NOVA
IMS
Information
Management
School

MGI

Master Degree Program in
Information Management

Strategic Selection of Charging Stations for Electric Vehicles

A Data-Driven Approach in Lisbon District

Yolanda Carvalho Ferretti

Master Thesis

presented as partial requirement for obtaining the Master Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Strategic Selection of Charging Stations for Electric Vehicles

A Data-Driven Approach in Lisbon District

by

Yolanda Carvalho Ferretti

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence.

Supervised by

Bruno Jardim, PhD, NOVA Information Management School

July 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, July 2024]

ABSTRACT

The rapid growth of electric vehicles (EVs) presents new challenges and opportunities for urban infrastructure, particularly in the development of a robust charging network. This thesis explores the current landscape of EV charging facilities in the Lisbon district of Portugal and proposes a strategic framework for their expansion. Utilizing data analytics methods, including K-means clustering, the study identifies high-demand areas lacking sufficient charging points and suggests optimal locations for new installations. Key findings underscore the necessity of an integrated approach involving policymakers, automakers, and community stakeholders. It is also highlighting the beneficial impact of data-driven planning on the successful integration of EVs into urban transport systems. The proposed solutions aim to not only fill the gaps in the existing infrastructure but also to stimulate the transition to eco-friendly mobility, with Lisbon serving as a case study for cities worldwide.

KEYWORDS

Electric vehicles; charging station; charging infrastructure; data analytics; K-means clustering

SUSTAINABLE DEVELOPMENT GOALS (SDG):

7 AFFORDABLE AND CLEAN ENERGY
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
11 SUSTAINABLE CITIES AND COMMUNITIES
13 CLIMATE ACTION

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
BEV	Battery-electric vehicles
BI	Business intelligence
CP	Charging Points
CS	Charging Stations
CVs	Conventional vehicles
EVCSs	Electric vehicles charging stations
EVs	Electric vehicles
GHG	Greenhouse gas
PHEVs	Plug-in hybrids vehicles

1. INTRODUCTION

1.1. MOTIVATION

As the environmental impacts of traditional fossil fuel-based vehicles become increasingly apparent, particularly regarding carbon dioxide (CO2) emissions, the urgency to transition to eco-friendly alternatives has never been greater (European Environment Agency, 2022). Electric vehicles (EVs) offer a compelling solution, releasing zero pollutants into the air and contributing to a cleaner environment (Hofer et al., 2018).

The move to electric vehicles is slowed down by a big issue—there aren't enough charging stations, and they're not spread out well. This study aims to tackle this key problem by looking at the Lisbon district in Portugal, with the goal of creating plans to improve how easy it is to reach and how widespread the EV charging network is. Using data analysis, this research tries to find the best spots to set up new charging stations. This is an important move to help more people choose electric vehicles, make transportation more sustainable, and keep the environment clean.

1.2. OBJECTIVES AND METHODOLOGY

Considering the pressing demand for sustainable transportation solutions, the objectives of this study are two: to conduct a thorough investigation into the current state of EV charging infrastructure in Portugal and to propose actionable strategies to expand this network effectively. This study identifies the scarcity of accessible charging points as a critical barrier to EV adoption and leverages data analytics to overcome these obstacles.

The central research questions guiding this investigation are:

- Where is the demand for new charging points most pronounced within the Municipalities of Lisbon?
- What are the strategic locations to install new charging stations?

To answer these questions, the research methodology incorporates an analysis of two datasets that, when combined, illuminate the areas within Lisbon's Municipalities that are underserved by the current charging station infrastructure. Subsequently, the study employs two additional datasets to identify specific locations with a high demand for charging points, providing valuable insights for strategic network expansion.

In summary, this study aims to bridge the gap between the environmental promise of electric vehicles and the practical challenges of developing a sufficient charging infrastructure. It explains the motivation for the research, goes deep into the specific objectives related to the expansion of the charging network, and details the methodology that combines environmental considerations with data driven practices. This is to support sustainable mobility and contribute to a healthier environment in Portugal.

The present project is organized in six different chapters. The Introduction chapter describes my motivations and brings the research questions of this project. On Chapter 2 is presented the Literature Review that outlines literatures that explain the urgency of choosing greener alternatives as EVs and discusses related works that implemented data analysis on this matter. Chapter 3 defines the Methodology used to identify the optimal locations for installing electric vehicle charging points in the municipalities of Lisbon. And includes data collection, understanding, processing and analysis developed on a Jupyter notebook with Python language. The Chapter 4 cover the Results and Discussions achieved with the project. Chapter 5 is addressed to the Conclusion, where the work it is summarized. And finally, chapter 6 explores the Limitations and Future Research that considers limitations found on the work and gives insight to future projects.

2. LITERATURE REVIEW

This chapter will explore significant works that provide insights into the optimal infrastructure for electric vehicle (EV) charging. Section 2.1 sets the context for greenhouse gas emissions and the need for sustainable choices, introducing the transition to EVs as a key strategy for achieving emission reduction targets. It will also briefly touch upon the development of charging infrastructure and the scenario in Portugal. In section 2.2, it will delve deeper into the analysis of previous studies that investigate the ideal network of chargers for EVs, examining works ranging from identifying points of interest to analyzing user preferences and their impact on EV adoption. Finally, in subsection 2.2.1, it will be discussed research that applies advanced optimization methods, such as genetic algorithms and clustering techniques, to determine the ideal location and capacity of charging points. These studies highlight the importance of a multifactorial approach in optimizing EV infrastructure to meet the needs of a society transitioning to electric mobility.

2.1. CONTEXTUALIZATION

2.1.1. Greenhouse gas emissions and the urge for sustainable choices

Fossil fuel combustion stands out as a significant contributor to the issue of urban air pollution, as elucidated by Mayr and Rentschler (2023). Recognizing the pressing need to address climate change, the European Green Deal was instituted with the ambitious aim of positioning Europe as the first climate-neutral continent. This landmark agreement aspires to achieve net-zero greenhouse gas (GHG) emissions by 2050, as outlined by the European Environment Agency (2022).

While the European Green Deal represents a crucial step towards climate neutrality, it is evident that the process of decarbonization demands a multifaceted approach, transcending a singular reliance on carbon pricing. Skjærseth's work (2021) highlights the need to examine the complex landscape of climate and energy policies within the European Union to effectively achieve these ambitious climate targets.

A key aspect of the broader decarbonization strategy involves reducing GHG emissions from the transport sector. Notably, Figure 1 illustrates that cars are mainly responsible for GHG emissions. Despite this, there is some progress, with car transport showing a commendable reduction of 15.4% between 2019 and 2020 (European Environment Agency, 2022).

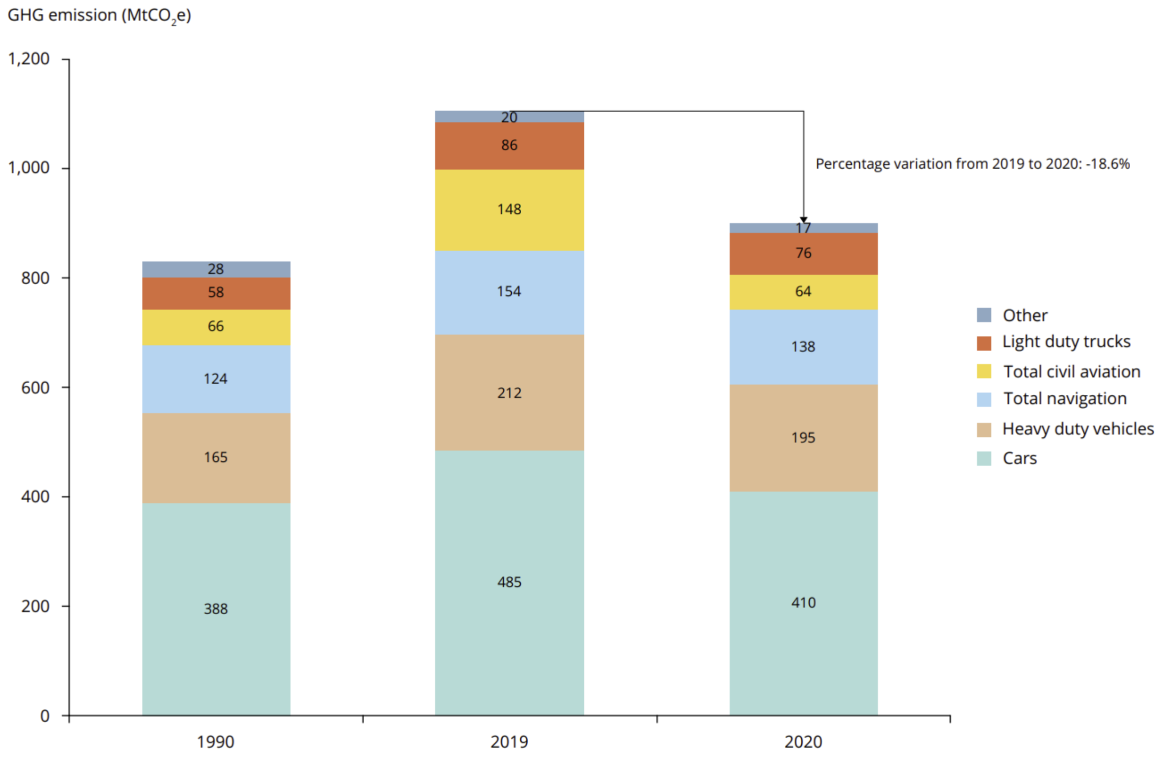


Figure 1 Greenhouse gas emissions from transports. Source: European Environment Agency (2022)

To continue this positive trend, the transition from traditional fossil fuel-driven cars to electric vehicles holds substantial promise. Such a shift can significantly impact the decline in CO2 emissions, thereby contributing to the overarching goal of achieving net-zero greenhouse gas emissions (Hofer et al., 2018).

The main technologies for light-duty electric vehicles (EVs) include hybrid electric vehicles (HEVs), plug-in electric vehicles (PHEVs), and battery electric vehicles (BEVs). HEVs combine an electric motor and an internal combustion engine for propulsion. It is important to note that the HEV battery cannot connect to the external power grid; instead, it is charged by the internal combustion engine (ICE) or through regenerative braking. PHEVs are like HEVs but have the additional capability of connecting the battery to the power grid for recharging. In contrast, BEVs exclusively use an electric motor for propulsion and can connect to the power grid for charging (Li et al., 2019).

While some car companies express skepticism about the feasibility of replacing CVs with EVs due to potential impacts on fleet performance, Romero-Ocaño et al. (2022) demonstrate that the replacement of CVs with EVs, without compromising fleet performance, is achievable, ranging between half and two-thirds of the total vehicle count.

Illustrated in Figure 2, the increasing demand for battery-electric vehicles (BEVs) over the past years is denoted by the green color. Additionally, forecasts for this category predict a substantial increase, with a prediction of 17 million light vehicle BEVs by 2035 (King, 2023).

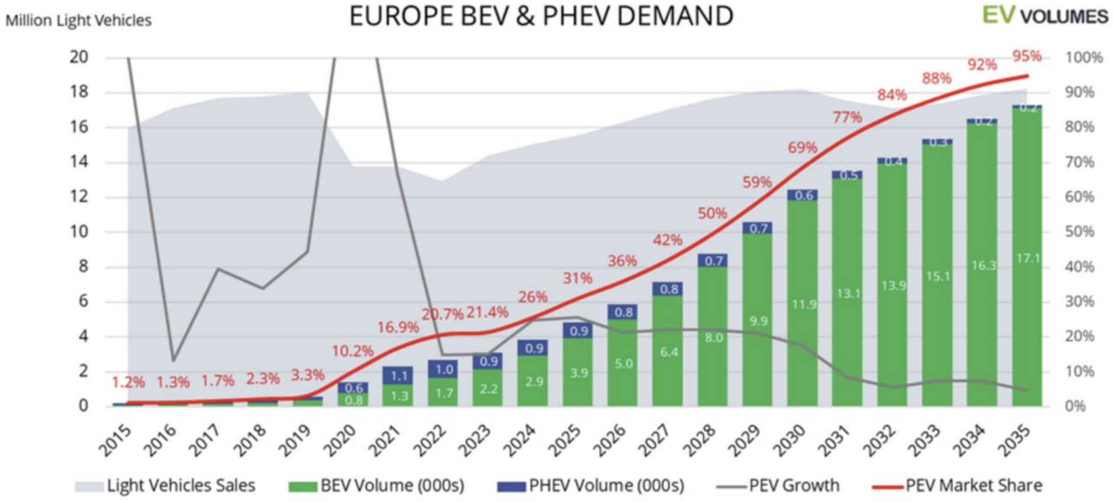


Figure 2 BEV Forecast. Source: King (2023)

This information highlights the trajectory of companies expanding their EV production, emphasizing the need for urban infrastructure to be adequately equipped to meet the increasing demand for charging. As the adoption of EVs continues to surge, urban planning and charging infrastructure development must be prioritized to facilitate a seamless transition towards sustainable transportation.

2.1.2. Charging Points for Electric Vehicles and Portugal Scenario

EV owners can decide whether to charge their vehicles publicly or privately. Jeczawitz (2021) explain that unlike traditional refueling approaches, EVs offer the flexibility to be charged at different places, including homes, parking lots, workplaces, or along highways. Statistical data emphasizes the escalating need for public charging stations, especially during extended travels.

However, the growing adoption of EVs and demand of charging points presents difficulties for the power grid, including voltage drops caused by the substantial charging load. Furthermore, improper positioning of EV charging stations (EVCSs) worsens these challenges, resulting in a clustering of charging demand on a limited number of stations (Woo et al., 2023).

The graph presented in Figure 3, sourced from E-redes (2024), illustrates the evolution of the installed connection points during the initial three quarters of 2023. Notably, there is a discernible upward trajectory in the number of these connection points with each passing period, indicating a consistent growth trend. This progression is anticipated to persist and even intensify in the forthcoming years, reflecting the expanding infrastructure to support.

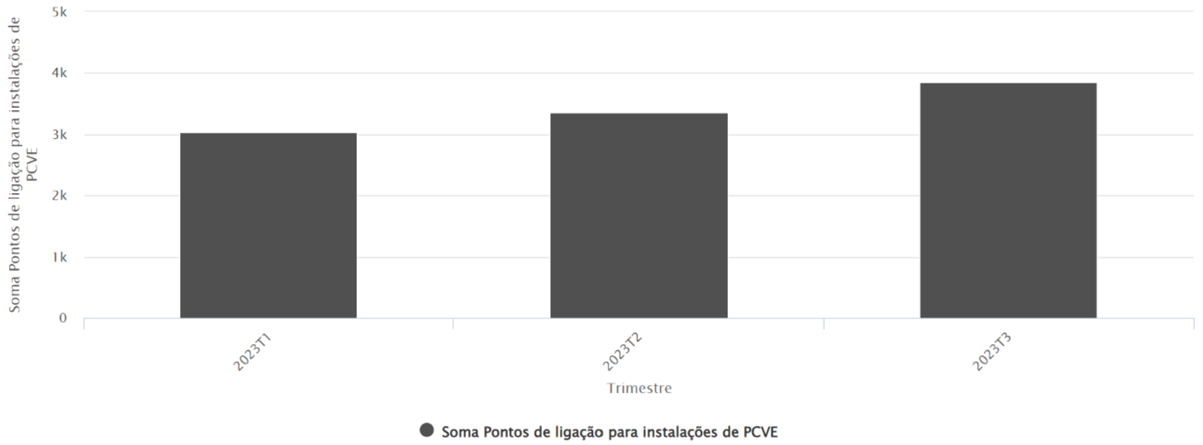


Figure 3 Electrical Vehicle Charging Points in Portugal. Source: E-redes (2024)

However, the study by Miranda & Delgado (2020) introduces an essential perspective, asserting that the decision to purchase an electric car in Portugal is significantly influenced by the accessibility of convenient charging locations. Recognizing this critical factor, it becomes crucial to strategically allocate charging points to ensure they are positioned in areas that meet the evolving needs of EV users and consequently, making electric vehicles more appealing to consumers in Portugal.

2.2. RELATED WORKS

Some studies have been conducted to understand the optimal network of electric vehicle (EV) chargers around the world. As early as 2013, academic work was already focused on understanding the optimal infrastructure for EV charging stations. In the cities of Amsterdam and Brussels, Wagner et al. (2013) worked on identifying points of interest to define optimal CP locations by using penalties to identify with more accuracy the points of interest.

One year later, Eisel et al. (2014) were analyzing how the EV adoption rates could increase, and they understood that to positively impact EV adoption rates the charging infrastructure should be influenced especially by user preferences. Also, EV Connect (2020) add that it should place CP close to local shops and restaurants so that people can charge their cars while getting groceries or eating dinner. One of the indirect benefits of public EV charging is contribution to the local economy. For instance, people tend to spend money on nearby shops when waiting on charging their EVs, adding value for the venues nearby.

Besides user preferences, are also many other factors that can may influence where is located a CP like economic problems of operators, power loss of vehicles, traffic congestion of transportation system and safety of power grid (Kong et al., 2019). Melissa (2024) adds the population factor, and states that on average, there were 106 public charging points for every 100,000 inhabitants in the EU in 2022.

2.2.1. Data Analysis Related Works

Applying their analysis on the city of Thessaloniki, Greece, Efthymiou et al., (2017) already started to use Genetic Algorithm (GA) to understand the best locations to install new CP. They explain GA is an artificial optimization algorithm that can be successfully applied on different optimization problems. Their research determinate the demand locations and was developed in R to being user friendly.

In the same year, Catalbas et al. (2017) used spectral clustering and Gaussian Mixture Model (GMM) to estimate the optimal locations for EV charging stations in Ankara, Turkey by using data mining methods. On the clustering part, they utilized the Spectral clustering, and they used three clustering objective functions: Minmax Cut, Ratio Cut, and the Normalized Cut as is possible to see on Figure 4.

$$J = \sum\nolimits_{1 \leqslant p < q \leqslant K} \frac{s(C_p, C_q)}{\rho(C_p)} + \frac{s(C_p, C_q)}{\rho(C_q)} = \sum\nolimits_{k=1}^{K} \frac{s(C_k, \overline{C}_k)}{\rho(C_k)}$$

$$\rho(C_k) = \begin{cases} |C_k| & Ratio\ Cut \\ \sum_{i \in \boldsymbol{C}_k} d_i & Normalized\ Cut \\ S(C_K, C_k) & Minmax\ Cut \end{cases}$$

Figure 4 Spectral clustering. Source: Catalbas et al. (2017)

Csiszár et al. (2020) brings a weighted multicriteria location optimization method with ranking and selection as we see on Figure 5. Where is possible to taking account several variables and parameters as it is used on their work. Which variables were traffic volume on adjacent roads, total population of nearby settlements, service level of nearby points of interest (e.g. supermarket, restaurant), effects of existing or installed fast-charging stations.

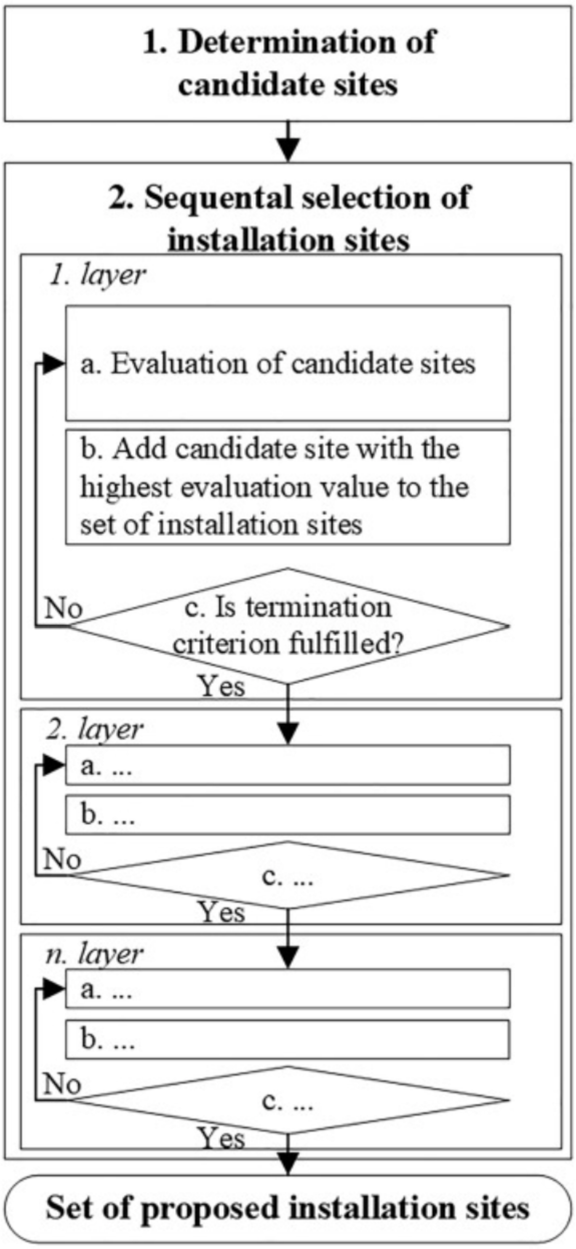


Figure 5 Weighted multicriteria location optimization method

Another project that uses GA to determine the best location and scale of the charging station were conducted by Zhou et al. (2022). However, on their project it is consider many parameters that brings more accuracy to the work.

Such parameters are: whether to build a charging station at the position of j, vehicle i chooses to go to charging station j for charging service, the number of chargers to be built in site j, the investment function of fixed cost, the annual, operating cost of site j, the discount rate, the depreciation period of charging stations, the, power consumption per unit distance to charging stations, the number of cars requiring, charging every day at point I, non-linear coefficient about roads, the space linear distance, between the point i to site j, the average electricity cost of an electric car currently, carbon emission, greenhouse gas (GHG) factor, the power consumption by a car running 1 km, the, charging efficiency, the probability of charging

per car per day, charging station j can provide, the maximum charging capacity and the maximum distance an electric car can travel.
On the other hand, Wang et al. (2022) bring a different approach. To define the location and capacity of Electrical Vehicle Charging Stations (EVCSs), they use a clustering algorithm, and to be more specific, the K-means algorithm. This algorithm is defined with four steps, the first one is having K randomly selected pieces of data as the initial center. The second step is calculating the distance between from non-center points to the clustering center with the Euclidean distance (formula below) to get the nearest centers that will form one group. The third step is calculating the new clustering center with the average distance from non-clustering centers to the clustering center in each group. And finally, the last step is repeat the steps 2 and 3 until the clustering center no longer changes or the maximum number of iterations is reached (J. Zhou et al., 2023).

$$dist(X_i, Y_j) = \sqrt{\sum_{k=1}^{n} (X_{ik} - Y_{jk})^2}$$

In synthesis, the literature review highlights significant insights into the optimal infrastructure for EV charging, emphasizing the importance of a multifactorial approach. It sets the context by discussing the urgent need for sustainable choices to reduce greenhouse gas emissions, focusing on the transition to electric vehicles (EVs) as a pivotal strategy. The review covers studies that identify key points for charger placement, analyze user preferences, and utilize advanced optimization methods like clustering techniques to determine ideal locations and capacities for charging points.

3. METHODOLOGY

In this chapter, it is presented the methodology used to identify the optimal locations for installing electric vehicle charging points in the municipalities of Lisbon. And it was followed the CRISP-DM methodology, or Cross-Industry Standard Process for Data Mining. That is a comprehensive methodology for conducting data science projects. It provides a step-by-step guide to efficiently and repeatably extract knowledge from data.

Furthermore, the process involves comparing databases, using APIs to extract coordinates and relevant information, and applying clustering technique to define the optimal installation points. Also, the project will be developed on a Jupyter notebook with Python language. The project logic 5 phases, the first consists of the collection and understanding of four different data sources, explored on session 3.1, the second is understanding the demand for charging points, as analyzed on session 3.2.1, the third and fourth understand candidate sites and defines sites on the session 3.2.2 and finally the recommended sites for installing new CP will be explored on chapter 4. The project logic is presented visually on the swimlane diagram below.

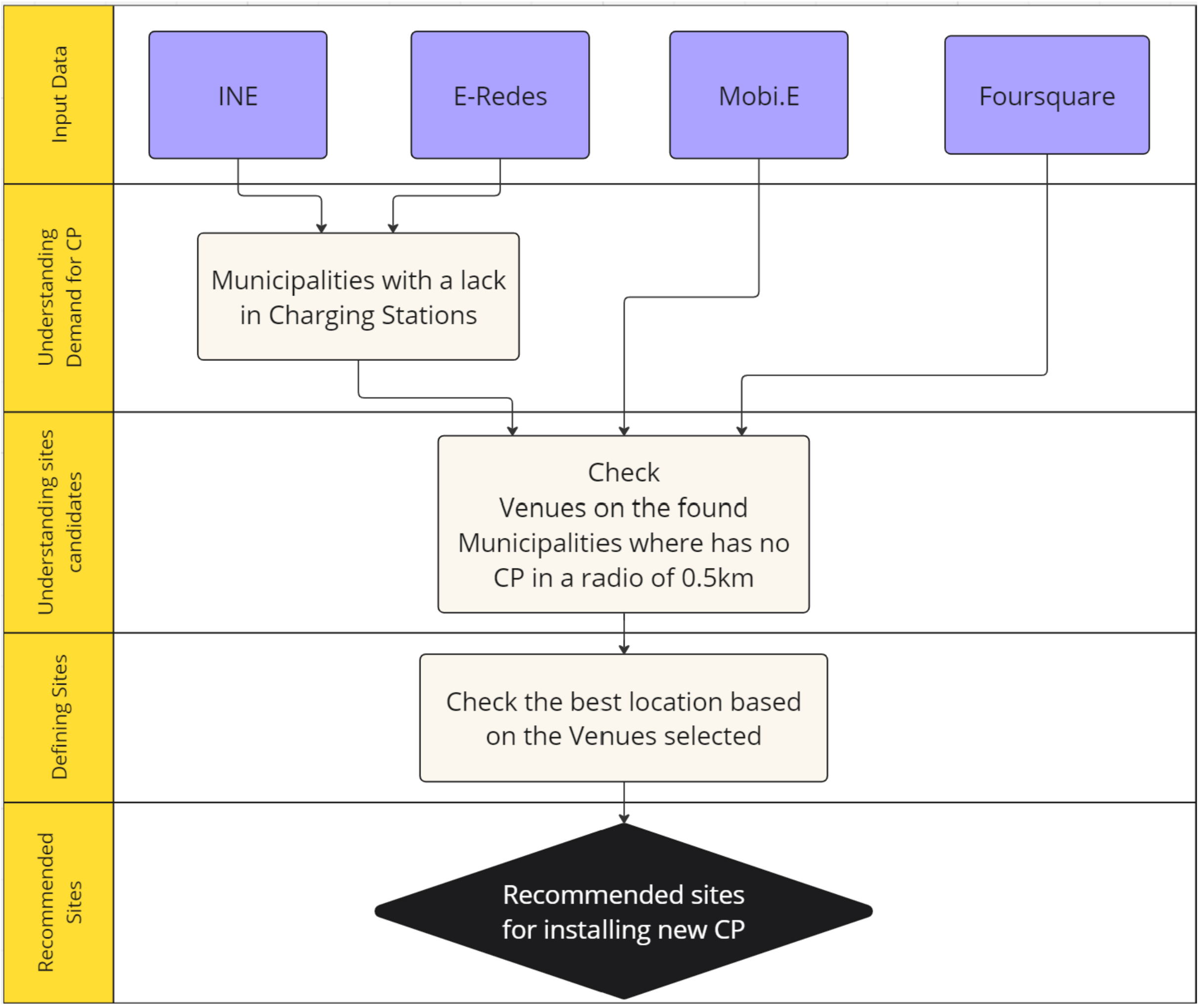


Figure 6 Project Logic

3.1. DATA COLLECTION AND UNDERSTANDING

On this session it will be explore how the data were collected and its purpose on the project. This project is supported by four datasets. The first two, from E-redes and Statistics Institute of Portugal, are responsible for bring demographic demand insights. While the last two, from Mobi.E and Foursquare, bring the specific information where new charging stations should be located.

The first part of the project consists of identify the Municipalities that has a lack of charging points. To achieve this goal, I bring two datasets, the *eredes* dataset from E-redes, that shares information about how many charging point already exist on each Municipality of Lisbon. And the *lisbon_population* dataset from Statistics Institute of Portugal, that share information about the total population of each Lisbon Municipality.

The second part of the project consists of comprehend where exactly the new charging point should be located. For this purpose, I bring the last two datasets, from Mobi.E, that brings the precisely location of all existent CP we already have on the district of Lisbon. And the dataset extracted from Foursquare, that share relevant venues data from a specific area and will be extremely important to define new CP installations.

3.1.1. E-Redes

E-redes is a company responsible for managing the National Electricity Transmission Grid (RNT) in Portugal. And its dataset was collected from E-redes website in csv format. To be possible to read the data as a DataFrame I imported pandas, an open-source library for data analysis and manipulation in Python. The data is composed by six columns and 7842 records. These records present information about the quarter when the data was collected and where each CP is located, in terms of District, Municipality and Parish. Also is provided the quantity of CP and Maximum Allowable Power(kW) allocated for each specific location as is possible to identify on the dataset schema (table 1).

Identifier	Description	Type	Sample
Trimestre	Quarter data was extracted	Text	2023T3
Distrito	District	Text	SETUBAL
Concelho	Municipality	Text	BARREIRO
Freguesia	Parish	Text	UF SEIXALINHO S ANDRE VERDEANA
Potência Máxima Admissível (kW)	Maximum Rated Power (kW)	Double	41.4
Pontos de ligação para instalações de PCVE	Electric Vehicle (EV) Charging Station Connection Points	Int	39

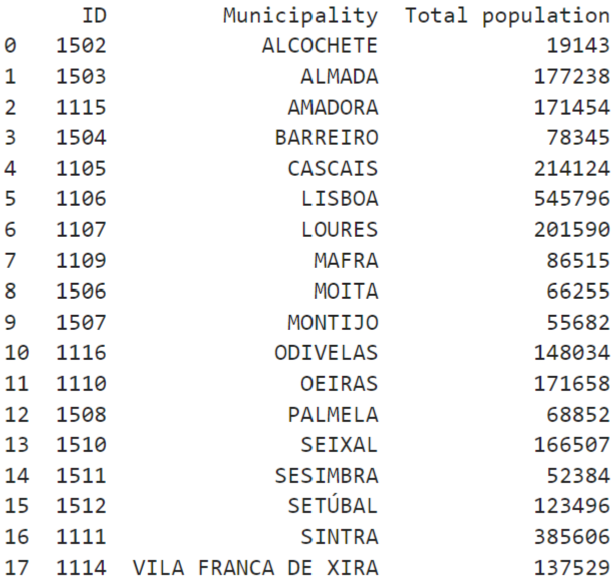
Table 1 Dataset schema

However, for this project it will be used only data from Lisbon District since is the area defined for the analysis. Therefore, I extracted only the necessary data, that means where District is equal Lisbon, and then I group by municipalities to have how many charging stations exist per municipality.

This data will be used to understand how many charging point already exist in each Municipality of Lisbon. And finally, comparing with *lisbon_population* dataset, will reveal potential gaps or oversaturation areas of charging points.

3.1.2. Statistics Institute of Portugal

The Instituto Nacional de Estatística (INE) is the national statistical office of Portugal. It is responsible for the production and dissemination of official statistics related to the country's economy, population, and society. The site has the possibility of filtering the specific regions you are looking for. That way was possible extract the *lisbon_population* dataset as I needed to use on this project as it is possible to see on Figure 7. Filtering by municipality and getting the total population of each one. And to visualize on a DataFrame, it was used panda's library as it was used with E-redes data.



```
      ID        Municipality  Total population
0   1502           ALCOCHETE             19143
1   1503              ALMADA            177238
2   1115             AMADORA            171454
3   1504            BARREIRO             78345
4   1105             CASCAIS            214124
5   1106              LISBOA            545796
6   1107              LOURES            201590
7   1109               MAFRA             86515
8   1506               MOITA             66255
9   1507             MONTIJO             55682
10  1116            ODIVELAS            148034
11  1110              OEIRAS            171658
12  1508             PALMELA             68852
13  1510              SEIXAL            166507
14  1511            SESIMBRA             52384
15  1512             SETÚBAL            123496
16  1111              SINTRA            385606
17  1114  VILA FRANCA DE XIRA            137529
```

Figure 7 Dataset lisbon_population

3.1.3. Mobi.E

Mobi.E is a company that acts as the managing entity of the electric mobility network, being responsible for the implementation and operation of the electric vehicle charging infrastructure in Portugal since 2015. The company works to expand the charging station network, facilitating access and promoting the adoption of electric vehicles as a sustainable alternative to conventional fossil fuel-powered transportation.

On Mobi.E’s website, was possible to collect the data with the information of the charging stations from Portugal in csv format. With this data I could extract only the ones from Lisbon District and create my data frame only with the columns would be useful for this project. The columns are ID, ADDRESS, LATITUDE and LONGITUDE, and it brings 1025 distinct CP information, as it is possible to see on Figure 8.

	ID	ADDRESS	LATITUDE	LONGITUDE
0	ACH-00002	Autoestrada do Sul do Tejo km 13, Alcochete, L...	38.718824	-8.956143
1	ACH-00008	Rotunda do Entroncamento, Alcochete, Lisboa	38.722252	-9.139337
2	ACH-00010	Av. Restauração, Alcochete, Lisboa	38.751776	-8.964013
3	ACH-00013	Avenida Euro 2004 - Sítio da Quebrada, Alcoche...	38.750266	-8.939431
4	ACH-00016	Lagoa do Lýparo, Alcochete, Lisboa	38.722252	-9.139337
...	...	...	...	...
1020	VFX-00053	Rua João da Silva Vitoriano, Vila Franca de Xi...	38.867074	-9.065763
1021	VFX-00054	Rua 1º de Maio, Vila Franca de Xira, Lisboa	38.952299	-9.013304
1022	VFX-00071	Avenida Flamenga, Vila Franca de Xira, Lisboa	38.872802	-9.083996
1023	VFX-00073	Rua 4 de Junho, Vila Franca de Xira, Lisboa	38.914191	-9.028192
1024	VFX-90001	Praceta da Justiça, Vila Franca de Xira, Lisboa	38.956103	-8.988305

1025 rows × 4 columns

Figure 8 Mobi.E dataframe

Furthermore, to visualize the results spatially was developed Figure 9. A heatmap that allows identify the areas with more charging points installed (red color) or less (blue color) on Lisbon district.

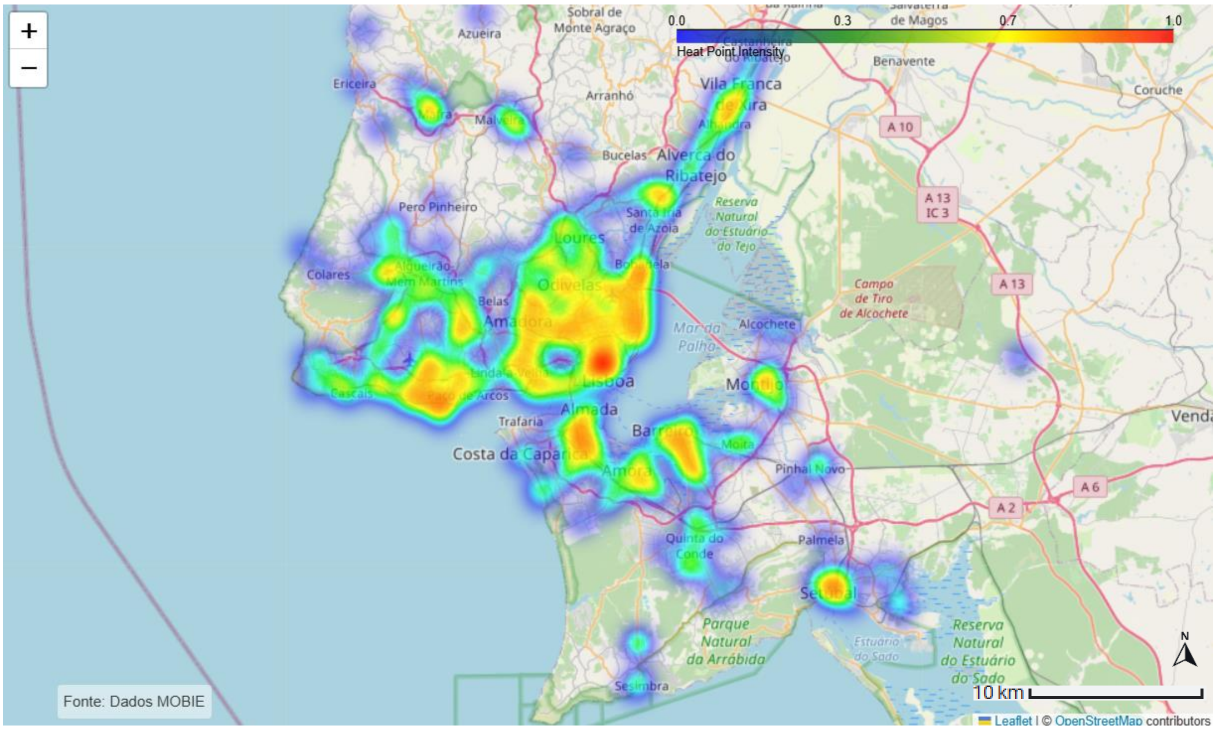


Figure 9 Heatmap Charging Points

3.1.4. Foursquare

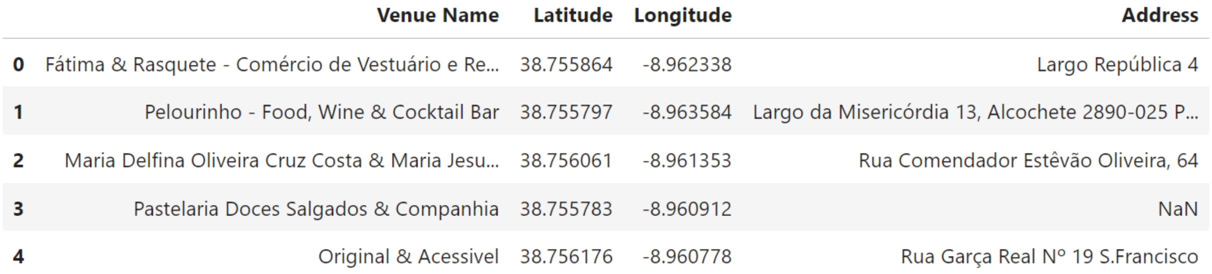
The fourth and final dataset to be collected comes from Foursquare, a location-based social network overflowing with user-generated content about various places such as restaurants, shops, cafes, and famous landmarks. Foursquare offers a Developer Console, which allows for the creation of credentials necessary to access data through its Application Programming Interface (API).

For this dataset, it is needed to gather information about venues in specific municipalities within the Lisbon district. The selection of these municipalities is informed by their identified shortage of charging points (CP), which is discussed in detail in section 3.2. To retrieve venue data from the targeted areas, first is necessary to determine the coordinates of each municipality. This was achieved using the Google API, which provided the latitude and longitude for each area under consideration. Additionally, it was essential to incorporate the Requests library in Python, which simplifies the process of interacting with APIs by enabling the sending and receiving of data in formats like JSON and XML.

Equipped with the latitude and longitude of each municipality in hand, the study progresses to the second step: extracting venue information using the Foursquare API. Due to API limitations, only 50 venues can be retrieved per query. As a result, 15 separate queries are conducted, corresponding to each municipality identified with a deficit in charging point (CP) infrastructure.

To filter the results and focus on the most relevant information, it is specifying the parameters. I set the "query" parameter to "Shops & Services" and “Food”, this ensures it is only received data about venues that fall under this category, such as restaurants, cafes, and shops – places that EV users might frequent. Next, it is provided the latitude from the Municipality (that was identified with Google API) using the "ll" parameter and the radius of 2000 meters. This tells the API that it is only necessary data specific to this specific Municipality, not the broader region. Finally, to prioritize locations with potentially high EV demand, I used the "sort" parameter set to "relevance." This means the API prioritizes venues based on how popular they are and how much user activity they see.

Once I had the data from Foursquare, it initially came in a format called JSON. While JSON can be useful, it's not always the easiest format to work with for analysis. So, I converted this data into Data Frame by import json_normalize from pandas. Then, I extracted only the columns relevant for this work and renamed to make them clearer in the context of our project. Finally, after applying the extraction process for all 15 municipalities, is time to merge these datasets to transform in one and remove the duplicates. The final of this processed Foursquare data brings 674 rows x 4 columns and it is presented in Figure 10.



	Venue Name	Latitude	Longitude	Address
0	Fátima & Rasquete - Comércio de Vestuário e Re...	38.755864	-8.962338	Largo República 4
1	Pelourinho - Food, Wine & Cocktail Bar	38.755797	-8.963584	Largo da Misericórdia 13, Alcochete 2890-025 P...
2	Maria Delfina Oliveira Cruz Costa & Maria Jesu...	38.756061	-8.961353	Rua Comendador Estêvão Oliveira, 64
3	Pastelaria Doces Salgados & Companhia	38.755783	-8.960912	NaN
4	Original & Acessivel	38.756176	-8.960778	Rua Garça Real Nº 19 S.Francisco

Figure 10 Foursquare sample Data

3.2. DATA PROCESSING AND ANALYSIS

On this session it will be explore how the data were processed and analyzed to support the findings of the project. Here, the first two datasets from E-redes and Statistics Institute of Portugal, are responsible for the Identification of the underserved areas, as it will be analyzed on session 3.2.1. While the last two datasets from Mobi.E and Foursquare, equipped with the information from the first two datasets, bring the specific information where on this identified underserved areas the new CP should be installed, as it is analyzed on session 3.2.2.

3.2.1. Identification of Underserved Areas

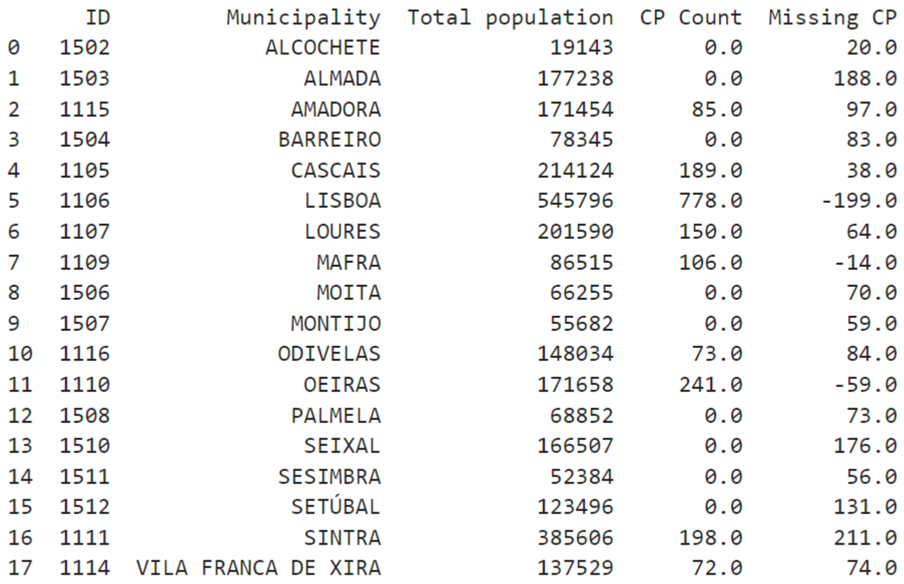
According to Melissa (2024), the average of charging point per inhabitant in EU in 2022 was 106 public charging points for every 100,000 inhabitants. With this number was possible to define which municipalities have a lack on charging point, delimitating the area where a CP should be located.

To encounter this number, first I combined both datasets through the columns Municipality. Then I create a column named CP Count (the counting of charging points per municipality). Finally, I could create the column of Missing CP (that counts how many CP is missing in each municipality) by applying the equation that multiplies per 0.00106 the number of inhabitants, to find how many charging points the municipality should have and subtract the respective number of the CP Count column.

The mathematical formula can be visualized in formula below. Where $MCP(Mi)$ is the number of missing charging points for municipality Mi. $P(Mi)$ is the population of municipality Mi. And $\sum j \in Mi$points$j$ the sum of the existing charging points in municipality Mi.

$$MCP(M_i) = (P(M_i) \times 0.00106) - \sum_{j \in M_i} points_j$$

Also, the final dataset with the merge data and the two new columns, CP Count and Missing CP, it is showing on Figure 11. On the column Missing CP, the positive numbers are the 15 municipalities where have a lack of charging points, and the negative numbers shows the 3 municipalities where the number of CP already meet the demand.



```
    ID        Municipality  Total population  CP Count  Missing CP
0   1502           ALCOCHETE             19143       0.0        20.0
1   1503              ALMADA            177238       0.0       188.0
2   1115             AMADORA            171454      85.0        97.0
3   1504            BARREIRO             78345       0.0        83.0
4   1105             CASCAIS            214124     189.0        38.0
5   1106              LISBOA            545796     778.0      -199.0
6   1107              LOURES            201590     150.0        64.0
7   1109               MAFRA             86515     106.0       -14.0
8   1506               MOITA             66255       0.0        70.0
9   1507             MONTIJO             55682       0.0        59.0
10  1116            ODIVELAS            148034      73.0        84.0
11  1110              OEIRAS            171658     241.0       -59.0
12  1508             PALMELA             68852       0.0        73.0
13  1510              SEIXAL            166507       0.0       176.0
14  1511            SESIMBRA             52384       0.0        56.0
15  1512             SETÚBAL            123496       0.0       131.0
16  1111              SINTRA            385606     198.0       211.0
17  1114  VILA FRANCA DE XIRA            137529      72.0        74.0
```

Figure 11 Missing Charge Points per Municipality

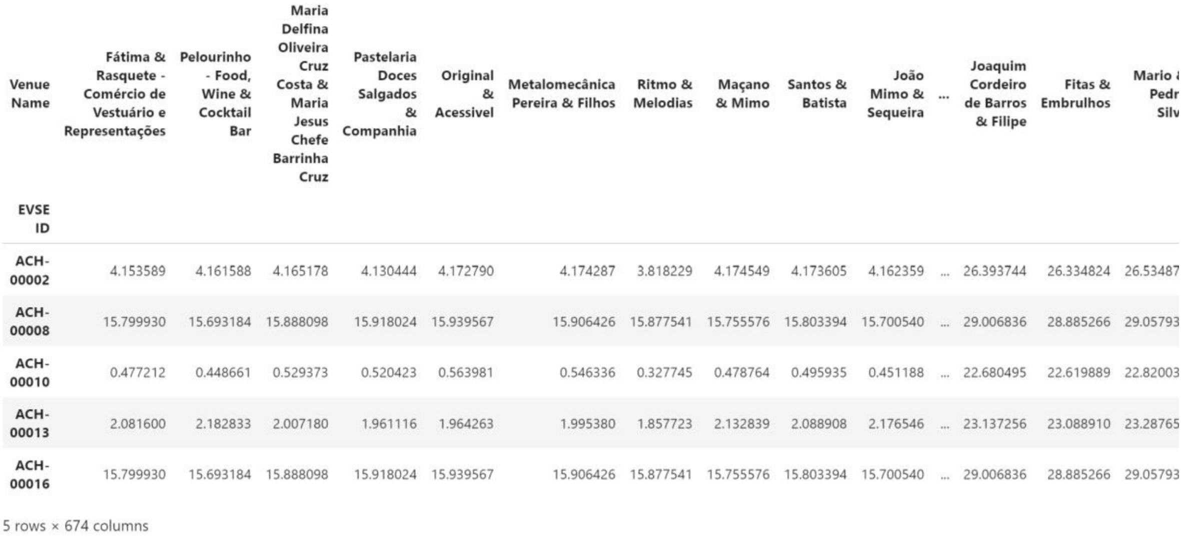
3.2.2. Defining location for new Charging Points

Once we know which municipalities have a lack on CP, now it is time to define where on these 15 municipalities would be the best location to install a new CP.

To this analysis it is combined the last two datasets, from Mobi.E and Foursquare. The Mobi.E dataset, as we could see on session 3.1.3, brings information about the CP locations. While the Foursquare dataset, explored on session 3.1.4, brings information about the relevant venues in the 15 municipalities where exist a lack on the CP infrastructure.

The combination of the two datasets has the objective to measure the distance between each venue obtained from the foursquare API and each CP location from MOBI.E data. And to be possible this analysis, I created two DataFrames, Locations_MOBIE and Locations_Venue, to store the locations of CP and venues, respectively. Also, the latitude and longitude values in these DataFrames had to be converted from degrees to radians. And to results in a distance matrix, the Haversine distance between each pair of CP and venue is computed.

This matrix is then stored in a DataFrame, df_dist_matrix, with CP IDs as the row index and Venue names as the column index. This DataFrame provides a convenient way to look up the distance between any given CP and venue and it is shown a sample with 5 charging points location on Figure 12. And it was necessary to install, besides pandas, numpy to perform operations such as converting degrees to radians and DistanceMetric from sklearn.metrics to calculate the haversine distance between the points.

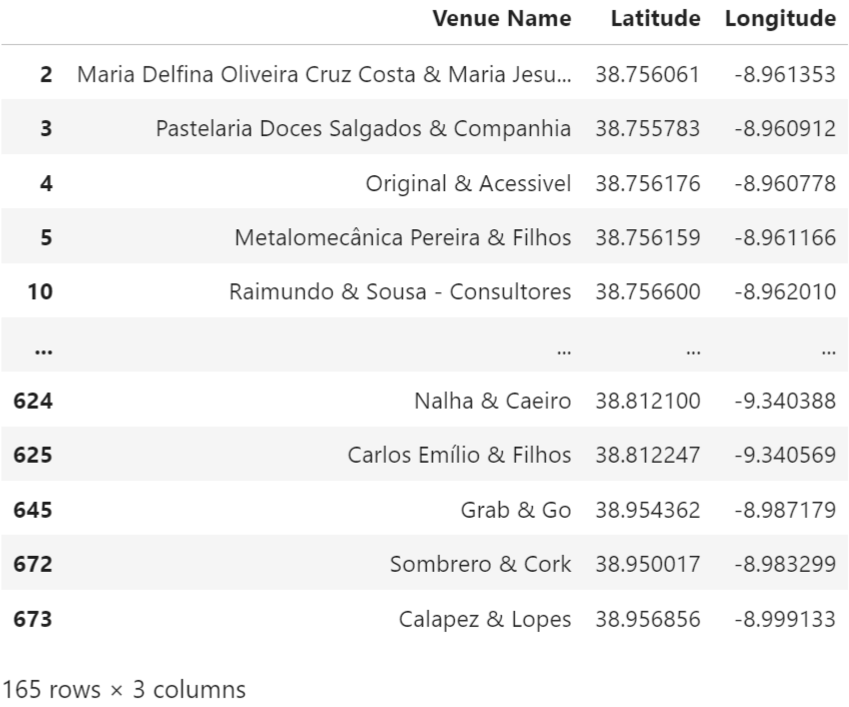


Venue Name	Fátima & Rasquete - Comércio de Vestuário e Representações	Pelourinho - Food, Wine & Cocktail Bar	Maria Delfina Oliveira Cruz Costa & Maria Jesus Chefe Barrinha Cruz	Pastelaria Doces Salgados & Companhia	Original & Acessivel	Metalomecânica Pereira & Filhos	Ritmo & Melodias	Maçano & Mimo	Santos & Batista	João Mimo & Sequeira	...	Joaquim Cordeiro de Barros & Filipe	Fitas & Embrulhos	Mario Pedr Silv
EVSE ID														
ACH-00002	4.153589	4.161588	4.165178	4.130444	4.172790	4.174287	3.818229	4.174549	4.173605	4.162359	...	26.393744	26.334824	26.53487
ACH-00008	15.799930	15.693184	15.888098	15.918024	15.939567	15.906426	15.877541	15.755576	15.803394	15.700540	...	29.006836	28.885266	29.05793
ACH-00010	0.477212	0.448661	0.529373	0.520423	0.563981	0.546336	0.327745	0.478764	0.495935	0.451188	...	22.680495	22.619889	22.82003
ACH-00013	2.081600	2.182833	2.007180	1.961116	1.964263	1.995380	1.857723	2.132839	2.088908	2.176546	...	23.137256	23.088910	23.28765
ACH-00016	15.799930	15.693184	15.888098	15.918024	15.939567	15.906426	15.877541	15.755576	15.803394	15.700540	...	29.006836	28.885266	29.05793

5 rows × 674 columns

Figure 12 Matrix dataset

Wagner et al. (2013) bring in their analysis that a distance of 500 meters is adequate as a walking distance from the CPs to the stores. This analysis is also useful for this project, once the objective is finding best locations for the EV owner park their car and walk until their destination. For this reason, the next step is identifying the number of Venues that has no charging station at distance which is less than or equal to 0,5 km from the prospective Venues. What has resulted on a dataset with 165 relevant Venues that needs a new CP installation close by, as it is exposed on Figure 13.



	Venue Name	Latitude	Longitude
2	Maria Delfina Oliveira Cruz Costa & Maria Jesu...	38.756061	-8.961353
3	Pastelaria Doces Salgados & Companhia	38.755783	-8.960912
4	Original & Acessivel	38.756176	-8.960778
5	Metalomecânica Pereira & Filhos	38.756159	-8.961166
10	Raimundo & Sousa - Consultores	38.756600	-8.962010
...	...	...	...
624	Nalha & Caeiro	38.812100	-9.340388
625	Carlos Emílio & Filhos	38.812247	-9.340569
645	Grab & Go	38.954362	-8.987179
672	Sombrero & Cork	38.950017	-8.983299
673	Calapez & Lopes	38.956856	-8.999133

165 rows × 3 columns

Figure 13 Venues with lack on CP

With the location of each relevant Venue with a lack on charging station infrastructure in Lisbon district, it is time for the next stage. At this stage, is decided which modeling technique is most suitable for the problem at hand. One such technique is clustering, an unsupervised learning approach where data points are grouped (or clustered) based on their similarities. K-Means, a popular clustering algorithm, falls under this category. It partitions the data into a pre-defined number of clusters, allowing us to identify inherent groupings and explore potential relationships within the data.

K-means clustering of the find Venues plays a pivotal role in this project, allowing us to group similar venues into distinct clusters. For the number of clusters k to use on the algorithm, was defined as the number of municipalities it is being working with, in other words, 15 clusters since it was identified the lack on charging points infrastructure in fifteen municipalities.

Afterward, all venues will be allocated to the nearest cluster. Also, the distance between each centroid in k-means is typically measured using Euclidean distance. The formula for Euclidean distance between two points in an n-dimensional space is given by:

$$D = \sqrt{\sum_{i=1}^{n} (X_{i1} - Y_{i2})^2}$$

Where:

D: is the Euclidean distance between the two points.

n: is the number of dimensions (or features) in the space.

Xi1 is the coordinate of point 1 in dimension i.

Xi2 is the coordinate of point 2 in dimension i.

By applying k-means method, it is also identified the centroids of each cluster. For this matter, I use geopy with Nominatim service from OpenStreetMap, which converts the centroids coordinates in to addresses. Resulting in namely addresses where I can set the exact location where should be located a new charging point. And to use the KMeans clustering algorithm on Jupyter notebook, it was necessary to import KMeans from sklearn.cluster.

4. RESULTS AND DISCUSSION

4.1. RESULTS

The expected outcome of this project is to pinpoint the optimal sites for the installation of new Charging Points (CPs) within the Lisbon district. To achieve this goal, it was first necessary to understand which areas were experiencing a deficiency in this infrastructure. As revealed in Section 3.2.1, the municipalities identified with a lack in CPs include Almada, Amadora, Barreiro, Cascais, Loures, Moita, Montijo, Odivelas, Palmela, Seixal, Sesimbra, Setúbal, Sintra, and Vila Franca de Xira.

Having identified the municipalities in question, the next step was to determine the most suitable locations within these areas for the establishment of new CPs. The methodology described in Section 3.2.2 enabled us to understand that there are 165 venues within these municipalities that lack a CP within a 0.5km radius. These potential sites for CP installation are distributed across the municipalities, as illustrated in Figure 14.

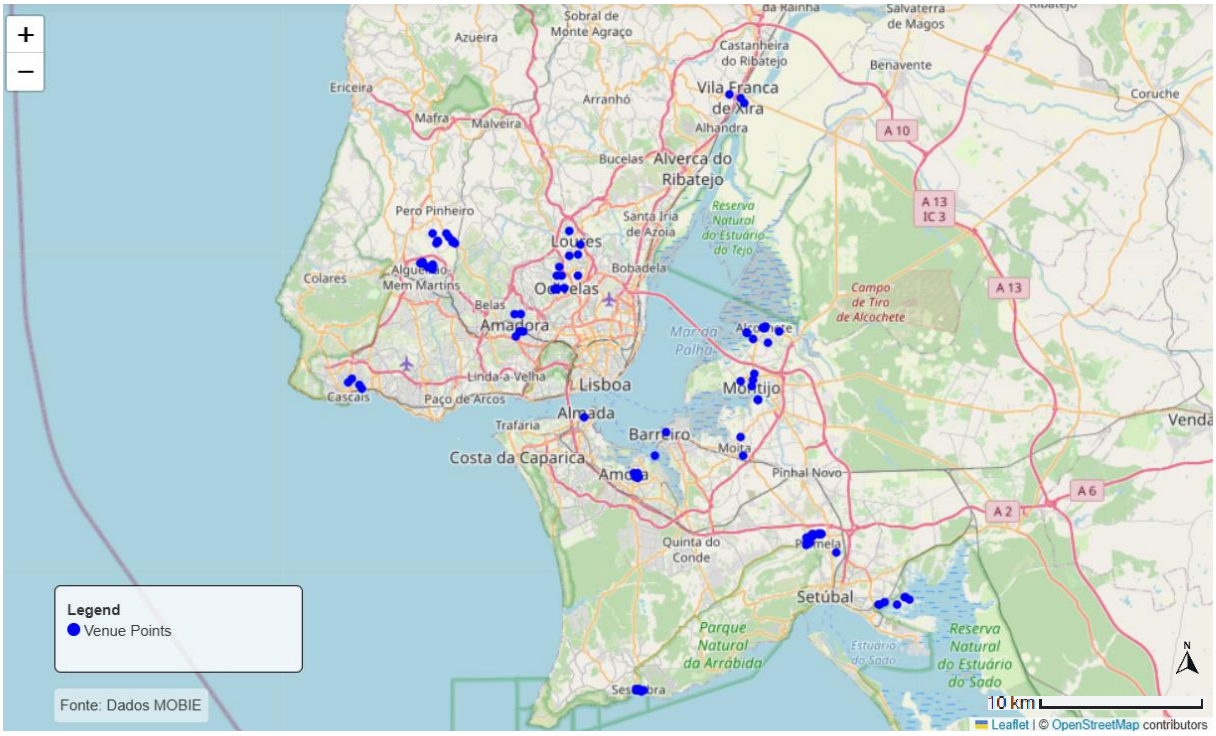


Figure 14 Venues without a near CP

Furthermore, it was necessary to apply the K-means clustering algorithm, as detailed in Section 3.2.2. The algorithm successfully delineated 15 distinct clusters, each differentiated by a unique color. Figure 15 offers a visual representation of the spatial distribution of these clusters.

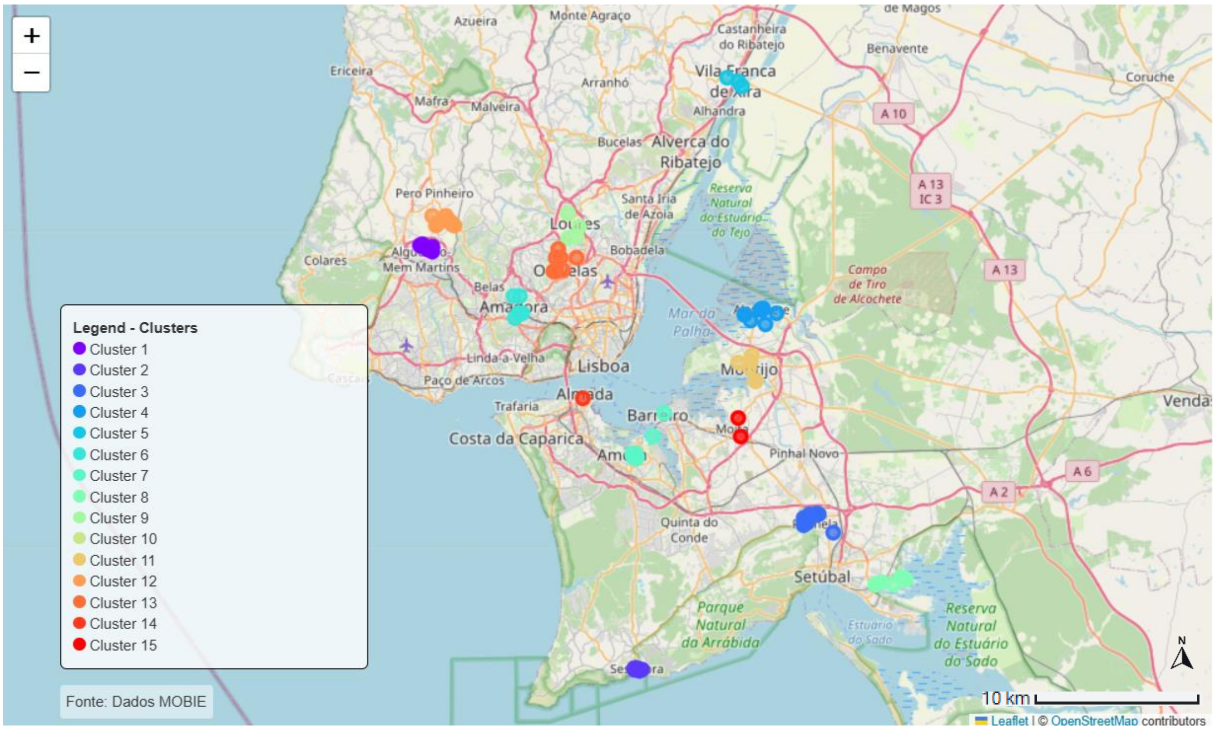


Figure 15 Venues K-means Clustering

The final step involves identifying the centroids of these clusters, which will serve as the proposed locations for installing new charging stations. Figure 16 displays a map visualization of these suggested locations, marked by blue pins with car icons. Additionally, the map includes green points to indicate the existing charging stations.

Below is a list of the centroid addresses, there are the recommended sites for new Charging Point (CP) installations:

- Rua dos Lavadouros, Sacotes, Algueirão-Mem Martins, Sintra, Lisboa, 2710-728, Portugal
- Rua Leão D'Oliveira, Santiago, Sesimbra, Setúbal, 2970-706, Portugal
- Escola Básica Ary dos Santos, Rua das Margaridas, Bairro do Corte Falcão, Bela Colónia, Montijo, Montijo e Afonsoeiro, Montijo, Setúbal, 2870-290, Portugal
- Avenida dos Bombeiros Voluntários, Palmela, Setúbal, 2950-201, Portugal
- Ribeirada, Casal do Chapim, Odivelas, Lisboa, 2675-563, Portugal
- Avenida dos Metalurgicos, Álamo, Arrentela, Seixal, Arrentela e Aldeia de Paio Pires, Seixal, Setúbal, 2840-266, Portugal
- CTT, Rua Luís de Camões, Cais de Vila Franca de Xira, Vila Franca de Xira, Lisboa, 2600-180, Portugal
- 804, Rua Marechal Carmona, Castelhana, Cascais Rivieira, Cascais e Estoril, Cascais, Lisboa, 2750-442, Portugal
- 126, Rua António Lourenço, Bairro Cooperativa, Faralhão, Sado, Setúbal, 2910-149, Portugal

- Bairro do Juncalinho, Moita, Setúbal, 2860-457, Portugal
- Mediogo - Mediação Imobiliária, Lda., 14, Rua Henrique Nogueira, Quinta Nova, Mina, Mina de Água, Amadora, Lisboa, 2700-449, Portugal
- Cortegaça, Almargem do Bispo, Pêro Pinheiro e Montelavar, Sintra, Lisboa, 2715, Portugal
- Praça Alexandre Herculano, Cidade Nova, Santo António dos Cavaleiros e Frielas, Loures, Lisboa, 2660-232, Portugal
- Ribs & Company, 11, Rua São Salvador da Bahia, Almada Velha, Cova da Piedade, Almada, Cova da Piedade, Pragal e Cacilhas, Almada, Setúbal, 2800-201, Portugal
- Paceta Padre Cruz, Urbanização dos Barris, Alcochete, Setúbal, Portugal

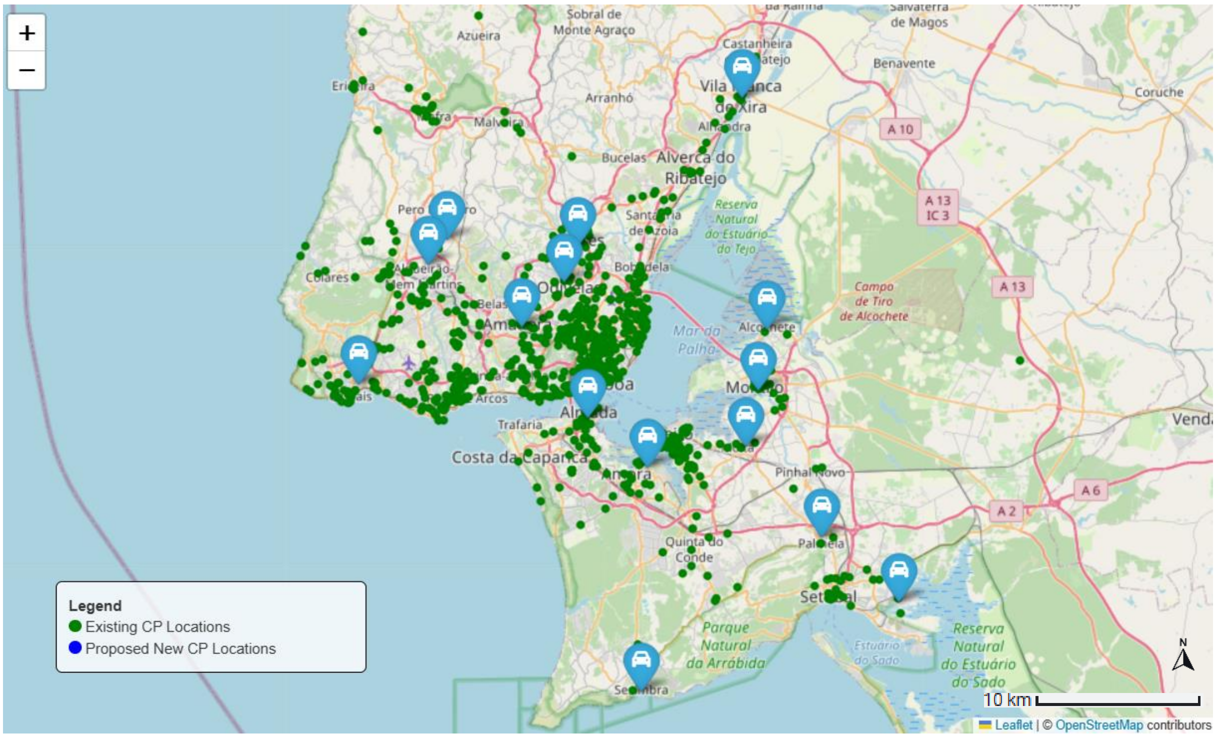


Figure 16 Map with recommended sites for new CP

In synthesis, this section responds the research questions defined that guided the investigation. The first question was about where the demand for new charging points is most pronounced within the Municipalities of Lisbon. The answer for this question was the municipalities of Almada, Amadora, Barreiro, Cascais, Loures, Moita, Montijo, Odivelas, Palmela, Seixal, Sesimbra, Setúbal, Sintra, and Vila Franca de Xira. The second question was what the strategic locations are to install new charging stations. To answer this question, we bring the map that shows that the strategic locations are the fifteen points marked by blue pins with car icons and the list with the proper complete address above on this same chapter.

5. CONCLUSION

This thesis has provided an insightful analysis and a set of viable strategies to enhance the electric vehicle (EV) charging infrastructure in the Lisbon district of Portugal. The findings highlight areas where the current system can be optimized and outline a concrete plan for the expansion of the charging network, aligned with identified needs.

Thanks to the effective application of data analysis tools, it was possible to map regions with unmet demand for charging points, providing crucial direction for future installations. This study thus significantly contributes to the energy transition effort, promoting solutions that can lead to greener and more sustainable mobility.

The results emphasize the importance of collaboration between the public sector, the automotive industry, and local communities. The implementation of the suggested strategies benefits from supportive policies, continuous investment, and data-driven planning.

It is recognized that the study has its limitations, including the use of more variables to estimates the demand for CP, the absence of an in-depth analysis of EV user preferences and impacts on the electrical grid. These areas represent opportunities for future research that can further enrich the field of electric mobility.

In summary, the insights gained in this thesis contribute to advancing knowledge in EV infrastructure planning and support the progression towards more sustainable transportation. While challenges remain, the work presented here establishes a solid foundation for the continued expansion of electric vehicles in urban environments.

6. LIMITATIONS AND FUTURE RESEARCH

This study provided important insights into the necessary infrastructure for electric vehicle (EV) charging points in certain municipalities within the Lisbon district, though it recognizes that there are limitations that need to be considered. One of the main limitations is that the demand for charging points was estimated based on the number of inhabitants, which may not reflect the actual demand, as factors such as average income per municipality, the quantity of EVs on that area and frequency of use are varied and relevant. Moreover, the study did not account for local variations in the adoption rate of electric vehicles, which can be influenced by municipal policies and incentives, as well as local culture concerning sustainability.

The analysis was also based on static data, without considering population growth or changes in the patterns of electric vehicle adoption over time. Preferences and behaviors of electric vehicle users were also not directly analyzed, which can affect the efficacy and acceptance of the charging points. Lastly, the potential impact that the installation of new charging points could have on the local electrical grid, including capacity and stability issues, was not addressed.

For future work, it would be beneficial to adopt models that consider multiple variables to estimate the demand for charging points, and to conduct longitudinal studies to monitor the adoption rates of electric vehicles. Furthermore, researching user preferences and charging patterns could help optimize the location and design of charging points. It is also essential to assess the impact on the existing electrical infrastructure to ensure that the grid can support the new demand. Finally, an economic feasibility analysis considering the costs associated with the installation and maintenance of charging points would be very useful for effective infrastructure planning.

In summary, this study serves as a starting point for understanding and planning the infrastructure for electric vehicle charging, and there is a vast field for more detailed and dynamic research that can build upon the foundation established here.

BIBLIOGRAPHICAL REFERENCES

Catalbas, M. C., Yildirim, M., Gulten, A., & Kurum, H. (2017). Estimation of optimal locations for electric vehicle charging stations. *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 1–4. https://doi.org/10.1109/EEEIC.2017.7977426

Csiszár, C., Csonka, B., Földes, D., Wirth, E., & Lovas, T. (2020). Location optimisation method for fast-charging stations along national roads. *Journal of Transport Geography*, *88*, 102833. https://doi.org/10.1016/j.jtrangeo.2020.102833

Efthymiou, D., Chrysostomou, K., Morfoulaki, M., & Aifantopoulou, G. (2017). Electric vehicles charging infrastructure location: A genetic algorithm approach. *European Transport Research Review*, *9*(2), 27. https://doi.org/10.1007/s12544-017-0239-7

Eisel, M., Schmidt, J., & Kolbe, L. M. (2014). Finding suitable locations for charging stations. *2014 IEEE International Electric Vehicle Conference (IEVC)*, 1–8. https://doi.org/10.1109/IEVC.2014.7056134

European Environment Agency. (2022). *Transport and environment report 2022: Digitalisation in the mobility system : challenges and opportunities.* Publications Office. https://data.europa.eu/doi/10.2800/47438

EV Connect. (2020). *How to Add Value to Parking Facilities With EV Parking and Charging*. https://www.evconnect.com/blog/how-to-add-value-to-parking-facilities-with-ev-parking-and-charging

Hofer, C., Jäger, G., & Füllsack, M. (2018). Large scale simulation of CO2 emissions caused by urban car traffic: An agent-based network approach. *Journal of Cleaner Production*, *183*, 1–10. https://doi.org/10.1016/j.jclepro.2018.02.113

Jeczawitz, K. (2021). *INCREASING CUSTOMER EXPERIENCE THROUGH THE OPTIMIZATION OF PREMIUM PUBLIC CHARGING BY MERCEDES-BENZ* [Nova School of Business and Economics]. https://run.unl.pt/bitstream/10362/144800/1/2020-21_fall_43311_katharina-jeczawitz.pdf

King, N. (2023). *Forecasting Lead at EV-volumes*. https://www.ev-volumes.com/

Kong, W., Luo, Y., Feng, G., Li, K., & Peng, H. (2019). Optimal location planning method of fast charging station for electric vehicles considering operators, drivers, vehicles, traffic flow and power grid. *Energy*, *186*, 115826. https://doi.org/10.1016/j.energy.2019.07.156

Li, Z., Khajepour, A., & Song, J. (2019). A comprehensive review of the key technologies for pure electric vehicles. *Energy*, *182*, 824–839. https://doi.org/10.1016/j.energy.2019.06.077

Mayr, K., & Rentschler, J. (2023). *Fossil Fuel Prices and Air Pollution: Evidence from a Panel of 133 Countries*. The World Bank. https://doi.org/10.1596/1813-9450-10397

Melissa, R. (2024). *Europe EV Charging Station Market*. https://statzon.com/insights/ev-charging-points-europe#:~:text=On%20average%2C%20there%20were%20106,in%20the%20EU%20in%202022.

Miranda, J. L., & Delgado, C. J. M. (2020). Determinants of Electric Car Purchase Intention in Portugal. In D. Crowther & S. Seifi (Eds.), *Governance and Sustainability* (Vol. 15, pp. 161–172). Emerald Publishing Limited. https://doi.org/10.1108/S2043-052320200000015009

Skjærseth, J. B. (2021). Towards a European Green Deal: The evolution of EU climate and energy policy mixes. *International Environmental Agreements: Politics, Law and Economics*, *21*(1), 25–41. https://doi.org/10.1007/s10784-021-09529-4

Wagner, S., Götzinger, M., & Neumann, D. (2013). *Optimal location of charging stations in smart cities: A point of interest based approach.* https://www.researchgate.net/publication/287014278_Optimal_location_of_charging_stations_in_smart_cities_A_point_of_interest_based_approach

Wang, D., Ge, Y., Lin, Q., Yang, S., & Chen, R. (2022). Location and Capacity of Charging Station for Electric Vehicles Based on Weighted K-Means Clustering. *2022 34th Chinese Control and Decision Conference (CCDC)*, 5305–5310. https://doi.org/10.1109/CCDC55256.2022.10034067

Woo, H., Son, Y., Cho, J., Kim, S.-Y., & Choi, S. (2023). Optimal expansion planning of electric vehicle fast charging stations. *Applied Energy*, *342*, 121116. https://doi.org/10.1016/j.apenergy.2023.121116

Zhou, G., Zhu, Z., & Luo, S. (2022). Location optimization of electric vehicle charging stations: Based on cost model and genetic algorithm. *Energy*, *247*, 123437. https://doi.org/10.1016/j.energy.2022.123437

Zhou, J., Wang, R., Huang, W., Zhao, Z., Zheng, F., & Wang, Z. (2023). Location and Layout of Electric Vehicle Charging Stations Based on K-Means Algorithm. *Journal of Physics: Conference Series*, *2592*(1), 012070. https://doi.org/10.1088/1742-6596/2592/1/012070

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa