffic

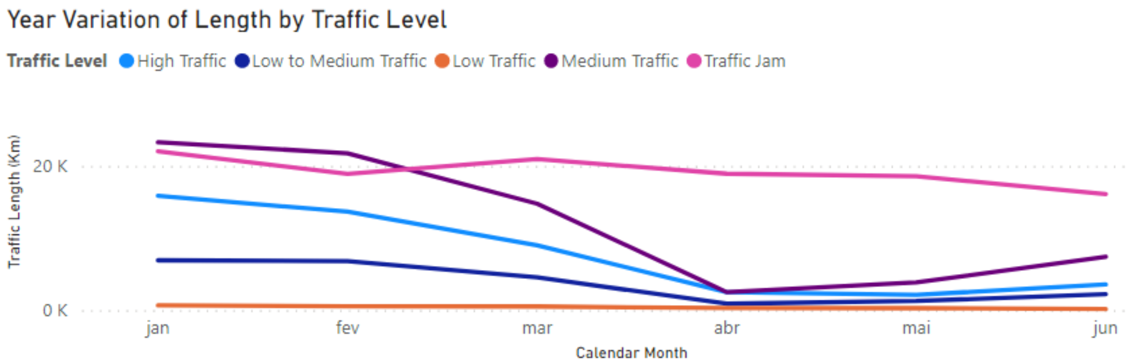

Figure 24 - Traffic Level Variation in 2020

The daily traffic occurrences (Figure 25) show that traffic is more severe at 8 am with 6.04% of traffic occurrences and then decreases between 9 am and 5 pm and reaches a new peak at 6 pm with 6.86% of traffic occurrences the period with less traffic is between 12 am and 6 am. with around 2% to 3% of traffic occurrences.

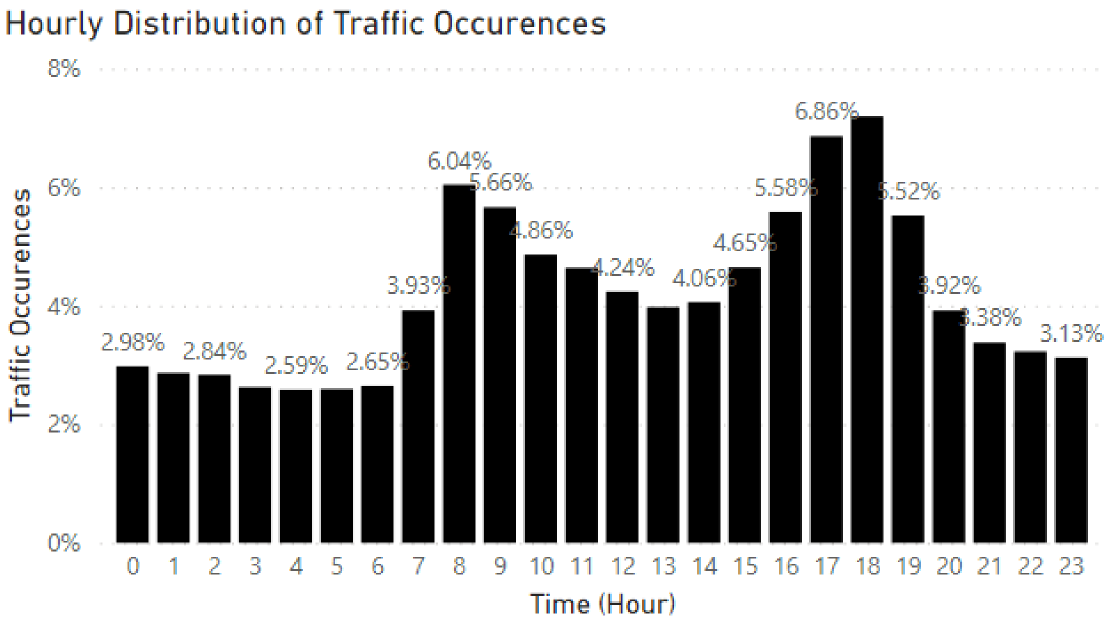

Figure 25 - Daily Traffic Occurrences by Hour

Figure 26 and Figure 27 show the traffic occurrences for two specific days, 1st January 2019 and 1st June 2019 this visualization displays where the traffic jams are located using the geojson coordinates and allow the user to see the traffic distribution in the Lisbon metropolitan area according to our data both these days had an average delay of 3 minutes and the average traffic length was between 0.2 kilometers and 0.3 kilometers and the roads with more traffic occurrences were Av. 5 de Outubro, Rua Viriato and Av. de Brasília for 1st January 2019 and for 1st June 2019 the roads with more traffic occurrences were R. Direita, Cç da Ajuda and Est. do Lumiar.

Figure 26 - Traffic Occurrences in 2019/01/01

Figure 27 - Traffic Occurrences in 2019/06/01

3.4.2. Facebook Prophet Application

When applying the Prophet method, we started by plotting the traffic delay raw data (Figure 28), and we found that the average of the delay varied between 140 seconds (2.33 minutes) and less than 300 seconds (5 minutes).

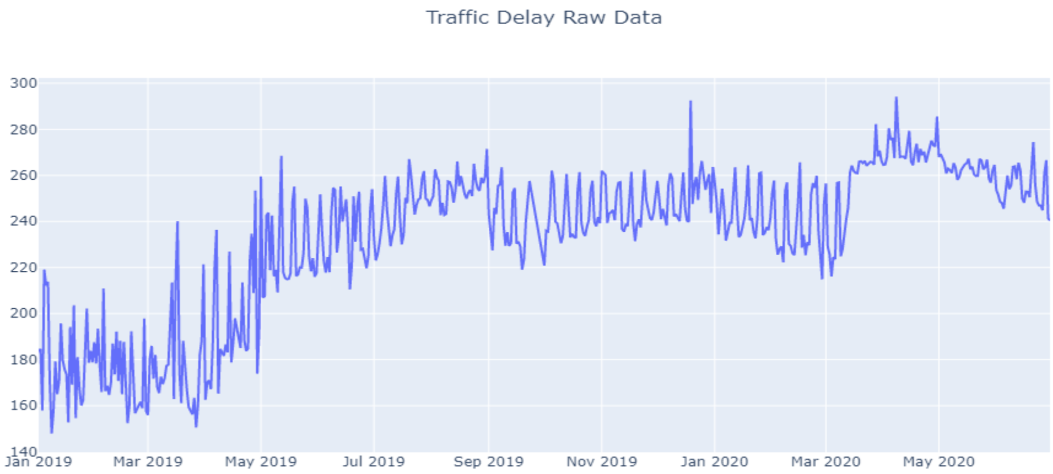

Figure 28 - Traffic Delay Variation

We noticed that some of the variability could be caused by the weekends when the majority of Lisbon citizens are not commuting to work. Therefore, we proceeded to remove the weekends (Figure 29):

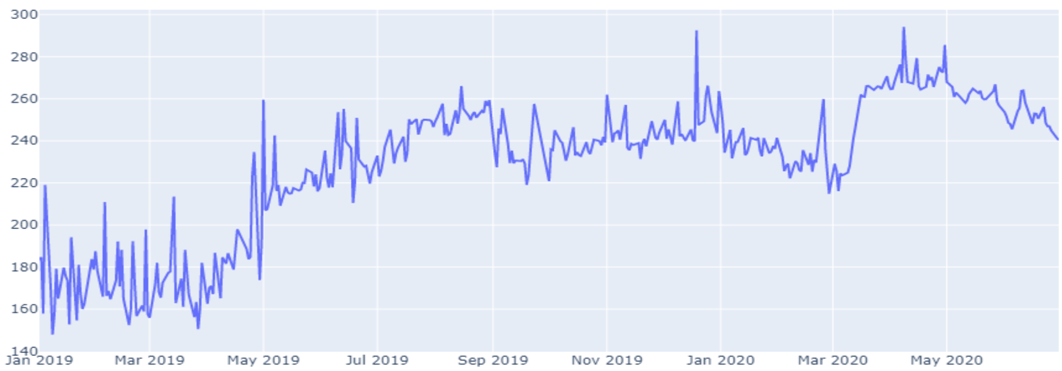

Figure 29 – Traffic Delay Variation Without Weekends

To achieve our prediction, we added some regressors, first we used the built-in holidays (New Year, Good Friday, Easter Sunday, Corpus Christi, Republic Day in Portugal, All Saints' Day, Portugal Restoration of Independence Day, Freedom Day, Labour Day, Portugal Day, Assumption of Mary, Feast of the Immaculate Conception,

Christmas Day), then we included the day of the Lisbon traditional festivities, called St. Anthony’s day (13th of June), added weekly and yearly seasonalities, and finally, we added the last regressor, traffic length, to our model. With these, we are able to find variations in the data and improve our prediction. Hence, we create a visualization with these regressors, temporal data, trend, actual value, and predicted value, resulting in Figure 30. Figure 31 shows the predicted value of Prophet and the actual value of our data.

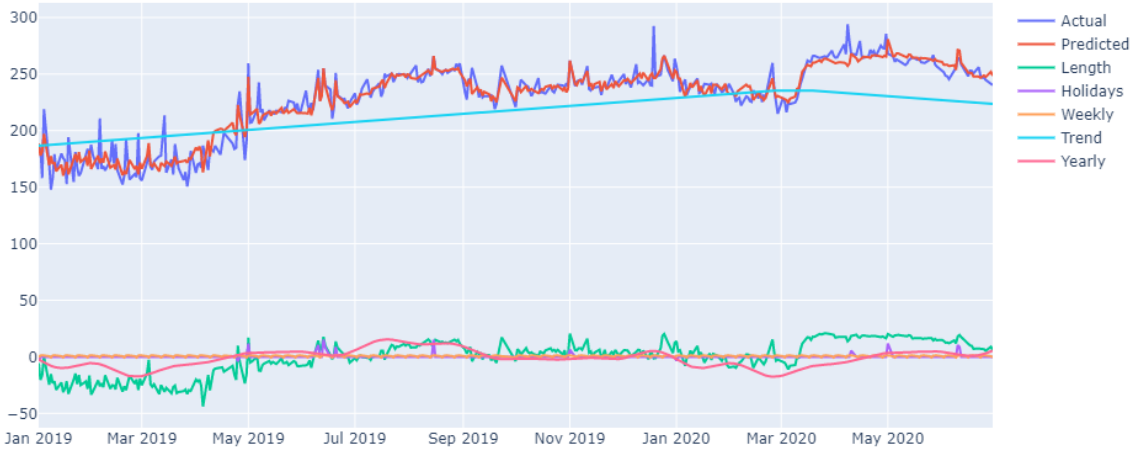

Figure 30 - Traffic Actual Value vs Prediction With Regressors, Temporal Data, and Traffic Length

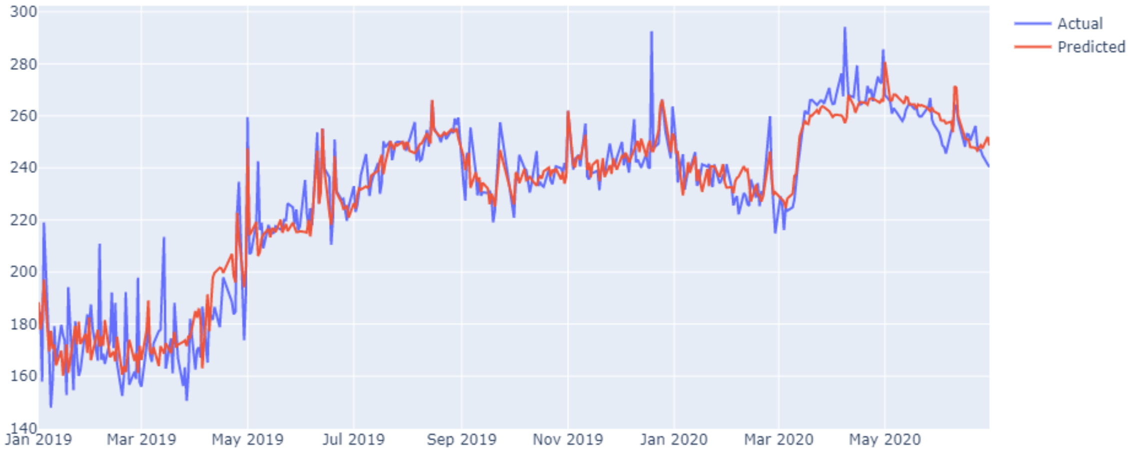

Figure 31 - Actual Data vs Predicted Data

3.5. Evaluation

The evaluation of our model relied on one meeting with CML, where we presented the results and was followed by sharing a questionnaire (Table 14) with the evaluation of our model. The model was evaluated by Engineer Sérgio Costa from CML, with five evaluation criteria, utility, understandability, consistency, level of detail, consistency, and robustness, and four levels of evaluation Not Achieved (NA), Partially Achieved, (PA), Largely Achieved (LA), Totally Achieved (TA).

Table 14 – Questionnaire Results

Criteria	Objective Statement	Evaluation	Evaluator
Utility	Can it help business decisions regarding traffic management?	(PA)	Sérgio Costa
Understandability	Provides understandable results	(LA)	Sérgio Costa
Consistency	Can be used without training	(PA)	Sérgio Costa
Level of detail	Provides knowledge from the mobility patterns	(TA)	Sérgio Costa
Consistency	Gives consistent results (compared with another country or company)	(LA)	Sérgio Costa
Robustness	Has enough detail to be used in other cases of Traffic Management	(PA)	Sérgio Costa

The main objective of our thesis was to provide knowledge of the patterns of traffic congestion within the city to the decision-makers. This was totally achieved by creating dashboards. Moreover, our results were considered understandable and consistent compared to other works in the same field. Our thesis was developed to provide an overview of the traffic and aimed to help in the decision-making process, but more information is needed to make these decisions. Furthermore, we provide knowledge of road traffic congestion management and not other use cases of traffic management, also our models are designed for policy-makers as such, it requires training to extract the value from the data provided. Even with these constraints, we partially achieved all these objectives of the Lx Data Lab, as shown in Table 14.

To evaluate our predictions, we applied the cross-validation technique and extracted some performance metrics for a horizon of 100 days (Table 15). After the calculation of these metrics, we applied the mean of the horizon. As a result, we got the following results: 423.88 Mean Square Error (MSE), 20.08 Root Mean Square Error (RMSE), 17.26 Mean Absolute Error (MAE), 0.07 Mean Absolute Percentage Error (MAPE), 0.07 Median Absolute Percentage

Error (MdAPE), 0.07 Symmetric Mean Absolute Percentage Error (SMAPE) and 0.42 of coverage.

Table 15 – Performance Metrics

Horizon	MSE	RMSE	MAE	MAPE	MdAPE	SMAPE	Coverage
10 days	270.35	16.44	13.87	0.06	0.05	0.06	0.50
11 days	299.43	17.30	14.83	0.06	0.07	0.06	0.42
12 days	327.86	18.11	15.85	0.06	0.07	0.07	0.37
13 days	351.50	18.75	16.67	0.07	0.07	0.07	0.32
14 days	361.31	19.01	16.97	0.07	0.07	0.07	0.32
				(…)			
98 days	463.68	21.53	19.12	0.08	0.06	0.08	0.46
99 days	526.58	22.95	20.59	0.08	0.06	0.08	0.39
100 days	578.97	24.06	21.61	0.09	0.07	0.09	0.38

According to Figure 31, our prediction line is very similar to our actual values. Our MAPE (Figure 32) demonstrated a variation between 2% and 10% when predicting the traffic delay for 100 days into the future.

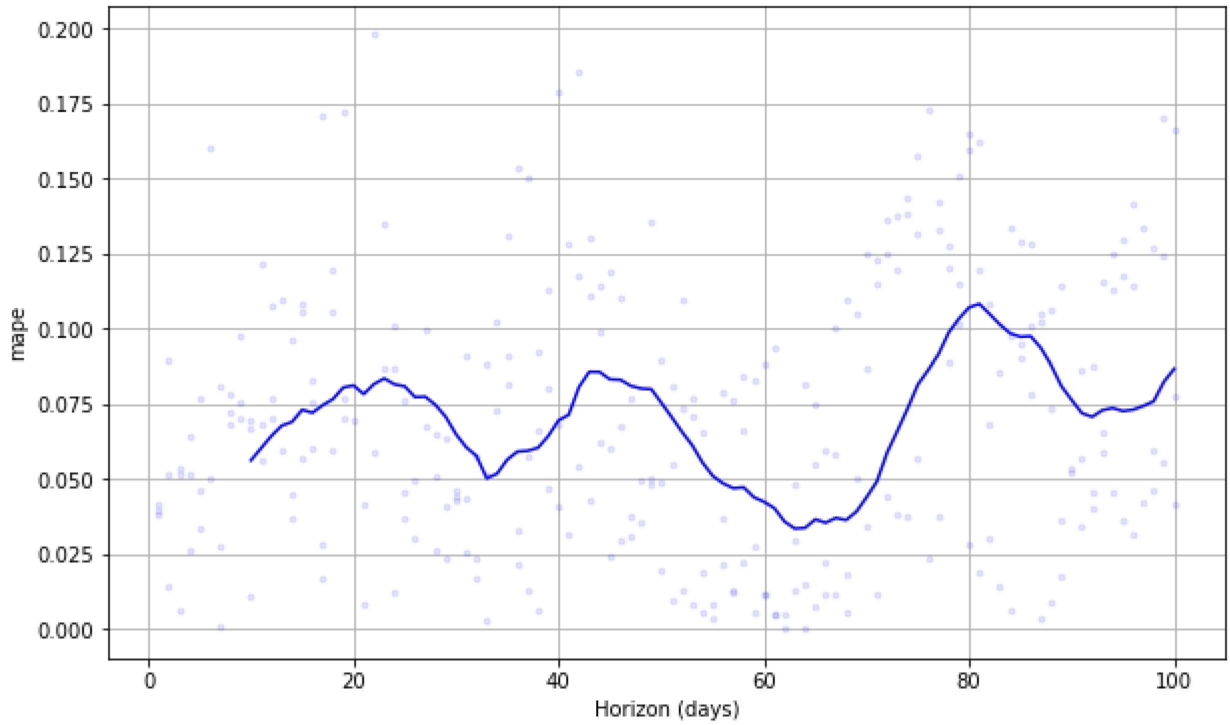

Figure 32- MAPE Variation

3.6. Deployment

The CRISP_DM methodology has a deployment phase, to achieve this, the dashboard and the visualizations were deployed in the following link: [Tese_BI DashBoards - Power BI] and are available for consultation.

In the link above is possible to consult the following dashboards:

- Traffic Indicator.
  - This Dashboard contains the main Indicator defined in the challenge's objectives from Lx Data Lab: Creation of general traffic indicator and mapping each of the main entrance routes into the city [73].
- Traffic Delay.
  - This Second Dashboard contains the traffic distribution and occurrences by Traffic level. It also shows the total delay in hours for the specific selected period and the total length of traffic in Lisbon.
- Traffic By Boroughs.
  - The traffic by boroughs dashboard displays the traffic grouped by the specific zones within the city of Lisbon, in this case, freeways have no assigned zone and appear as “Lisboa”.
- Traffic By Road.
  - The traffic by road display of traffic by specific road also shows the end node where the traffic ends.
- Traffic By End Node.
  - This dashboard is similar to the traffic by road the main difference is in the aggregation, in this dashboard, the aggregation is done by the end node instead of the road.
- Yearly Traffic.
  - The Yearly traffic dashboard displays the traffic variations in the specified year it also shows how the traffic length varies according to the level of traffic for the given month.
- Year Comparison.
  - The year comparison dashboard shows the variation of traffic and compares it to the previous year.
- Monthly Traffic.
  - The monthly traffic dashboard displays the information from a chosen month, and a given day of the month, it also displays the weekends and weekdays, allowing the comparison of traffic occurrences by day.
- Daily Traffic.

  - The Daily traffic dashboard allows us to see the variation of traffic during the day it also enables us to find when and where the rush hour happens and the traffic distribution during the day.

- Holiday Traffic.

Finally, we have the holiday dashboard where we can compare the holiday traffic against other specific holidays allowing us to see which have an increase or a reduction in traffic.

Chapter 4 – Conclusions

4.1. Discussion

The major finding addressing our research question 1 on Lisbon traffic patterns is that the predominance of traffic occurrences in primary streets originates most of the traffic events in the Lisbon metropolitan area.

Our research led us to discover that the majority of traffic events are traffic occurrences characterized as level 5 (traffic jam). This means that the majority of traffic occurrences lead to a stop on the traffic flow for the road where the incident occurred.

Another finding showed that in freeways when a traffic jam (level 5 traffic level) occurs, the delay in average is inferior to the delay when the level of traffic is high (level 4 traffic level).

Furthermore, findings show that the average driver is likely to spend between 3 to 5 minutes more on each road – primary street or freeway - when a traffic event occurs. These traffic occurrences tend to peak between 8 am, and 9 am and between 5 pm and 6 pm.

Observing the data patterns, it is possible to assess that Lisbon boroughs, at the city limits, have the highest concentration of traffic congestion. Most of these boroughs are neighboring municipalities, part of Lisbon's entrances and exits networks.

Our analysis also showed that the majority of traffic occurrences happen in the last three months of the year and the month with fewer traffic occurrences is the month of April.

Our dataframe had data for the years 2019 and 2020 according to this data, the delay and the length of the traffic suffered a reduction in the year 2020 this reduction is especially evident in March up until the end of our dataset in July 2020.

According to the data analysis, freeway traffic tends to have a higher length and delay, but these traffic occurrences tend to be less severe than primary street traffic also, freeway traffic tends to have a jam speed higher than in primary streets. However, traffic occurrences in freeways tend to be less severe than in primary streets some freeways have slower and lengthier traffic jams than others, such as IC 17, A5, and Eixo Norte-Sul.

The primary streets are the type of road that displayed the majority of traffic occurrences the traffic occurrences in this type of street tend to have smaller lengths and delays but are more frequent when compared to freeways. Some of these primary streets that outstand with more traffic congestion are R. Direita, R. Alta da Palma Carlos and Calçada da Ajuda.

When analyzing the maps, we can see that when a traffic event occurs, neighboring roads also tend to have traffic events this is especially verifiable in the central part of Lisbon and on the roads close to the Tagus river.

Traffic congestion is susceptible to pendular variations in its intensity, with critical peaks in the city's main entrances, especially in the morning and evening peak hours. We addressed this by developing data analytics and visualization, providing effective communication of research insights and visualization tools to policymakers.

Following previous studies [28], [75] insights, we developed data analysis and visualization based on different and combined factors, such as temporal, spatial, Spatio-temporal, and multi-variable.

The significant finding addressing our research question 2 on how to predict traffic delay in Lisbon with continuous variables is that our model suffered from the fact that our data was from a period with significant abnormal traffic variation because of the COVID-19 pandemic, and for that reason, the seasonality was affected. However, even with this constraint, we used other variables (weekend removal, holidays, traffic length), and our model was able to predict the traffic delay patterns, especially when comparing the data from the original dataset with the data predicted and the predictions for 100 days into the future.

4.2. Main Conclusions

Lisbon traffic congestion is an ongoing issue in Lisbon's urban mobility, carbon emissions 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 is one of the key challenges to improve citizens' quality of life. The developed multi-variable analysis and visualization provided essential insights in urban mobility traffic finding patterns on traffic congestion.

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

Traffic congestion analysis and visualization provide knowledge and insights on the more problematic roads and where the traffic bottlenecks occur, providing data-driven guidelines and knowledge about traffic in Lisbon to the city authorities and policymakers in the framework of a traffic management and visualization tool to help them mitigate such phenomena.

This traffic management and visualization tool, allied with Prophet or other new machine learning prediction techniques, can give an overall picture of what is expected in traffic behavior for the present and can give some knowledge of what can happen in the future and how to tackle the traffic problem with new and updated policies.

4.3. Research Limitations

Our study's limitations are related to incomplete data features and lack of information regarding external factors data – weather, air quality, events, sports, music, and the Covid-19 pandemic. Waze traffic data needs to be understood in the context of Lisbon traffic, particularly, what percentage does it represent of the overall Lisbon traffic reality. Additionally, more data is required, namely the number of cars in the city, cars that commute to the city, car speed, accidents occurrence, and external factors data, such as public events - soccer games, music festivals – as well as Covid-19 pandemic data. These additional data sources are needed to provide a broader analysis and visualization of traffic patterns in the city and make a more accurate prediction using Prophet.

4.4. Future Work

For future work, the following proposals are presented:

To understand better the results' implications, especially regarding external factors, such as weather, air quality, events (concerts, football games, and other events), crowd flow, and bike data. Moreover, cars, pedestrian, and bike accident data, as well as walkability, cycleway, bike station data, public transportation data, and more Spatio-temporal data are of interest to correlate with this study findings and henceforth understand the overall Lisbon urban mobility scenario.

To this aim, an integrated urban mobility dashboard with data analysis, visualization, and prediction methods can provide city management authorities and policymakers with a general picture of the city’s urban mobility, enabling smart solutions implementation towards a more resilient city.

This urban mobility tool would allow city management authorities and decision-makers to explore and better understand Lisbon city and metropolitan area commuters’ profile by means of an interactive dashboard depicting georeferenced data with different features such as boroughs, geographical, demographic, economic, social, planning and environment, 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) [76], an existing data platform of the City Hall, further developing PGIL capacity to process and provide helpful information for the operational and strategic management of the city to the various stakeholders.

References

[1] Z. Allam and P. Newman, 'Redefining the Smart City: Culture, Metabolism and Governance', *Smart Cities,* vol. 1, no. 1, Art. no. 1, Dec. 2018, doi: 10.3390/smartcities1010002.

[2] V. Albino, U. Berardi, and R. M. Dangelico, 'Smart Cities: Definitions, Dimensions, Performance, and Initiatives', *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015, doi: 10.1080/10630732.2014.942092.

[3] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, 'Road Traffic Forecasting: Recent Advances and New Challenges', *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2. Institute of Electrical and Electronics Engineers, pp. 93–109, Jun. 2018. doi: 10.1109/MITS.2018.2806634.

[4] A. M. Nagy and V. Simon, 'Survey on traffic prediction in smart cities', *Pervasive and Mobile Computing*, vol. 50. Elsevier B.V., pp. 148–163, Oct. 2018. doi: 10.1016/j.pmcj.2018.07.004.

[5] J. Smith, 'Intelligent transport systems', *Mobility and Transport - European Commission*, Sep. 22, 2016. https://ec.europa.eu/transport/themes/its_en (accessed Jul. 17, 2021).

[6] N. Petrovska and A. Stevanovic, 'Traffic Congestion Analysis Visualisation Tool', *2015 IEEE 18th Int. Conf. Intell. Transp. Syst.*, 2015, doi: 10.1109/ITSC.2015.243.

[7] G. Dimitrakopoulos and P. Demestichas, 'Intelligent transportation systems: Systems based on cognitive networking principles and management functionality', *IEEE Veh. Technol. Mag.*, vol. 5, no. 1, pp. 77–84, Mar. 2010, doi: 10.1109/MVT.2009.935537.

[8] D. Nunes, '42 minutos de fila por dia: Lisboa é a cidade ibérica com mais trânsito', *Diário de Notícias*. Jun. 2019. Accessed: Feb. 09, 2021. [Online]. Available: https://www.dn.pt/dinheiro/42-minutos-de-fila-por-dia-lisboa-e-a-cidade-iberica-com-mais-transito-10976480.html

[9] 'Covid-19: trânsito em Lisboa cai 30%; Porto é a cidade mais "engarrafada" do País'. https://www.jornaldenegocios.pt/economia/coronavirus/detalhe/covid-19-transito-em-lisboa-cai-30-porto-e-a-cidade-mais-engarrafada-do-pais (accessed Jul. 17, 2021).

[10] J. Vieira, 'Lisboa teve menos tráfego automóvel em 2020', *Jornal das Oficinas*, Jan. 15, 2021. https://jornaldasoficinas.com/pt/2021/01/15/lisboa-teve-menos-trafego-automovel-em-2020/ (accessed Jul. 17, 2021).

[11] 'Poluição atmosférica custa a cada residente em Lisboa 1.159 euros por ano'. https://www.jornaldenegocios.pt/economia/ambiente/detalhe/poluicao-atmosferica-custa-a-cada-residente-em-lisboa-1159-euros-por-ano (accessed Jul. 18, 2021).

[12] 'Relatório Anual de Segurança Rodoviária', p. 154, 2019.

[13] R. Wirth and J. Hipp, 'CRISP-DM: Towards a Standard Process Model for Data Mining'.

[14] 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', *undefined*, 2000, Accessed: Sep. 01, 2021. [Online]. Available: https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfbd2d5a72

[15] 'LxDataLab – Laboratório de Dados Urbanos de Lisboa'. https://lisboainteligente.cm-lisboa.pt/lxdatalab/ (accessed Sep. 01, 2021).

[16] 'Welcome to Python.org'. Accessed: Mar. 24, 2021. [Online]. Available: https://www.python.org/

[17] 'Data Visualization | Microsoft Power BI'. Accessed: Mar. 20, 2021. [Online]. Available: https://powerbi.microsoft.com/en-us/

[18] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, 'The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration.', *BMJ*, vol. 339, Jul. 2009, doi: 10.1136/bmj.b2700.
[19] Y. Xiao and M. Watson, 'Guidance on Conducting a Systematic Literature Review', *J. Plan. Educ. Res.*, vol. 39, no. 1, pp. 93–112, Mar. 2019, doi: 10.1177/0739456X17723971.
[20] 'Google Scholar'. https://scholar.google.com/ (accessed Jul. 19, 2021).
[21] 'Zotero | Your personal research assistant'. https://www.zotero.org/ (accessed Aug. 13, 2021).
[22] 'VOSviewer - Visualizing scientific landscapes', *VOSviewer*. https://www.vosviewer.com// (accessed Aug. 23, 2021).
[23] N. J. van Eck and L. Waltman, 'Software survey: VOSviewer, a computer program for bibliometric mapping', p. 16.
[24] 'Scimago Journal & Country Rank'. https://www.scimagojr.com/ (accessed Aug. 15, 2021).
[25] I. Kalamaras, A. Zamichos, A. Salamanis, A. Drosou, D. D. Kehagias, G. Margaritis, S. Papadopoulos, and D. Tzovaras, 'An Interactive Visual Analytics Platform for Smart Intelligent Transportation Systems Management', *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 487–496, Feb. 2018, doi: 10.1109/TITS.2017.2727143.
[26] C. Lee, Y. Kim, S. Jin, D. Kim, R. Maciejewski, D. Ebert, and S. Ko, 'A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion', *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 11, pp. 3133–3146, Nov. 2020, doi: 10.1109/TVCG.2019.2922597.
[27] I. Alam, M. F. Ahmed, M. Alam, J. Ulisses, D. M. Farid, S. Shatabda, and R. J. F. Rossetti, 'Pattern mining from historical traffic big data', in *2017 IEEE Region 10 Symposium (TENSYMP)*, Jul. 2017, pp. 1–5. doi: 10.1109/TENCONSpring.2017.8070031.
[28] W. Chen, F. Guo, and F. Y. Wang, 'A Survey of Traffic Data Visualization', *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2970–2984, Dec. 2015, doi: 10.1109/TITS.2015.2436897.
[29] J. Zhang, F. Y. Wang, K. Wang, W. H. Lin, X. Xu, and C. Chen, 'Data-driven intelligent transportation systems: A survey', *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011, doi: 10.1109/TITS.2011.2158001.
[30] S. Zhang, Z. Kang, Z. Hong, Z. Zhang, C. Wang, and J. Li, 'Traffic Flow Prediction Based on Cascaded Artificial Neural Network', in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2018, pp. 7232–7235. doi: 10.1109/IGARSS.2018.8518853.
[31] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, 'A hybrid deep learning based traffic flow prediction method and its understanding', *Transp. Res. Part C Emerg. Technol.*, vol. 90, pp. 166–180, 2018, doi: 10.1016/j.trc.2018.03.001.
[32] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, 'Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting', *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4883–4894, Nov. 2020, doi: 10.1109/TITS.2019.2950416.
[33] W. Xiangxue, X. Lunhui, and C. Kaixun, 'Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN', *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3043–3060, 2019, doi: 10.1007/s13369-018-3390-0.

[34] S. Kaffash, A. T. Nguyen, and J. Zhu, ‘Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis’, *Int. J. Prod. Econ.*, vol. 231, 2021, doi: 10.1016/j.ijpe.2020.107868.
[35] G. Ochoa-Ruiz, A. A. Angulo-Murillo, A. Ochoa-Zezzatti, L. M. Aguilar-Lobo, J. A. Vega-Fernández, and S. Natraj, ‘An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment’, *Appl. Sci.*, vol. 10, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/app10113974.
[36] N. Andrienko, G. Andrienko, and S. Rinzivillo, ‘Experiences from Supporting Predictive Analytics of Vehicle Traffic’.
[37] Q. Wang, M. Lu, and Q. Li, ‘Interactive, multiscale urban-traffic pattern exploration leveraging massive gps trajectories’, *Sens. Switz.*, vol. 20, no. 4, 2020, doi: 10.3390/s20041084.
[38] S. A. Angayarkanni, R. Sivakumar, and Y. V. Ramana Rao, ‘A review on traffic congestion detection methodologies and tools’, *Int. J. Adv. Sci. Technol.*, vol. 28, no. 16, pp. 1400–1414, 2019.
[39] M. Chen, X. Yu, and Y. Liu, ‘PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction’, *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018, doi: 10.1109/TITS.2018.2835523.
[40] K. N. Bui, H. Oh, and H. Yi, ‘Traffic Density Classification Using Sound Datasets: An Empirical Study on Traffic Flow at Asymmetric Roads’, *IEEE Access*, vol. 8, pp. 125671–125679, 2020, doi: 10.1109/ACCESS.2020.3007917.
[41] L. Zhu, R. Krishnan, F. Guo, J. W. Polak, and A. Sivakumar, ‘Early Identification of Recurrent Congestion in Heterogeneous Urban Traffic’, in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, pp. 4392–4397. doi: 10.1109/ITSC.2019.8916966.
[42] B. Liu, X. Tang, J. Cheng, and P. Shi, ‘Traffic flow combination forecasting method based on improved LSTM and ARIMA’, *Int. J. Embed. Syst.*, vol. 12, no. 1, pp. 22–30, 2020, doi: 10.1504/IJES.2020.105287.
[43] J. Zheng and M. Huang, ‘Traffic flow forecast through time series analysis based on deep learning’, *IEEE Access*, vol. 8, pp. 82562–82570, 2020, doi: 10.1109/ACCESS.2020.2990738.
[44] D. -H. Shin, K. Chung, and R. C. Park, ‘Prediction of Traffic Congestion Based on LSTM Through Correction of Missing Temporal and Spatial Data’, *IEEE Access*, vol. 8, pp. 150784–150796, 2020, doi: 10.1109/ACCESS.2020.3016469.
[45] Z. Wang and P. Thulasiraman, ‘Foreseeing Congestion using LSTM on Urban Traffic Flow Clusters’, in *2019 6th International Conference on Systems and Informatics (ICSAI)*, Nov. 2019, pp. 768–774. doi: 10.1109/ICSAI48974.2019.9010150.
[46] Q. Chu, G. Li, R. Zhou, and Z. Ping, ‘Traffic Flow Prediction Model Based on LSTM with Finnish Dataset’, in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Apr. 2021, pp. 389–392. doi: 10.1109/ICSP51882.2021.9408888.
[47] Y. Hu, ‘Research on City Traffic Flow Forecast Based on Graph Convolutional Neural Network’, in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Mar. 2021, pp. 269–273. doi: 10.1109/ICBAIE52039.2021.9389951.
[48] X. Di, Y. Xiao, C. Zhu, Y. Deng, Q. Zhao, and W. Rao, ‘Traffic Congestion Prediction by Spatiotemporal Propagation Patterns’, in *2019 20th IEEE International Conference on Mobile Data Management (MDM)*, Jun. 2019, pp. 298–303. doi: 10.1109/MDM.2019.00-45.

[49] A. Sadeghi-Niaraki, P. Mirshafiei, M. Shakeri, and S.-M. Choi, ‘Short-Term Traffic Flow Prediction Using the Modified Elman Recurrent Neural Network Optimized through a Genetic Algorithm’, *IEEE Access*, vol. 8, pp. 217526–217540, 2020, doi: 10.1109/ACCESS.2020.3039410.

[50] M. A. Lusiandro, S. M. Nasution, and C. Setianingsih, ‘Implementation of the Advanced Traffic Management System using k-Nearest Neighbor Algorithm’, in *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, Oct. 2020, pp. 149–154. doi: 10.1109/ICITSI50517.2020.9264952.

[51] B. Priambodo, A. Ahmad, and R. A. Kadir, ‘Spatio-temporal K-NN prediction of traffic state based on statistical features in neighbouring roads’, *J. Intell. Fuzzy Syst.*, vol. 40, no. 5, pp. 9059–9072, 2021, doi: 10.3233/JIFS-201493.

[52] B. Priambodo, A. Ahmad, and R. A. Kadir, ‘Predicting Traffic Flow Propagation Based on Congestion at Neighbouring Roads Using Hidden Markov Model’, *IEEE Access*, vol. 9, pp. 85933–85946, 2021, doi: 10.1109/ACCESS.2021.3075911.

[53] T. Sun, Z. Huang, H. Zhu, Y. Huang, and P. Zheng, ‘Congestion Pattern Prediction for a Busy Traffic Zone Based on the Hidden Markov Model’, *IEEE Access*, vol. 9, pp. 2390–2400, 2021, doi: 10.1109/ACCESS.2020.3047394.

[54] G. Boquet, A. Morell, J. Serrano, and J. L. Vicario, ‘A variational autoencoder solution for road traffic forecasting systems: Missing data imputation, dimension reduction, model selection and anomaly detection’, *Transp. Res. Part C Emerg. Technol.*, vol. 115, 2020, doi: 10.1016/j.trc.2020.102622.

[55] Y. Liu and H. Wu, ‘Prediction of Road Traffic Congestion Based on Random Forest’, in *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Dec. 2017, vol. 2, pp. 361–364. doi: 10.1109/ISCID.2017.216.

[56] Y. Tao, P. Sun, and A. Boukerche, ‘A Hybrid Stacked Traffic Volume Prediction Approach for a Sparse Road Network’, in *2019 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2019, pp. 1–6. doi: 10.1109/ISCC47284.2019.8969710.

[57] S. Kundu, M. S. Desarkar, and P. K. Srijith, ‘Traffic Forecasting with Deep Learning’, in *2020 IEEE Region 10 Symposium (TENSYMP)*, Jun. 2020, pp. 1074–1077. doi: 10.1109/TENSYMP50017.2020.9230762.

[58] A. Sinha, R. Puri, U. Balyan, R. Gupta, and A. Verma, ‘Sustainable Time Series Model for Vehicular Traffic Trends Prediction in Metropolitan Network’, in *2020 6th International Conference on Signal Processing and Communication (ICSC)*, Mar. 2020, pp. 74–79. doi: 10.1109/ICSC48311.2020.9182755.

[59] R. Zhou, H. Chen, H. Chen, E. Liu, and S. Jiang, ‘Research on Traffic Situation Analysis for Urban Road Network Through Spatiotemporal Data Mining: A Case Study of Xi’an, China’, *IEEE Access*, vol. 9, pp. 75553–75567, 2021, doi: 10.1109/ACCESS.2021.3082188.

[60] M. M. Chowdhury, M. Hasan, S. Safait, D. Chaki, and J. Uddin, ‘A Traffic Congestion Forecasting Model using CMTF and Machine Learning’, in *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, Jun. 2018, pp. 357–362. doi: 10.1109/ICIEV.2018.8640985.

[61] Z. He, G. Qi, L. Lu, and Y. Chen, ‘Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data’, *Transp. Res. Part C Emerg. Technol.*, vol. 108, pp. 320–339, 2019, doi: 10.1016/j.trc.2019.10.001.

[62] X. Song, W. Li, D. Ma, D. Wang, L. Qu, and Y. Wang, ‘A Match-Then-Predict Method for Daily Traffic Flow Forecasting Based on Group Method of Data Handling’, *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 33, no. 11, pp. 982–998, 2018, doi: 10.1111/mice.12381.

[63] Z. Bartlett, L. Han, T. T. Nguyen, and P. Johnson, ‘A Novel Online Dynamic Temporal Context Neural Network Framework for the Prediction of Road Traffic Flow’, *IEEE Access*, vol. 7, pp. 153533–153541, 2019, doi: 10.1109/ACCESS.2019.2943028.

[64] E. Husni, S. M. Nasution, Kuspriyanto, and R. Yusuf, ‘Predicting Traffic Conditions Using Knowledge-Growing Bayes Classifier’, *IEEE Access*, vol. 8, pp. 191510–191518, 2020, doi: 10.1109/ACCESS.2020.3032230.

[65] S. Guo, Y. Lin, S. Li, Z. Chen, and H. Wan, ‘Deep Spatial–Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting’, *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3913–3926, Oct. 2019, doi: 10.1109/TITS.2019.2906365.

[66] P. Sun, N. AlJeri, and A. Boukerche, ‘A Fast Vehicular Traffic Flow Prediction Scheme Based on Fourier and Wavelet Analysis’, in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–6. doi: 10.1109/GLOCOM.2018.8647731.

[67] A. Izhar, S. M. K. Quadri, and S. A. M. Rizvi, ‘Hybrid Feature Based Label Generation Approach for Prediction of Traffic Congestion in Smart Cities’, in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Dec. 2020, pp. 991–997. doi: 10.1109/ICISS49785.2020.9316085.

[68] Z. Wang, L. Zhang, and Z. Ding, ‘Multi-Step Prediction of Traffic Flow Based on Wavelet Decomposition Correlation Matrix’, in *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, May 2020, pp. 444–448. doi: 10.1109/ICECTT50890.2020.00102.

[69] L. Zhang, K. Gong, M. Xu, A. Li, Y. Dong, and Y. Wang, ‘How to Identify Patterns of Citywide Dynamic Traffic at a Low Cost? An In-Depth Neural Network Approach with Digital Maps’, *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6648116.

[70] X. Ma, Y. Li, and P. Chen, ‘Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative matrix factorization algorithm’, *J. Traffic Transp. Eng. Engl. Ed.*, vol. 7, no. 4, pp. 529–539, 2020, doi: 10.1016/j.jtte.2018.12.002.

[71] A. Rao, A. Phadnis, A. Patil, T. Rajput, and P. Futane, ‘Dynamic Traffic System Based on Real Time Detection of Traffic Congestion’, in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Aug. 2018, pp. 1–5. doi: 10.1109/ICCUBEA.2018.8697838.

[72] Vighnesh and S. D, ‘Traffic Prediction Using a Supervised Learning Approach’, in *2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEECCOT)*, Dec. 2018, pp. 1219–1223. doi: 10.1109/ICEEECCOT43722.2018.9001658.

[73] ‘Criação indicador de tráfego geral e indicadores para cada uma das principais vias de entrada na cidade’, *LxDataLab*. 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 Sep. 01, 2021).

[74] ‘Prophet | Forecasting at scale.’ https://facebook.github.io/prophet/ (accessed Sep. 14, 2021).

[75] H. R. Surya, N. Raju, and S. S. Arkatkar, ‘Stability Analysis of Mixed Traffic Flow using Car-Following Models on Trajectory Data’, in *2021 International Conference on COMmunication Systems NETworkS (COMSNETS)*, Jan. 2021, pp. 656–661. doi: 10.1109/COMSNETS51098.2021.9352915.

[76] ‘Plataforma de Gestão Inteligente de Lisboa’, *Lisboa Inteligente*. https://lisboainteligente.cm-lisboa.pt/lxi-iniciativas/plataforma-de-gestao-inteligente-de-lisboa/ (accessed Sep. 02, 2021).

x