INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sustainability Measurement at Logistics Transportation Company

Fábio Cordeiro Pereira

Master in **Integrated Decision Support Systems**

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

Co-supervisor:
MSc Lídia Vitória Pires de Albuquerque, Researcher
NOVA IMS

October, 2022

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Sustainability Measurement at Logistics Transportation Company

Fábio Cordeiro Pereira

Master in **Integrated Decision Support Systems**

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

Co-supervisor:
MSc Lídia Vitória Pires de Albuquerque, Researcher
NOVA IMS

October, 2022

“Without data you are just another person with an opinion”

W. Edwards Deming

Acknowledgments

The conclusion of this thesis dictates the end of my Master's degree in Integrated Decision Support Systems that I entered in 2020. As a working student, this phase of my life represents two years of extreme dedication in professional and personal terms, due to what I consider to be a very high accumulation of knowledge in data science.

It is certain that the success of this journey was due to the people who shared this path with me. During these two years, I owe a special thanks to my supervisors, José Miguel Dias and Vitória Albuquerque, who since day one have been tireless in sharing knowledge and guidance. I must also thank them for the great availability shown throughout this period and the patience to discuss the endless doubts and questions. To both of them, a big thank you.

To my parents…

I leave a special thank you for everything, literally everything. A thank you for the values that you passed on to me, the ideals and the support. But more than that, I thank you for the example. With you I learned that success comes from hard work and that the result is proportional to our dedication. With you, I learned that the details, the small details are essential and that the way we face the small things is reflected in the way we face the big things, the rest is exponential. Thank you for everything and even more, much more.

To my brothers...

To my first and always true best friends I thank you for all your unconditional support. You made me realize that no matter how far away I may be, I will never be alone, no matter where I am. I want to leave a special thank you to each one of you, for the advice, the company, the friendship, and even the typical and natural fights between brothers that, even after 27 years, are still funny.

To my friends...

To my friends, those who are really like brothers to me, I leave a thank you. A thank you for telling me what I need to hear when I need to hear it, for giving me a head's up when I need to hear it. A thank you for showing me the way, even when the way is the wrong one, especially there.

To my co-workers...

I must also thank Deloitte and my co-workers for their availability and flexibility throughout this journey. I also thank you for your attention and sharing of experiences. With all of you I can learn new things every day.

To You…

Finally, to You, who never doubted me, a special thank you. You made me realize how important is the meaning of the sentence “that itch you can’t scratch”. You made me realize that all stones are to be kicked, even those that are not in front of me, especially those. You made me want to discover, seek, live, more and more, and above all, seek to scratch that so called little itch. To you, thank you, for everything.

To everyone who has crossed this path, my unconditional thanks.

"What we do in life echoes in eternity".

Fábio Pereira

Resumo

A preocupação generalizada pela redução de emissões de CO_2 tem vindo a crescer de forma exponencial. As empresas têm vindo a desenvolver mecanismos e métricas que permitam a redução generalizada da sua pegada carbónica, através de renovação de fontes de energia, e desenvolvimento de atividades energeticamente mais eficientes.

O nosso estudo tem dois objetivos principais. O primeiro passa pelo desenvolvimento de um modelo para cálculo da pegada de carbono de uma empresa de transportes a operar na península ibérica, tendo por base as metodologias de cálculo previstas no programa Lean&Green, as quais se regem pelos protocolos ISO14064 e EN16258. O segundo objetivo consiste na identificação de caraterísticas, padrões e fatores que influenciem a emissão de carbono e consequentemente contribuam para a o cálculo da pegada de carbono.

Desta forma, para este trabalho de investigação identificámos indicadores associados à atividade desenvolvida pela empresa de logística. Implementámos uma abordagem de medição das emissões de CO_2 para verificar o objetivo de redução assumido de pelo menos 20% durante um período de 5 anos face às emissões de CO_2 emitidas no ano de 2017.

Nos processos de trabalho foi seguida a metodologia CRIPS-DM, tendo sido os dados recolhidos através de uma API disponibilizada pela empresa para o primeiro semestre de 2022.

Decorrente das técnicas de visualização adotadas e consequente análises desenvolvidas, identificámos os principais fatores que contribuem para o cálculo da pegada de carbono. Desenvolvemos um modelo de cálculo da pegada de carbono e um conjunto de métricas de gestão por forma a monitorizar as emissões de CO_2.

Palavras-chave: *Pegada de carbono; Transporte logístico; Sustentabilidade, Normas*

Abstract

The concern for reducing CO2 emissions has been growing exponentially. Companies have been developing mechanisms and metrics that allow the generalized reduction of their carbon footprint, through the renewal of energy sources, and development of more energy efficient activities.

Our study has two main objectives. The first is to develop a model to calculate the carbon footprint of a transport company operating in the Iberian Peninsula, based on the calculation methodologies provided by the Lean&Green program, which are governed by the ISO14064 and EN16258 protocols. The second objective is to identify characteristics, patterns and factors that influence carbon emissions and therefore contribute to the calculation of the carbon footprint.

Thus, for this research work we identified indicators associated with the activity developed by the logistics company. We implemented an approach to measure CO2 emissions to verify the assumed reduction goal of at least 20% over a period of 5 years compared to the CO2 emissions emitted in 2017.

In the work processes, the CRIPS-DM methodology was followed, with data collected through an API made available by the company for the first half of 2022.

As a result of the visualization techniques adopted and consequent analyses developed, we identified the main factors that contribute to the calculation of the carbon footprint. We developed a carbon footprint calculation model and a set of management metrics in order to monitor CO2 emissions.

Keywords: *Carbon footprint; Logistic transportation; Sustainability, Standards*

Index

Tables index

Figures index

List of abbreviations

CF – Carbon Footprint

CLECAT – European Association for Forwarding, Transport, Logistics and Customs Services

CO – A company of the transport sector, operating in Iberian market

CO_2 - Carbon dioxide

CRISP-DM – Cross Industry Standard Process for Data Mining

DSS – Decision Support System

GA – Genetic Algorithm

GGE – Greenhouse Gas Emissions

GHG – Greenhouse Gas

GHG Protocol – Greenhouse Gas Protocol Product Standard

LG – Lean & Green

Kg - Kilogram

Km - Kilometer

ML – Machine Learning

NN – Neural Networks

NO – Nitric Oxide

PAS – Publicly Available Specification

RQ – Research Question

SCR – Selective Catalytic Reduction

SLR – Systematic Literature Review

1. Introduction

1.1. Topic context

With the increase in greenhouse gas emissions and consequently the emergence of adverse climate changes, international regulators have delivered laws , namely the Paris Agreement in 2015 [1], to enforce the reduction of carbon footprint emissions by companies and organizations through the implementation of fiscal or other measures, to promote the development of more environmentally friendly practices.

This study aims to model a carbon footprint calculation method to be implemented by in a logistic company (CO) operating in Iberian market. CO has interposed a strategic objective under the Lean & Green (LG) initiative, to reduce at least 20% carbon dioxide (CO_2) emissions for a period of 5 years, starting in 2017 in the following business areas: Logistics, Transport and Distribution.

The LG initiative is the largest European collaboration platform that aims to reduce CO_2 emissions and to achieve decarbonization associated to the supply chain, forming a network of leading organizations in sustainable logistics.

LG is Europe's largest collaborative platform encourages companies to achieve a higher level of sustainability in logistics. It is currently implemented in 13 countries with more than 600 companies joining, which has already allowed the reduction of 2.5 megatonnes of CO_2. In Portugal Global System of Standards (GS1) [2] is the Portuguese entity responsible for this initiative.
The process begins with the definition of an action plan, for a maximum of 5 years, where the company is subjected to an initial audit that assesses the carbon footprint of logistics activities as well as the relevance of the plan. Over the 5 years, the company is monitored by the responsible entity - in the Portuguese case, GS1 Portugal - with meetings every 6 months to evaluate the implementation of the defined measures.

CO joined GS1 Portugal, in a trusted partnership on issues related to sustainability, with a 5-year action plan.

CO was established in 1982 and has set the mission of analyzing and offering the best logistics solution to its customers. Currently the CO operates in the areas of Distribution, Logistics and Transport with an efficient operating model, specifically developed for the Iberian market.

CO currently has 21 Logistics Platforms in mainland Portugal and generates annually around 65 million euros of revenue (data from 2021). It has about 1,200 employees and a fleet of over 500 vehicles. CO makes more than 8,500 daily deliveries, corresponding to about 750,000 tons per year, and has a logistical area of more than 185,500 sqm.

1.2. Research questions and objectives

Our research aims to calculate the carbon footprint associated to logistic transport and to perform data analysis, identifying patterns in CO data, as a case study.

CO operates in the transport and logistics sector and that set up as a strategic goal and objective, under the LG initiative, the reduction of its carbon footprint of at least 20%, for a period of 5 years. This is based on the search for continuous and sustainable improvement through the implementation of more efficient practices, as well as more sustainable from an environmental point of view. Furthermore, and despite the higher structural costs associated to an environmentally friendly strategy y, CO assumed this goal as a social commitment to future generations.

3.5In this way, the research aims to answer the following research questions (RQ):

RQ 1: How can we calculate the carbon footprint of a freight company?

RQ 2: What are the main factors of logistics transportation that significantly influence the carbon footprint of a freight company?

To address RQ 1, we developed metrics to analyze and monitor the evolution of CO_2 emissions, based on CO's data. This supported the decision-making process of CO on aspects related to the carbon footprint produced on its activity. For the development of this system, contributions to the calculation of the carbon footprint were considered, such as methodologies proposed in Greenhouse Gas Protocol ISO 14064 and EN 16258.

Also, the LG program defined conversion factors from fuel used to CO_2, which allowed the calculation of the carbon footprint from fuel consumption, as analyzed in more detail in section 3.5.

The data analysis regarding the main factors that influence the carbon footprint (RQ 2) aim to identify the patterns of CO transport activity.

Additionally, using data mining we extracted knowledge from the data, namely patterns, correlations, and other interesting characteristics in order to understand the company's activity and the whole decision support process.

1.3. Methodology

In our study we applied the Cross Industry Standard Process for Data Mining (CRISP-DM) [3] methodology, an iterative process that aims to describe the data analysis process to improve the planning, organization and implementation of data analysis projects. The methodology applied is explained in more detail within Section 3 (section 3.1.)

1.4. Dissertation structure

This thesis dissertation is structured into four Sections. In Section 0 we introduced the topic, the context, the research questions, objectives, as well as the applied methodology, and the structure. This is followed by Section 2 literature review, where we applied the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [4]. Then the VOSviewer tool [5] was used to visualize scientific landscapes in our literature dataset and to find the latest state-of-the-art methodologies applied on the logistic carbon footprint. In Section 0, we applied the CRISP-DM methodology on our case study dataset, available through the CO API using the Python programming language [6]. Finally, in Section 0, we presented the discussion of the results, our conclusions, and limitations, as well as proposals for future work.

2. Literature review

2.1. Methods

2.1.1. Literature review methodology

The Systematic Literature Review (SLR) was developed in order to answer the RQs defined in section 1.2 by using the PRISMA methodology [4]. Therefore, the PRISMA Statement was applied, considering the checklist of twenty-seven items and a four-phase flowchart. The checklist includes items considered essential for the transparent communication of a systematic review, considering relevant empirical studies and methodological literature.

Our search was performed during the month of April 2022 and was limited to a five-year period, between 2018 and 2022, restricting our literature review to the most recent studies carried out by the scientific community.

2.1.2. Keywords and research query

Our literature search started with the definition of the RQs as defined in section 1.2, on which and used to perform the queries of our research. As a result of our analysis, we identified the following Keywords:

- “Carbon Footprint” OR
- “Ecological Footprint” AND
- “Freight Transport” AND
- “Data Analytics” OR “Machine Learning” OR “Prediction”

("Carbon Footprint" OR "Ecological Footprint") AND "Freight Transport" AND ("Data Analytics" OR "Machine Learning" OR "Prediction") AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (LANGUAGE , "English"))

2.2. Literature review

2.2.1. PRISMA results

PRISMA flow diagram follows four steps: Identification, Screening, Eligibility and Included.

Our SLR PRISMA flow diagram (Figure 1, where "n" is the number of retained articles, and “e” is the number of excluded articles), depicts the application of the query defined in Section 1.2 performed on the Scopus and ScienceDirect repositories, corresponding to the "Identification" phase. As result, we obtained 117 journal papers (62 papers from Scopus and 55 ScienceDirect), and 15 conference papers (12 conferences from Scopus and 3 ScienceDirect). Only journal and conference papers published in English were considered. Additionally, 3 articles were included, that were recommended by my Supervisors and Master program director.

Considering that we used more than one repository, 16 duplicate papers were identified and validated by using the Mendeley tool.

In the "Screening" phase, we had 119 articles, based on the abstract reading.

Subsequently, in the "Eligibility" phase, we proceeded to a complete reading of the papers, and identified the ones to include in our search. The records assessed for eligibility were 48 articles. In the last phase "Included", after full reading and content analysis, we included 24 articles in our literature review.

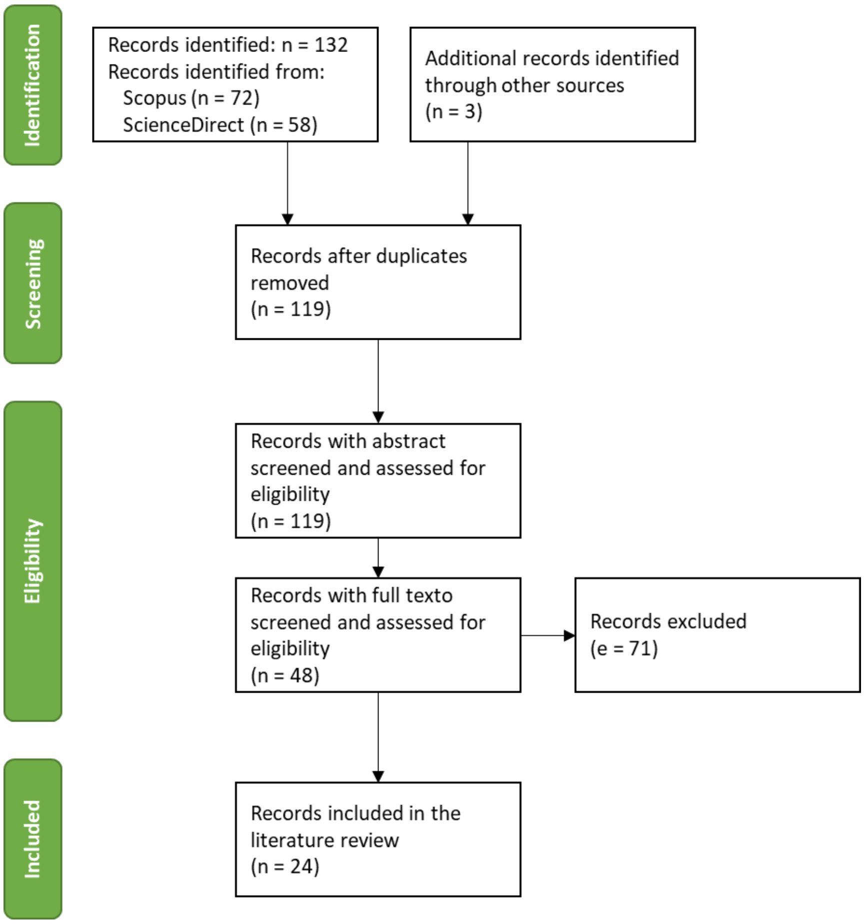

Figure 1 - Flowchart Diagram - PRISMA

2.2.2. Main journals identification

Our SLR resulted in a total of 24 included papers. These 24 papers cover a wide area of research, namely about Business, Management and Accounting, Energy, Engineering, Environmental Science, Computer Science, among others, as shown in more detail in Table 1.

The 24 papers were published in 18 journals. It should be noted that one of the articles read, which did not result from the query carried out within the scope of the literature review, is not published. The journal with the highest number of articles considered within our literature review is the Journal of Cleaner Production (with 3 journal articles), followed by Sustainability (Switzerland), Transportation Research Part D: Transport and Environment, and Science of the Total Environment, (all with 2 articles).

Additionally, according to Scimago [7], most of the journals are Q1. A total of 16 journal papers were included that are Q1, representing about 67% of the number of journal papers considered. 6 journal papers are Q2, representing about 25%, and the rest are Q4, representing only 8% of the number of journal papers considered.

Finally, at the publisher level, Elsevier Ltd is the publisher with the largest number of journal papers considered in our literature review, with 12 journal papers representing about 50% of the journal papers considered. Followed by Elsevier BV and MDPI with 3 journal papers each, individually representing about 13% of the number of journal papers considered.

Table 1 - Main journals

Journal	**Quartile Rank**	**Fields**	**Publisher**	**Publisher Country**	**No**
Journal of Cleaner Production	Q1	Business, Management and Accounting; Energy; Engineering; Environmental Science	Elsevier Ltd	United Kingdom	3
Sustainability (Switzerland)	Q2	Chemical Engineering; Energy; Engineering; Environmental Science; Social Sciences	MDPI	Switzerland	2
Science of the Total Environment	Q1	Environmental Science	Elsevier BV	Netherlands	2

Transportation Research Part D: Transport and Environment	Q1	Engineering; Environmental Science; Social Sciences	Elsevier Ltd	United Kingdom	2
Renewable and Sustainable Energy Reviews	Q1	Energy	Elsevier Ltd	United Kingdom	1
Energy Policy	Q1	Energy; Environmental Science	Elsevier Ltd	United Kingdom	1
International Journal of Environmental Research and Public Health	Q2	Environmental Science; Medicine	MDPI	Switzerland	1
Energy Conversion and Management	Q1	Energy	Elsevier Ltd	United Kingdom	1
Computers and Industrial Engineering; Engineering	Q1	Computer Science; Engineering	Elsevier Ltd	United Kingdom	1
Transportation Research Interdisciplinary Perspectives	Q2	Decision Sciences; Engineering; Social Sciences	Elsevier Ltd	United Kingdom	1
PLoS ONE	Q1	Multidisciplinary	Feng Chen	United States	1
Transportation Engineering	Q2	Engineering; Social Sciences	Elsevier Ltd	United Kingdom	1
International Journal of Sustainable Transportation	Q1	Energy; Engineering; Environmental Science; Social Sciences	Taylor & Francis	United Kingdom	1
Journal of Environmental Management	Q1	Environmental Science; Medicine	Elsevier Ltd	United Kingdom	1
Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering	Q4	Computer Science	Springer Nature Switzerland	Switzerland	1

Structural Change and Economic Dynamics	Q1	Economics, Econometrics and Finance	Elsevier BV	Netherlands	1
Environmental Science and Pollution Research	Q2	Environmental Science; Medicine	Springer Nature Germany	Germany	1
International Journal of Advanced Manufacturing Technology	Q1	Computer Science; Engineering	Springer Verlag London	United Kingdom	1

2.2.3. Publications ranked by the number of citations

In this section we identified the number of citations for each of the journal papers included in the scope of our literature review. This exercise allowed us to identify which papers had the greatest popularity and influence within research communication.

Altogether, the journal papers read and included in our SLR were cited 669 times, as presented in Table 2.

The following 5 papers stand out: Umar [8] with 162 citations, followed by García-Olivares [9] with 151 citations, Leach [10] with 87 citations, Cunha [11] with 85 citations and [12] with 51 citations.

Table 2 - Publications ranked by the number of citations

Title	Author	Year of Publication	Publisher	Citations
COP21 Roadmap: Do innovation, financial development, and transportation infrastructure matter for environmental sustainability in China?	Umar, Muhammad Ji, Xiangfeng Kirikkaleli, Dervis Xu, Qinghui	2020	Elsevier Ltd	162
Transportation in a 100% renewable energy system	García-Olivares, Antonio Solé, Jordi Osychenko, Oleg	2018	Elsevier Ltd	151
The scope for improving the efficiency and environmental impact of internal combustion engines	Leach, Felix Kalghatgi, Gautam Stone, Richard Miles, Paul	2020	Elsevier Ltd	87

A new approach to reduce the carbon footprint in resistance spot welding by energy efficiency evaluation	Cunha, Carla Ferreira Andrade de Oliveira Gomes, Jefferson de Carvalho, Hugo Marcelo Bezerra	2022	Springer Verlag London	85
Evaluation of sustainable transport research in 2000–2019	Zhao, Xianbo Ke, Yongjian Zuo, Jian Xiong, Wei Wu, Peng	2020	Elsevier Ltd	51
Well-to-wheels approach for the environmental impact assessment of road freight services	Osorio-Tejada, Jose Luis Llera-Sastresa, Eva Hashim, Ahmad Hariza	2018	MDPI	28
Towards sustainable wood procurement in forest industry – The energy efficiency of larger and heavier vehicles in Finland	Palander, Teijo Haavikko, Hanna Kärhä, Kalle	2018	Elsevier Ltd	19
The development of a carbon footprint model for the calculation of GHG emissions from highways: the case of Egnatia Odos in Greece	Roukounakis, Nikolaos Valkouma, Efthaleia Giama, Efrosini Gerasopoulos, Evaggelos	2020	Taylor & Francis	19
Optimising truckload operations in third-party logistics: A carbon footprint perspective in volatile supply chain	Wong, Eugene Y.C. Tai, Allen H. Zhou, Emma	2018	Elsevier Ltd	19
Comparison of product carbon footprint protocols: Case study on medium-density fiberboard in China	Wang, Shanshan Wang, Weifeng Yang, Hongqiang	2018	MDPI	10
Life cycle assessment of mid-range passenger cars powered by liquid and gaseous biofuels: Comparison with greenhouse gas emissions of electric vehicles and forecast to 2030	Ternel, Cyprien Bouter, Anne Melgar, Joris	2021	Elsevier Ltd	9

Biomass feedstock transport using fuel cell and battery electric trucks improves lifecycle metrics of biofuel sustainability and economy	Baral, Nawa Raj Asher, Zachary D. Trinko, David Sproul, Evan Quiroz-Arita, Carlos Quinn, Jason C. Bradley, Thomas H.	2021	Elsevier Ltd	8
Energy use and emissions scenarios for transport to gauge progress toward national commitments	Schmitz Gonçalves, Daniel Neves Goes, George Vasconcelos de Almeida D'Agosto, Márcio Albergaria de Mello Bandeira, Renata	2019	Elsevier Ltd	8
Trading off cost, emission, and quality in cold chain design: A simulation approach	Fan, Yun de Kleuver, Caroline de Leeuw, Sander Behdani, Behzad	2021	Elsevier Ltd	3
Data analytics for sustainable global supply chains	Mangina, Eleni Narasimhan, Pranav Kashyap Saffari, Mohammad Vlachos, Ilias	2020	Elsevier Ltd	3
Modelling carbon emissions of diesel trucks on longitudinal slope sections in China	Dong, Yaping Xu, Jinliang Gu, Chenwei	2020	Feng Chen, Tongii University, CHINA	2
An Intelligent Visualisation Tool to Analyse the Sustainability of Road Transportation	de Armiño, Carlos Alonso Urda, Daniel Alcalde, Roberto García, Santiago Herrero, Álvaro	2022	MDPI	1
Simulation study on carbon emission of China's freight system under the target of carbon peaking	Wen, Lei Song, Qianqian	2022	Elsevier BV	1

A new artificial neural networks algorithm to analyze the nexus among logistics performance, energy demand, and environmental degradation	Magazzino, Cosimo Mele, Marco Schneider, Nicolas	2022	Elsevier BV	1
Integrated effects of SCR, velocity, and air-fuel ration on gaseous pollutants and CO2 emissions from China V and VI heavy-duty diesel vehicles	Li, Xueyao Ai, Yi Ge, Yunshan Qi, Jingyu Feng, Qian Hu, Jie Porter, William C. Miao, Yaning Mao, Hongjun Jin, Taosheng	2022	Elsevier BV	1
Research on application of a hybrid heuristic algorithm in transportation carbon emission	Li, Yanmei Dong, Hong Kai Lu, Shuangshuang	2021	Springer Nature	1
Transition of mobility in companies – A semi-systematic literature review and bibliographic analysis on corporate mobility and its management	Gorges, Tobias Holz-Rau, Christian	2021	Elsevier Ltd	0
Optimising Supply Chain Logistics System Using Data Analytics Techniques	Mangina, Eleni Narasimhan, Pranav Kashyap Saffari, Mohammad Vlachos, Ilias	2020	Springer Nature Switzerland	0
Driver profile classification on heavy trucks based on telemetry	Alves, Daniel Janeiro	2019	-	-

2.2.4. Methods and applications of carbon footprint measurement in logistic transport

Methods and applications of carbon footprint measurement were explored the SLR papers in the scope of their various influence factors on CF. Li [13] carried out a study to test emissions from heavy-duty, diesel-powered vehicles in China, to evaluate the integrated effects of Selective Catalytic Reduction (SCR), speed, and air-fuel ratio on CO_2 and nitrogen oxide (NO) emissions. Li's results revealed that CO_2 emission levels based on average distance at high speeds (50-90 km/h) were lower than those based on low speeds.

Li [14] developed another study where he explored the influence of some factors on CO_2 emissions in the transportation industry and developed future CO_2 emission forecast scenarios. Li used four algorithms to develop his analysis: neural network (NN), extreme learning machine, genetic algorithm (GA) optimized neural network, and genetic algorithm optimized extreme learning machine.

In the research conducted by Fan [15], associated with the simulation of the financial impacts, quality and CO_2 emission in the transport of refrigerated goods by ship, a formula for the calculation of CO_2 was studied, and a set of sensitivity analyses were performed on it through the variation of several components of the equation. From this analysis, it was concluded that the speed of transport is intrinsically related to CO_2 emissions.

Ternel [16] developed a study on the current and future predictions of the Greenhouse Gas Emissions (GGE) analysis of mid-range passenger cars. The study showed that plug-in hybrid electric vehicles were the best solution. Ternel also concluded that biofuels were able to provide significant benefits regarding GGE emissions in a short period of time.

Osorio-Tejada [17] developed a Well-to-wheels approach for assessing the environmental impact of road freight services. In this study he concludes that speed, load, and road gradient generate variations of up to 145% in the estimated GGE emissions.

Mangina [18][19] evaluated models for freight transport using data analysis. Three algorithms (Horizontal Cooperation, Pooling and Physical Internet) were developed, using historical road freight transport data from the European market, for the period between 2011 and 2014. To calculate the total emissions, a simplified formula proposed by the European Association for Forwarding, Transport, Logistics and Customs Services (CLECAT) was used,

$$G_T = T_{cap} * g_t \tag{1}$$

$$T_{cap} = \frac{(W_g - \phi + \chi) * d}{1000} \tag{2}$$

where Gt is the GGE per tonne kilometres (g CO2/t- km), Tcap represents the transport capacity, gt represents the emissions factor depending on the vehicle weighting, Wg represents the gross weight, and d represents the distance.

Mangina concluded that using the pooling algorithm, there was a 12% reduction in road freight transport emissions and an increase in transport efficiency of about 23%.

Alves [20] proposed a method for classifying truck drivers based on time series to describe their driving efficiency. The study aimed to develop a methodology to reward drivers who presented more ecologically sustainable driving. Correlation analyses were developed to identify the variables with the highest correlation when compared to fuel consumption, as follows: Acceleration, Acceleration Pedal Use, Brake Pedal Use, Engine Speed, Vehicle Speed.

Palander [21] developed a quantitative analysis of energy performance from the optimization of timber transportation. Palander demonstrated that there was an upward trend in the increase of the average load weight and a transition to 7 to 9 axle vehicles, which allowed the transport of higher loads. As a result of this transition, Palander sees a significant increase in energy efficiency, but this does not, however, achieve the goals set by the government, demonstrating that the success of the goal involved other measures besides using vehicles with higher transport capacity.

Wong [22] analyzed the operation of three logistics companies in Hong Kong and developed a metric analysis of measuring GGE. Truckload utilization and transportation routes were analyzed in detail, as well as the correlation on truckload utilization against truck capacity, cargo volume, fuel consumption, truck size, travel distances and the number of destinations. Wong developed an integrated model, upon which CF achievement goals defined, targeting CF reduction initiatives by penalizing transport times and distance, minimizing the number of trucks, and fostering improved truck utilization.

Schmitz [23], developed a study that aims to model scenarios of energy use and GGE from the transport sector up to 2030. Three scenarios were developed based on different policy commitments. The method was based on a bottom-up approach, requiring multi-sector collaborative efforts to not only account for direct energy use, but also balance transport and energy activity across modes, justifying each case in terms of development stage and energy supply capacity.

In the study by Armiño [24], the sustainability of the transportation activity was analyzed, proposing the application of new Machine Learning techniques. In this research, Hybrid Unsupervised Exploratory Plots is used, along with new Exploratory Projection Pursuit techniques, allowing the development of a set of visualizations to support decision making.

Baral [25] performed a stochastic analysis based on transport vehicles powered by various energy sources, namely, diesel, hybrid and electric trucks. Baral concluded that the use of hybrid and fully electric trucks powered by hydrogen and renewable sources of electricity, respectively, allowed achieving a large CF reduction, especially in the context of long-distance transportation.

Magazzino [26] critically evaluated the effect of dependence on fossil fuels and pollutant emissions from the transport sector on the performance of logistics operations. In the study a NN algorithm was implemented in a multivariate framework in order to analyze the correlation between Logistics Performance Indices, demand for petroleum products, and CO_2 emissions from fuel combustion in the transportation sector. The study demonstrated a high correlation between logistics performance indices and fuel consumption, as well as high GGE.

Cunha [11] applied statistical analysis, ANOVA to develop characterizations of energy performance and manufacturing quality of products in a factory. Cunha's study allowed inducing the influence of certain factors on energy performance, allowing developing predictive scenarios regarding energy consumption and manufacturing quality.

Table 3 represents the application and the method used in each paper included in our SLR.

Table 3 - Methods and applications

Title	Author	Application	Method
Transportation in a 100% renewable energy system	García-Olivares, Antonio Solé, Jordi Osychenko, Oleg	Propose a framework based, to compute the energy and monetary expenditures associated with the deployment of the new transport system	Discuss the main proven technologies, new infrastructure and policy measures that would make an optimal transition strategy; Compute the energy and monetary expenditures associated with the deployment of the new transport system
Trading off cost, emission, and quality in cold chain design: A simulation approach	Fan, Yun de Kleuver, Caroline de Leeuw, Sander Behdani, Behzad	Present an agent-oriented simulation framework to support decision-making in the design and operation of cold chains to trade off cost, emission, and quality	A case study of a global banana supply chain is presented and discussed; Simulation model is developed in Python with the Mesa package.
Towards sustainable wood procurement in forest industry – The energy efficiency of larger and heavier vehicles in Finland	Palander, Teijo Haavikko, Hanna Kärhä, Kalle	Discuss how the local biofuel cycling through larger and heavier vehicles may affect the sustainability of wood procurement in the industrial ecosystem by focusing on transport efficiency, cost-efficiency and energy efficiency	Quantitative energy-performance analysis from the optimization of results of the multi-objective dynamic biofuel cycle model; Deterministic energy-performance analysis of the industrial ecosystem was made using optimization results of the multi-objective dynamic energy-supply cycle model
Transition of mobility in companies – A semi-systematic literature review and bibliographic analysis on corporate mobility and its management	Gorges, Tobias Holz-Rau, Christian	Review literature between 2016 and mid-2020 and critically review research conclusions and their implication to determine research gaps for potential alternative ways of system transition	Semi-systematic review of international scientific journals and conference papers independently from their research field with a set of a-priori defined terms on corporate mobility and logistics
Data analytics for sustainable global supply chains	Mangina, Eleni Narasimhan, Pranav Kashyap Saffari, Mohammad Vlachos, Ilias	Acquire patterns in logistic operations based	Used algorithms Horizontal Cooperation; Pooling and Physical Internet

Modelling carbon emissions of diesel trucks on longitudinal slope sections in China	Dong, Yaping Xu, Jinliang Gu, Chenwei	Propose a carbon emission quantification model for diesel trucks on longitudinal slope sections and investigate the influence of gradient on the carbon emissions of trucks for use in the low-carbon highway design	Application of the law of conservation of mechanical energy; Predictive model to quantify the carbon emissions of diesel trucks
The scope for improving the efficiency and environmental impact of internal combustion engines	Leach, Felix Kalghatgi, Gautam Stone, Richard Miles, Paul	Define the scope for improving the efficiency and environmental impact of internal combustion engines	Discuss the basic principles that govern engine efficiency and the technologies to control exhaust pollution
Life cycle assessment of mid-range passenger cars powered by liquid and gaseous biofuels: Comparison with greenhouse gas emissions of electric vehicles and forecast to 2030	Ternel, Cyprien Bouter, Anne Melgar, Joris	Comparison with greenhouse gas emissions of electric vehicles and forecast to 2030	Current and future forecast comparative analysis of the greenhouse gas emissions of mid-range passenger cars, obtained using life cycle assessment
The development of a carbon footprint model for the calculation of GHG emissions from highways: the case of Egnatia Odos in Greece	Roukounakis, Nikolaos Valkouma, Efthaleia Giama, Efrosini Gerasopoulos, Evaggelos	Delineates the operation review of three third-party logistics firms in Hong Kong and develops an organisation-based carbon emission measurement metrics for logistics operations	Excel Solver and Visual Basic Application (VBA) programming tools
An Intelligent Visualisation Tool to Analyse the Sustainability of Road Transportation	de Armiño, Carlos Alonso Urda, Daniel Alcalde, Roberto García, Santiago Herrero, Álvaro	Analyse the Sustainability of Road Transportation	Artificial Intelligence techniques, Machine Learning
Simulation study on carbon emission of China's freight system under the target of carbon peaking	Wen, Lei Song, Qianqian	Simulation study on carbon emission	Application of 13 scenarios of carbon peaking to explore the paths of carbon peaking before 2030
Well-to-wheels approach for the environmental impact assessment of road freight services	Osorio-Tejada, Jose Luis Llera-Sastresa, Eva Hashim, Ahmad Hariza	Proposes a method for the energy consumption and emissions estimation based on vehicle operating conditions	The accounting of emissions was elaborated based on the equations, and coefficients from the EMEP/EEA

COP21 Roadmap: Do innovation, financial development, and transportation infrastructure matter for environmental sustainability in China?	Umar, Muhammad Ji, Xiangfeng Kirikkaleli, Dervis Xu, Qinghui	Explain the long-run and causal effects of innovation, financial development, and transportation infrastructure on CO2 emissions	Bayer-Hanck cointegration test
Evaluation of sustainable transport research in 2000–2019	Zhao, Xianbo Ke, Yongjian Zuo, Jian Xiong, Wei Wu, Peng	Identify the hot research topics, explore knowledge gaps and recommend future directions in the domain of sustainable transport	Bibliometric records of journal articles were searched from the Web of Science core collection database
Optimising Supply Chain Logistics System Using Data Analytics Techniques	Mangina, Eleni Narasimhan, Pranav Kashyap Saffari, Mohammad Vlachos, Ilias	Evaluates models for European freight transport logistics actions utilising advanced data analytics solutions	Algorithms of horizontal collaboration, pooling, and physical internet
Biomass feedstock transport using fuel cell and battery electric trucks improves lifecycle metrics of biofuel sustainability and economy	Baral, Nawa Raj Asher, Zachary D. Trinko, David Sproul, Evan Quiroz-Arita, Carlos Quinn, Jason C. Bradley, Thomas H.	Presents the first detailed stochastic techno-economic analysis and life-cycle assessment of biomass feedstock supply systems with diesel, fuel cell hybrid electric, and fully electric trucks and determines their impacts on biofuel production considering butanol as a representative biofuel	Biomass Supply Analysis and Logistics, developed by Oak Ridge National Laboratory; Uniform-Format Solid Feedstock Supply System, developed by Idaho National Laboratory
A new artificial neural networks algorithm to analyze the nexus among logistics performance, energy demand, and environmental degradation	Magazzino, Cosimo Mele, Marco Schneider, Nicolas	Assesses the effect of fossil fuel dependence and polluting emissions from the transport sector on the performance of logistics operations in the context of Green Supply Chain Management	Artificial Neural Networks
Energy use and emissions scenarios for transport to gauge progress toward national commitments	Schmitz Gonçalves, Daniel Neves Goes, George Vasconcelos de Almeida D'Agosto, Márcio Albergaria de Mello Bandeira, Renata	Develop and model scenarios of energy use and GHG emissions from the transport sector until 2030	Bottom-up and top-down approaches based on IPCC

Optimising truckload operations in third-party logistics: A carbon footprint perspective in volatile supply chain	Wong, Eugene Y.C. Tai, Allen H. Zhou, Emma	Delineate the operation review of three third-party logistics firms in Hong Kong and develops an organisation-based carbon emission measurement metrics for logistics operations	Excel Solver and Visual Basic Application (VBA) programming tools
Comparison of product carbon footprint protocols: Case study on medium-density fiberboard in China	Wang, Shanshan Wang, Weifeng Yang, Hongqiang	Compare the criteria and implications of the three protocols to quantify Carbon footprint	Sensitivity analysis to assess the implications of different end-of-life disposals was conducted
Integrated effects of SCR, velocity, and air-fuel ration on gaseous pollutants and CO2 emissions from China V and VI heavy-duty diesel vehicles	Li, Xueyao Ai, Yi Ge, Yunshan Qi, Jingyu Feng, Qian Hu, Jie Porter, William C. Miao, Yaning Mao, Hongjun Jin, Taosheng	Evaluate the integrated effects of Selective Catalytic Reduction (SCR), velocity, and air-fuel ratio on carbon dioxide (CO2) and nitrogen oxide (NOx) emissions	Correlation analysis
Driver profile classification on heavy trucks based on telemetry	Alves, Daniel Janeiro	Create profiles for heavy-duty truck drivers describing their driving efficiency	Time series, clustering, feature engineering
Research on application of a hybrid heuristic algorithm in transportation carbon emission	Li, Yanmei Dong, Hong Kai Lu, Shuangshuang	Analyze the main sources of carbon emissions in the transportation industry, including nine major energy consumption sources such as coal, gasoline, and diesel, and obtains the carbon emission values from 2000 to 2017. Secondly, a linear regression analysis was performed on 13 pre-selected influencing factors and CO2 emissions in the transportation industry	Algorithms, Neural network, Extreme learning machine, Genetic algorithm optimized neural network, and Genetic algorithm optimized extreme learning machine

A new approach to reduce the carbon footprint in resistance spot welding by energy efficiency evaluation	Cunha, Carla Ferreira Andrade de Oliveira Gomes, Jefferson de Carvalho, Hugo Marcelo Bezerra	Propose a method to improve the resistance spot welding process's energy performance without compromising its quality	Statistical analysis (ANOVA)

2.3. Network analysis and visualization with VOSviewer

We used the VOSviewer [5] tool to generate and visualize the bibliometric networks considering citations, co-authorship, and text mining capabilities enabling to extract of co-occurrence networks of terms from our SLR. This tool also allowed the creation of graphical information such as mapping of relevant information regarding authors, abstracts, titles, and keywords.

2.3.1. Keyword occurrence analysis

We performed with VOSViewer tool, a keyword occurrence analysis to the SLR papers. This analysis was performed considering the full counting option, and we defined as minimum number of occurrences of the keywords to 1. Due to the factors considered, the tool identified a total of 104 keywords, of which only 30 met the threshold, and were selected.

Table 4 illustrates the results obtained based on their occurrence and relevance. From the keywords identified by VOSViewer, it is noteworthy the fact that most of the keywords were related to some characteristics intrinsic to the transportation process. Some keywords related to data analysis methodologies were also identified. The words with the highest number of occurrences were road transportation (with 2 occurrences and a link strength of 12), life cycle assessment (with 2 occurrences and a link strength of 10) and carbon footprint (with 2 occurrences and a link strength of 7).

Table 4 - Keyword occurrence

Keyword	Occurrences	Total link strength
road transportation	2	12
life cycle assessment	2	10
carbon footprint	2	7
2030	1	6
age of transport means	1	6
artificial intelligence	1	6

biofuels	1	6
clustering	1	6
electrified vehicles	1	6
energy mix	1	6
exploratory projection pursuit	1	6
passenger car	1	6
transport sustainability	1	6
unsupervised machine learning	1	6
sustainability	2	4
polluting emissions	1	4
road transport fuels	1	4
truck emissions	1	4
logistics operations journal: journal of cleaner p	1	3
mobility management	1	3
optimisation	1	3
road freight transport	1	3
supply chain efficiency	1	3
transition	1	3
carbon emission	2	1
co2 emissions	2	1
logistics	1	1
road freight transportation	1	1
transportation infrastructure	1	1
efficiency	2	0

VOSViewer identified 8 clusters (Figure 2), with a total of 30 keywords and 67 links. Resulting from the analysis of the visualization, we highlight the words identified in red, green, and blue, which integrate the largest clusters of words. The referred 3 clusters present the same number of links and relate terms associated with the logistics and transport activity as well as its CO_2 emission.

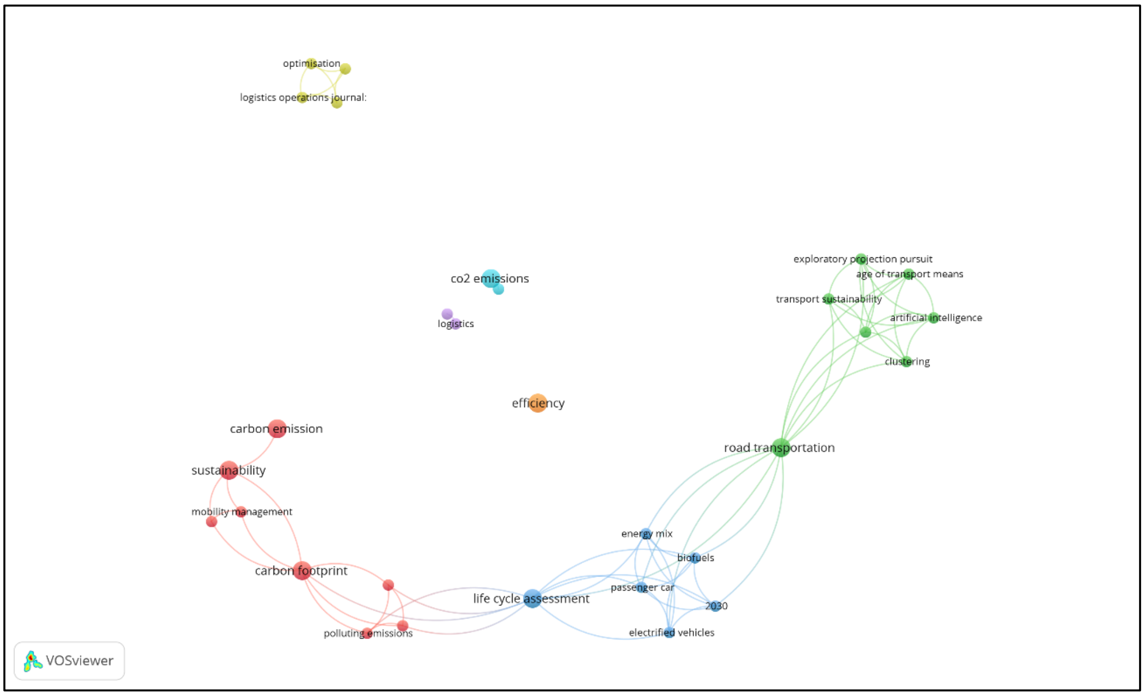

Figure 2 - VOSviewer clustering by Keywords

2.3.2. Title and abstract occurrence analysis

VOSViewer was also applied to perform a title and abstract occurrence analysis. In this analysis we considered the options full counting, and a minimum number of occurrences of 5. As a result, we obtained 965 words, of which only 36 met the threshold. Later, we selected for analysis 22 words, which represent 60% of the most relevant words as suggested by the VOSViewer tool. Table 5 illustrates the results obtained regarding the number of occurrences and the relevance of each of the terms identified.

The most relevant words are quality, environmental impact, technology, energy efficiency, transport, financial development, SCR (Selective Catalytic Reduction), and innovation, which all have a relevance score higher than 1.00. It should be noted that these are the most relevant words used in the documentation of the papers by the research communication.

Table 5 - Title and abstract occurrence

Term	Occurrences	Relevance
quality	10	1.59
environmental impact	7	1.43
technology	11	1.33
energy efficiency	5	1.24
transport	21	1.10
financial development	5	1.02

scr	5	1.01
innovation	6	1.01
etp	5	1.00
freight system	6	1.00
carbon peaking	7	1.00
efficiency	14	0.98
energy	11	0.97
term	5	0.97
transport sector	7	0.96
transportation industry	7	0.96
period	5	0.84
logistic	7	0.78
carbon emission	17	0.73
sustainability	11	0.71
c02 emission	11	0.70
china	11	0.68

Figure 3 illustrates, the word network visualization associated with the exercise in scope. VOSviewer developed a network with 22 terms, 4 clusters, 81 links and 920 strength links.

The main nodes were carbon emission, CO2 emissions and transport. These were the words that, as shown in Table 5, had the highest number of occurrences, although they were not the most relevant words used by the research community, as previously mentioned in the analysis of Table 5.

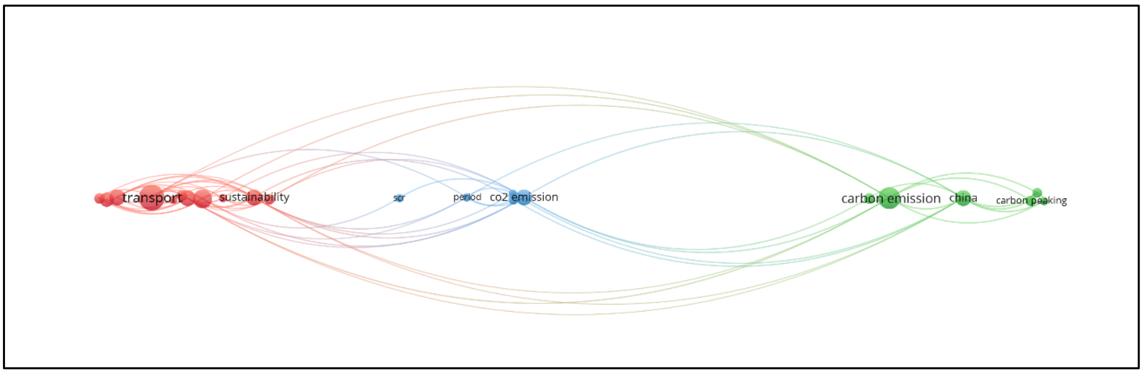

Figure 3 - VOSViewer network visualization

2.3.3. Author co-authorship analysis

An Author Co-Authorship Analysis was also performed using the VOSViewer. We considered the options full counting and a minimum number of occurrences of 1. As a result, we obtained 86 identified authors, 86 met the threshold, and 21 were selected (those with a link strength higher or equal to 6), according to Table 6.

The authors in Table 6 presented a higher level of link strength and were subsequently grouped into 3 clusters as visible in Figure 4 through network visualization, which is composed of 72 links and a total of 78 link strengths.

Table 6 - Author co-authorship

Author	Documents	Total link strength
ai, yi	1	9
asher, zachary d.	1	6
baral, nawa raj	1	6
bradley, thomas h.	1	6
feng, qian	1	9
ge, yunshan	1	9
hu, jie	1	9
jin, taosheng	1	9
li, xueyao	1	9
mangina, eleni	2	6
mao, hongjun	1	9
miao, yaning	1	9
narasimhan, pranav kashyap	2	6
porter, william c.	1	9
qi, jingyu	1	9
quinn, jason c.	1	6
quiroz-arita, carlos	1	6
saffari, mohammad	2	6
sproul, evan	1	6
trinko, david	1	6
vlachos, ilias	2	6

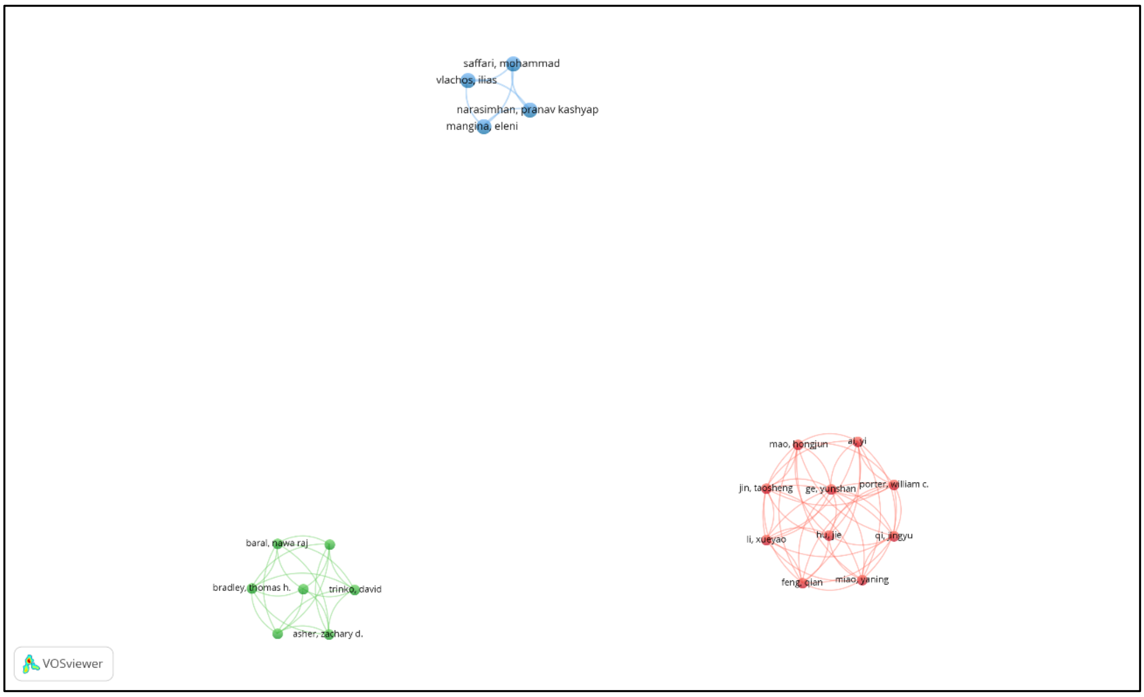

Figure 4 - Author and co-author network visualization

2.4. Influence of logistic transport on carbon footprint

Logistics and passenger transport by combustion vehicles contribute about 25% to CO_2 emissions [10]. The logistics transport sector has an extensive environmental, social, and economic impact on society. Due to environmental degradation, regarding the increase in the level of greenhouse gas emissions (GGE), such as CO_2, several measures have been implemented. We found that some articles demonstrate the importance of the logistic transport sector, namely, land transport, as one of the main drivers of GGE.

According to Wen [27], in 2019 in China, about 4,713 billion tons were transported, of which 72.9% through land vehicles. According to Wen, in 2019, about 4,713 billion tonnes were transported in China, of which 72.9% was through land vehicles, which has the highest levels of pollution.

As shown in the study of Leach [10], currently, about 99.8% of vehicles used for land transport were powered by internal combustion engines, and 95% of this energy comes from fuels produced from oil. The Leach study concluded that the reduction of pollution levels associated with the logistics and transport activity involves changing some existing practices in the market, namely related to the development of engine types with lower energy consumption and the use of hybrid vehicles.

In our SLR, we also identified articles where the authors identify and analyse, through holistic research, research topics related to the carbon footprint. Zhao [12], developed a holistic research that started on the analysis of publications of research articles between the years 2000 and 2019, and explored gaps and possible aspects of improvement, in order to identify recommendations and future orientations for the development ofsustainable transport. This analysis resulted in the following research topics: 1) indicators of sustainable transport and performance model, 2) sustainable transport policy, 3) stakeholder involvement, 4) supply chain management and logistics, 5) environmental impact, 6) travel behaviour, 7) new vehicle fuels, 8) the strategic planning of transport, and 9) cycling and public transport.

In the research work of Gorges [28], which sought to study the mobility transition in companies, based on a literature review analysis on corporate mobility and its management, identified that the main goal of companies was to achieve zero emissions. For the reduction of the carbon footprint, Gorges identified the following business motivations: to analyse the modes of transport and their impact on the carbon footprint; to analyse the drivers' behaviour; to identify measures that allow the reduction of emissions; to develop tools for monitoring the carbon footprint; to examine the factors and their relationship with the carbon footprint; to understand the impact of information systems; and to develop sustainable organisations.

In the study developed by Umar [8] on the long run and casual effects of innovation, financial development and transport infrastructure on CO_2 emissions using the combination cointegration approaches during the period 1971 to 2018, it was demonstrated that innovation is observed to be a significant predicter of CO_2 emissions during the period 2007 to 2013; in the long run, there are negative relationships between CO_2 emissions and financial development, showing that in the long run CO_2 emissions will result in lower financial returns; during the periods from 2000 to 2015, and 1985 to 1989, transport caused significantly more CO_2 emissions, suggesting that enhancements in innovation and transport infrastructure should be made to achieve environmental sustainability goals.

García-Olivares [9] analysed the technologies and systems being proposed or proven as an alternative to fossil fuel based transport, and their prospects for entering the post carbon era, both from a technological and energy perspective. The analysis concluded that 100% renewable transport was achievable, but not necessarily in line with an indefinite increase in resource consumption.

There was a high relationship between logistic transportation activity and CO2 emission as demonstrated in the articles presented. Transportation, namely land transportation influences in a very significant and negative way the pollution levels concerning GGE. In the next section, we analyze further the methodologies proposed for the calculation of the carbon footprint.

2.5. Carbon footprint calculation

We identified studies on methodologies for calculating the carbon footprint as well as the other protocols that support them. Wang [29] in his study identified and developed a comparative analysis between the various existing protocols for carbon footprint calculation. The carbon footprint is used to quantify GGE. Various protocols, such as Publicly Available Specification (PAS) 2050, GHG Protocol Product Standard (GHG Protocol), and ISO 14067 Carbon Footprint (CF) of Products (ISO 14067), have been developed for CF calculations.

These protocols follow the life cycle thinking approach. The calculation formula as provided in the protocols is the same (Equation 3), but its interpretation may vary according to the protocol adopted,

$$Ef,GHG = FCf \times EFf,GHG \tag{3}$$

where Ef, GHG is the GHG emissions from fuel combustion in the stationary sources (kg CO_2 e), f represents the types of energy employed, FCf represents the energy consumption of the type of energy, and EFf,GHG is the emission factor for the type of energy f by GHG.

Wang identified the differences associated with calculating the carbon footprint, mainly related to the interpretation of some factors considered in the calculation formula (Table 7).

Table 7 - An overview of key aspects specified in carbon footprint protocols: Publicly Available Specification (PAS) 2050, Greenhouse Gas Protocol Product Standard (GHC Protocol), and ISO 14067 Carbon Footprint of Products (ISO 14067).

Specifications and Requirements	**PAS 2050**	**GHG Protocol**	**ISO 14067**
Goals	To provide a uniform specification for GHG emissions of goods and services	To provide detailed guidelines on accounting and reporting	To standardize the quantification process and the communication of GHG emissions
Life cycle stage included	Cradle-to-grave	Cradle-to-grave	Cradle-to-grave Cradle-to-grate Gate-to-gate
	Cradle-to-gate	Cradle-to-gate	Partial life cycle
Cut-off criteria	Exclusion based on materiality (<1%); at least 95% of the complete product life cycle must be included; no scale-up requirement to account for 100%	No cut-off criteria exist, because 100% completeness is necessary	No specific criteria available

Capital goods		Excluded	Excluded, but encouraged to be included when relevant	Excluded if they do not significantly affect the overall conclusions
Biogenic carbon	Carbon storage	Stored carbon Whitin 100 years shall be recorded and accounted for in the CF calculations	For cradle-to-gate system, credit is given to biogenic carbon storage	If carbon storage is calculated, then it shall be separately reported but not included in the CF result
	Delayed emissions	A weighting factor is included and proposed	(a) Shall not be included	Shall not be included
Other exclusions	Land-use change	Specific procedure and provides default soil emissions per country	Provides guidance for determining attributable	Direct land-use change shall be separately documented; indirect land-use change should be considered
	Others	Other exclusions include the transport of workers to their workplace and consumers to purchase sites, human energy inputs to the process, and animals providing transport services		
Allocation		(1) Avoiding allocation by process subdivision or system boundary expansion (2) Supplementary requirements (3) Economic allocation	(1) Avoiding allocation by process subdivision and redefining the functional unit or system expansion (2) Physical relationships (3) Economic or other allocation methods	
Global warming potential			100 years	

(a) Shall mean recommendation

From the study, Wang concluded that the CF calculation varied depending on the interpretation of the factors considered in the formula, especially the ones related to the Cut-off criteria, other boundary issues, and the Biogenic Carbon.

Roukounakis [30] developed in his study methodology for calculating the CF of a highway road. The total GGE was calculated based on several factors, namely fuel and electricity consumption, air conditioning associated gases and waste disposal. Regarding the calculation of emissions, the model related to the following variables: Vehicle-kilometers; Average Speed; and Vehicle categories associated with transport capacity in tons.

Dong [31] developed a study where he proposed a model for quantifying CF for diesel trucks and investigated the influence of road gradient on trucks' carbon emissions. Dong concluded that gradient provides higher levels of GGE emissions and therefore, transportation routes and road construction with severe gradients should be avoided.

3. Data analysis and modeling

3.1. Data mining with CRISP-DM

The Cross Industry Standard Process for Data Mining (CRISP-DM) [32] on a six-step iterative process that aims to describe (Figure 5), in a methodological way, the data analysis process in order to improve the planning, organisation and implementation of data analysis projects.

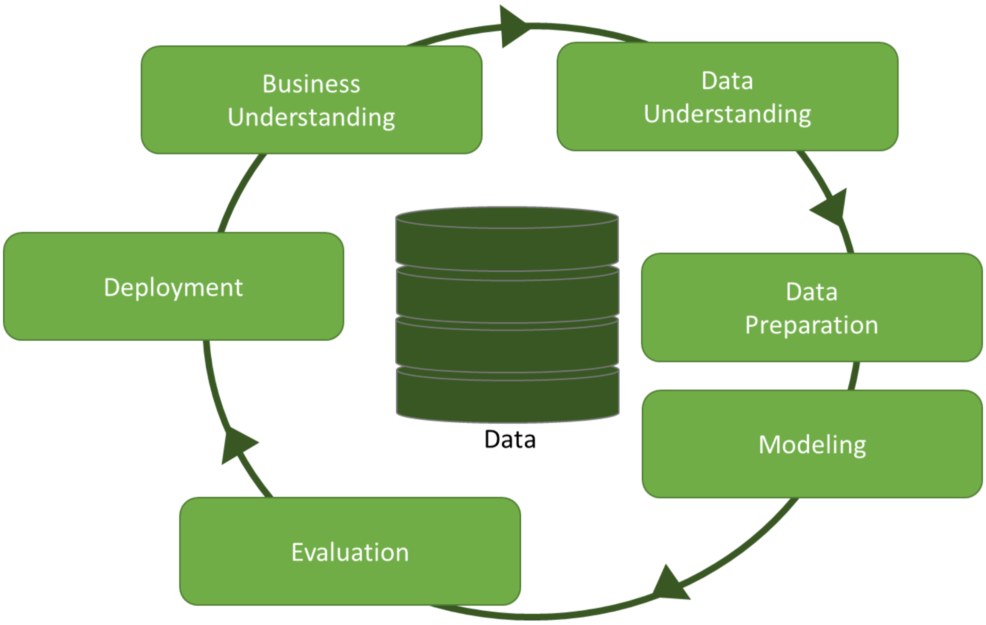

Figure 5- CRISP-DM Workflow

The first phase, Business Understanding, has as its main objective the clear definition of the problem under analysis, as well as the clear definition of objectives and methods to be applied to achieve the defined objective. This is followed by the Data Understanding phase, where the aim is to identify the potential of the data and its limitations to understand whether the data obtained allows for the resolution of the problem in question. Then follows the Data Preparation phase, where the aim is to perform some data preparation as necessary, namely data conversion, identification of incorrect formatting, missing data, among other aspects, through various data extraction methodologies. After data preparation, the Modeling phase follows, in which the objective is to define the best models to apply according to the defined objective. The models are then applied and evaluated in the Evaluation phase, to carry out a thorough assessment of the conclusions resulting from the application of the models, and to evaluate whether the applied model allows the defined objective to be achieved. Finally, there follows the application of the selected model in the Deployment phase, where the selected model is applied to solve the problem, which becomes the final solution to the problem.

The aim of this thesis is to develop an Information System adapted to the CO. In the Business Understanding phase we developed several interactions with the CO, namely interviews with the main decision makers, to define the study objective. The databases made available by the TRACKiT API were analyzed and interpreted in the Data Understanding and Data Preparation phases. We selected the best model that allowed us to quantify the carbon footprint considering the various available variables under study. Finally, the evaluation of the developed solution was evaluated in the Evaluation phase. For the purpose of this study, we end in the Evaluation phase.

3.2. Business understanding

CO was founded in 1982 and its mission is to analyze and consequently offer the best logistics solution to its customers, regarding the transportation of goods and merchandise.

With reference to 2021, its fleet consists of more than 500 vehicles, and 1200 employees spread over 21 operational platforms ("hubs") in Portugal, responsible for more than 8500 deliveries daily, annually moving approximately 750,000 tons of cargo.

3.2.1. Lean & Green

The LG initiative is the largest European collaboration platform specifically aimed to reduce CO_2 emissions associated with the supply chain, to achieve decarbonization, providing a network of sustainable logistics leaders.

In the strategic definition of CO, the environment is defined as one of the strategic pillars. In this context, CO joins GS1 Portugal as the Portuguese entity responsible for the LG initiative, in a partnership of trust on issues related to sustainability, in which it establishes a 5-year action plan. In its action plan, CO assumes a reduction target of at least 20% for a period of 5 years, regarding CO_2 emissions corresponding to the base line year of 2017, for the following business areas: Logistics, Transport and Distribution.

The action plan created by CO includes a description of the company and the main logistics activities, the scope of the project objective, the calculation methods and the data sources. The Action Plan also includes information regarding the definition of measures to be implemented to achieve the goal of a 20% reduction in the carbon footprint, the efficiency indicators and the monitoring chosen to analyze and verify the results.

3.2.2. Interactions with CO

In the scope of this research, meetings were held with the CO, namely interviews and a Focus Group session which was attended by the Administrator the Director of Development and Operations and the Director of Information Systems.

As a result of the interactions, the company's main objectives were defined in terms of sustainability and environment as follows:

CO's target aims to reduce CO_2 emissions per ton of freight transported by at least 20% between 2017 and 2022. CO_2 emissions were indicated in absolute value (t CO2) and in units of kg CO2 / t of cargo transported and kg CO2/km.

As part of the procedures developed in terms of Business Understanding, we also promoted interactions to obtain a better understanding of the data sources responsible for the management of the data, documented in detail in the Data understanding Data understanding.

3.3. Data understanding

The API responsible for creating and managing the data is TRACKiT. The API is managed externally, and activity reports are generated whenever requested by CO.

The data extraction was done through the API, extracting individual trips data by license plate of the vehicle. As a result, we obtained JSON format files, every month for each license plate. For each license plate we obtained twelve JSON files for all the months that comprised the year 2022, regardless of whether there were records or not.

Furthermore, CO has provided a list of 133 vehicles, related to the extracted data through the API, obtaining information regarding their activity for the first 6 months of the year 2022.

In order to answer the research questions, we first analyze the CO’s activity patterns and consequently, correlations between the various variables. Later, this allowed us to identify the main factors influencing the calculation of the carbon footprint. It is important to highlight, as concluded in the Literature review (Section 2) that the main source of CO_2 emissions in transport and logistics is fuel consumption. Thus, our main variable under study is fuel consumption. This analysis is documented in more detail in section 3.4.6.

After identifying the main factors influencing the carbon footprint, we developed a model that calculates the carbon footprint. This is documented further developed section 3.5.

After aggregating all the collected data in JSON format and transforming to dataframe, we obtained a dataset of 37 columns and 167,105 records.

Each of the records were represented as a "moment", from the moment the vehicle was turned on and off. Thus, a "rout" corresponded to a set of several aggregated moments, i.e., more than one dataset record.

The database metadata made available by the TRACKiT API is shown in Table 8.

Table 8 - Metadata CO's database

#	Column	Type	Description
1	Plate	string	The vehicle Plate
2	mid	Integer	The vehicle unique ID
3	ymd	string	Travel date (Y-m-d)
4	total_drive	integer	Total driving time (seconds)
5	total_stop	integer	Total time standing between trips (seconds)
6	total_km	integer	Total distance travelled (km)
7	total_fuel	float	Total of litres consumed (L)
8	max_speed	integer	Maximum recorded speed (Km/h)
9	distance_between_legs	float	Distance between the end of the previous trip and the starting point of the current trip (meters)
10	avg_speed	float	Average speed (km/h)
11	consumption	float	The average consumption in the trip (L/100Km)
12	ini_timestamp	string	Start date of trip (Y-m-d H:i:s)
13	ini_km	integer	Early in the trip odometer
14	ini_poi	integer	POI ID registered at the beginning of the journey
15	ini_poi_img	integer	POI recorded image at the beginning of the journey
16	ini_fractal	integer	Address of start of trip
17	ini_lat	integer	Latitude in decimal degrees of the beginning of the journey
18	ini_lng	integer	Longitude in decimal degrees of the beginning of the journey
19	ini_flv	integer	Fuel level (%)

20	ini_thermo_temp1	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
21	ini_thermo_temp2	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
22	ini_thermo_temp3	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
23	ini_thermo_temp4	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
24	ini_timestampUTC	string	Event date (Y-m-d H:i:s)
25	end_timestamp	string	End date of trip (Y-m-d H:i:s)
26	end_km	integer	Odometer value at the end of the trip
27	end_poi	integer	POI ID registered at the end of the journey
28	end_poi_img	integer	POI recorded image at the end of the journey
29	end_fractal	integer	Address of the end of trip
30	end_lat	integer	Latitude in decimal degrees of the end of the journey
31	end_lng	integer	Longitude in decimal degrees of the end of the journey
32	end_flv	integer	Fuel level (%)
33	end_thermo_temp1	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
34	end_thermo_temp2	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
35	end_thermo_temp3	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
36	end_thermo_temp4	integer	{pos} The position of the probe (1 to 4). First temperature reading in °C
37	end_timestampUTC	string	Event date (Y-m-d H:i:s)

To obtain a database containing only the routes performed, it was necessary to process it as an aggregated database by "route".

The procedures associated with data preparation are documented in section 3.4.

3.4. Data preparation

As introduced in the previous section, we developed an aggregated database, that was used to perform statistical analysis by route. However, we had first to develop some procedures for database treatment, namely: metadata analysis to identify eventual variables that had redundant information and did not add value to the scope of this analysis; identification and treatment of NULL records;

aggregation of records by route; identification of patterns and treatment of eventual inconsistencies if applicable.

3.4.1. Metadata analysis

The Metadata analysis proved to be fundamental to identify a set of variables that had redundant information on the scope of our analysis:

- [‘ini_poi’] and [‘end_poi’] – source and end route identifier ID;
- [‘ini_poi_img’] and [‘end_poi_img’] – code that identifies the image at the beginning and end of the path;
- [‘ini_fractal’] and [‘end_fractal’] – address of the origin and destination of the route;
- [‘ini_timestamp’] and [‘end_timestamp’] – start and end date of trip;
- [‘ini_timestampUTC’] and [‘end_timestampUTC’] - date of event record;
- [‘ini_thermo_temp1’], [‘ini_thermo_temp2’], [‘ini_thermo_temp3’], [‘ini_thermo_temp4’], [‘end_thermo_temp1’], [‘end_thermo_temp2’], [‘end_thermo_temp3’], [end_thermo_temp4’] – temperature.

As such, these variables were dropped from the dataset.

3.4.2. NULL removal

Next, we proceeded to the identification and removal of NULL data. In this context it was possible to verify that the variables [‘total_stop’], [‘total_fuel’], [‘avg_speed’] and [‘consumption’] present NULL data, as shown in Table 9.

Table 9 - NULL data count summary

Column	Nulls (Count)
total_stop	414
total_fuel	85
avg_speed	39
consumption	55,656

The number of NULL data, except for the variable [‘consumption’], was reduced and did not influence the results of our analysis. Thus, we proceeded to their removal.

As for the consumption variable, it resulted from the application of a calculation formula. Considering that in the following section we proceeded to the aggregation of the records by route, this was only done afterwards the data aggregationwas completed.

3.4.3. Aggregation of records by route

In order to obtain a database that aggregates all the records per route, we used the "groupby" command to summarize all the numeric variables by day and license plate. As such, each record in our database represented a daily route performed per license plate.

The new database created consisted in 24 columns and 13,057 records. The columns that constituted the database are presented in Table 10.

Table 10 - Metadata aggregated database

#	Column
1	total_drive
2	total_stop
3	total_km
4	total_fuel
5	distance_between_legs
6	ini_km
7	ini_lat
8	ini_lng
9	ini_flv
10	end_km
11	end_lat
12	end_lng
13	end_flv
14	avg_speed_agg
15	max_speed

16	avg_speed
17	consumption
18	Plate
19	Date
20	Type of vehicle
21	Brand
22	Model
23	Tag
24	Status

After aggregating the database, we processed the consumption variable, which results from formula application, as presented below,

$$consumption = \frac{total\_fuel}{total\_km} \times 100 \quad (4)$$

Once this process was concluded, we observed the existence of some NULL data, as shown in Table 11.

Table 11 - NULL data count summary

Column	Nulls (Count)
avg_speed	279
consumption	149

We concluded that there were records in our database whose variable number of kilometers traveled on the route [‘total_km’], the was null. This is due to the fact that there were some daily trips made in the CO hubs with no trips out of the distribution centers.

Thus, we eliminated all records whose [‘total_km’] is null (279 records), obtaining a database without null records and consisting of 12,778 records and 26 columns.

In this process, we also included some information that is not in the database, namely the year of registration ([‘Year_Plate’]); vehicle fuel ("[‘Fuel’]"), resulting in a database with 12,778 and 26 columns.

It should be noted that the information regarding the year of registration and vehicle fuel was obtained by making the following assumptions:

- Year of Registration - Obtained through information made available by the National Automotive Association (ARAN)
- Vehicle Fuel - Information provided by CO.

Subsequently, we re-evaluated the data and identified a set of variables that lost their context after the data aggregation. These variables are the following: ini_km, ini_lat, ini_lng, ini_flv, end_km, end_lat, end_lng, and distance_between_legs.

Having completed this process, we obtained a final database with 12,778 records and 18 columns.

3.4.4. Identification of patterns and treatment of eventual inconsistencies

We proceed to the identification of patterns to understand which variables were more relevant for fuel consumption and, consequently, CO_2 emissions.

Throughout this section we also intended to identify and explore possible inconsistencies and outliers that may arise from the analyses performed.

For this purpose, we performed time series, fuel type, and year plate analysis.

Time series analysis

With this analysis we aimed to identify any temporal patterns regarding the journeys. The distribution of transport during the first semester of 2022 is presented in Figure 6.

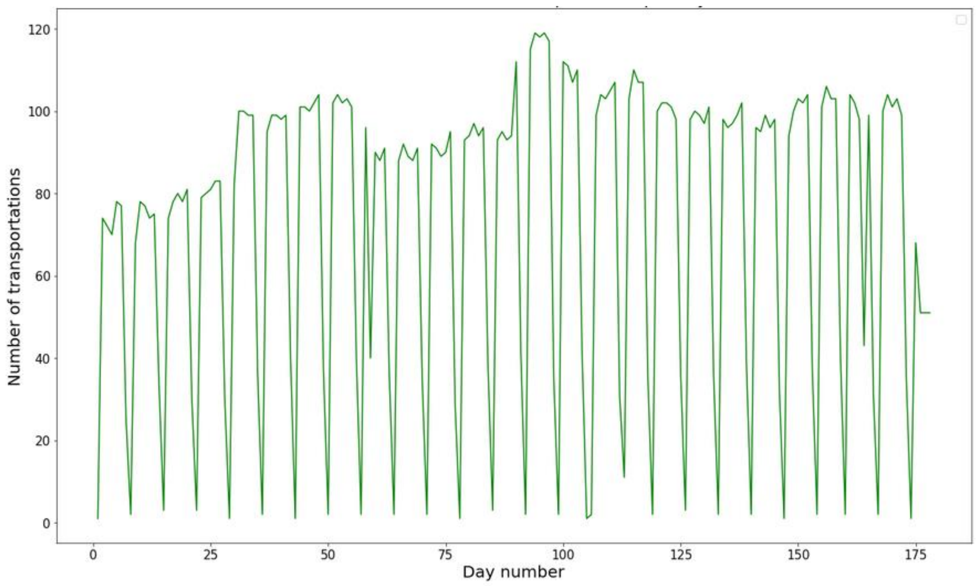

Figure 6 - Distribution of the number of transports performed per day, during the first semester of 2022

As shown in the visualization above, there were some days where the number of jouneys was almost null, showing a pattern. We performed an analysis of the monthly distribution, in order to better understand the pattern, as presented in Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11 and Figure 12.

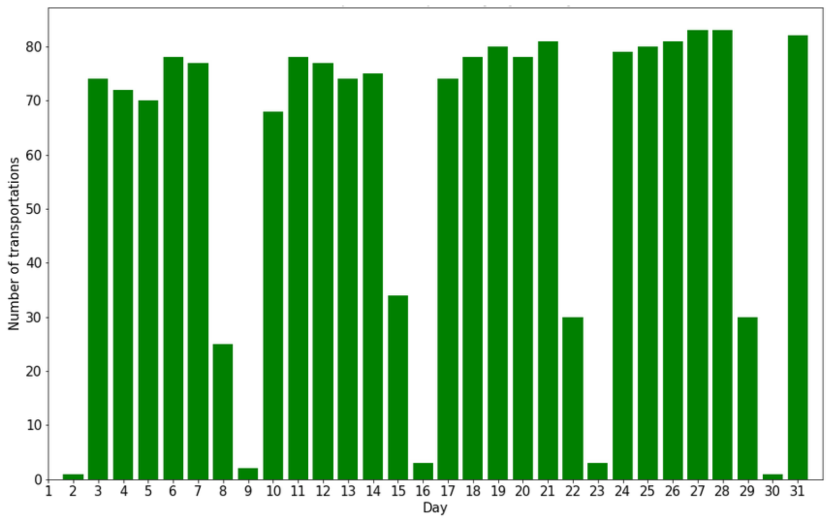

Figure 7 - Distribution of the number of transports performed per day, during the month of January 2022

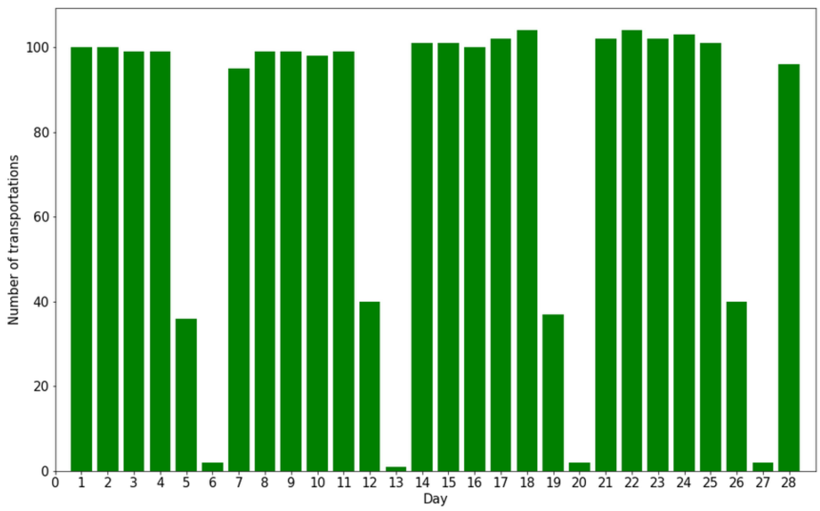

Figure 8 - Distribution of the number of transports performed per day, during the month of February 2022

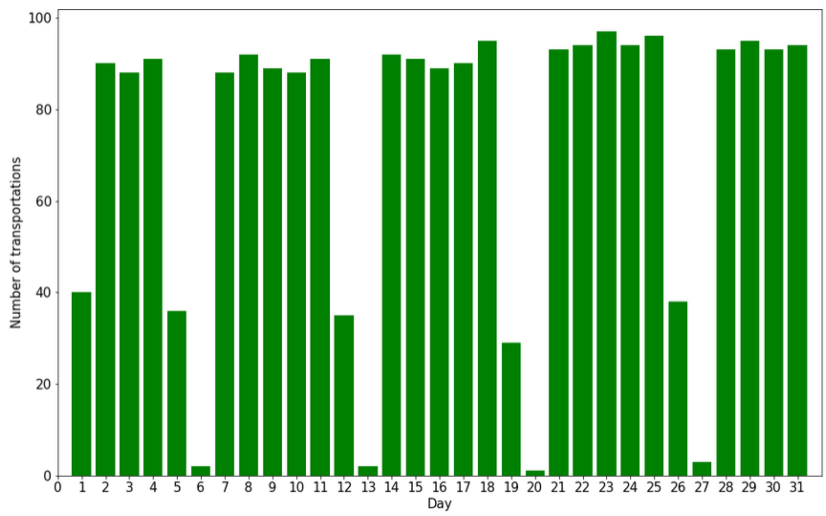

Figure 9 - Distribution of the number of transports performed per day, during the month of March 2022

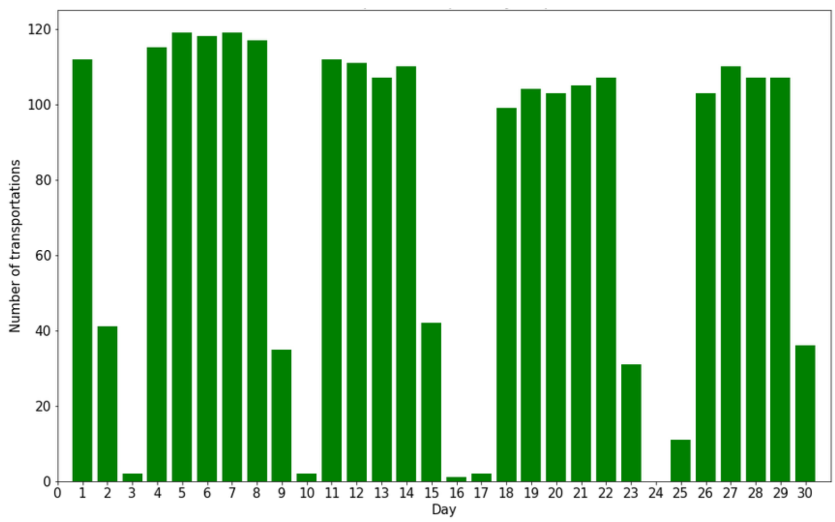

Figure 10 - Distribution of the number of transports performed per day, during the month of April 2022

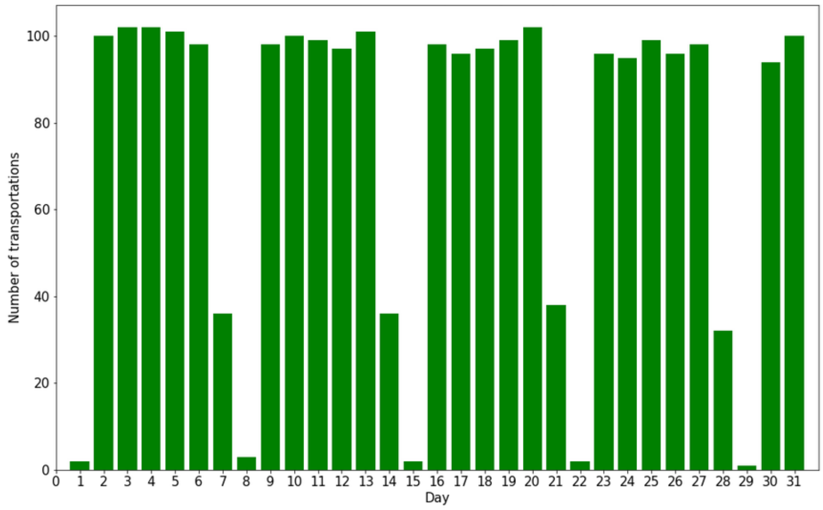

Figure 11 - Distribution of the number of transports performed per day, during the month of May 2022

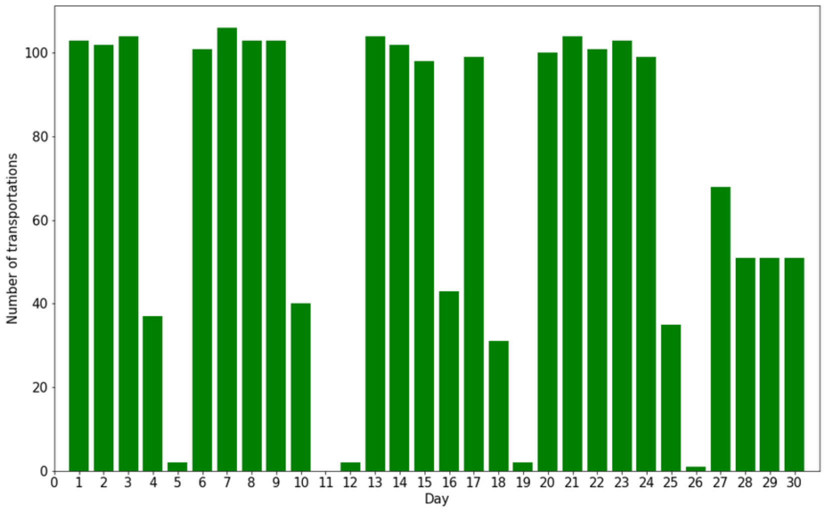

Figure 12 - Distribution of the number of transports performed per day, during the month of June 2022

In Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11 and Figure 12, we were able to confirm that the 7^{th} day of the week (Sunday) presented almost no transportation. The remaining days of the week showed a fairly constant level of transports.

Table 12 presents the average number of transportations per day of the week:

Table 12 - Average number of transportations per day

Weekday	Number of transportations per day
1	89.73
2	92.96
3	94.54

4	92.50
5	92.72
6	33.29
7	1.92

Fuel type analysis

As previously mentioned, the fuel type used was obtained through information provided by CO and later included in our database. Next, we present t In Figure 13we present he distribution of vehicles by fuel type.

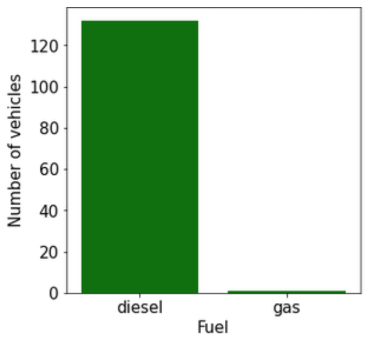

Figure 13 - Distribution of the number of vehicles by fuel type

The database visualization showed 132 diesel-powered vehicles and only 1 natural gas-powered vehicle.

Year plate analysis

Mentioned previously, the year of registration was obtained through an assumption. The Figure 14 shows the distribution of vehicles by year of registration ("Year_Plate"):

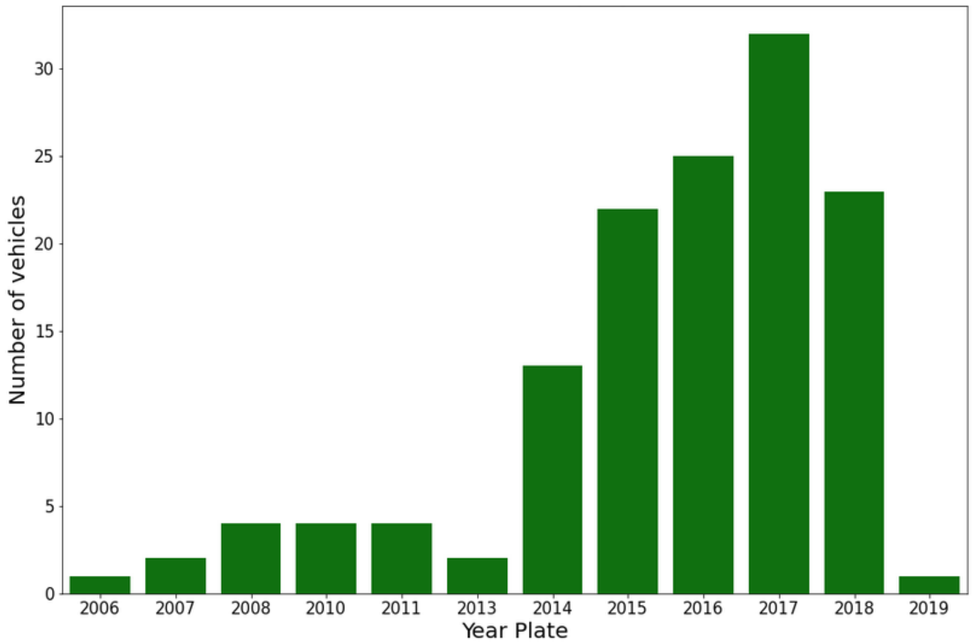

Figure 14 - Truck fleet distribution by registration plate year

The visualization presented in Figure 14 allowed us to conclude that the transport records were essentially of vehicles acquired in 2015, 2016 and 2017. The year with the least relevance is 2006.

Concerning the consumption level of the vehicles (liters/100km), as shown in Figure 15, we observed that the vehicles whose ['Year_Plate'] was more recent, presented a lower level of consumption. This was mainly due to the fact that the newer vehicles were technologically more advanced and had mechanisms that allowed a more efficient and less polluting activity. Nevertheless, the year 2010 had a reduced expression of consumption levels. This aspect was not representative of the database under analysis, considering that the number of vehicles with 2010 registration date was quite small.

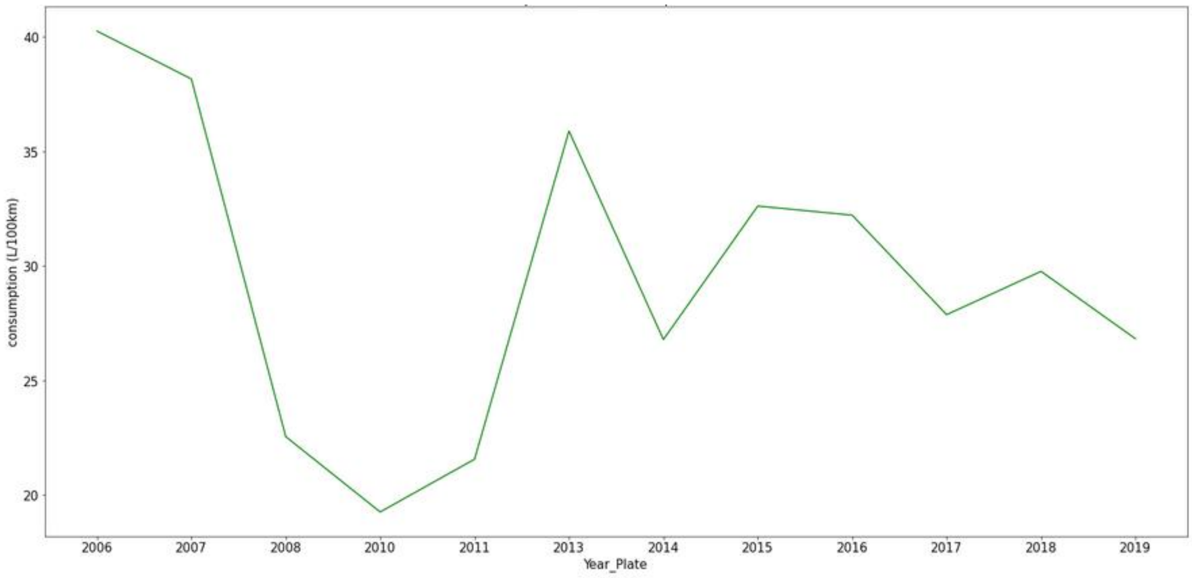

Figure 15 - Distribution of consumption (L/100Km) per Year Plate of the vehicle

3.4.5. Distribution analysis

In this section performed the distribution of a set of variables such as ['total_drive'], ['total_km'], ['total_fuel'], ['consumption'], ['max_speed'] and ['avg_speed'].

Distribution analysis of the variable ['total_drive']

The variable ['total_drive'] represents the duration of time spent on the route in seconds. The distribution of the variable is presented in Figure 16.

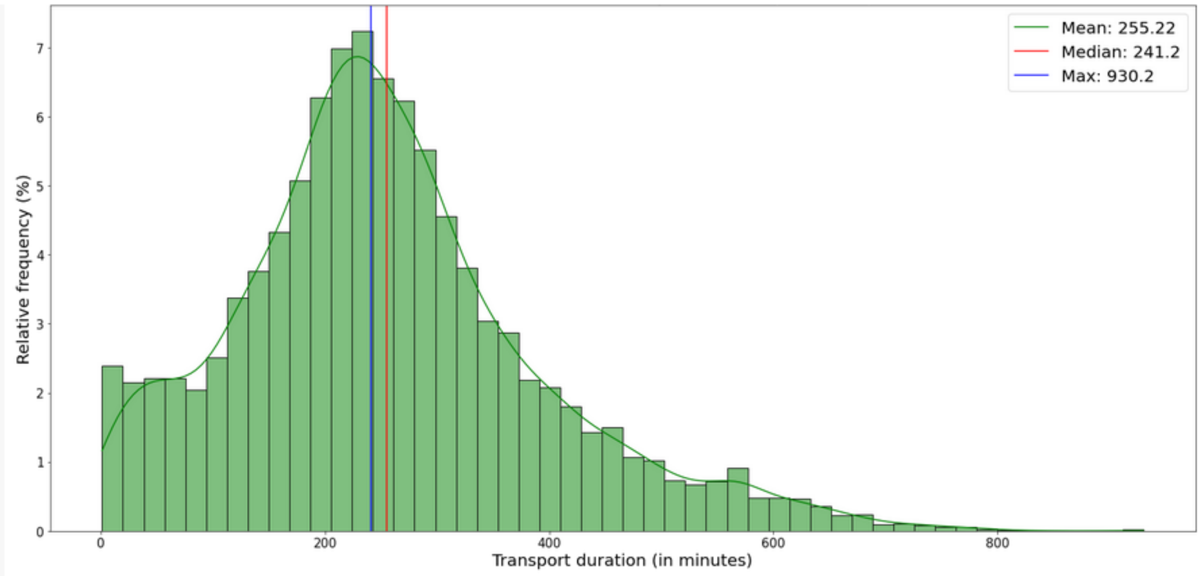

Figure 16 - Distribution of the variable - Transport durations in seconds (in minutes)

The distribution of the time spent on the route showed a mean of about 255.22 seconds (about 4 hours and 15 minutes) and a median of 241 (about 4 hours and 1 minute). We verified that the mean and median are quite close, showing a normal distribution.

It was also understood a pattern of short trips. We proceeded to discretize the variable by time intervals, in order to more easily detect the existence of outliers, as shown in Figure 17.

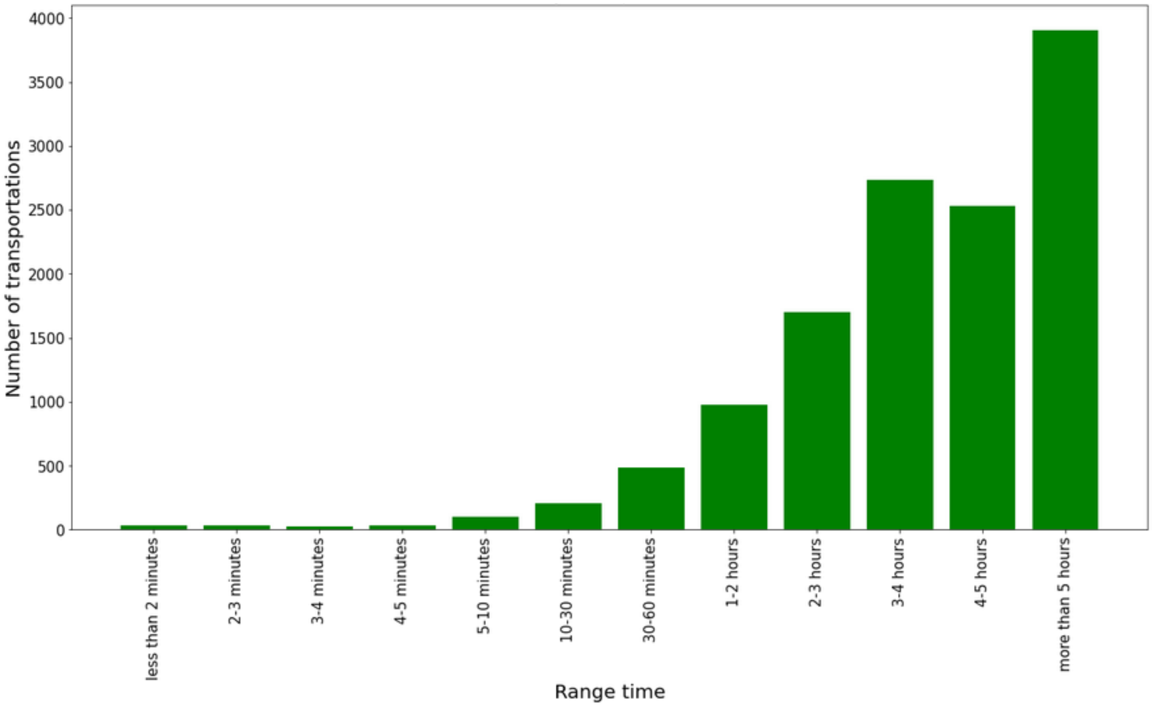

Figure 17 - Distribution of the variable - Transport durations by range time

Distribution analysis of the variable [‘total_km’]

The variable [‘total_km’] represents the distance of the route traveled in kilometers. The distribution of the variable is presented in Figure 18.

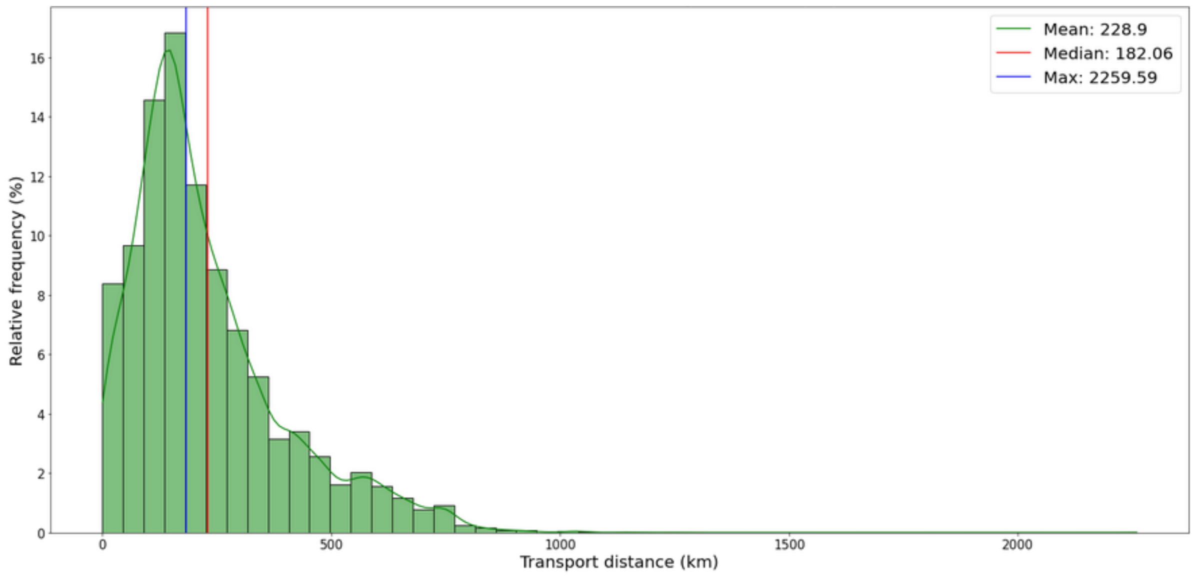

Figure 18 - Distribution of the variable - Transport distance in KM

In Figure 18, the routes taken by CO had an average of about 229 km and a median of 182 km. Only a trip had a greater travel distance of 2,260 km.

We proceeded to discretize the variable by distance intervals, to easily detect the existence of outliers, as shown in Figure 19.

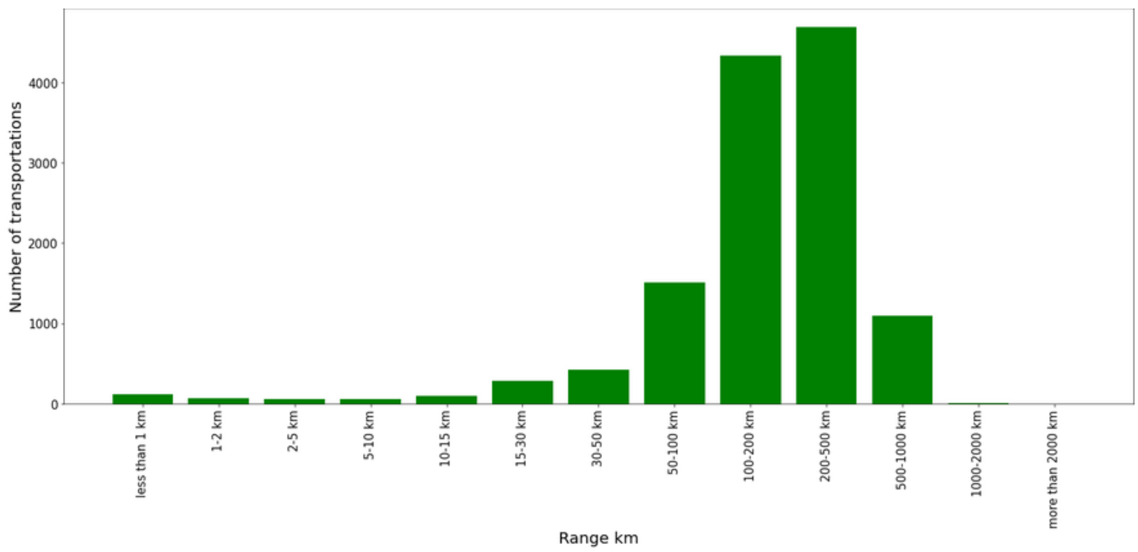

Figure 19 - Distribution of the variable - Transport distance by range of km

Distribution analysis of the variable [‘total_fuel’]

The [‘total_fuel’] variable represents the total number of liters consumed on the route. The distribution of the variable is presented in Figure 20.

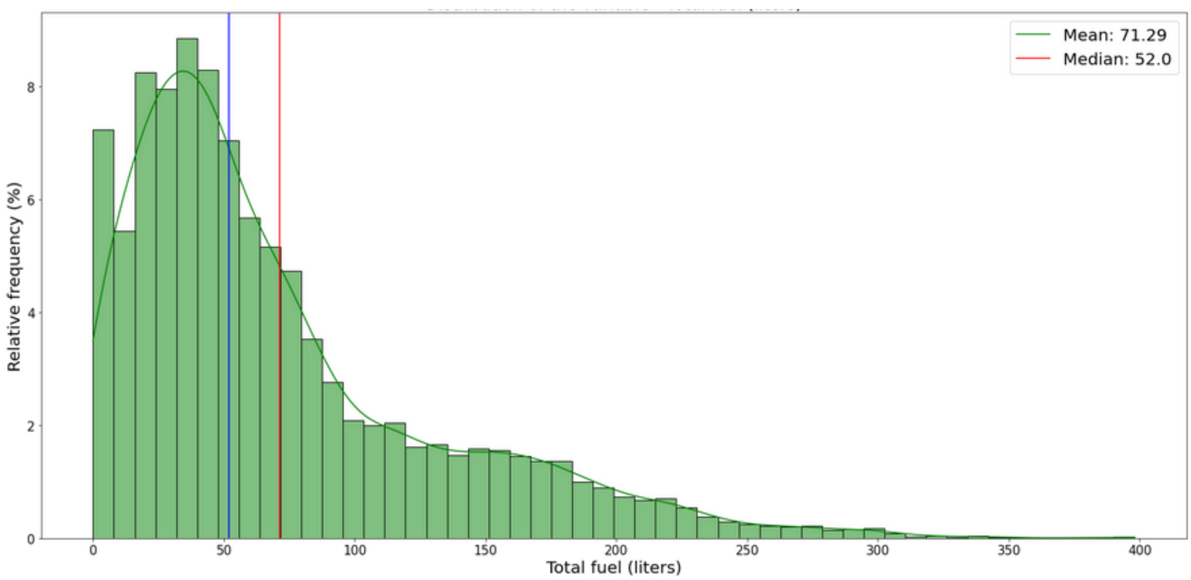

Figure 20 - Distribution of the variable - Total fuel in liters

As far as fuel consumption is concerned, the average number of liters consumed per route is 71.3, and the median is 52. It is important to highlight the existence of a relevant deviation between the average and the median of the distribution, showing disparity in the fuel levels per route, which is intrinsically related to the distances traveled, as can be seen in more detail in Figure 24

Figure 24 - Correlation matrix

Distribution analysis of the variable [‘consumption’]

The [‘consumption’] variable represents the average number of liters consumed per 100 km. Before proceeding to the distribution of the variable visualization, we determined statistical indicators, as detailed in Table 13.

Table 13 - Statistic information about the variable [‘consumption’]

Statistic	Describe
count	12778.00
mean	87.11
std	3472.09
min	0.00
0,25	25.00
50%	29.95
75%	35.63
max	364615.38

In this way, we eliminated all records whose variable value exceeds the consumption of 200 liters per 100 km. As a result of this procedure, we eliminated 39 records, resulting in 12,739 total.

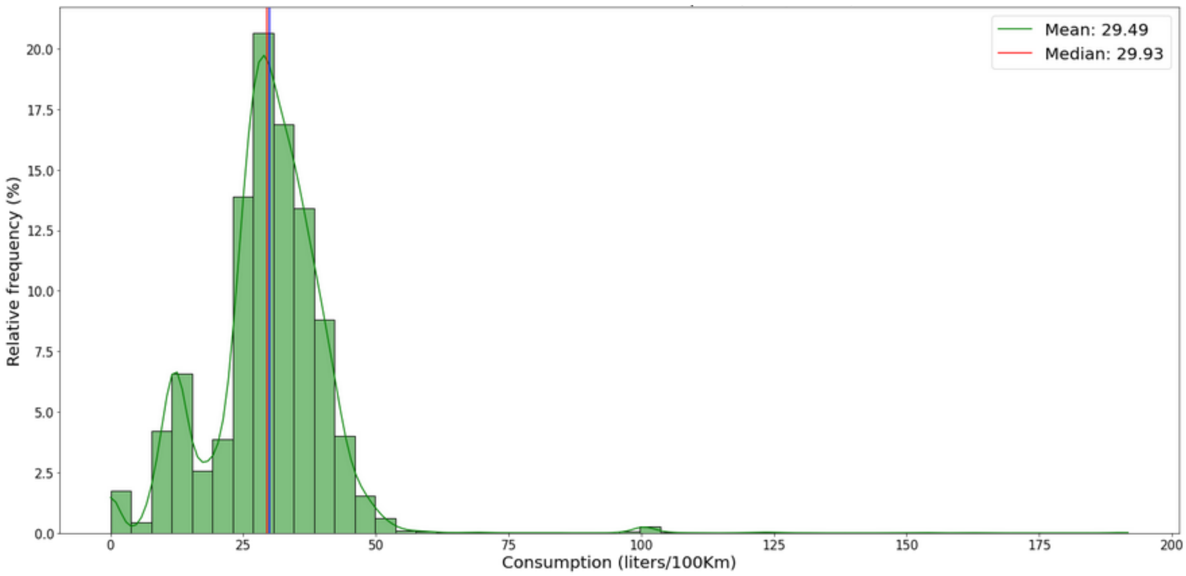

Figure 21 - Distribution of variable [‘consumption’] (liters/100kmh).

As can be seen in Figure 21, regarding the number of liters consumed per 100/km, the distribution presented a mean and median that are quite similar, showing a normal distribution. Nevertheless, it is equally noticeable that there were some registered peaks for lower levels of consumption, although with lower frequencies.

Also, the number of liters consumed per 100/km was naturally aligned with other variables that did not exist in our database, such as the inclination of the roads or the weight of the cargo transported which, consequently, led to greater consumption effort of the vehicles and an increase in the level of fuel consumption, and to the driving patterns behavior.

Distribution analysis of the variable [‘max_speed’]

The [‘max_speed’] variable represents the average number of liters consumed per 100 km. Before proceeding to visualize the distribution of the variable, we determined some statistical indicators, as detailed in Table 14.

Table 14 - Statistic information of variable [‘max_speed’]

Statistic	Describe
count	12739.00
mean	93.85

std	15.23
min	9.00
Quartile 25%	90.00
Quartile 50%	93.00
Quartile 75%	98.00
max	295.00

As shown inTable 14, there were records with a maximum value of 295 Km/h. Thus, we considered the indicated speed inconsistent, so we decided to eliminate all records whose value is higher than 160 (24 records), f resulting in 12,715 records.

The distribution of this variable is presented in Figure 22.

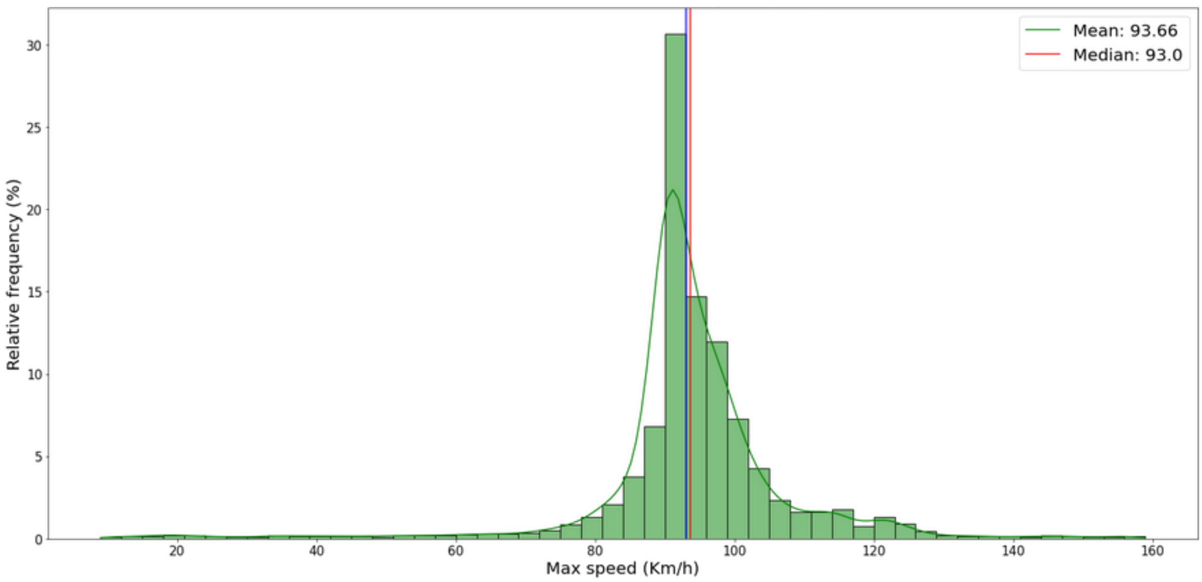

Figure 22 - Distribution of variable ['max_speed'] (Km/h).

The distribution of the variable presented a mean and median around 93 Km/h, showing a normal distribution.

Distribution analysis of the variable [‘avg_speed’]

The [‘avg_speed’] variable represents the average speed.

Before we proceed to visualize the distribution of the variable, we computed statistical indicators, as detailed in Table 15.

Table 15 - Statistic information of variable [‘avg_speed’]

Statistic	Describe
count	12715.00
mean	55.01
std	21.37
min	0.00
Quartile 25%	43.08
Quartile 50%	54.46
Quartile 75%	65.43
max	941.00

We decided to eliminate all records with values higher than 160 Km/h (24 records), resulting in 12,673 records. The distribution of the variable is presented in Figure 23:

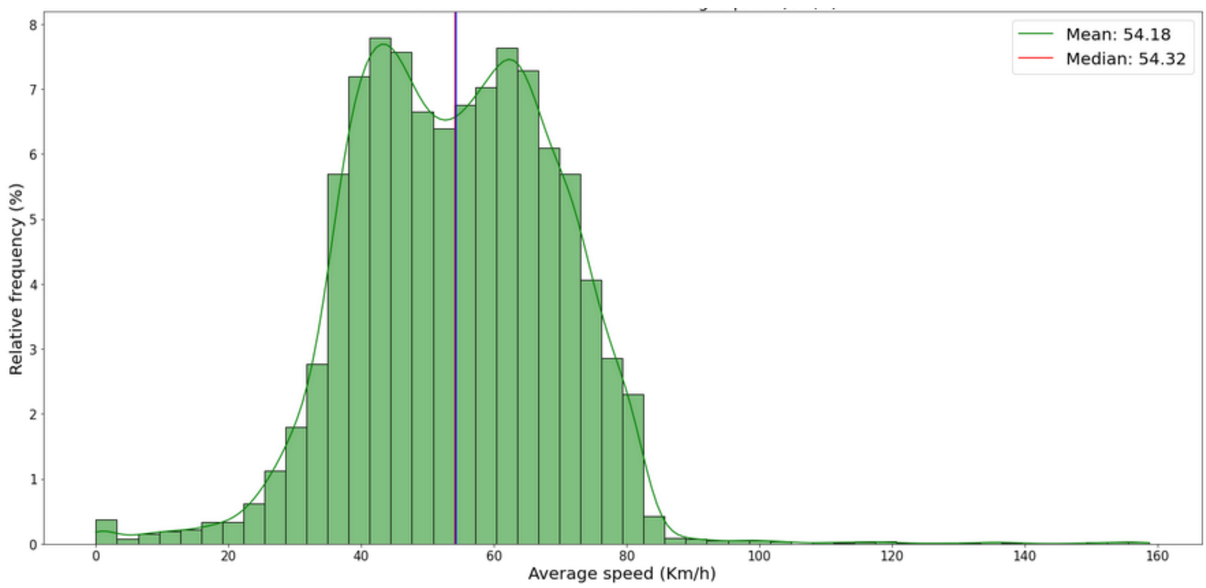

Figure 23 - Distribution of variable [‘average speed'] (Km/h).

In terms of average speed, it presented a very similar average and median, around 54 Km/h.

In the next section we analyze correlations between variables. The most important variable under analysis was ['fuel_total'], which determines the number of liters consumed per trip and consequently the CO_2 emission. This variable has an added importance considering that fuel consumption is the main CO_2 emitter.

3.4.6. Correlation analysis

Correlation matrix

As shown in the correlation map, presented in Figure 24, the ['total_fuel'] variable showed a strong correlation with ['total_drive']. Likewise, with ['total_km'], and ['avg_speed'], as expected, resulting from the graphical interpretation of the previous visualizations.There is also a strong correlation for the consumption's variable, although smaller than the other variables.

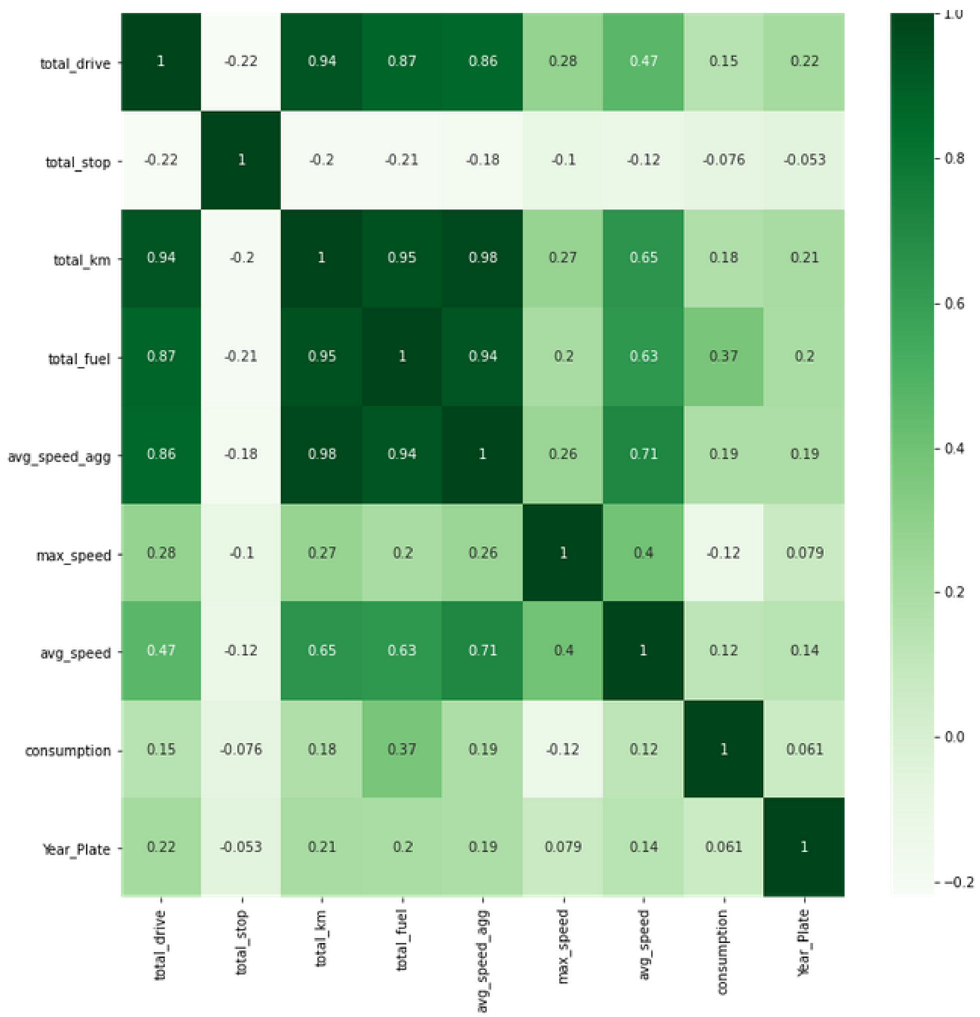

Figure 24 - Correlation matrix

In this section we present the correlation analysis of the following sets of variables: correlation matrix, [‘total_fuel’] and [‘consumption’], [‘total_fuel’] and [‘total_drive’], and [‘total_fuel’] and [‘avg_speed’].

Correlation analysis – [‘total_fuel’] and [‘consumption’]

The visualization below correlates the total liters of fuel consumed and the level of fuel consumption. As shown in Figure 25, the trips where the number of liters consumed was higher, had a more moderated average consumption of about 45 liters/100 km. It was also evident that there were some trips whose consumption at 100 km was quite high, but the liters consumed was small.

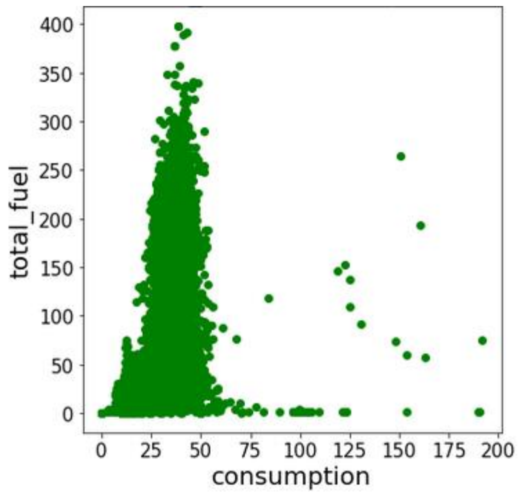

Figure 25 - Correlation analysis between the variables [‘total_fuel’] (liters) and [‘consumption’] (liters/100km)

Correlation analysis – [‘total_fuel’] and [‘total_drive’]

Regarding the total number of liters consumed compared to the total time spent on the trip, in Figure 26, it can be seen that the number of liters consumed is greater the longer the time spent on the trip.

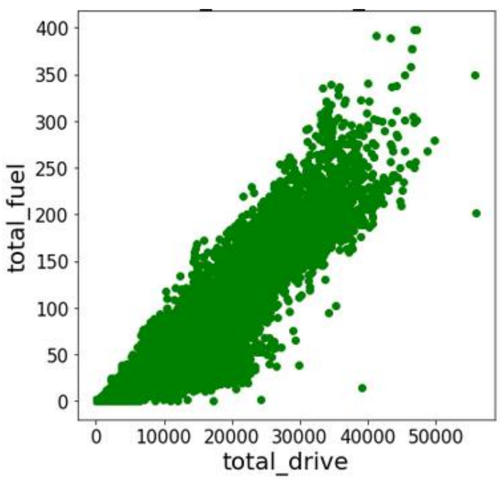

Figure 26 - Correlation analysis between the variables [‘total_fuel’] (liters) and [‘total_drive’] (in seconds)

Correlation analysis – [‘total_fuel’] and [‘total_km’]

Regarding the total liters consumed compared to the distance traveled on the route, in Figure 27, the liters consumed was higher the longer the distance of the route.

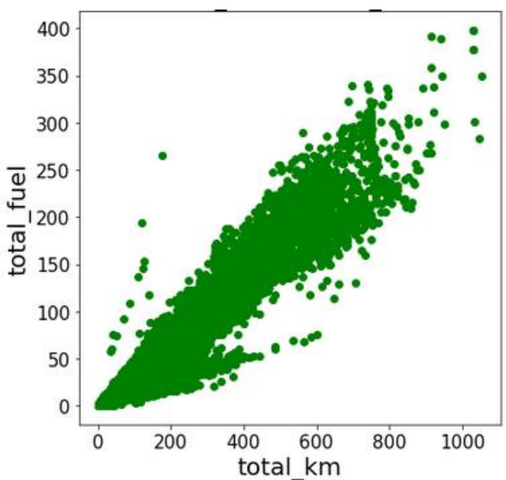

Figure 27 - Correlation analysis between the variables [‘total_fuel’] (liters) and [‘total_km’]

Correlation analysis – [‘total_fuel’] and [‘avg_speed’]

Regarding the total number of liters consumed compared to the total time spent on the trip, inFigure 28, the liters consumed was greater the longer the average speed.

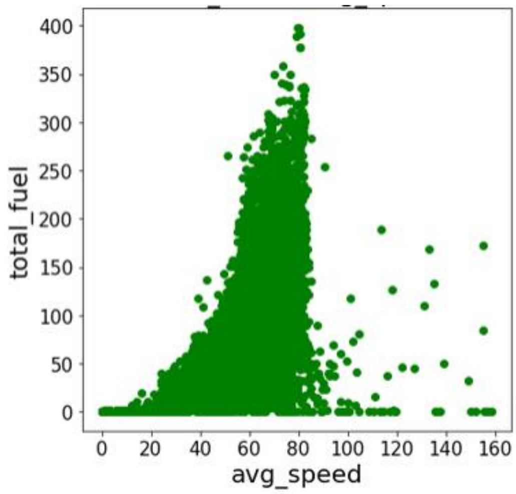

Figure 28 - Correlation analysis between the variables [‘total_fuel’] (liters) and [‘avg_speed’]

Correlation analysis – [‘total_fuel’] and [‘max_speed’]

In Figure 29,there were higher levels of consumption at higher speeds. Nevertheless, the speeds were limited by vehicle capabilities and, of course, by roud traffic regulation, which is why speeds typically do not exceed 120 Km/h.

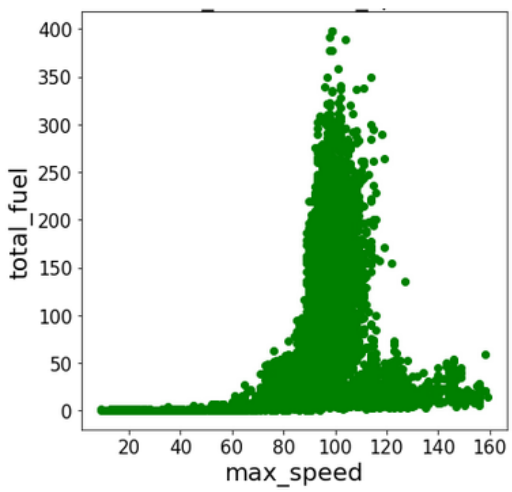

Figure 29 - Correlation analysis between the variables [‘total_fuel’] (liters) and [‘max_speed’]

3.5. Data modelling

3.5.1. Lean & Green conversion factor

In the Lean & Green program, the fuel consumption is converted into CO_2 emissions. The conversion factors by fuel type are shown in Table 16.

Table 16 - Conversion factor by liter as described in Lean & Green

Fuel type	Conversion factor/liter
Natural gas	1.887
Diesel	3.230

The formula for calculating the CO_2 emission per route is presented below:

$$Total\ KgCO2 = Conversion\ factor/liter\ \times total\_fuel \quad (5)$$

Therefore, the greater the number of liters consumed by the vehicle, the greater the CO_2 emission. Thus, the longer trips had a higher level of CO_2 emissions, the greater carbon footprint was generated by CO.

The distribution of the carbon footprint by transport is shown in Figure 30.

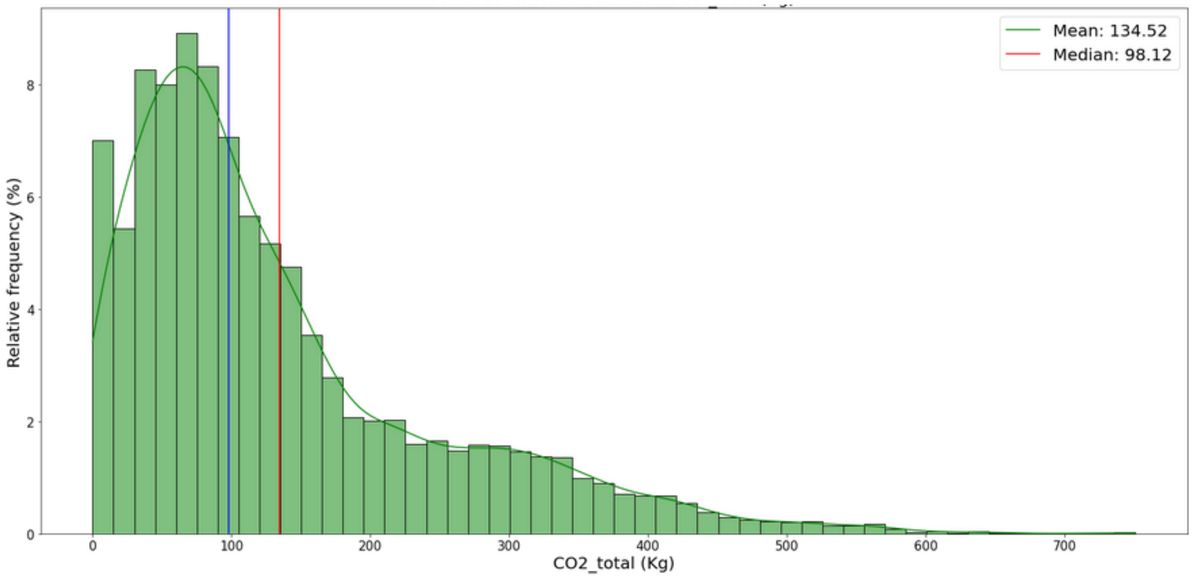

Figure 30 - Distribution of total CO2 emissions (Kg)

Within the scope of CO's Action Plan, CO determined as the primary metric for monitoring its carbon footprint, the indicator kg CO_2/t, which represents the total kg of CO_2 emitted per ton of cargo transported. However, considering that the database did not have information on the transported cargo, we developed a set of additional metrics.

Also, the Action Plan defined other metrics to monitor CO_2 emissions associated with the transport activity, namely those indicated below:

- Fleet emissions per liter of fuel consumed (kg CO_2/l), which represents the total kg of CO_2 emitted per liter of fuel
- Fleet emissions per km traveled (Kg CO_2/km), which represents the total kg of CO_2 emitted per km traveled

Additionally, although not a metric identified in the CO's Action Plan, we analyzed the following metrics:

- Fleet emissions per route (kg CO_2/r), which represents the total kg of CO_2 emitted per route
- Fleet emissions per day (kg CO_2/d), which represents the total kg of CO_2 emitted per day

The information needed to calculate the metrics above is available in the database provided by CO. Below we present a set of analyses on these metrics.

Figure 31shows the distribution of the number of routes performed per month.

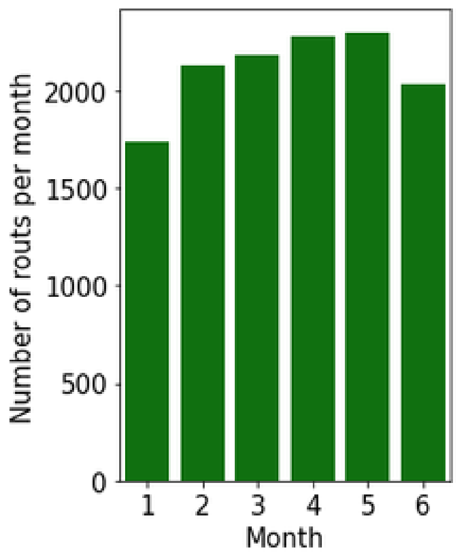

Figure 31 - Distribution of number of routs per month

As shown in figure Figure 31, the number of routes was typically over 1,700, with fewer routes occurring during January and June 2022. The lower number of routes in January was mainly due to the lower economic activity triggered by the COVID-19 pandemic, but its effect was not very significant.

Table 17 presents the summary of the indicators above mentioned per month.

Table 17 - CO2 Metrics

Month	kg CO_2/l	Kg CO_2/km	kg CO_2/r	kg CO_2/d
1	1.887	0.594	136.567	7652.181
2	1.887	0.585	132.477	10077.739
3	1.887	0.589	131.930	9298.972
4	1.887	0.583	133.535	10148.632
5	1.887	0.583	133.562	9918.042
6	1.887	0.590	139.842	9504.586

As shown in table Table 17, the kg CO_2/l indicator presents the same value for all months. This results from the fact that the database only had 1 vehicle not diesel. Regarding the remaining indicators, there was fluctuation, especially in the kg CO_2/d indicator, which determined, as previously mentioned, the total CO_2 emitted per day. This indicator varies depending on the number of routes taken per day, so the greater the number of routes taken, the higher the metric. For the Kg CO_2/km indicator, it had a small variation on the first 6 months of 2022, the indicator oscillates between 0.583 and 0.590. Finally, the kg CO_2/r indicator, which determines the emission per route, was between 131.930 and 139.842. This last indicator was higher the longer the distance traveled on the route.

In order to better visualize the evolution of the indicators, their distribution is presented in

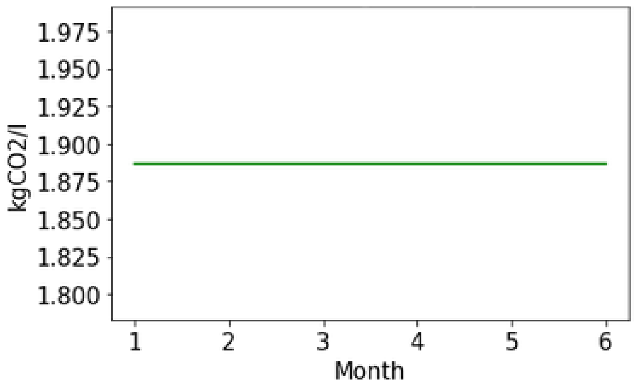

Figure 32 - Distribution of KgCO2/l per month

Figure 32 illustrates the evolution of the Kg CO_2/l metric per month. As the vehicle fleet was almost entirely composed by diesel vehicles, it was not possible to identify peaks in the level of CO_2 emission per liter.

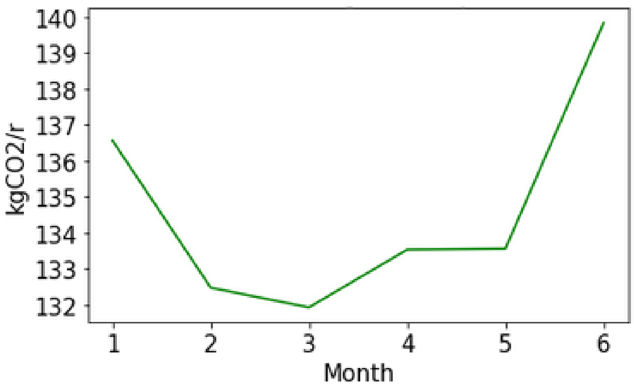

Figure 33 - Distribution of KgCO2/r per month

Figure 33 represents the evolution of CO_2 emissions per route during the first 6 months of 2022. The emission varied between about 132 Kg and 140 Kg, the C depending on the distances traveled and the number of routes.

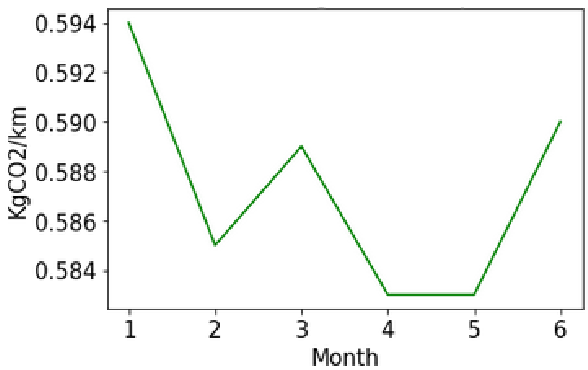

Figure 34 - Distribution of KgCO2/km per month

Figure 34 represents the CO_2 emission per Km, for the first 6 months of 2022. The range of emission varied between 0.583 and 0.594 KgCO_2/Km, showing a quite small metric variation. It was also possible to verify that January, which is the month with the lowest number of routes as shown in Figure 31, was the month with the highest metric. This factor is associated with other constraints, including driving performance, among other variables, not provided in our database. This indicator also allowed us to identify among other aspects, the need to promote better driving practices, to promote the reduction of CO_2 emissions per traveled Km.

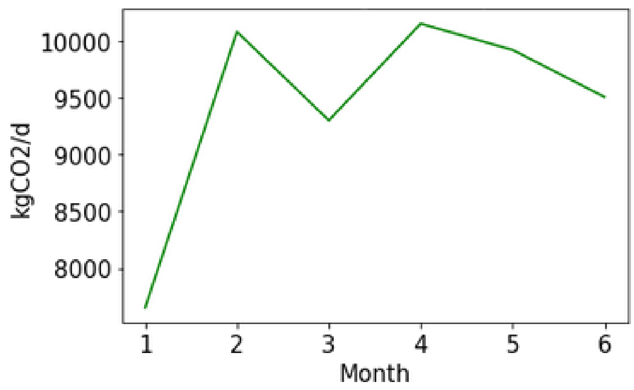

Figure 35 - Distribution of KgCO2/d per month

Figure 35 represents the CO_2 emission per day for the first 6 months of the year 2022. This indicator varies naturally depending on the routes developed as well as their distances. Thus, it can be seen that the month of January is the one that presents the lowest value for the metric under study, considering that January is the month that had the fewest routes, according to Figure 31.

After analyzing the evolution of the kg CO_2/l, Kg CO_2/Km, kg CO_2/r, kg CO_2/d metrics, shown in Figure 32, Figure 33, Figure 34 and Figure 35 it is possible to verify that the one that best supports the decision is the Kg CO_2/Km metric. This metric clearly shows the CO_2 emission per Km, which may vary naturally from several circumstances, including the way the driver drives, the vehicle's condition, the weight of the load carried, the existence of traffic, or even the road condition.

3.6. Evaluation

3.6.1. End-user evaluation

In order to evaluate our study, particularly with regard to the metrics defined and quantified for the purpose of monitoring the company's CO_2 emissions and consequent support for decision making, we conducted a questionnaire to the company representatives. However, considering scheduling restrictions, it was not made available to us in time, so we developed a self-assessment of the project results considering the feedback from the company.

The results obtained can be seen in Table 18.

Table 18 - Method assessment questionnaire

Criteria	Objective statement	Evaluation	Evaluator
Utility	It can help business decisions regarding the behaviour of the fleet and hub expansion?	Largely Achieved	Largely Achieved
Understandability	Provides understandable results.	Largely Achieved	Largely Achieved
Accessibility	Can be used without training	Largely Achieved	Partially Achieved
Level of detail	Provides knowledge from the mobility of the fleet and detailed location for expansion.	Partially Achieved	Partially Achieved
Consistency	Gives consistent results.	Largely Achieved	Largely Achieved

The development of the questionnaire follows the standards defined by the ISO/IEC TS 33061, primarily used to assess software development processes. Four levels of the NLPF were employed for evaluation:

- Not Achieved (NA) - [0-15%]
- Partially Achieved (PA) -]15-50%]
- Largely Achieved (LA) -]50-85%]
- Fully Achieved (FA) -]85-100%]

3.7. Deployment

The models created were not applied in a real production environment. Observations and results obtained were compiled in the thesis writing and summarized in a report presented in power point format to the CO. All software development was done in Python on a personal computer. The packages used were detailed in section 3.4. The reproducibility of the whole process can be accomplished by running the Jupyter Notebooks and auxiliary files provided along with the thesis. All the developed software material and datasets are available to the CO and for further academic research purposes. All software development was done in Python on a personal computer equipped with Windows 10 (64bits) operating system, Intel$^{(R)}$ Core$^{(TM)}$ i5-10TH GEN, i5-10210U CPU @ 1.60GHz, with 8Gb of memory ram. We adopted the python programming language on Jupyter Notebooks extension. The reproducibility of the whole process can be accomplished by running the Jupyter Notebooks and auxiliary files provided along with the thesis.

4. Conclusions

4.1. Discussion

In this thesis, we sought to develop data analysis that allowed us to answer our research questions:

RQ 1: How can we calculate the carbon footprint of a freight company?

RQ 2: What are the main factors of logistics transportation that significantly influence the carbon footprint of a freight company?

To answer RQ1, we performed SLR, detailed in section 2. Based on the SLR , we identified the preferred methodologies for calculating the carbon footprint, most of them referred by Wang [32]. The form identified is arithmetically simple, consisting in weighting the total liters of fuel consumed by the CO_2 conversion factor. The CO_2 conversion factor is specifically defined in the LG framework as detailed in Table 16.

Thus, after applying the formula suggested by the author, we calculated for each route the total Kg of CO_2 released into the atmosphere, associated with the distribution analysis that determined the company's carbon footprint. The distribution CO_2 emissions is presented in Figure 30Figure 30 - Distribution of total CO2 emissions (Kg)

To enable monitoring the CO_2 emission levels as well as timely decision making, we developed a set of metrics. These metrics were subjected to detailed analysis in section 3.5. It is important to highlight that, according to the interactions developed with the CO and which are described in section 3.6.1, the metric favored by the CO combines the CO_2 emissions per transported ton. According to the CO, this metric allowed combining the evaluation of business evolution on a financial basis and, simultaneously, monitoring CO_2 emissions.

However, the information on transported cargo was not available in the database, which is why this metric was not calculated. Because of this, we have calculated the following metrics:

- Fleet emissions per liter of fuel consumed (Figure 32), which represents the total Kg of CO_2 emitted per liter of fuel
- Fleet emissions per Km traveled (Figure 34), which represents the total Kg of CO_2 emitted per km traveled

Additionally, although not a metric identified in the Action Plan of the company, we will also analyze the following metrics:

- Fleet emissions per route (Figure 33), which represents the total kg of CO_2 emitted per route

- Fleet emissions per day (Figure 35), which represents the total kg CO_2 emitted per day

During the development of our project, we carried out several interactions with the CO in order to obtain an alignment between the development work and the intended objectives. Among the interactions developed, we highlight the physical visit to the company, where we met the Director of Operations, IT Director and one of the Directors of the company and that allowed us to collect information and knowledge at the company level, namely also with regard to the identification of the best indicators for monitoring the evolution of the carbon footprint.

From the analysis on the metrics, as well as from the CO’s comments on the metrics results, we concluded the metric that best monitored CO_2 emissions was the Kg CO_2/Km metric, which allows quantifying the impact in terms of carbon emissions per Km. However, this metric was influenced by other data factors that were not considered in our database, such as: consumer behavior; supply chain management and logistics; new vehicle fuels; strategic transport planning.

Regarding the RQ2, we sought through the CRISP-DM methodology [32] to develop data analysis, to define the main factors that contribute to the CO's carbon footprint.

The data referred to the first six months of 2022 (between January and June), for a set of 133 vehicles, with about 13 thousand routes. In terms of patterns, it was possible to notice that the month of January presents a lower number of routes than the other months, as shown in Figure 31. This aspect, according to an interaction with the CO, was associated with business characteristics, as January was a typically less active month in logistics activity. However, the other months showed an average distribution of reduced variations. Regarding the activity throughout the week, we concluded that there was a lower number of routes on weekend, especially on Sunday, which presents an average of about 2 routes per day, as shown in more detail in Table 12.

It is also important to emphasize that no relevant patterns have been identified for the purpose of predictive analyses of data evolution, so no exploratory analyses were developed in this sense.

Additionally, in terms of data cleaning, we came across outliers, which were removed from the database. These referred to possible errors, and given their reduced expressiveness, do not condition or impact the conclusions of this research.

As far as the correlation analysis is concerned, and as shown in the correlation matrix presented in Figure 24, it was easy to perceive the following aspects: the total amount of liters of fuel was higher as the higher the consumption levels of the vehicle, the distance traveled, the speeds practiced, and the time of the trip. All these characteristics resulted from the exploratory analyses and were subjected to interpretation by the company and were positively validated.

As a result of the conclusions and the interactions with the CO, we identified limitations to the work as well as proposals for the development of future work, as analyzed in Section 4.2 and 4.3.

It was also proposed, within the scope of the final meeting with the company, and resulting from the presentation of conclusions, the survey assesses the alignment of the project against the identified objectives. However, considering scheduling restrictions, it was not made available to us in time, so we developed a self-assessment of the project results considering the feedback from the company, as described in more detail in section 3.6.1.

4.2. Research limitations

This research work was the first carried on CO's data, only existing other types of analysis developed by the CO for the purpose of participation in the LG program.

The research work was based specifically on the analysis of data for the first six months of the 2022, thus making it impossible to identify temporal patterns, or to identify other type of characteristic that could be highlighted if the database analyzed had a larger temporal spectrum.

Also, the database did not present relevant information on other characteristics intrinsic to the logistics activity, namely the transported cargo which, as defined in CO's Action Plan, is induced as the main metric for the management of CO_2 emissions and consequent monitoring of the goal set by the company, which is to reduce 20% of CO_2 emissions within five years. Furthermore, there were other characteristics that can influence the CO_2 emission in the scope of logistics activity, associated with the drivers' driving behavior, as evidenced in the study of Zhao [12] and which are not included in our database.

4.3. Future work

As a result of our thesis research, as well as the limitations listed in section 4.2, the following are aspects should be considered in future research work:

- Expand the number of observations analyzed to detect long-term trends and produce more insightful results;
- Include in the database other relevant information, namely the weight of cargo transported;
- Reflect in the data analysis other characteristics inherent to the transport activity, namely the drivers' driving behavior;
- Analyze other sources of CO_2 emissions associated with the logistics activity, namely vehicle wear and tear.

5. References

[1] J. Delbeke, A. Runge-Metzger, Y. Slingenberg, and J. Werksman, “The paris agreement,” *Towar. a Clim. Eur. Curbing Trend*, pp. 24–45, 2019, doi: 10.4324/9789276082569-2.

[2] “GS1.” https://gs1pt.org/ (accessed May 17, 2022).

[3] D. Sengupta, A. Das, and U. K. Bera, “A gamma type-2 defuzzification method for solving a solid transportation problem considering carbon emission,” *Appl. Intell.*, vol. 48, no. 11, pp. 3995–4022, 2018, doi: 10.1007/s10489-018-1173-7.

[4] 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 health care interventions: Explanation and elaboration,” *PLoS Med.*, vol. 6, no. 7, 2009, doi: 10.1371/journal.pmed.1000100.

[5] “VOSviewer.” https://www.vosviewer.com// (accessed Mar. 15, 2022).

[6] “Welcome to Python.” https://www.python.org/ (accessed Mar. 15, 2022).

[7] “Scimago Journal & Country Rank.” https://www.scimagojr.com/ (accessed Mar. 15, 2022).

[8] M. Umar, X. Ji, D. Kirikkaleli, and Q. Xu, “COP21 Roadmap: Do innovation, financial development, and transportation infrastructure matter for environmental sustainability in China?,” *J. Environ. Manage.*, vol. 271, no. July, p. 111026, 2020, doi: 10.1016/j.jenvman.2020.111026.

[9] A. García-Olivares, J. Solé, and O. Osychenko, “Transportation in a 100% renewable energy system,” *Energy Convers. Manag.*, vol. 158, no. January, pp. 266–285, 2018, doi: 10.1016/j.enconman.2017.12.053.

[10] F. Leach, G. Kalghatgi, R. Stone, and P. Miles, “The scope for improving the efficiency and environmental impact of internal combustion engines,” *Transp. Eng.*, vol. 1, no. April, 2020, doi: 10.1016/j.treng.2020.100005.

[11] C. F. A. Cunha, J. de Oliveira Gomes, and H. M. B. de Carvalho, “A new approach to reduce the carbon footprint in resistance spot welding by energy efficiency evaluation,” *Int. J. Adv. Manuf. Technol.*, vol. 119, no. 9–10, pp. 6503–6520, 2022, doi: 10.1007/s00170-021-08472-7.

[12] X. Zhao, Y. Ke, J. Zuo, W. Xiong, and P. Wu, “Evaluation of sustainable transport research in 2000–2019,” *J. Clean. Prod.*, vol. 256, p. 120404, 2020, doi: 10.1016/j.jclepro.2020.120404.

[13] X. Li, Y. Ai, Y. Ge, J. Qi, Q. Feng, J. Hu, W. C. Porter, Y. Miao, H. Mao, and T. Jin, “Integrated effects of SCR, velocity, and air-fuel ration on gaseous pollutants and CO2 emissions from China V and VI heavy-duty diesel vehicles,” *Sci. Total Environ.*, vol. 811, pp. 1–8, 2022, doi: 10.1016/j.scitotenv.2021.152311.

[14] Y. Li, H. K. Dong, and S. Lu, “Research on application of a hybrid heuristic algorithm in

transportation carbon emission," *Environ. Sci. Pollut. Res.*, vol. 28, no. 35, pp. 48610–48627, 2021, doi: 10.1007/s11356-021-14079-y.

[15] Y. Fan, C. de Kleuver, S. de Leeuw, and B. Behdani, "Trading off cost, emission, and quality in cold chain design: A simulation approach," *Comput. Ind. Eng.*, vol. 158, p. 107442, 2021, doi: 10.1016/j.cie.2021.107442.

[16] C. Ternel, A. Bouter, and J. Melgar, "Life cycle assessment of mid-range passenger cars powered by liquid and gaseous biofuels: Comparison with greenhouse gas emissions of electric vehicles and forecast to 2030," *Transp. Res. Part D Transp. Environ.*, vol. 97, no. May, p. 102897, 2021, doi: 10.1016/j.trd.2021.102897.

[17] J. L. Osorio-Tejada, E. Llera-Sastresa, and A. H. Hashim, "Well-to-wheels approach for the environmental impact assessment of road freight services," *Sustain.*, vol. 10, no. 12, pp. 1–27, 2018, doi: 10.3390/su10124487.

[18] E. Mangina, P. K. Narasimhan, M. Saffari, and I. Vlachos, *Optimising Supply Chain Logistics System Using Data Analytics Techniques*, vol. 310 LNICST. Springer International Publishing, 2020. doi: 10.1007/978-3-030-38822-5_6.

[19] E. Mangina, P. K. Narasimhan, M. Saffari, and I. Vlachos, "Data analytics for sustainable global supply chains," *J. Clean. Prod.*, vol. 255, p. 120300, 2020, doi: 10.1016/j.jclepro.2020.120300.

[20] D. J. Alves, "Driver profile classification on heavy trucks based on telemetry," 2019.

[21] T. Palander, H. Haavikko, and K. Kärhä, "Towards sustainable wood procurement in forest industry – The energy efficiency of larger and heavier vehicles in Finland," *Renew. Sustain. Energy Rev.*, vol. 96, no. July, pp. 100–118, 2018, doi: 10.1016/j.rser.2018.07.043.

[22] E. Y. C. Wong, A. H. Tai, and E. Zhou, "Optimising truckload operations in third-party logistics: A carbon footprint perspective in volatile supply chain," *Transp. Res. Part D Transp. Environ.*, vol. 63, no. July 2018, pp. 649–661, 2018, doi: 10.1016/j.trd.2018.06.009.

[23] D. N. Schmitz Gonçalves, G. V. Goes, M. de Almeida D'Agosto, and R. Albergaria de Mello Bandeira, "Energy use and emissions scenarios for transport to gauge progress toward national commitments," *Energy Policy*, vol. 135, no. July, p. 110997, 2019, doi: 10.1016/j.enpol.2019.110997.

[24] C. A. de Armiño, D. Urda, R. Alcalde, S. García, and Á. Herrero, "An Intelligent Visualisation Tool to Analyse the Sustainability of Road Transportation," *Sustain.*, vol. 14, no. 2, 2022, doi: 10.3390/su14020777.

[25] N. R. Baral, Z. D. Asher, D. Trinko, E. Sproul, C. Quiroz-Arita, J. C. Quinn, and T. H. Bradley, "Biomass feedstock transport using fuel cell and battery electric trucks improves lifecycle metrics of biofuel sustainability and economy," *J. Clean. Prod.*, vol. 279, p. 123593, 2021, doi: 10.1016/j.jclepro.2020.123593.

[26] C. Magazzino, M. Mele, and N. Schneider, “A new artificial neural networks algorithm to analyze the nexus among logistics performance, energy demand, and environmental degradation,” *Struct. Chang. Econ. Dyn.*, vol. 60, pp. 315–328, 2022, doi: 10.1016/j.strueco.2021.11.018.

[27] L. Wen and Q. Song, “Simulation study on carbon emission of China’s freight system under the target of carbon peaking,” *Sci. Total Environ.*, vol. 812, p. 152600, 2022, doi: 10.1016/j.scitotenv.2021.152600.

[28] T. Gorges and C. Holz-Rau, “Transition of mobility in companies – A semi-systematic literature review and bibliographic analysis on corporate mobility and its management,” *Transp. Res. Interdiscip. Perspect.*, vol. 11, p. 100462, 2021, doi: 10.1016/j.trip.2021.100462.

[29] S. Wang, W. Wang, and H. Yang, “Comparison of product carbon footprint protocols: Case study on medium-density fiberboard in China,” *Int. J. Environ. Res. Public Health*, vol. 15, no. 10, pp. 1–14, 2018, doi: 10.3390/ijerph15102060.

[30] N. Roukounakis, E. Valkouma, E. Giama, and E. Gerasopoulos, “The development of a carbon footprint model for the calculation of GHG emissions from highways: the case of Egnatia Odos in Greece,” *Int. J. Sustain. Transp.*, vol. 14, no. 1, pp. 74–83, 2020, doi: 10.1080/15568318.2018.1523509.

[31] Y. Dong, J. Xu, and C. Gu, “Modelling carbon emissions of diesel trucks on longitudinal slope sections in China,” *PLoS One*, vol. 15, no. 6, pp. 1–17, 2020, doi: 10.1371/journal.pone.0234789.

[32] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[33] H. Liu, R. Yang, Y. Wang, and Q. Zhu, “Measuring performance of road transportation industry in China in terms of integrated environmental efficiency in view of Streaming Data,” *Sci. Total Environ.*, vol. 727, p. 138675, 2020, doi: 10.1016/j.scitotenv.2020.138675.