INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Electric Vehicles Charging Patterns

Rui Simão Rodrigues

Master's in Integrated Decision Support Systems

Supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with aggregation,
ISCTE

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

September, 2021

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Electric Vehicles Charging Patterns

Rui Simão Rodrigues

Master's in Integrated Decision Support Systems

Supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with aggregation,
ISCTE

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

September, 2021

“*Our greatest weakness lies in giving up.*

The most certain way to succeed is always to try just one more time.”

Thomas Edison

Acknowledgments

To my advisors, Prof. João Ferreira and Prof. Miguel Dias. Their insight, expertise, and commitment in answering a never-ending number of doubts whenever the path forward was nebulous. Thank you both for all the fantastic support and guidance over these last months.

To Mr. Manuel Reis, from the UVE association, and to the DECO-Proteste magazine, for publishing the surveys to their respective userbases.

To Prof. Lúcia Martins and Paulo Pereira for providing valuable assistance in conducting the necessary statistical analysis and Arch. Vitória Albuquerque, for the availability and support given.

To my mother, for all the support given. Her advice and encouragement throughout this dissertation have been an immense help.

To my beta readers Nuno and Rui, thank you for taking the time to review, error-check, and provide valuable input. Without them, this thesis would lack the necessary clarity and delivery, and I thank them for their effort.

Lastly, to Joana. Her support has been unwavering, even though she has so much on her shoulders. Without your support, I would have been lost.

Resumo

O aumento das emissões de gases de efeito de estufa na atmosfera, e os seus efeitos negativos no ambiente veio instigar a procura de fontes de energia alternativas aos combustíveis fósseis. Uma das soluções que tem vindo a ganhar terreno, é a eletrificação de diversas atividades humanas, tais como o setor dos transportes. Esta tendência tem fomentado uma crescente necessidade por armazenamento de energia elétrica em baterias de lítio. Saber com exatidão o grau de degradação que este tipo de baterias acumula ao longo do seu tempo de vida útil, é uma necessidade que poderá trazer vantagens económicas, tanto para as empresas como para os cidadãos.

O presente trabalho propõe duas perguntas de investigação sobre os veículos automóveis elétricos, para responder à necessidade existente: a primeira incide sobre hábitos praticados pelos donos de veículos elétricos, que poderão ter um efeito negativo na vida útil das baterias, e a segunda sobre fatores que poderão afastar os consumidores da compra deste tipo de veículos. Esta tese procurou responder a essas duas perguntas, recorrendo a uma metodologia da área da ciência dos dados, e a análise estatística, aplicadas a três inquéritos realizados a proprietários de veículos elétricos.

Os resultados permitiram concluir que à exceção da variável Ano (*Year*), todos os outros fatores tiveram um efeito marginal na degradação da autonomia real dos veículos. No que respeita aos obstáculos à adoção de veículos elétricos, o maior obstáculo encontrado foi o da insuficiente cobertura da rede de postos de carregamento.

Palavras-Chave: veículos elétricos, processo de carregamento, comportamento.

Abstract

The increase in greenhouse gas emissions into the atmosphere, and their adverse effects on the environment, has prompted the search for alternative energy sources to fossil fuels. One of the solutions gaining ground is the electrification of various human activities, such as the transport sector. This trend has fueled a growing need for electrical energy storage in lithium batteries. Precisely knowing the degree of degradation that this type of battery accumulates over its useful life is necessary to bring economic benefits, both for companies and citizens.

This paper aims to answer the current need by proposing two research questions about electric motor vehicles. The first focuses on habits EV owners practice, which could harm the battery life, and the second on factors that could keep consumers from purchasing this type of vehicle. This thesis sought to answer these two questions, using a methodology from data science and statistical analysis, applied to three surveys carried out on electric vehicle owners.

The results allowed us to conclude that, except for the Year variable (Year), all other factors had a marginal effect on the vehicles' absolute autonomy degradation. About obstacles to the adoption of electric vehicles, the biggest obstacle encountered was the insufficient coverage of the network of charging stations.

Keywords: electric vehicles, charging process, behavior.

Table of Contents

List of Tables

List of Figures

Glossary of Acronyms

AI	Artificial Intelligence
ANOVA	Analysis of Variance
CRISP-DM	Cross Industry Standard Process for Data Mining
EV	Electric Vehicle
GHG	Greenhouse gases
KNN	K-Nearest Neighbors
Li-ion	Lithium-ion
LR	Logistic Regression
ML	Machine Learning
NB	Naïve-Bayes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RUL	Remaining Useful Life
SoC	State of Charge
SoE	State of Energy
SoH	State of Health
SLR	Systematic Literature Review
SPSS	Statistical Product and Service Solutions
SVM	Support Vector Machine

1. Introduction

The central topic of this thesis is aligned with the urgent need to electrify most of the human activities, which is an ongoing European and national policy. For example, the European Green Deal of 2020 [1] aims to reduce greenhouse gases (GHG) in 2030 to at least 55% of 1990 values. Likewise, the Portuguese National Plan for Energy and Climate PNEC 2030 (2019) [1] foresees until 2030 a reduction between 45% and 55% in greenhouse gas emissions of 2005 levels and a 20% incorporation of renewable energies in the transportation sector.

Human activities have been emitting considerable amounts of GHG to the atmosphere, notably in the last century, from trade to transportation, industry, and even agriculture. The 2021 Intergovernmental Panel on Climate Change (IPCC) [2] report recently stated with high confidence that there is a near-linear relationship between cumulative anthropogenic CO2 emissions and the global warming consequences. Therefore, the current living generations must take a decisive step to accelerate the transition to power our energy needs with renewable sources.

Hence, the theme of this thesis focuses on a small part of the bigger problem, the transition to electric passenger vehicles. Any energy storage solution developed with optimization could diminish the importance of the current lackluster battery capacities. Moreover, this study aims to find the critical factors regarding the usage and charging of an electric vehicle (EV) that could most negatively influence its battery's remaining useful life (RUL).

To tackle this problem and answer its following research questions, which will be detailed in sub-chapter 1.3.2 below, we adopted a Data Science oriented approach, applying the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology [3]. CRISP-DM was applied to one of the available data sources, a dataset with Tesla due to its volume of data. In addition, a more traditional exploratory statistical analysis was used on three surveys.

Additionally, a systematic literature review was performed based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology [4], which focused on Machine Learning (ML) classification models, later applied in our analysis.

Regarding data collection, we started by retrieving data from a public inquiry to Tesla vehicles owners. This Tesla dataset was published on Elektrek [5], an electrical news website, on April 14, 2018. It depicted a downward trend curve for vehicle degradation that stabilized at a deficit of 10% of battery total capacity after one hundred and sixty thousand miles, which was promising news. Furthermore, confirming the Elektrek article findings, in early 2020, Tesla Inc. published a report stating their batteries would retain 90% of their original capacity after 200,000 miles of usage [6].

Moreover, we applied the CRISP-DM methodology to the Tesla dataset by creating a classification model based on the variables available in the raw file obtained by the Elektrek article. The CRISP-DM method was adapted to create a labeling model based on the pertinent variables available in the raw file. The end goal was to analyze the distinct factors at play when operating or charging an EV.

Additionally, three surveys were created in the thesis framework and shared online with current and prospective EV users to further enrich and diversify the Tesla dataset's answers. Data from these three surveys were merged with the Tesla dataset and were subject to statistical analysis with Statistical Product and Service Solutions (SPSS) software from IBM. The aim was to uncover valuable insight into the satisfaction degree of the current charging network in Portugal.

The main results from the battery degradation analysis showed that factors like the frequency of charging an EV in fast-charging stations, such as the Tesla Supercharger network, might not have a measurable impact on battery degradation. Instead, based on the results of this sample, the car's age might be the central factor. Future research could produce more precise and detailed models and better explain these factors' impact on car range.

The statistical analysis results about the EV adoption obstacles allowed us to conclude that the insufficient coverage of the charging station network is a critical factor. A second factor is often mentioned and has a deterrent role: the cost of recharging at these same stations.

1.1. Motivation

The electrification of most human activities is nowadays a necessity. It is a crucial action towards reducing GHG emissions — targeting the larger goal of decarbonizing human society. The application of energy storage technology in the transportation sector, mainly adopted in electric passenger vehicles, is a strategic step towards the widespread adoption of this type of mobile technology and the subsequent decarbonization of society. This research on lithium-ion batteries aims to know more about a subject we understand is not as disseminated as it ideally should be. Also, it aims to collect and obtain insights into the most updated state-of-the-art research on the topic and extract information from data relinquished by EV (Electric vehicles) owners. This information is expected to discuss the satisfaction degree that electric vehicle users have with the current solutions in Portugal, potentially repel consumers from purchasing this type of vehicle.

1.2. Problem Description

In recent decades, the increase of erratic climatic changes has made it crucial to find alternative forms of energy conservation to conventional methods, such as fossil fuels, that significantly contribute to greenhouse gas emissions. Therefore, it is of particular importance to adopt these alternatives with the utmost celerity.

Fortunately, actions to reduce GHG emissions have been implemented, even before the covid-19 pandemic, resulting in CO2 emissions decrease by 1.8 percent, from 2018 to 2019 [7]. The US Environmental Protection Agency explains that this result was primarily due to a drop in total energy use in 2019 compared to 2018, brought down by fossil fuel emissions reduction. In tandem with GHG emissions reduction, an ongoing shift from coal to natural gas (the most GHG-emitting fossil fuel) occurs in the energy sector.

One alternative solution to the burning of fossil fuels in the transportation sector is adopting Li-ion batteries. This type of battery is currently empowering EVs. Technological improvements have been implemented in the last ten years to increase these batteries' energy capacity and efficiency. However, because its capacity is finite, any factor that decreases its energy retention ability is crucial. The degradation of the energy capacity of this type of battery, which is observed, for instance, in cell phones, is one of the main problems faced by energy experts.

Battery aging is currently a problem that crosscuts all sectors of activity that depend on it now or may depend on it soon. The transportation sector is one of the most affected sectors of activity, specifically in private vehicles. For example, electric mobility is an emerging, ever-growing mode of transport that causes an increased demand for Li-ion batteries in vehicles. However, these batteries have a limited useful life and are usually grouped in packs that make them difficult to replace. Additionally, the recycling of batteries' toxic components has proven to be a hazard to the environment. Thankfully, there is a growing need to find methods that can extend the life of these battery packs to reduce their environmental footprint [8] and find non-toxic elements that can be favored in their manufacturing process.

In the automotive industry, this premature aging of batteries is adverse in two other ways: firstly, it limits the range autonomy of the private vehicle. Moreover, it also affects its acceptance and adoption by the public in general. Therefore, knowing the exact pace of battery degradation is necessary and often motivates information campaigns for technology adoption, academic research, and industrial R&D to improve its performance and longevity [8].

Future potential owners of vehicles powered by Li-ion batteries are starting to require accurate information on how long their vehicle batteries will last [8]. Hence, consumers are interested in determining whether it is advantageous to invest in this new technology and pay extra fees for its early adoption.

Battery early aging often depends on the Li-ion battery materials' chemical composition, namely its anode, cathode, and electrolyte. In addition, external factors, such as voltage, discharge intensity, temperature, and the number of charging-discharging cycles performed, are also considered important factors. However, the reference literature does not quantify how relevant these factors are to the overall battery longevity. For instance, the Tesla manufacturer applies solutions to mitigate premature battery aging; all its vehicles have a management system whose primary function is to control the battery’s temperature to remain below 55 degrees Celsius [8].

However, behavioral factors associated with the operation and charging of electric cars and their storage are significantly considered to impact the degradation of batteries [8]. As already observed with mobile phones, car batteries are subjected to premature aging if left unused. This concern regards that both cases use the same technology and materials. On the other hand, their continued use also leads to a progressively shorter service life. May [9] suggest that both technologies' similarities would not stop at that point, and the EV would be as prevalent as the mobile phone. May also envisions that one day, everyone would be able to have one.

The evaluation of Li-ion batteries’ performance is still an ongoing process. This technology continues to be studied and matured iteratively by the scientific community that seeks different methods to measure its capacity, internal resistance, and voltages as well as its influence in charge and discharge cycles [10][11][12].

According to Yun, [11] the high complexity of practical solutions brings difficulty in measuring the variables mentioned above, especially in controlling the internal variables related to the consistency of the manufacturing quality of the various components of the batteries. Thus, it becomes necessary to assess batteries' health status or State of Health (SoH).

The SoH of the battery, expressed as a percentage, represents its current capacity in Watts, concerning its original capacity. This value weighs various parameters of Li-ion batteries, such as their voltage, current, and capacity. Currently, few articles [13][14][15][16][17][18] can accurately predict the actual value of SoH.

There are two types of battery capacity forecasting methods to determine the SoH: model-based methods and data-based methods. Model-based methods were always related to the chemical composition of batteries, and there is plenty of reference literature available on this subject.

Regarding data-based prediction models, these sometimes use the parameters referred to earlier [13][14][15][16][17][18] to monitor the SoH and forecast the state of the RUL [19][20]. Compared to prediction methods based on chemical models, these data-based methods [21] are faster, more convenient, and less complex [22]. Moreover, Machine Learning (ML) methods can be used, resulting in improving the accuracy of these models. These prediction methods have raised a growing interest in verifying the SoH of batteries [23].

1.3. Thesis Objectives

The challenges humanity faces in sustainable mobility and the current European Union [24] and National policies [1] triggered innovative technologies to minimize fossil fuels in transport. To this end, intelligent computing techniques, such as ML, are advantageous to achieve this goal. The benefits of ML are especially relevant when it comes to alternative technologies such as battery-powered electric vehicles. However, this technology's maturity has not yet been reached, as there is room for improvement in the quantity of energy stored and in recharging speed. These two significant disadvantages impact the vehicles' autonomy and the convenience of their users, who may constitute additional barriers to their adoption, besides EV prices.

1.3.1. Research Gap

Regarding the systematic literature review (SLR), it was possible to identify a gap in the current state of the art: just a single article [25] refers to the degradation of batteries from the standpoint of EV user behavior and battery charging patterns. All other articles mention battery degradation solely from the point of view of the electrochemistry field, explaining in detail how the batteries' components and the environmental conditions affected battery longevity. Furthermore, they also mention that Li-ion batteries lose capacity depending on the intensity of their use or even lack thereof. These drawbacks add to the list of disadvantages mentioned previously, may further deepen EV adoption hesitancy by consumers, which is an exciting topic discussed in sub-chapter 4.5.2.

1.3.2. Research Questions

After reviewing the literature, we found the following gaps: just one paper on battery degradation behavioral factors and none on existing dissuasive factors before purchasing electric vehicles that might dissuade potential buyers. The following two research questions address these lapses found in the reference literature were formulated:

RQ1: Which behavioral habits from the electric vehicles may negatively impact lithium-ion battery capacity?

RQ2: Which factors might present themselves as a hindrance to the adoption of EV vehicles by citizens?

In Table 1-1, the methodology to answer the two formulated research questions is proposed as follows:

Table 1-1 - Research question methodology.

Research Questions	Objectives	Methodology
RQ1 -Which behavioral habits from the electric vehicles may negatively impact Li-ion battery capacity?	OBJ 1 - To determine if there are factors in the data sample that might influence battery degradation.	CRISP-DM plus Descriptive Statistics and Content Analysis
RQ2 - Which factors might present themselves as a hindrance to the adoption of EV vehicles by citizens?	OBJ 2 - To identify and understand user satisfaction and how that might hinder the adoption of electric vehicles.	Descriptive Statistics and Content Analysis

The two main objectives of this thesis were stated as follows: first to determine if there are factors in the sample that might influence battery degradation (by answering the RQ1), and secondly, to identify and understand user satisfaction and how that might hinder the adoption of electric vehicles by pursuing an answer to RQ2.

Descriptive statistical analysis in SPSS was the chosen method to obtain a general and exploratory overview of the distribution of the several datasets collected and meet the second objective (OBJ2). In addition, to fulfill the first objective (OBJ 1), we performed a second analysis based on the CRISP-DM methodology due to the higher volume of data present in the Tesla dataset.

1.4. Dissertation Structure

This document is structured into five chapters as defined in the following structure:

- *Chapter 1* presents the context and the methodological steps taken.
- *Chapter 2* showcases state of the art, which introduces the subject's present situation. The PRISMA systematic literature review method and the VOSviewer bibliometric visualization tool are applied in this chapter.
- *Chapter 3* describes the case study and the application of the CRISP-DM methodology to the Tesla dataset and discusses the results.
- *Chapter 4* introduces the complementary creation of three online surveys, their objectives, targeted audiences, questions made and designed, and the performed statistical analysis.
- *Chapter 5* describes the research conclusions; the limitations found and provides insights for future research to improve the CRISP-DM battery degradation model's accuracy and performance.

2. Literature Review

The upcoming sub-chapters document the process of our systematic literature survey. Although the topic of electric vehicles is recent, there is much interest in it, and as a result, there is plenty of pertinent, available literature online. Therefore, it was necessary to employ a systematic analysis method to efficiently filter out the works less relevant and highlight the ones most related to the theme of the thesis.

2.1. PRISMA Systematic Literature Survey Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[4] is a method for obtaining literature reviews with systematic and objective results. It is a method where findings can be reproduced and verified easily by other researchers. PRISMA is used to write systematic research reviews, particularly in the Medicine, Social Sciences, and Exact Sciences areas.

It thus presents as a set of recommendations followed by authors who wish to publish and report comprehensively and transparently how they reached their conclusions at the time of their bibliographic research. In other words, the PRISMA guidelines help authors to describe the conducted literature work best, findings achieved, and what they are planning to do.

2.2. Keyword’s Identification

Keyword identification is the first step of PRISMA and was performed by an iterative search for specific keywords on the collected articles from the selected repositories. The final set of collected academic papers was determined by inserting the following logical query on their databases:

“Electric vehicle” AND “Predictive model” AND “Aging” AND” Degradation” AND “Battery” OR “SoH.”

2.3. Repositories

The search for keywords was performed on known academic repositories: Scopus, Institute of Electrical and Electronics Engineers (IEEE), and Web of Science.

IEEE is a repository introduced in 1952 and whose primary focus is the electrical and electronics engineering and computer science fields. It publishes roughly 200 peer-reviewed journals and more than 1,200 conference proceedings every year.

Web of Science is a research tool that allows access to repositories containing peer-reviewed papers about science, social science, arts, and humanities. It stores papers from the year 1900 to the present and encompasses 12,000 journals and 160,000 conference proceedings.

Scopus is a database of abstracts and article citations for academic journals. It covers approximately 19,500 titles from more than 5,000 international publishers, including 16,500 peer-reviewed journals in the scientific, technical, medical, and social sciences (including arts and humanities) fields. It is available on the web to subscribers. Searches on SciVerse Scopus incorporate scientific searches of web pages through Scirus, another Elsevier product, as well as patent databases.

The same query parameters mentioned in the previous sub-chapter were used on all three repositories to get relevant and comparable results.

2.4. Bibliometric Analysis

We end with a paper set for further quantitative and qualitative analysis: our SLR collection by applying PRISMA. This one was structured using the Mendeley reference manager tool [26] to extract papers' metadata and duplicated entries. The following metadata elements were extracted from each publication: author's name, number of publications, publication data, references, and number of citations.

2.5. Bibliometric Research Tool

We used VOSviewer [27] to map and visualize bibliometric networks of the SLR publications. This tool allowed us to identify network properties, such as clusters and node centrality, and derive characteristics of the SLR papers. These networks were built based on the number of common citations, bibliographic coupling, co-citations, and co-authorship relations, visually representing our scientific literature survey's bibliographic data.

2.6. Literature Review Results

2.6.1. PRISMA Flow Diagram

The PRISMA flow diagram in Figure 2-1 Illustrates our SLR process for further quantitative and qualitative analyses. In the first step, we identified the publications through a database search, using the logical query described previously, resulting in 149 publications (Scopus: 30; IEEE: 69; Web of Science: 50).

The two main factors for selection were papers written in English and published by peer-reviewed journals during the last five years, the 2017-2021 period. Additionally, we manually added 12 extra papers that proved to be relevant for this paper's scope. Finally, the research did not include review papers, conferences, position papers, and reports.

In the next step, we removed the exact duplicates. In this case, there were none (n = 0). Afterward, we performed vetting of the collected abstracts. In the first review, the methodology excluded articles from our research scope (n = 80). The second review excluded articles not related to prediction models (n = 4). Finally, the remaining 47 full texts were read, assessed, and fitted on the research scope. All papers were considered and eligible for a systematic review and analysis in the full-text screening phase. As such, this eligibility phase excluded a total of zero of such remaining papers.

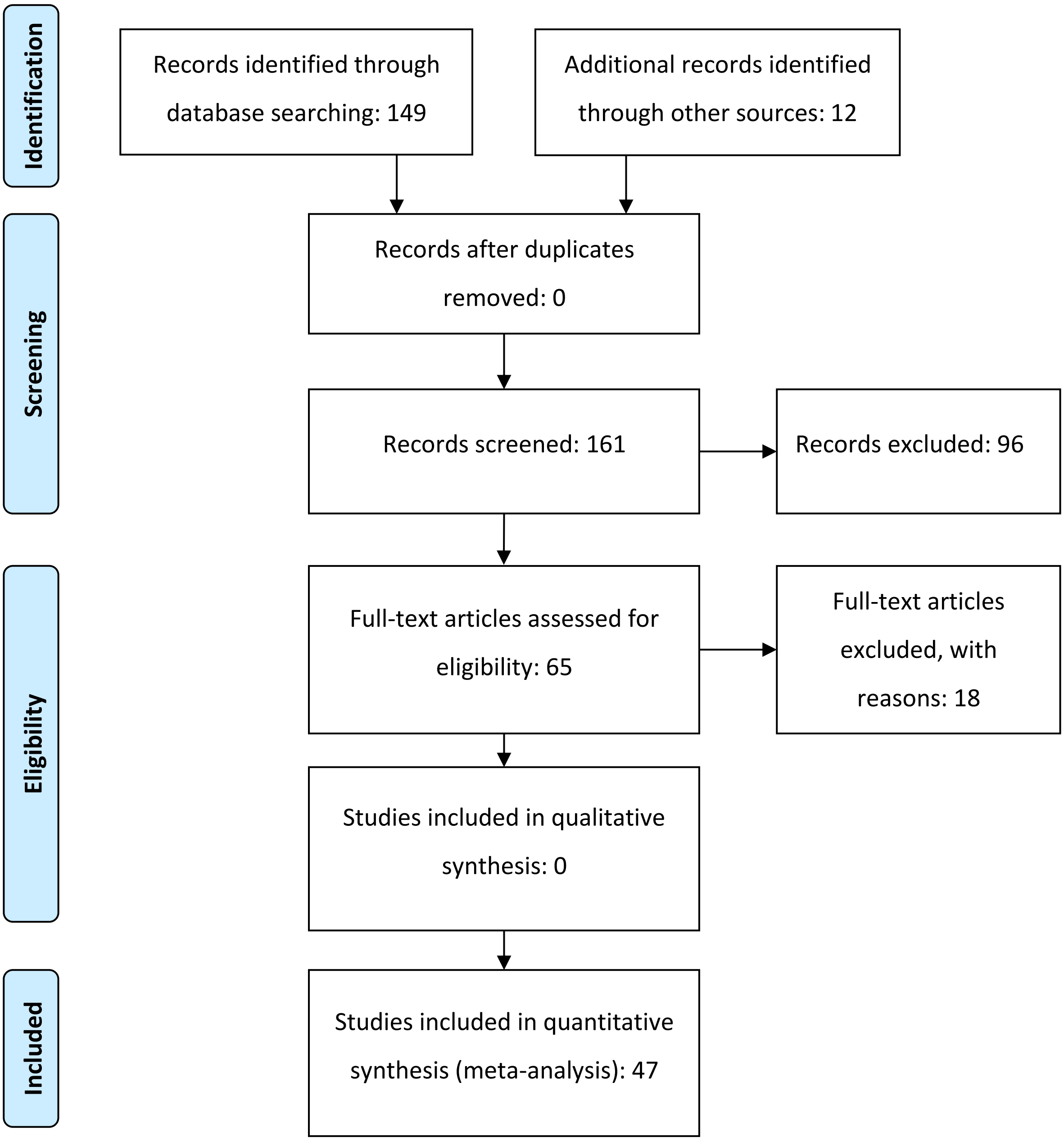

Figure 2-1 - PRISMA methodology flowchart.

The PRISMA flow diagram in Figure 2-1 illustrates our SLR process for further quantitative and qualitative analyses. In the first step, we identified the publications through a database search, using the logical query described previously, resulting in a total of 149 publications (Scopus: 30; IEEE: 69; Web of Science: 50).

2.6.2. Papers with Full-Text Reading

Based on the systematic literature review mentioned in the previous sub-chapter, a collection of 47 academic articles was arrived at and subjected to full-text reading and analysis. The themes were closely linked to the chosen keywords. Table 2-1 represents a list of the collected papers and the methods used.

Table 2-1 - PRISMA literature review results.

Title	Author	Method
Lifetime of Self-Reconfigurable Batteries Compared with Conventional Batteries [28].	Bouchhima, N., Gossen, M., Schulte, S., Birke, K.P.	Semi-empirical aging model.
Hybrid VARMA and LSTM Method for Lithium-Ion Battery State-of-Charge and Output Voltage Forecasting in Electric Motorcycle Applications [29].	Caliwag, A-C, Lim, W.	Neural Networking State of Charge (SoC) prediction method.
Predicting Life-Cycle Estimation of Electric Vehicle Battery Pack through Degradation by Self Discharge and Fast Charging [10].	Singh Ceng, M., Janardhan Reddy, K.	Battery pack SoC estimation of self-discharge simulation.
State of Health Estimation for Lithium-Ion Batteries Based on Fusion of Autoregressive Moving Average Model and Elman Neural Network [30].	Chen, Z, Xue, Q., Xiao, R., Liu, Y., Shen, J	Neural Network prediction method.
Lifecycle Comparison of Selected Li-Ion Battery Chemistries under Grid and Electric Vehicle Duty Cycle Combinations [31].	Crawford, A.J., Huang, Q., Kintner-Meyer, M.C.W., Zhang, J.-G., Reed, D.M., Sprenkle, V.L., Viswanathan, V.V., Choi, D.	Neural Network State of Energy prediction method.
State of Health Diagnosis and Remaining Useful Life Prediction for Lithium-Ion Battery Based on Data Model Fusion Method [32].	Cui, X., Hu, T	Neural Network prediction method.
Battery Health Prognosis Using Brownian Motion Modeling and Particle Filtering [33].	Dong, G., Chen, Z., Wei, J., Ling, Q.	Particle Filtering prediction model.

Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine [34].	Feng, X., Weng, C., He, X., Han, X., Lu, L., Ren, D., Ouyang, M.	SoH and SVM prediction model.
Co-Estimation of State of Charge and the State of Health for Lithium-Ion Batteries Based on Fractional-Order Calculus [13].	Hu, X S., Yuan, H., Zou, C F., Li, Z., Zhang, L	SoC and SoH estimation model.
Charging, Power Management, and Battery Degradation Mitigation in Plug-in Hybrid Electric Vehicles: A Unified Cost-Optimal Approach [35].	Hu, X., Martinez, C.M., Yang, Y.	Battery Management System (BMS) estimation model.
State Estimation for Advanced Battery Management: Key Challenges and Future Trends [36].	Hu, X S, Feng, F., Liu, K L., Zhang, L., Xie, J. L., Liu, B	BMS estimation model.
Bayesian Network-Based State-of-Health Estimation for Battery on Electric Vehicle Application and Its Validation Through Real-World Data [37].	Huo, Q., Ma, Z., Zhao, X., Zhang, T., Zhang, Y.	Bayesian Network SoH estimation model.
State of Health Estimation for Lithium-Ion Batteries Using Empirical Degradation and Error Compensation Models [14].	Jiang, Y., Zhang, J., Xia, L., Liu, Y.	SoH predictive model.
Batteries State of Health Estimation via Efficient Neural Networks with Multiple Channel Charging Profiles [38].	Khan, N., Ullah, F. U. M., Afnan, Ullah, A., Lee, M. Y., Baik, S. W.	SoH Neural Network prediction model.
Data-Driven State of Health Estimation of Li-Ion Batteries With RPT-Reduced Experimental Data [16].	Kim, J., Chun, H., Kim, M., Yu, J., Kim, K., Kim, T., Han, S.	SoH prediction model.
Reliable Online Parameter Identification of Li-Ion Batteries in Battery Management Systems Using the Condition Number of the Error Covariance Matrix [39].	Kim, M., Kim, K., Han, S.	SoH prediction model.
A Practical Lithium-Ion Battery Model for the State of Energy and Voltage Responses Prediction Incorporating Temperature and Ageing Effects [17].	Li, K., Wei, F., Tseng, K.J., Soong, B.H.	SoE predictive model.
State-of-Health Estimation for Li-Ion Batteries by Combing the Incremental Capacity Analysis Method with Grey Relational Analysis [40].	Li, X. Y., Wang, Z.P., Zhang, L., Zou, C.F., Dorrell, D.D.	SoH and RUL predictive model.
Lithium-Ion Battery State of Health Monitoring Based on Ensemble Learning [41].	Li, Y., Zhong, S., Zhong, Q., Shi, K.	Grey Relational Analysis model.
Optimal BP Neural Network Algorithm for State of Charge Estimation of Lithium-Ion Battery Using PSO with PCA Feature Selection [42].	Hossain Lipu, M.S., Hannan, M.A., Hussain, A., Saad, M.H.M.	SoC estimation Back-Propagation Neural Network model.
A Review of State of Health and Remaining Useful Life Estimation Methods for Lithium-Ion Battery in Electric Vehicles: Challenges and Recommendations [43].	Lipu, M.S.H., Hannan, M.A., Hussain, A.,	SoH prediction model.

	Hoque, M.M., Ker, P.J., Saad, M.H.M., Ayob, A.	
An On-Line State of Health Estimation of Lithium-Ion Battery Using Unscented Particle Filter [18].	Liu, D., Yin, X., Song, Y., Liu, W., Peng, Y.	Neural Networking SoH prediction model.
Modified Gaussian Process Regression Models for Cyclic Capacity Prediction of Lithium-Ion Batteries [44]	Liu, K., Hu, X., Wei, Z., Li, Y., Jiang, Y.	Gaussian Process Regression model.
Remaining Useful Life Prediction of Lithium-Ion Battery Based on Gauss-Hermite Particle Filter [45].	Ma, Y., Chen, Y., Zhou, X W., Chen, H.	RUL and SoH prediction model.
Battery-Degradation Model Based on the ANN Regression Function for EV Applications [9].	May, G., El-Shahat, A.	Neural Network-based prediction model.
Lithium-Ion Batteries Health Prognosis Considering Aging Conditions [19].	El Mejdoubi, A., Chaoui, H., Gualous, H., Van Den Bossche, P., Omar, N., Van Mierlo, J.	RUL predictive model.
A Design-Based Predictive Model for Lithium-Ion Capacitors [46].	Moye, D.G., Moss, P.L., Chen, X.J., Cao, W.J., Foo, S.Y.	A predictive model of Capacitors.
A Neural-Network-Based Method for RUL Prediction and SoH Monitoring of Lithium-Ion Battery [22].	Qu, J., Liu, F., Ma, Y., Fan, J.	Neural Network-based model.
Empirical Electrical and Degradation Model for Electric Vehicle Batteries [8].	Saldaña, G., Martín, J.I.S., Zamora, I., Asensio, F.J., Oñederra, O., González, M.	Degradation model based on actual LG battery cell.
The Co-Estimation of State of Charge, State of Health, and State of Function for Lithium-Ion Batteries in Electric Vehicles [47].	Shen, P., Ouyang, M.G., Lu, L.G., Li, J.Q., Feng, X. N.	SoH predictive model.
Real-Time State-of-Health Estimation of Lithium-Ion Batteries Based on the Equivalent Internal Resistance [12].	Tan, X., Tan, Y., Zhan, D., Yu, Z., Fan, Y., Qiu, J., Li, J.	SoH and SoC predictive model.
A Health Monitoring Method Based on Multiple Indicators to Eliminate Influences of Estimation Dispersion for Lithium-Ion Batteries [48].	Tang, J., Liu, Q., Liu, S., Xie, X., Zhou, J., Li, Z.	SoH estimation of Lithium-ion batteries.
Fractional-Order Model-Based Incremental Capacity Analysis for Degradation State Recognition of Lithium-Ion Batteries [49].	Tian, J.P., Xiong, R., Yu, Q. Q	SoH predictive model.
Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art [23].	Vidal, C., Malysz, P., Kollmeyer, P., Emadi, A.	ML-based prediction model.
State-of-Health Estimation for Lithium-Ion Batteries Based on the Multi-Island	Wang, Z., Ma, J., Zhang, L.	SoH estimation of Lithium-ion batteries.

Genetic Algorithm and the Gaussian Process Regression [50].		
Remaining Useful Life Prediction and State of Health Diagnosis for Lithium-Ion Batteries Using Particle Filter and Support Vector Regression [20].	Wei, J.W., Dong, G.Z., Chen, Z.H.	RUL and SoH estimation model.
State of Health Estimation for Lithium-Ion Batteries Based on Healthy Features and Long Short-Term Memory [51].	Wu, Y., Xue, Q., Shen, J., Lei, Z., Chen, Z., Liu, Y.	Neural Network SoH estimation model.
State-of-Health Prognosis for Lithium-Ion Batteries Considering the Limitations in Measurements via Maximal Information Entropy and Collective Sparse Variational Gaussian Process [52].	Xiang, M., He, Y., Zhang, H., Zhang, C., Wang, L., Wang, C., Sui, C.	Neural Network SoH estimation model.
State-of-Health Estimation for Lithium-Ion Batteries Based on Wiener Process with Modeling the Relaxation Effect [53].	Xu, X., Yu, C., Tang, S., Sun, X., Si, X., Wu, L.	SoH estimation of Lithium-ion batteries.
Novel Lithium-Ion Battery State-of-Health Estimation Method Using a Genetic Programming Model [54].	Yao, H., Jia, X., Zhao, Q., Cheng, Z., Guo, B.	SoH estimation model.
Remaining Useful Life Estimation of Lithium-Ion Batteries Based on Optimal Time Series Health Indicator [11].	Yun, Z, Qin, W.	Bayesian Monte Carlo prediction model.
Capacity Prognostics of Lithium-Ion Batteries Using EMD Denoising and Multiple Kernel RVM [55].	Zhang, C., He, Y., Yuan, L., Xiang, S.	Monte Carlo prediction model.
Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Exponential Model and Particle Filter [56].	Zhang, L., Mu, Z., Sun, C.	RUL prediction model.
Lithium-Ion Battery Remaining Useful Life Prediction with Box-Cox Transformation and Monte Carlo Simulation [57].	Zhang, Y., Xiong, R., He, H., Pecht, M.G.	RUL prediction model.
Hybrid Lithium Iron Phosphate Battery and Lithium Titanate Battery Systems for Electric Buses [25].	Zhang, X., Peng, H., Wang, H., Ouyang, M.	Novel hybrid battery system accounting for behavior.
State-of-Health Prediction for Lithium-Ion Batteries with Multiple Gaussian Process Regression Model [58].	Zheng, X., Deng, X.	SoH Grey Relational Analysis.
State of Health Monitoring and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Temporal Convolutional Network [59].	Zhou, D., Li, Z., Zhu, J., Zhang, H., Hou, L.	RUL prediction model.

2.6.3. Identification of Research Themes

The identified journals papers covered a broad spectrum of research fields ranked by importance: Engineering, Energy, Computer Science, Materials Science, Chemistry, Physics, and Mathematics, as shown in Table 2-2. Again, this demonstrates a broad interest in this dissertation subject across many fields of research.

Table 2-2 - Literature review by Publishers and Ranks.

Journals	No.	Rank	Publisher	Country	Field
IEEE Access	25	Q1	Institute of Electrical and Electronics Engineers Inc.	United States	Computer Science
IEEE Transactions on Industrial Electronics	5	Q1		United States	Computer Science and Engineering
IEEE Transactions on Vehicular Technology	4	Q1	Institute of Electrical and Electronics Engineers Inc.	United States	Aerospace Engineering, Applied Mathematics, Automotive Engineering, Electrical and Electronic Engineering
Journal of Power Sources	3	Q1	Elsevier	Netherlands	Chemistry, Energy, and Engineering
Journal of Energy Storage	3	Q1	Elsevier	Netherlands	Energy and Engineering
IEEE Global Humanitarian Tech. Conference	1			United States	Business, Engineering, Management, and Accounting
IEEE Transactions on Control Systems Technology	1	Q1	Institute of Electrical and Electronics Engineers Inc.	United States	Engineering

IEEE Transactions on Transportation Electrification	1	Q1	Institute of Electrical and Electronics Engineers Inc.	United States	Energy, Engineering and Social Sciences
Journal of Cleaner Production	1	Q1	Elsevier Ltd.	Netherlands	Business, Energy, Engineering and Environmental Science
Mechanical Systems and Signal Processing	1	Q1	Academic Press Inc.	United States	Computer Science, Engineering
Renewable and Sustainable Energy Reviews	1	Q1	Elsevier Ltd.	Netherlands	Energy

From an initial collection of 149 papers, 47 journal papers were analyzed, including IEEE Access (25) and IEEE Transactions on Industrial Electronics (5). As shown in Table 2-2, most journals are ranked as Q1-quartile (45), representing 96%, while the remaining (2) are Q3 articles. The five primary areas of expertise identified in the analysis were Computer Science, Engineering, Environmental Science, Transportation, and Mathematics.

The 46 selected articles' publishers originate from two countries, with the most extensive set coming from the United States (40), followed by the Netherlands (7). The top publishers identified are Elsevier Ltd. (6), Institute of Electrical and Electronics Engineers Inc. (3), the American Institute of Physics (1), Academic Press Inc. (1), and SAE International (1).

2.7. Network Analysis and Visualization

The most frequent keywords used in the titles and abstracts of the articles collected are represented in Figure 2-2. Thus, it is possible to assess three distinct groups, in three assorted colors, related to different subjects: in red, we show words linked with the theme of forecasting models; in green, the keywords related to the second model that refers to the theme of electric batteries, and in blue color, the third model associated with the temperature conditions under which the batteries operate. In this case, the most relevant topic present in the literature was the “model.”

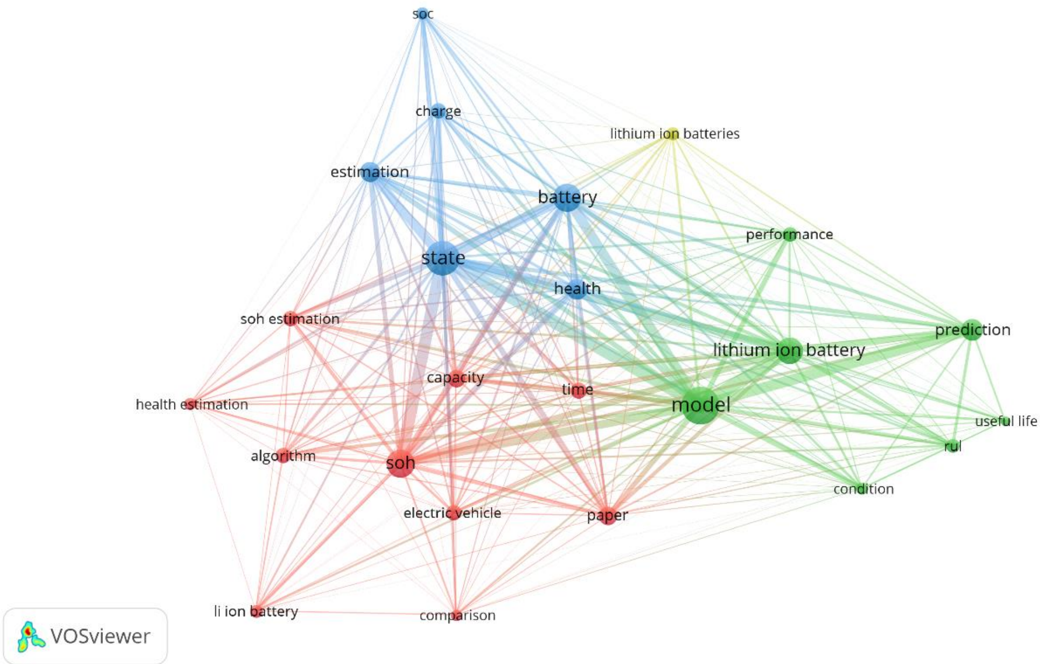

Figure 2-2 – Keywords collected from Title and Abstract fields by the literature review.

In Figure 2-3, we depict the paper´s authors. To have a fitted representation of the authors, we did an initial visualization with all 1826 authors. However, the visualization of all the authors showed a too dense and complex network, making it hard to visualize. Hence, we selected authors with a minimum of 5 published works, resulting in 60 authors. This analysis showed that the scientific community is mostly of Chinese origin. Therefore, no filtering on nationality was applied in the papers screening.

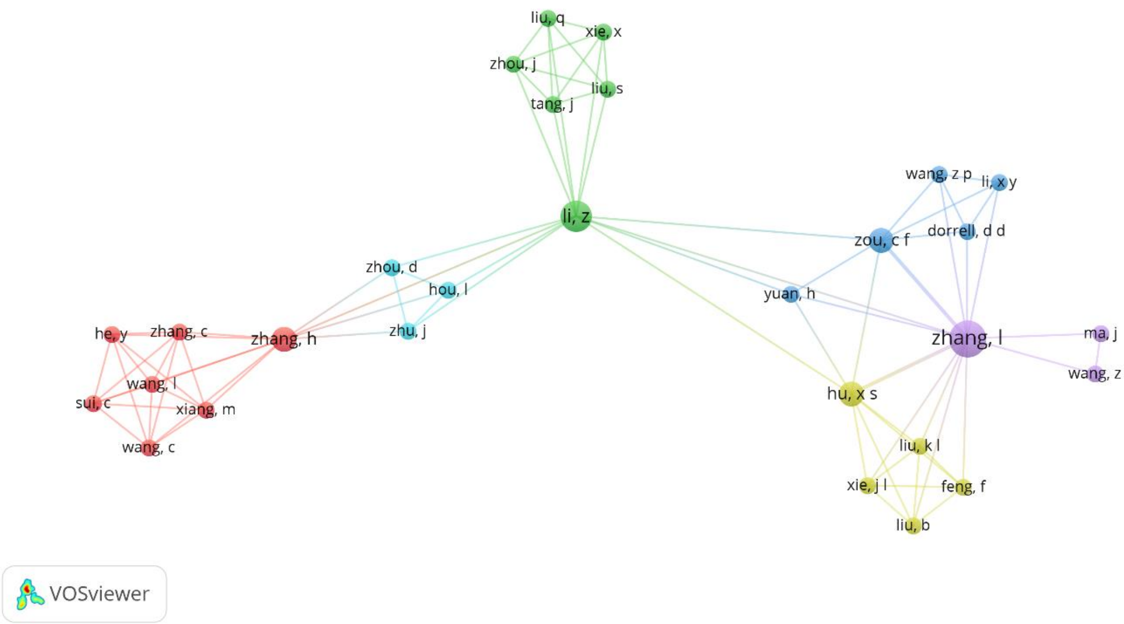

Figure 2-3 - Most central authors in the literature review.

Figure 2-4 shows the different themes present in the text of the consulted articles. There were four distinct themes, highlighted by different colors- a less prevalent theme linked to Electronics. There was a second cluster, a slightly more important theme, colored in blue on the theme of ML; the third theme colored in green focuses on EVs; and finally, a fourth theme displayed in red on lithium batteries.

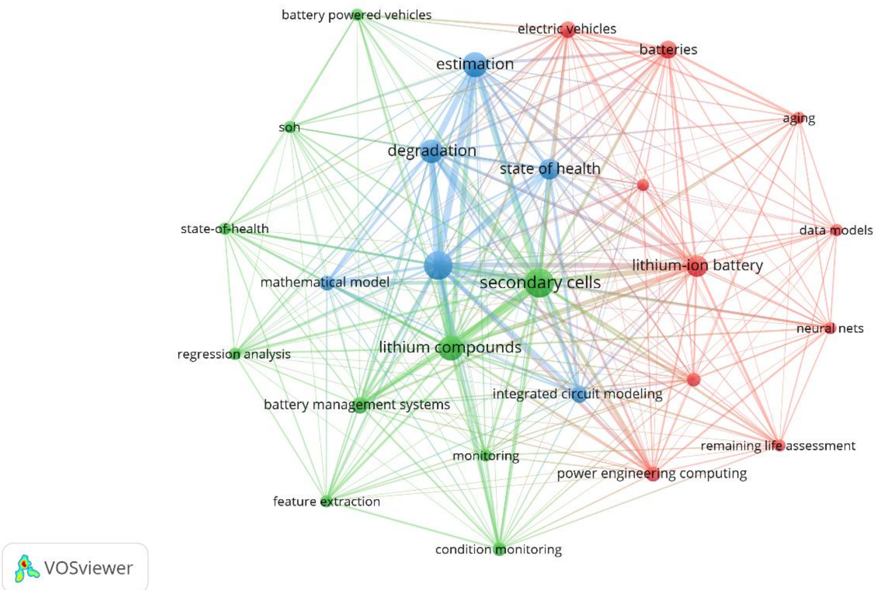

Figure 2-4 - Keywords co-occurrence in the literature review.

In Table 2-3, the most important contributors are listed, coupled with their quartile rank and the insight they provided to this paper:

Table 2-3 - List of most important papers and their specific contributions.

Author	Title	Rank	Contributions
Qu, J., Liu, F., Ma, Y., Fan, J.	A Neural-Network-Based Method for RUL Prediction and SoH Monitoring of Lithium-Ion Battery [22].	Q1	• Data-driven methods are faster and less complex than model-based methods. • ML methods can improve model accuracy.
Saldaña, G., Martín, J.I.S., Zamora, I., Asensio, F.J., Oñederra, O., González, M.	Empirical Electrical and Degradation Model for Electric Vehicle Batteries [8].	Q1	• Aging depends on the level of electrical current, the depth of discharge, and the number of cycles made. • Temperature, current level, and cycles are the variables that have the most significant impact on battery degradation. • The driving environment is also relevant to battery degradation.
Singh Ceng, M., Janardhan Reddy, K.	Predicting Life-Cycle Estimation of Electric Vehicle Battery Pack through Degradation by Self Discharge and Fast Charging [8].	Q1	• Rising need to increase the working life of battery packs used in electric vehicles.
Vidal, C., Malysz, P., Kollmeyer, P., Emadi, A.	Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art [23].	Q1	• AI and ML have actively contributed to an increase in research and development of new methods to estimate the states of EVs. • Few studies focus on SoC and SoH at negative temperatures.
Zhang, X., Peng, H., Wang, H., Ouyang, M.	Hybrid Lithium Iron Phosphate Battery and Lithium Titanate Battery Systems for Electric Buses [25].	Q1	• Novel hybrid battery system. • Accounts for driving behavior and charging patterns.

The most significant contributions retained from this literature review were from authors, such as Zhang [25] that proposed a hybrid model of electric buses powered by one Lithium Iron Phosphate (LPF) battery and one Lithium-Titanate-Oxide (LTO) battery. When coupled together and under strict temperature and voltage controls, these battery types would enable the charging of buses in 20 minutes at fast-charging stations. In addition, Zhang concluded that buses built with this hybrid battery pack would suffer a maximum decay of 20% of their total capacity after eight years of operation until they no longer remain viable for service. Finally, he also accounted for behavioral factors like charging/discharging cycles in his model [13].

Vidal [23] reinforced the importance of applying ML techniques to precisely ascertain the batteries' SoH. Additionally, he frequently mentioned very few studies about the battery pack performance under cold temperatures. This finding is of notable consequence in, for instance, in colder regions and, to a lesser degree, vehicles directly exposed to the exterior elements (i.e., parked outside).

Saldaña [8] and Vidal [23] also alluded to the essential nature of knowing the rate of battery degradation. Furthermore, Saldaña stated that battery endurance depends not only on electro-chemical factors such as temperature, the degree to which EV users let their packs discharge but also on the total number of charging/discharging cycles completed throughout the useful life of the car. Finally, he mentioned that the driving environment where the car is used is of great relevance to determine how well the batteries will mitigate their decay.

3. Data Analytics – CRISP-DM

The Cross-Industry Standard Process for Data Mining (CRISP-DM) [3] is a methodology widely used by data science specialists to develop solutions for business problems based on data. CRISP-DM can be understood as a cross-industry standard process for data mining. It can transform a company's data into management information and knowledge, appropriate for decision making.

This methodology was created over 20 years ago due to the need for Data Mining professionals. Although several tools can guide these professionals, they fell short in Big Data and its large data volume requirements. Thus, CRISP-DM emerged precisely to meet the projects directly involved with processing and analyzing a large volume of data.

Data Mining is part of Data Science, which uses statistics and mathematics as a basis for crossing and correlating data, using induction techniques to propose assumptions and solve business issues. The CRISP-DM methodology gathers the best practices so that the DM is as productive and efficient as possible, analyzing financial data, human resources, production, customer habits, and other data sources to propose models for business improvement or problem-solving. It defines a project's life cycle, dividing it into the six stages shown below in

Figure 3-1, which follows a linear progression.

It is essential to emphasize the theoretical character of this thesis. The application of the CRISP-DM phases is therefore limited. For this reason, there are phases of this methodology that are less explored than others.

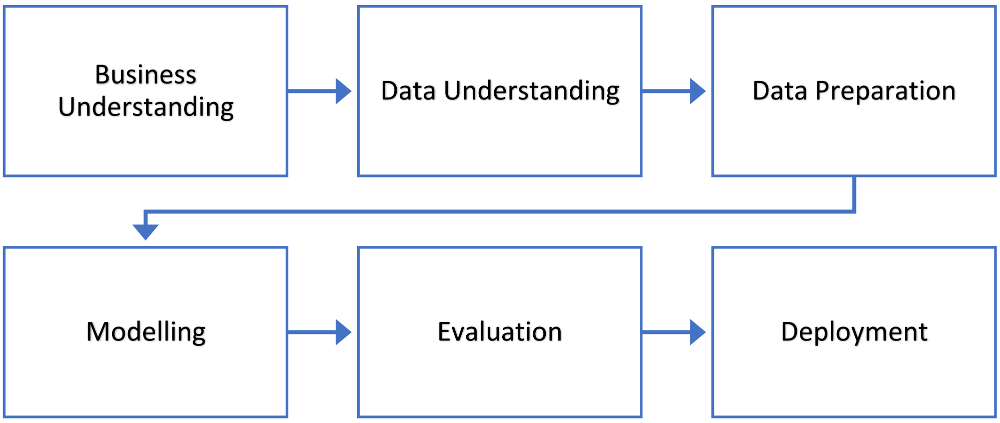

Figure 3-1 - CRISP-DM methodology flowchart.

3.1. Business Understanding

The first phase of CRISP-DM aims to understand the entity's primary needs and business requirements where the methodology will be implemented. It looks for all the details about its internal organization, terminology, marketing strategies, target audience, and available products. The conclusion of this step defines guidelines for the steps that will follow, such as the selection, cleaning, and interpretation of the information retrieved for the implemented data mining project. In the case of this study, the final client is an individual electric car driver. For example, a potential EV car buyer needs estimates of what an EV car's range could be. The EV car buyer could be interested in predicting the range of a car on a full charge based on its attributes. More precise, he could need the answer to the following business questions:

- Is the battery degrading at a regular or abnormal rate?
- Which factors had the most impact on said degradation? (RQ1)

3.2. Data Understanding

The second step of CRISP-DM consists of organizing and documenting all the available data sources relevant to the institution or client. This documentation implies the identification of a target audience and the selection of sources of data. This stage is an iterative process that switches between searching for data sources and the data essential for their selection. It is expected to obtain an extensive dataset with the potential of obtaining meaningful information from EV users and about their vehicles. Ideally, it would reach a diversity of responses high enough to ensure a comprehensive analysis of the business in focus.

However, in this thesis, the original dataset was obtained through a single source. It came from an international news blog called Elektrek [5]. This blog shares a dataset compiling answers to a survey from a forum of Tesla users, who registered their range entries and other data in an excel spreadsheet, collecting a total of 1425 observations, structured in 43 variables.

3.3. Data Preparation

The data preparation phase aims to transform the collected data into clean, structured, and integrated data. To this aim, we developed procedures with the Python programming language [60] using the Jupyter Notebook tool [61]. Our data preparation included the following steps:

Matching variable formats: normalizing dataset variables with formats (e.g., dates, distance units in the Imperial system). Figure 3-2 represents all the variables from the Tesla dataset. Some of them in object format (mostly text ones), which had to be transformed from their original format to a numerical form, so this step processes the variables across all observations.

•

```
Location                              object
Model                                 object
Mileage in miles                     float64
Mileage per day                      float64
EPA rated range                      float64
RNG Mode On/Off                      float64
EPA range Mode off                   float64
Battery rplc Y/N                      object
Batt mileage after rplc              float64
Batt days after rplc                 float64
Total avg energy C                   float64
Rated range when new                 float64
Remaining original range             float64
Wh cap until range is zero           float64
Freq SCHG                             object
Freq 100%                             object
Freq empty                            object
Daily charge level                   float64
Daily charge power in W              float64
100% range when new                  float64
Range mode on/off previous reading   float64
Vehicle age (days)                   float64
Cycles                               float64
Total Km                             float64
Wh/mi to Wh/Km                       float64
Avg cap all cars at this mileage     float64
Cap below trendline                  float64
dtype: object
```

Figure 3-2 - Tesla dataset variables listed by their data format.

Elimination of blank records: some of the observations from the dataset had missing fields. Figure 3-3 shows a visual representation of each variable's number of null values. The higher the grey bars, the fewer null values its variable has. Given that these missing variables could create bias, leading to wrong conclusions, all responses with missing variables had to be excluded from the dataset for further analysis.

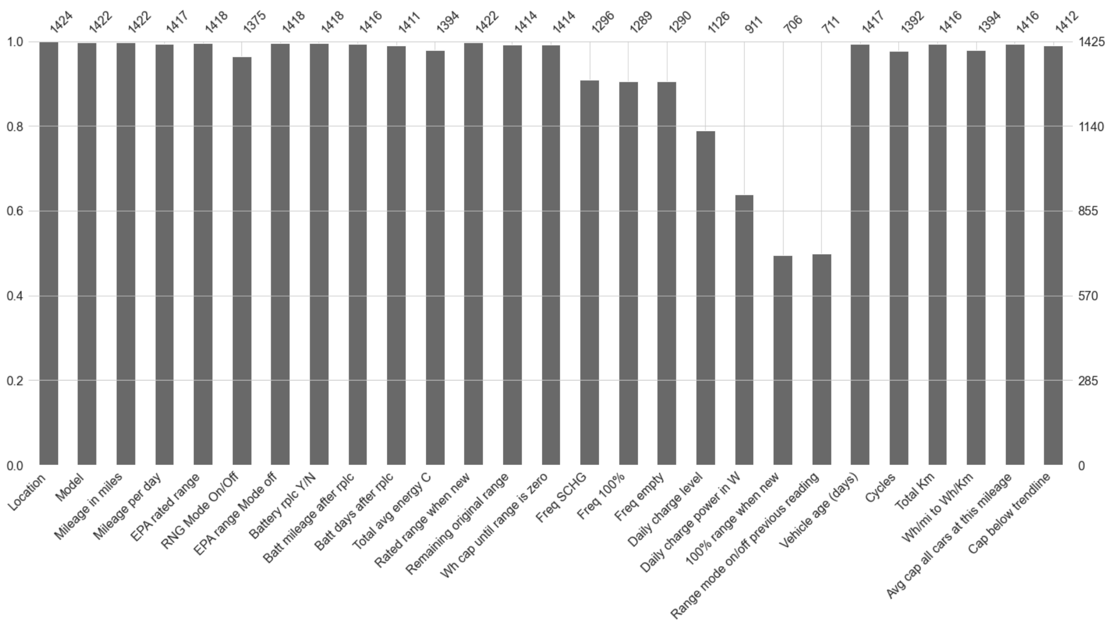

Figure 3-3 - Total amount of null values per variable, in percentage.

Elimination of outliers: when existing variable values were too far from the remaining observations, they were considered outliers and needed to be removed. The method used to detect outliers was based on percentiles. With the percentile's method, all data variables outside an interval formed by the 5th and 95th percentiles were considered potential outliers and removed. Figure 3-4 shows an example of detected outliers for the charging frequencies variables, where the deviating values are shown in red.

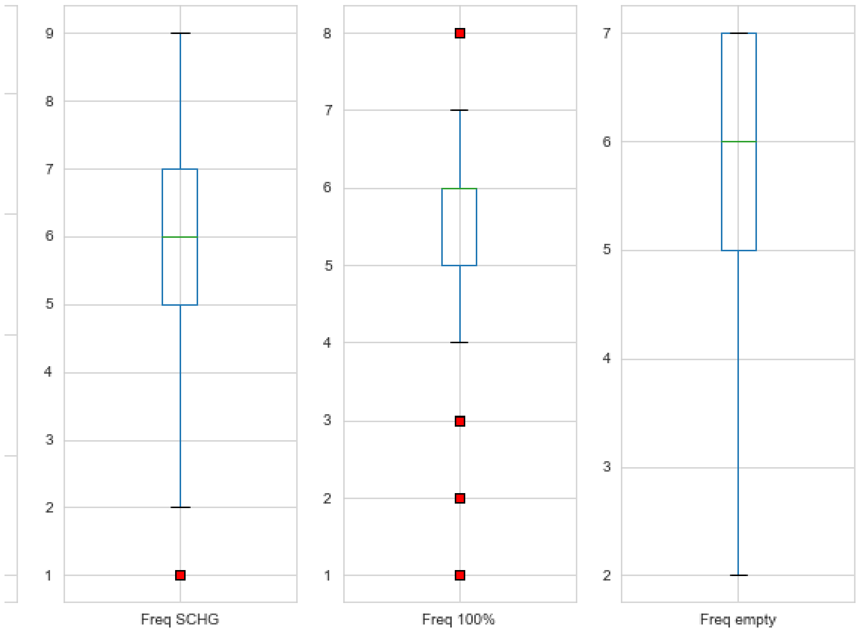

Figure 3-4 - Example of outliers' identification of the three charging frequency variables.

Removal of unnecessary variables: the original dataset (with 43 variables) has unnecessary variables for this study. Table 3-1 below lists all the variables present in the Tesla dataset, as collected, before any data cleaning operations, their data type, exclusion status from the study, and the reason behind the exclusion.

Table 3-1 - Listing of all excluded Tesla dataset variables.

Variable	Type	Excluded	Reason
Username	String	Yes	No other demographic information.
Location	String	No	The dataset was too unbalanced
Vehicle manufacture date	Date	No	
Date of range reading	Date	Yes	No insights were found.
Model	String	No	

Mileage in miles	Int	No	
Mileage per day	Float	Yes	A duplicate variable of Mileage in miles
EPA rated range at 100% charge in miles	Int	Yes	No correlations were found with other variables.
Range mode on/off at the time of reading?	Boolean	Yes	No correlations were found with other variables.
EPA range after correction if range mode was off	Float	Yes	No correlations were found with other variables.
Did you have a battery replacement?	Boolean	Yes	Very few vehicles had their battery replaced.
What happened to the EPA range after replacement?	String	Yes	Excessive number of null values.
At what miles did you replace the battery?	Int	Yes	Excessive number of null values.
Mileage in mi after correction if the battery was replaced	Int	Yes	Dependent on battery replacement.
Battery age (days) after correction if the battery was replaced	Int	Yes	Dependent on battery replacement.
Lifetime average energy consumption at the time of reading Wh/mi	Int	Yes	Irrelevant for this study.
Rated range of this model when new	Int	Yes	Like variable *Mileage in miles*.
Remaining original range	Float	Yes	Variable replaced by *Average Capacity*
Remaining usable Wh capacity until typical range shows zero		Yes	
Unanswered questions	Int	Yes	Majority of no answers.
Frequency of supercharging	String	No	
Frequency of 100% charge	String	No	
Frequency of almost empty (5mi or less)	String	No	
Daily charge level	Float	Yes	Excessive number of null values.
Daily charge power in watts	Float	Yes	Excessive number of null values.
What was the 100% rated range when the car was new?	Int	Yes	Excessive number of null values.
Range mode on/off at the time of reading the previous column?	Boolean	Yes	Excessive number of null values.
Rated range at the beginning of the trip	Int	Yes	Excessive number of null values.
Rated range at the end of the trip	Int	Yes	Excessive number of null values.
Consumption for this trip	Float	Yes	Excessive number of null values.

Range mode on/off when reading these trip numbers?	Boolean	Yes	Excessive number of null values.
Typical range consumption for the trip	Float	Yes	Excessive number of null values.
Typical range after correction if range mode was off	Int	Yes	Excessive number of null values.
Remaining usable capacity until typical range shows zero according to trip data	Float	Yes	Excessive number of null values.
Remaining original capacity	Float	Yes	Excessive number of null values.
Trip based battery capacity calculation explained	Float	Yes	Excessive number of null values.
100% range when the car was new after range mode adjustment	Int	Yes	Dependent on range mode.
Vehicle age (days)	Int	No	
Cycles	Int	No	
Mileage in miles	Int	No	
Wh/mi to wh/km	Float	Yes	The unit conversions were made in Python and SPSS.
Average capacity of all cars at this mileage according to chart trendline	Float	Yes	Irrelevant for this study.
Your capacity minus chart trendline at this mileage	Float	Yes	Irrelevant for this study.

The next step was the merging of the three surveys into a single dataset. As seen in Figure 3-5, most vehicles from the sample traveled a few kilometers because the vast majority have a range below 100,000 kilometers. This occurrence aligns with the fact that most vehicles in the sample are less than ten years old. Thus, it is no surprise that the least populated group of vehicles are the vehicles with more range, in this case, more than 300 thousand kilometers. In contrast, the most numerous vehicles are precisely the group of vehicles with less range, meaning those whose range is below fifty thousand kilometers.

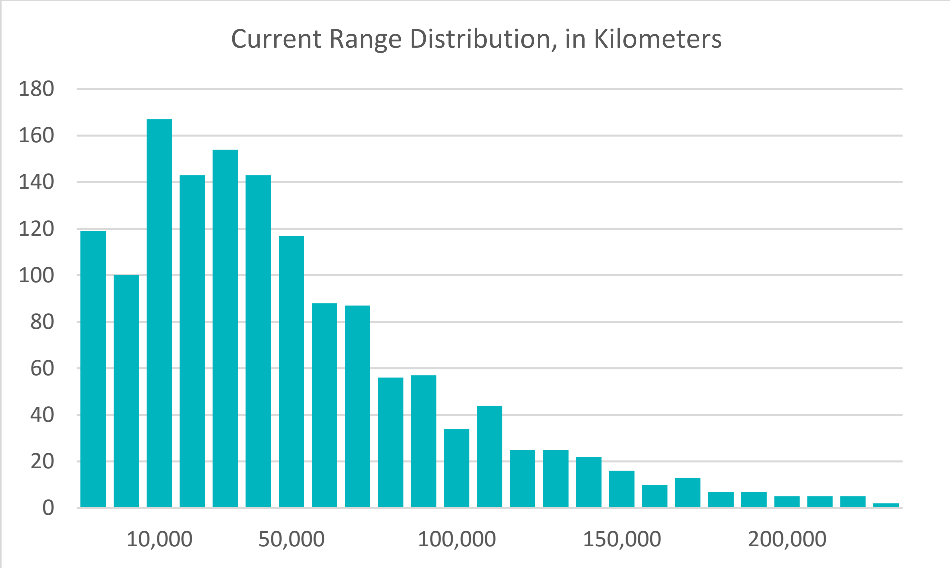

Figure 3-5 - Car sample range distribution.

Considering the vehicles' maximum range distribution, it follows a normal distribution (Figure 3-6). Most of the vehicles in this sample have a range between three hundred and eight hundred kilometers. This significant autonomy is explained as most of the vehicles in this sample are Tesla's Model S, a specific model with greater autonomy than most electric vehicles currently on sale.

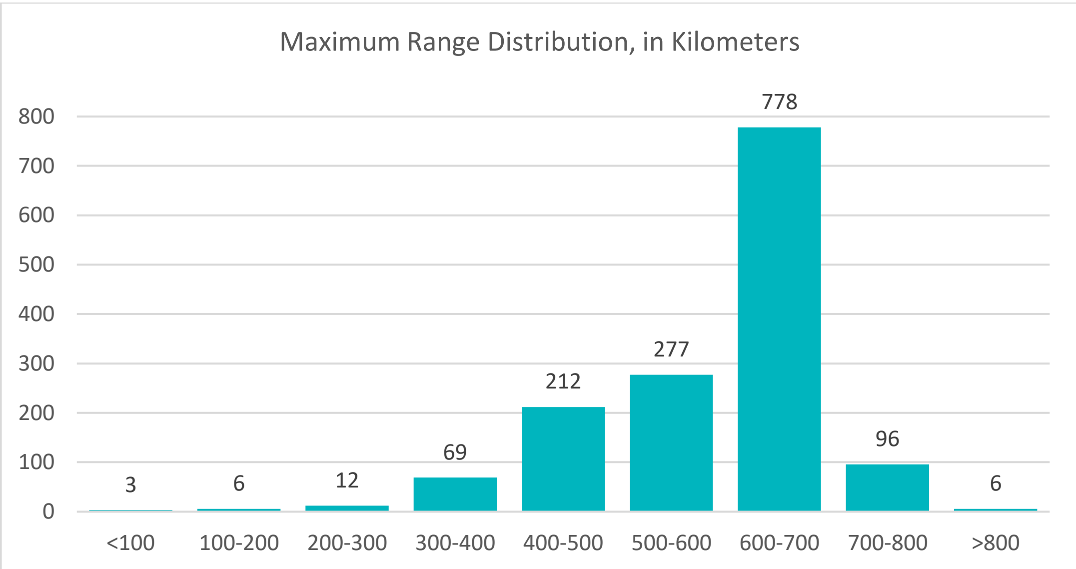

Figure 3-6 - Car maximum range distribution.

Regarding the distribution of vehicles in the sample by country, the original dataset disproportionately represents vehicles from Asia and the Pacific. They are mainly vehicles from mainland China. The three surveys conducted aimed to add new vehicles to the sample and thus obtain vehicles from other continents/countries with different ages and ranges. However, due to the low participation in the survey, this objective was not successfully achieved. Figure 3-7 below shows the sample composition by country.

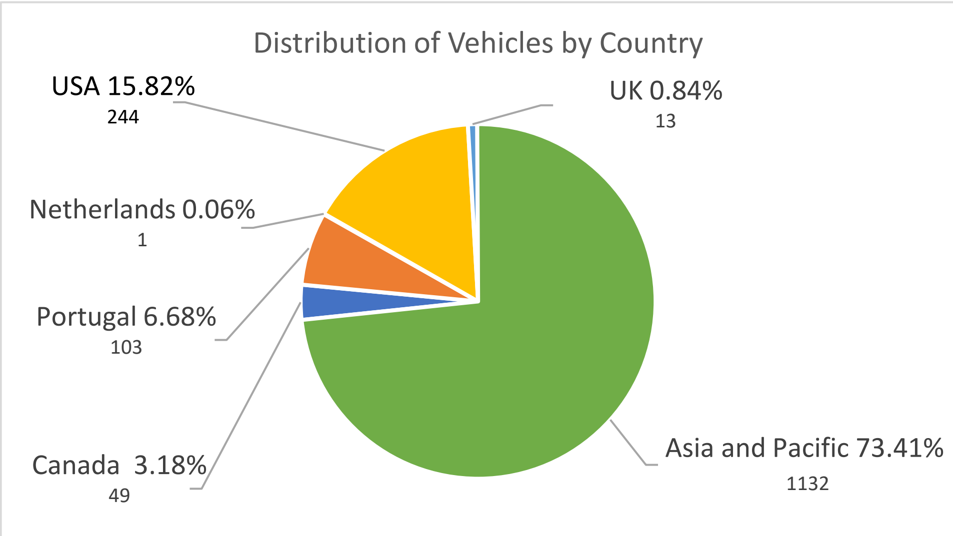

Figure 3-7 - EV vehicles distribution by country.

In Figure 3-7, it is possible to see that the surveys carried out managed to add 103 observations to the dataset for the country of Portugal. However, this number was small due to the necessary data processing that included, for instance, the removal of outlier values and duplicates.

After processing the data from the three surveys, we merged the statistical analysis about the vehicle distribution according to their range.

3.4. Modeling

In this phase, ML techniques were applied to the Tesla dataset to understand how several factors affected the range of the EVs present in the dataset. It was necessary to take the following steps to know which variables affect a car maximum range with a full charge:

Classification analysis: Several ML supervised classification algorithms from the Scykit-learn package [62] were employed to classify which of the included variables influenced the cars' maximum range on a full charge, addressing the first research question (RQ1). The objective of these classifiers was to label the current range that each car had against the original range value announced by their respective manufacturers. A new discrete variable called "*Degradation*" was created, containing two labels: "Normal" and "Abnormal." These two labels represented the batteries' degree of energy capacity loss in a percentage of the original maximum range. Degradation levels lower than 10% were labeled as "normal" and higher than 10% as "abnormal." The 10% threshold was used based on the

Elektrek article claiming that 10% was the average capacity degradation of EVs after 160,000 miles [5]. The following ML classification models were chosen to perform the labeling task:

- K-Nearest Neighbours (KNN) [63]
- Logistic Regression [64]
- Naïve-Bayes [65]
- Support Vector Machine - Linear (L-SVM) [66]
- Support Vector Machine - Radial (R-SVM) [67]

These models were selected because they are low-complexity models, and as such, they should generalize better when dealing with small datasets. Since the Tesla dataset is a small dataset with just 1425 observations, the decision boundary of complex models such as Decision Trees or the Random Forest would change wildly. As a result, the results from those more complex models would have high degrees of variance. Simpler models such as those chosen above perform better as they have more minor degrees of freedom.

Following best practices from the literature [68], each classification model was preceded by a split of the original dataset into a training set and a test set (the split chosen was 80% - 20%, respectively) so that fitted models could be evaluated, regarding their performance and compared using a confusion matrix.

Cross-Validation: An iterative cross-validation technique [69], employed to compare model performance, obtain the best model, and avoid overfitting [70]. Overfitting happens when a model obtains near-perfect scores after being trained with known training and testing sets but cannot make accurate predictions when using new data. In our case, the training set was split into ten smaller equal sets. A model was trained using nine sets as training data and judged its results against the tenth set. Then, a loop was created to switch the testing set between all ten sets. The average of the values computed in the loop reports the global performance of the model.

3.5. Evaluation

The evaluation phase aims to evaluate the results of the labeling process defined in sub-chapter 3.4. First, the cars' current range was checked to see if the vehicles retained more or less than 10% of the initial total battery capacity. Afterward, the classification models and the cross-validation technique labeled the vehicles with an abnormal or regular decline. Finally, the factors that had the most negative impact on the batteries were pinpointed, answering research question 1 (RQ1). Generally, the models' performance presented accuracy results ranging between 57% and 62%.

Table 3-2 shows the models' accuracy and their standard deviation to classify the degradation rate of the vehicles. The accuracy values were calculated using the formula shown in Equation 1, which is the sum of true positive (TP) and true negative (TN) values, divided by the sum of true positive, true negative, false positive (FP), and false-negative (FN) values.

Table 3-2 - Training Accuracy results from Cross-validation.

Classifiers	Median	Standard Deviation
K-Nearest Neighbours	0.57	0.07
Logistic Regression	*0.69*	*0.11*
Naïve-Bayes	0.59	0.08
L-SVM	0.59	0.11
R-SVM	*0.62*	*0.11*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 1 - Accuracy formula.

The two best-fitted models identified using cross-validation are highlighted in italic. The key takeaway from the cross-validation evaluation results was the Logistic Regression and the Radial Support Vector Machine (Radial SVM). These were the two best performing models from the group, having reached 62.84 and 61.75% median accuracy, respectively, for the labeling task. Moreover, it demonstrates that from the 500 observations (vehicles), the best classification model, the Logistic Regression, correctly guessed 62.84% of the degradation labels.

At this point in the analysis, each independent variable's weight over the dependent remains unknown to answer research question number one (RQ1). However, Table 3-3 below does just that, showing both the weight and significance of each variable.

Table 3-3 - Degradation - Coefficients of the independent variables.

Degradation	Regression Coefficient	Chi-Squared	p-Value
Country	0.384	4.52	0.052
Year	2.191	12.6	0.037
Maker	0.137	0.03	0.1839
Model	0.012	0.01	0.982
Freq. Fast	-0.02	0.90	0.814
Freq. Full	0.09	0.55	0.555
Freq. Empty	0.027	0.28	0.370
Mileage	.0586	0.10	0.627
Max Range	0.061	0.88	0.837

From the table above, it is possible to conclude that there were two predominant variables: the *Year* and *Country*. Furthermore, these two predictors seem to have positive and strong correlations with the dependent variable Degradation. Lower values for the *Year* variable correspond to older vehicles, and an “Abnormal” value of the *Degradation* variable, meaning a higher than usual degrading of the battery.

Although insignificant, there seems to be a correlation regarding the variable Country, possibly due to this dataset's imbalance. It contains a large volume of vehicles from Asia and the Pacific regions. However, it would be necessary to have more data to draw more elaborate conclusions to explain this phenomenon. As for the remaining variables listed, there seems to be no relationship between them and *Degradation*.

In conclusion, the variable variables *Year* was the most impactful factor on the rate of degradation. However, none of the behavioral factors were significant, which is the answer we could reach to answer the RQ1 (Which behavioral habits from the electric vehicles may negatively impact lithium-ion battery capacity?). For this reason, we expand the dissertation in Chapter 4 to include another methodology to try to find new insights with new data.

3.6. Deployment

In the deployment phase, knowledge extracted from the data is delivered and applied. From this moment, the processes within the organization might be changed or new products created. The ML classification algorithms were the final prototype. They aimed to determine which behavioral habits from the EV drivers negatively impacted the Li-ion battery capacity of the cars, answering research question 1 (*Which behavioral habits from the electric vehicles may negatively impact Li-ion battery capacity?*).

The chosen models allowed us to gather and label the degree of battery degradation in two categories and get a general sense of the degradation trend of all vehicles. For the entities to whom this study is aimed, i.e., the owners of EV vehicles, this information can be crucial because it sheds light on the current rate of degradation of their car batteries and anticipates the need for future maintenance events, such as a complete battery replacement.

4. Surveys

Following the work and analysis done in the previous chapters with the Tesla dataset, we conducted three surveys to find the habits of electric vehicles drivers. However, after evaluating the results of the CRISP-DM methodology, we concluded that this methodology partially answered the RQ1. Furthermore, after cleaning its data, we had to work with a dataset that was reduced to a third of its original size. Thus, we decided to obtain more data through surveys to seek more confident conclusions drawn from a dataset with a more significant number of observations.

4.1. Survey Creation

These questionnaires were developed in either Portuguese or English languages, depending on their target audiences. The surveys were created on Qualtrics [69], a survey software tool, and comprised an introduction for the study objectives and its scope and the questionnaire. Unfortunately, personal data from the respondents was scrubbed from the database following the European Union General Data Protection Regulation [70]. Overall, all three surveys had in everyday choice eleven multiple items, three of which on a Likert scale, scored from 1 to 8 (1 - Never; 2 - Once or twice a year; 3 - A few times a year; 4 - Monthly; 5 - Twice a month, 6 - Weekly, 7 - Twice a Week and 8 - Daily).

In the first survey, the target audience was members of the Portuguese Association of Electric Vehicle Users (UVE - Utilizadores de Veículos Elétricos) [71]. This institution's mission is to promote electric mobility and inform the public about the use of electric vehicles. This association plays an essential role in the success of the adoption of electric vehicles in Portugal. It has a thousand associated members, and for this reason, it was considered a potentially attractive data source.

At the suggestion of my supervisors, we contacted a representative of the association to better understand its mission and possibilities of cooperation. Since there was an agreement between their mission and the purpose of this work, we requested them to disclose the survey on their associates' loading and parking habits.

The survey's decision to disclose to this group was taken because the original dataset only had Tesla vehicles. With the contribution of this community, we aimed to obtain more recent data on the same types of vehicles. Likewise, the Tesla survey, the UVE survey, was shared in the last week of February 2021, two months available to collect responses, resulting in 54 valid responses.

The second survey reached a more international and broader audience of Tesla drivers to complement the original dataset for this study that included Tesla vehicles only. Most of the Tesla vehicles of this group were mainly from before 2017 and exclusively Model S vehicles. In order to minimize this sample imbalance of Tesla Model S vehicles, an attempt was made to diversify the data by querying specific Tesla fan groups to get a more diverse and newer sample of new Tesla models owners. The aim was to get responses from owners of the newer Model 3 vehicles, which have improvements to their range from previous generations, and some occasional responses from Model Y owners, the most recent model released in early 2020, before the CoVID-19 pandemic. The second survey's target audience comprised the Reddit online platform subscribers, namely the *TeslaMotors* [72] and *TeslaLounge* [73] Reddit subgroups. By the end of February 2021, the first Reddit group mentioned had about one million subscribers (equivalent to the Cyprus population), and the second group had around thirty thousand subscribers. Due to the international nature of the previously mentioned Reddit groups, the Tesla survey was done in English. Therefore, some of the original questions were removed, such as inquiring about the EV car manufacturer. This step was necessary since these two Reddit group forums were exclusive for fans of Tesla vehicles.

The *TeslaMotors* group had strict publication rules, prohibiting the sharing of surveys within their group. For this reason, this specific survey was limited to the *Tesla Lounge* group only. On the other hand, a few questions were added to the Tesla survey. An example was an additional question regarding the distance units used by the car. Some car owners were originally from the United States of America, so this specific question was added to cater to them. This survey was created at the end of February 2021, two months available, and resulted in 30 valid responses.

A third survey was created and shared with subscribers of the Portuguese magazine DECO-Proteste [74], a magazine in Portugal. This magazine has a leading role in informing and advising consumers in Portuguese about the quality and price of various types of services and goods marketed in Portugal. The DECO survey was created because of the identified need to obtain more diverse data and more participation. In addition, this survey was more extensive and sought to obtain answers on electric batteries and the degree of satisfaction that current EV owners have about the charging infrastructure operating in Portugal. There was a suspicion that the charging network's coverage would hinder the massification of electric vehicles adoption by the public.

The dissertation supervisors contacted a representative of DECO, a journalist, and scheduled an interview. Their objective was to raise DECO awareness on the relevance and growing international interest that this topic has sparked. Before the interview, we had preparation meetings to discuss the specific questions and details of the survey, tailoring them specifically to DECO's userbase. As in the previous surveys, the DECO canvas was also created with the Qualtrics platform. The inquiry was tested on its quality, length, and conciseness.

The survey had the following requirements: it had to be in Portuguese to satisfy DECO's readers; had to be shared on an online link due to the ongoing Covid pandemic; user anonymity needed to be assured by The Qualtrics platform by not registering the answerers' IP addresses; the questionnaire had to be completed in eight or fewer minutes for the user convenience and be answerable on both desktop and mobile devices, to assure access. After the requirements were met, we had a meeting with DECO. The DECO survey was done after the previous two and published on the last week of March

2021, with availability through a link. This link was open until the date of writing this thesis (June 2021). This survey yielded ten valid responses, despite being available online for more time and published in a national reference magazine.

During the past months, several efforts have been made to circumvent the scarcity of data. The publication of an article by DECO magazine, coupled with a link to the survey, was an excellent opportunity to populate and diversify the replies of the original dataset. At the same time, it served the purpose of raising awareness in society on this subject. The objective was to obtain a substantial number of answers that could lead to more solid conclusions, which it managed to accomplish, but with modest results.

4.2. Survey Shared Questions

The three surveys followed the same structure process, created in different stages of our research. The surveys had different total numbers of questions, sharing eleven common questions. These common questions were tailored and matched among all three questionnaires to enable future comparisons and aggregations. The list below shows the eleven common question items, written as variables, with the following reasoning behind their inclusion:

Country - This was a critical variable, which was included mainly due to the international nature of the Tesla survey with responses from foreign users. It is a nominal qualitative variable, which was a simple list of countries. We did not expect a direct relationship between the country of a vehicle and its anticipated range.

Year - A continuous quantitative variable, included due to the need to distinguish different versions of identical electric vehicles. Its values ranged from the year 2010 to 2021. It was expected that this variable would find a direct relationship between the age of a vehicle and increasing battery degradation.

Maker - The carmaker was a nominal qualitative variable, including the non-exclusive Tesla surveys, such as the UVE and DECO surveys.

Model - A nominal qualitative variable, which collected data specific to each electric vehicle model. Like the previous variable, it is qualitative and nominal. It simply distinguishes different models so that each car manufacturer is not treated as a homogeneous group. It was expected that different models, even from the same car brand, had different behavior.

Freq. Fast - This categorical qualitative variable gave the respondent the chance to give one of eight answers, following the Likert scale (answers ranged from 1 to 8, where 1 means a sporadic event, and eight a recurring occurrence). This variable asked people about the frequency with which they charged their vehicle at a fast public charging station. In Portugal, fast-charging stations have power delivery exceeding the 22 kW mark [75].

Freq. Full - A categorical qualitative variable adopting the same Likert scale as the previous question. This question asked people how often they charged to the maximum charge their vehicle battery would allow. It was expected that the vehicles frequently charged beyond 80% up to their maximum capacity would have worse results in their range [76].

Freq. Empty - A categorical qualitative variable, with a Likert scale of eight answers, like the previous two questions. This question asked electric vehicle owners how often their vehicle battery is discharged below the ten percent charge threshold. The purpose of including this variable was to assess how this discharging practice could negatively impact the battery charge [76].

Charging Place - This was a nominal qualitative variable that asked people where they recharged their vehicles. Charging options included their garage, a condominium garage or a box, public charging stations, or a company's station. This variable was expected to explain in part whether vehicles parked outside, and therefore exposed to more significant variations in ambient temperatures, had a more remarkable degradation of their maximum range.

Parking Place - The Parking Place variable sought to capture the different solutions that electric vehicle owners have found to park their vehicles during the day when they are not on the road. The purpose of this question was to find a relationship between the vehicles being parked and a conjectured degenerative effect on their batteries' longevity since lithium-ion batteries do not fare well in extreme temperatures.

Mileage is a nominal, continuous variable that measures the vehicles' current range, measured in kilometers. This variable was expected to have an inverse relationship with vehicles' range.

Max Range was a continuous nominal variable representing the maximum range of electric vehicles, measured by their owners and expressed in kilometers. This variable was used as a dependent variable. Through linear regressions, attempts were made to find relationships between it and all the previously mentioned variables.

It was questioned if the frequency of charging and vehicle age variables would impact the maximum range, at the very least. This assumption would translate into a significant diminishing effect on the vehicle range.

In Chapter 3, the data cleaning process reduced the Tesla dataset from 1,425 initial responses to a small sample of 500 useable observations. This lower number of observations was the outcome of the removal of null values and outliers. Due to the shortage of useable observations and alternative data sources, it was necessary to conduct additional surveys to provide more data and substance to the conclusions.

Based on our literature review, a gap in state of the art was identified: very few articles [8][25][77][78] referred to the degradation of batteries from the point of view of electric vehicles user behavior. In most cases, articles only mentioned the point of view of Electrochemistry. It is known that lithium batteries lose capacity depending on their use, or even if they are not in use [79]. These drawbacks might constitute a barrier to the adoption of electric vehicles by EV users.

4.3. Survey Unique Questions

Some surveys had to include specific questions tailored to their different target audiences to their purposes:

UVE survey - this survey was prepared thinking of a national public that is more familiar with the reality of Portuguese charging stations. Thus, their specific questions were more related to satisfaction with the current service provided by the various charging stations. In addition, it tried to know their enthusiasm for a possible expansion of the network through financing a charging station in the garages of their condominiums. Therefore, the questions included were as follows:

- What are the biggest obstacles you face when loading?
- Indicate the degree of satisfaction with the following charging options for Electric Vehicles.
- How do you see the possibility of having a shared charging system in your condominium's garage with personal consumption accounting (through the condominium account)?
- How much would you be willing to pay to install a charging system that would make your day-to-day easier?
- How much would you be willing to pay for the monthly fee for a charging system that would make your day-to-day easier?
- How do you see the possibility of having a shared charging system in your company's garage with personal consumption accounting (through the company's account)?
- How often do you take the following types of trips?

Tesla survey - the Tesla inquiry was aimed at an international audience. For that reason, there was a high probability of getting responses from American respondents. Hence, the Tesla survey included the following question to lead respondents to answer in their favorite measurement system: *How many Miles/Kilometers does your car have, roughly?* Later unit conversions were done in Python and SPSS.

DECO survey - the DECO survey sought to obtain responses from a wider audience, both from enthusiastic and experienced users of electric vehicles, as well as from people who have never tried them. He also tried to capture their feelings and beliefs about this type of vehicle, particularly the advantages and disadvantages. The following questions were more opinionated than the reports of their habits or the counting of kilometers traveled with their vehicles, as in the previous surveys:

- Do you have an electric vehicle?
- Have you ever driven an electric vehicle?
- What are, in your opinion, the main advantages of a 100% electric vehicle?
- Indicate your degree of agreement on each of the following statements:
    - EVs are quieter than other vehicles
    - EVs have a great acceleration
    - EVs are environmentally friendly because they produce zero direct emissions
    - The cost of charging an EV is less than the fuel cost of internal combustion vehicles
    - EVs cost the same as ICE vehicles
    - The technology of electric vehicles has improved, and now they have a much longer range
    - Charging EVs is difficult
- What is the probability that the next vehicle you buy will be an electric vehicle?
- Please indicate the importance you place on each of the following factors when purchasing an Electric Vehicle:
    - EV environmental benefits
    - EV performance
    - EV looks
    - Number of available charging stations
    - Operating costs
    - Charging costs
    - EV maximum range
    - EV purchase cost
    - Ease of buying a second-hand electric vehicle
- What is the maximum charging time you consider acceptable in an electric vehicle (in hours)?
- How far must an electric vehicle be able to travel on a single charge for you to consider buying one (in kilometers)?

4.4. Surveys Data Cleaning and Merging Process

The methodology for processing these three datasets in SPSS followed a set of steps before the analysis. Like the process done in Chapter 3, when the original dataset was cleaned with Python procedures. The following steps were performed:

Standardization of units and formats: Values of dates and distances appear in different units. This step focused mainly on the conversion of variable *Mileage* from the Imperial system to the metric system.

Variable Type adjustments: Importing some variables into SPSS resulted in data type errors. It was necessary to manually adjust its type according to nominal, categorical, and continuous data.

Elimination of null records: Some of the observations from the original dataset had missing fields. These cases were excluded from the dataset.

Removal of outliers: some values were too far apart from the others, hence the need to remove them. Like the data cleaning done in Python, the outlier removal process adopted the sample percentiles technique. All observations outside the interval formed by the 5th and 95th percentiles were considered potential outliers and removed.

4.5. Results and Discussion

The results obtained from the three conducted surveys were focused on two research questions: *RQ1* - "*Which behavioral habits from the electric vehicles may negatively impact lithium-ion battery capacity?"*, and RQ2 - "*Which factors might present themselves as a hindrance to the adoption of EV vehicles by citizens?"*. The results of both research questions are presented in this sub-chapter using a similar methodology, performed with the SPSS tool owned by IBM and dedicated to statistical analysis.

4.5.1. Battery Degradation – All Surveys

This sub-chapter aimed to respond solely to *RQ1* – "*Which behavioral habits from the electric vehicles may negatively impact lithium-ion battery capacity?"*. After studying the distribution of variables, an analysis was made on how the various variables influenced the continuous variable *Mileage*. Then, using the SPSS program, a linear regression was performed. According to the specifics of the problem, this is a suitable approach to measure future values from a continuous variable. Linear regression is a primary type of regression in ML. It needs a predictor variable and a dependent variable (*Mileage*) linearly related to each other and involves using a best-fit line. Linear regression is instrumental in this case because the variables are related linearly. Therefore, the more significant the effect of charging, the greater the battery degradation effect. Also, since the linear regression analysis is susceptible to outliers, it should not analyze big data sets. However, this dataset had few observations after being cleaned of outliers, so it should not pose any problems.

Table 4-1 shows that many variables have a positive relationship with the dependent variable, *Mileage*. Those variables were *Country, Year, Maker, Charging Place, Freq. Full* and *Parking Place*. Their Unstandardized B values are positive, which seems to confirm a positive correlation between these and the independent variable.

Table 4-1 - SPSS coefficients with all variables.

Model	Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
(Constant)	-57,621.02	22,213.14		2.59	0.14
Country	18.19	28.67	.080	.63	0.53
Year	28.63	11.02	.482	2.60	**0.014**
Maker	7.25	3.60	.28	2.01	0.05
Model	-0.54	4.63	-.02	-0.13	0.91
Charging Place	6.26	8.60	.10	0.73	0.47
Freq. Fast	-12.15	9.88	-.14	1.23	0.23
Freq. Full	6.29	6.37	.11	0.99	0.33
Freq. Empty	-5.93	7.13	-.10	-0.83	0.41
Parking Place	27.34	17.14	.18	1.60	0.12
Mileage	0.00	.00	-.07	-0.38	0.70

However, there is a smaller group of variables, *Model, Freq. Fast* and *Freq. Empty* that is inversely proportional to the *Mileage* variable. This result meant that the more prominent the values of these variables are, the less range the car is likely to travel. In addition, these variables are not significant due to their p-values (listed on the *Sig*. column) being higher than 0.05, and because of it, we must reject the null hypothesis. The null hypothesis is the opposite assumption we want to prove (these variables explain the dependent variable *Max Range*). Since the p-values of these variables are higher than 0.05, we cannot assume they correlate with the *Max Range* metric. Finally, the variable *Year* had enough significance (0.014) to confirm our hypothesis. This outcome seems to hint that the older a vehicle is, the more likely it is to have less range available on a full charge. The model's effectiveness was analyzed, considering all the variables, as shown in Table 4-2 below.

Table 4-2 - Model Summary - Max Range model.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.799a	0.639	0.529	102.81

a. Predictors: (Constant), Mileage, Model, Freq. Fast, Freq. Full, Parking Place, Freq. Empty, Charging place, Maker, Year.

b. Dependent Variable: Max Range.

The *Max Range* variable was the dependent variable, while the remaining ones acted as independent variables. By employing a linear regression method, results were obtained and expressed in Table 4-2. It shows that the ten independent variables can obtain a result of R Squared of 63.9%, meaning these variables explain almost two-thirds of the *Max Range* variability.

The model summary results shown in Table 4-2 did not account for each variable's weight on the model and achieved such a high level of performance. This lack of detail is because it treats all variables as a homogeneous group. A new analysis of variance (ANOVA) test was done to determine the relative weight of each variable, which considered the variation of the R squared value. This way, it was possible to discover the change in R Squared for each predictor. All was done without the need for extra computation and fitting every variable on a single model.

Table 4-3 below shows in the *Sig* column the individual importance each variable had. It is vital to mention that concerning the significance of the variables studied, the *Maker* variable has not proven to be significant on the variability of the *Max Range* variable. However, the *Maker* variable reached a value of 0.52, above the significance 0.05 threshold, as shown in Table 4-3 below.

The *R Square Change* column displays the change in *R Square* resulting from a new predictor (or block of predictors). It highlights the reduction in the explanatory power of the model if each of the variables is removed. It is a helpful way to assess the unique contribution of new predictors (or blocks) explaining variance in the outcome. In the case of this sample, the most impactful variable was the *Year*.

Table 4-3 - Analysis of Variance - ANOVA

ANOVA [b]

		Sum of Squares	df	Mean Square	F	Sig.	R Square Change
Subset Tests	Country	4,253.93	1	4,253.91	0.40	0.53[a]	0.004
	Year	7,1381.45	1	7,1381.47	6.75	0.01[a]	0.074
	Maker	4,2770.42	1	42,770.43	4.05	0.05[a]	0.044
	Model	145.61	1	145.68	0.01	0.91[a]	0.000
	Charging place	5,592.27	1	5,592.26	0.53	0.47[a]	0.006
	Freq. Fast	15,981.43	1	15,981.48	1.51	0.23[a]	0.017
	Freq. Full	10,310.09	1	10,310.14	0.98	0.33[a]	0.011
	Freq. Empty	7,315.59	1	7,315.58	0.69	0.41[a]	0.008
	Parking Place	26,985.23	1	26,985.24	2.55	0.12[a]	0.028
	Mileage	1,548.56	1	1,548.59	0.15	0.70[a]	0.002
Regression		616,605.59	10	61,660.56	5.83	0.00[c]	
Residual		348,797.27	33	10,569.61			
Total		965,402.87	43				
a. Tested against the entire model.							
b. Dependent Variable: Max Range							
c. Predictors in the Full Model: (Constant), Mileage, Model, Freq. Fast, Freq. Full, Country, Parking Place, Freq. Empty, Charging place, Maker, Year.							

4.5.2. EV Adoption Obstacles – UVE and DECO Surveys

The UVE and DECO surveys were carried out to understand how satisfied respondents were with electric vehicles. The objective of this sub-chapter was to obtain data and to answer the second research question, RQ2 – *"Which factors might present themselves as a hindrance to the adoption of EV vehicles by citizens?"*.

A demographic analysis of the people surveyed was carried out. Ranging from Figure 4-1 to Figure 4-4 show visualization of the demographic distribution results. Most people that responded to the survey were male, representing 77% of the responses.

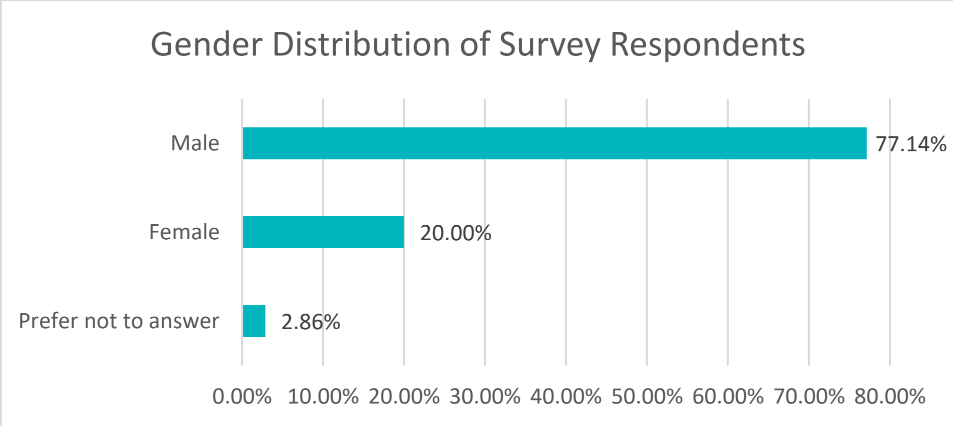

Figure 4-1 - Gender distribution of survey respondents.

Concerning age distribution, Figure 4-2 shows that nearly two-thirds were between 40 and 59 years old. The second-largest group was young adults aged between 26 and 39 years old. Lastly, adults over 60 years of age represent 14.29% of the total respondents.

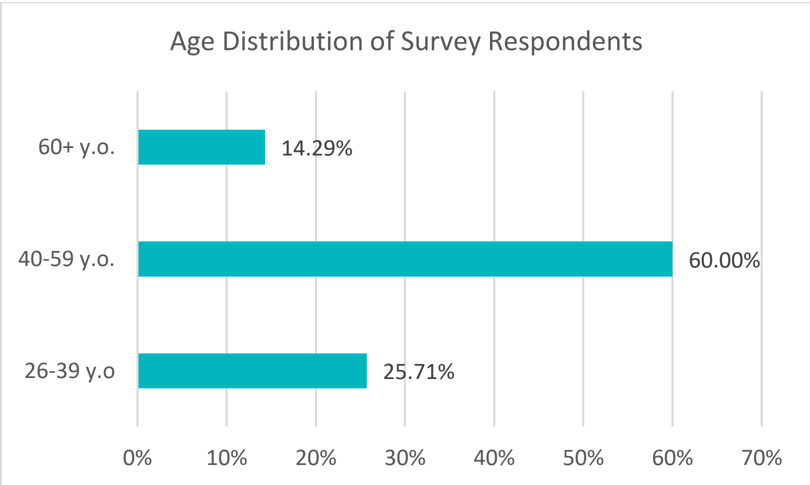

Figure 4-2 - Age distribution of survey respondents.

The following distribution was obtained regarding the household size of the people surveyed, shown in Figure 4-3. First, the majority, comprising more than half of the responses received, referred to people whose households had three or more members. Secondly came households with two elements per household, and finally, people living alone. Therefore Figure 4-3 may indicate a need for more than one vehicle per household, confirmed in Figure 4-4.

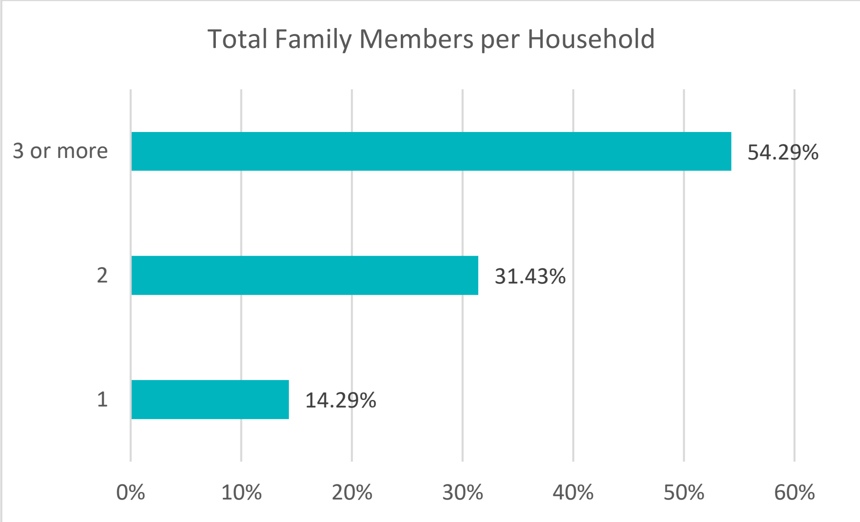

Figure 4-3 - Total family members per household.

As expected, more people per household is equivalent to more vehicles per household. Figure 4-4 corroborates this assumption: a majority between homes with two and three vehicles and a small minority of households with only one vehicle.

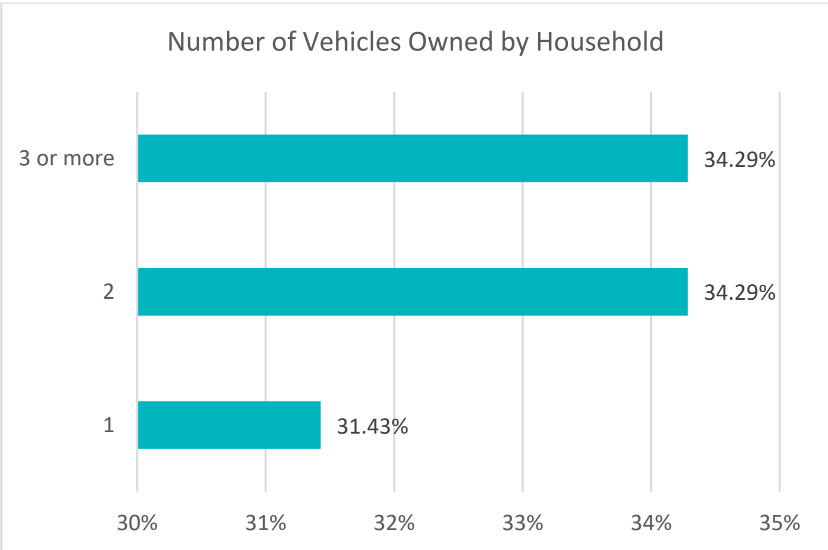

Figure 4-4 - Number of vehicles owned by the household.

This survey showed that the public is becoming sensitive to environmental issues. For instance, in a multiple-choice question on the main advantages of an electric vehicle, the most voted answer (23) was the Environmental advantage (Figure 4-5). The second most selected answer was the low operating cost that an EV has. These results are encouraging because they suggest a will for EV adoption and savings opportunities for the consumer. Opposite to this trend, a minority of nine people responded that EVs do not have any advantage over internal combustion vehicles.

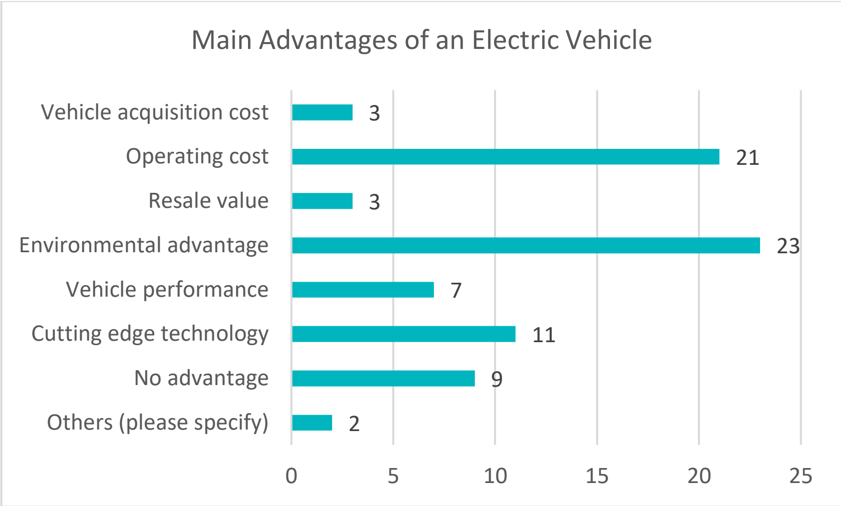

Figure 4-5 – Perceived main advantages of an electric vehicle.

Figure 4-6 below reports the main difficulties that current electric vehicles users face during the daily operation of their vehicles. The most frequent problem was the great distance to the charging stations (public and private networks). It indicates a significant deficit in national infrastructure. “Range anxiety” is a frequent problem mentioned by the scientific community [80][81], referred to as a source of stress due to the combination of reduced autonomy with inefficient charging stations network.

Secondly, the high cost of charging in these stations was indicated as an obstacle. However, the surveys' results make it impossible to confirm what price would suit Portuguese consumers. Also, it was not identified which charging network was the most expensive.

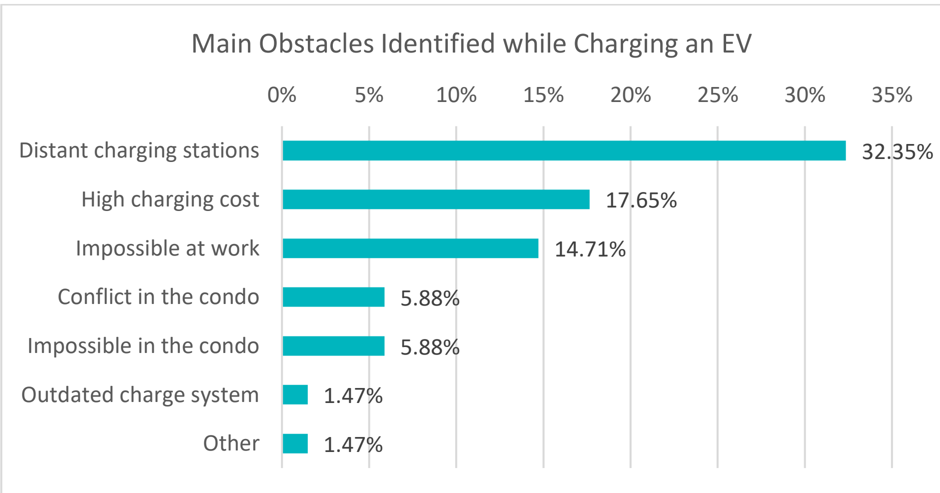

Figure 4-6 - Main obstacles identified by UVE users while charging an EV.

Thirdly, 14.71% of the users indicated the impossibility of charging EVs at their workplace. It suggests that companies still resist transitioning to electric vehicles and providing the necessary charging infrastructure to their employees. Instead, companies should provide electric charging stations in their parking areas. This suggestion is appliable to Municipalities that should provide charging stations in adjacent car parks to office buildings. Other inconveniences, such as condominium issues, were reported regarding the lack of available features to install charging stations in Portuguese condominiums.

The most preferred charging locations by UVE members are presented below in Figure 4-7. Half of the respondents preferred to make their charging from home. Considering the sizable distance between charging stations previously mentioned (Figure 4-6) and an inefficient national charging coverage network, this preference for home charging is not surprising. The second largest group preferred to charge on the public networks, such as Mobi.E that has a greater expression in the country.

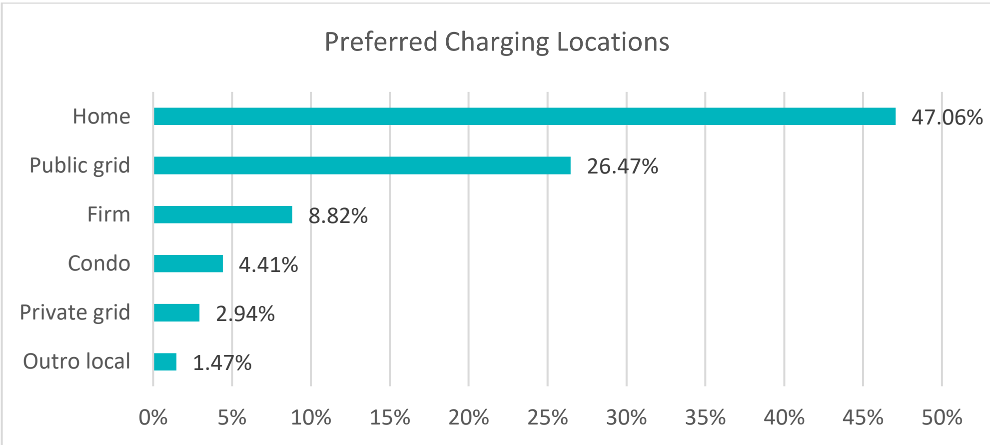

Figure 4-7 - Preferred charging locations.

There is a preference for collective solutions, such as loading at the firm (8.82%) and the condominium garage (4.41%). However, these solutions still show low levels of adherence. The least used solution is charging in a private network (2.94%), with few users. The Tesla network, an example of a private grid, has very few charging stations in Portugal. For example, when this thesis was done, there was no Tesla charging stations operating in the country's two main cities, Lisbon and Oporto. Despite that, it has eight stations that allow the country's crossing from end to end with Tesla vehicles.

Drivers' driving habits can perhaps explain the preferred use of charging at home. Most drivers travel short distances. Figure 4-8 below shows that most drivers (61.76%) make daily trips less than 50m in length that might be the commuting route from home to work. For this reason, in the survey, most expressed the need to have a charging station close by, and in the absence, they preferred to charge their vehicles at home.

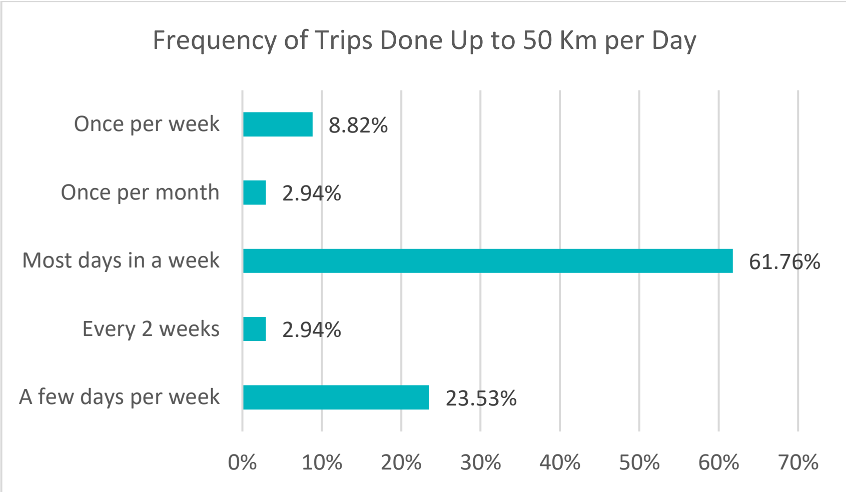

Figure 4-8 - Frequency of trips up to 50 Km per daily.

Regarding the users' satisfaction with the existing charging solutions, the consumers' preferences are represented in Figure 4-9. In first and second places are Mobi.E network (20.41%) and its Miio application (19.39%), an optional service, followed by the recent charging network of the Continente group and by other undifferentiated service stations not belonging to any of the leading suppliers (16.33%).

Figure 4-9 below represents users and their degree of satisfaction with the country's different available charging station networks. Again, most users were disappointed with the quality of the charging stations networks, and only a tiny minority is happy or does not use them at all.

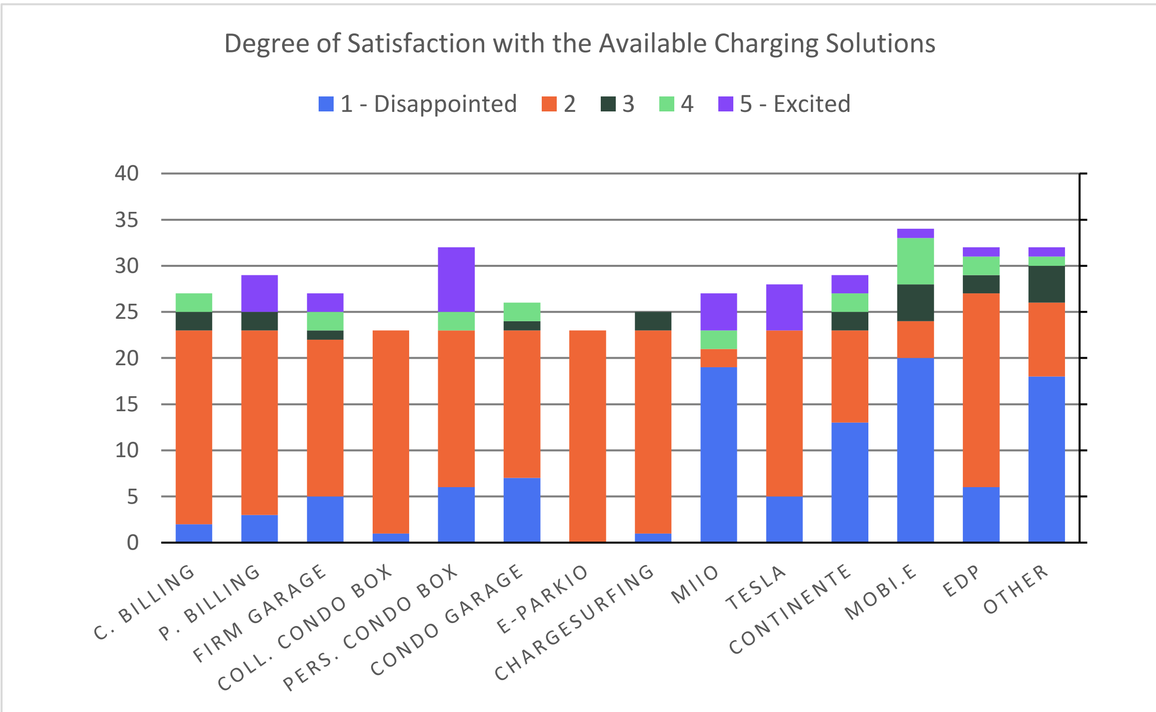

Figure 4-9 - Degree of Satisfaction with the available charging solutions.

Below 7% are the least popular options in this study: the Tesla-branded charging network, private charging solutions at home, collective garage solutions, workplace solutions, or startup ChargeSurfing's app.

There is a relationship between the number of charging stations that a network has available to its customers and its popularity: the Mobi.E network, the most popular solution, has 2390 charging stations in mainland Portugal [13], the most significant number of stations per network. In addition, there is also their proprietary application, the Miio app, which allows users to locate charging stations. Furthermore, the Miio app informs users in real-time about whether the station is available or has any impediment, such as breakdowns or users occupying the station. In addition, the application allows the user to make payments and integrates all steps in the same platform as an extra convenience.

Respondents' second most popular option is the Continente network, the solution with the second broadest national coverage, with roughly 27 charging stations operating throughout the country, strategically installed next to Continente supermarkets.

Tesla's network has the lowest coverage of all the networks listed and has only eight superchargers operating in the country. This number of Tesla chargers might seem low, but the spacing between them is enough to drive across the country, both transversely and longitudinally. Plus, Tesla has a network of partner stations compatible with Tesla vehicles, called Destination Chargers, usually located in hotels and supermarkets to complement the superchargers, but with the added drawback of providing a lower voltage and, thus, a slower charging speed.

4.5.3. Potential for Expansion – DECO Survey

DECO asked us to include a different subject in our study, tangential to this dissertation's topic and to document the respondents' answers: EV owners' opinion regarding the current service provided by the Portuguese EV charging networks, their acceptance of future network coverage expansions, and an overall improvement of the user charging experience. These different topics do not aim to answer any of the two research questions. Instead, they were included as a compromise to get our survey published to their subscriber base.

Regarding the attractiveness of the existing network expansion, Figure 4-10 shows that most survey participants expressed a very high interest in having a shared charging system installed in their condominium's garage. All that had a positive feeling were added to this first group. People with a favorable opinion represent most reported cases, meaning they had a greater interest in adopting this domestic solution. However, the second biggest group voiced no significant interest in receiving this technology for unspecified reasons. It could probably be because of this offshoot of people charging their EV cars elsewhere, other than their homes/condominiums. Still, it is worthy of mentioning the interest this solution raised among the respondents. Furthermore, it tried to quantify this interest in costs for users.

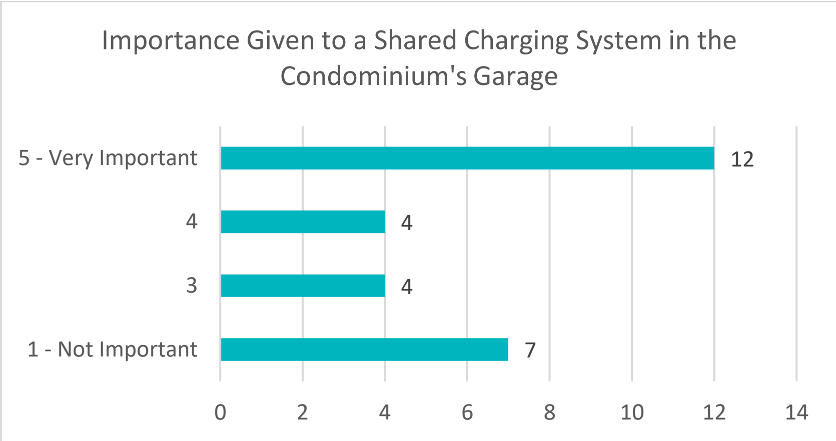

Figure 4-10 - Importance given to a shared charging system in the condominium's garage.

Another common topic among respondents was the importance that people gave to charging stations in the vicinity of their homes. In Figure 4-11, most people responded very positively to this question, suggesting a need to be fulfilled to potential buyers of EV. It confirms the previous Figure interpretation that EV owners of this sample are charging their vehicles more at any charging station near the condominiums. This outcome indicates considerable interest in the expansion and densification of charging networks, notably in public roads.

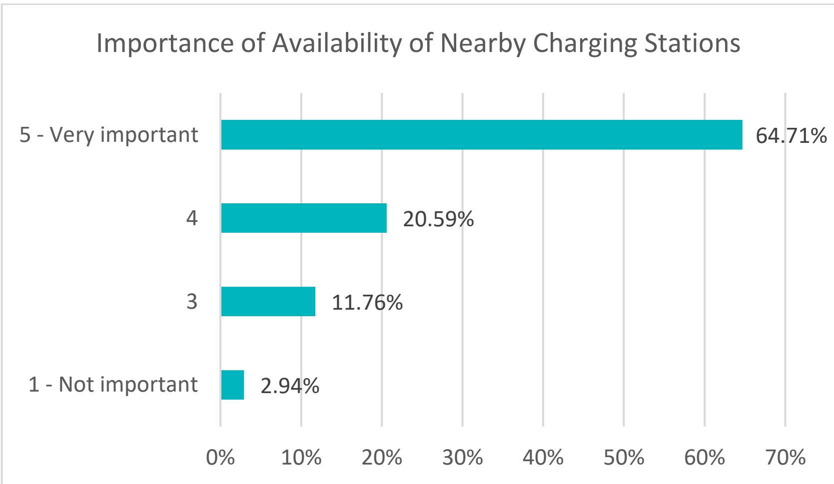

Figure 4-11 - Importance of nearby available charging stations.

Next, people were asked what monetary amount they would consider acceptable for installing a charger in their condominium garage, regardless of other existing factors. Most respondents (46%)

replied that they would not accept to pay any amount. This response had a dominant expression and was expected when the survey was launched. However, these answers may be related to multiple factors that the DECO survey did not predict, such as the economic situation of respondents or simply a pure lack of interest in the announced solution. Thus, we are interested in analyzing the people willing to pay for this service, representing 54%. A smaller group of 31% declared they would accept a one-time expenditure of up to 200 euros, and a more favored minority declared willingness to pay above 1,000 euros, as shown in Figure 4-12.

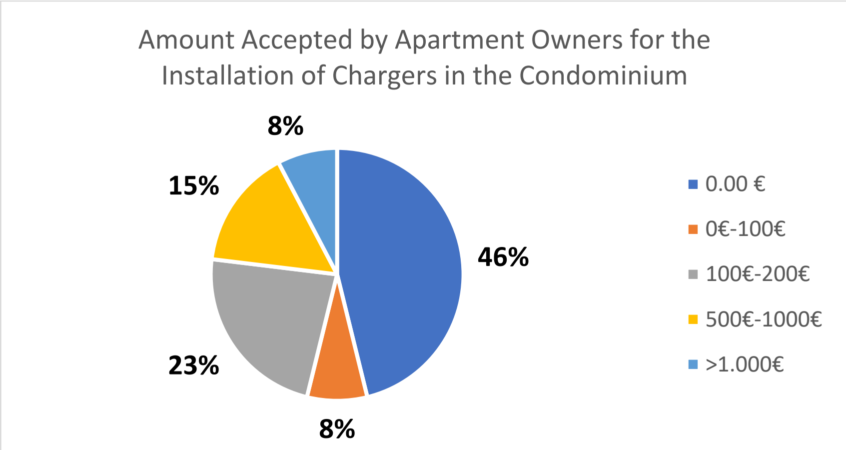

Figure 4-12 - Amount accepted for the installation of chargers.

Nevertheless, the installation of a charger is just one of the expenses involved. It is also necessary to account for the monthly cost of its everyday use. This cost is a monthly fee that each owner must pay for the service maintenance. The following survey question, "*How much would you be willing to pay for the monthly fee for a charging system that would make your day-to-day easier?"* responses show that most people do not want to pay again for this service. Only a third of people declared they would accept to pay a monthly fee, according to the distribution made in Figure 4-13.

Furthermore, only 5% would agree to pay more than €40 extra per month. This reluctance in spending makes it challenging to accept the service as a first step.

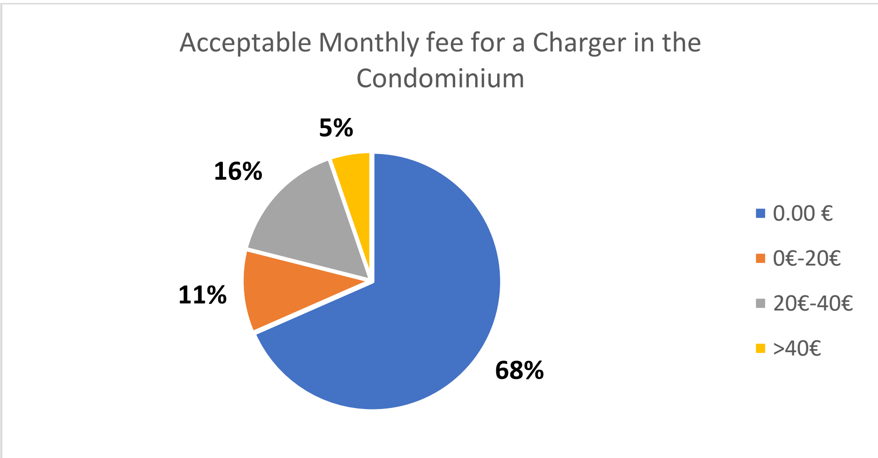

Figure 4-13 - Acceptable monthly fee for the installation of a charger.

Following this descriptive statistical study, a search was done between all variables for strong, positive, and statistically significant correlations. To this end, a few correlations were found that can attest to the potential and market to expand the current EV station coverage.

The first correlation found was between the *Indoor/Outdoor* variable, a new variable created through feature engineering, and the variables *Importance: of EV maximum range, Agreement: Charging EVs is difficult*, and *Importance: number of charging stations*.

The *Indoor/Outdoor* variable expresses in percentage the probability that the user prefers to charge at public stations rather than at home. Values of this variable closer to zero represent a greater appetence to charging the car with solely domestic solutions (wall chargers at home or the condominium). In contrast, values closer to 1 represent the probability of using domestic, public, and private charging solutions.

In

Table 4-4, the model summary compares the dependent variable *Indoor/Outdoor* against all its predictors. SPSS reached an R Squared value of 0.444. This value means that the model had a positive and moderate correlation and can explain 44.4% of the variability between its independent and dependent variables.

Table 4-4 - Model Summary -Usage of outdoor charging solutions and obstacles found.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.667[a]	0.444	-1.222	1.63299	0.444	0.267	3	1	0.852

a. Predictors: (Constant), Importance: EV maximum range, Agreement: Charging EVs is difficult, Importance: number of charging stations

By analyzing variance (Table 4-5), the model reached a significance or p-value of 0.012, a value below 0.05, which confirms it is statistically significant. A value of 0.012 signals evidence against the null hypothesis (that the variables are not correlated), as there is less than a 5% probability of being correct. Therefore, the null hypothesis is rejected, and the alternative hypothesis is accepted (the dependent and independent variables are correlated).

Table 4-5 - Analysis of variance - Outdoor charging solution usage.

ANOVA [a]

Model		Sum of Squares	df	Mean Square	F	p-value.
1	Regression	2.133	3	0.711	7.267	.012[b]
	Residual	2.667	1	2.667		
	Total	4.800	4			

a. Dependent Variable: Indoor/Outdoor charging

b. Predictors: (Constant), Importance: EV maximum range, Agreement: Charging EVs is difficult, Importance: number of available charging stations

After a few more experiments, we found another correlation between a dependent variable *Importance: having a shared charging system*, and the following variables as independent ones: *Age distribution, Agreement: Charging EVs is difficult, Agreement: EV charging cost is low, Frequency of trips between 50 and 100 Km per day, Frequency of trips up to 50 Km per day, Frequency of trips with more than 100 Km per day, Gender distribution, Importance: EV maximum range, Importance: operating costs, Importance: operating costs, Maximum acceptable charging time for an EV (Hours), Number of elements per household, Minimum range in kilometers to consider buying an EV* and lastly *Number of vehicles per household*. With this selection of independent variables, this model obtained an R Squared of 69.0%, suggesting a positive and strong correlation, as shown in Table 4-6 below.

Table 4-6 - Model Summary - Shared charging system.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.831[a]	0.690	0.450	1.317	0.690	2.867	14	18	0.019

a. Predictors: (Constant), Number of vehicles per household, Frequency of trips between 50 and 100 Km per day, Maximum acceptable charging time for an EV (Hours), Frequency of trips up to 50 Km per day, Gender distribution, Agreement: Charging EVs is difficult, Frequency of trips with more than 100 Km per day, Agreement: EV charging cost is lower, Number of elements per household, Importance: operating costs, Age distribution, Minimum range in kilometers to consider buy an EV, Importance: EV maximum range, Importance: charging costs
b. Dependent Variable: Importance of having a shared charging system in the condominium's garage with individual consumption accounting (through the condominium account)?

As with the previous model, it is necessary to analyze its variance to determine whether the results of the surveys are significant. In addition, we need to know whether we reject the null hypothesis (whether the predictor elements explain the dependent variable) or not. The results of this analysis are shown in Table 4-7.

Table 4-7 - Analysis of Variance - Importance of having a shared charging system.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	p-value.
1	Regression	69.667	14	4.976	2.867	.019[b]
	Residual	31.242	18	1.736		
	Total	100.909	32			

a. Dependent Variable: Importance of having a shared charging system in the condominium's garage with individual consumption accounting (through the condominium account)?
b. Predictors: (Constant), Number of vehicles per household, Frequency of trips between 50 and 100 Km per day, Maximum acceptable charging time for an EV (Hours), Frequency of trips up to 50 Km per day, Gender distribution, Agreement: Charging EVs is difficult, Frequency of trips with more than 100 Km per day, Agreement: EV charging cost is lower, Number of elements per household, Importance of operating costs, Age distribution, Minimum range in kilometers to consider buying an EV, Importance of EV maximum range, Importance of charging costs

As shown in Table 4-7, the analysis of variance obtained a p-value of .019, and therefore below 0.05, which allows us to confirm that this model has a positive, strong, and significant correlation.

It is possible to gauge which variables had the most significant weight in the correlation found. Studying the coefficients represented in Table 4-8 makes it possible to determine which variables are most important. However, only two variables from this model were statistically significant, with a significance value below 0.05. These variables were: *Agreement: EV charging cost is lower and 100 Km per day and Frequency of trips with more than 100 Km per day.*

Table 4-8 - Coefficients table - Importance of having a shared charging system.

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)	-5.693	5.288		-1.077	0.296			
	Agreement: Charging EVs is difficult	-0.319	0.259	-0.199	-1.233	0.233	-0.270	-0.279	-0.162
	Importance: charging costs	0.752	0.776	0.269	0.968	0.346	0.382	0.222	0.127
	Importance: EV maximum range	-0.242	0.514	-0.117	-0.471	0.644	0.045	-0.110	-0.062
	Maximum acceptable charging time for an EV (Hours)	0.059	0.137	0.103	0.429	0.673	0.113	0.101	0.056
	Agreement: EV charging cost is lower	0.593	0.230	0.465	2.577	0.019	0.435	0.519	0.338
	Importance: operating costs	1.033	0.791	0.357	1.307	0.208	0.443	0.294	0.171
	Minimum range in kilometers to consider buy an EV	-0.001	0.002	-0.119	-0.605	0.552	-0.286	-0.141	-0.079
	Frequency of trips up to 50 Km per day	0.204	0.265	0.114	0.771	0.451	-0.068	0.179	0.101
	Frequency of trips between 50 and 100 Km per day	0.009	0.138	0.012	0.068	0.946	0.075	0.016	0.009
	Frequency of trips with more than 100 Km per day	0.275	0.126	0.327	2.185	0.042	0.429	0.458	0.287
	Gender distribution	0.178	0.708	0.047	0.251	0.805	-0.110	0.059	0.033
	Age distribution	-0.185	0.515	-0.068	-0.359	0.723	-0.143	-0.084	-0.047
	Number of elements per household	-0.062	0.425	-0.025	-0.146	0.886	-0.241	-0.034	-0.019
	Number of vehicles per household	0.107	0.425	0.050	0.253	0.803	-0.221	0.060	0.033

a. Dependent Variable: Importance: having a shared charging system in the condominium's garage with individual consumption accounting (through the condominium account)?

Therefore, this allows us to state that respondents' importance to having a joint charging system is even more significant as they agree that EVs are cheaper to use than internal combustion vehicles. Moreover, this seems more critical; the greater the frequency of trips a person makes corresponds to a longer distance of more than 100 kilometers a day.

5. Conclusions

This thesis aimed to identify the main behavioral factors that impact lithium-ion battery performance by studying human actions when interacting with EVs. Based on the three surveys created and an approach backed by the CRISP-DM methodology, it was possible to conclude that charging and parking habits are negligible at best and might even be irrelevant to the decay of batteries. Furthermore, charging habits did not negatively impact the batteries' longevity. The results from the CRISP-DM methodology have shown that the cars' increasing age (variable *Year*) was by far the most significant variable. Vehicles that were regularly fully charged did not display significantly lower *Max Range* values than the other vehicles that avoided that practice. To this end, we answered the RQ1 by identifying the vehicle's release year as the determining factor for battery degradation, but no behavioral habits play a meaningful part in the decay.

The second research question, RQ2 - "*Which factors might present themselves as a hindrance to the adoption of EV vehicles by citizens?"* focused on the need for mass EV acceptance in society. Its objective was to establish the factors that might be present and hinder EV adoption by citizens. To this end, the respondents of the three surveys were asked questions to measure their degree of satisfaction with the current electrical charging network, identifying the main obstacles they encountered, and measuring their acceptability in funding the installation of smaller chargers in their condominiums garage. The results from the statistical analysis showed the following points: first, the Mobi. E network was the preferred station network by the EV users. Secondly, a small portion of the EV userbase was willing to fund the installation of chargers in their condominiums' garages. Third, the importance of sharing a charging system in the condominium is correlated with the frequency of making long-range trips above 100 Km. Finally, the main obstacle mentioned was the poor coverage by public and private chargers, followed by their high charging cost. This last finding allowed us to answer RQ2 directly.

Table 5-1 summarizes the conclusions reached by this dissertation and identifies which methodologies applied allowed us to answer the research questions.

Table 5-1 - Summary of research questions findings.

Research Question (RQ)	**CRISP-DM**	**Survey Tesla**	**Survey UVE**	**Survey DECO**
RQ1 - Which behavioral habits do EV drivers take that more negatively impact lithium-ion battery capacity	Variable Year	Variable Year	Variable Year	Variable Year
RQ2 - What are the factors that present themselves as barriers to the adoption of EV vehicles?			• Poor charging station coverage • High charging costs at stations	
Extra objective - Determine the acceptance of EV charging network expansions/funding a charger in the condominium				Positive reception to both questions

5.1. Contributions

This paper sought answers to the two main research questions: RQ1 - "Which behavioral habits do EV drivers take that more negatively impact lithium-ion battery capacity?" and RQ2 - "What are the factors that present themselves as barriers to the adoption of EV vehicles?". The two questions were answered by analyzing the Tesla dataset and the three questionnaire results. Holistically, in the scope of the Tesla dataset comprised of 1.400 vehicles, the factors determined to be the most critical predictor factor of battery decay were the vehicles' age (variable *Year*). The research allowed us to verify that all the factors under study have a marginal effect on the total autonomy of the vehicles. The correlations found were positive but not very significant.

Regarding the disadvantages of EVs that discourage potential buyers from purchasing them (RQ2), it was shown in Chapter 4 that the most significant variable for this disappointment was the high distance between charging stations.

Concerning the literature review conducted in Chapter 2, the work clarifies that just one scientific paper [25] quantified the battery degradation from the drivers' charging and operating habits of this type of vehicle. Most of the papers collected by employing the PRISMA methodology were centered around the Electrochemistry field of expertise. These studies focused primarily on physical factors the batteries were subjected to, such as voltage, materials composition, and the varying range of temperatures

During the research, it was possible to reach several conclusions that challenged some initial questions we had. First, the question of a higher frequency of using a fast-charging station harms the vehicle's range was questioned. Likewise, the assumption that the frequency with which drivers let the batteries charge to 100% or discharge below 10% would each hurt autonomy was also challenged. For the sample studied, these two theories seem to have had no adverse effect on battery autonomy. The effect of these three categorical variables (*Freq. Fast, Freq. Full, and Freq. Empty*) was unexpected, as it was anticipated that they would affect full range to some degree of significance. Instead, the values of these three variables in this sample seemed to disprove that belief. Finally, the idea that Li-ion batteries in EVs degrade in the same way as cell phone lithium batteries after a few years pass by or after a few hundred cycles are completed seems to be refuted. Additional evidence on this will be needed to draw definitive conclusions.

On the other hand, the research allowed to state that the price of EVs is the main factor preventing the purchase of these types of vehicles (RQ2). However, while the price is a vital factor since EVs and ICE cars have not yet reached price parity, it was found that there were other factors at play, specifically the EV's short-range and the high distance between charging stations. These observations can be confirmed in Chapter 4, where a correlation was found between the following questions: 1 - *EVs cost the same as ICE vehicles*, 2 - *EVs have great acceleration*, and 3 - *The cost of charging an EV is less than the fuel cost of internal combustion vehicles*, and the probability that the respondents' next car will be an EV.

5.2. Future Work

This section briefly describes some important research topics, which are worth investigating further. Finally, based on the previously mentioned conclusions, practitioners should consider the following future extensions to this work:

An excellent suggestion for future works is to expand the dataset's total of observations. Hopefully, this way, a more extensive dataset will have a more significant number of participants.

Another suggestion is the inclusion of research from the standpoint of the electrochemistry field. That task should be easy, as there is plenty of available literature produced in this area. It is also more exhaustive about BMS (Battery Management Systems) and the role that the SoH and the SoC's monitoring has on battery long life.

Based on the previous two points, it would be interesting to establish guidelines for EV users to adopt beneficial actions to lessen premature aging on batteries.

References

[1] C. de Ministros, “Plano Nacional de Energia e Clima 2021-2030,” 2019. https://www.portugalenergia.pt/setor-energetico/bloco-3/], (accessed Aug. 16, 2021).

[2] R. Y. and B. Z. (eds. . Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, “Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,” Cambridge, 2021. [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf

[3] “Cross Industry Standard Process for Data Mining (CRISP-DM),” 1996. https://www.datascience-pm.com/crisp-dm-2/ (accessed Aug. 18, 2021).

[4] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, and M. M. Lalu, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, no. 71, p. 9, 2021, doi: 10.1136/bmj.n71.

[5] F. Lambert, “Tesla battery degradation at less than 10% after over 160,000 miles, according to latest data.,” *Elektrek*, 2018. https://electrek.co/2018/04/14/tesla-battery-degradation-data/ (accessed May 04, 2021).

[6] I. Tesla, “2020 Tesla Impact Report,” 2020. https://www.tesla.com/ns_videos/2020-tesla-impact-report.pdf (accessed Aug. 16, 2021).

[7] EPA, “EPA and NHTSA Finalize Historic National Program to Reduce Greenhouse Gases and Improve Fuel Economy for Cars and Light Trucks,” no. April 2010, pp. 1–9, 2012.

[8] G. Saldaña, J. I. S. Martín, I. Zamora, F. J. Asensio, O. Oñederra, and M. González, “Empirical Electrical and Degradation Model for Electric Vehicle Batteries,” *IEEE Access*, vol. 8, pp. 155576–155589, 2020, doi: 10.1109/ACCESS.2020.3019477.

[9] G. May and A. El-Shahat, “Battery-degradation model based on the ANN regression function for EV applications,” in *GHTC 2017 - IEEE Global Humanitarian Technology Conference, Proceedings*, 2017, vol. 2017-Janua, pp. 1–3. doi: 10.1109/GHTC.2017.8239301.

[10] M. Singh Ceng and K. Janardhan Reddy, “Predicting Life-Cycle Estimation of Electric Vehicle Battery Pack through Degradation by Self Discharge and Fast Charging,” *SAE Tech. Pap.*, no. 2020, 2020, doi: 10.4271/2020-28-0435.

[11] Z. Yun and W. Qin, “Remaining Useful Life Estimation of Lithium-Ion Batteries Based on Optimal Time Series Health Indicator,” *IEEE Access*, vol. 8, pp. 55447–55461, 2020, doi:

10.1109/ACCESS.2020.2981947.

[12] X. Tan, Y. Tan, D. Zhan, Z. Yu, Y. Fan, J. Qiu, and J. Li, "Real-Time State-of-Health Estimation of Lithium-Ion Batteries Based on the Equivalent Internal Resistance," *IEEE Access*, vol. 8, pp. 56811–56822, 2020, doi: 10.1109/ACCESS.2020.2979570.

[13] X. S. Hu, H. Yuan, C. F. Zou, Z. Li, and L. Zhang, "Co-Estimation of State of Charge and State of Health for Lithium-Ion Batteries Based on Fractional-Order Calculus," *Ieee Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10319–10329, 2018, doi: 10.1109/tvt.2018.2865664.

[14] Y. Jiang, J. Zhang, L. Xia, and Y. Liu, "State of Health Estimation for Lithium-Ion Battery Using Empirical Degradation and Error Compensation Models," *IEEE Access*, vol. 8, pp. 123858–123868, 2020, doi: 10.1109/ACCESS.2020.3005229.

[15] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, *et al.*, "Search for long-lived, massive particles in events with displaced vertices and missing transverse momentum in root S=13 TeV pp collisions with the ATLAS detector," *Phys. Rev. D*, vol. 97, no. 5, 2018, doi: 10.1103/PhysRevD.97.052012.

[16] J. Kim, H. Chun, M. Kim, J. Yu, K. Kim, T. Kim, and S. Han, "Data-Driven State of Health Estimation of Li-Ion Batteries With RPT-Reduced Experimental Data," *IEEE Access*, vol. 7, pp. 106987–106997, 2019, doi: 10.1109/ACCESS.2019.2932719.

[17] K. Li, F. Wei, K. J. Tseng, and B.-H. Soong, "A Practical Lithium-Ion Battery Model for State of Energy and Voltage Responses Prediction Incorporating Temperature and Ageing Effects," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6696–6708, 2018, doi: 10.1109/TIE.2017.2779411.

[18] D. Liu, X. Yin, Y. Song, W. Liu, and Y. Peng, "An On-Line State of Health Estimation of Lithium-Ion Battery Using Unscented Particle Filter," *IEEE Access*, vol. 6, pp. 40990–41001, 2018, doi: 10.1109/ACCESS.2018.2854224.

[19] A. El Mejdoubi, H. Chaoui, H. Gualous, P. Van Den Bossche, N. Omar, and J. Van Mierlo, "Lithium-ion batteries health prognosis considering aging conditions," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6834–6844, 2019, doi: 10.1109/TPEL.2018.2873247.

[20] J. W. Wei, G. Z. Dong, and Z. H. Chen, "Remaining Useful Life Prediction and State of Health Diagnosis for Lithium-Ion Batteries Using Particle Filter and Support Vector Regression," *Ieee Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5634–5643, 2018, doi: 10.1109/tie.2017.2782224.

[21] C. Vidal, M. Hausmann, D. Barroso, P. M. Shamsabadi, A. Biswas, E. Chemali, R. Ahmed, and A. Emadi, "Hybrid Energy Storage System State-Of-Charge Estimation Using Artificial Neural Network for Micro-Hybrid Applications," in *2018 IEEE Transportation and Electrification Conference and Expo, ITEC 2018*, 2018, pp. 868–873. doi: 10.1109/ITEC.2018.8450251.

[22] J. Qu, F. Liu, Y. Ma, and J. Fan, "A Neural-Network-Based Method for RUL Prediction and SoH Monitoring of Lithium-Ion Battery," *IEEE Access*, vol. 7, pp. 87178–87191, 2019, doi:

10.1109/ACCESS.2019.2925468.

[23] C. Vidal, P. Malysz, P. Kollmeyer, and A. Emadi, "Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, vol. 8, pp. 52796–52814, 2020, doi: 10.1109/ACCESS.2020.2980961.

[24] European Comission, "European Green Deal," 2021. https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en (accessed Aug. 18, 2021).

[25] X. Zhang, H. Peng, H. Wang, and M. Ouyang, "Hybrid Lithium Iron Phosphate Battery and Lithium Titanate Battery Systems for Electric Buses," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 956–965, 2018, doi: 10.1109/TVT.2017.2749882.

[26] "Mendeley." https://www.mendeley.com/search/ (accessed May 22, 2021).

[27] L. Waltman and N. J. van Eck, "VOSviewer - Visualizing scientific landscapes." https://www.vosviewer.com/ (accessed Jul. 18, 2021).

[28] N. Bouchhima, M. Gossen, S. Schulte, and K. P. Birke, "Lifetime of self-reconfigurable batteries compared with conventional batteries," *J. Energy Storage*, vol. 15, pp. 400–407, 2018, doi: 10.1016/j.est.2017.11.014.

[29] A. C. Caliwag and W. Lim, "Hybrid VARMA and LSTM Method for Lithium-ion Battery State-of-Charge and Output Voltage Forecasting in Electric Motorcycle Applications," *IEEE Access*, vol. 7, pp. 59680–59689, 2019, doi: 10.1109/ACCESS.2019.2914188.

[30] Z. Chen, Q. Xue, R. Xiao, Y. Liu, and J. Shen, "State of Health Estimation for Lithium-ion Batteries Based on Fusion of Autoregressive Moving Average Model and Elman Neural Network," *IEEE Access*, vol. 7, pp. 102662–102678, 2019, doi: 10.1109/ACCESS.2019.2930680.

[31] A. J. Crawford, Q. Huang, M. C. W. Kintner-Meyer, J.-G. Zhang, D. M. Reed, V. L. Sprenkle, V. V. Viswanathan, and D. Choi, "Lifecycle comparison of selected Li-ion battery chemistries under grid and electric vehicle duty cycle combinations," *J. Power Sources*, vol. 380, pp. 185–193, 2018, doi: 10.1016/j.jpowsour.2018.01.080.

[32] X. Cui and T. Hu, "State of Health Diagnosis and Remaining Useful Life Prediction for Lithium-ion Battery Based on Data Model Fusion Method," *IEEE Access*, vol. 8, pp. 207298–207307, 2020, doi: 10.1109/ACCESS.2020.3038182.

[33] G. Dong, Z. Chen, J. Wei, and Q. Ling, "Battery health prognosis using brownian motion modeling and particle filtering," *IEEE Trans. Ind. Electron.*, vol. 65, no. 11, pp. 8646–8655, 2018, doi: 10.1109/TIE.2018.2813964.

[34] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, and M. Ouyang, "Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8583–8592, 2019, doi: 10.1109/TVT.2019.2927120.

[35] X. Hu, C. M. Martinez, and Y. Yang, "Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: A unified cost-optimal approach," *Mech. Syst. Signal Process.*, vol. 87, pp. 4–16, 2017, doi: 10.1016/j.ymssp.2016.03.004.

[36] X. S. Hu, F. Feng, K. L. Liu, L. Zhang, J. L. Xie, and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," *Renew. Sustain. Energy Rev.*, vol. 114, 2019, doi: 10.1016/j.rser.2019.109334.

[37] Q. Huo, Z. Ma, X. Zhao, T. Zhang, and Y. Zhang, "Bayesian Network-Based State-of-Health Estimation for Battery on Electric Vehicle Application and its Validation Through Real-World Data," *IEEE Access*, vol. 9, pp. 11328–11341, 2021, doi: 10.1109/ACCESS.2021.3050557.

[38] N. Khan, F. U. M. Ullah, Afnan, A. Ullah, M. Y. Lee, and S. W. Baik, "Batteries State of Health Estimation via Efficient Neural Networks With Multiple Channel Charging Profiles," *IEEE Access*, vol. 9, pp. 7797–7813, 2021, doi: 10.1109/ACCESS.2020.3047732.

[39] M. Kim, K. Kim, and S. Han, "Reliable Online Parameter Identification of Li-Ion Batteries in Battery Management Systems Using the Condition Number of the Error Covariance Matrix," *IEEE Access*, vol. 8, pp. 189106–189114, 2020, doi: 10.1109/ACCESS.2020.3031500.

[40] X. Y. Li, Z. P. Wang, L. Zhang, C. F. Zou, and D. D. Dorrell, "State-of-health estimation for Li-ion batteries by combing the incremental capacity analysis method with grey relational analysis," *J. Power Sources*, vol. 410, pp. 106–114, 2019, doi: 10.1016/j.jpowsour.2018.10.069.

[41] Y. Li, S. Zhong, Q. Zhong, and K. Shi, "Lithium-Ion Battery State of Health Monitoring Based on Ensemble Learning," *IEEE Access*, vol. 7, pp. 8754–8762, 2019, doi: 10.1109/ACCESS.2019.2891063.

[42] M. S. Hossain Lipu, M. A. Hannan, A. Hussain, and M. H. M. Saad, "Optimal BP neural network algorithm for state of charge estimation of lithium-ion battery using PSO with PCA feature selection," *J. Renew. Sustain. Energy*, vol. 9, no. 6, 2017, doi: 10.1063/1.5008491.

[43] M. S. H. Lipu, M. A. Hannan, A. Hussain, M. M. Hoque, P. J. Ker, M. H. M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *J. Clean. Prod.*, vol. 205, pp. 115–133, 2018, doi: 10.1016/j.jclepro.2018.09.065.

[44] K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, "Modified Gaussian Process Regression Models for Cyclic Capacity Prediction of Lithium-Ion Batteries," *IEEE Trans. Transp. Electrif.*, vol. 5, no. 4, pp. 1225–1236, 2019, doi: 10.1109/TTE.2019.2944802.

[45] Y. Ma, Y. Chen, X. W. Zhou, and H. Chen, "Remaining Useful Life Prediction of Lithium-Ion Battery Based on Gauss-Hermite Particle Filter," *Ieee Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1788–1795, 2019, doi: 10.1109/tcst.2018.2819965.

[46] D. G. Moye, P. L. Moss, X. J. Chen, W. J. Cao, and S. Y. Foo, "A design-based predictive model

for lithium-ion capacitors," *J. Power Sources*, vol. 435, 2019, doi: 10.1016/j.jpowsour.2019.226694.

[47] P. Shen, M. G. Ouyang, L. G. Lu, J. Q. Li, and X. N. Feng, "The Co-estimation of State of Charge, State of Health, and State of Function for Lithium-Ion Batteries in Electric Vehicles," *Ieee Trans. Veh. Technol.*, vol. 67, no. 1, pp. 92–103, 2018, doi: 10.1109/tvt.2017.2751613.

[48] J. Tang, Q. Liu, S. Liu, X. Xie, J. Zhou, and Z. Li, "A Health Monitoring Method Based on Multiple Indicators to Eliminate Influences of Estimation Dispersion for Lithium-Ion Batteries," *IEEE Access*, vol. 7, pp. 122302–122314, 2019, doi: 10.1109/ACCESS.2019.2936213.

[49] Y. Wang, G. Gao, X. Li, and Z. Chen, "A fractional-order model-based state estimation approach for lithium-ion battery and ultra-capacitor hybrid power source system considering load trajectory," *J. Power Sources*, vol. 449, 2020, doi: 10.1016/j.jpowsour.2019.227543.

[50] Z. Wang, J. Ma, and L. Zhang, "State-of-Health Estimation for Lithium-Ion Batteries Based on the Multi-Island Genetic Algorithm and the Gaussian Process Regression," *IEEE Access*, vol. 5, pp. 21286–21295, 2017, doi: 10.1109/ACCESS.2017.2759094.

[51] Y. Wu, Q. Xue, J. Shen, Z. Lei, Z. Chen, and Y. Liu, "State of Health Estimation for Lithium-Ion Batteries Based on Healthy Features and Long Short-Term Memory," *IEEE Access*, vol. 8, pp. 28533–28547, 2020, doi: 10.1109/ACCESS.2020.2972344.

[52] M. Xiang, Y. He, H. Zhang, C. Zhang, L. Wang, C. Wang, and C. Sui, "State-of-Health Prognosis for Lithium-Ion Batteries Considering the Limitations in Measurements via Maximal Information Entropy and Collective Sparse Variational Gaussian Process," *IEEE Access*, vol. 8, pp. 188199–188217, 2020, doi: 10.1109/ACCESS.2020.3029276.

[53] X. Xu, C. Yu, S. Tang, X. Sun, X. Si, and L. Wu, "State-of-Health Estimation for Lithium-Ion Batteries Based on Wiener Process With Modeling the Relaxation Effect," *IEEE Access*, vol. 7, pp. 105186–105201, 2019, doi: 10.1109/ACCESS.2019.2923095.

[54] H. Yao, X. Jia, Q. Zhao, Z. Cheng, and B. Guo, "Novel Lithium-Ion Battery State-of-Health Estimation Method Using a Genetic Programming Model," *IEEE Access*, vol. 8, pp. 95333–95344, 2020, doi: 10.1109/ACCESS.2020.2995899.

[55] C. Zhang, Y. He, L. Yuan, and S. Xiang, "Capacity Prognostics of Lithium-Ion Batteries using EMD Denoising and Multiple Kernel RVM," *IEEE Access*, vol. 5, pp. 12061–12070, 2017, doi: 10.1109/ACCESS.2017.2716353.

[56] L. Zhang, Z. Mu, and C. Sun, "Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Exponential Model and Particle Filter," *IEEE Access*, vol. 6, pp. 17729–17740, 2018, doi: 10.1109/ACCESS.2018.2816684.

[57] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Lithium-Ion Battery Remaining Useful Life Prediction with Box-Cox Transformation and Monte Carlo Simulation," *IEEE Trans. Ind. Electron.*, vol. 66,

no. 2, pp. 1585–1597, 2019, doi: 10.1109/TIE.2018.2808918.

[58] X. Zheng and X. Deng, “State-of-Health Prediction For Lithium-Ion Batteries With Multiple Gaussian Process Regression Model,” *IEEE Access*, vol. 7, pp. 150383–150394, 2019, doi: 10.1109/ACCESS.2019.2947294.

[59] D. Zhou, Z. Li, J. Zhu, H. Zhang, and L. Hou, “State of Health Monitoring and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Temporal Convolutional Network,” *IEEE Access*, vol. 8, pp. 53307–53320, 2020, doi: 10.1109/ACCESS.2020.2981261.

[60] “Python.org.” https://www.python.org/ (accessed Aug. 24, 2021).

[61] “Project Jupyter.” https://jupyter.org/ (accessed Aug. 24, 2021).

[62] F. Pedregosa, G. Varoquaux, and A. Gramfort, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[63] · N S Altman and N. S. Altman, “BU-1065MA An Introduction to Kernel and Nearest Neighbor Nonparametric Regression An Introduction to Kernel and Nearest Neighbor Nonparametric Regression,” 1991.

[64] R. A. Stine, “Sentiment analysis,” *Annu. Rev. Stat. Its Appl.*, vol. 6, 2019, doi: 10.1146/annurev-statistics-030718-105242.

[65] G. H. John and P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers”, Accessed: Aug. 20, 2021. [Online]. Available: http://robotics.

[66] “sklearn.svm.LinearSVC — scikit-learn 0.24.2 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC (accessed Aug. 20, 2021).

[67] “RBF SVM parameters — scikit-learn 0.24.2 documentation.” https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html (accessed Aug. 20, 2021).

[68] G. James, D. Witten, T. Hastie, and R. Tibshirani, “Springer Texts in Statistics An Introduction to Statistical Learning”, Accessed: Aug. 21, 2021. [Online]. Available: http://www.springer.com/series/417

[69] “Qualtrics XM // The Leading Experience Management Software.” https://www.qualtrics.com/uk/ (accessed Sep. 01, 2021).

[70] “General Data Protection Regulation (GDPR) – Official Legal Text.” https://gdpr-info.eu/ (accessed Aug. 19, 2021).

[71] “UVE - Associação de Utilizadores de Veículos Eléctricos.” https://www.uve.pt/page/ (accessed Aug. 19, 2021).

[72] “r/TeslaMotors - Your Tesla Community on Reddit.” https://www.reddit.com/r/teslamotors/ (accessed Sep. 01, 2021).

[73] “r/TeslaLounge - Your Relaxed Tesla Community on Reddit.”

https://www.reddit.com/r/TeslaLounge/ (accessed Sep. 01, 2021).

[74] Y. Zuo, W. Song, Y. Lu, Q. Gu, and R. Sun, “Interior noise sound quality prediction based on least square support vector machine-sample entropy | 基于LSSVM-样本熵的车内噪声声品质预测,” *Jiangsu Daxue Xuebao (Ziran Kexue Ban)/Journal Jiangsu Univ. (Natural Sci. Ed.*, vol. 41, no. 1, pp. 1–7, 2020, doi: 10.3969/j.issn.1671-7775.2020.01.001.

[75] “Como Carregar um Veículo Elétrico - UVE.” https://www.uve.pt/page/como-carregar-um-veiculo-eletrico/ (accessed Jun. 07, 2021).

[76] E. D. Kostopoulos, G. C. Spyropoulos, and J. K. Kaldellis, “Real-world study for the optimal charging of electric vehicles,” *Energy Reports*, vol. 6, pp. 418–426, Nov. 2020, doi: 10.1016/J.EGYR.2019.12.008.

[77] J. Du, X. Zhang, T. Wang, Z. Song, X. Yang, H. Wang, M. Ouyang, and X. Wu, “Battery degradation minimization oriented energy management strategy for plug-in hybrid electric bus with multi-energy storage system,” *Energy*, vol. 165, pp. 153–163, 2018, doi: 10.1016/j.energy.2018.09.084.

[78] Q. Cui, S. Du, C. Liu, L. Zhang, and G. Wei, “A stochastic optimal energy management strategy considering battery health for hybrid electric bus,” *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 234, no. 13, pp. 3112–3127, 2020, doi: 10.1177/0954407020924285.

[79] H. Qiao and Q. Wei, “Functional nanofibers in lithium-ion batteries,” *Funct. Nanofibers their Appl.*, pp. 197–208, 2012, doi: 10.1533/9780857095640.2.197.

[80] X. Tang, K. Liu, X. Wang, F. Gao, J. MacRo, and W. D. Widanage, “Model Migration Neural Network for Predicting Battery Aging Trajectories,” *IEEE Trans. Transp. Electrif.*, vol. 6, no. 2, 2020, doi: 10.1109/TTE.2020.2979547.

[81] X. Wu, X. Li, and J. Du, “State of Charge Estimation of Lithium-Ion Batteries over Wide Temperature Range Using Unscented Kalman Filter,” *IEEE Access*, vol. 6, pp. 41993–42003, 2018, doi: 10.1109/ACCESS.2018.2860050.