Data - Driven Analysis of EV Charging Infrastructure for a Smart City - Medium/Heavy Duty Vehicles

A Project Report

Presented to

The Faculty of the Department of Applied Data Science San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science in Data Analytics

By

Mahe Jabeen Abdul, Pranavi Sandrugu, Rohan Naga Venkata Mayukh Ungarala, Lohitha Vanteru

December 15, 2023

APPROVED FOR DEPARTMENT OF APPLIED DATA SCIENCE

Dr. Jerry Gao, Project Advisor

Dr. Lee C. Chang, Department Chair

ABSTRACT

The advent of electric vehicles (EVs) has ushered in a transformative era in urban transportation.In response to the rising prominence of electric vehicles (EVs) in urban transportation, this project conducts a meticulous data-driven analysis, concentrating on the unique challenges and opportunities presented by the charging infrastructure tailored for medium/heavy-duty vehicles within the dynamic context of a smart city. The overarching objectives of the study encompass forecasting the expanding demand for EVs, mitigating concerns related to driving range limitations, optimizing energy resource allocation, and strategically situating charging stations. Leveraging advanced forecasting models, such as the Prophet model for market growth and a Stacking Ensemble Regressor model with weighted fusion for range prediction, the study anticipates and addresses critical issues associated with EV adoption. Additionally, the Temporal Fusion Transformer model is applied to optimize energy resource allocation at EV charging stations, ensuring sustainability and cost-effectiveness. The placement of charging infrastructure is strategically addressed through the integration of a PuLP linear programming optimization model with K-means clustering, enabling the prediction of optimal locations for new charging stations that cater to diverse vehicle types and align with the evolving needs of urban mobility. This comprehensive project, employing cutting-edge machine learning methodologies, encapsulates a forward-thinking analysis of the evolving landscape of EV charging infrastructure. The findings aim to contribute actionable insights for the enhancement and strategic expansion of EV charging infrastructure, thereby propelling sustainable urban mobility into the future.

Keywords: *Machine Learning, Stacking Ensemble Regressor, Temporal Fusion Transformer*

ACKNOWLEDGMENTS

We would like to express our gratitude and special thanks to our project guide, Dr. Jerry Gao (Zeyu), Department of Applied Data Science, San José State University, and supervisor Dr. Lee C.Chang who is the Director of the Department of Applied Data Science at San José State University for providing us with a wonderful opportunity to study, research, and work on the project related to Natural Language Processing in the field of Artificial Intelligence. We would like to sincerely appreciate Dr. Jerry Gao (Zeyu), efforts in guiding the entire team in the right direction to carry out this project successfully. His motivation and guidance helped us to push ourselves in the most critical stages of our project. We would also like to appreciate all his efforts and the time he has given to our team to make this project a successful one by providing all the problem-solving guidelines.

Towards the end of the project, we were confident that we are technically skilled enough to take up any real-time project in a company and work on it. All the credits go to our project guide, Dr. Jerry Gao (Zeyu), Department of Data Science, San José State University. He taught us how to be open-minded and accept challenges and work on them step by step to achieve the end results successfully.

Furthermore, we would like to appreciate each other's efforts in the team for working in different timings coordinating well within the team, and performing the tasks that we assigned each other. The Data Collection part was the crucial stage for this project and our teammates together catered a lot of time and effort in forming the dataset without which getting good output results for our project would not be possible.

We would also like to thank our supervisor Dr. Lee C. Chang for designing this program and allowing us to understand our technical capabilities. Without this program, we would not have been able to upscale our knowledge and problem-solving skills. We are also grateful to our spouses for all their understanding and support throughout our project tenure. Finally, we would like to thank our family and friends for all their support and encouragement throughout the completion of this project work.

TABLE OF CONTENTS

1. Introduction

1.1 Project Background and Executive Summary

The transportation industry is undergoing a substantial shift towards environmentally conscious and sustainable methods, with a particular focus on electric vehicle technology. In this context, medium- and heavy-duty electric vehicles (EVs) are beginning to show promise as viable alternatives to their combustion engine counterparts. However, the successful integration of electric vehicles into the transportation system depends on the construction of a robust infrastructure for charging them. Authorities in the United States and around the world have recently been more and more skeptical about global warming, particularly in light of rising levels of greenhouse gas emissions. According to the U.S. Environmental Protection Agency (EPA), transportation is the main cause of greenhouse gas (GHG) emissions in the country, producing about 29% of all GHG emissions [1]. Promoting the use of electric cars (EVs), which are a well-liked remedy for the pollution produced by automobiles running on fossil fuels, has grown more important as the hazard of climate change grows. Electric vehicles are essential for California to achieve its aspirational environmental and air quality targets. Without these cars, which have no tailpipe emissions, the state's goal of having 1.5 million zero-emission vehicles on Californian roads by 2025 would not be accomplished. Nevertheless, many important private sector automakers, such as Ford, BMW, Volvo, and others, have stated that they would switch to totally electric vehicles within the next ten years. If all of these changes in the public and private sectors were implemented, the demand for electric vehicles would skyrocket (EVs). To facilitate this shift, the state will need to invest a large amount of money in the infrastructure for electric vehicle charging. The California Energy Commission estimates that by 2030, the state may need up to 1.2 million EV chargers to support the eight million electric passenger cars that are anticipated, as well as an additional 157,000 chargers to support non-passenger vehicles like trucks and buses. The charging needs of trucks and buses differ greatly from those of individual automobiles in terms of power demands, locations, and access. In addition to the $384 million the state would receive from the federal government over the following four years, the California Energy Commission (CEC) and the California Air Resources Board (CARB) recently announced that the state of California would contribute more than $5.5 billion to this effort.

In the current landscape of urban transportation [2], heavyweights in the form of medium/heavy-duty electric vehicles (EVs) find themselves in the spotlight due to their pivotal role in reducing emissions. However, a notable gap exists in the availability of tailored charging infrastructure to support these vehicles. The existing charging infrastructure predominantly caters to light-duty vehicles, revealing a disparity that calls for a robust and dedicated solution. This underscores the urgency for a comprehensive analysis and enhancement of the charging ecosystem, ensuring that all segments of the EV market are adequately served.

Within the broader context of smart city initiatives, our project aligns seamlessly with the vision of a sustainable and energy-efficient urban environment. By contributing to the development of a green urban blueprint, we transcend the conventional notion of charging infrastructure as a mere means of powering vehicles. Instead, our focus extends to shaping a smarter and greener urban landscape. Beyond the act of charging, our project seeks to integrate seamlessly into the fabric of smart cities, fostering sustainability and contributing to the holistic development of urban spaces. In essence, it encapsulates a forward-thinking approach that not only addresses the charging needs of heavy-duty EVs but also envisions a future where urban mobility is intricately linked with ecological balance and energy efficiency. Medium and heavy-duty electric vehicles, such as buses and trucks, play a crucial role in mitigating carbon emissions [3] and improving overall energy efficiency within the transportation industry. In the dynamic urban environment of San Jose, continual construction endeavors and infrastructure enhancements are prevalent. Essential for the transportation of materials, execution of construction duties, and playing a pivotal role in the city's urban advancement, heavy-duty vehicles, including construction trucks and equipment, are indispensable. Medium and heavy-duty vehicles play a vital role in San Jose's public transit network, serving as essential elements. They offer a crucial means of transportation for residents, aiding in the alleviation of traffic congestion and diminishing reliance on individual car usage.

In the dynamic landscape of sustainable transportation, pioneers in the medium and heavy-duty electric vehicle market, including trailblazers like Blue Bird, Lion Electric, and BYD, are at the forefront of driving innovation as shown in Figure 1. These companies are instrumental in the electrification of school buses, trucks, and diverse commercial vehicles, catalyzing a shift toward environmentally friendly practices. Their commitment to advancing electric commercial vehicles underscores a collective endeavor to reduce carbon footprints and enhance the sustainability of transportation networks. The sector is witnessing notable technological trends, including advancements in fast-charging technologies, battery-swapping solutions, and smart grid integration, underscoring a commitment to efficiency and innovation in the EV charging ecosystem. Moreover, as environmental consciousness [4] intertwines with economic considerations, consumer adoption of electric medium and heavy-duty vehicles is on the rise. Fleet operators, logistics companies, and municipal services are increasingly drawn to electric vehicles, recognizing the dual benefits of addressing environmental concerns and realizing potential cost savings over time. This collective momentum among industry pioneers, charging infrastructure leaders, and a growing consumer base reflects a transformative shift towards a more sustainable and electrified future in commercial transportation.

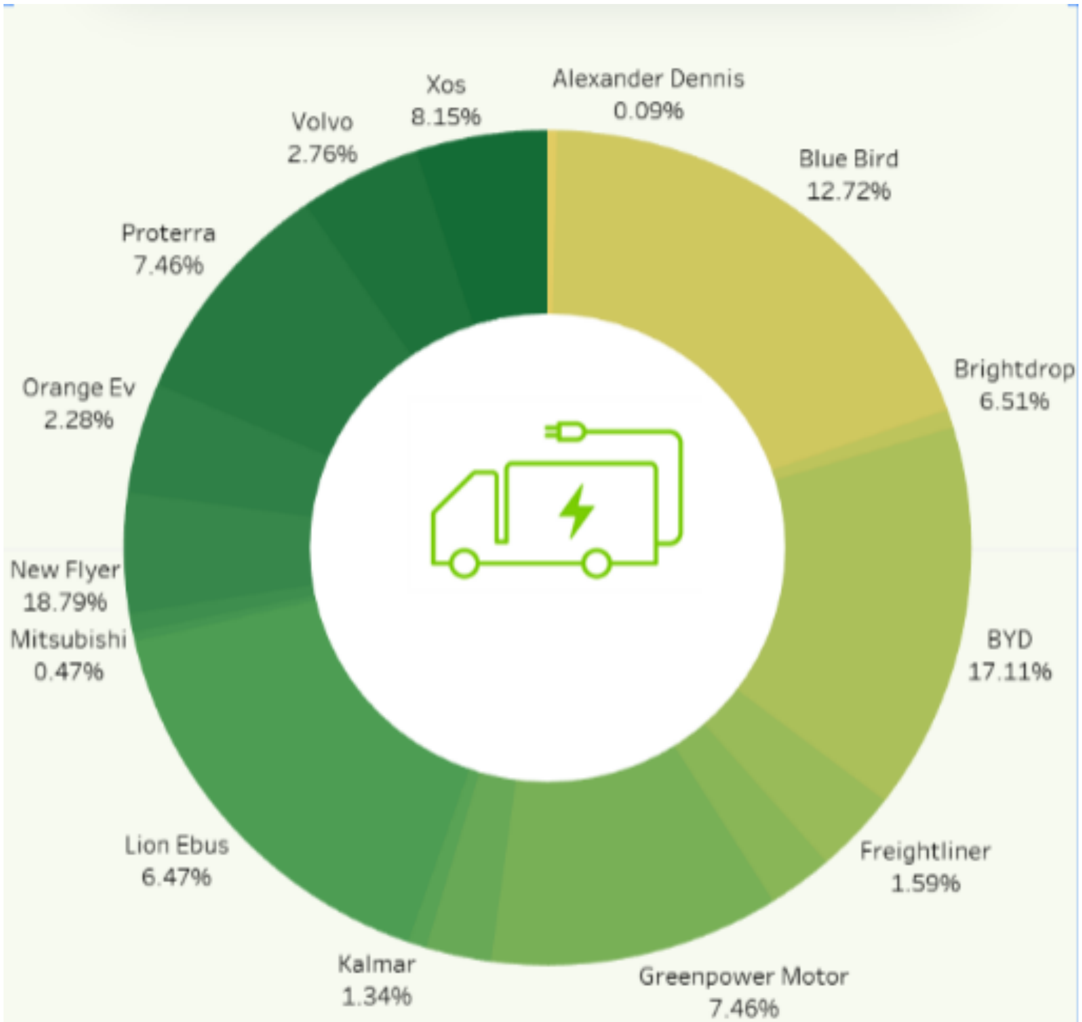


Figure 1. Pioneers of Electric Commercial Vehicles

This project seeks to perform an extensive data-driven examination of the charging infrastructure that supports medium and heavy-duty EVs. By exploring essential metrics, usage trends, and challenges linked to charging infrastructure, the project aims to offer valuable insights that will guide strategic decision-making, policy creation, and the development of infrastructure. The primary problem targeted for the project is to address the limited availability of EV charging stations for commercial transportation by Placing optimal locations of New Charging Stations for each vehicle type such as Transit Buses, Delivery Trucks, and School Buses in the community, and to support the transition to be a more sustainable and environmentally friendly transportation system. There are three objectives of the project Firstly, to reduce carbon footprint and improve air quality. Secondly, to promote the widespread adoption of public electric vehicles by providing an affordable and accessible charging infrastructure. Furthermore, it will help to support the growth of the EV markets, creating new jobs and economic opportunities in the city. To achieve these goals, the project approach involves four tasks. Initially, the proposal entails forecasting the demand for Heavy Duty and Medium Duty Electric Vehicles (EVs) along with the required charging stations. The plan involves developing a predictive model using the Prophet model to anticipate EV demand, encompassing vehicle counts and the corresponding number of needed charging stations, up until the year 2035. We will assemble an extensive dataset containing information on vehicle quantities, zip codes, fuel types, vehicle brands, duty classifications, and production years. Secondly, Predicting Vehicle Range of Heavy Duty and Medium Duty EVs. Forecasting the range of a vehicle through charging data involves grasping the variations in the vehicle's state of charge (SOC) over time and comprehending its consequential impact on driving distance. Enhancing the operational efficiency of electric vehicles, including transit buses, school buses, and delivery trucks, entails making predictions about their range using charging data. Our research is centered on leveraging analytics of charging data to estimate the range of these three vehicle categories, aiming to enhance decision-making in fleet management. The third goal is Predicting Short and Long-term Energy Demand for Heavy Duty and Medium Duty EVs. The effective utilization of energy resources, the streamlined operation of diverse vehicle fleets—

ranging from transit buses to school buses and delivery trucks—and the optimization of charging infrastructure all hinge on the capability to anticipate the energy consumption of electric vehicles (EVs). This use case incorporates both long-term and short-term projections to meet various planning and operational goals. Finally, strategically locating new charging stations to align with the charging needs of diverse vehicle types to promote the uptake of electric vehicles (EVs) and advance sustainable transportation practices. After finishing these tasks, we will assess the performance of our models using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R2. Additionally, precision, recall, and F1 score will be considered for predictions on a station-wise, site-wise, and zipcode-wise basis. Subsequently, our attention will shift towards the development and implementation of the system, incorporating predictions for new charging station locations tailored to each vehicle category, namely school buses, transit buses, and delivery trucks, within a specified zip code in San Jose city of California.

Our project prioritizes the advancement of electric vehicles (EVs) for use in public transportation systems within the community. Through these initiatives, we anticipate fostering innovation in clean energy technology and the infrastructure of EV charging stations within the community.

The primary contributions of our project are outlined below:

1. Strategic Infrastructure Planning: Through a thorough examination of charging infrastructure, the project plays a pivotal role in strategic planning for the establishment of charging stations catering to medium and heavy-duty electric vehicles. This encompasses discerning insights into ideal locations, charging capacities, and the specific types of vehicles accommodated.

2. Anticipating Future EV Demand: A key contribution lies in our capacity to predict the demand for both heavy-duty and medium-duty electric vehicles. This foresight empowers stakeholders, including policymakers, manufacturers, and infrastructure developers, to anticipate the expanding market and adapt their strategies accordingly.
3. Forecasting Energy Consumption: The project's ability to predict the energy consumption of heavy and medium-duty vehicles is a significant feature. By understanding the fluctuations in the state of charge (SOC) over time and how it impacts driving distance, the project aids in forecasting the energy needs of these vehicles, which is pivotal for ensuring efficient fleet management.
4. Long-Term and Short-Term Planning: Forecasting the energy demand for medium and heavy-duty vehicles in both the long and short term has multifaceted benefits. It aids in strategic planning, enhances operational efficiency, facilitates optimal resource allocation, promotes cost optimization, fosters environmental sustainability, encourages technology innovation, and allows adaptability to market changes. The inclusion of both long-term and short-term projections within the project is a significant asset, ensuring that the predictions address immediate operational requirements while also facilitating strategic planning for the future.

The findings from our project, which include forecasting EV count and charging station demand, predicting vehicle range, and anticipating both long and short-term energy demand, offer numerous benefits. Initially, the application of the Prophet model to forecast EV count and charging station demand creates a data-driven guide for future EV infrastructure development. This holds significant value for legislators, utility corporations, and infrastructure planners, facilitating more efficient resource allocation and investment decisions through precise estimates of car counts and charging station needs. The second advantage lies in predicting vehicle range through the application of a stacked regressor model. This proves advantageous in comprehending the fluctuations in the vehicle's state of charge (SOC) over time and how it influences driving distance. Such understanding is essential for enhancing the operational efficiency of electric vehicles such as transit buses, school buses, and delivery trucks. Our research concentrates on leveraging charging data analytics to gauge the range of these vehicle types, thereby enhancing decision-making in fleet management. Finally, the prediction of both short and long-term energy demand using a temporal fusion transformer model offers significant advantages by addressing both temporal scopes. Long-term energy demand forecasting provides vital insights for strategic planning and resource allocation, encompassing weekly, monthly, and yearly perspectives. This equips fleet managers and infrastructure planners with the ability to proactively manage variations in energy demand and ensure the availability of ample charging resources. On the other hand, short-term energy demand forecasting is focused on predicting energy needs for the upcoming day or within a shorter timeframe, typically within the next 6 hours. This form of forecasting plays a critical role in energy load balancing and making real-time operational decisions, contributing to heightened operational efficiency, minimized energy waste, and the promotion of environmentally friendly transportation practices. The targeted problems along with the project deliverables are outlined in the following Table 1.

Table 1. Targeted Problems and Project Deliverables

Targeted Problems	**Project Deliverables**
Anticipating EV Market Growth	Count Forecast of Heavy/Medium Duty EV Vehicles and Charging stations using Prophet.

Table 1. *Cont.*

Targeted Problems	Project Deliverables
Addressing Range Anxiety	Range Prediction using Stacking Ensembler Regressor Model.
Efficient Energy Resource Allocation	EV Charging Station Energy Demand Prediction using Temporal Fusion Transformer Model.
Strategic Charging Infrastructure	Optimal Location Prediction for New Charging Station using PuLP linear programming optimization model with K-means clustering model.

1.2 Project Requirements

The aim of this research focused on the data-driven analysis of EV charging station infrastructure for medium and heavy-duty vehicles, is to utilize data and analytical methods to comprehend the current status and anticipate future requirements of EV charging infrastructure. To identify locations in need of additional charging stations, it is imperative to analyze data on the current charging station locations, usage patterns, and capacities, as well as demographic and geographic information. The project involves the application of predictive modeling and simulation techniques to forecast future demand for charging infrastructure and optimize the strategic placement and design of additional charging stations. Ultimately, the project seeks to offer insights to policymakers, infrastructure planners, and other stakeholders, facilitating the creation of a more efficient and effective charging infrastructure to support the transition to electric vehicles.

Functional requirements refer to the attributes and capabilities that a system must possess or be able to achieve. Regarding data collection and management, the system should have the capacity to gather data from diverse sources, including information from charging stations, vehicle registrations, traffic statistics, California zip codes, transit bus stops, and vehicle fleet transactions, consolidating it in a centralized database. To assess the system's capability to handle extensive data while upholding data integrity, it becomes crucial to scrutinize the completeness and quality of the acquired data. The system should possess the ability to analyze the gathered data, identifying patterns, trends, and insights about EV charging station usage, such as peak hours, popular locations, and charging durations. The accuracy of the analysis and the relevance of the generated insights can be used to evaluate this requirement. Additionally, the system should be equipped to forecast future requirements for EV charging infrastructure based on current usage patterns and other relevant factors like industry expansion, shifts in consumer behavior, and modifications in governmental regulations. This can be verified by comparing anticipated demand with actual consumption trends over time. Taking into account variables such as location, capacity, and user requirements, the system must possess the capability to optimize the placement and design of EV charging stations to meet both current and future demand. The effectiveness and efficiency of the optimized charging infrastructure can be assessed to verify this requirement. Additionally, the system should be adept at generating reports and visualizations that communicate outcomes and recommendations to stakeholders. The evaluation of the usefulness and relevance of these reports and visualizations to stakeholders serves as a test for this requirement. Furthermore, the initiative should outline specifications for seamlessly integrating data and analysis outputs with existing systems and platforms. Testing these specifications involves assessing the ease of integration with current systems and the accessibility and utility of shared data and analytical outputs for end-users. The functional requirements of our project are comprehensively outlined in Table 2.

Utilizing AI features enhances the system's sophistication and autonomy. The forecasting capabilities of the Prophet model and Neural network architecture particularly in forecasting

Table 2. Functional requirements of the project

Task	Functional Requirement
Data collection and management	• Ability to collect data from multiple sources. • Data cleaning and preparation to ensure accuracy and consistency. • Robust data management system to store and organize data. • Ability to update data regularly.
Data analysis and modeling	• Ability to perform statistical analysis on the data. • Ability to develop predictive models to forecast future demand for charging infrastructure. • Ability to simulate different scenarios to optimize the placement and design of charging stations. • Ability to generate visualizations and reports to communicate insights and recommendations.
Efficiency Evaluation	• Ability to assess the efficiency and efficacy of the optimized charging infrastructure to ensure optimal performance. • Ability to assess how well the optimized infrastructure contributes to reducing carbon emissions and promoting sustainable transportation practices. • Ability to use Continuous monitoring and feedback mechanisms to contribute to ongoing improvements and the sustained efficiency of the EV charging system.
Data Visualization	• Ability to display charging station locations on an interactive map. • Ability to display real-time data on charging station availability for each vehicle type, occupancy, and charging speed. • Ability to customize the visualization based on user preferences.
Integration with existing systems	• Ability to integrate with geographic information systems (GIS) to visualize data on maps. • Ability to integrate with transportation planning and decision-making tools. • Ability to share data and analysis results with stakeholders.
Ease of Integration Assessment	• Ability to assess the ease of integration with the existing systems and the availability of shared data and analytical outputs for customers. • Ability to assess the robustness and flexibility of the system's Application Programming Interface (API).

charging station usage, energy demand, and availability, can be assessed by comparing predicted values with actual results and evaluating predictive accuracy. The efficacy of optimization algorithms in improving the positioning and design of new charging stations can be gauged by contrasting enhanced solutions with real-world outcomes. Pattern recognition algorithms driven by AI can be evaluated for precision and efficiency in identifying patterns and anomalies in charging station data. This includes recognizing stations frequently out of service or detecting consumption patterns inconsistent with typical usage. The effectiveness of AI-powered recommendation engines in suggesting optimal charging stations based on customer preferences is assessable by comparing recommendations with actual usage patterns. To test the system's robustness, performance, and reliability under different stress and load conditions, experiments can be conducted. This evaluation determines how well the AI-powered system can manage substantial data volumes and operate reliably and efficiently across various scenarios. Details regarding the AI requirements of the project are outlined in Table 3.

To assess the distribution and concentration of charging stations across diverse zip codes and counties in California, obtaining precise location data for charging stations is essential. A comprehensive understanding of electric vehicle counts and the demand for charging infrastructure, especially in areas with either underutilized or overburdened charging stations, necessitates access to vehicle registration data. This includes information on fuel types and vehicle manufacturers. Analyzing the energy demand of charging stations requires charging transaction data, encompassing details such as the starting state of charge (SOC) and ending SOC. For effective analysis and decision-making tailored to specific locations, geographic and environmental data play a pivotal role. This data may include population density, traffic patterns, and environmental considerations, aiding in pinpointing areas that require additional charging

Table 3. AI requirements of the Project

AI Requirement	Description	Measurable Metric
Predictive modeling	• The ability to use historical data to forecast Electric Vehicle count and future demand for Charging stations. • The ability to use EV transaction data to predict vehicle range to optimize the operating efficiency of EV.	Prediction accuracy, error rates, and correlation coefficients.
Sequence to Sequence modelling	The ability to use EV transaction data and consider the temporal elements like day of the week, month, and year to determine EV charging station energy demand.	Prediction accuracy, temporal accuracy, seasonal pattern recognition, peak demand prediction, utilization efficiency.
Linear Programming plus Clustering analysis	The ability to utilize the capabilities of linear programming and clustering method to suggest optimal locations for EV charging stations	Clustering analysis, precision values, recall values and F1 score.
Visualization	The ability to present data in a clear, intuitive, and interactive way, such as maps, graphs, and heatmaps	User engagement metrics, such as time spent on the platform, clicks, zooms, and searches.
Optimization algorithms	The ability to optimize the placement and design of new charging stations based on various criteria, such as vehicle range, EV demand, and accessibility	Charging station utilization, coverage and accessibility, EV range coverage, demand satisfaction and user satisfaction.

infrastructure based on requirements and constraints. To predict the demand for charging infrastructure and optimize the design and capacity of charging stations, detailed information on the usage of electric vehicles becomes paramount. This includes data on the quantity, types, and battery capacities of medium and heavy-duty electric vehicles. Understanding public transportation involves sourcing transit bus stop data for San Jose and providing insights into the locations of bus stations. Additionally, obtaining school bus terminal data for San Jose contributes to a comprehensive overview of school bus terminal locations. For a deeper understanding of commercial fleet operations, heavy and medium-duty vehicle fleet transaction data has been acquired. This dataset encompasses real-world data for various weight classes of commercial fleet vehicles. In summary, the table labeled Table 4, provides a breakdown of the diverse data aspects essential for the successful execution of this project.

Table 4. Data requirements of the project

Data Requirement	Description	Source
CA Vehicle registration data	Number of vehicles registered each year for every zip code and county in CA	California Department of Motor Vehicles
CA EV charging stations data	Real-time data of EV charging stations for each zip code in CA	U.S. Alternate Fuel Data Center
Transit bus stops data	Data of public bus station locations for transit bus in San Jose	CA open data portal
School bus terminals data	Data of school bus terminal station locations in San Jose	CA Department of Education
Heavy/Medium vehicle fleet transaction data	Real-world data of commercial fleet vehicle operating data for each weight class	National Renewable Energy Laboratory

The team comprises members dedicated to specific use cases outlined in the project, including: 1) Forecasting electric vehicle count and charging station demand for Medium and Heavy-duty EVs using time series modeling, 2) Predicting the vehicle range of Medium-duty and Heavy-duty EVs through ensemble machine learning techniques, 3) Projecting short and long-term energy demand of medium and heavy-duty EVs using a sequence-to-sequence modeling approach, and 4) Optimizing the placement of new charging stations for each vehicle type using linear programming and clustering methods. Responsibilities for these tasks have been distributed among team members to ensure the timely achievement of objectives. Each member's roles and responsibilities are detailed as follows:

Table 5. Team members and their allocated roles

Task	**Assigned**
Data Extraction	Lohitha, Mahe, Pranavi
Data Cleaning	Pranavi, Rohan
Data Transformation	Pranavi, Lohitha
Feature Selection	Mahe, Lohitha, Rohan
Exploratory Data Analysis	All
Forecasting EV count and charging station demand of Medium and Heavy duty EVs based on electric vehicle registrations	Pranavi, Mahe
Predicting vehicle range of Medium and Heavy duty EVs based on Medium/Heavy duty transactions	Lohitha, Mahe
Predicting short and long-term energy demand of Medium and Heavy duty EVs based on Medium/Heavy duty transactions	Pranavi, Rohan
Optimal placement of new charging stations for each vehicle type such as school bus, transit bus and delivery truck	Lohitha, Rohan
Final Report and Presentation	All

1.3 Project Deliverables

As part of the project, deliverables include a project proposal that defines the research problem statement, related references, and a background survey. Using the Cross-Industries Standard Process for Data Mining (CRISP-DM), these deliverables are divided into various stages. A literature review is conducted to identify the research gap, followed by an analysis of possible technology solutions to the problem statement. The project's next phase begins with an introductory chapter that describes the project's requirements and defines the scope of the project. Among the deliverables of the third phase is the Data and Management Plan for the project, which discusses how data is collected, as well as the insights gained by studying the given data, as well as how the tools and software will be utilized with cost justifications.

As part of this phase, a Work Breakdown Structure (WBS) will be provided, displaying each task's timeline. WBS can be created with project management tools such as JIRA or ClickUp. According to CRISP-DM, this matrix shows the hierarchical representation of work packages, tasks, and deliverables of a project. A research problem requires a certain amount of effort to solve. This is the effort estimation. Based on the project's complexity and the availability of individual team members, various processes are estimated. In agile methodologies, work is estimated using story points. You can calculate the time required to complete a user story based on the complexity of the task. There are two types of project scheduling charts: Gantt charts and PERT charts. There is significance to each of them. A Gantt chart is a graphical representation of timelines, tasks, sub-tasks, dates, milestones, and resources allocated to each task/sub-task. By using a Gantt chart, it shows how the subtasks are related to one another. The PERT chart is used to organize and schedule projects by breaking down, mapping, and visualizing tasks. An estimate of the time needed to complete a project can be made using a PERT chart.

As soon as the data and project management plan are complete, the next phase will be data engineering. In this phase, data is preprocessed, transformed including dimensionality reduction, and provides the final data set for training and testing the modeling models. In this phase, algorithms' performance parameters are measured and compared. Each team member will submit a report explaining the model's features, framework, efficiency, and limitations.

The final deliverables include California, will be able to manage its electric vehicle recharging infrastructure comprehensively. The proposed solution comprises an interactive website interface that is designed to facilitate user interactions by allowing queries for analytical results and providing a user-friendly mechanism for suggesting new charging station locations. Fleet managers can use this information to ensure the smooth and effective running of their EV fleets by learning the best charging locations. Users can benefit from the system's accuracy in predicting optimal charging station locations, enabling them to plan routes and optimize charging stops. The results of the various experiments on the datasets will be compared in a table comparing the results of these models. At the end of the project, the group submits a report and presentation. The project deliverables are listed in Table 6, along with their corresponding descriptions and timelines.

Table 6. Project Deliverables with their respective description and due dates.

Deliverable	Description	Due Date
Project Abstract	Contains a literature survey and a sample data file. In addition, it contains the background, examples of data sources, motivation, and the problem statement for the study.	17 February 2023
Data Preparation Plan	The study of data generation models and the understanding of the public transportation management system helps fleet operators better manage charging schedules.	10 March 2023
Project Introduction	The document includes the project background and executes summary, project requirements, project deliverables, technology, and literature surveys.	15 March 2023

Table 6. *Cont.*

Deliverable	Description	Due Date
Work Breakdown Structure	In a hierarchical structure, the work must be done by the project team to produce the required deliverables. This breakdown includes the start and end dates of each task.	22 March 2023
Clean Data Set	An exploratory data set using dimensionality reduction and exploratory analysis, ready for modeling.	10 April 2023
Data Engineering	This stage provides a complete explanation of the data process, data collection, data pre-processing, data preparation, data transformation, and data statistics steps.	14 April 2023
Research Report	The report contains all the necessary details about the research, such as how the data was transformed, project requirements, and methodologies.	12 May 2023
System Design and Implementation	This document contains the system development process, where the detailed structure and functionality of the system are defined.	02 October 2023
Model Evaluation and System Visualization	This stage provides model improvements, system components, and processes, facilitating a clearer understanding of the system's architecture and functionality.	13 November 2023
Web system design and development	Provides a user interface for a web-based application, creating a functional and interactive web system.	27 November 2023
Research Report Presentation	ıThe final research report will be presented in a PowerPoint presentation on this day.	11 December 2023
Final Project	The project integrates with the public transportation management system to empower users to easily locate optimal charging stations. By using the website, users can locate charging stations nearby, making it easier for them to adopt electric vehicles.	15 December 2023

1.4 Technology and Solution Survey

In this section, configurations of multiple types of charging stations were examined to select the relevant technology for our use case. Later, the suitable models were scrutinized and compared on a higher level and then an in-depth comparison analysis among all the existing technological surveys was done which will give a clear overview of the model requirements of this research.

Table 7 provides a summary of different EV charging technologies and their respective specifications. It can be used to compare the effectiveness and efficiency of different charging technologies in a data-driven analysis of EV charging infrastructure. Here, various technologies for electric vehicle (EV) charging infrastructure are briefly summarized. These technologies include Level 1 and Level 2 charging, DC fast charging, wireless charging, smart charging, solar-powered charging, and battery swapping. Power output, connector type, network accessibility, location, utilization rate, and availability are a few of the criteria used to evaluate each technology. In both residential and commercial settings, level 1 and level 2 charging are typical. They provide medium power outputs and have affordable installation and maintenance expenses. DC fast charging is better suited for high-traffic locations due to its higher power outputs but higher installation and maintenance expenses. Physical connectors are not required for wireless charging; however, installation costs are higher and usage rates are lower. Smart charging features high utilization rates, low installation, and maintenance costs, and optimizes charging based on demand and grid circumstances. Although solar-powered charging has lower power outputs and moderate installation costs, it decreases pollution and reliance on the grid. While battery swapping offers quick charging times, it also necessitates greater room and more expensive infrastructure. Finding the best options for various locations and use cases requires comparing and assessing various EV charging systems. To support the rising demand for electric vehicles, this information can assist in guiding decisions regarding where and how to invest in EV charging infrastructure.

Table 7. Technology Comparison Table for EV Charging

Technology	**Level 1**	**Level 2**	**DC Fast Charging**	**Wireless Charging**
Power Output	1.4 kW	7.2 kW	50-350 kW	3.7-7.7 kW
Charge Time (0-100%)	8-12 hrs	4-8 hrs	20-60 min	3-4 hrs
Connector Types	NEMA 5-15	J1772	CCS1, CCS2, CHAdeMO, Tesla Supercharger	Qi, SAE J2954
Compatibility	Light Duty	All EVs	Medium and Heavy Duty	Limited
Cost	Low	Medium	High	Very High
Convenience	Low	High	Medium	High
Availability	High	High	Low-Medium	Low

Table 8 offers a thorough overview of the models utilized to analyze the infrastructure of electric vehicle (EV) charging stations. It describes the different models, their advantages and disadvantages, and particular use cases. Traditional regression models are used to determine the number of stations required, while more sophisticated methods such as Temporal Fusion Transformer (TFT) are used to forecast the demand for energy consumption in the smart grid. Each model is designed to a different set of requirements, providing advantages as well as limitations. Some of the models anticipate the occupancy of infrastructure, others identify clusters with similar usage patterns, and yet others optimize the placement of charging stations inside road networks. A model is chosen based on the specific use case and the intended result, considering variables like interpretability, computational complexity, and handling of various data kinds.

Table 8. Technology Survey on Models Used for EV Charging Station Infrastructure Analysis

Ref ID	Model	Description	Pros	Cons	Use case
[5]	Regression models	A statistical model that analyzes the relationship between dependent and independent variables.	Easy to interpret, can be used with different data types.	May not capture complex relationships or nonlinear patterns in data.	Predicting the number of EV charging stations needed in a specific location.
[6]	Time-series forecasting models	Models that analyze historical data to make predictions about future trends.	Can capture seasonal patterns and make predictions based on historical data.	May not account for sudden changes or external factors that could impact trends.	Forecasting the demand for EV charging stations in a specific location.
[7]	Neural networks	A model that simulates the behavior of the human brain to process complex data.	Can capture complex and non-linear relationships, can handle large datasets.	Requires a large amount of data to train, can be computationally intensive, and can be difficult to interpret.	Predicting the optimal placement of EV charging stations.
[8]	Clustering	A technique that groups similar data points based on their characteristics.	Can identify patterns and relationships in data that may not be immediately apparent.	Can be difficult to interpret and may require subjective determination of the number of clusters.	Identifying clusters of EV charging stations that have similar usage patterns.

Table 8. *Cont.*

Ref ID	Model	Description	Pros	Cons	Use case
[9]	Prophet	The prophet model is used to identify and simulate the time-dependent patterns and structures that are present in the charging load data for EVs.	Can manage unavailable data efficiently, distribution of uncertainty intervals.	Over-simplification of intricate patterns. Limited attention to outside influences.	Load Forecasting of Battery Electric Vehicle Charging Stations.
[10]	Gradient Boosting Classifier	A model that predicts the occupation status of charging infrastructure, utilizing binary training data.	Manages complex interactions well and provides excellent predicted accuracy.	Can be computationally demanding and prone to overfitting, particularly when dealing with noisy data.	Predict charging infrastructure occupancy.
[11]	Linear Programming	An optimization technique based on integer linear programming is used to place electric vehicle charging stations strategically throughout a network of roads.	By providing mathematical rigor and efficiency, linear programming ensures the best possible options for the location of EV charging stations.	Practical applications may face issues related to sensitivity to input data accuracy and significant computing complexity.	Efficient Placement of Electric Vehicles Charging Stations.
[12]	Temporal Fusion Transformer Model	To address the increasing energy consumption in the residential sector	TFT provides flexibility in predicting day, weekly, and monthly energy use prediction horizons.	The complex internal workings of the model present interpretability issues.	Forecasting energy consumption demand of customers in the smart grid

Table 9 provides a thorough summary of various research initiatives targeted at tackling significant components of integrating electric vehicles (EVs) into transportation networks. Every topic describes the objective, methods, and important findings of the corresponding studies and covers a variety of subjects, including energy-efficient routing, charging optimization, infrastructure planning, and charging station placement techniques. The research emphasizes the significance of evaluating different charging systems, investigating the expansion of electric vehicles in commercial sectors, and utilizing cutting-edge techniques such as SHapley Additive exPlanations (SHAP) to improve energy efficiency. It also recognizes the difficulties and constraints these studies face, from operational adaptability issues and computational complexity to the requirement for more thorough analyses, like comparing the evaluations of urban and rural routes or validating machine learning models against external datasets. In the ever-changing field of electric vehicle integration and sustainable transportation planning, the survey provides a useful resource by highlighting the developments, gaps, and possible directions for further research.

Table 9. Technology survey methods of different Electric Vehicles (EVs)

Objective	Methods	Description	Negatives
Planning Electric vehicle infrastructure for various types of vehicle	Heavy- and medium-duty vehicles [13]	Uses the Return-to-base model and the on-route charging model, and summarizes the challenges of charging commercial electric vehicles (CEVs) at public locations.	Need for a balanced exploration of alternative charging strategies beyond the return-to-base model and V2G technology.
	Heavy-duty trucks only [14]	Joint routing and charging (JRC) scheduling approach for electric trucks, to minimize costs and delivery delays.	Computational complexity of the optimization problem, operational adaptability in environments.

Table 9. *Cont.*

	Methods	Description	Negatives
	E-Trucks & Buses [15]	Growth of electric trucks and buses by industry engagement, facilitating the electrification of vehicles.	There is a need for adjustments in utility rates and a lack of adequate charging infrastructure.
Utilizing the Energy Demand of Electric vehicles (EVs)	Shapley Additive exPlanations (SHAP) [16]	Examines (BEVs) to identify factors influencing BEV energy consumption and provide insights for enhancing energy efficiency and informing transport policies.	More detailed comparative analysis of urban and rural routes, and validation of ML models against external datasets.
	Energy-efficient route planning [17]	Estimating energy consumption in electric vehicles compared to conventional methods.	Can improve by providing more detailed results of statistical tests.
	Hybrid simulation model [18]	Examine EVs through strategies addressing greenhouse emissions from electricity generation and EV adoption in passenger transportation.	Oversimplified alternative options, lack of a detailed transportation demand forecast, and unclear model validation.
Methods used for locating charging stations	Activity-based approach [19]	Locate the (sub)optimal locations for charging stations using multiday travel activity constraints.	Lack of information on grid stability and electricity demand charging infrastructure utilization.
	Route-based analysis [20]	Analyze the infrastructure development and EV efficiency in minimizing travel time and maximizing user comfort.	Uncertainty in optimal charging stop selection and waiting times was not included in the study.
	GIS-based approach [21]	Identify the optimal locations for EVSE in terms of improved efficiency and lower emissions.	Lack of available specialists in electric mobility and incomplete vehicle registration data.

1.5 Literature Survey of Existing Research

The objective of this section is to organize the literature survey on sustainable and efficient deployment of Heavy Duty and Medium Duty EVs: forecasting demand for EV and charging stations, predicting vehicle range, estimating short and long-term energy demand, and optimizing the placement of new charging stations to identify gaps in the field and, more importantly, to establish a standard for measuring our model execution performance. Several pertinent studies have already been cited in earlier sections. To categorize the overall strengths and shortcomings of those methodologies, we assess our findings holistically and suggest prospective paths for research into the field of heavy-duty and medium-duty electric vehicles (HEVs) and their associated charging infrastructure.

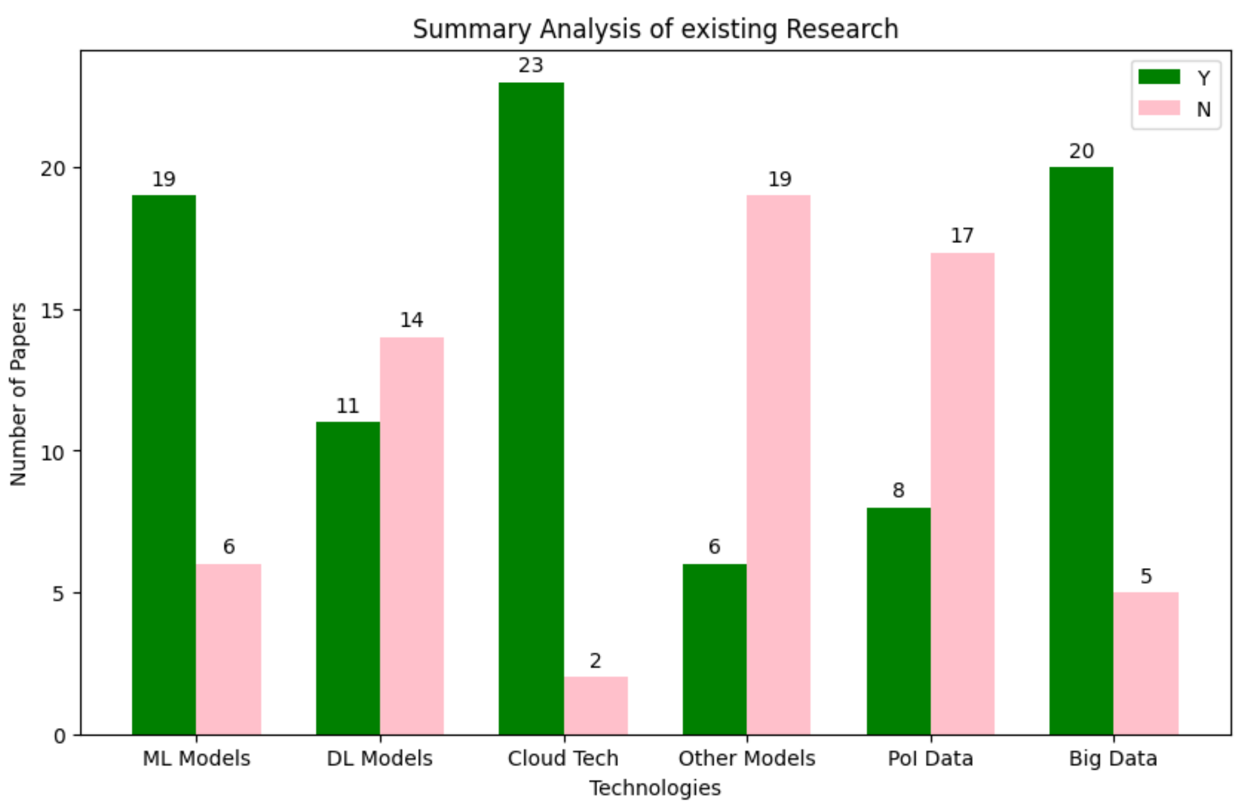


Figure 2. Analysis of existing research

1.5.1 Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand

The increasing use of Heavy Duty and Medium Duty Electric Vehicles (EVs) makes a detailed analysis of their demand necessary in the future. The current literature on predicting demand for these EVs and the related charging stations is examined in this review of the literature. A significant portion of research employing the ANN, RNN, LSTM, and Transformer models is accessible. To have a better understanding of implementation, a comparison of research publications based on various models is done in Table 10.

Table 10. Comparison of existing research on EV and Charging Stations Demand

Ref ID	Objective	Region	Dataset	ML Models	DL Models	Other models	Eval Metrics
[22]	Forecasting medium-term public EV charging demand	Scotland, UK	Eleven charging stations	LinReg, RF, SVM, KNN	ANN	ARIMA	MAE, RMSE, SMAPE
[23]	Forecasting demand for EV charging post-session start	Morocco	Two Public charging stations	N	ANN, RNN, LSTM, GRU	N	MAPE, RMSE
[24]	EV charging demand predictions for short-term and long-term forecasting	Boulder, Colorado	Twenty-five Public charging stations	N	ANN, LSTM, Transformer	ARIMA, SARIMA	RMSE, MAE
[25]	Addresses the challenges of mass electric vehicle (EV) adoption on power systems	Shenzhen China	real-world EV charging station datasets	N	DNN, RNN, LSTM, GRU	N	NRMS, normalized mean absolute error (NMAE)
[26]	Smart EVC solution—intelligent charging station management platform based on AI	United Kingdom	real-time data on the energy consumption and charging demand.	N	N	Custom Reservation Algorithm Reservation	Charging time

Table 10. *Cont.*

Ref ID	Objective	Region	Dataset	ML Models	DL Models	Other models	Eval Metrics
[27]	Forecast electric vehicle charging demand	Atlanta, USA	EV charging dataset of Georgia Tech	N	Deep Learning - based LSTM, DLSTM	arithmetic optimization algorithm (AOA), empirical mode decomposition (EMD).	MAE, RMSE, Accuracy

1.5.2 Predicting Vehicle Range of Heavy-Duty and Medium-Duty EV

Precisely forecasting the range of Heavy Duty and Medium Duty Electric Vehicles (EVs) is essential to performance optimization in the ever-ch anging world of electric transportation. Through an exploration of techniques and innovations, this study seeks to provide information that is essential for improving operational effectiveness and resolving customer issues. A comparison of research publications based on different models is done in Table 11 to provide a detailed comprehension of implementation.

Table 11. Comparison of existing research on Vehicle Range

Ref ID	Objective	Region	Dataset	ML Models	Deep Learning	Other models	Eval Metrics
[28]	Analyze the effects of different environmental parameters on energy consumption and driving range	Various regions of China	Real-world BEV	N	N	Micro-simulation model	Accuracy, Relative error, Absolute error.

Table 11. *Cont.*

Ref ID	Objective	Region	Dataset	ML Models	Deep Learning	Other models	Eval Metrics
[29]	Improve the driving range prediction accuracy	Beijing	Real-world data from Baic New Energy Automobile	Classification and Regression tree, MLR, GBDT	N	N	MAE, Accuracy Maximum, minimum error
[30]	Reduce driver's range anxiety by estimating the real-time energy consumption of EVs	United States	Nissan Leaf 2013, Argonne National Laboratory (ANL)	N	Deep Convolutional Neural Networks	N	RMSE, MAE, K-fold correlation
[31]	Predict the remaining driving range of EVs	Major cities in China	National Big Data Alliance of New Energy Vehicles	XGBT, Boosting Regression Tree, XGBoost	N	Blended Model	MAE, RMSE, MAPE
[32]	Accurately estimating the driving range of electric vehicles	China	National Monitoring and Management Platform for New Energy Vehicles	Gradient boosting decision tree (LGBM), SVM	N	N	MAE, MSE, RMSE, R2 (R-Square)
[33]	Impact of driving electric vehicles (EVs) at highway speeds, using auxiliary loads	Australia	Mitsubishi i-MiEV and Nissan Leaf test car data	N	N	Vehicle mathematical modeling	Mean unsigned error, Energy consumption

1.5.3 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV

Predicting the energy consumption of Heavy Duty and Medium Duty Electric Vehicles (EVs) both in the short and long term is critical as we move toward a sustainable transportation paradigm. In addition to addressing the changing environment of charging behavior and grid interactions, this research examines research methodology and insights into factors driving energy usage patterns. Table 12 presents an extensive overview of implementation through a comparison of research findings based on various models.

Table 12. Comparison of existing research on Energy Demand

Ref ID	**Objective**	**Region**	**Dataset**	**ML Models**	**Other models**	**Eval Metrics**
[34]	Predicting the energy usage during charging sessions for plug-in electric vehicles (PEVs)	Nebraska, USA	public charging stations	RF, SVM, Xgboost, LinReg	N	RMSE, MAE, R2
[35]	Predicting the energy demand for EVs	Dundee city, UK	Charging stations data(CS)	RF, DT, KNR, SVR, SGDR	Federated Energy Demand Learning (FEDL), Clustering based EDL,	RMSE
[36]	Predicting the energy consumption in electric vehicles (EVs)	Brussels	Controller Area Network (CAN) bus signals of a 2012 Nissan Leaf	Multiple linear regressio n	Macro and hybrid model	correlation coefficient (R2), Accuracy
[37]	China's road transport sector's energy consumption and greenhouse gas emissions at the provincial level up to 2050	Mainland China	National Bureau of Statistics of China, Annual statistic of population	N	CPREG model, GHG Emissions Analysis model	BAU scenario, LC scenario, GDP growth rate

Table 12. *Cont.*

Ref ID	Objective	Region	Dataset	ML Models	Other models	Eval Metrics
[38]	Analyze future trends in China's road transport sector, evaluating direct and life cycle energy demand and greenhouse gas emissions	China	China Automotive Technology and Research Center (CATARC) data	N	Gompertz curve model, bottom-up model	BAU scenario, HEV application rate
[39]	Reliability and sustainability of smart city transportation systems	United States	Electric Vehicle Charging Dataset Kaggle	KNN, DT, RF, SVM	Deep Neural Networks (DNN), LSTM	MSE, Gradient loss, Accuracy, MAPE

1.5.4 Optimal Placement of New Charging Stations for Heavy Duty and Medium Duty EV

The positioning of charging stations is crucial for the successful integration of Heavy Duty and Medium Duty Electric Vehicles (EVs) in the context of electric mobility. The research on the best locations for new charging stations considers grid efficiency, user convenience, and accessibility. The study's analysis of current approaches aims to put an insight into important considerations for the construction of strategic infrastructure. Similar to previous modules, a brief comparison of studies on the optimal placement of new charging stations given in Table 13.

Table 13. Comparison of existing research on Optimal Placement of New Charging Stations

Ref ID	Objective	Region	Dataset	ML Models	Deep Learning	Other models	Eval Metrics
[40]	Optimizing the selection of EV charging station	SingaporeUSA, UK	Road network data, charging station data	N	N	Proximal Policy Optimization algorithm, RL	profit increase, travel time, charging time

Table 13. *Cont.*

Ref ID	Objective	Region	Dataset	ML Models	Deep Learning	Other models	Eval Metrics
[41]	Resolving EV charging infrastructure planning	Sydney, Australia	Traffic flow Distribution network data	N	CNN (GCN)	Cournot competition game model	Travel time, wait time, charging time
[42]	Determining the placement of charging stations	Germany	Charging Station data	LinReg	N	N	MAE, MAPE, RMSE
[43]	Optimization model using GIS data and grid partitioning to locate and size public charging stations for EVs.	Thailand	GIS data, Vehicle Population	N	N	Optimization and spatial analysis models	EV–CP ratio, Service time
[44]	Allocating charging stations in large-scale transportation networks for electric vehicles (EVs)	Southern Sweden	The Swedish National Road Database (NVDB)	N	N	Probabilistic rule, Integer programming for solving the optimization problem	Maximal route cost
[45]	Address the multi-stage placement of electric vehicle (EV) charging stations	San Pedro District of Los Angeles.	zip code tabulation area (ZCTA), Road network data	N	N	Nested logit model, Bayesian game analysis	Charging Demand, Total Profit, Overall Utility.

The literature review on the analysis of electric vehicle (EV) charging infrastructure offers a thorough look at previous studies and research that have been done on different facets of EV charging networks is given in Table 14. The research was conducted on forecasting demand for EVs, including heavy- and medium-duty models, as well as the vehicle's range and short- and long-term energy requirements. Also, explored the optimal placements for new charging stations, considering factors such as geographic locations, traffic patterns, and user behavior. An increasing amount of information about the intricacies of EV charging infrastructure, including user behaviors, technological developments, and the effect on electricity grids, is available in the literature survey. Geospatial analyses have been used to investigate the best locations for new charging stations, highlighting the significance of key areas to fulfill the increasing demand for electric mobility. By synthesizing these diverse strands of research, the literature review serves as a valuable resource for shaping the future of electric vehicle charging infrastructure and fostering sustainable transportation solutions.

Table 14. Literature Review in Analyzing Electric Vehicle Charging Infrastructure

Ref ID	Region	Purpose	Models	Evaluation Metrics
[46]	Adaptive Charging Network	Finding clusters of EV charging behavior and classifying future sessions to predict which cluster they belong.	K- means. KNN	Accuracy: 97.9 AUC: 0.99
[47]	Santa Monica, California	Predicting charging session duration and energy consumption	Hybrid estimator using GKDE and DKDE	MED: 0.74 hr MED: 1.68 Kw/hr
[48]	Shenzhen China	Predicting hourly charging load of public stations	RNN Based Models	GRU NRMSE: 2.89 %

Table 14. *Cont.*

Ref ID	Region	Purpose	Models	Evaluation Metrics
[49]	England, U.K	Estimating Traffic flow and based on that predicting EV arrival rates	CNN	MAPE: 3.21 %
[50]	Austin, Texas	Classifying Charging profiles of EV users	DGM	Accuracy: 0.98 F1- Score:0.8
[51]	UCLA Campus, Santa Monica, California	Predicting charging behavior based on the labels obtained from clustering	ANN	Accuracy
[52]	Beijing, China	Classifying whether the user will use fast charging or not	Binary Log regression	Accuracy: 0.894
[53]	Netherland	Predict the time to the next plug for residential charging.	SVR with radial basis K-means	MAE: 0.124 RMSE:0.158
[54]	Dutch metropolitan area	Creating EV Profiles that capture charging behavior	GMM	ARI:0.6
[55]	UCLA Campus, California	Predicting session duration and energy consumption in both residential and non-residential areas	Ensemble Model with SVR, RF, DKDE	SMAPE: 10.4 Reduced peak load by 27%, charging cost by 4%

2. Data and Project Management Plan

2.1 Data Management Plan

The data management plan is designed to articulate the strategies for data collection, methods of management, storage procedures, and utilization mechanisms for the project. The data will be collected from a variety of sources. Initially, vehicle registration data will be obtained, providing annual counts of registered vehicles for each zip code and county in California. Subsequently, real-time information on EV charging stations across each zip code in California will be collected. Additionally, data on transit bus stop locations in San Jose, school bus terminal locations in San Jose, and real-world operational data for heavy/medium-duty commercial fleet vehicles across different weight classes will be gathered. This comprehensive dataset aids in understanding the various factors influencing charging station usage.

To improve the quality of data and maintain consistency in information, the following data management process can be applied: The first step is data cleaning, which involves identifying and removing inaccuracies, inconsistencies, and errors in the data. It is an essential step to improve data quality and ensure that the data is reliable and accurate. The second step is data transformation, it helps convert the data from one format to another to make sure the data is in a standardized format. For example, converting all other data types to a structured format CSV. So that can be easily analyzed and interpreted. The last step is to combine data from multiple sources into a single, comprehensive view. In addition to direct data collection, the project explores data partnerships and collaborations with relevant stakeholders. Collaborative efforts with school districts, commercial fleet operators, and local authorities contribute to a more holistic understanding of the EV ecosystem. These collaborations may involve data requests, sharing agreements, or even joint initiatives for data collection. By fostering partnerships, the project aims to enrich its datasets and ensure a comprehensive representation of the diverse facets of the medium/heavy-duty EV landscape. The combination of these data collection approaches ensures that the project starts with a robust foundation of accurate, diverse, and timely datasets, setting the stage for insightful analyses and informed decision-making in the realm of EV charging infrastructure. The entire Data management plan is summarized in below Table 15.

Table 15. Summary of Data Management Plan

Aspect	Description
Data Collection Sources	• Vehicle registration data for annual counts by zip code and county in California. • Real-time EV charging station data for each zip code in California. - Transit bus stop locations data in San Jose. • School bus terminal locations data in San Jose. • Real-world operational data for heavy/medium-duty commercial fleet vehicles across different weight classes.
Data Management Process	• Data cleaning: Identify and remove inaccuracies, inconsistencies, and errors. • Data transformation: Convert data to standardized formats (CSV). • Data integration: Combine data from multiple sources into a single, comprehensive view.
Data Storage Methods	• Cloud-based storage on Amazon Web Services (AWS S3) for scalability and accessibility. • Structured database for efficient organization. • Version control (GitHub) for traceability.
Data Utilization Mechanisms	• Analytical models for forecasting vehicle count, range predictions, and energy demand. • Optimization models for strategic placement of new charging stations. • Visualization tools(Tableau) for effective communication of insights.

2.1.1 Data Collection Approaches

For primary data sources, collaboration with essential stakeholders is paramount. Engaging with charging station operators, city authorities, and heavy/medium-duty vehicle owners will provide real-time insights into charging station usage, energy demand patterns, and vehicle counts. Secondary data sources will complement primary data, drawing from publicly available datasets, government reports, and industry publications. Additionally, sensor networks will be deployed at existing charging stations to capture live data on usage behaviors and charging dynamics.In the pursuit of a comprehensive Data-Driven Analysis of EV Charging Infrastructure for a Smart City for Medium/Heavy Duty Vehicles, our data collection strategy employs a multi-faceted approach. Primary data collection involves establishing robust partnerships with key stakeholders in the EV ecosystem. Through engagements with charging station operators, city authorities, and owners of medium/heavy-duty electric vehicles, we aim to extract real-time insights into the dynamic landscape of charging station usage, energy demand fluctuations, and the daily vehicular counts. This primary data, rich in its immediacy and specificity, forms the backbone of our analysis.

In addition to primary sources, secondary data collection is integral to our strategy. This involves tapping into publicly available datasets, government reports, and industry publications to augment and validate our primary data. By leveraging these authoritative secondary sources, we ensure a holistic and well-rounded understanding of the broader context in which our project operates. Furthermore, the deployment of sensor networks at existing charging stations adds granularity to our dataset, capturing live data and providing detailed insights into usage behaviors and charging dynamics. This combination of primary, secondary, and sensor-derived data enhances the depth and accuracy of our dataset, forming the basis for informed decision-making.

2.1.2 Data Technology Selection

The information is securely housed in an Amazon Web Services (AWS) cloud-based environment, ensuring both scalability and adaptability in data storage. Accessibility to the data is facilitated from any location with an internet connection. In terms of utilizing the data on

Amazon Web Services (AWS), there are various services available. Dataflow serves the purpose of collecting data from diverse sources, while Cloud Data Prep aids in data preparation. Additionally, AWS SageMaker is employed for real-time data processing, enabling the continuous monitoring of usage patterns and immediate issue detection. These functionalities make use of machine learning and deep learning models to extract insights into usage patterns and enhance capacity planning.

To examine the EV charging infrastructure, Python will serve as the primary programming language, leveraging several data science libraries like Pandas, Numpy, Scikit-learn, and TensorFlow. TensorFlow, specifically, will be applied for the implementation of deep learning models, including temporal fusion transformers. Additionally, time series modeling techniques will be employed to scrutinize the seasonality and trends in electric vehicles, aiding in the prediction of demand.

2.1.3 Data Engineering Process

The data engineering process for the project encompasses multiple stages. Initially, we identified pertinent data sources crucial for our analysis, including EV charging station data, EV transaction data, transit bus data, school bus terminal data, and fleet transaction data. Subsequently, we procured and extracted the necessary data from diverse sources such as the CA open data portal, the CA Department of Education, the National Renewable Energy Laboratory, and others. The next steps involved data cleaning and preprocessing to ensure the data's quality and suitability for analysis. After collecting the data, we performed exploratory data analysis to gain insights and identify any issues or anomalies in the data. We also performed data transformation, data aggregation, and feature engineering to create new variables that would be useful for our analysis. Once the data was cleaned and transformed, we stored it in a suitable data storage system over the cloud. Finally, we developed data pipelines and automated processes to update the data and perform regular analysis, ensuring that the insights and recommendations from the project are based on up-to-date and reliable data. Overall, this data engineering process ensures that the project is based on a solid foundation of relevant, accurate, and well-organized data. The entire data engineering process which is the ELT pipeline is shown in Figure 3.

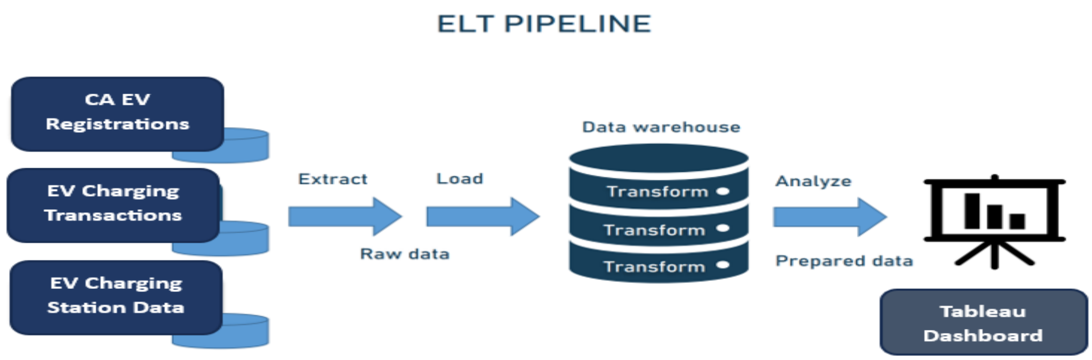


Figure 3. Data Engineering Process - ELT Pipeline

2.2 Project Development Methodology

A CRISP-DM methodology is used for this project, expanding into a CRISP approach Industry Standard Data Mining Process. This approach provides explanations of the duties involved in each project stage, descriptions of typical stages, and the connections between these activities. The CRISP-DM process model provides an overview of the data mining life cycle. The life cycle model contains six phases, and arrows indicate their most important and frequent dependencies. The phases can occur in any order. Projects frequently move back and forth between stages as needed. The model is highly flexible and easily adaptable. To put it another way, CRISP-DM aids in developing a data mining model that satisfies particular needs.

The CRISP-DM methodology includes the following phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The following subsections go into great detail about each of these phases as they relate to this project. The different phases of CRISP-DM Methodology are shown in Figure 4 below.

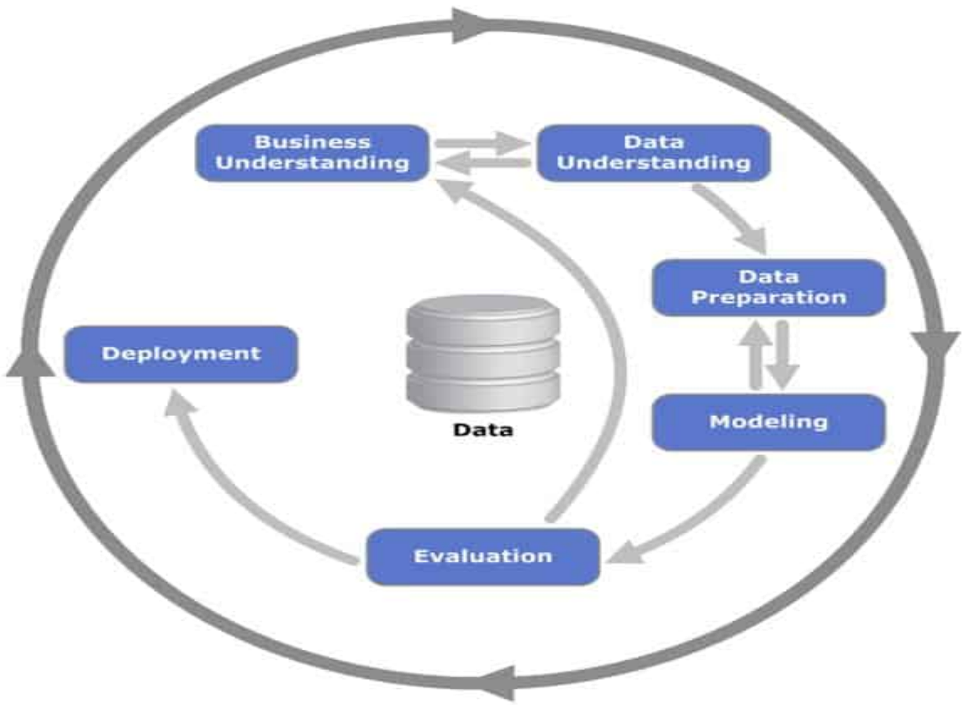


Figure 4. CRISP-DM Methodology Life Cycle

2.2.1 Business Understanding

The initial and most significant phase in strategic planning is to have a thorough grasp of the business environment. Here, the emphasis is on defining precise requirements and work objectives for the data-driven examination of the infrastructure supporting electric vehicle (EV) charging, especially for medium- and heavy-duty automobiles. The development of a reliable and widely accessible charging infrastructure is imperative given the growing demand for electric vehicles (EVs), particularly those in the medium- and heavy-duty segment. The problem is where to put charging stations that can effectively provide the electricity needed to recharge

EVs. Public spaces, shopping centers, and tourist spots are prime places to put them. More EV charging stations that are specifically designed to meet the charging requirements of medium- and heavy-duty vehicles must be planned, designed, and built to promote the wider adoption of electric vehicles.

2.2.2 Data Understanding

Understanding the nature and characteristics of the data pertaining to the infrastructure for charging electric vehicles (EVs) in medium- and heavy-duty vehicles is the primary objective at the data understanding stage. Acquiring knowledge about the distribution, structure, and essential characteristics of the dataset is necessary for this. It is crucial to comprehend the underlying trends, patterns, and possible difficulties in the data. Significant factors are analyzed, including utilization trends, charging capacities, geographic distribution of charging stations, and any external influences on the data. To gain useful insights for optimizing EV charging infrastructure, appropriate issues, hypotheses, and modeling methodologies must be formulated. This fundamental understanding sets the basis for succeeding steps in data-driven research.

2.2.3 Data Preparation

The main objective of the Data Preparation stage is to prepare the data for further modeling; the Google Cloud Platform is used to store the raw data. In this step, careful data wrangling for EV charging station data is done, and any missing, erroneous, or duplicate values are checked. Moreover, any unnecessary or redundant data that is not relevant to the research is removed. The process of data transformation involves altering the data types, standardizing the formats, and sometimes adding computed columns once the data wrangling is finished. Subsequently, the data is partitioned into training and testing sets. The preprocessed data is saved on Google Cloud, serving as the input for the subsequent models in the analysis of EV charging infrastructure for Medium/Heavy-duty vehicles.

2.2.4 Modeling

The modeling method in this analysis employs a diverse set of models to uncover patterns and insights within the data. More specifically, the temporal fusion transformer model, PuLP Linear Programming Optimization with K-means Clustering, and Stacking Ensemble Regressor (Weighted Fusion) ensemble methods, used alongside the time-series forecasting model Prophet, have been chosen meticulously, are used to predict, and forecast the future count of Heavy Duty and Medium Duty Vehicles and the required EV charging stations to cater the demand. Applying the Temporal Fusion Transformer Model's capabilities, enables adaptable forecasts of energy consumption on a daily, weekly, and monthly basis, considering the fluctuating demands for EV charging. The Stacking Ensemble Regressor integrates multiple algorithms and gives each one a weight to improve prediction accuracy and forecast the range of heavy and medium-duty vehicles of each weight class using characteristics like charging duration, and state of charge to the battery. Furthermore, the strategic placement of charging stations within the road network is optimized by the integration of PuLP Linear Programming Optimization and K-means Clustering. Each model utilizes historical charging data, geographical information, and other pertinent factors to provide a comprehensive and accurate assessment of the charging infrastructure requirements.

2.2.5 Evaluation

Through this phase, we will assess our built models by passing them through a test data set. First, model performance checks with unknown data. Then, to analyze the model performance of the algorithms, we have chosen these cluster performance indicators, such as

Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2 Score. We also evaluate how well the optimization strategy works to deploy new charging stations in optimal locations for all kinds of vehicles. The Objective Function Value (Cost), Coverage Metric, Utilization Metric, and Equity Metric are the four metrics applied to our optimization problems. We analyze these measures in a variety of data models to determine which model has the best-predicted accuracy and best fit. The most reliable data-driven model is chosen for deployment based on this careful evaluation, which ensures that it is prepared for practical significance in the context of EV charging infrastructure for medium- and heavy-duty vehicles.

2.2.6 Deployment

Finally, in the Deployment phase, we continuously monitor and maintain models in order in order to ensure predictable and consistent performance after deployment. This includes meticulously tracking the models' behavior, quickly identifying any possible problems, and, if necessary, retraining the models to improve accuracy. A website is also created to make it easier to install EV charging stations in areas where they are most required. This forward-thinking strategy makes sure that the website contributes to a sustainable and effective transportation ecosystem by anticipating and adapting to the changing environment of electric mobility in addition to meeting current infrastructure needs. The project's dedication to accuracy, flexibility, and creativity in influencing the direction of EV infrastructure planning is demonstrated by the website. The culmination of this phase involves the compilation of all requisite data for the project report and presentation, summarizing the comprehensive data-driven analysis of EV charging infrastructure for Medium/Heavy-duty Vehicles.

2.2.7 Intelligent System Engineering Process

To summarize and create a well-defined intelligent system for our project, the first step in the process was to define the problem statement and gather the necessary data from various sources. The data was then cleaned and preprocessed using various techniques such as text normalization and feature extraction. Next, we employed various machine learning algorithms such as Prophet model, Stacking Regressor model, Temporal Fusion Transformer and Linear Programming and Clustering model to perform different tasks such as demand forecasting, range prediction, energy demand forecasting and finding optimal locations of new charging stations. We also utilized various time series modeling techniques to analyze the seasonality and trends of electric vehicle demand. The entire flow of the project is shown as below Figure 5.

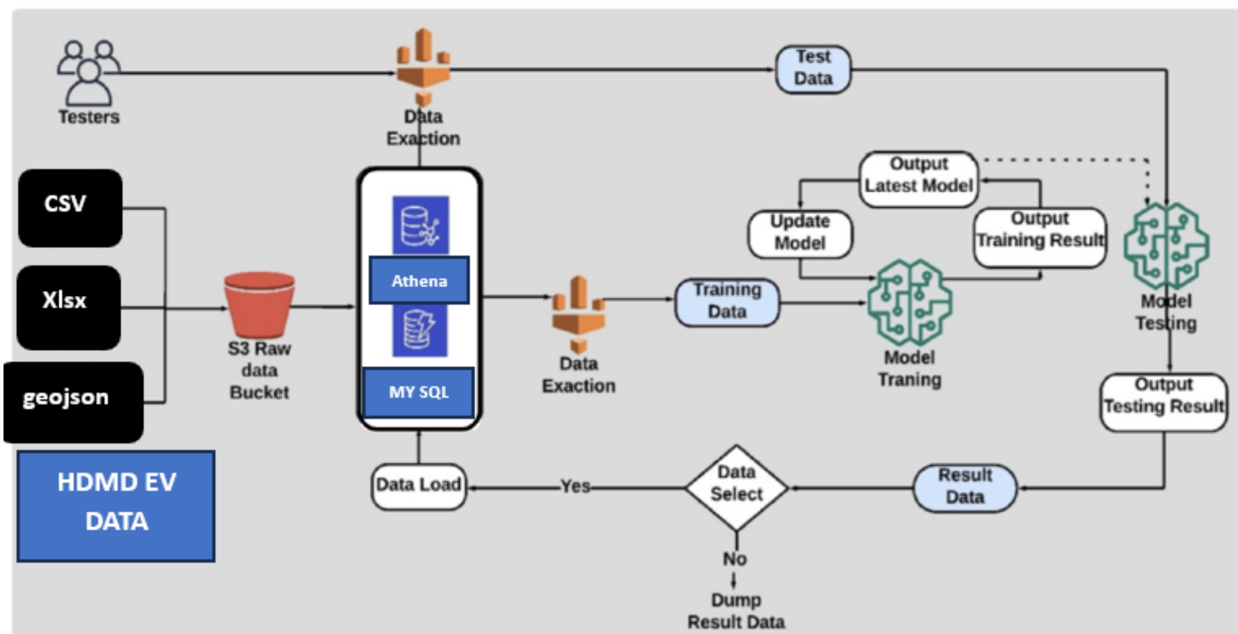


Figure 5. Project Data Flow Methodology

The performance of these algorithms was evaluated using various metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2 Score. Based on the performance, we selected the best-performing algorithms for each task.

Finally, we deployed the selected algorithms in an integrated system that could take input data from different sources and provide output predictions in real-time. The system was tested for various use cases and the results were validated with real-world data. Throughout the process, we followed best practices such as using open-source tools, documenting the code and processes, and collaborating with domain experts to ensure the accuracy and usefulness of the insights generated by the system.

2.3 Project Organization Plan

We are using the CRISP-DM technique to define the work breakdown structure for our project. A WBS tool is used to manage the objectives of a project by breaking it down into multiple tasks, and team members can be assigned different tasks to finish the tasks on time. As depicted in Figure 6, our project has been divided into several phases.

The Business Understanding phase is the first and most important step in the process. During this phase, we have gathered background information for our data-driven research on the deployment of EV infrastructure in urban areas. Our research into ML/DL efforts for this issue led us to establish the project's objectives. Here, our project aims to provide a comprehensive overview of how models such as Prophet, Stacking Ensemble Regressor (Weighted Fusion), the temporal fusion transformer model, PuLP Linear Programming Optimization with K-means Clustering can be used to analyze and forecast charging behavior and demand to enable the necessary infrastructure even in remote areas and communities that are too frequently ignored.

In the second phase, known as Data Understanding, we looked into the pertinent EV charging station data sources that were needed for the project. In this stage, we have selected California’s EV charging station usage data in accordance with the requirements for determining whether it is adequate for training and testing. We have also performed a quality analysis on the dataset and developed a data management strategy to store our image data in the Google Cloud Platform (GCP).

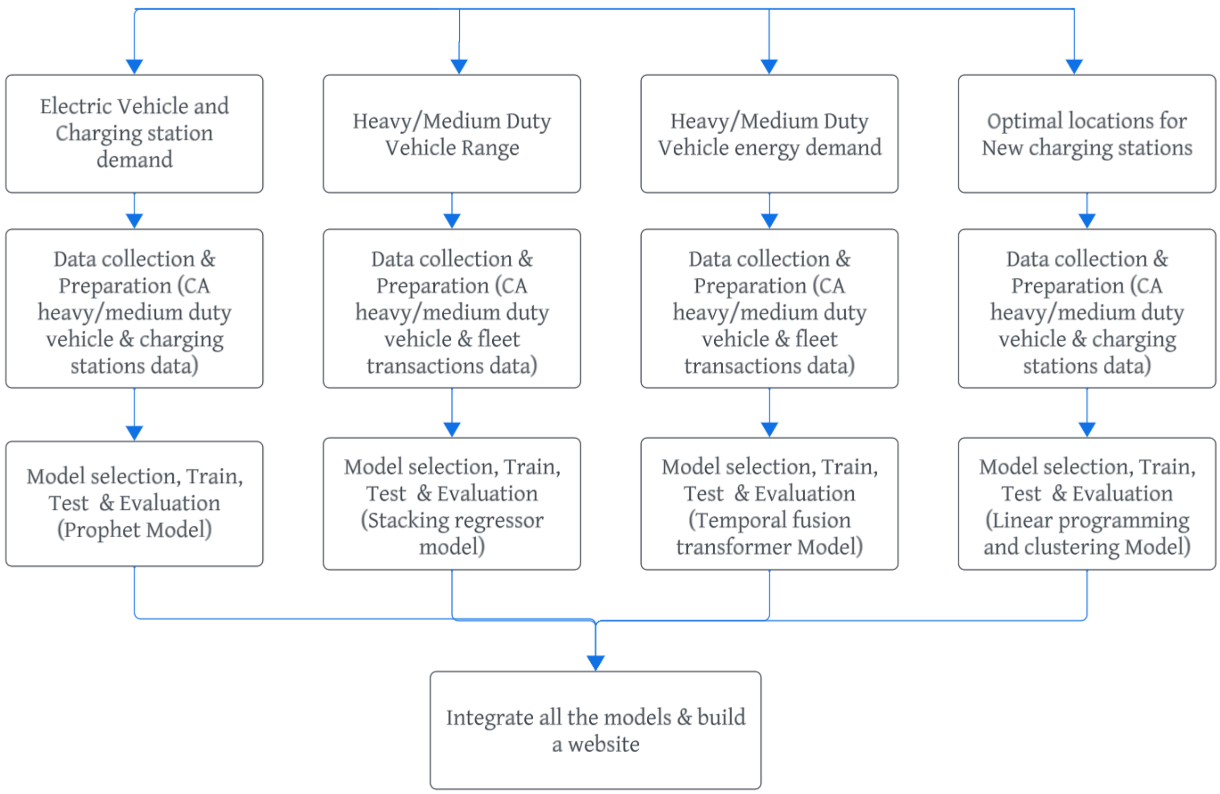


Figure 6. Flow chart of project phases

The third stage is Data Preparation, which involves processing and cleaning the data. Performing feature engineering enhances the performance of the ensemble learning model by selecting the appropriate features for the model and preparing the attributes in a way that is suitable for the model and helpful for the following phases. The data is divided for the models' training and testing and is then used in the following stages.

The Modeling stage comprises selecting an ensemble model and training that model with our preprocessed EV charging stations data set. specifically utilized and enhanced by the time-series forecasting model Prophet, the Temporal Fusion Transformer model, PuLP Linear Programming Optimization with K-means Clustering, and Stacking Ensemble Regressor

(Weighted Fusion) ensemble approaches are utilized. These models are used to forecast and predict the number of Heavy Duty and Medium Duty Vehicles in the future, as well as the number of EV charging stations needed to meet demand. By utilizing the Temporal Fusion Transformer Model's features, it is possible to generate flexible energy consumption estimates on a daily, weekly, and monthly basis, which can be adjusted to meet the changing needs for EV charging. And using PuLP Linear Programming Optimization with K-means Clustering, can find the optimal locations for the charging stations.

In the last stage of deployment, we will deploy the ensemble model and create a website that can be useful for installing EV charging stations wherever they are required and doing the necessary maintenance and monitoring tasks. The last stage will be to compile a project report and presentation with all the required information.

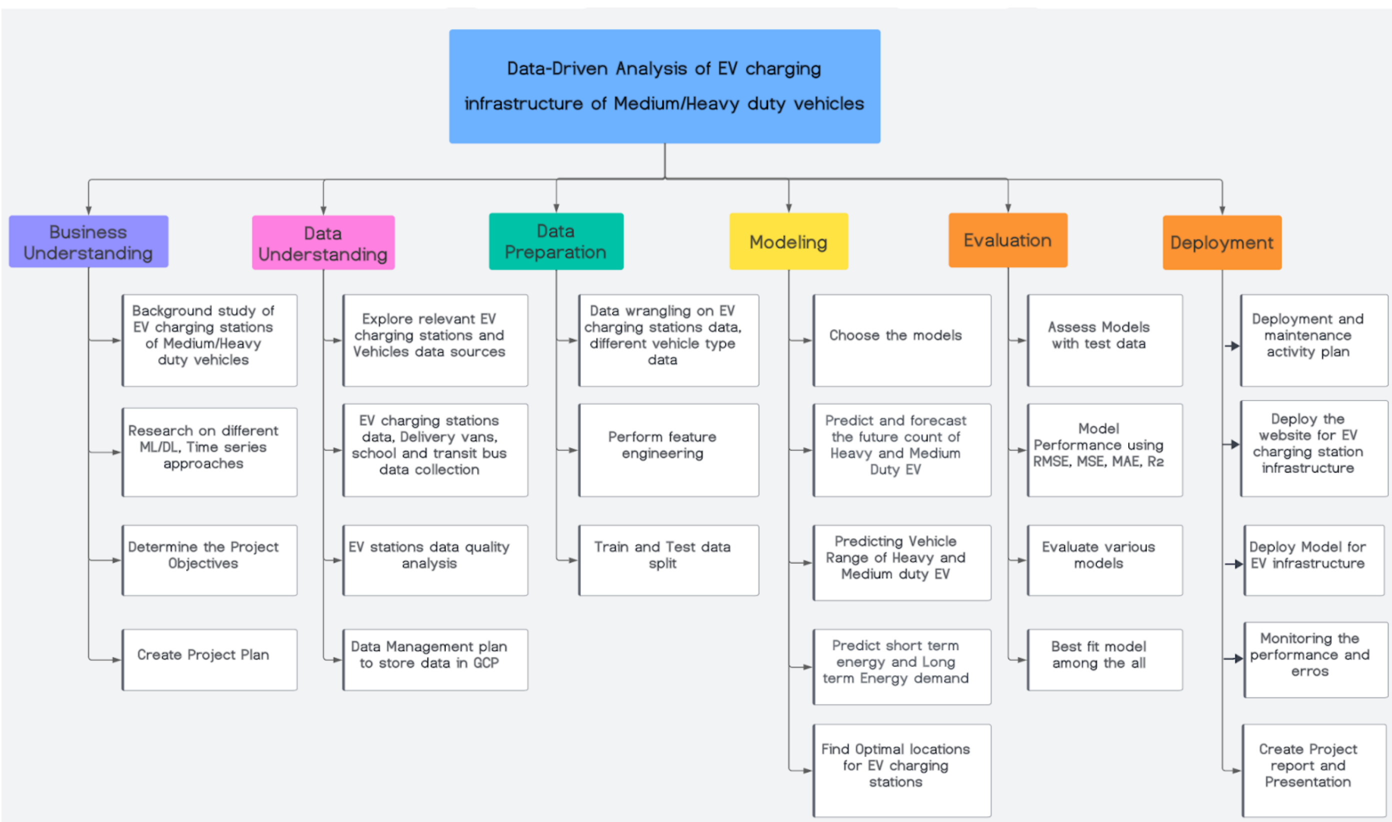


Figure 7. WBS Chart

2.4 Project Resource Requirements and Plan

In this project, it is essential to find a suitable dataset of electric charging stations and to select appropriate software and hardware resources. The project's budget must be planned with an understanding of the costs and resources involved as given in Table 16. This project uses Amazon Web Services (AWS). It is a free public dataset that we collected from the source of data that we used in our study. AWS environments support platform-as-a-service, infrastructure-as-a-service, and other cloud computing execution models. Affordability, adaptability, and traceability are all advantages of these technologies. A consistent approach is required to manage the different services offered by AWS, which include a variety of hardware and software resources. The tool known as the "AWS CLI" is the Amazon Web Services command-line interface. This unified tool offers a command-line interface for users to interact with a range of AWS services. With the AWS CLI, users can effectively manage their AWS resources, configure services, and automate tasks through scripting.

The first thing we need in this project is a virtual machine. We require two or more GPU worker nodes, 64 GB of RAM, and 16 virtual CPUs to run machine learning models uninterrupted. A bucket and folder will organize all project-related data on Aamazon S3. The AWS is a centralized repository for storing files so that you can access them quickly. Because we store data at each stage, such as raw, processed, trained, tested, and archived, we need 30 GB of storage.

Various data cleaning and wrangling procedures are used to prepare the gathered data for analysis. The free Google Colab software will be used for data preparation and modeling. For data wrangling and data transformation, we will use Python programming language. We can code and develop the model quickly using all Python 3.9 packages and modules. In order to analyze and visualize data effectively, Pandas, NumPy, and Matplotlib libraries will be used. PIL (Python Imaging Library) and Torchvision libraries are given in Table 17. The platform facilitates the development, testing, and deployment of machine-learning models.

Table 16. Hardware Requirements

Hardware	Configuration	Purpose
GCS- Standard Dual Region (us-central1 (Iowa) and us-west1 (Oregon))	20 GB	To store the collected data in each stage like raw, processed, train and test data.
GCS-Standard (us-east4 (Northern Virginia))	10 GB	To archive the data in a separate region for disaster recovery.
GC Virtual machine	OS: Red Hat Enterprise Linux n1-standard-16 (vCPUs: 16, RAM: 60GB), NVIDIA TESLA T4 GPU:2	To run machine learning models uninterrupted for the high-resolution image data.

Table 17. Software Requirements

Software	Version	Purpose
Python Programming Language	3.9.0	Performing exploratory data analysis
Google Colab	Python Libraries: Pandas version 1.5.0 Matplotlib version 3.60 Touchvision 0.13.1 PIL 8.32	Performing data wrangling and data transformations
AWS Sage Maker	NumPy version 1.21.0 Scikit-learn version 1.1.2 SciPy version 1.9.2	Developing, testing, and deploying unsupervised clustering machine learning models.

In accordance with CRISP-DM methodology, we prepared the Workbench structure, Pert, and Gantt charts using the Click Up tool given in Table 18. To collaborate online, we used Zoom, a cloud-based, free video conference application. To redistribute our preprocessed data and metadata documentation, we use GitHub, an open-source platform. As part of our project, we also use the MS Office 365 suite under a student license to prepare reports and presentations.

Table 18. Tools and Licenses

Tool	License	Purpose
ClickUp	Unlimited (Paid)	For our project management, we need to create a WBS, a Gantt chart, and a Pert chart using CRISP-DM.
Zoom	Free	Project planning and discussion can be done online.
Github	Free	Data acquisition, usage, and distribution documentation and preprocessed data will be redistributed.
MS Office 365 suite	Student	As part of our project, we will create reports and presentations

Based on the configurations and the duration of the various tools and resources we use. Table 19 provides an overview of the budget projections for these tools and resources. In total, it is estimated that the calculation services will cost approximately $2875 in total. A Google Cloud Pricing Calculator is a tool that estimates costs for all of the available GCP services.

Table 19. Project Cost Estimation and Justification

Purpose	Resource Type	Resource	Time Duration	Cost in USD
Cloud Service Management Tool	Software	GCloud CommandLine	02/17/2023 – 11/30/2023 ~ 12 months	Free

Table 19. *Cont.*

Purpose	Resource Type	Resource	Time Duration	Cost in USD
Data Storage	Hardware	AWS S3	02/17/2023 – 12/17/2023 ~ 10.5 months	$15.20 ($ 3.80 per 1 month for 50 GiB)
Virtual Machine	Hardware	GCVM n1-standard- 16 (vCPUs: 16, RAM: 60GB), NVIDIA TESLA T4 GPU:2	02/20/2023 – 12/15/2023 ~ 10 months	$2400($2.36 per GPU/hour, we will shut down the virtual machine as much as possible. There may be a need for 2 GPUs)
Data Preprocessing	Software	Google Colab Python 3.7 Version	02/19/2023 – 12/15/2023 ~ 10 months	Free
Model Development	Software	AWS Sage Maker	02/22/2023 – 12/15/2023 ~ 10 months	$70.75
Project Management	Tool	ClickUp version 3.0	03/18/2023 – 12/03/2023 ~ 9 months	$36 ($7.2 per each member)
Data Redistribution	Tool	Github version 3.6.3	03/29/2023 – 11/29/2023 ~ 9 months	Free
Project Work Collaboration	Tool	Zoom version 5.1.2	01/29/2023 – 12/15/2023 ~ 2.5 months	Free
Project Documentation	Tool	Ms Office 365 Suite version 2209	02/15/2023 – 12/15/2023 ~ 10 months	Free (Student License)

2.5 Project Schedule

The most well-known tool for organizing project goals or tasks is the Gantt chart which helps to organize the project by establishing completion dates for the activities and allocating resources who will be in charge of those tasks. We made a chart with the task name, start date, deadline date, estimated completion time, and team member responsible for the assignment. The start and end dates, as well as their durations, are important since they show how long the activity will last. We have not included weekends, Thanksgiving break, or any other holidays that fell on our timetable when allocating these deadlines. We have coordinated the deadlines with the given deadlines for the deliverables. Also, there are still many tasks that need to be completed. Figure 8 is shown in the Gantt chart.

One of the most often used project management tools is a PERT chart shown in Figure 9, which displays the tasks, duration, and assignees inside rectangular shapes. PERT chart is used to organize and determine the approximate duration of a project's tasks. Tasks on the critical path in project management are essential to finishing a project. For the PERT chart, we used a task-oriented methodology. Here, the dependencies between the tasks are defined, and assigned to the team members, and the time required to complete each task has been computed. The red arrows indicate the critical path; each task has a task number.

3 Data Engineering

3.1 Data Process

The raw form of data lacks utility, and data processing is the essential procedure of taking this raw data, refining it, and converting it into usable information. Commencing with the collection of raw data, the process involves enhancements, organization, processing, evaluation, archival, and ultimately presenting the data in a more structured manner.

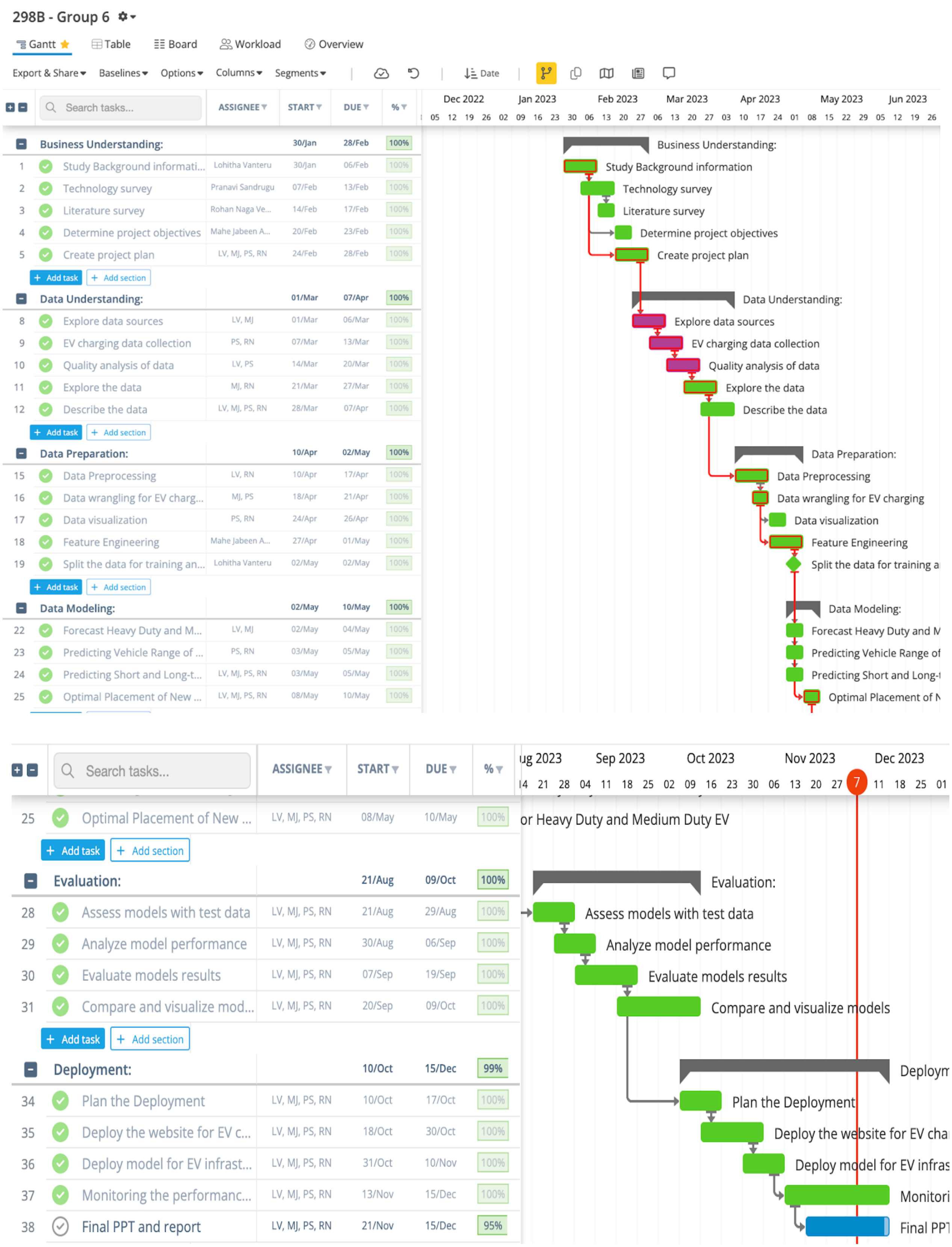


Figure 8. Gantt Chart

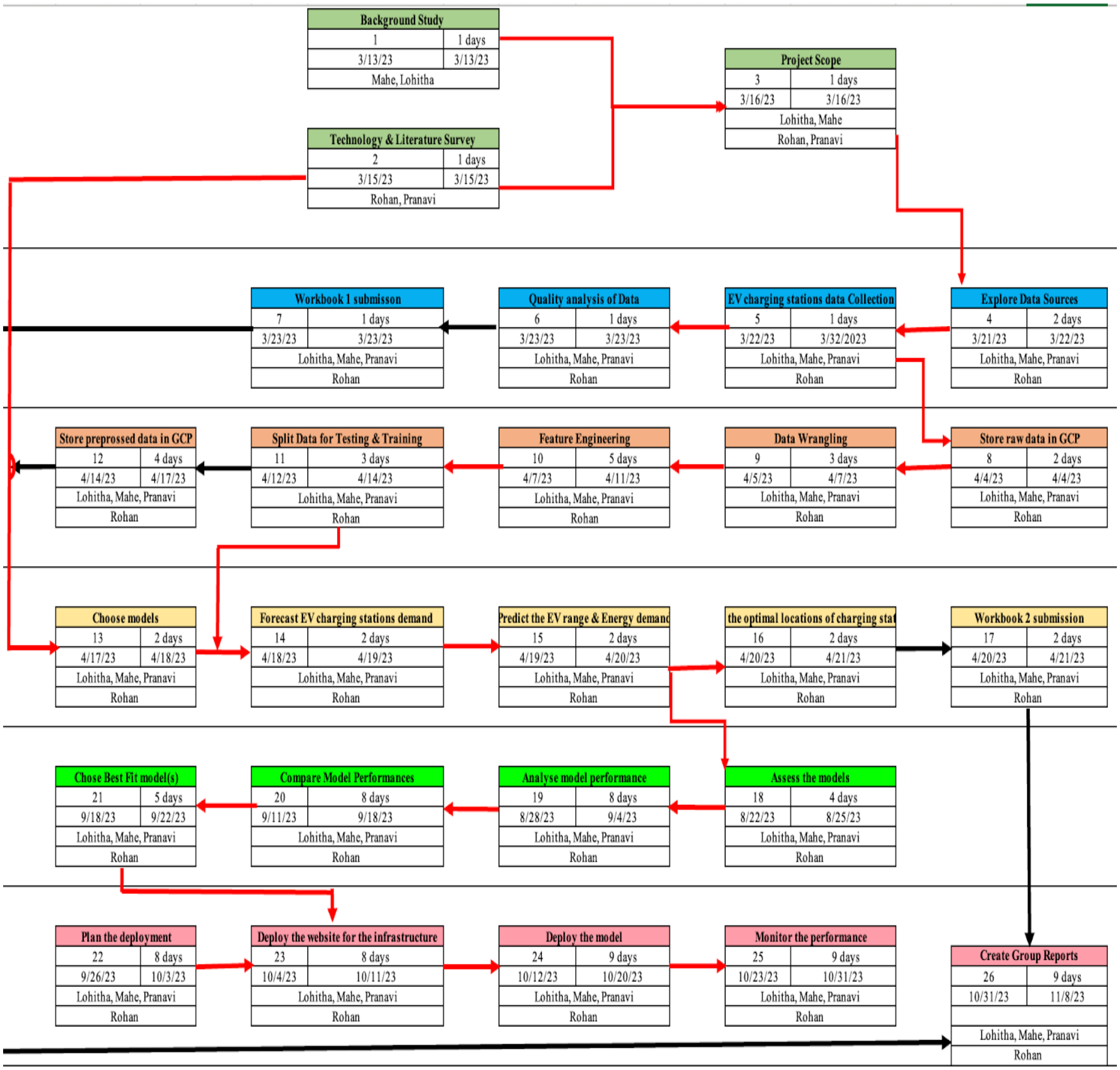


Figure 9. Pert Chart

This transformation provides the necessary shape and context for machines to analyze and derive meaningful insights, uncovering patterns or trends that might be challenging to discern otherwise. It holds significance as a crucial stage that facilitates informed decision-making and underscores the informational value available in the process. In broad terms, data processing for this project encompasses data gathering, data exploration, data cleaning, data transformation, and data preparation. The entire data process is summarized as a flow chart in the following Figure 10.

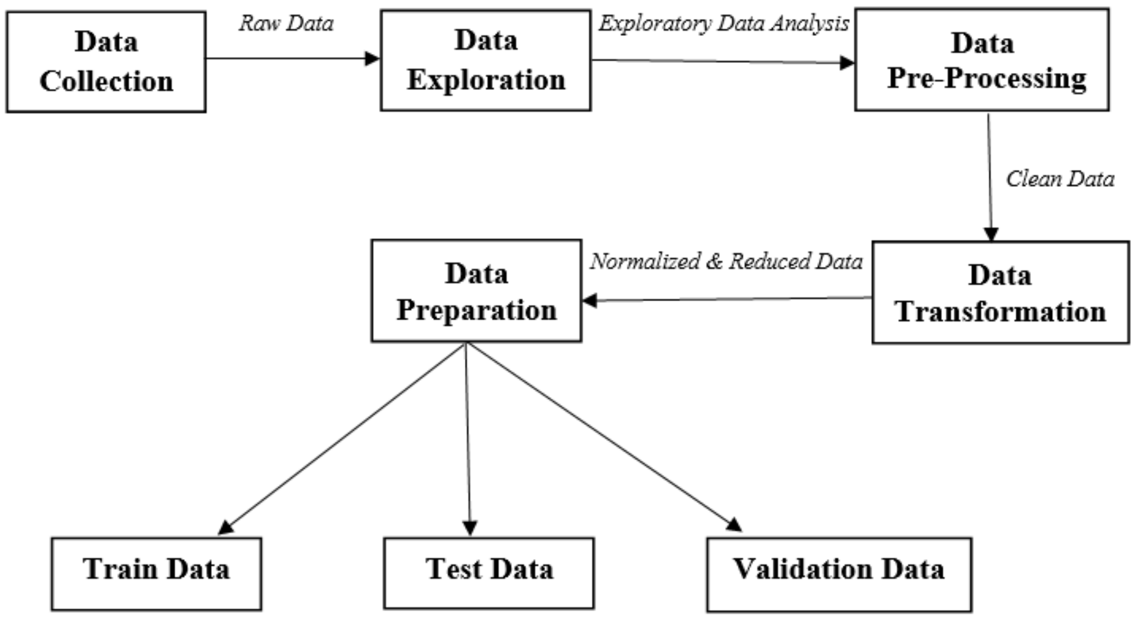


Figure 10. Data Process Flow Chart

The initial step in the data processing cycle is the collection of raw data, and the accuracy and relevance of the final results are heavily dependent on the quality of this unprocessed information. It is crucial to source preliminary data from trustworthy and dependable sources to ensure the validity and significance of the outcomes. In our project, the assessment of potential locations for new charging stations relies solely on real-world data within the Electric Vehicle Charging Infrastructure framework. Although acquiring data in this domain can pose challenges, real-world data provides insights into the actual behaviors of EV owners and potential EV users, making it more realistic than simulation-based results. The process of addressing machine learning problems commences with data acquisition, followed by exploration to assess its quality and accuracy, and comprehension through the interpretation of various visualizations. Post-cleaning, the data may transform to ensure suitability for analysis, involving tasks such as converting categorical variables to numerical ones, scaling or normalizing data, and employing feature engineering techniques to create new variables. After transformation, the data is segmented into training, validation, and test sets. The training set is employed for model training, the validation set aids in tuning hyperparameters, and the test set is instrumental in evaluating model performance. The adequacy of training data is crucial for model consistency, while an insufficient amount of testing or validation data can lead to increased variability in performance metrics. Adhering to these principles, the modified data is prepared for modeling, with an 80% allocation to the Train set and the remaining 20% divided into 10% for Test and 10% for Validation sets, used for model evaluation. The final dataset is then employed for both descriptive and predictive analyses after completing the data pre-processing stage. The established methodology can propose optimal locations for new charging stations by leveraging real-world data, particularly focusing on minimizing the average drop in charging usage within a charging zone. This essentially translates to maximizing the overall aggregate usage within a charging zone once the new charger is operational. Beyond pinpointing suitable locations for new charging stations, this methodology sheds light on the usage patterns of existing charging stations and the charging behaviors of EV owners. Consequently, providers of EV charging infrastructure can employ this proposed technique as a decision-support tool for expanding their charging network.

The California Open Data Portal is an online platform dedicated to fostering transparency and open government by providing public access to a diverse array of datasets related to the state of California. It covers various topics such as demographics, education, health, and transportation. The datasets encompass information on electric vehicle (EV) charging infrastructure, adoption rates of electric vehicles, emissions reductions, and environmental data.

The portal serves a broad audience, including researchers, policymakers, developers, and the general public, who can freely access, download, and analyze datasets in machine-readable formats, facilitating seamless integration and analysis for gaining insights into different facets of California's public life. The California Department of Motor Vehicles (DMV) is a government agency responsible for vehicle registration, driver's license issuance, and maintaining driving records to ensure safety and proper documentation in the state. It also contributes data on electric vehicles (EVs) through registration and licensing records, providing statistics on the quantity and types of electric vehicles registered. This information is valuable for tracking the expansion of the electric vehicle market in California, evaluating the popularity of various EV models, and understanding demographic trends in electric vehicle ownership across the state. The National Renewable Energy Laboratory (NREL) operates as a research institution under the U.S. Department of Energy with a primary focus on advancing renewable energy and energy efficiency technologies. It serves as a crucial source of research and data on various renewable energy technologies, including electric vehicles (EVs). NREL conducts comprehensive studies on EV technology, charging infrastructure, battery advancements, and the overall integration of electric vehicles into the broader energy framework. The laboratory offers valuable insights into the performance, efficiency, and environmental advantages of electric vehicles. Researchers and professionals in the industry frequently depend on NREL's data to comprehend the current state of EV technology and guide future advancements. The U.S. Alternative Fuel Data Center (AFDC) is a comprehensive platform managed by the U.S. Department of Energy (DOE), offering a wealth of information, datasets, and tools related to alternative fuels and advanced vehicles. Serving as a central repository, it covers various data sets related to alternative fuels, with a specific focus on electric vehicles (EVs). The datasets include valuable information such as charging station data, fueling station data, and vehicle data. Researchers and policymakers frequently leverage the AFDC's data to analyze patterns, strategize infrastructure development, and make informed decisions concerning the promotion and adoption of alternative fuel vehicles, particularly electric vehicles.

The data processing for forecasting the demand for electric vehicle charging stations and predicting the count of Electric Vehicles (EVs) for Medium and Heavy-duty vehicles involves obtaining data from the California Department of Motor Vehicles for vehicle details and the U.S. Alternate Fuel Data Center for charging station information across CA zip codes. This dataset includes details such as year, model, make, vehicle types, and zip codes within California, spanning the years 2010 to 2022, with a count of 35,742. The aim is to predict future EV counts, forecast charging station demand, and visualize seasonal patterns. Additionally, we gathered medium and heavy-duty vehicle fleet transaction data from the National Renewable Energy Laboratory, encompassing charging start and end times, starting, and ending State of Charge (SOC), and total energy delivered. This dataset covers the period from 2018 to 2022, with a count of 428,963. The objective here is to predict the range of both heavy and medium-duty vehicles and visualize daily, weekly, and monthly energy demand forecasts. To enhance our understanding of public transportation, we also collected transit bus stop data from the CA Open Data portal and school bus terminal data from the CA Department of Education. The summary of open-source datasets that are collected for this project is outlined in the following Table 20.

These diverse data sources collectively serve as the foundation for our data-driven analysis, enabling us to extract valuable insights into electric vehicle adoption and the deployment of charging stations. Following the data collection phase, we performed data cleaning, essential aggregation, and transformation. Subsequently, we prepared the datasets, divided them into an 80:10:10 ratio, and proceeded to the modeling stage. We applied the model fitting process to generate accurate predictions, and in addition, we conducted fine-tuning procedures to enhance the efficiency of the models.

Table 20. Open-Source Datasets

Dataset	Description
CA Vehicle Registration Data	Number of vehicles registered each year for every zip code and county in CA.
CA EV charging Stations Data	Real-time data of EV Charging stations for each zip code in CA.
Transit Bus Stops Data	Data of public bus stations locations for transit bus in San Jose.
School Bus Terminal Data	Data of school bus terminal stations locations in San Jose.
Heavy/Medium Fleet Vehicle Transactions	Real-world data of commercial fleet vehicle operating data for each weight class.

3.2 Data Collection

The process of collecting data follows a methodical approach to obtain pertinent information from various sources. The fundamental steps in this data collection process are elucidated below:

1. **Identifying Data Sources:** The initial phase involved identifying the primary data sources crucial for the analysis. These sources included the California Department of Motor Vehicles, supplying details on vehicle specifics; the U.S. Alternate Fuel Data Center, providing data on charging stations; the CA Open Data portal, serving as a source for transit bus stops; and the CA Department of Education, acting as a source for school bus terminal data. Additionally, the National Renewable Energy Laboratory offered data on Heavy/Medium fleet transactions.

2. **Retrieving Data from the California Department of Motor Vehicles:** Information about electric vehicles, encompassing details like the year, model, make, vehicle types, and zip code information within the state of California, was sourced from the California Department of Motor Vehicles. This dataset covered the period from 2010 to 2022. This dataset has 35, 742 records.
3. **Utilizing the U.S. Alternate Fuel Data Center:** Real-time data concerning charging stations in every zip code in California was acquired from the U.S. Alternate Fuel Data Center. This dataset encompassed essential information, including the year, cities, zip codes, latitude, and longitude, in addition to charger types such as EV level 1, level 2, and DC fast count throughout the state. This dataset spans from 2010-2022 and has 14,862 records.
4. **Exploring the CA Open Data Portal:** To gain a deeper understanding of public transportation, information on transit bus stops was sourced from the CA Open Data portal. This dataset played a crucial role in comprehensively grasping the dynamics of public transit. Both datasets cover the period of 2022. Transit bus stops data has 1289 records.
5. **Retrieving Data from the CA Department of Education:** For a more profound insight into public transportation, data on school bus terminals was obtained from the CA Department of Education. This dataset proved essential in gaining a comprehensive understanding of public transit dynamics. The dataset encompasses the timeframe of 2022 and has 3526 records.
6. **Deepening Commercial Fleet Analysis:** Lastly, to enhance our understanding of commercial fleet operations, we procured 'Heavy/Medium Vehicle Fleet Transactions

Data' from the National Renewable Energy Laboratory. This dataset furnishes real-world data about different weight classes of commercial fleet vehicles and was acquired from the National Renewable Energy Laboratory. The dataset spans from 2018-2022 and has 4,28,963 records.

In conclusion, the process of collecting data entailed a careful identification of sources, obtaining access to pertinent datasets, and establishing a thorough groundwork for the ensuing data-driven analysis. The detailed information on datasets is summarized in below Figure 11.

Name	Vehicle Registration-Zip code	Charging Stations- Zip Code	Vehicle Fleet Transactions-Weight Class	Transit Bus Stations-City	School Bus Terminals-City
Source	California Department of Motor Vehicles	U.S. Alternate Fuel Data Center	National Renewable Energy Laboratory	CA Open Data Portal	CA Department of Education
Data Type	.xslx	.csv	.xslx	.geojson	.csv
Size	28.8 MB	16.7 MB	126.8 MB	1.6 MB	3.4 MB
Count	35,742	14,862	4,28,963	1289	3526
Yearly	Y	Y	Y	Y	Y
Weekly	N	N	Y	N	N
Monthly	N	N	Y	N	N
Duration	2010-2022	2010-2022	2018-2022	2022	2022

Figure 11. Summarized information of the datasets

Raw dataset samples offer an initial look at the unprocessed and varied data that underpins the Data-Driven Analysis of EV Charging Infrastructure for Medium/Heavy Duty Vehicles project. These samples cover electric vehicle details, charging station information, transit bus stops, and fleet transactions, encompassing essential variables like a year, make, model, location, and charging specifics. These samples act as the foundational material for subsequent processes such as data processing, analysis, and modeling, providing a starting point to derive meaningful insights into the intricate dynamics of electric vehicle adoption and charging infrastructure deployment. The figures below illustrate raw samples of the data collected.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3424 entries, 0 to 3423
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Year        3424 non-null   int64
 1   Zip Code    3424 non-null   int64
 2   Model Year  3424 non-null   object
 3   Fuel        3424 non-null   object
 4   Make        3424 non-null   object
 5   Duty        3424 non-null   object
 6   Vehicles    3424 non-null   int64
dtypes: int64(3), object(4)
memory usage: 187.4+ KB
```

Figure 12. Showing the raw sample data from vehicle registration data

	Year	Zip Code	Model Year	Fuel	Make	Duty	Vehicles
0	2015	90018	2010	Battery Electric	OTHER/UNK	Heavy	1
1	2015	90031	2018	Battery Electric	OTHER/UNK	Heavy	1
2	2015	90212	2011	Battery Electric	OTHER/UNK	Heavy	1
3	2015	90247	2009	Battery Electric	OTHER/UNK	Heavy	5
4	2015	90247	<2008	Battery Electric	NEWFLYER	Heavy	16

Figure 13. Summary of the vehicle registration data

	Fuel Type Code	Station Name	Street Address	Intersection Directions	City	State	ZIP	Plus4	Station Phone	Status Code	...	EV On-Site Renewable Source	Restricted Access	RD Blends	RD Blends (French)	RD Blended with Biodiesel	RD Maximum Biodiesel Level
0	ELEC	Los Angeles Convention Center	1201 S Figueroa St	West hall and South hall	Los Angeles	CA	90015	NaN	213-741-1151	E	...	NaN	False	NaN	NaN	NaN	NaN
1	ELEC	California Air Resources Board	9530 Telstar Ave	NaN	El Monte	CA	91731	NaN	626-575-6800	E	...	NaN	False	NaN	NaN	NaN	NaN
2	ELEC	Scripps Green Hospital	10666 N Torrey Pines Rd	Patient Parking Structure, level G	La Jolla	CA	92037	NaN	NaN	E	...	NaN	False	NaN	NaN	NaN	NaN
3	ELEC	Galpin Motors	15421 Roscoe Blvd	NaN	Sepulveda	CA	91343	NaN	800-256-6219	E	...	NaN	True	NaN	NaN	NaN	NaN
4	ELEC	Galleria at Tyler	1299 Galleria at Tyler	NaN	Riverside	CA	92503	NaN	951-351-3110	E	...	NaN	False	NaN	NaN	NaN	NaN

5 rows × 73 columns

Figure 14. Showing the raw sample data from EV charging stations

	Vehicle ID	Charger ID	Local Connect Time	Local Disconnect Time	Local Charge Start Time	Local Charge End Time	Average Power	Max Power	Total Energy Delivered	Starting SOC	Ending SOC	Starting SOC	Date	Number of Charging Sessions	Connection Time	Charging Time
0	EV026	CH007-CH008	NaN	NaN	8/10/2018 18:48	8/10/2018 18:52	193.386	342.28	13.910	81.0	99.5	NaN	NaN	NaN	NaN	NaN
1	EV026	CH007-CH008	NaN	NaN	8/10/2018 18:14	8/10/2018 18:20	308.800	363.64	24.091	65.0	100.0	NaN	NaN	NaN	NaN	NaN
2	EV026	CH007-CH008	NaN	NaN	8/10/2018 17:11	8/10/2018 17:19	327.830	364.82	31.054	57.5	100.0	NaN	NaN	NaN	NaN	NaN
3	EV026	CH007-CH008	NaN	NaN	8/10/2018 16:04	8/10/2018 16:09	321.533	361.06	25.935	62.5	99.5	NaN	NaN	NaN	NaN	NaN
4	EV026	CH007-CH008	NaN	NaN	8/10/2018 15:02	8/10/2018 15:10	230.942	360.10	32.010	56.5	100.0	NaN	NaN	NaN	NaN	NaN

Figure 15. Showing the raw sample data of vehicle fleet transactions

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55695 entries, 0 to 55694
Data columns (total 73 columns):
 #   Column                                   Non-Null Count  Dtype
---  ------                                   --------------  -----
 0   Fuel Type Code                           55695 non-null  object
 1   Station Name                             55695 non-null  object
 2   Street Address                           55695 non-null  object
 3   Intersection Directions                  1991 non-null   object
 4   City                                     55695 non-null  object
 5   State                                    55691 non-null  object
 6   ZIP                                      55695 non-null  object
 7   Plus4                                    0 non-null      float64
 8   Station Phone                            55089 non-null  object
 9   Status Code                              55695 non-null  object
 10  Expected Date                            0 non-null      float64
 11  Groups With Access Code                  55695 non-null  object
 12  Access Days Time                         47076 non-null  object
 13  Cards Accepted                           4401 non-null   object
 14  BD Blends                                0 non-null      float64
 15  NG Fill Type Code                        0 non-null      float64
 16  NG PSI                                   0 non-null      float64
 17  EV Level1 EVSE Num                       181 non-null    float64
 18  EV Level2 EVSE Num                       48925 non-null  float64
 19  EV DC Fast Count                         7865 non-null   float64
 20  EV Other Info                            36 non-null     object
 21  EV Network                               55695 non-null  object
 22  EV Network Web                           49568 non-null  object
 23  Geocode Status                           55695 non-null  object
 24  Latitude                                 55695 non-null  float64
 25  Longitude                                55695 non-null  float64
 26  Date Last Confirmed                      55629 non-null  object
```

Figure 16. Summary of EV charging stations data

	DataYear	County	Fuel	MakeName	Model	Class	BodyStyle	VehicleType	Number of Vehicles
0	2022.0	Alameda	Electric	Blue Bird	All American / All Canadian	8	School Bus	Bus	18.0
1	2022.0	Alameda	Electric	BYD	Electric Truck	8	Tractor Truck	Truck	2.0
2	2022.0	Alameda	Electric	BYD	Electric Truck	6, 7, 8	Straight Truck	Truck	1.0
3	2022.0	Alameda	Electric	BYD	Transit Bus 40F	8	Transit Bus	Bus	1.0
4	2022.0	Alameda	Electric	Greenpower Motor	Evc210	4	Commercial / Shuttle Bus / Coach	Bus	1.0

Figure 17. Showing sample raw data samples from transit bus stop and school bus terminals data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 59295 entries, 0 to 59294
Data columns (total 18 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Vehicle ID                    59295 non-null  object
 1   Charger ID                    59295 non-null  object
 2   Local Connect Time            0 non-null      float64
 3   Local Disconnect Time         0 non-null      float64
 4   Local Charge Start Time       43055 non-null  object
 5   Local Charge End Time         36950 non-null  object
 6   Average Power                 55900 non-null  float64
 7   Max Power                     39660 non-null  float64
 8   Total Energy Delivered        59290 non-null  float64
 9   Starting SOC                  39415 non-null  float64
 10  Ending SOC                    43055 non-null  float64
 11  Starting SOC                  2780 non-null   float64
 12  Date                          16240 non-null  object
 13  Number of Charging Sessions   9810 non-null   float64
 14  Connection Time               0 non-null      float64
 15  Charging Time                 16240 non-null  float64
 16  SOC Charged                   14660 non-null  float64
 17  Local Charge E/Time           6105 non-null   object
dtypes: float64(12), object(6)
memory usage: 8.1+ MB
```

Figure 18. Summary of vehicle fleet transaction data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 9 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   DataYear             280 non-null    float64
 1   County               280 non-null    object
 2   Fuel                 280 non-null    object
 3   MakeName             280 non-null    object
 4   Model                280 non-null    object
 5   Class                280 non-null    object
 6   BodyStyle            280 non-null    object
 7   VehicleType          280 non-null    object
 8   Number of Vehicles   280 non-null    float64
dtypes: float64(2), object(7)
memory usage: 23.3+ KB
```

Figure 19. Summary of the transit bus stop and school bus terminals data

3.3 Data Pre-processing

In the data preprocessing stage, for all the four models namely, the time-series forecasting model Prophet, Temporal Fusion Transformer (TFT) model, PuLP Linear Programming Optimization with K-means Clustering, and Stacking Ensemble Regressor (Weighted Fusion), a meticulous and customized approach is used in order to guarantee the precision and dependability of the ensuing analyses. The initial step involves the storage of raw data in a robust platform, providing a secure and scalable environment for data handling. The selected approach offers a scalable framework that is suitable for managing large datasets typical of electric vehicle charging infrastructure assessments, in addition to guaranteeing the confidentiality and dependability of the data.

In the next step, in-depth data wrangling techniques are applied to the data to handle a variety of problems, such as resolving missing, erroneous, or duplicate values. Each model has specific preprocessing processes that are meticulously built to meet its specific needs and characteristics. The customized preprocessing processes created for every model are one of the distinguishing features of this technique. Considering the distinct needs and attributes of every model, particular modifications and improvements are implemented.

3.3.1 Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand

In this step, for the heavy and medium-duty EV charging dataset, the data preprocessing is conducted to enhance its consistency and effectiveness. This preparatory step is undertaken to ensure that the dataset is optimized for subsequent analyses, fostering greater usability and coherency.

- **Define Column Data Types:**

To ensure correct data interpretation, the initial step is to define specific data types for columns using a 'column_data_types' dictionary. This is useful for ensuring that the columns are interpreted correctly.

- **Clean Column Names:**

  The function, 'clean_column_names' is defined to is to replace underscores for spaces in column names. By doing this, a standardized naming system is established, which will facilitate data manipulation. This is done to create a consistent naming convention, making it easier to work with the data.

- **Convert Year to Datetime:**

  The 'Year' column in the data frame is transformed using 'pd.to_datetime' to datetime format. This transformation is likely applied to facilitate temporal analysis or plotting.

- **Removing irrelevant columns:**

  The 'Model_Year' column is removed from the data frame, this is due to the non-relevance of this intended analysis. This sequence of preprocessing steps aims to refine the dataset, making it more amenable to subsequent analyses and modeling tasks in the realm of electric vehicle data exploration.

3.3.2 Predicting Vehicle Range of Heavy-Duty and Medium-Duty EV

Below are the data preprocessing steps involved in the predicting Vehicle Range of heavy-duty and medium-duty electric vehicles.

- **Datetime conversion:**

  Date and time stamped columns such as Local Disconnect Time, Local Connect Time, Local Charge Start Time, Local Charge End Time, Date, should be converted to datetime format for more straightforward manipulation and examination.

- Some of the categorical columns like ‘Vehicle ID’, ‘Charger ID’ are converted to numerical columns using the label encoder.

The sample data is shown in Figure 20 after completing all the data preprocessing steps.

```
Sampled Data:
           Year            Fuel      Make   Duty  Vehicles Zip_Code  \
552  2016-01-01  Battery Electric  OTHER/UNK  Heavy         2    90221
2991 2022-01-01  Battery Electric  OTHER/UNK  Heavy         1    92706
1545 2018-01-01  Battery Electric  OTHER/UNK  Heavy         2    90661
191  2016-01-01  Battery Electric  OTHER/UNK  Heavy         1    93625
2015 2020-01-01  Battery Electric  OTHER/UNK  Heavy         1    92161
2780 2022-01-01  Battery Electric  OTHER/UNK  Heavy         1    94801
1037 2022-01-01  Battery Electric  OTHER/UNK  Heavy        10    95482
2162 2020-01-01  Battery Electric  OTHER/UNK  Heavy         1    92331
2695 2021-01-01  Battery Electric  OTHER/UNK  Heavy         6    93447
2742 2021-01-01  Battery Electric  OTHER/UNK  Heavy         1    94066

     primary_city state                 county
552       Compton    CA      Los Angeles County
2991    Santa Ana    CA           Orange County
1545  Pico Rivera    CA      Los Angeles County
191        Fowler    CA           Fresno County
2015    San Diego    CA        San Diego County
2780     Richmond    CA     Contra Costa County
1037        Ukiah    CA        Mendocino County
2162      Fontana    CA   San Bernardino County
2695  Paso Robles    CA  San Luis Obispo County
2742    San Bruno    CA        San Mateo County

Data Types:
Year            datetime64[ns]
Fuel                    object
Make                    object
Duty                    object
Vehicles                 int64
Zip_Code                object
primary_city            object
state                   object
county                  object
dtype: object
```

Figure 20. Sample data after the data preprocessing

3.3.3 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV

The below are the data preprocessing steps involved in the predicting short- and long-term energy demand of heavy-duty and medium-duty electric vehicles.

- **Clean Column Names:**

clean_column_names' function is created to in order to replace column names with underscores instead of spaces. The 'charging_data' dataframe is needed to be used with the function to ensure uniform column naming.

- **Remove the irrelevant columns:**

  Local_Connect_Time" and "Local_Disconnect_Time" columns are removed from the dataframe.

- **Renaming the column names:**

  Some of the column names are changed for the better understanding of the data such as, 'Local_Charge_E/Time' is changed to 'Local_Charge_End_Time'.

- **Convert to Datetime:**

  The columns 'Local_Charge_Start_Time' and 'Local_Charge_End_Time' are converted to datetime format.

The sample data snippet is shown in Figure 21 after completing all the data preprocessing steps.

	Vehicle_ID	Charger_ID	Local_Charge_Start_Time	Local_Charge_End_Time	Average_Power	Max_Power	Total_Energy_Delivered	Starting_SOC	Ending_SOC
0	EV032	CH007-CH008	2018-08-13 17:51:00	2018-08-13 18:00:00	337.797	359.58	33.940	54.0	100.0
1	EV032	CH007-CH008	2018-08-13 16:47:00	2018-08-13 16:55:00	24.101	132.10	33.551	53.5	99.0
2	EV032	CH007-CH008	2018-08-13 15:48:00	2018-08-13 15:56:00	346.784	372.34	32.824	54.0	98.5
3	EV032	CH007-CH008	2018-08-13 14:46:00	2018-08-13 14:52:00	195.148	339.36	29.633	55.5	94.5
4	EV032	CH007-CH008	2018-08-13 13:46:00	2018-08-13 13:54:00	302.232	358.74	29.293	59.5	100.0
...	...	...	...	...	...	...	...	...	...
1184	EV032	CH007-CH008	2018-01-05 06:25:00	2018-01-05 06:34:00	254.277	337.78	36.742	40.0	88.5
1185	EV032	CH007-CH008	2018-01-03 15:28:00	2018-01-03 15:52:00	22.699	41.42	13.089	61.5	86.0
1186	EV032	CH007-CH008	2018-01-03 15:10:00	2018-01-03 15:16:00	-0.763	-0.70	0.027	60.0	62.0
1187	EV032	CH007-CH008	2018-01-02 14:49:00	2018-01-02 14:51:00	-0.683	-0.66	0.009	60.5	60.5
1188	EV032	CH007-CH008	2018-01-02 14:31:00	2018-01-02 14:35:00	-0.778	-0.74	0.016	61.0	60.5

Figure 21. Sample data after the data preprocessing

3.3.4 Optimal Placement of New Charging Stations for Heavy Duty and Medium Duty EV

The below are the data preprocessing steps for the Optimal Placement of New Charging Stations for Heavy Duty and Medium Duty EV.

- **Change column names:**

  The column names having spaces are replaced with underscores using replace function in Python, is done for the better understanding of the data.

- Filtering based on the city column where we have filtered the dataframe to include only rows where the 'City' column is equal to 'San Jose'.

The sample data snippet is shown in Figure 22 after completing all the data preprocessing steps.

```
    Fuel_Type_Code                        Station_Name      Street_Address  \
146           ELEC                      Premier Nissan  1120 W Capitol Expy
171           ELEC                  Cupertino Electric       1132 N 7th St
202           ELEC  Homewood Suites - San Jose Airport      10 W Trimble Rd
510           ELEC                  American Red Cross       2731 N 1st St
515           ELEC                 Elements Apartments    1201 Parkmoor Ave

         City State    ZIP  EV_Level2_EVSE_Num  EV_DC_Fast_Count   Latitude  \
146  San Jose    CA  95136                 1.0               2.0  37.275144
171  San Jose    CA  95112                12.0               NaN  37.362180
202  San Jose    CA  95131                 2.0               NaN  37.385140
510  San Jose    CA  95134                 2.0               NaN  37.390356
515  San Jose    CA  95126                 1.0               NaN  37.317246

      Longitude     ID Facility_Type
146 -121.877656  39935    CAR_DEALER
171 -121.899539  46482       UTILITY
202 -121.928800  49342         HOTEL
510 -121.932236  89034   OFFICE_BLDG
515 -121.909487  89626   PARKING_LOT
```

Figure 22. Sample data after the data preprocessing

3.4 Data Transformation

One of the most important stages of data analysis is data transformation, involving the systematic conversion of raw data into a more structured and suitable format for analysis and modeling. This multifaceted process encompasses a range of operations aimed at enhancing data compatibility with specific analytical tools or models. Common data transformation tasks include Standardizing data formats, correcting outliers, and handling missing, erroneous, or duplicate values. In addition, transformations could entail disaggregating or aggregating data, feature-engineering new variables, or changing data types in accordance to analytical specifications. Getting the data ready for further analysis in a way that best facilitates insightful conclusions, precise forecasts, and robust decision-making in subsequent stages of analysis.

3.4.1 Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand

In time series analysis, grouping and aggregating temporal data is a common transformation as seen in the forecasting Heavy Duty and Medium Duty EV and Charging Stations Demand.

- **Grouping and Aggregation:**

  The 'Year' column in the original dataframe (df) is used to group the data, and groupby and sum are then used to determine the total number of 'Vehicles' for each year.
- Once the data has been converted, the function 'fit_and_forecast' is used to set the growth, seasonality mode, and changepoint scales, among other parameters for the modelling part.
- Extraction and printing of the anticipated EV car counts for each subsequent year until 2045 constitute the last phase, showcasing the effective use of data transformation.
- **Merging DataFrames:**

  Using the 'Zip_Code' column, the 'df' and 'zc' dataframes are combined as part of the data transformation. Next, the 'zip' column is removed and the column order is changed in the resulting'vehicle_data' dataframe. The desired column order is specified in the 'new_order' list, and the'vehicle_data' dataframe is reorganized in accordance with that order. The modified dataframe now has columns labeled 'Year,' 'Fuel,' 'Make,' 'Duty,' 'Vehicles,' 'Zip_Code,' 'primary_city,''state,' and 'county' in the following sequence.

The sample data snippet is shown in Figure 23 after merging the dataset.

	Year	Fuel	Make	Duty	Vehicles	Zip_Code	primary_city	state	county
0	2015-01-01	Battery Electric	OTHER/UNK	Heavy	1	90018	Los Angeles	CA	Los Angeles County
1	2015-01-01	Battery Electric	OTHER/UNK	Heavy	1	90031	Los Angeles	CA	Los Angeles County
2	2016-01-01	Battery Electric	OTHER/UNK	Heavy	1	90031	Los Angeles	CA	Los Angeles County
3	2016-01-01	Battery Electric	OTHER/UNK	Heavy	1	90031	Los Angeles	CA	Los Angeles County
4	2017-01-01	Battery Electric	OTHER/UNK	Heavy	1	90031	Los Angeles	CA	Los Angeles County
...	...	...	...	...	...	...	...	...	...
3006	2022-01-01	Battery Electric	OTHER/UNK	Heavy	1	95334	Livingston	CA	Merced County
3007	2022-01-01	Battery Electric	OTHER/UNK	Heavy	1	95553	Miranda	CA	Humboldt County
3008	2022-01-01	Battery Electric	NEW FLYER	Heavy	16	95822	Sacramento	CA	Sacramento County
3009	2022-01-01	Battery Electric	OTHER/UNK	Heavy	1	95822	Sacramento	CA	Sacramento County
3010	2022-01-01	Battery Electric	OTHER/UNK	Heavy	2	96105	Chilcoot	CA	Plumas County

3011 rows × 9 columns

Figure 23. Sample data after the data merge

3.4.2 Predicting Vehicle Range of Heavy-Duty and Medium-Duty EV

- **Dropping Null/Missing values:**

  Dropping the null or missing data values present in the dataset.

- **Creating new columns:**

  A new column named 'Range' created which later uses in predicting the vehicle range of heavy and medium duty EV. The column 'Range' is created by applying a formula involving 'Delta_SOC' to estimate the range.

- **Feature Selection:**

  Features that are relevant to the prediction task are selected. "Charging_Duration," "Average_Power," and "Delta_SOC" are a few of these.

Charger_ID	Local_Charge_Start_Time	Local_Charge_End_Time	Average_Power	Max_Power	Starting_SOC	Ending_SOC	Delta_SOC	Charging_Duration	Range
CH007-CH008	2018-08-13 17:51:00	2018-08-13 18:00:00	337.797	359.58	54.0	100.0	46.0	0.150000	110.4
CH007-CH008	2018-08-13 16:47:00	2018-08-13 16:55:00	24.101	132.10	53.5	99.0	45.5	0.133333	109.2
CH007-CH008	2018-08-13 15:48:00	2018-08-13 15:56:00	346.784	372.34	54.0	98.5	44.5	0.133333	106.8
CH007-CH008	2018-08-13 14:46:00	2018-08-13 14:52:00	195.148	339.36	55.5	94.5	39.0	0.100000	93.6
CH007-CH008	2018-08-13 13:46:00	2018-08-13 13:54:00	302.232	358.74	59.5	100.0	40.5	0.133333	97.2

Figure 24. Sample data after performing the data transformation.

3.4.3 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV

- **Handling missing values:**

  Removed rows from the Data Frame charging_data that contain missing (null) values. SimpleImputer is used to handle missing values in the resampled data. Missing values are filled with the mean value of the respective column using the fit_transform and transform methods for daily, weekly, and monthly forecasts, respectively.

- **Creating new columns:**

  The new variables such as 'Charging_Duration,' which is created by finding the time difference in hours between 'Local_Charge_Start_Time' and 'Local_Charge_End_Time'. And, Average_Power,' and 'Delta_SOC.' are created and added to the existing data to better represent and summarize aspects of the data.

- **Feature Engineering:**

  Feature engineering done selecting the relevant columns which are created such as 'Delta_SOC' and 'Charging_Duration', feature selection, and defining the target variable ('Range’).

The sample data snippets are shown in Figure 25 before and after handling the null values.

```
Vehicle_ID                 0
Charger_ID                 0
Local_Charge_Start_Time    0
Local_Charge_End_Time      0
Average_Power              0
Max_Power                  0
Total_Energy_Delivered     0
Starting_SOC               2
Ending_SOC                 0
dtype: int64
```

```
Vehicle_ID                 0
Charger_ID                 0
Local_Charge_Start_Time    0
Local_Charge_End_Time      0
Average_Power              0
Max_Power                  0
Total_Energy_Delivered     0
Starting_SOC               0
Ending_SOC                 0
dtype: int64
```

Figure 25. Sample data before and after handling the null values.

The sample data snippet is shown in Figure 26 after performing the data transformation techniques.

```
  Vehicle_ID   Charger_ID Local_Charge_Start_Time Local_Charge_End_Time  \
0      EV032  CH007-CH008     2018-08-13 17:51:00   2018-08-13 18:00:00
1      EV032  CH007-CH008     2018-08-13 16:47:00   2018-08-13 16:55:00
2      EV032  CH007-CH008     2018-08-13 15:48:00   2018-08-13 15:56:00
3      EV032  CH007-CH008     2018-08-13 14:46:00   2018-08-13 14:52:00
4      EV032  CH007-CH008     2018-08-13 13:46:00   2018-08-13 13:54:00

   Average_Power  Max_Power  Total_Energy_Delivered  Starting_SOC  Ending_SOC  \
0        337.797     359.58                  33.940          54.0       100.0
1         24.101     132.10                  33.551          53.5        99.0
2        346.784     372.34                  32.824          54.0        98.5
3        195.148     339.36                  29.633          55.5        94.5
4        302.232     358.74                  29.293          59.5       100.0

   Delta_SOC  Charging_Duration  Range
0       46.0           0.150000  110.4
1       45.5           0.133333  109.2
2       44.5           0.133333  106.8
3       39.0           0.100000   93.6
4       40.5           0.133333   97.2
```

Figure 26. Sample data after performing the data transformation.

3.4.4 Optimal Placement of New Charging Stations for Heavy Duty and Medium Duty EV

The data transformation steps involved in creating the new columns, handling the missing values are handled effectively prior to additional analysis or modeling. In this instance, replacing missing values with zeros or a designated string preserves the dataset's integrity of the dataset and prevents potential issues in subsequent computations or visualizations.

- **Creating new columns:**

  Based on the values in the 'EV Level2 EVSE Num' column, a new column called 'Charger_Type' is formed. A lambda function is used to determine the values in this new column. The value in 'EV Level2 EVSE Num' is 0, the 'Charger_Type' is set to 0 (indicating a DC Charger); otherwise, it is set to 1 (indicating a Level 2 Charger).

- **Conversional to categorical:**

  The pd.Categorical function is then used to transform the 'Charger_Type' column into a categorical variable. This might be helpful in some kinds of studies or visualizations and is frequently done to optimize memory usage.

- **Handling null/missing values:**

  Some of the columns in the dataframe having null/missing values such as 'EV_Level2_EVSE_Num', 'EV_DC_Fast_Count', 'Facility_Type' are removed and replaced with zero in the dataset respectively using the 'fillna' method in python.

- **Creating new columns:**

  A new GeoDataFrame named 'proposed_charging_stations' is created with columns 'Latitude', 'Longitude', and 'Capacity'. And, the extraction of Latitude and Longitude from 'geometry' column and stored in new columns.

- **Dropping irregular columns:**

The 'geometry' column is removed from the GeoDataFrame since latitude and longitude information has been extracted. And also dropped the rows having the null values from the data frame to prepare the data for the next steps.

The sample data snippets are shown in Figure 27 before and after handling the null values.

```
Fuel_Type_Code          0
Station_Name            0
Street_Address          0
City                    0
State                   0
ZIP                     0
EV_Level2_EVSE_Num     29
EV_DC_Fast_Count      516
Latitude                0
Longitude               0
ID                      0
Facility_Type         524
Charger_Type            0
dtype: int64
```

```
Fuel_Type_Code        0
Station_Name          0
Street_Address        0
City                  0
State                 0
ZIP                   0
EV_Level2_EVSE_Num    0
EV_DC_Fast_Count      0
Latitude              0
Longitude             0
ID                    0
Facility_Type         0
Charger_Type          0
dtype: int64
```

Figure 27. Sample data before and after handling of the null values

The sample data snippet is shown in Figure 28 after performing the data transformation.

bus	local_ref	...	network:wikidata	surface	covered	lit	old_name	wikidata	operator:wikidata	wikipedia	Latitude	Longitude
None	None	...	None	None	None	None	None	None	None	None	37.336490	-121.884501
None	None	...	None	None	None	None	None	None	None	None	37.336254	-121.885414
None	None	...	None	None	None	None	San Fernando & Almaden	None	None	None	37.333168	-121.891784
None	8	...	Q1456861	None	None	None	None	None	None	None	37.368459	-121.927689
None	None	...	None	None	None	None	None	None	None	None	37.351472	-121.902293

Figure 28. Sample data after performing the data transformation.

3.5 Data Preparation

Data preparation is a pivotal phase in the lifecycle of any analytical or machine learning project, involving a series of steps to transform raw data into a suitable format for modeling or the analysis. This vital process starts with gathering data from various sources and moves forward through stages of integration, cleansing, and transformation. Integration creates a cohesive format by merging data from several sources, enabling a thorough understanding of the data. Data transformation entails transforming data into an appropriate format, which may include generating derived features, standardizing formats, and altering data types. These meticulous steps aim to enhance data quality and coherence, laying a solid foundation for subsequent analyses and model development.

3.5.1 CA Vehicle Registration Data

From the CA vehicle registration dataset, is used to predict the count of vehicles and charging station demand. The vehicle registration data having 35,742 x 9 columns entities of raw data transactions respectively of data to be processed. There are three phases, which will consist of Training, Validation, and Testing of these data, and these phases will be run in phases as explained below. Consequently, the data was divided into three categories: training, validation, and testing. The training ratio was 80 percent, and the validation and testing ratios were 10 percent each. The entire dataset has been assigned a proportion of 80:10:10.

As illustrated in Table 21, 80% of the data is for training and 10% is for testing and validation. There would be 28571 data fields in the trained model, and 3571 data fields in the test and validation model for the CA vehicle registration dataset. The reason why it was selected was because it is widely used and accepted as an acceptable split ratio. It is because of this that it strikes the right balance between providing enough data for training a model, but also providing enough data to test and validate it, that it strikes the right balance between providing both enough data for training and testing. As far as training a model is concerned, it is always better to keep a higher proportion for training and keep the other two constant at all times. In the next steps, created fit_and_forecast function with the yearly_vehicle_count Data Frame and specified parameters for interval width. This prepares a Prophet model, fits it to the data, and forecasts EV vehicle counts until 2045.

Table 21. Data preparation for predicting vehicle count.

Dataset	Training and Test Data	Classes to Predict
CA Vehicle Registration data	Training: 28571 Test: 3571 Validation: 3571	'Vehicle count'

	Year	Zip_Code	Vehicle_Count
1561	2021.0	94124	298
279	2017.0	94124	288
707	2019.0	94124	322
424	2018.0	94124	5
1990	2022.0	94124	122

Figure 29. Sample data snippet of training data

3.5.2 CA EV charging station data

The same ratio 80:10:10 will be used for the CA electric vehicle charging station dataset where there would be 11869 data fields in the trained model and 1483 data fields for the test model as shown in Table 22. Among machine learning tasks, the 80/10/10 split ratio is widely used and accepted as an acceptable split ratio. The reason why it was selected was because it is widely used and accepted as an acceptable split ratio. It is because of this that it strikes the right balance between providing enough data for training a model, but also providing enough data to test and validate it, that it strikes the right balance between providing both enough data for training and testing.

The process of preparing data in order to find potential sites for future charging stations in San Jose, California. First, a 15 km radius around San Jose is the only area in which the charging stations dataset is included. Then, to find possible sites for future charging stations, a customized k-means clustering method is implemented. Assuring the exclusion of current charging station sites, the method starts with random centers based on transit bus stop coordinates. Overall, these data preparation steps integrate spatial analysis and clustering techniques to strategically identify and visualize new charging station locations in the San Jose area.

Table 22. Data preparation for finding optimal locations.

Dataset	Training and Test Data	Classes to Predict
CA EV charging station data	Training: 11869 Test: 1483 Validation: 1483	‘New charging station’

3.5.3 Vehicle Fleet Transactions data

The vehicle fleet transaction data is used to predict the short- and long-term energy demand. Likewise, the vehicle fleet transactions dataset will employ the same ratio, with 342887 data fields in the trained model and 42861 data fields in the test and validation models, as indicated in Table 23. In the process of preparing data for a model in the context of vehicle fleet transactions, several key steps are taken. First, a subset of the charging_data is chosen to include elements such as 'Average_Power,' 'Max_Power,' 'Starting_SOC,' 'Ending_SOC,' and

'Charging_Duration,' which are essential for forecasting the overall energy delivered during charging. Next, the variable 'Total_Energy_Delivered' is recognized as the target, which stands for the expected result (y). In order to create precise and broadly applicable models for forecasting energy demand, this division makes sure the model can be trained on a subset of the data and evaluated for its predictive performance on an independent subset.

In the next steps, the 'Range,' the target variable, is set up for prediction (y), and the attributes 'Charging_Duration,' 'Average_Power,' and 'Delta_SOC' are designated as input features (X) for the model. Dividing the dataset into separate training and testing sets is the next crucial step. The train_test_split function from scikit-learn is used to accomplish this, using a given random state for reproducibility and a test size of 10%. This division facilitates the training of the machine learning model on one subset and the evaluation of its performance on an independent subset, ensuring robustness and accuracy in predicting EV range.

Table 23. Data preparation for predicting energy demand and range

Dataset	Training and Test Data	Classes to Predict
Vehicle fleet transactions data	Training: 342887 Test: 42861 Validation: 42861	‘Total_energy_delivered’, ‘range

	Vehicle_ID	Charger_ID	Charging_Duration	Local_Charge_Start_Time	Average_Power	Delta_SOC	Max_Power
298	EV032	CH007-CH008	0.150000	2018-06-30 16:43:00	249.335	42.5	364.96
711	EV032	CH007-CH008	0.133333	2018-04-09 08:21:00	169.412	28.0	210.78
654	EV032	CH007-CH008	0.166667	2018-04-17 18:16:00	293.226	44.0	366.46
275	EV032	CH007-CH008	0.150000	2018-07-03 11:02:00	227.246	40.5	278.32
599	EV032	CH007-CH008	0.083333	2018-05-02 08:21:00	224.093	21.5	367.28

Figure 30. Sample data snippet of training data

Figure 31 illustrates the overall data preparation details such as total number of train, test and validation rows for all the datasets used in this research.

Dataset/Task	Raw	Preprocessed	Prepared			
			Total Rows	Train Rows	Test Rows	Validation Rows
CA Vehicle Registration Data	35,742	35713	35713	28571	3571	3571
CA EV Charging Station Data	14862	14835	14835	11869	1483	1483
Vehicle Fleet Transactions Data	4,28,963	4,28,609	4,28,609	3,42,887	42861	42861

Figure 31. Dataset details after Data Preparation

3.6 Data Statistics and Data Analytics Results

The data statistics offer a numerical summary of diverse aspects related to electric vehicles and charging infrastructure. These metrics provide valuable insights into the features and patterns present in the datasets. For instance, these statistics encompass details such as the count of electric vehicles, charging stations, transit bus stops, and school bus terminals. Essential metrics may involve parameters like the number of vehicles, charging station capacities, geographical spread, and patterns of energy consumption. Analyzing these statistics provides insights into the present scenario of electric vehicle adoption, the geographic distribution of charging infrastructure, and the requirements for sustainable transportation solutions. Additionally, the quantitative measures derived from data statistics serve as a fundamental basis for further analyses, empowering researchers and stakeholders to recognize patterns, make well-informed decisions, and contribute to the continuous advancement of electric mobility solutions.

These metrics are valuable instruments for strategic planning, optimizing infrastructure, and advocating sustainable practices within the domain of medium and heavy-duty electric vehicles.

The datasets used in this project are open-source datasets that must be ready for training, validation, and testing to create a reliable and accurate categorization model for EV user comments. Our transformed dataset consists of 479,157 rows, which will be divided into three sets in an 80:10:10 ratio: training, validation, and testing. As part of the data preparation process, the rows are first randomly shuffled to ensure an equal distribution of comments throughout the three sets. As a result, there will be a reduction in bias in the distribution of classes in the dataset. As a result of applying the aforementioned ratios to the shuffled data, three sets will be created. A classification model will be trained based on the training set, which contains 80% of the data. There are approximately 383,325 records. There will be approximately 47,915 rows in the validation set, representing 10% of the data. By doing so, we will be able to fine-tune the model's hyperparameters and prevent overfitting. With the test set, the performance of the model will be assessed with respect to unobserved data. As well as this, there are 47,915 records in the data that represent 10% of the data. The following figures present statistics for both raw and processed data.

Dataset	Vehicle Registration-Zip code	Charging Stations- Zip Code	Vehicle Fleet Transactions-Weight Class	Transit Bus Stations-City	School Bus Terminals-City
Data Type	.xslx	.csv	.xslx	.geojson	.csv
Size	28.8 MB	16.7 MB	126.8 MB	1.6 MB	3.4 MB
Count	35,742	14,862	4,28,963	1289	3526
Features	15	65	48	12	16
Duplicate Rows	0	0	96	0	0
Missing Values	29	27	258	0	0

Figure 32. Raw Data Statistics

Dataset	Vehicle Registration-Zip code	Charging Stations- Zip Code	Vehicle Fleet Transactions-Weight Class	Transit Bus Stations-City	School Bus Terminals-City
Data Type	.csv	.csv	.csv	.csv	.csv
Size	14.2 MB	8.9 MB	87.1 MB	89 KB	1.4 MB
Count	35,713	14,835	4,28,609	1289	3526
Features	7	15	14	5	4
Duplicate Rows	0	0	0	0	0
Missing Values	0	0	0	0	0

Figure 33. Processed Data Statistics

3.6.1 Data Analytics results

The data analytics results encompass a comprehensive examination of various aspects related to electric vehicles (EVs) and charging stations. This analytical process entails the processing and interpretation of data to extract valuable insights. The results of the data analytics provide a comprehensive understanding of the present status of EV adoption, charging infrastructure, and energy demand. These insights play a crucial role in making well-informed decisions, promoting sustainable practices, and facilitating the ongoing shift towards electric mobility in the medium and heavy-duty vehicle sectors.

3.6.1.1 Medium and Heavy-duty Vehicles Data Analysis

Understanding data about Medium and Heavy-duty vehicles means learning about vehicles that are notable for their size and weight and are frequently used for business and industrial purposes. This understanding is critical for a variety of applications, including fleet management, transportation planning, and environmental impact assessment.

Figure 34 shows the vehicle type that is bus, truck, and delivery vans population data across the counties in California state. Heavy and medium-duty vehicle manufacturers and configurations are related to the companies that make these vehicles and the unique body type or layout that is specific to the uses for which they are intended. For a variety of customers, including fleet managers, transportation planners, regulatory agencies, and manufacturers themselves, knowing the manufacturer and body type is crucial. To accommodate a variety of transportation needs, manufacturers may offer a range of options or specialize in particular body types.

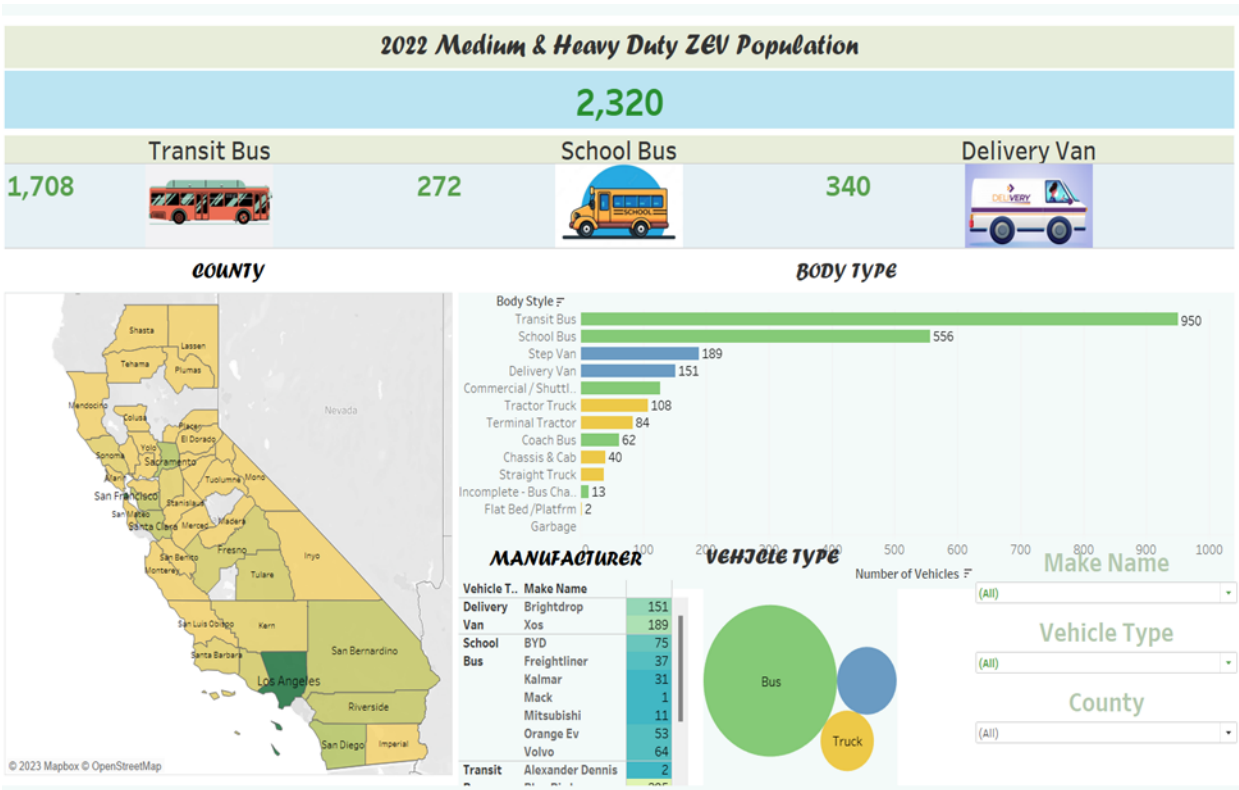


Figure 34. Showing Medium and Heavy-Duty population data for a total 2022

The Dashboard boasts an engaging and insightful visual representation that vividly captures the distribution of Medium and Heavy-duty vehicle data statistics. Through visually appealing charts, graphs, and interactive data visualizations, users can explore a comprehensive overview of the spatial distribution of these vehicles, gaining a deep understanding of their prevalence and usage patterns. The visual depiction not only facilitates an at-a-glance understanding of the concentration of Medium and Heavy-duty vehicles across different regions but also allows for nuanced insights into variations in their counts and demands. This dynamic visual component serves as a powerful tool for stakeholders, providing a clear and accessible means to interpret complex data, make informed decisions, and shape strategies for optimizing charging infrastructure to meet the specific needs of these specialized vehicles. It aligns with the website's commitment to delivering actionable insights through a visually compelling and user-friendly interface, enhancing the overall user experience and promoting informed decision-making in the realm of electric vehicle infrastructure planning.

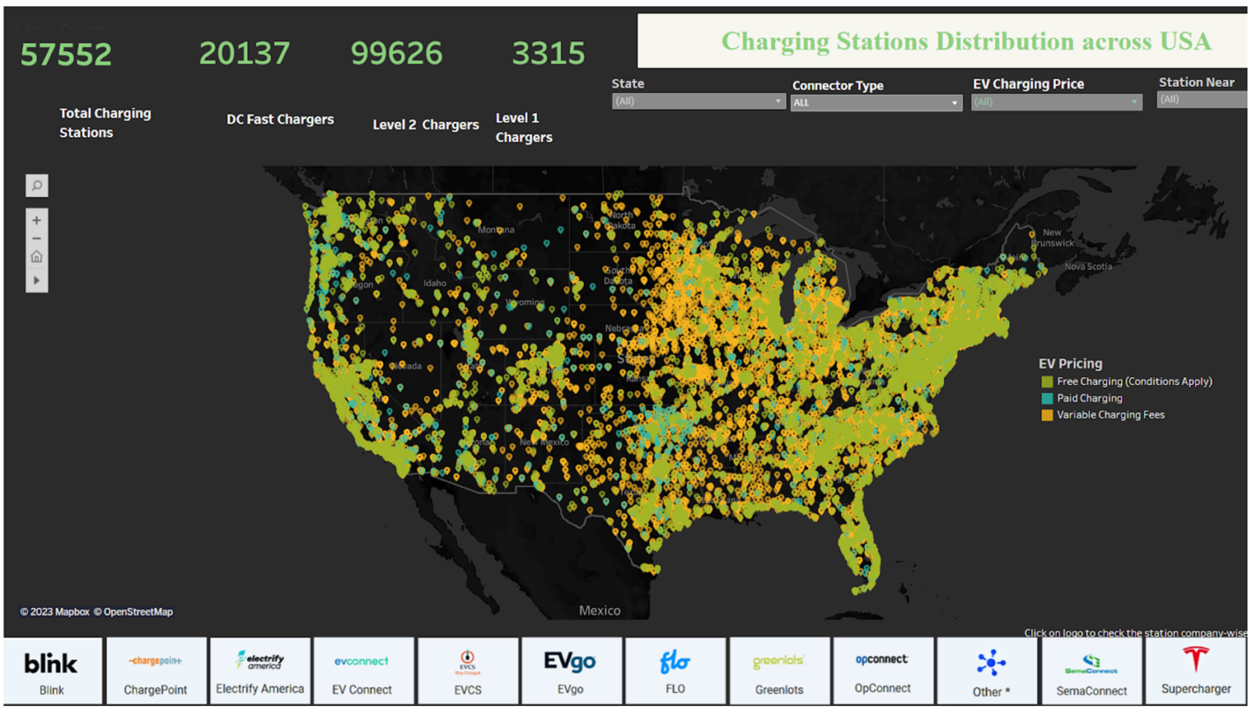


Figure 35. Charging Station Distribution across United States

3.6.1.2 Electric Vehicle charging types with respect to the cities in California

Understanding the variety of charging infrastructure found in urban areas is necessary to understand the relationship between EV charging types and cities. Different charging methods satisfy different requirements and inclinations. The successful integration of electric vehicles into urban transportation ecosystems is facilitated by charging types that are tailored to the specific needs of a diverse population living in cities.

The Figure 36 shows the EV DC Fast chargers count and EV Level 2 chargers count by zip codes in the cities of San Francisco and San Jose in California state. From the insights we can see that EV Level 2 charging types are the most widely used charger types in both San Francisco and San Jose cites. It will be beneficial for the users with more EV Level 2 charging types at new locations in the cities of California state.

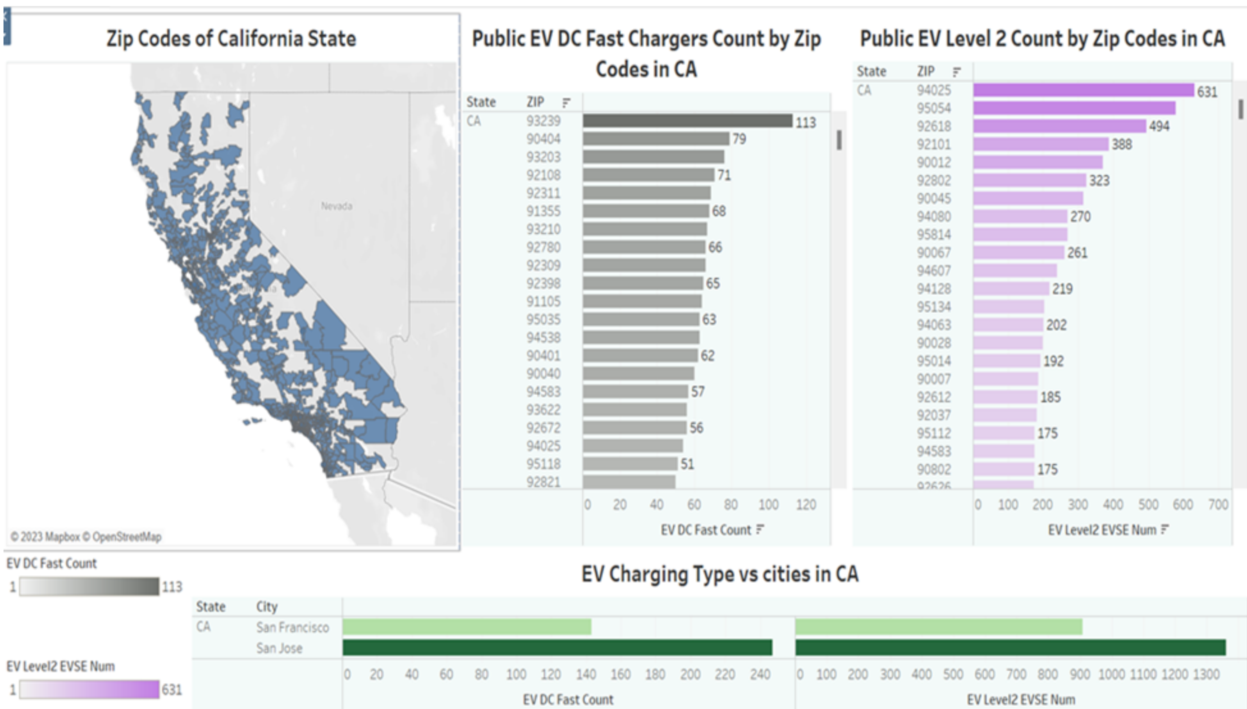


Figure 36. Showing EV charging types with respect to the cities in California

3.6.1.3 Insights about the EV charging types with different facility groups in California state

To create a flexible and easily accessible charging network, it is important to have different facility groups centered around different types of EV charging. By addressing charging challenges in a variety of scenarios and contexts, this approach not only supports the current needs of EV owners but also helps to promote the growth and acceptance of electric vehicles.

A number of benefits come from positioning different EV charging stations close to popular destinations, which supports the growing popularity and practicality of electric cars. Incorporating charging infrastructure near points of interest encourages EV adoption, enhances user experience, addresses range anxiety, boosts local businesses, increases charging accessibility, aligns with urban planning goals.

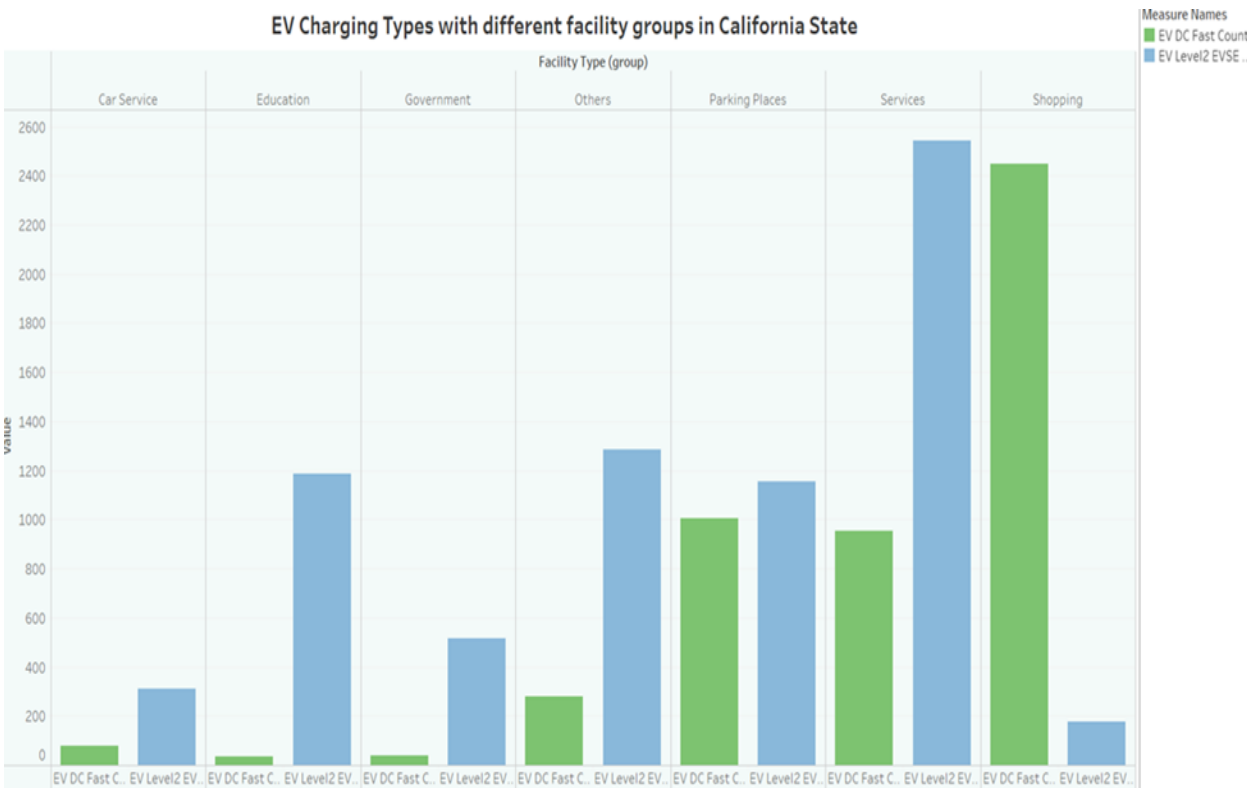


Figure 37. Showing different facilities around EV charging types.

3.6.1.4 Insights about the energy delivered at each charging station

This analysis entails extracting valuable details and observations related to the quantity of energy provided at each charging station. The analysis explores patterns, trends, and fluctuations in energy delivery, offering insights into the charging dynamics at particular locations. These findings aid in optimizing the placement of charging stations, identifying usage patterns, and enhancing the overall management of energy efficiency in the electric vehicle charging network.

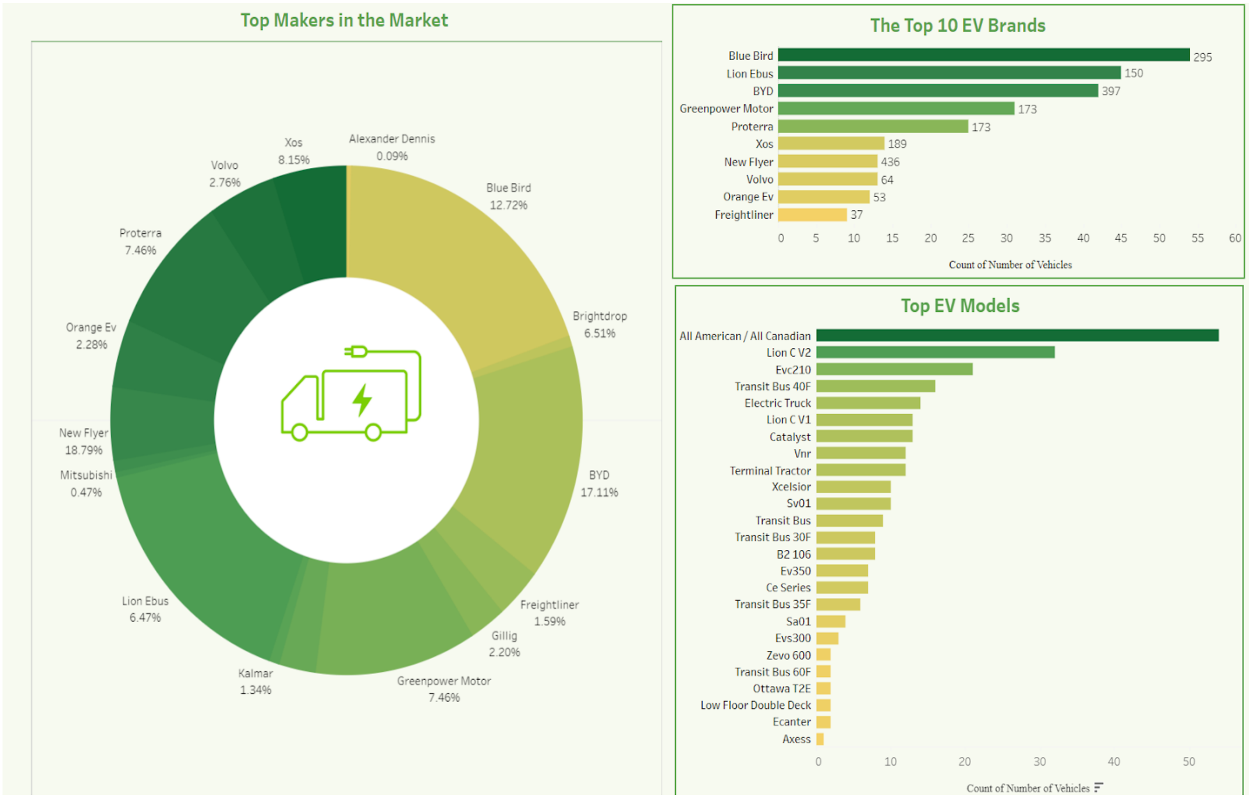


Figure 38. Distribution of Heavy and Medium-Duty Vehicles in the Market

3.6.1.5 Optimal placement of new charging stations

Placing new, ideal EV charging stations is crucial to increasing the usability, convenience, and appeal of electric vehicles for a wider variety of consumers. It helps create a robust and user-friendly EV infrastructure and promotes the shift to environmentally friendly modes of transportation.

Figure 40 shows the newly placed charging stations for the zip code 94124. The previously mentioned techniques have shown themselves to be very helpful in the planning of new EV charging infrastructure inside of a given zip code. These methods improve our understanding of how drivers behave when attempting to charge their electric cars and provide important information for well-thought-out infrastructure development.

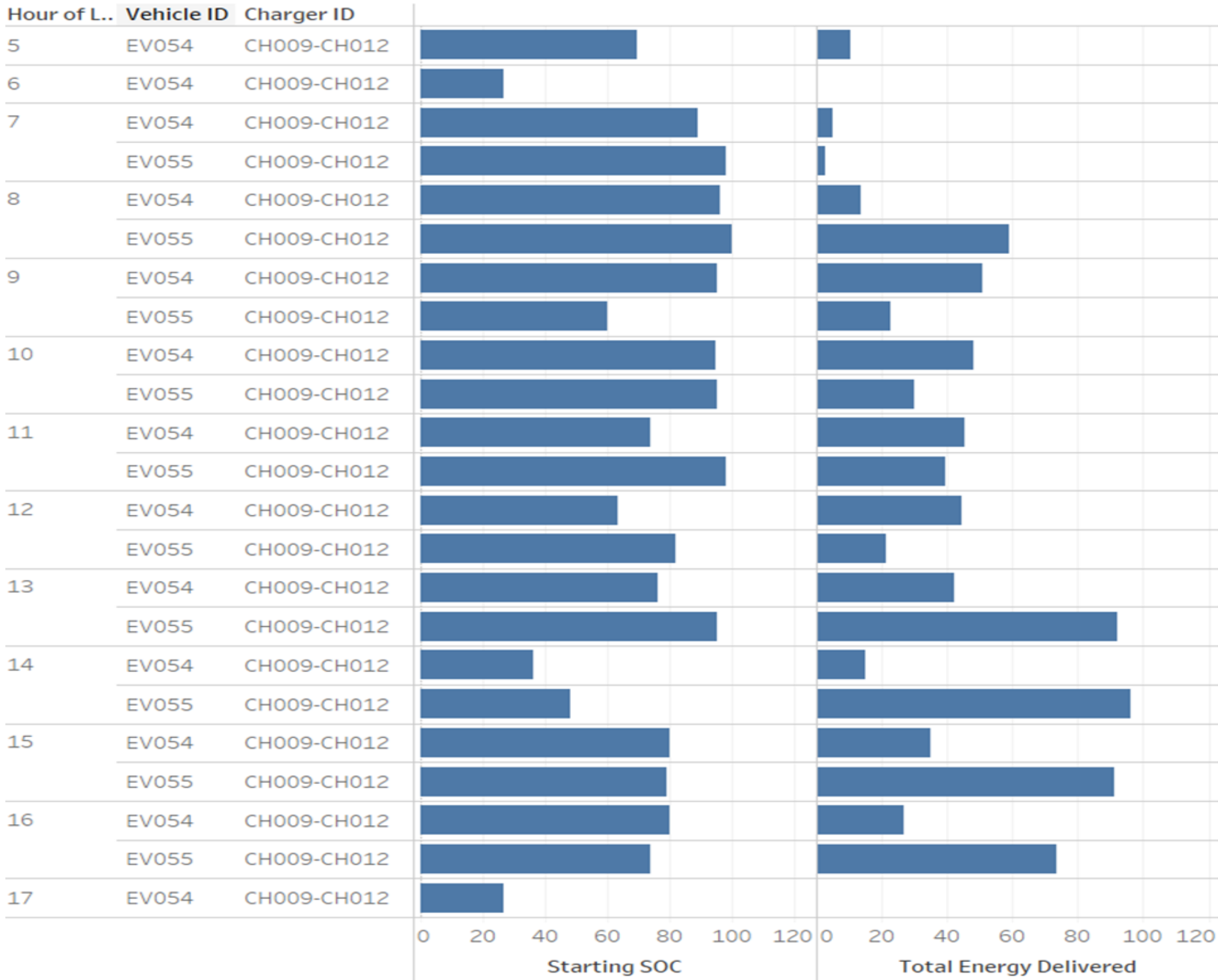


Figure 39. Machine Learning Results of Energy Delivered at Each Charging Station

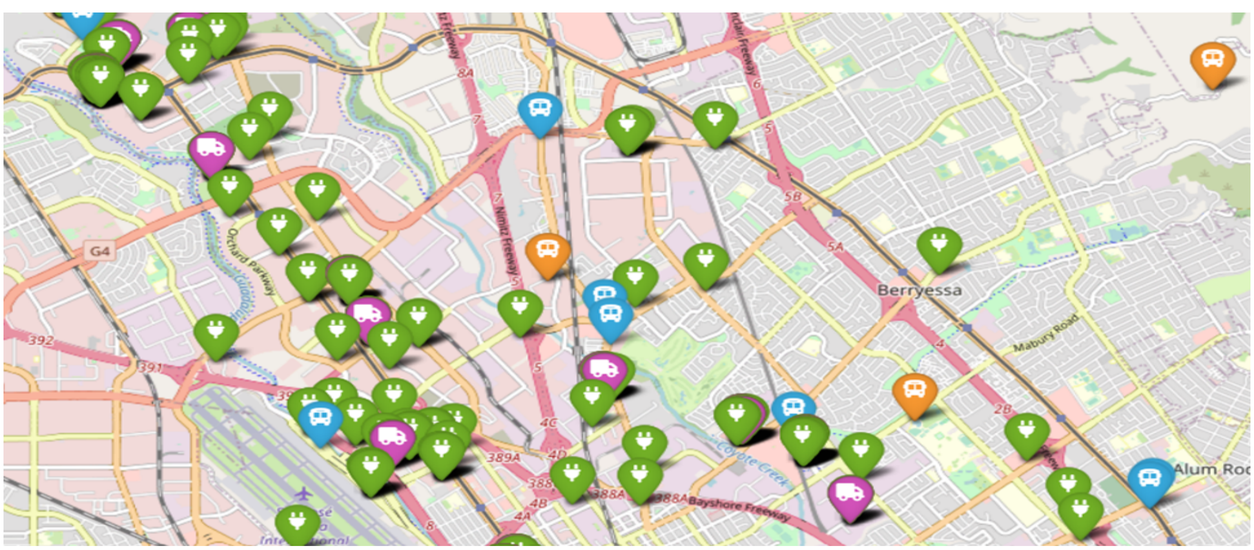


Figure 40. Webpage Displaying the Newly Suggested Location for EVCS

4. Model Development

4.1 Model Proposals

Electric vehicles (EVs) are quickly taking over as the preferred mode of transportation, so it is critical to strategically build out a reliable charging infrastructure. The availability of EV charging stations is crucial for the smooth transition to electric mobility, but their strategic location is also crucial for maximizing accessibility and use. This proposal describes a ground-breaking initiative that would use machine learning to develop a smart system for the accurate planning and deployment of EV charging infrastructure, designed especially for medium and heavy-duty motorized vehicles.

Our strategy envisions the design of an intelligent system that will revolutionize the deployment and planning of EV charging stations, particularly the kind that is customized to the requirements of medium- and heavy-duty cars. The ability to forecast the anticipated number of vehicles expected to be on the road in the years preceding 2035 serves as the basis for our intelligent EV charging infrastructure system. This predictive modeling strategy is based on time series machine learning algorithms that examine historical data and demographic trends of EV registrations, allowing us to forecast the rising demand for EV charging. We can calculate the number of charging stations needed to meet this demand by precisely calculating the number of vehicles, ensuring that EV users have easy access to charging stations.

Understanding the unique needs of various vehicle fleets operating in various regions, in addition to calculating the necessary number of charging stations, is one of the main issues in the planning of EV infrastructure. As a result, our model goes a step further by including information on the energy requirements and driving range of three different categories of heavy-duty vehicles: transit buses, school buses, and delivery vans. Due to the granularity, we can customize charging station recommendations to meet the specific needs of each fleet type, resulting in efficient and affordable solutions. Additionally, the proposal contemplates providing specific location recommendations for positioning new charging stations both within and across zip codes. These placement suggestions will be supported by data-driven insights that take into consideration variables like the anticipated number of vehicles, energy use, and traffic patterns. By carefully locating charging stations, we hope to promote programs for sustainable mobility while also meeting present and future demand for EVs.

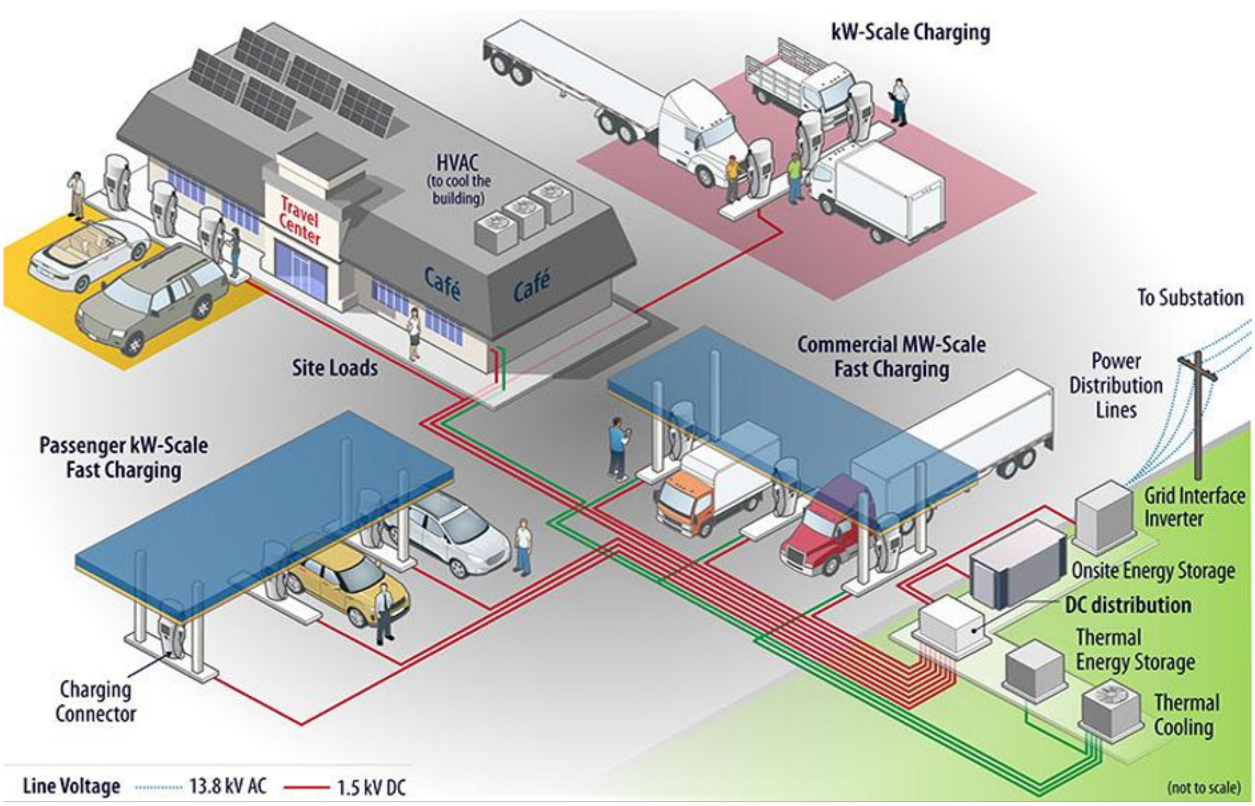


Figure 41. High-Power Medium- and Heavy-Duty Electric Vehicle Charging

In a nutshell the suggested smart electrical vehicle (EV) charging system offers a comprehensive strategy to deal with the difficult issues involved in making proposals regarding the development of electric transportation in the future. We seek to guarantee the availability, usability, and effectiveness of EV charging infrastructure for medium and heavy-duty vehicles by utilizing machine learning and data-driven insights. This program encourages the widespread use of electric vehicles and the switch to cleaner transportation, which will help create a more environmentally conscious and enduring future. The complete model architecture is outlined in the following Figure 42.

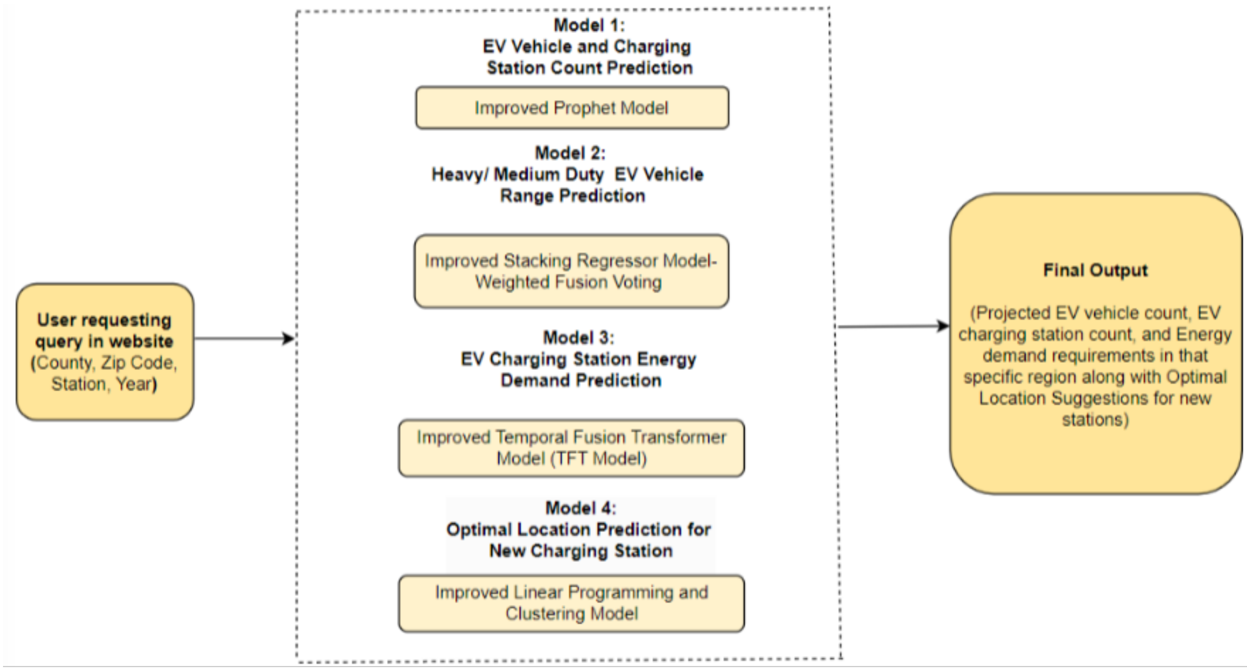


Figure 42. Model Architecture Overview

4.1.1 Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand

The availability of a dependable and widely distributed charging infrastructure is crucial as we move towards a sustainable future with electric transportation, especially for heavy and medium-duty electric vehicles (EVs). This proposal describes the creation of a prediction model using the Prophet model for predicting EV demand, including vehicle counts and the matching number of necessary charging stations, up to the year 2035. This artificial intelligence will provide insights necessary for improving the installation of EV charging infrastructure and maintaining accessibility and comfort for EV users by relying on historical information. At the outset, we compile a large dataset with details on the number of vehicles, zip codes, fuel kinds, vehicle brands, duty types, and years.

The Prophet time series forecasting model, recognized for its accuracy in capturing seasonality, trends, and holiday effects in data, will be the foundation of the suggested approach. The model will be trained on an extensive historical dataset that includes data on vehicle counts, charging station utilization, demographic trends, and other relevant factors in order to forecast future EV demand. We filter the data for each distinct zip code in the dataset, group the vehicle counts by years, then add the counts to get yearly totals. The Prophet model-compatible format is created from this aggregated data. The model has been developed to recognize seasonality, growth tendencies, and changepoints in the data, enabling precise forecasting. The model will generate projections for both the predicted numbers of heavy and medium-duty EVs on the roads and the accompanying demand for charging stations by examining this data. We will be able to catch subtle variations in EV adoption patterns over time and across different regions owing to the Prophet model's adaptability. As a result, we will be able to make more accurate predictions about the overall growth of electric vehicles and the demand for charging stations. This will help us to better plan for the future and ensure that the infrastructure needed to support electric vehicles is in place.

The implementation of this method of forecasting could have several advantages. First of all, it will offer a data-driven roadmap for future EV infrastructure development to legislators, utility corporations, and infrastructure planners. More effective resource allocation and investment choices will be made with accurate estimates of car counts and charging station requirements. Additionally, it will hasten the adoption of EVs by ensuring that the expansion of the charging infrastructure keeps pace with the increase in demand, improving the overall EV

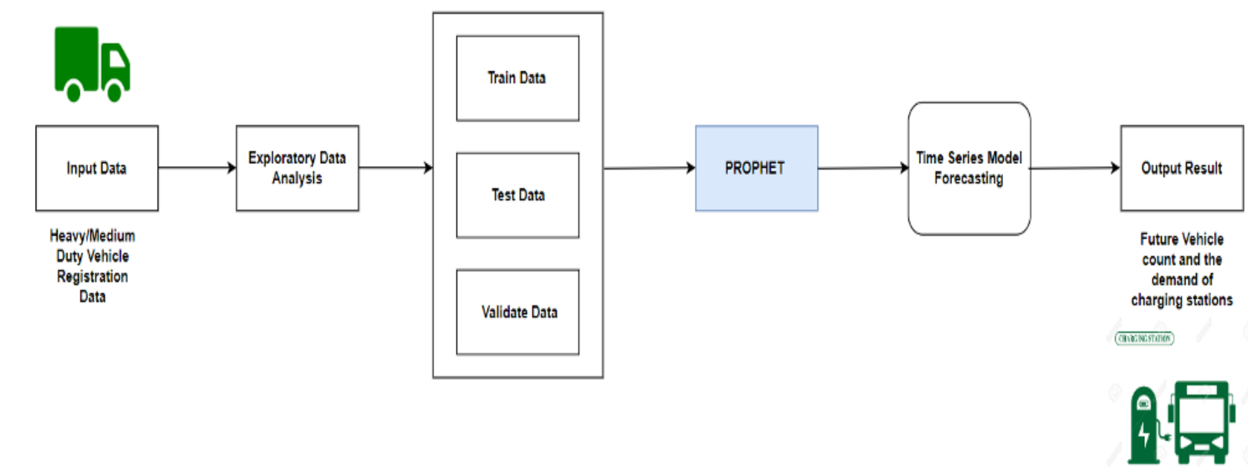


Figure 43. Model for Forecasting Heavy /Medium Duty EV Count and Charging Stations Demand

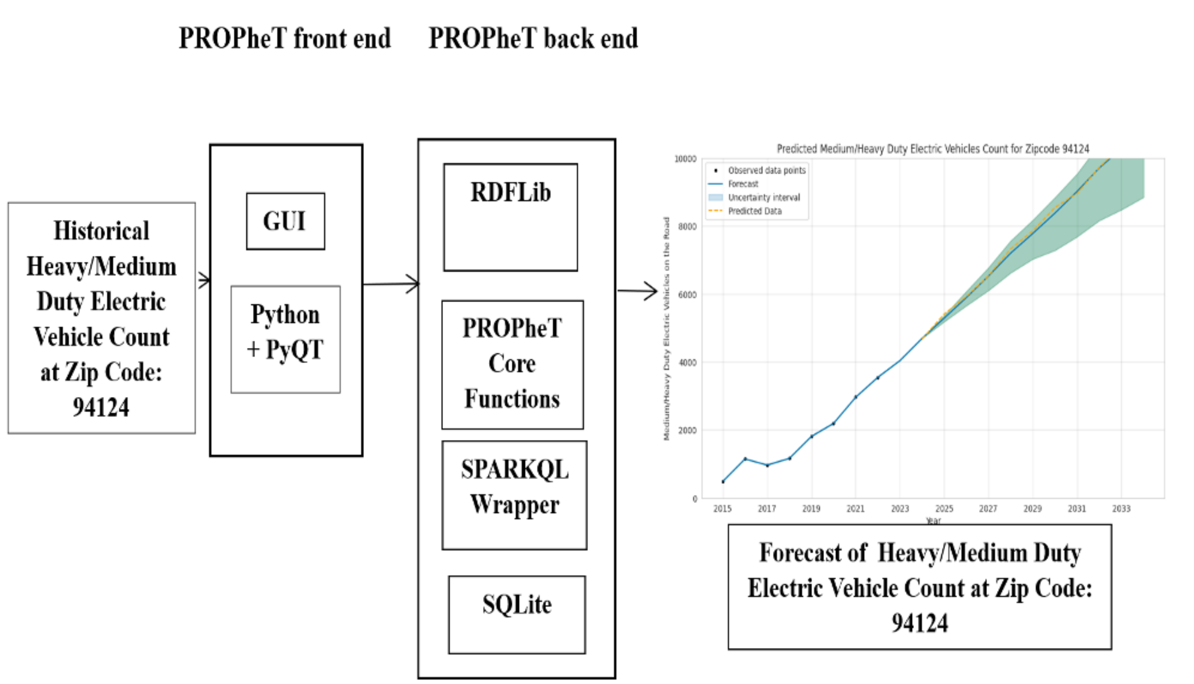


Figure 44. PROPhet Architecture is utilized for forecasting Electric Vehicle count at a Zip code.

driving experience for heavy- and medium-duty drivers. The resulting projections enable stakeholders to plan for sustainable transportation, deploy charge stations, and anticipate energy demand with confidence. This strategy supports the overarching objective of lowering greenhouse gas emissions and encouraging the use of battery-powered automobiles on a large scale to create a cleaner, more sustainable future.

4.1.2 Predicting Vehicle Range of Heavy-Duty and Medium-Duty EV

Predicting vehicle range based on charging data necessitates understanding how the vehicle's state of charge (SOC) varies over time and how this impacts the distance it can drive. Optimizing the operating efficiency of electric vehicles, such as transit buses, school buses, and delivery trucks, requires making range predictions based on charging data. In order to improve fleet management decisions, our research focuses on using charging data analytics to estimate the range of these three vehicle types.

The computation of the State of Charge (SOC) change over charging sessions, designated as "Delta_SOC," commences the research procedure. The range of the vehicle is mostly determined by the amount of energy contributed to the battery during charging, which is revealed by this measure. In addition, we calculate "Average_Power," which measures the rate of energy supply during charging, and "Charging_Duration," which represents the amount of time spent charging. The effectiveness of the charging process is greatly influenced by these characteristics. We use a straightforward method to estimate the range, where "Range" is expressed as a percentage of "Delta_SOC" in relation to the battery's total capacity. Range (in miles or kilometers) = Battery Capacity (in kWh) * Efficiency (in miles or kilometers per kWh) is a formula used to express the unique features of the battery system in use. Battery capacity, which is commonly expressed in kilowatt-hours (kWh), is the total amount of energy that can be stored in the electric vehicle's battery pack, while efficiency refers to how far the car can go on a single kWh of electricity.

We leverage an ensemble of machine learning techniques, such as Random Forest, Gradient Boosting, Linear Regression, Bagging Regressor, Stacking Regressor, and a Voting Regressor that integrates the results of all models, for predictive modeling. The objective is to develop a solid model that can correctly forecast vehicle ranges based on charging characteristics. The weighted fusion of these models' forecasts ensures that each model's output is taken into account under its unique performance. This method accounts for each model's advantages while minimizing its disadvantages, producing a more accurate prediction of vehicle range.

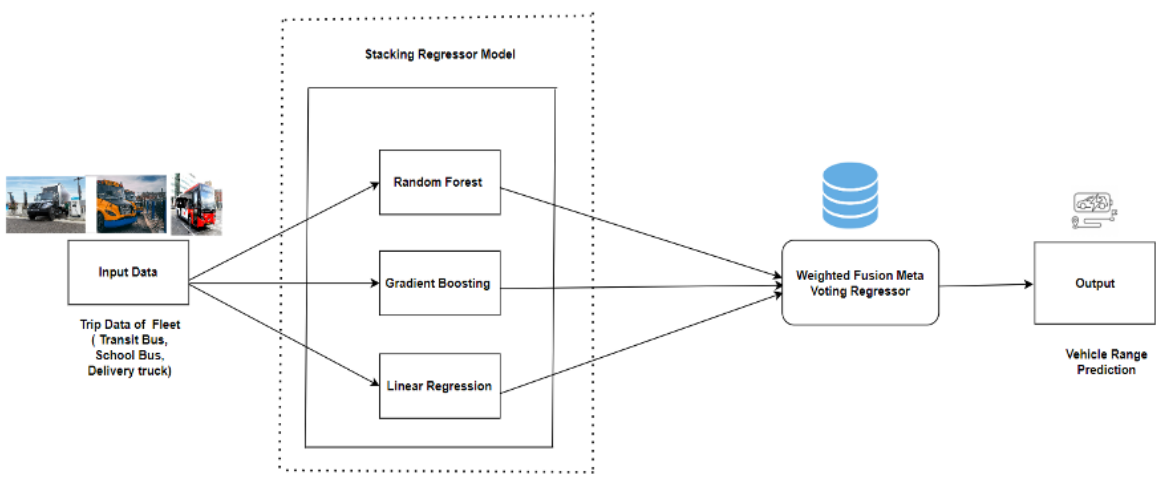


Figure 45. Model For Predicting Vehicle Range of Heavy Duty and Medium Duty EV

Complex correlations in the data are exceptionally well captured by Random Forest, a collection of decision trees. Another ensemble technique, gradient boosting, creates a model that improve sequentially and is very powerful for regression applications. A foundational model that offers simplicity and interpretability is provided by linear regression. We use ensemble approaches to increase forecasting accuracy. A type of bootstrap aggregation known as bagging uses the Random Forest model as its foundation estimator. Bagging lowers overfitting and boosts the stability of the model. In order to provide a more reliable final predictor, the Stacking Regressor combines predictions from various models (Random Forest, Gradient Boosting, and Linear Regression) with a meta-estimator (Linear Regression). This method minimizes the constraints of particular models while maximizing the strengths of several models. A Voting Regressor, which aggregates the results from all models while taking into account each one's performance individually, produces the final prediction. This strategy makes use of weighted fusion, allocating various weights to each model's predictions based on their historical reliability and accuracy. Weighted fusion makes sure that the most important models make a larger contribution to the final prediction, improving the accuracy of the final prediction as a whole.

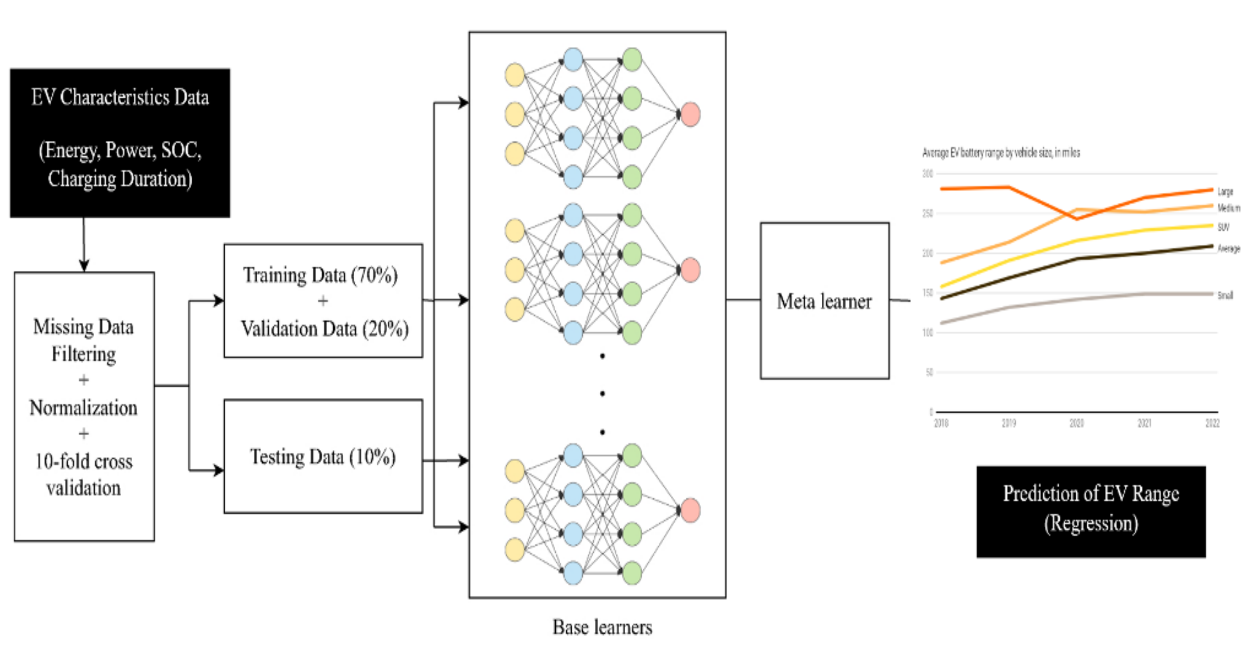


Figure 46. Stacked Regressor Model Architecture for Predicting EV Range

4.1.3 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV

EV energy demand forecasting is estimating how much energy will be required to charge the car to a specific SOC or estimating how much charging is still necessary to get a full charge. The efficient management of energy resources, the optimization of charging infrastructure, and the smooth operation of different vehicle fleets, such as transit buses, school buses, and delivery trucks, all depend on the ability to predict the energy consumption for electric vehicles (EVs). Both long-term and short-term projections are implemented in this use case to fulfill various planning and operational objectives.

The Temporal Fusion Transformer (TFT) model is a cutting-edge neural network architecture built for time series forecasting tasks. The TFT model employs self-attention mechanisms and multi-head attention to capture complicated temporal connections within sequential data, drawing inspiration from the Transformer architecture's success in natural language processing. This makes the model particularly effective for forecasting jobs that need knowledge of historical context since it enables it to recognize long-term patterns and linkages. The TFT model considers temporal elements like the day of the week, the month, and the year. These characteristics are crucial for identifying patterns and seasonality in time series data. The TFT model can produce more precise and context-aware forecasts by taking both the temporal context and the order of the data points into account. The TFT model's capacity to combine both temporal and static elements, offering a comprehensive perspective of the data, is one of its standout characteristics. The TFT model can produce more precise and context-aware predictions by adding extra contextual data, such as categorical variables or metadata linked to each time series. Additionally, the TFT model is flexible and capable of multi-horizon forecasting, which allows it to make simultaneous predictions of future values at various time periods. This is particularly helpful when making forecasts on different time horizons, such as when estimating both the short- and long-term energy consumption in the context of charging electric vehicles.

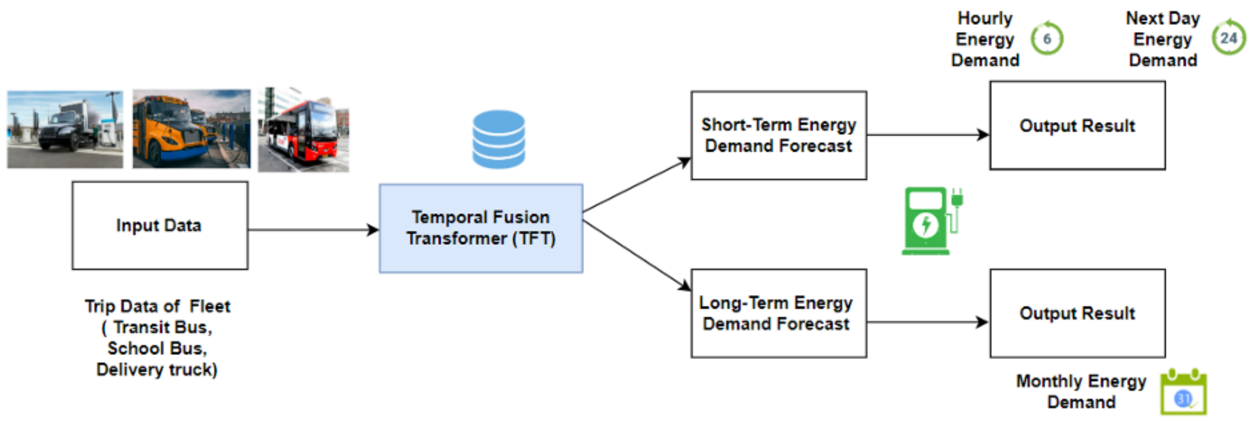


Figure 47. Model for Predicting Energy Demand of Heavy Duty and Medium Duty EV

4.1.3.1 Long-Term Energy Demand Forecasting

Long-term energy demand forecasting includes weekly, monthly, and yearly estimates, which provide useful information for strategic planning and resource allocation. A useful tool in this situation is the Temporal Fusion Transformer (TFT) concept. The TFT model can capture complicated linkages and multiple temporal characteristics of the data, which makes it useful for identifying trends and patterns in long-term energy consumption. For instance, the TFT model is capable of performing detailed analyses of previous data, such as charger ID, average power, starting SOC, ending SOC, and charging duration, to create precise forecasts regarding future energy use. We can develop a reliable long-term forecasting model by training the model on this historical data and optimizing it for MSE loss. The output of this model enables fleet managers and infrastructure planners to plan for fluctuations in energy demand and guarantee that there are enough charging resources available.

4.1.3.2 Short-Term Energy Demand Forecasting

Short-term energy demand forecasting, on the other hand, concentrates on estimating energy needs for the following day or within a shorter time frame, often in the order of next 6 hours. The use of this kind of forecasting in energy load balancing and real-time operational decisions is important. The TFT model uses its ability to record temporal dependencies and adapt to changing situations to be equally skilled at short-term forecasting. The model can forecast the amount of energy required for brief charging sessions by taking features like charger ID, average power, and SOC indicators as inputs. This forecast is essential for maximizing the use of energy, particularly in situations where many vehicles with various energy needs share charging infrastructure.

In a nutshell fleet managers and infrastructure planners are empowered by the combination of long-term and short-term energy demand predictions utilizing the TFT model to optimize the charging process for transit buses, school buses, and delivery trucks. By ensuring that the appropriate quantity of energy is accessible when and where it is most required, these forecasts will improve operational effectiveness, eliminate energy waste, and support sustainable transportation practices.

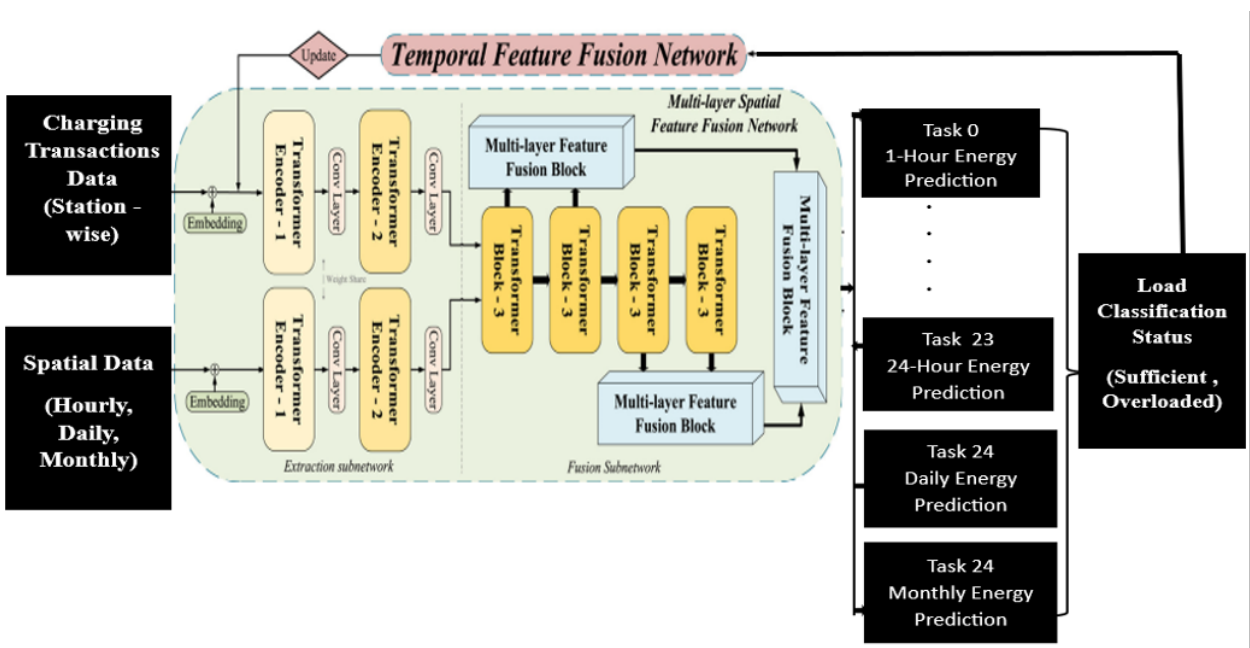


Figure 48. Temporal Fusion Transformer Architecture for Energy Demand Prediction at each Station

4.1.4 Optimal Placement of New Charging Stations for each vehicle type

Our research focuses on finding a solution to the crucial problem of strategically positioning new charging stations to meet the charging requirements of various vehicle types in the context of promoting the adoption of electrically powered automobiles (EVs) and enhancing sustainable transportation practices. To accomplish this, we built an optimization model that considers Points of Interest (POI), the proximity of current charging stations, and location effectiveness in addition to cost-effectiveness. The main goal is to reduce installation costs while making sure that EV users have easy access to the infrastructure for charging, considering their preferences for adjacent amenities. Our approach utilizes the capabilities of the PuLP library and linear programming to create a model of optimization for the best location of charging stations. To assist in guiding the decision-making process, this model considers several important characteristics and limitations.

We construct a pair of possible variables. Whether a charging station is put up for a specific vehicle type at a particular location is represented by the first set, marked by the symbol x. The second set, y, indicates if a site is chosen for the installation of a charging station. Important factors include installation costs, demand for charging per vehicle type, the highest possible number of charging stations that can be built (N) , and the greatest distance that a vehicle will travel to reach a charging station (R) as well as details on local POIs and available charging stations. Our objective function, which now considers installation costs as well as user-friendliness and location effectiveness, aims to reduce the overall cost of establishing new charging stations. It includes the price of installation for every charging station, accounting for consumers' preferences for adjacent conveniences and their closeness to already-existing charging stations. By focusing on the needs and convenience of EV users, the location is guaranteed to be appropriate. The following is a definition of the objective function as defined in equation (1):

Minimize: Total Cost = Σ [(Installation Cost at Location + PoI + Proximity to Existing Stations) * x[Location, Vehicle Type]] (1)

Installation Cost at Location: The cost of installing a charging station for a particular vehicle type. *User Preference*: A factor that represents user preferences for nearby amenities (POI) at the location. *Proximity to Existing Stations:* A factor representing the proximity of existing charging stations to the location. *x [Location, Vehicle Type]:* A binary variable that indicates whether a charging station is installed at a specific location for a particular vehicle type.

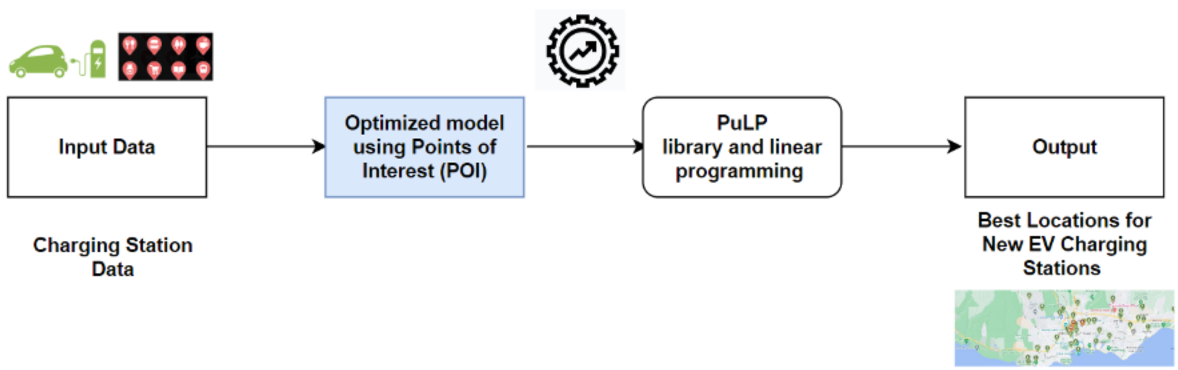


Figure 49. Model for Optimized Locations for New EV Charging Stations

Several limitations are incorporated to ensure a complete and user-friendly solution. Budgetary restrictions, coverage specifications, and a cap on the overall number of charging stations are included in these constraints. Additionally, we include restrictions that promote choosing a location based on POI proximity and an efficient charging station distribution. After resolving the optimization model, we obtain the best location for charging stations for various car kinds. This positioning maximizes both location efficacy and user-friendliness while minimizing installation expenses. While taking into account their preferences for local amenities, it makes sure that EV customers have simple access to charging infrastructure. The following equations collectively define the constraints considered for our optimization problem:

Budget Constraint: Ensure that the overall cost of installing the charging stations does not go beyond the allotted budget.

Σ [(Installation Cost at Location) * x[Location, Vehicle Type]] $\leq$ Budget (2)

Coverage Constraint: Ensure that each type of vehicle has access to at least one charging station within the range of their maximum travel.

Σ [x[Location, Vehicle Type]] $\geq$ 1 for all Vehicle Types (3)

Number of Charging Stations Constraint: The total number of charging stations that may be installed must not be greater than N

Σ[y[Location]] $\leq$ N (4)

Charging Station Location Decision Constraint: Make sure that if a place is chosen (y[place] = 1), at least one charging station for each type of car is installed there. This is the charging station location decision constraint.

Σ [x[Location, Vehicle Type]] $\geq$ y[Location] for all Locations (5)

Our improved optimization approach delivers a comprehensive method for strategically placing EV charging infrastructure. It encourages the use of EVs, makes it easier for people to use charging stations, and supports sustainable mobility practices by taking into account a variety of criteria, such as installation costs, user preferences for local facilities, and the distribution of existing charging stations. This user-centered and location-aware strategy makes sure, that the placement of the charging stations is not only economical but also improves the entire EV user experience.

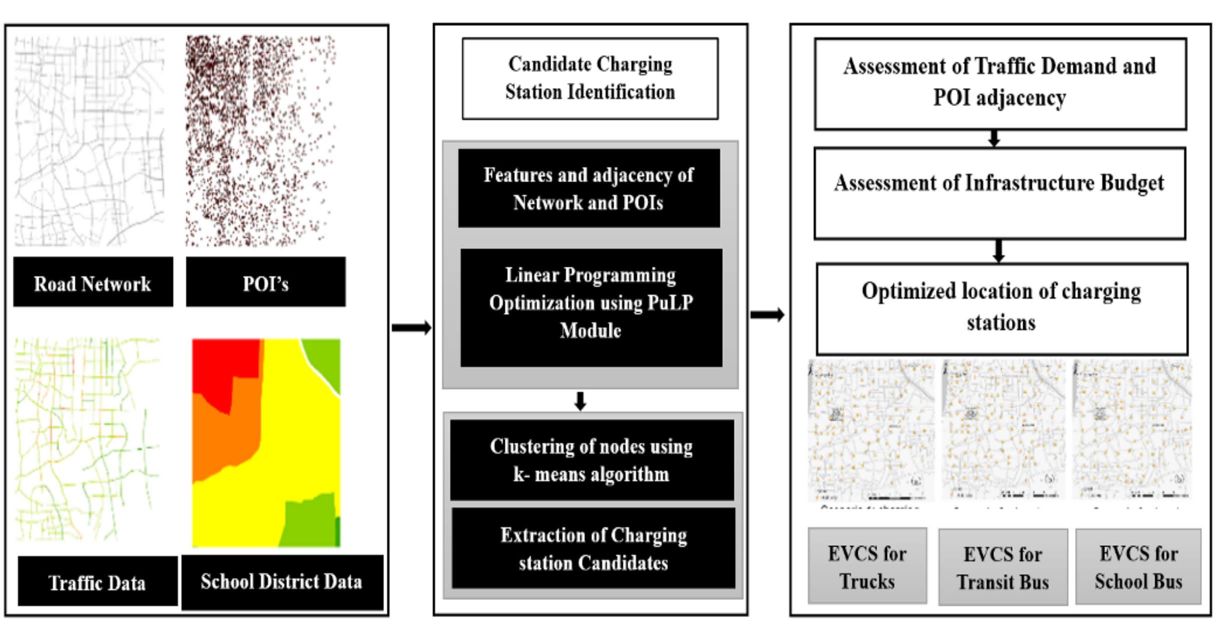


Figure 50. Model Architecture to Suggest Optimized Location for New EVCS

4.2 Model Supports

4.2.1 Environment, Platform, Tools

A stable environment, appropriate platforms, and specific tools are necessary for the efficient execution of this integrated project, which includes forecasting EV and charging station demand, projecting vehicle range, forecasting energy use, and optimizing charging station placement. An overview of the essential elements is provided below.

The execution of various use cases within a single workflow, which includes forecasting EV and charging station demand, projecting vehicle range, short and long-term energy need, and optimizing charging station placement, necessitates a comprehensive environment, platform, and set of tools. The foundation is constructed using the Python programming language and its vast libraries for machine learning, optimization, and data analysis. The interactive development and documentation platform is Jupyter Notebook. The set of tools includes TensorFlow or PyTorch deep learning frameworks for Temporal Fusion Transformer (TFT) model implementation, Scikit-Learn for regression models and ensemble approaches, and Facebook's Prophet for demand forecasting. The PuLP package makes it easier to create linear programming models for charging station placement. Pandas, NumPy, and data visualization tools like Matplotlib and Seaborn are used for data management and analysis. Version control and collaboration are aided by Git/GitHub, while task organization is facilitated by project management software like Jira. With a particular emphasis on AWS SageMaker for machine learning and data analysis, the entire workflow is smoothly organized within the AWS environment. AWS S3 is used to store and manage historical data and Facebook Prophet is used to forecast time series data. SageMaker is a flexible platform for creating, honing, and deploying machine learning models, including the Temporal Fusion Transformer (TFT) model application for energy demand forecasting and vehicle range prediction. Data retrieval and model execution are automated by AWS Lambda, computational operations are hosted by EC2, optimization data is managed by DynamoDB, and core AWS components are monitored by CloudWatch. AWS Glue and Athena are utilized for data transformation, and Tableau is used for data visualization.

Table 24. Hardware Requirements

Resource	Minimum Configuration
CPU	N1-standard-16 vCPU 5.6GHz, 32Cores
GPU	NVIDIA GTX 1050 or equivalent
RAM	128 GB DDR4-3200 MHz
Storage	Amazon S3 (12 GB)

Git is used as the version control system for collaboration and code management. Docker is the containerization tool used to package the application into a container that can run on any

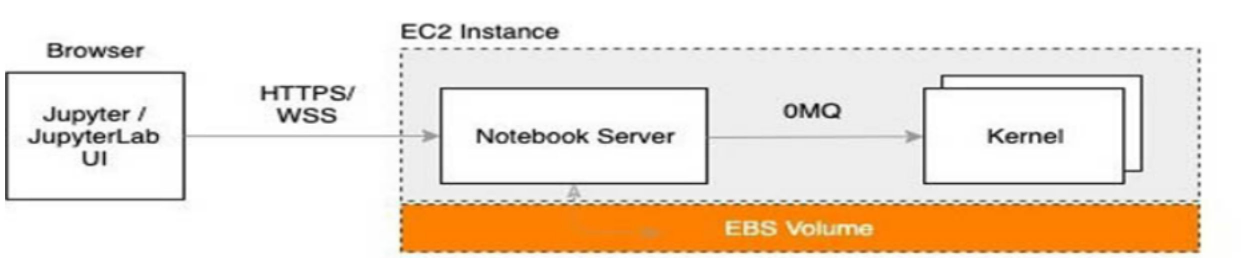


Figure 51. High-Level Architecture of Jupyter

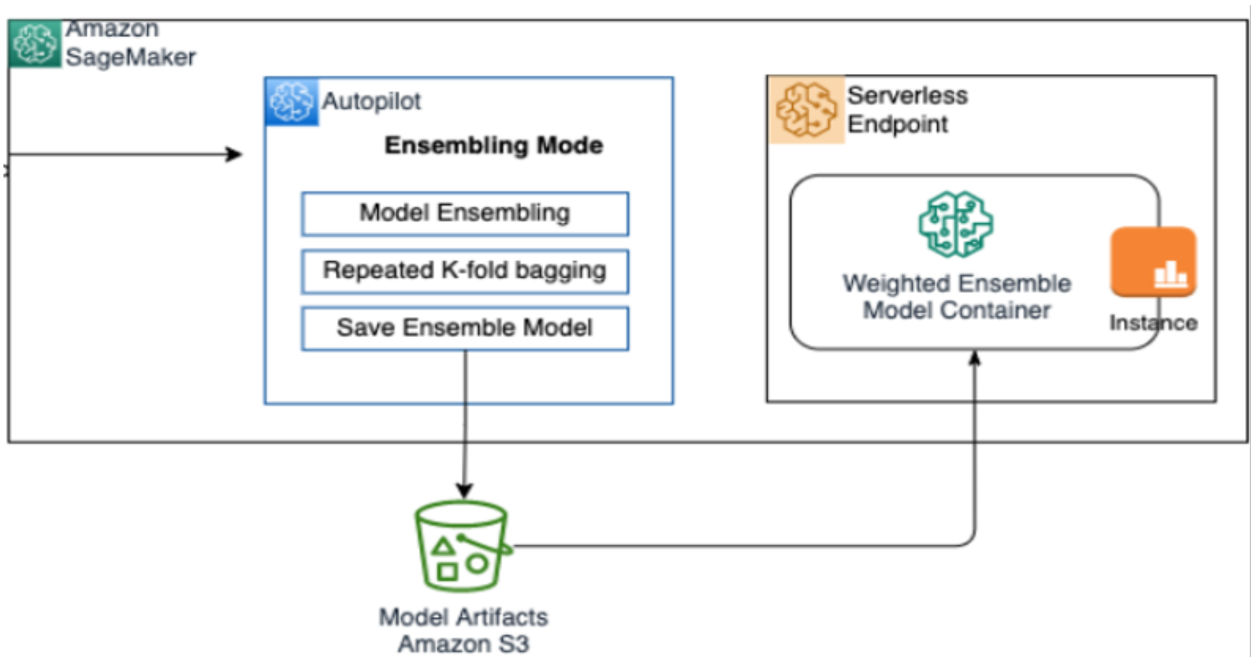


Figure 52. Ensemble Modeling using Sage Maker Auto Pilot

system without any modification. Flask is the web framework used for building RESTful APIs that can be easily integrated into the platform. Swagger UI is the API development tool used for documentation and testing of the RESTful APIs. This combination of tools and platforms provides a robust and scalable platform for building and deploying machine learning models for data-driven analysis of EV charging infrastructure. These tools and platforms provide an integrated development environment that streamlines the development process, allowing developers to focus on the modeling and analysis rather than worrying about infrastructure and deployment. Using specialized features, this platform facilitates the development, testing, and deployment of supervised regression machine learning models. The specifics of the software configurations utilized in this research are listed below.

Table 25. Software Requirements

Resource	Minimum Configuration
Operating System	Red Hat Enterprise Linux (Amazon EC2)
Programming Language	Python 3.9
Integrated Development Environment (IDE)	JupyterLab v3.6.0a4
Virtual Environment Tool	Anaconda 4.10
Cloud Platform	AWS Sage Maker
Version Control System	Git 2.32
Containerization Tool	Docker 20.10.8
Web Framework	Flask 2.0
API Development	Swagger UI

With the aid of all the packages and modules that Python 3.9 has to offer, this project can write and develop the model swiftly. It will be able to analyze, analyze, and visualize the data properly thanks to the Pandas, NumPy, and matplotlib libraries. Modeling is done using the Scikit-Learn package by importing several techniques. Data processing and numerical computations including EV charging transaction data and point-of-interest data would require the use of Pandas and NumPy. For data visualization and exploratory analysis, Matplotlib and Seaborn would be necessary. Building and training machine learning and deep learning models to forecast the best locations for EV charging stations based on historical data of charging transactions and accessibility to adjacent POIs will require the use of Scikit-Learn and TensorFlow. The geographical analysis and mapping of EV charging stations and POIs would also require GeoPandas. The development team would also be able to produce interactive maps and visualizations of EV charging stations and POIs thanks to Folium. These libraries give us access to a wide range of tools and features that enable us to conduct in-depth analyses and come to wise judgments about where to locate EV charging stations. Table 3 gives a thorough overview of the Python libraries used for modeling in this study as well as their function.

Table 26. Python Library Requirements

Library	Method	Purpose
Pandas	info, shape, sample, head, drop, merge	Before modeling, reading training data into a dataframe, manipulating it, and running statistics over it.
NumPy	array, math, ravel, random	Manipulating and processing numerical data, such as distance and time calculations
Garbage Collect	GC collect	Clearing the memory
timeit	default_timer	Calculating Computation time of various regression algorithms.
Matplotlib	pyplot. Plot	Creating visualizations such as scatter plots, line plots, and bar charts to analyze the distribution of charging transactions and station locations
Seaborn	scatterplot, histplot	Creating more advanced and informative visualizations such as heat maps, regression plots, and distribution plots

Table 26. *Cont.*

Library	Method	Purpose
Geopy	distance	Calculating distances between charging station locations and points of interest for further analysis
GeoPandas	read_file, sjoin	Working with geospatial data such as charging station and point of interest locations, and performing spatial joins and operations
Folium	Map, Marker, Choropleth	Creating interactive maps of charging station locations and points of interest for visual analysis
Scikit-Learn	train_test_split, GridSearchCV	Splitting the data into training and testing sets, and tuning hyperparameters to build accurate and robust machine learning models for predicting optimal charging station locations
	RandomForestRegressor, GradientBoostingRegressor, BaggingRegressor, StackingRegressor, VotingRegressor, Kmeans,	Prophet Building machine learning models for more complex data analysis tasks of EV charging stations including Range, Demand and Energy prediction and suggesting optimized locations for new stations.

4.2.2 Platform Architecture and Data Flow

The diligently designed system that optimizes the processing of data and the completion of complicated activities is demonstrated by the platform architecture and data flow that underlie the comprehensive use cases. It starts with data ingestion, a crucial phase in which historical information on the demand for EVs and charging stations is routinely and systematically ingested into specific AWS S3 buckets. AWS Glue is used to seamlessly automate complex data transformation and cleansing procedures, enhancing data quality and consistency. These pre-processing steps make sure that the data has been meticulously cleaned up and geared up, thereby rendering it appropriate for subsequent research and modeling initiatives.

Table 27. Model Support

Model	Platform	Environment	Tools
EV Count Demand Forecasting	JupyterLab	Python 3.x with GPU Compute	Pandas, Numpy, Prophet
EV Range Prediction	JupyterLab	Python 3.x with GPU Compute	math, sklearn,matplotlib
EV Energy Demand Classification	JupyterLab	Python 3.x with GPU Compute	sklearn.ensemble, sklearn.neural_network transformers
Optimal Location Recommendation	JupyterLab, AWS Sage Maker	Python 3.x with GPU Compute	geopy, folium, pandas.io.json, sklearn.neighbors , geopandas ,sklearn.clusters , PuLP, sklearn.base

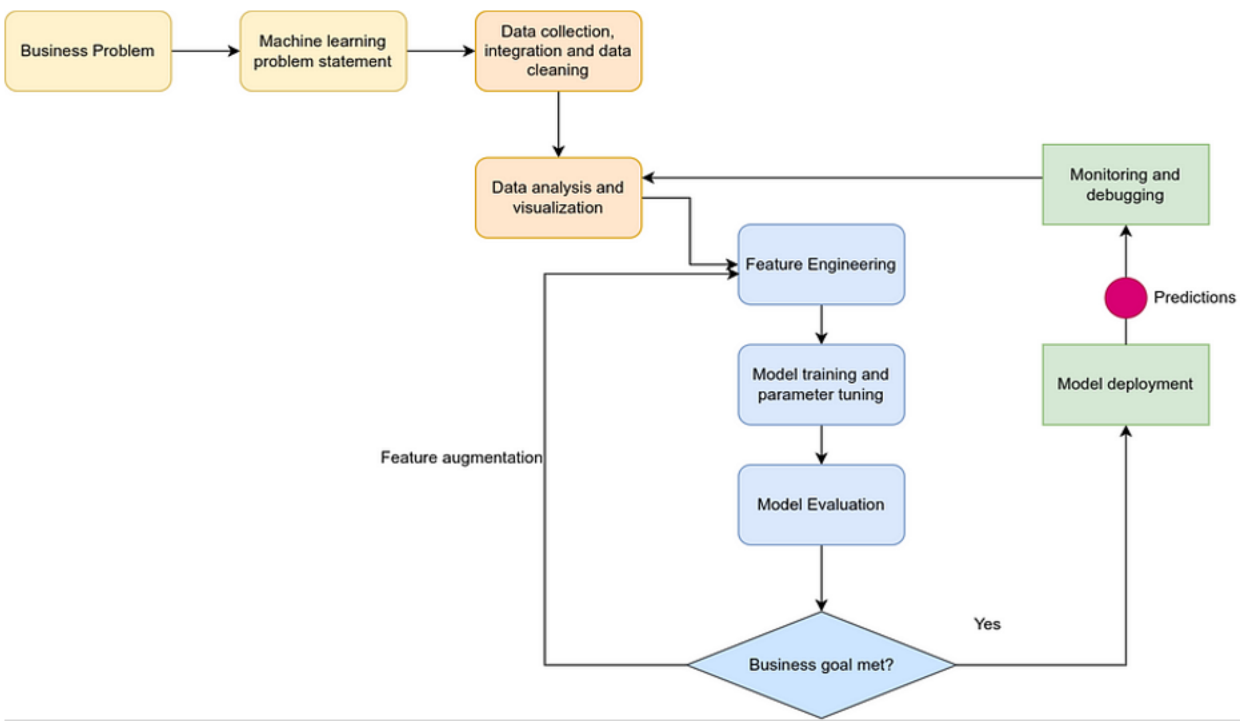


Figure 53. Model Platform Architecture

Following that, data analysis and modeling operations take center stage, all while being orchestrated within the limitations of AWS SageMaker Jupyter Notebooks. For data scientists, this setting acts as a dynamic center, giving them the freedom to work together to explore data, carry out complex preprocessing, and execute in-depth analyses of the data. SageMaker develops into a flexible platform for the comprehensive development, training, and rigorous evaluation of machine learning models in addition to its data analysis capabilities. It includes the creation and testing of models for estimating a vehicle's range as well as the application of the Temporal Fusion Transformer (TFT) model, an advanced deep learning technique designed specifically for the accurate forecasting of both short- and long-term energy requirements. It's noteworthy that the architecture includes a feature that optimizes the placement of new charging stations. This feature makes use of AWS Lambda functions, which are skilled at coordinating automatic data retrieval and swiftly and accurately performing the optimization model. On AWS EC2 instances, the optimization model that supports the data-driven decision-making process is housed. Additionally, AWS DynamoDB's meticulous organization and storage of the optimization data, which is essential for the model's effectiveness, ensures smooth and efficient data flow dynamics.

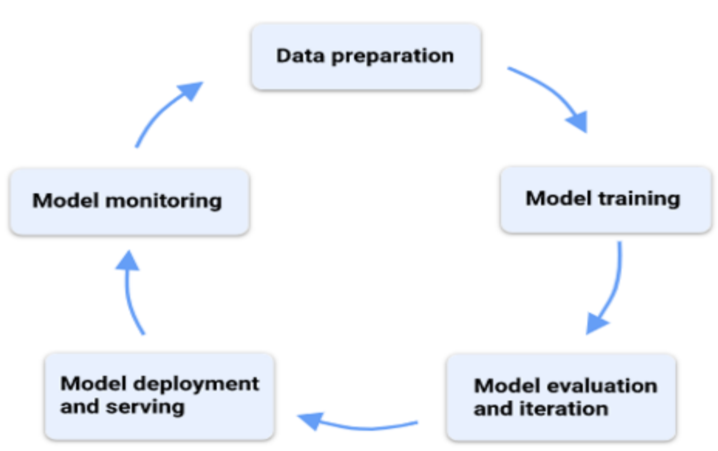


Figure 54. Machine Learning Data Flow

Additionally, the platform architecture creates a special path just for reporting and data insights. A powerful analytics and visualization tool called Tableau enters the fray by expertly creating interactive dashboards and visualizations. Data analysts and stakeholders alike can gain deeper insights and make wiser decisions thanks to these artifacts. Jupyter Notebooks simultaneously play a crucial role, acting as the main method for coding documentation and creating in-depth project reports. The traceability and reproducibility of analyses, which are crucial in data-driven projects, are ensured by this documentation component. The workflow culminates with the deployment and scaling aspects. The smooth deployment of applications and models, which meets the demand for flexibility and scalability in real-world settings, is made possible by AWS Elastic Beanstalk and Amazon ECS. AWS Auto Scaling is particularly notable for being seamlessly integrated to dynamically distribute and manage resources, adjusting to the changing workloads with accuracy and efficiency.

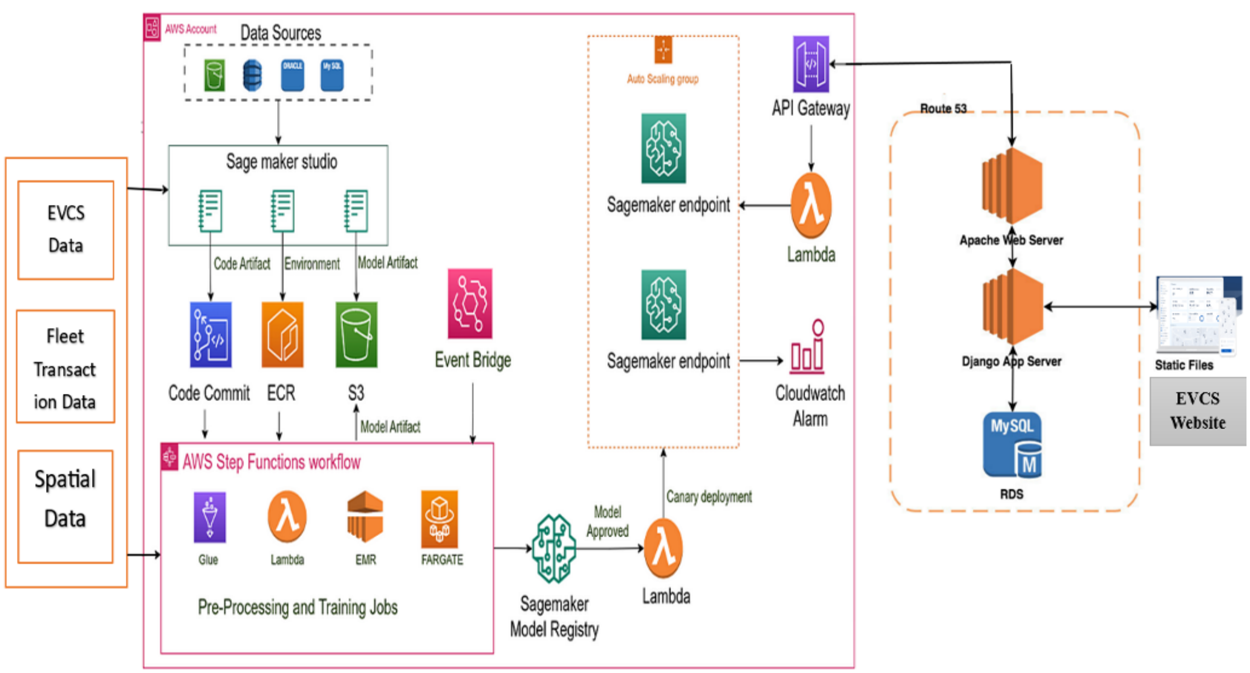


Figure 55. Proposed Platform Architecture Components

In essence, the platform architecture and data flow comprise a finely woven tapestry of AWS services, which have been painstakingly created to give a cogent, scalable, and effective solution for the whole range of use cases. It makes use of SageMaker's flexible capabilities to enable the seamless integration of reporting, optimization, and machine learning. The end result is a solid platform that not only handles the challenging tasks at hand but also guarantees the flexibility to develop and adapt to shifting demands in the field of managing EVs and charging stations.

4.3 Model Comparison and Justification

For each project component and module, we must analyze a number of models. As mentioned earlier, there are: Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand, Predicting Vehicle Range of Heavy Duty and Medium Duty EV, Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV, Optimal Placement of New Charging Stations for each vehicle type. Model comparison and justification are crucial in determining the best approaches for the four use cases. Let's examine each use case in greater detail.

A detailed study of numerous models has been conducted in the field of EV and charging station demand forecasting to determine their viability for this challenging task. ARIMA, Long Short-Term Memory (LSTM), and Seasonal Autoregressive Integrated Moving-Average (SARIMA) have all been taken into consideration in addition to the Prophet model, which has been selected as the final model. The Prophet model stands out for its effective management of multiplicative seasonality, flexibility around changepoints, and innate capacity to accurately capture seasonality, holidays, and special events. It is an appealing option for predicting the demand for EVs and charging stations due to its adaptability in handling varied data patterns. Historical information on the demand for EVs and charging stations has been used in an empirical study to examine the performance of various models. We have carefully evaluated the ability of each model to reflect seasonality, short-term volatility, and long-term trends. Their ability to adapt to different data patterns has also received attention. The choice of the Prophet model fits with the intricacy of the forecasting assignment and the intrinsic complexity of the data. The results of these models are outlined in the table below, which highlights their advantages and disadvantages in predicting EV and charging station demand.

Table 28. Model Comparison for EV Demand Forecasting

Model	Strengths	Limitations	Performance
Prophet	Seasonality and special events are well captured	It may be difficult to handle complex data patterns	Excellent
ARIMA	Effective for short-term fluctuations	Capacity to handle complex seasonality is limited	Moderate
LSTM	Captures long-range dependencies effectively	A substantial amount of data and tuning is required	Good
SARIMA	Effectively handles strong seasonal patterns	Less seasonal data may not perform well	Good

Our selection of models within our ensemble for the task of forecasting vehicle range for Heavy Duty and Medium Duty Electric Vehicles (EVs) is justified by their unique capabilities and the requirements of the prediction task. This ensemble of models was chosen because of its comprehensive approach to prediction. We want to maximize the power of these algorithms by mixing models of different strengths. By considering both linear and nonlinear patterns in the data, the ensemble is built to ensure reliable predictions. With this method, the ensemble is better able to respond to a variety of conditions, making it a well-rounded option for estimating the vehicle range for heavy-duty and medium-duty EVs. Based on the relative effectiveness and reliability of each model within the ensemble, weighted regression includes giving each model's predictions a distinct weight. We can provide more weight to models that exhibit superior forecast accuracy and consistency since the weights indicate our trust in each model's predictions. By multiplying each model's forecast by the set weight and then adding the weighted predictions, the weighted fusion of predictions is calculated. It can be modeled mathematically as follows:

Weighted Fusion = (Weight_Model1 * Prediction_Model1) + (Weight_Model2 * Prediction_Model2) + ... + (Weight_ModelN * Prediction_ModelN)

This weighted fusion, which represents a consensus opinion and makes use of the ensemble's collective predictive power while minimizing the effects of individual models' shortcomings, is the ultimate forecast for vehicle range. In conclusion, the strategic decision to include weighted regression in the group of models for estimating vehicle range is intended to maximize the prediction accuracy and resilience. The ensemble develops a well-balanced and flexible predictive system that can successfully address the challenges of vehicle range prediction for Heavy Duty and Medium Duty EVs by allocating weights to models based on their performance. A table that compares the various Voting Regressor versions and explains why the Weighted Fusion method was chosen is mentioned below.

A comprehensive review of numerous models has been done in the area of projecting the short- and long-term energy consumption for Heavy Duty and Medium Duty Electric Vehicles (EVs) in order to determine their viability for this challenging task. Additional models, such as Seasonal Decomposition of Time Series (STL) and Long Short-Term Memory (LSTM), have been taken into consideration in addition to the initial Temporal Fusion Transformer (TFT) model. Because of its innate capacity to successfully handle complicated time series data, capture temporal dependencies, and allow both short- and long-term forecasting horizons, the TFT model has been chosen as the basic model. Its aptitude for capturing complex nonlinear interactions and adaptability to different data patterns fit well with the multifarious nature of energy demand prediction.

While the LSTM and STL models thrive in particular areas, such as capturing long-term dependencies and significant seasonality, the TFT model excels in giving a comprehensive approach to energy demand forecasting in diverse scenarios. The following table highlights the effectiveness of various models in projecting the short- and long-term energy requirements for Heavy Duty and Medium Duty EVs while also noting their individual strengths and weaknesses.

Table 29. Comparison of various Voting Regressor versions for EV Range Predictions

Voting Regressor Variant	Strengths	Limitations	Justification for Weighted Fusion
Simple Voting Regressor	The majority vote principle is simple to implement	Ignores model performance differences	The optimization of performance is limited
Bagging Regressor	Reduces model variance and is robust to overfitting	Focuses on a single base model	Model diversity is limited
Stacking Regressor	Diverse models are combined, nuances are captured	Optimization requires a meta-learner	Diversity of models enhanced
Weighted Fusion	Model performance is incorporated as a weight.	Adaptable to variations and strengths of models.	The performance of the ensemble has been optimized

In order to solve this challenging optimization problem, a thorough comparison of numerous models has been conducted in the context of locating new charging stations for various vehicle types. Other complex methods, such Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO), have been investigated and added into the analysis in addition to the initial linear programming model. The linear programming model, which makes use of the PuLP

Table 30. Comparison of models in projecting the short- and long-term energy requirements

Model	Strengths	Limitations	Performance (Short- Term)	Performance (Long-Term)
TFT Model	Effectively captures temporal dependencies	Adaptable to various patterns of data	Excellent	Excellent
LSTM	Ensures long-range dependencies are accurately modeled	Fine-tuning and substantial data are required	Good	Good
STL	Effectively handles strong seasonal components	Data with weak seasonality may not perform well	Moderate	Moderate

library's features and linear programming approaches serve as the fundamental methodology. To find the best locations for charging stations, it formulates the optimization problem efficiently, considering variables like charging demand, station capacity, and site constraints. In this comparison study, Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) stand out as formidable competitors, showcasing their effectiveness in determining the best locations for charging stations for various kinds of heavy and medium-duty vehicles. While the Linear Programming model provides a strong foundation for optimization, Gas, and PSO offer a sophisticated and adaptive approach that effectively considers a variety of parameters to discover the best placements for charging stations.

Table 31. Model Comparison and Justification

Use Case	Model	Features	Approach	Strengths	Justification	Limitations
Count Forecast of Heavy/Medium Duty EV Vehicle	Prophet	Designed for time series forecasting	Additive model with components for trend, seasonality, and holidays	Handles strong seasonality; robust to outliers and missing data	Chosen for its effectiveness in capturing temporal patterns	May struggle with irregular or changing seasonality patterns; manual tuning is needed.

Table 31. *Cont.*

Use Case	Model	Features	Approach	Strengths	Justification	Limitations
Count Forecast of Heavy/Medium Duty EV Vehicle	ARIMA	Statistical time series model	Autoregressive, Integrated, Moving Average components	Handles linear relationships; captures seasonality	Common alternative for time series forecasting	Sensitive to model assumptions; may not handle complex non-linear patterns well
	SARIMA	Seasonal ARIMA model	Extension of ARIMA with seasonal components	Suitable for data with clear seasonality patterns	Effective for datasets with clear seasonal patterns	Similar limitations as ARIMA; assumes stationarity and linearity in data.
Range Prediction	Stacking Ensemble Regressor (Weighted Fusion)	Combines predictions from multiple models; Weighted Fusion for model combination	Ensemble approach for more accurate predictions	Leverages strengths of various models; Weighted Fusion adapts to individual model strengths	Chosen for accuracy improvement by combining diverse models	Complexity in implementation; sensitivity to base model choice
	Gradient Boosting	Sequentially builds a series of weak learners	Each learner corrects errors of the previous one	Robust to outliers; adapts well to complex relationships	Effective for improving accuracy with ensemble methods	Sensitive to noisy data; requires careful tuning of hyperparameters
	Neural Networks	Deep learning model with interconnected layers	Sequential processing of data through multiple layers	Captures complex non-linear relationships	Suitable for capturing intricate patterns in data	Requires substantial computational resources; potential for overfitting

Table 31. *Cont.*

Use Case	Model	Features	Approach	Strengths	Justification	Limitations
EV Charging Station Energy Demand Prediction	Temporal Fusion Transformer Model	Specifically designed for time series forecasting	Transformer architecture for capturing complex temporal patterns	Effective in capturing long-range dependencies; handles multiple seasonalities	Chosen for advanced architecture tailored for time series forecasting	Requires substantial computational resources; challenging interpretability
	LSTM	Recurrent Neural Network (RNN) architecture	Sequential processing of input data over time	Captures long-range dependencies; suitable for sequential data	Powerful for complex time series patterns; learns non-linear relationships	Sensitive to noisy data; requires careful tuning of hyperparameters
	GRU	Gated Recurrent Unit (RNN) architecture	Simplified version of LSTM with similar functionality	Efficient for capturing dependencies in sequential data	Faster training compared to LSTM; suitable for simpler tasks	May not perform as well as LSTM on datasets with complex dependencies
Optimal Location Prediction for New Charging Station	PuLP Linear Programming Optimization with K-means Clustering	Linear programming for optimization; K-means for spatial pattern recognition	Integrates linear programming with clustering for optimal location determination	Linear programming provides systematic optimization; K-means identifies spatial patterns	Chosen for mathematically optimized solutions based on spatial clustering	Assumes optimal locations can be determined purely through mathematical optimization; sensitivity to data quality and clustering algorithm
	Genetic Algorithms	Evolutionary optimization algorithm	Iteratively evolves potential solutions	Effective for searching large solution spaces	Suitable for global optimization; handles non-linear objective functions	Requires careful tuning of parameters; may not guarantee global optimum

Table 31. *Cont.*

Use Case	Model	Features	Approach	Strengths	Justification	Limitations
Optimal Location Prediction for New Charging Station	Particle Swarm Optimizatio n	Swarm-based optimizat ion algorithm	Particles move through solution space iteratively	Efficient for optimization in continuous search spaces	Simplicity and ease of implementati on; parallelizable	Sensitive to choose of hyperparameters; may converge to local optima.

Table 32. Comparison of Proposed Models

Characteristic	Prophet	Stacking Ensemble Regressor	Temporal Fusion Transformer	PuLP Linear Programming with K-means Clustering
Architecture	Additive Time Series	Ensemble of Regressors	Transformer	Linear Programming + K-means Clustering
Dataset	Time Series	General	Time Series	Spatial and Temporal Data
Data Type	Continuous and Discrete	Continuous	Continuous and Categorical	Continuous and Discrete
Real-time	No	Yes	No	No
Targeted Problems	Time Series Forecasting	Regression	Time Series Forecasting	Location Optimization
Approaches	Time Series Decomposition	Ensemble Learning	Sequence-to-Sequence Model	Linear Programming + Clustering
Training Speed	Fast	Moderate to Slow	Moderate	Fast

Table 32. *Cont.*

Characteristic	Prophet	Stacking Ensemble Regressor	Temporal Fusion Transformer	PuLP Linear Programming with K-means Clustering
Training Time	Fast	Moderate to Long	Long	Fast
Accuracy	Good	High	High	Highly Dependent on Model and Clustering Quality
Preprocessing required	Minimal	Moderate	Moderate	Moderate
Space Complexity	Low	Moderate	High	Low
Computational Complexity	Low	Moderate to High	High	Moderate to low
Strengths	Seasonality Patterns	Ensemble Synergy	Temporal Dependencies	Spatial Optimization, Linear Programming
Known Issues	May Struggle with Irregular Seasonality Patterns	Sensitive to Base Models	Computational Resource Intensive	Sensitivity to Input Data and Clustering Algorithm

4.4 Model Evaluation Methods

To fit models, training data will be needed. Model comparison will be done using validation data. Following model selection, testing data will be used to confirm findings and create the final reportable metrics. A significant step in machine learning and predictive modeling is model evaluation. It enables us to evaluate the effectiveness, dependability, and generalizability of models, resulting in better decision-making. Numerous intelligent applications are built on it, and the ongoing evaluation of the effectiveness of machine learning models is crucial to their success. The performance of a model depends on several variables, including the algorithm used and how well it was trained. While choosing a model, metrics for model evaluation are used to compare models, evaluate how well a model fits the data, and forecast how accurate forecasts will turn out using a certain model and set of data. Model evaluation is the practice of employing several evaluation measures to comprehend the performance strengths and weaknesses of a machine learning model.

Simply training a model using a problem-specific training machine learning algorithm does not ensure that the resulting model fully captures the underlying concept concealed in the training data or that the best parameter values were chosen for the model training. If a model's performance isn't tested, it might be released on the actual system with unreliable predictions. It's uncertain to select a model out of the many accessible alternatives based solely on intuition. In the early stages of research, it is crucial to evaluate a model's effectiveness. The model evaluation also aids with model monitoring.

There are various methods for evaluating models, some of which are used in our analysis. The below-mentioned evaluation methods are used for this project to Forecast Heavy-duty and Medium Duty EV and Charging Stations Demand, and in Predicting Vehicle Range of Heavy Duty and Medium Duty EV, and in Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV.

Root Mean Squared Error

The Root Mean Squared Error (RMSE) is a performance metric that measures the average magnitude of the errors made by a predictive model. Like the MAE, it considers the differences between predicted and actual values, but RMSE calculates these differences by squaring them first, averaging them, and then taking the square root. Essentially, it measures the standard deviation of the differences between predicted and actual values, with higher values indicating a greater average error magnitude. RMSE is a useful metric to employ in case of significant errors. To use RMSE, if the model over- or underpredicted a few points in the prediction (since the residual will be square, resulting in a significant error). The RMSE is a preferred evaluation tool for regression issues since it not only determines the average distance between the forecast and the actual value but also highlights the impact of significant errors. The RMSE outcome will be impacted by significant mistakes. Its formula is given below.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Figure 56. Root mean squared equation.

Mean Absolute Error

Mean Absolute Error (MAE) is a metric used to evaluate the performance of predictive models. Regardless of their direction, MAE determines the average magnitude of errors a model makes when predicting a set of data. Specifically, MAE is calculated as the average of the absolute differences between the predicted values and the actual values, with each difference receiving equal weight in the calculation. MAE is used to determine the model's average absolute distance when making a forecast. In other words, it is interesting in how closely the forecasts on average match the actual model. The low MAE values signify accurate prediction from the model. Larger MAE values show that the model performs poorly in terms of prediction. MAE is calculated by the given formula.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Figure 57. Mean absolute error equation.

Mean Squared Error

A measure of an estimator's mean squared error (MSE), or the average squared difference between the estimated values and the actual value, is the average of the squares of the errors. The expected value of the squared error loss is represented by a risk function called MSE. Randomness or the estimator's failure to take into consideration data that could lead to a more accurate estimate is the reason why MSE is nearly always strictly positive (and not zero). While a smaller MSE suggests the opposite, a bigger MSE shows that the data points are widely scattered about the central moment (mean). Because a smaller MSE suggests that your data points are distributed tightly around the center moment (mean), it is preferable. In addition to not being skewed and reflecting the center distribution of your data values, it also has less errors as determined by how far apart the data points are from the mean than the original.

The MSE is calculated by the given formula.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Figure 58. Mean squared error equation.

R2 Score

The R2 score, also known as the coefficient of determination, is a statistical measure used to evaluate the performance of regression models. R-squared is used as an estimate of how well the regression model explains the observed data. It is possible to compare R2 against other models trained on the same dataset because it is a relative metric. Better fit is denoted by a higher value. R2 can also be utilized to provide an approximation of the model's general performance.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y_i})^2}$$

Figure 59. R2 score equation.

We assess the efficacy of your optimization model for determining the ideal placement of new charging stations for every type of automobile. The four metrics we used for our optimization issues are shown below.

1. Objective Function Value (Cost): The cost incurred by the charging station placement model is shown by this metric, which is called the objective function value (cost). This cost, which also includes installation and operating costs, has to be kept to a minimum. In terms of budget allocation and resource use, a lower cost denotes a more effective solution.

2. Coverage Metric: The percentage of car types for which charging stations have been effectively installed is measured by the coverage metric. A more thorough solution that supports a wider range of vehicle types and has a larger coverage percentage ensures better service for a wide range of consumers.

3. Utilization Metric: Utilization gauges the effectiveness with which the capacity of the charging stations is being utilized. The ratio of the total demand met to the total station capacity is quantified. A higher rate of utilization indicates that the infrastructure is being used efficiently, which lowers the possibility of either underutilized or overloaded stations.

4. Equity Metric: The equity metric assesses how fairly the load is distributed among the charging stations. By dividing the maximum load on a station by the lowest load on any station, it is calculated. A more balanced distribution results in a lower equity score, lowering the possibility of some stations being overwhelmed while others are underutilized. A more dependable and fair charging infrastructure is made possible by achieving a balanced load distribution.

4.5 Model Validation and Evaluation Results

Model validation and evaluation results for the following analysis include forecasting heavy-duty and Medium Duty EV and Charging Stations Demand, Predicting Vehicle Range of Heavy Duty, and Medium Duty EV, Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV, and finding the Optimal Placement of New Charging Stations for each vehicle type, can give important insights into the efficacy and efficiency of different initiatives and techniques.

4.5.1 Forecasting heavy-duty and Medium Duty EV and Charging Stations Demand

For evaluating the precision and dependability of the predictive models, data from model validation and evaluation for anticipating the demand for heavy- and medium-duty electric vehicles (EV) and charging stations are crucial. In the case of heavy-duty and medium-duty vehicles in particular, these models seek to forecast demand for EVs and charging infrastructure. We have created instructive visualizations after training a Prophet model and collecting predictions for upcoming data are shown in the figure below. The objective is to examine and break down the forecast's numerous elements. While "forecast_df" includes the model's forecasts, "model" refers to the previously trained Prophet forecasting model. When the function is used, distinct subplots are generated to illustrate important elements including yearly trend prediction uncertainty. More in-depth insights into the prediction and its influencing elements are made possible by these visualizations, which aid users in understanding the underlying patterns, trends, and seasonal variations in time series data.

The below Figure 60 shows that usually, the trend component displays the data's general trend. If the line slopes upward, it indicates that the variable has generally increased with time. If the line, on the other hand, a downward-sloping line shows a declining tendency.

From the below Figure 61, we have created customized visualizations of forecasted data. Although the initial forecast offered a core estimate, the additional noise illustrated the underlying ambiguity and fluctuation in actual data. The prediction intervals are represented by the green shading that covers the uncertainty zone.

This is vital for comprehending the model's level of assurance in the predicted values and aids users in determining the risks involved. This plot shows the effective way of communicating forecasting insights while keeping adaptable to different data circumstances and forecasting requirements.

4.5.2 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV

Validation entails examining the models' accuracy in determining how far an electric car can travel after charging. The results of the evaluation might demonstrate how closely the predicted ranges match actual driving data, enabling customers to decide whether their EVs can accommodate their travel needs. In this analysis, the expected energy demand is plotted alongside the data on the actual energy demand for three different time periods—daily, weekly, and monthly. To visually evaluate the model's performance and the precision of forecasts, this is a typical approach in time series forecasting and data analysis. We aim to illustrate how closely the forecasting model matches the observed energy demand data at various time scales (daily, weekly, and monthly). It helps in evaluating the overall accuracy of forecasts and rapidly finding areas where the model may need improvements.

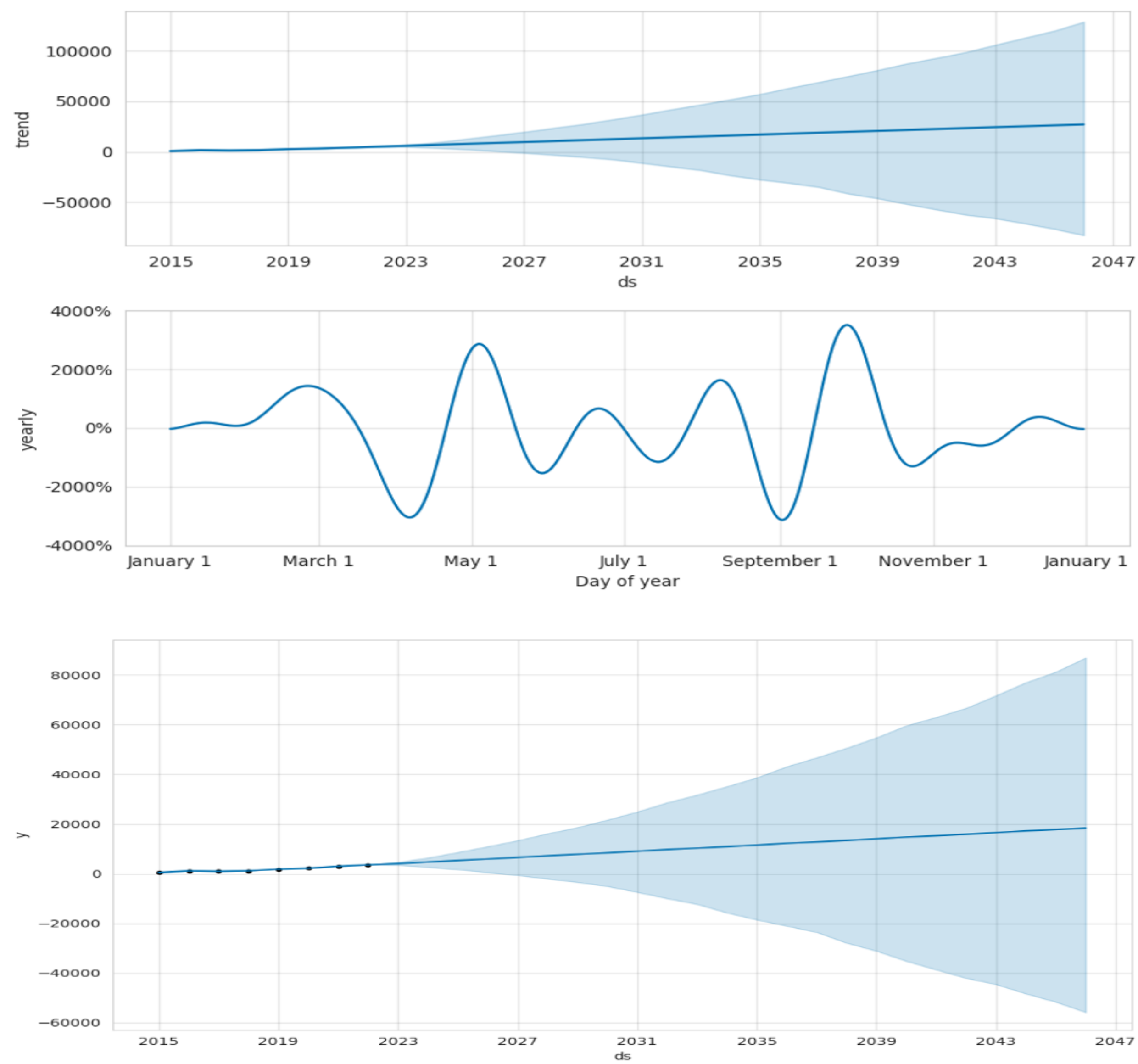


Figure 60. Forecasting yearly trends for EV charging station demand

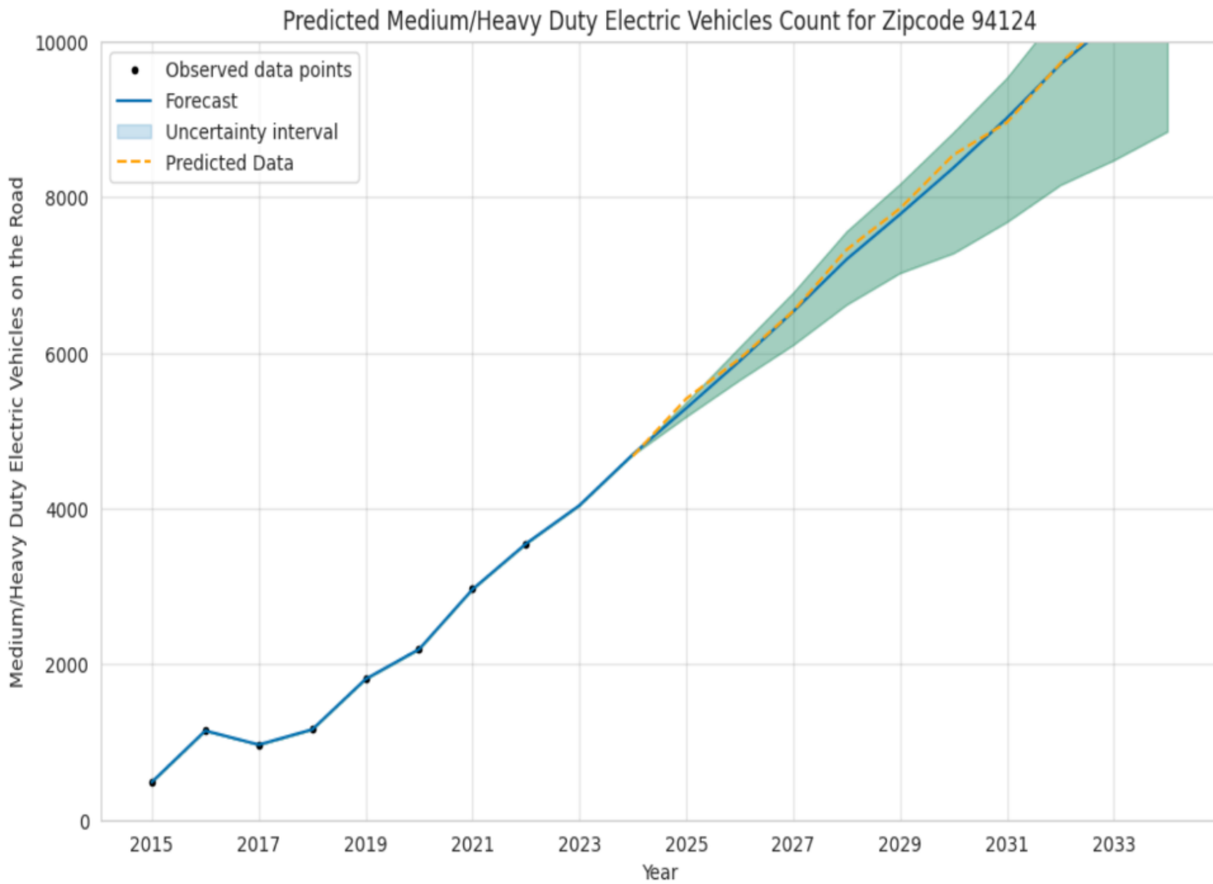


Figure 61. Forecasting Medium/Heavy Duty EV count for Zipcode 94124

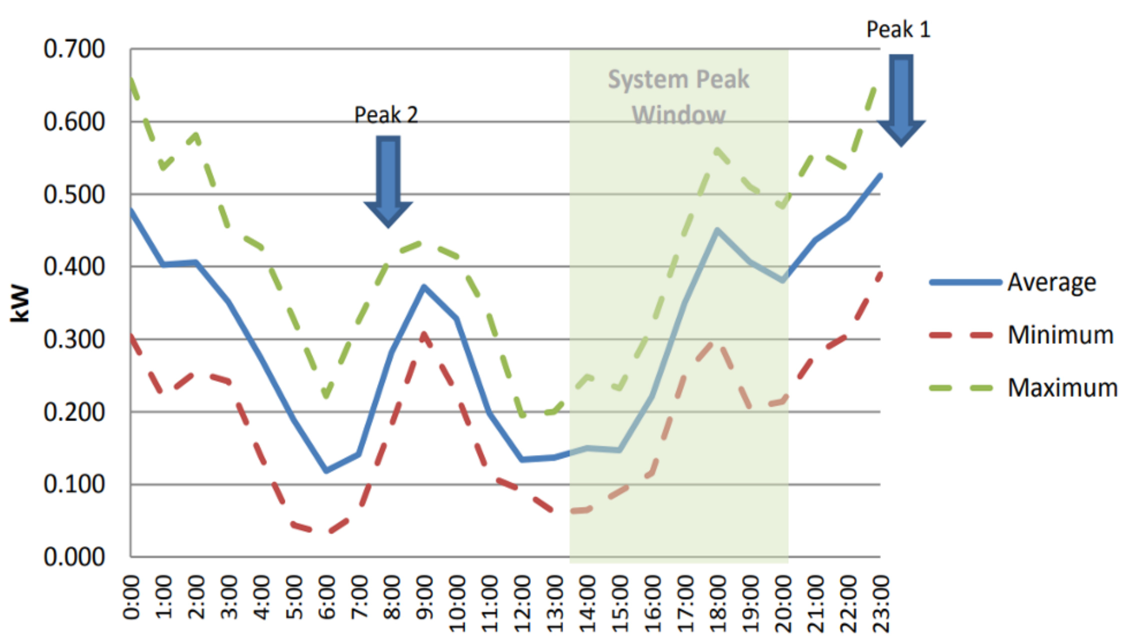


Figure 62. Peak hours Window Profile of EV Energy Demand

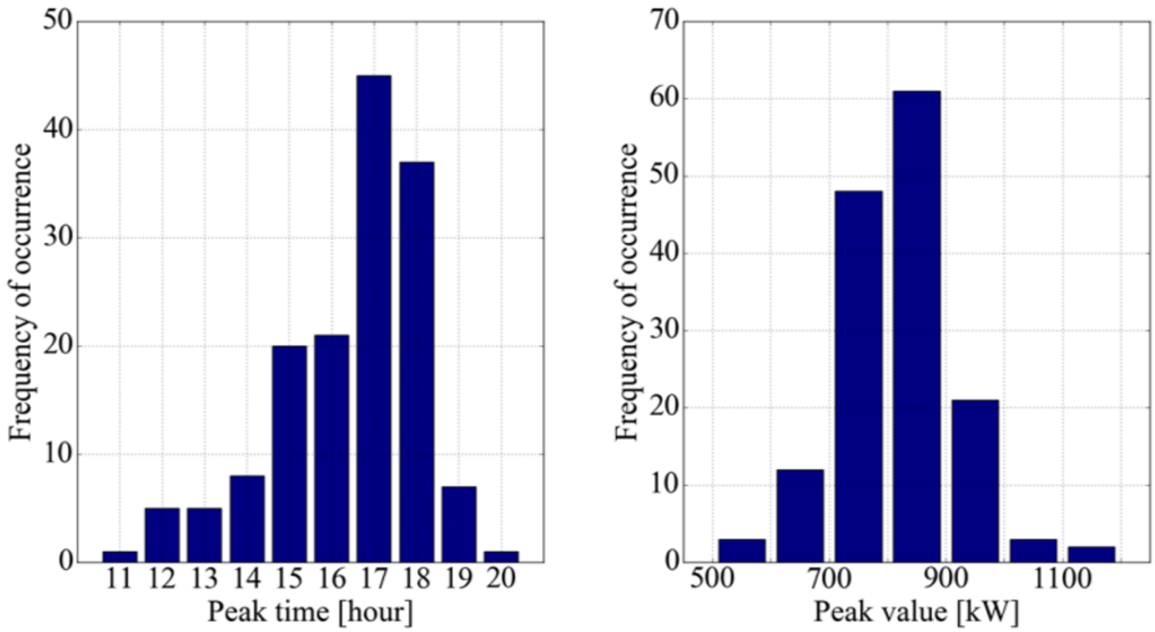


Figure 63. Distribution of Peak Hours and Values for EV Energy Demand

From Figure 66, we can evaluate how well the forecasting model performs by comparing the expected energy demand (orange dashed line) with the actual energy demand (green line). We can see the anticipated line closely matches the observed data; the model is producing reliable projections. We can also identify the temporal patterns in energy consumption by analyzing the plot at various time intervals (daily, weekly, monthly). Also, observe the weekly oscillations, monthly trends, or daily peaks and valleys in energy demand. The distribution of resources and making decisions can both benefit from these patterns. An alteration in resource allocation may be necessary if the model continually overestimates energy demand during particular time periods in order to ensure enough energy supply during peak hours. This chart is a useful tool for assessing how well a forecasting model for energy demand performs. It delivers actionable information for improving the allocation of energy resources and increasing the model's accuracy over various time periods in addition to providing a visual comparison of predictions and actual data.

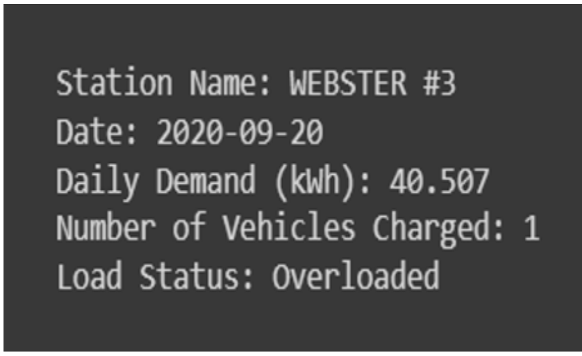


Figure 64. Load Status - Station wise

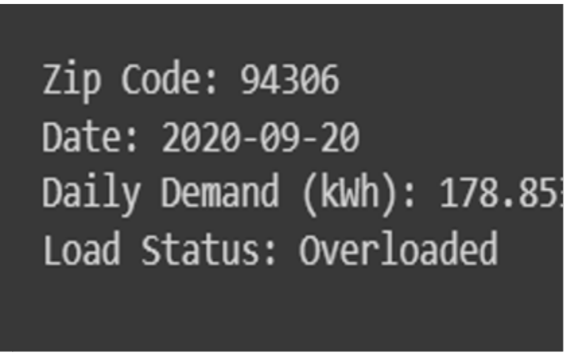


Figure 65. Load Status - Zip Code Wise

The weighted fusion model's performance indicators are summarized in the below table, demonstrating how well it predicts the energy needed for electric vehicle charging sessions. The metrics, which provide a thorough assessment of the model's prediction abilities, include Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2). The low MSE 4.36, MAE 0.86, and RMSE 2.08 values imply that, overall, the model's predictions are relatively comparable to the actual values.

Additionally, the high R-squared value which is 93 percent shows that the model has great predictive power and can account for a significant portion of the variance in energy provided. The following Table 12 summarizes the list of values of the evaluation metrics of all the use cases we mentioned earlier to design a smart EV charging infrastructure system.

4.5.3 Optimal Placement of New Charging Stations for each vehicle type

The PuLP optimization linear programming model's outcomes for model validation and evaluation for the best placement of new charging stations for each vehicle type are as follows: By successfully reducing the overall cost of setting up and running the charging stations, the optimization model has shown effective resource allocation. The coverage metric indicates that a significant portion of car types have strategically positioned charging stations, resulting in a comprehensive and inclusive infrastructure. The utilization statistics also show effective station

Figure 66. Daily, Weekly, and Monthly Energy Demand Forecast

use, lowering the possibility of under- or over-utilization. The equity statistic shows an equitable load distribution among stations, improving access fairness and reliability. To determine how resistant the model is to changes in the parameters, sensitivity analysis has been carried out. These findings show how the model may be used to choose a placement strategy for charging stations that is economical, inclusive, effective, and fair. This helps to create a balanced charging infrastructure. Figure 72 shows the overview of the newly proposed electric vehicle charging station along with the existing stations for the Zip code 94124.

Table 33. Summarized View of Evaluation Metric Values

Feature	Vehicle Type	Time Frame	Model	MSE	MAE	RMSE	R^2
EV Demand Forecast	Heavy/ Medium Duty Vehicles	Yearly	Prophet	12.753	11.695	26.874	0.87
EV Range Prediction	Transit Bus	On Demand	Weighted Fusion Meta Regressor Model	9.063	9.874	88.113	0.75
	School Bus	On Demand		8.947	9.113	87.248	0.76
	Delivery Truck	On Demand		11.345	12.278	83.903	0.82
EV Energy Demand prediction	Transit Bus	Daily	Temporal Fusion Transformer	0.828	0.957	4.36	0.93
		Weekly		1.794	1.897	9.36	0.91
		Monthly		0.984	0.932	7.64	0.91
	School Bus	Daily		0.865	0.847	5.96	0.92
		Weekly		0.897	0.828	4.98	0.92
		Monthly		1.895	1.952	9.68	0.90
	Delivery Truck	Daily		2.969	2.643	12.36	0.88
		Weekly		2.453	2.387	11.97	0.89
		Monthly		2.874	2.775	13.01	0.88

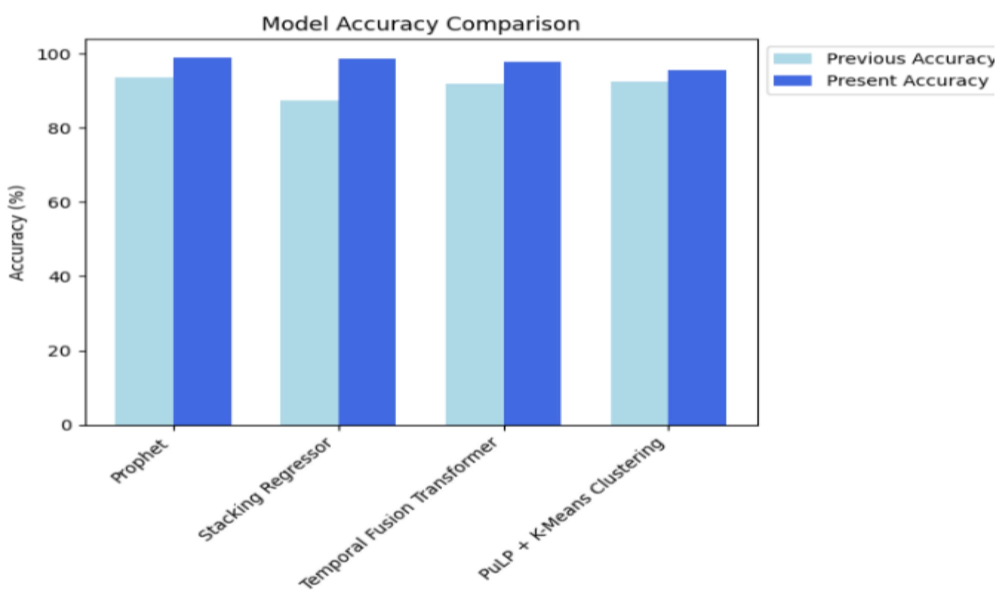


Figure 67. Summary of Model Performance

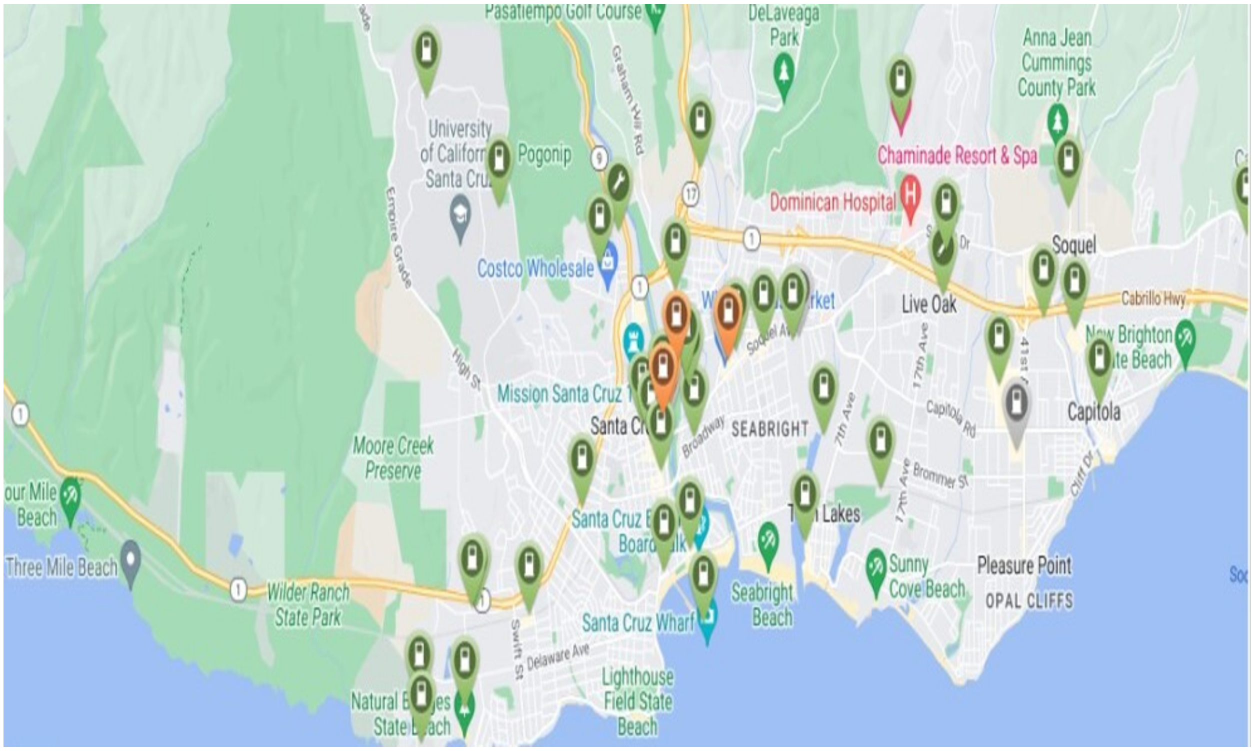


Figure 68. Newly Suggested EV Charging Station Locations for zip code 94124

Green- Existing EV charging Stations Red- Predicted New Location for EV Charging Station

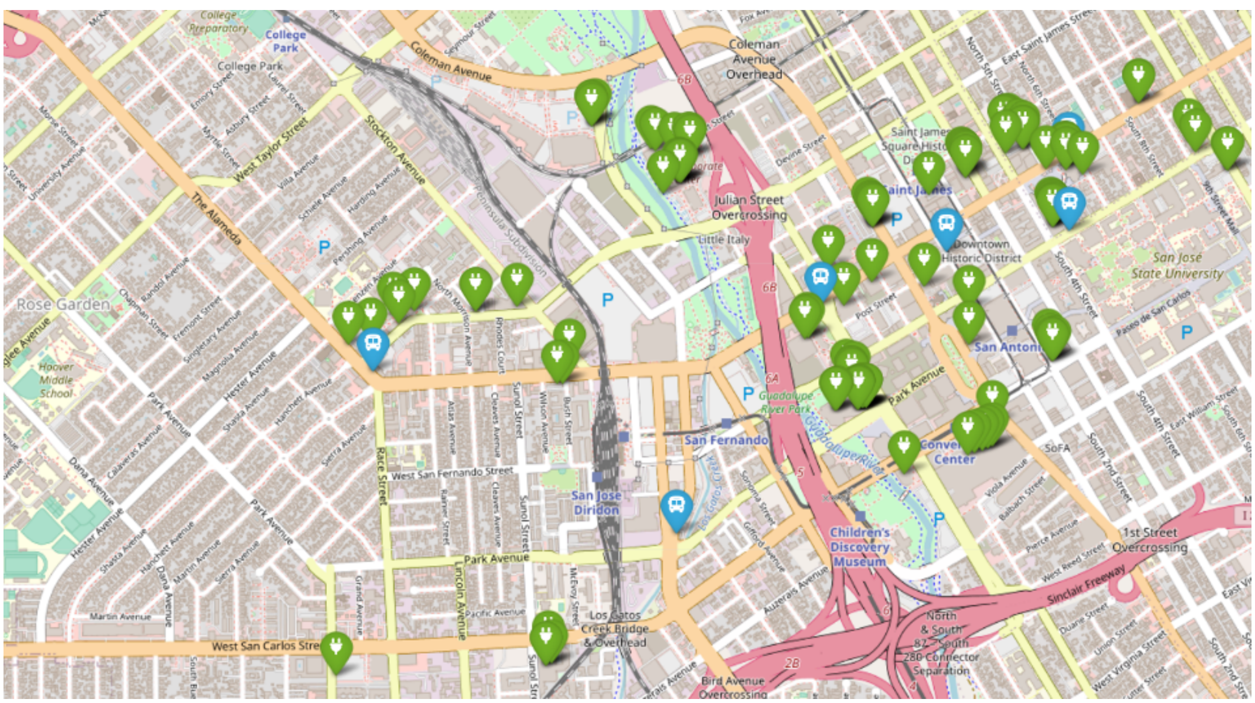


Figure 69. Optimized Location Suggestions for Transit Bus EV charging Stations.

Green- Existing EV charging Stations Blue- Predicted Locations for EV Charging Station

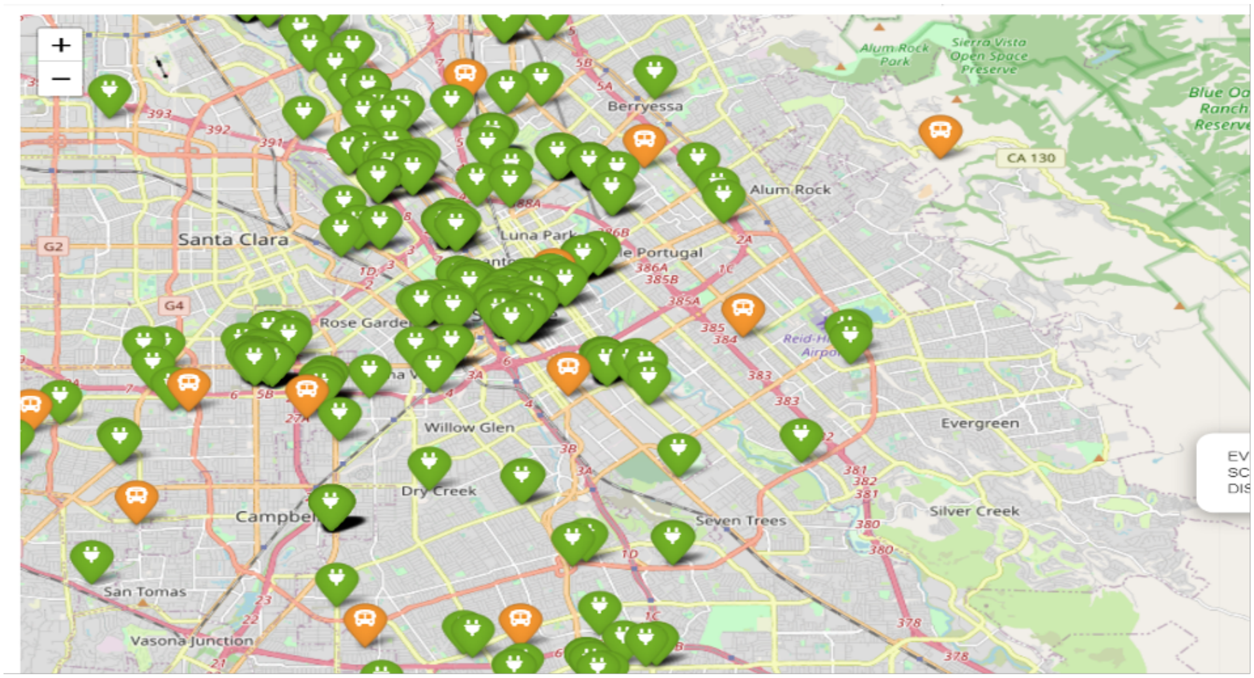


Figure 70. Optimized Location Suggestions for School Bus EV charging Stations.

Green- Existing EV charging Stations Orange - Predicted Locations for EV Charging Station

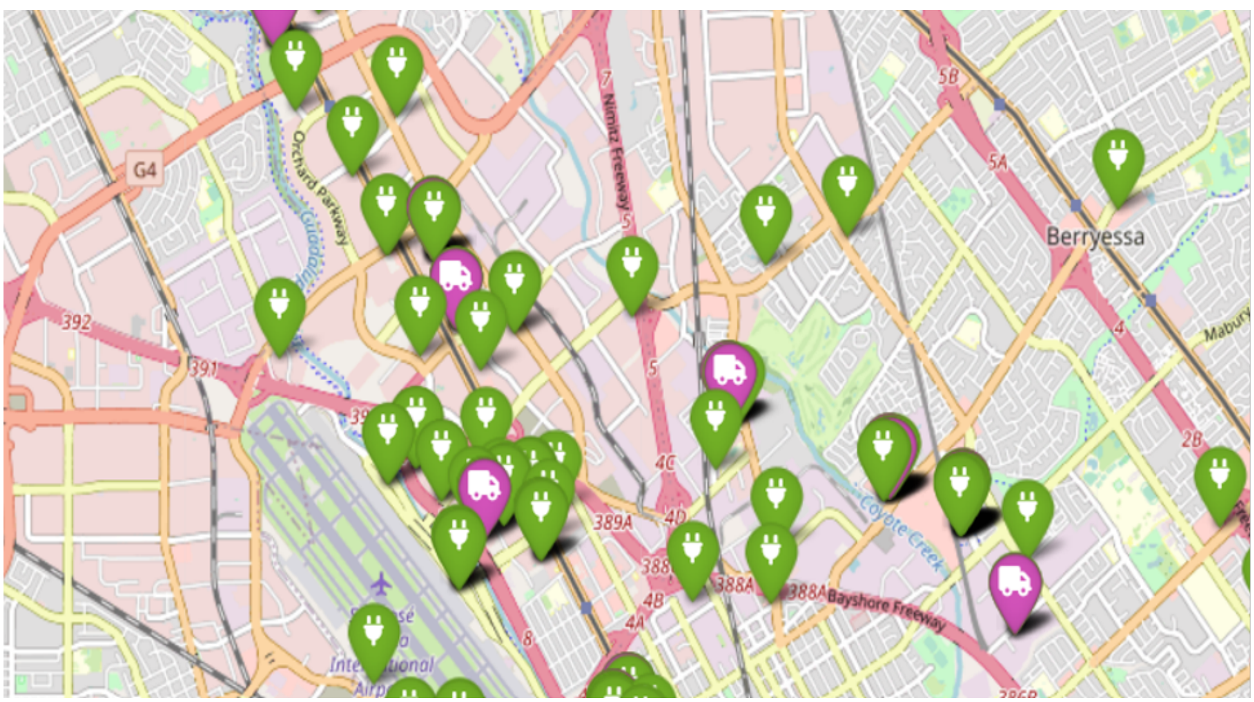


Figure 71. Optimized Location Suggestions for Delivery Truck EV charging Stations.

Green- Existing EV charging Stations Purple - Predicted Locations for EV Charging Station

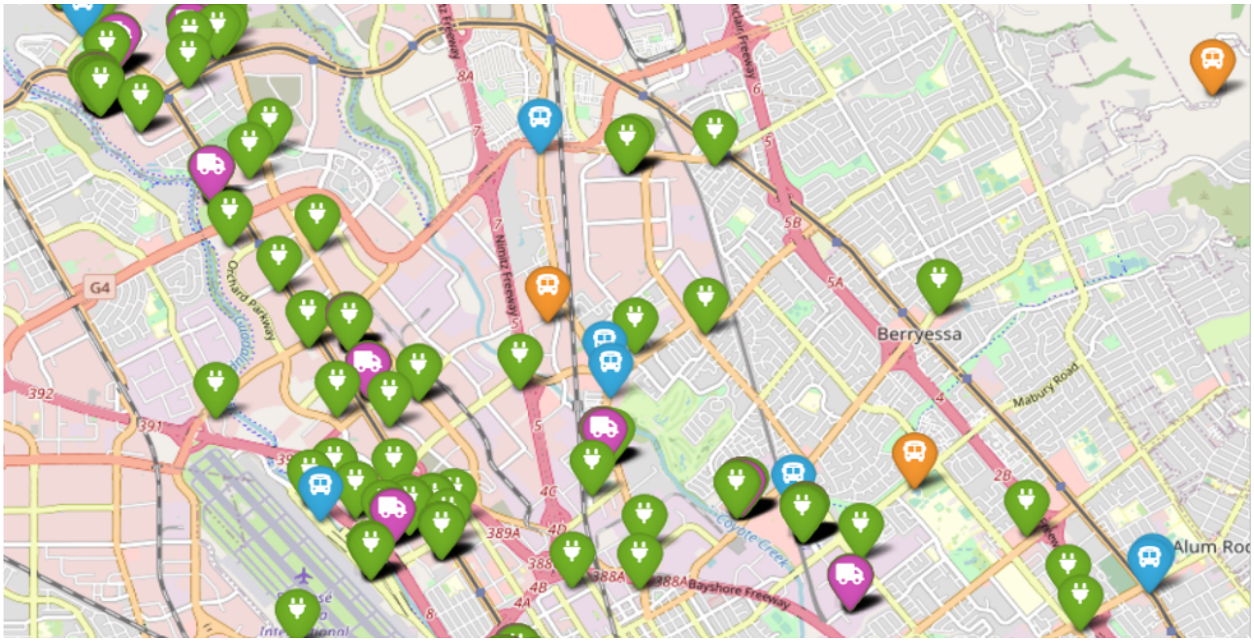


Figure 72. Optimized Location Suggestions for Heavy/Medium EV charging Stations – Zipcode

5.1 System Requirement Analysis

5.1.1 System boundary and Use cases.

The infrastructure needed to plan and implement EV charging station deployment would be included in the system boundary. The potential use cases of this project could involve the deployment of smart EV charging stations in various locations such as parking lots, shopping centers, residential complexes, and highway rest areas. The users for this system would be EV charging network administrators and other staff, property managers, or owners of the locations where the charging stations are placed. These applications, which range from public transportation to commercial logistics and specialized services, demonstrate the adaptability and benefits of transit buses, delivery vans, and school buses in a variety of settings. The versatility makes the system useful in a variety of scenarios and applies to a wide range of vehicle types.

A possible scenario of this use case could be a driver of an electric vehicle utilizing the system to locate a charging station that is available or get suggestions for new charging station locations inside particular zip codes. The system considers variables such as energy use, vehicle range and expected vehicle counts. The system's capabilities include an advanced understanding of factors like anticipated vehicle counts and patterns of energy consumption, guaranteeing that the suggestions meet the requirements of EV users in the short term as well as the general objectives of sustainability and effectiveness in the installation of charging stations.

5.1.2 System high-level data analytics requirements

In this project, to guarantee the efficacy and efficiency of the system devoted to the creation and implementation of EV charging infrastructure of medium/heavy-duty vehicles, high-level data analytics requirements are essential. Firstly, to have an effective method for gathering and integrating data, this process includes combining data from a variety of sources.

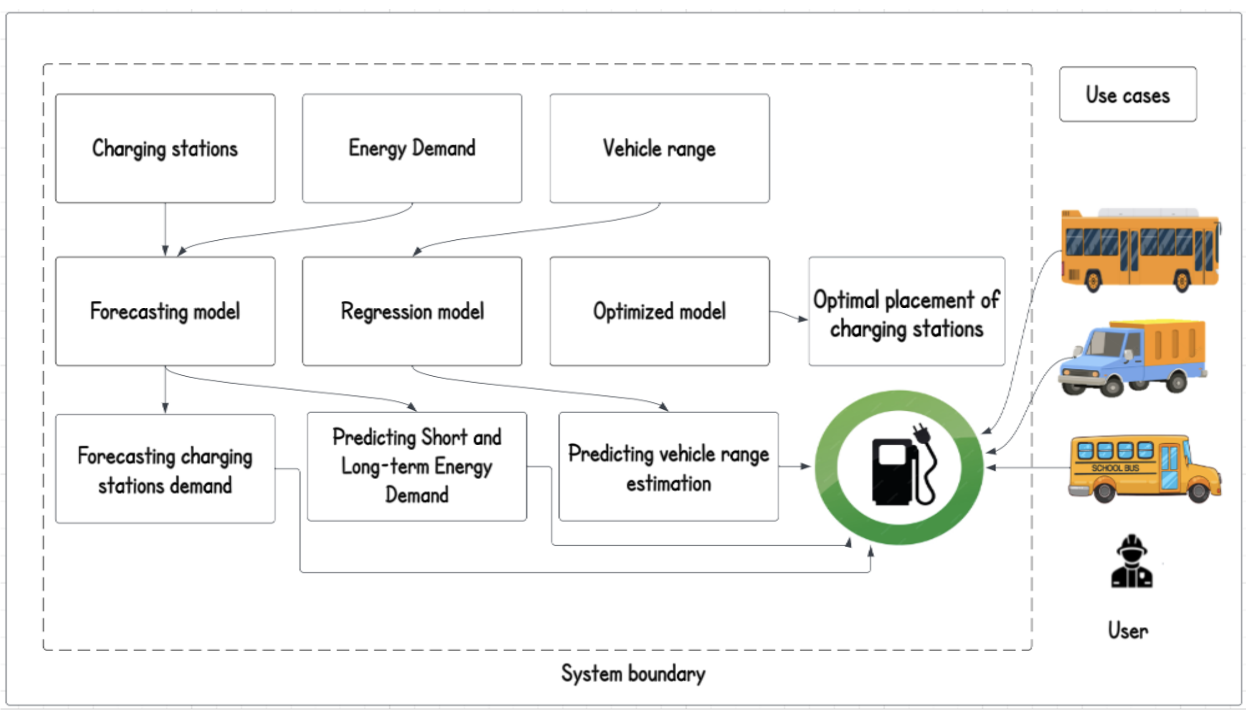


Figure 73. Systemic Boundary, Use Cases and Users

Table 34. System Requirements - Actors

Actors	Description
User	Interacts with the website interface to query analytical results and provides input for new charging station suggestions
Data Analytics Engine	Performs high-level data analytics, generates insights, and manages interactions with machine learning models
Machine Learning Models	Receives input from the data analytics engine and is utilized for forecasting heavy/medium-duty EV vehicle counts, predicting energy demand, and suggesting optimal charging station locations

The system uses comprehensive and real-world data collected from various sources such as commercial fleet vehicle operating data of delivery vans, school buses, and transit buses, charging transactions data, load and energy supply, and geospatial data for predicting the optimal locations. These datasets are necessary to train predictive models, which ensures the system's ability to predict and assess different EV infrastructure aspects with accuracy. Precise elicitation of data is crucial for the development of a reliable and efficient analytics system.

Table 35. System Requirements - Use Cases

Use Case	Description	Actors Involved	Flow
Analytical Results Query	Users query station-wise, site-wise, and zip code-wise analytical results through the website.	User, Data Analytics Engine, Database Management System	User interacts with the website, submits queries, Data Analytics Engine retrieves and processes relevant data from the database, and analytical results are displayed to the user.
New Charging Station Suggestion	Users input parameters through the website for suggesting new charging station locations.	User, Data Analytics Engine, Machine Learning Models	User provides input parameters, Data Analytics Engine processes the input, and Machine Learning Models suggest optimal charging station locations based on forecasting models.

Table 36. Components utilized in the System Boundary

Internal Components	Website Interface
	Data Analytics Engine
	Machine Learning Models
	Database Management System
External Components	EV Charging Station Database
	Geographic Information System (GIS) Data
	User Input (through the website)

The data analytics functions within our system encompass a range of tasks aimed at providing users with meaningful insights into the current state and future trends of EV charging stations. These functions include:

Descriptive Analytics: Descriptive analytics within our system serves as the foundational layer, summarizing and illuminating the current state of EV charging stations. This involves extracting meaningful insights from historical data, enabling users to comprehend the characteristics and trends associated with existing charging infrastructure. The system employs various statistical measures to achieve this, including measures of central tendency, dispersion, and frequency distribution. Station-wise, site-wise, and zip code-wise descriptive statistics are meticulously calculated, offering users a comprehensive view of factors such as usage patterns, peak charging times, and the distribution of heavy/medium-duty EVs across different regions. By leveraging descriptive analytics, the system provides a snapshot of the current landscape, aiding stakeholders in making informed decisions based on a thorough understanding of existing charging station dynamics.

Predictive and Prescriptive Analytics: Moving beyond the present, our system incorporates predictive analytics to forecast future trends in the realm of EV charging infrastructure. Employing advanced machine learning models, the system predicts heavy/medium-duty EV vehicle counts and anticipates the energy demand at charging stations. These predictions are crucial for proactive decision-making and resource allocation. Furthermore, the system engages in prescriptive analytics by suggesting optimal actions based on the forecasted scenarios. Specifically, it recommends strategic locations for new charging stations to accommodate the projected demand effectively. This integration of predictive and prescriptive analytics empowers stakeholders with forward-looking insights, enabling them to plan and expand EV infrastructure in alignment with anticipated trends and demands. The dynamic combination of these analytics functions positions the system as a valuable tool for sustainable and strategic growth in the electric vehicle charging ecosystem.

In the training process, for accurate forecasting, the system's core depends on machine learning modeling techniques. To identify and use important connections and patterns in the datasets, ongoing training on data is essential. For the analysis stage, this step entails assessing the precision and dependability of predictive models created to anticipate the demand for heavy- and medium-duty EVs as well as charging infrastructure, specifically with regard to EV and charging station demand forecasts. The analysis employs visualizations created after training the Prophet model, providing insights into time series data, trends, and seasonal variations. These visualizations enable data patterns, thereby making informed decisions based on accurate forecasts. The process of iteration assures that the predictive capabilities of the system consistently develop and adapt to the demands of EV’s and the requirements for charging infrastructure.

Data specifications include time series data to capture seasonal fluctuations and temporal trends. These specifications are essential for ensuring that the input data is of the highest quality and appropriate for precise forecasting based on real-world data. Information on charging habits and time-related trends in energy usage is essential for maximizing the distribution of resources. The system offers insightful information that directs its flexibility to the changing environment of electric vehicle usage as it explores all the nuances of charging behavior.

Table 37. System Scenario Use Cases

Scenario	**Description**
Peak Usage Scenario	Assesses the system's response to peak user queries and demands.
New Charging Station Recommendation Scenario	Analyzes the accuracy and efficiency of machine learning models in suggesting optimal locations for new charging stations.
Real-time Data Update Scenario	Examines the system's response to real-time updates in the charging station dataset.
Geographical Variation Scenario	Tests the system's ability to handle variations in geographical data for different regions.

The continuous processes of validation and verification are essential to preserving the accuracy and dependability of the system. These phases make sure that ML models are constantly retrained to adjust to changing circumstances and that training data stays in line with real-world data. When evaluating the predictive power of the system, validation and verification are essential components in determining metrics like MSE, MAE, RMSE, and R-squared (R2). These metrics offer a numerical representation of the predicted accuracy, allowing for a thorough evaluation of the system's operation and which allows ongoing improvement for the optimal outcomes. Maintaining the system's effectiveness and which ensures its capacity to offer reliable insights into heavy- and medium-duty EV demand and charging infrastructure projections are made possible by the systematic validation and verification procedures.

5.2 System design

5.2.1 System Architecture and Infrastructure

Figure 74, illustrates the architecture and infrastructure in order to execute multiple use cases within a single workflow, such as forecasting future EV and charging station demand count at particular zip code, vehicle range prediction, determining short- and long-term energy needs like hourly energy demand, next day energy demand and monthly energy demand respectively, and optimizing charging station best location placement, a comprehensive environment, platform, and set of tools must be available.

Table 38. System Infrastructure Components and Technology

Component	**Description**	**Technology Used**
Web Servers	Hosts the website interface and handles user requests.	AWS EC2, Load Balancer, Apache/Nginx
Database Management System	Stores and manages the dataset of existing EV charging stations.	Amazon RDS (Relational Database Service)
Data Analytics Engine	Central component for executing high-level data analytics functions	AWS Lambda, Python, Pandas
Machine Learning Model Servers	Hosts and manages the deployed machine learning models	AWS SageMaker, TensorFlow, Scikit-Learn
Geographic Information System (GIS)	Integrates geographical data for mapping and location-based analyses.	AWS Location Service, Google API
User Interface (UI)	The front-end interface where users interact with the system.	React, JavaScript, HTML, CSS, Flask
Communication Protocols	Establishes secure communication channels between different system components.	HTTPS, API Gateway

Users send a request on the website using Flask Architecture. Then the backend process begins where it comprises of collection of all datasets and to identify all the parameters required for the machine learning models and all these datasets are hosted and stored in Amazon S3 and then collecting real - time data which will send the parameters to predict the desired results to the machine learning models and predict the required output. Then the system will display the desired output to the user on the website.

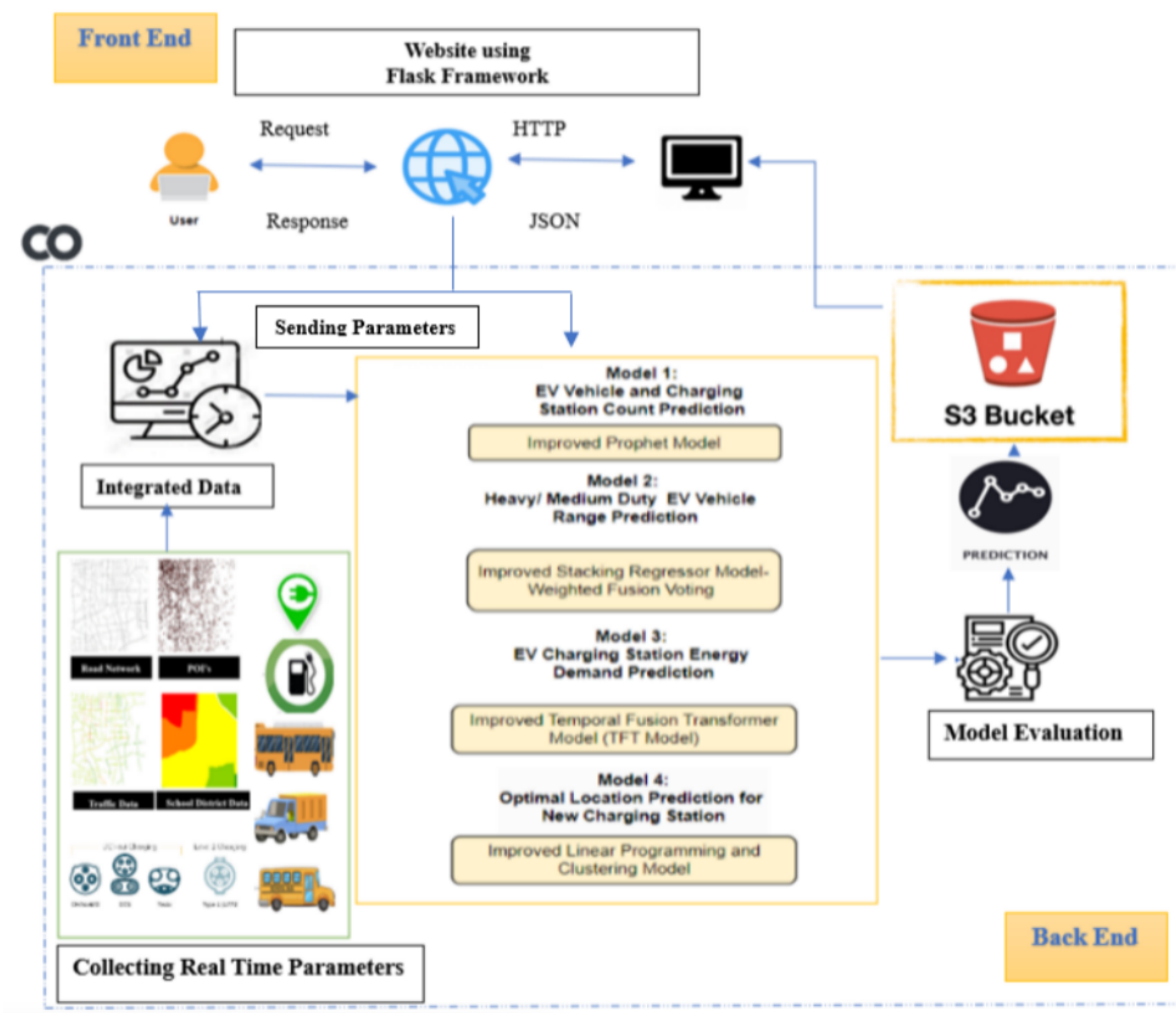


Figure 74. System Architecture

5.2.2 System Supporting Platforms and Cloud Environment

The foundation is constructed using the Python programming language and its vast libraries for machine learning, optimization, and data analysis. The interactive development and documentation platform is Jupyter Notebook. The set of tools includes TensorFlow or PyTorch deep learning frameworks for Temporal Fusion Transformer (TFT) model implementation, Scikit-Learn for regression models and ensemble approaches, and Facebook's Prophet for demand forecasting. The PuLP package makes it easier to create linear programming models for charging station placement. Pandas, NumPy, and data visualization tools like Matplotlib and Seaborn are popular tools for data management and analysis. Version control and collaboration are aided by Git/GitHub, while task organization is facilitated by project management software like Jira. With a particular emphasis on AWS SageMaker for machine learning and data analysis, the entire workflow is smoothly organized within the AWS environment. AWS S3 stores and manages historical data and Facebook Prophet forecasting time series data. SageMaker is a flexible platform for creating, honing, and deploying machine learning models. Among these applications are the Temporal Fusion Transformer (TFT) model, which can be used for energy demand forecasting and to estimate the driving range of vehicles. By using AWS Lambda, data retrieval and model execution can be automated, computational operations can be hosted by EC2, optimization data can be managed by DynamoDB, and cloud monitoring can be done by CloudWatch, which monitors the core AWS components. The transformation of data is handled by AWS Glue and Athena, while the visualization of the data is handled by Tableau.

There are numerous use cases that illustrate the meticulously designed data flow and platform architecture that support the comprehensive use cases illustrating how the system is designed to optimize data processing and automate the completion of complicated tasks. As soon as the data is ingested, it is routinely and systematically entered into a specific bucket of the AWS S3 service, where the historical information on the demand for EVs and charging stations is stored. As a result of AWS Glue, complex data transformations and cleansing procedures can be seamlessly automated, which will improve the quality and consistency of the data. These preprocessing steps are undertaken to ensure that the data is thoroughly cleansed and prepared for further research and modeling to take place after they have been preprocessed. Furthermore, the platform architecture is designed in such a way that data insights and reporting can be carried out via a specific route.

It is possible to create expert dashboards and visualizations using Tableau, a tool for analytics and visualization that is used by many analysts. The Jupyter Notebook is also an essential tool when it comes to producing detailed project reports and coding documentation. Scaling and deployment are the final steps in the workflow. With the help of Elastic Beanstalk and Amazon Elastic Compute Cloud (ECS), deployment is easy. By implementing AWS Auto Scaling, workloads can be adjusted efficiently and correctly as workloads vary, since resources are distributed and managed seamlessly so there are no interruptions in service.

5.2.3 System Data Management Solution

As a result of the introduction of a reliable and widely distributed charging infrastructure, an electric transportation system that can create a sustainable future, especially for heavy and medium-duty electric vehicles (EVs), should be developed in a way that is conducive to the development of an electric transportation system. Using AWS SageMaker Jupyter Notebooks as a next step, we will be able to orchestrate the analysis and modeling of the data with the use of AWS SageMaker Jupyter Notebooks. By using SageMaker software, you will be able to create, train, and rigorously evaluate machine learning models. It is possible to optimize the model in a timely and accurate manner by coordinating automatic data retrieval and restoring data promptly by using Lambda functions in AWS. An optimization model is available on Amazon EC2 which is designed to support the decision-making process based on the data available. Additionally, Amazon DynamoDB can be used to organize and store optimization data meticulously, allowing models to run smoothly and efficiently, thanks to its outstanding storage capabilities.

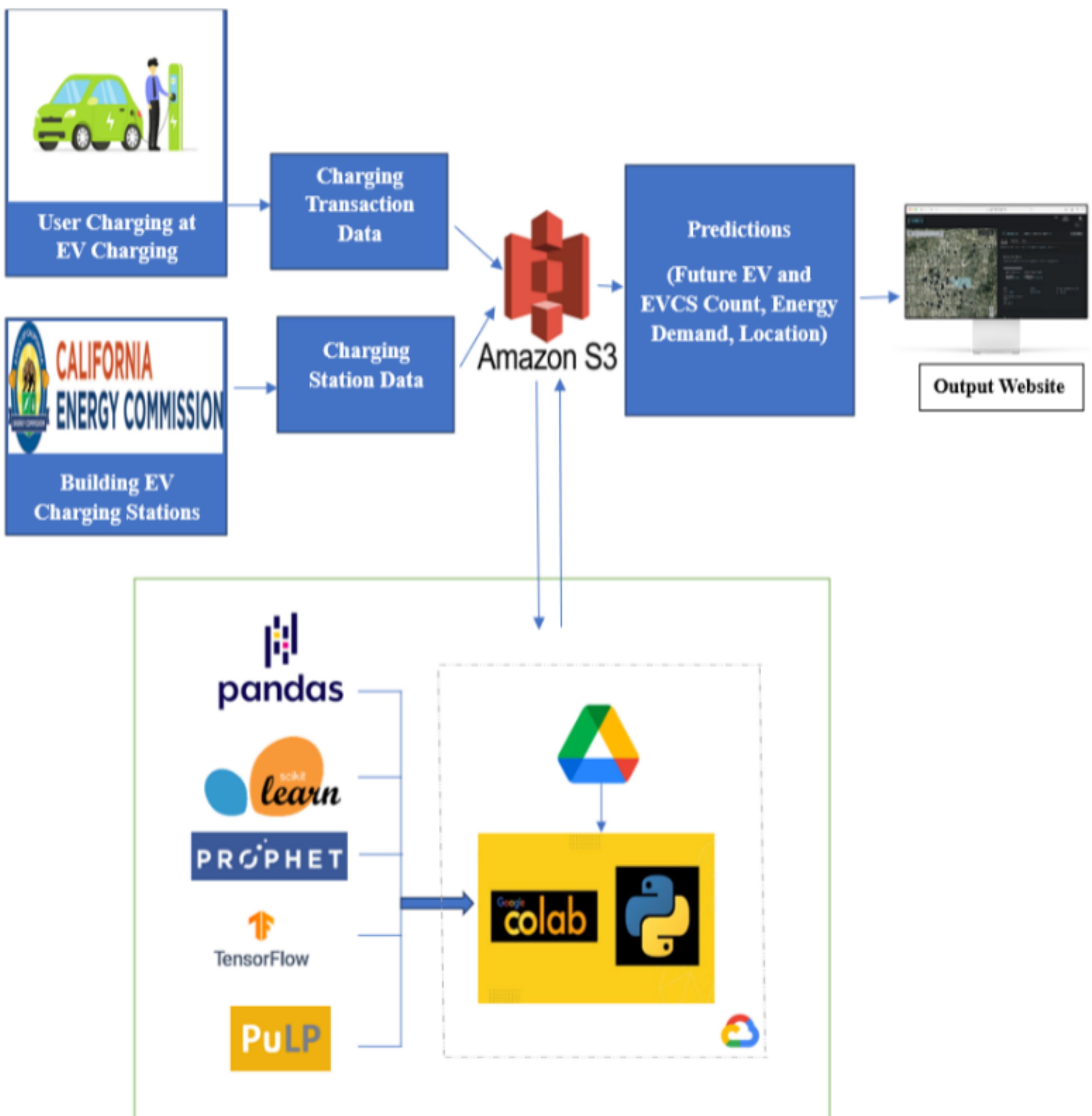


Figure 75. System Supporting Platforms and Cloud Environment

5.2.3.1 System Database Design

The system's database design is structured to efficiently store and manage the diverse data required for Electric Vehicle (EV) Charging Infrastructure Planning. The central table, Charging Stations, captures essential information about each charging station, including its unique identifier (StationID), geographical location, charging capacity, usage statistics in JSON format, and a timestamp indicating when the data was recorded. User queries and corresponding analytical results are stored in separate tables—User Queries and Analytical Results. The User Queries table logs each query submitted by a user, with a unique identifier (QueryID), user identifier (UserID), the query itself, and a timestamp. The Analytical Results table links to the user queries and stores the analytical results, providing a reference to the original query through the QueryID foreign key. This structured approach ensures an organized and accessible database that accommodates user interactions and stores analytical insights effectively.

5.2.3.2 System User Interface

Figure 76 shows an example of the system deployment architecture of the charging station interface in California State and San Jose city with respect to the desired code for medium and heavy-duty vehicles such as transit bus, school bus and delivery truck. A map of available EV charging stations is displayed in the front-end interface, which is built with HTML, CSS and Javascript. In addition to that, the zip code's current population and the number of EV vehicles registered in the zip code are also displayed. The desired query given by the user will be searched with the database and connected to the Amazon S3 and the respective machine learning models which were trained will be predicting the stations as shown in the below diagram either the school bus or transit bus or delivery trucks charging stations, their respective locations will be displayed so that the user can get charged to the nearest charging station.

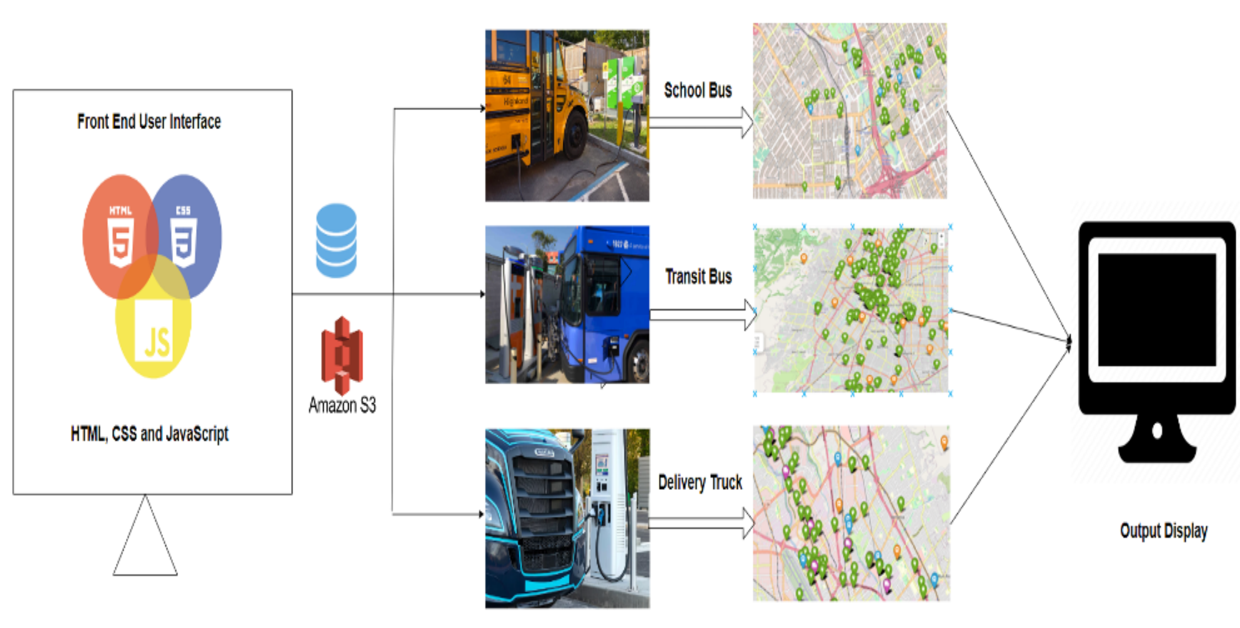


Figure 76. System User Interface Clone

5.2.3.3 System UI Design

The system's user interface (UI) is meticulously designed to provide a seamless and intuitive experience for users engaging with the EV Charging Infrastructure Planning System. The dashboard serves as a comprehensive overview, featuring visually appealing charts and graphs for station-wise, site-wise, and zip code-wise analytical results. A user-friendly query submission form facilitates the input of parameters for heavy/medium-duty EV vehicle count, energy demand, and optimal charging station locations. The integration of interactive maps, powered by Google API, allows users to explore charging station locations and geographical insights visually. The UI is crafted with a responsive design, ensuring optimal functionality across a variety of devices, including desktops, tablets, and mobile phones. Visualization widgets, such as pie charts, bar graphs, and heatmaps, enhance the presentation of analytical results, contributing to an intuitive understanding of the data.

Table 39. Description of System UI Design Components

UI Component	Description
Dashboard Overview	Visualizes station-wise, site-wise, and zip code-wise analytical results through charts and graphs.
Query Submission Form	User-friendly form for submitting queries with parameters for forecasting and analysis.
Interactive Maps	Google-powered maps for exploring charging station locations and geographical insights.
Responsive Design	Ensures a seamless and consistent experience across devices, including desktops and mobile devices.
Visualization Widgets	Pie charts, bar graphs, and heatmaps enhance the presentation of analytical insights.

5.3 System Development and Implementation

Building a thorough system for predicting heavy-duty and medium-duty EV demand and charging station requirements necessitates a multidisciplinary approach that includes data analysis, modeling, stakeholder engagement, and continuous improvement to offer trustworthy and beneficial insights for decision-makers.

We have developed a system that combines different models, each individually trained and tested, to produce a set of outputs, such as forecasting the charging station demand, predicting vehicle range, and forecasting short- and long-term energy demand. This will enable us to achieve the project's objective of identifying the best locations for new charging stations tailored to different vehicle types, including transit buses, school buses, and delivery trucks. Our strategy is built around this integrated system, which unifies the many parts and models created for various vehicle kinds. The main elements and procedures that make up this system are summarized as follows:

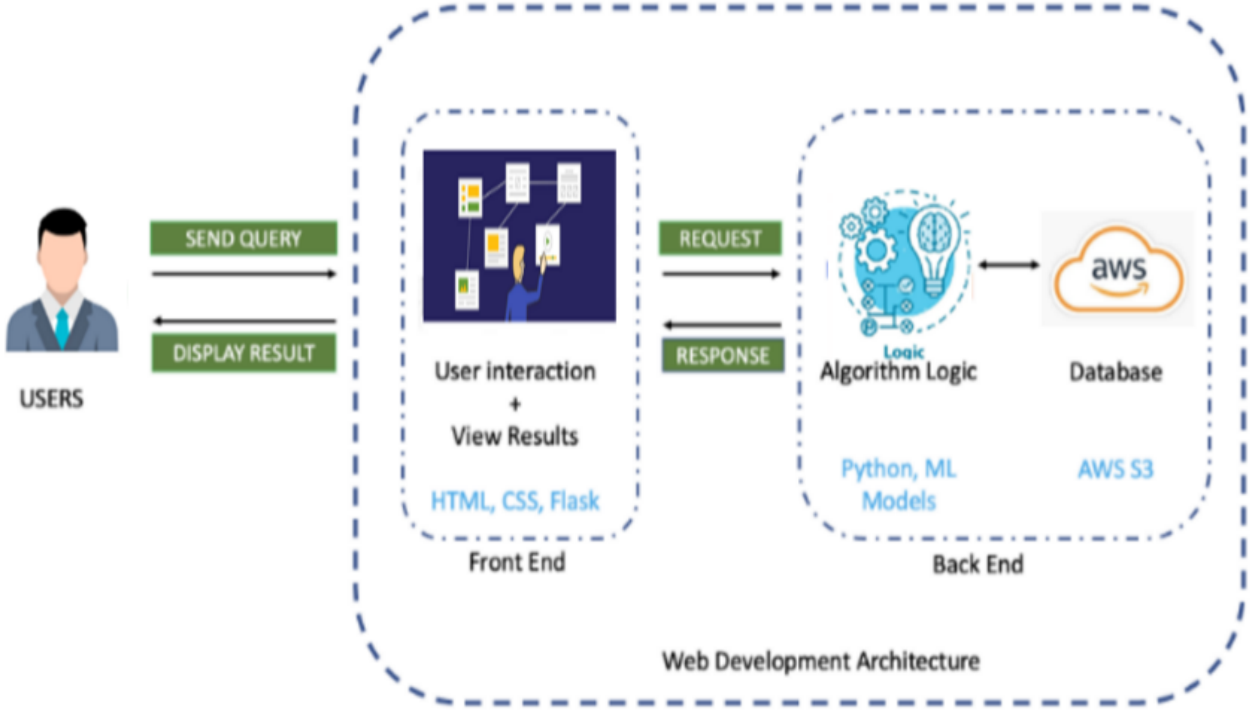


Figure 77. User Interface

5.3.1 Gathering and preparing data for analysis and modeling

We have gathered and preprocessed pertinent data for transit buses, school buses, and delivery trucks. This includes details on the paths taken by vehicles, traffic patterns, the capacity of the energy system, and other relevant factors.

5.3.2 Developing and formulating models tailored for a particular objective or function.

We have developed unique forecasting and optimization models such as Prophet model, ensemble machine learning models and Temporal fusion transformer for forecasting the charging stations demand, predicting the vehicle range and long and short-term demand of Heavy Duty and Medium Duty EV. These models consider the characteristics of each vehicle that are unique to each category, such as driving habits, energy needs, and practical limitations. To ensure accuracy, these models trained and validated individually.

5.3.3 Developing an Optimization Algorithm.

We have developed an optimization algorithm by utilizing the capabilities of the PuLP library and linear programming that uses the outcomes of different models to identify the best placements for charging stations. This algorithm considers various factors, including closeness to routes, Points of Interest (POI), energy needs, grid capacity, and cost efficiency.

5.3.4 Implement Designed System

We have developed a comprehensive implementation strategy that outlines the timeline, budgetary requirements, and procedural steps required for the deployment of the new charging stations with respect to the zip codes.

Table 40. Developed AI and Machine Learning Solutions

Targeted Problem	Machine Learning Model	Description
Count Forecast of Heavy/Medium Duty EV Vehicles	Prophet	Utilizes the Prophet time series forecasting model for predicting the count of heavy/medium-duty EV vehicles. Integrated into the data analytics engine.
Range Prediction	Stacking Ensemble Regressor	Employs a Stacking Ensemble Regressor model with weighted fusion to predict the range of EV vehicles. Integrated into the data analytics engine.
EV Charging Station Energy Demand Prediction	Temporal Fusion Transformer (TFT)	Leverages the Temporal Fusion Transformer (TFT) model for forecasting energy demand at EV charging stations. Integrated into the data analytics engine.
Optimal Location Prediction for New Charging Stations	PuLP Linear Programming Optimization with K-Means Clustering	Combines PuLP linear programming optimization with K-Means clustering to suggest optimal locations for new charging stations. Integrated into the data analytics engine.

We have considered various scenario analyses to account for many future possibilities, such as fleet size changes, technological developments, and legislative changes that can be easier to create adaptable recommendations. We have utilized Geographic Information Systems (GIS) technology to graphically evaluate geographic data, assisting in the selection of the optimum places for charging stations. Spatial restrictions and environmental factors can also be taken into account in this process using GIS.

Table 41. Expected Outputs of our Proposed System

Expected Output	**Description**
Count Forecast of EV Vehicles	Time-series predictions for the count of heavy/medium-duty EV vehicles.
Range Predictions	Accurate forecasts of the range for different EV vehicles.
Energy Demand Predictions	Predictions of energy demand at individual charging stations.
Optimal Location Suggestions	Recommendations for optimal locations for new charging stations.

We have utilized a Tableau dashboard which is an effective tool for data visualization that enables users to convert raw data into engaging and insightful insights. You can design interactive dashboards with Tableau that let users examine data, see trends, and make wise decisions. We Established a continuous monitoring system to assess the effectiveness of charging stations and adjust the positioning strategy as necessary in response to actual usage patterns and shifting environmental factors. The end product of the project includes an interactive website with detailed information of newly placed charging stations with respect to the zip codes and considering various scenarios.

These components can be efficiently combined to create a comprehensive set of recommendations for tactically positioning charging stations for transit buses, school buses, and delivery trucks. The charging infrastructure is placed strategically to meet the specific needs of each vehicle category using this strategy, which maximizes project efficiency. The Electric Vehicle (EV) Charging Infrastructure Planning System integrates advanced AI and machine learning solutions to address key challenges in forecasting, demand prediction, and optimal location planning. To enhance the system's adaptability and functionality, a suite of Solution APIs has been developed, enabling seamless integration and communication between different components. These Solution APIs play a crucial role in the system's architecture, providing a bridge between the developed models, the data analytics engine, and external services. The RESTful Machine Learning Model APIs are implemented using Python and Flask, enabling easy access and utilization of predictive models. Additionally, Geographical APIs leverage services like Google and AWS Location Service, enhancing the system's capabilities for spatial analysis and visualization. This comprehensive integration ensures a holistic and intelligent solution for EV charging infrastructure planning.

Table 42. Solution APIs utilized in the Proposed System

API	Description	Technology Used
Machine Learning Model APIs	Expose functionalities of developed machine learning models, allowing integration into the data analytics engine.	RESTful APIs, Python, Flask, AWS SageMaker (for model hosting)
Geographical APIs	Facilitate communication with GIS and mapping services for spatial analysis and map rendering.	Google API, AWS Location Service

5.4 System Supporting Environment

In this project, the specific environment consists of a customized combination of programming languages, machine learning frameworks, and collaborative platforms that are all carefully coordinated to predict energy consumption, optimize the infrastructure for charging, and strategically locate charging stations. Within this context, promoting the development and implementation of sophisticated models, including the Temporal Fusion Transformer (TFT), and offering an extensive toolbox for machine learning, data analysis, and visualization.

Temporal Fusion Transformer (TFT):

The TFT model is an essential component of effective energy resource management, optimized charging infrastructure, and smooth operation of heterogeneous vehicle fleets. It is built as a cutting-edge neural network architecture for time series forecasting. Its ability to take into account temporal aspects, identify long-term trends, and grasp complex temporal linkages.

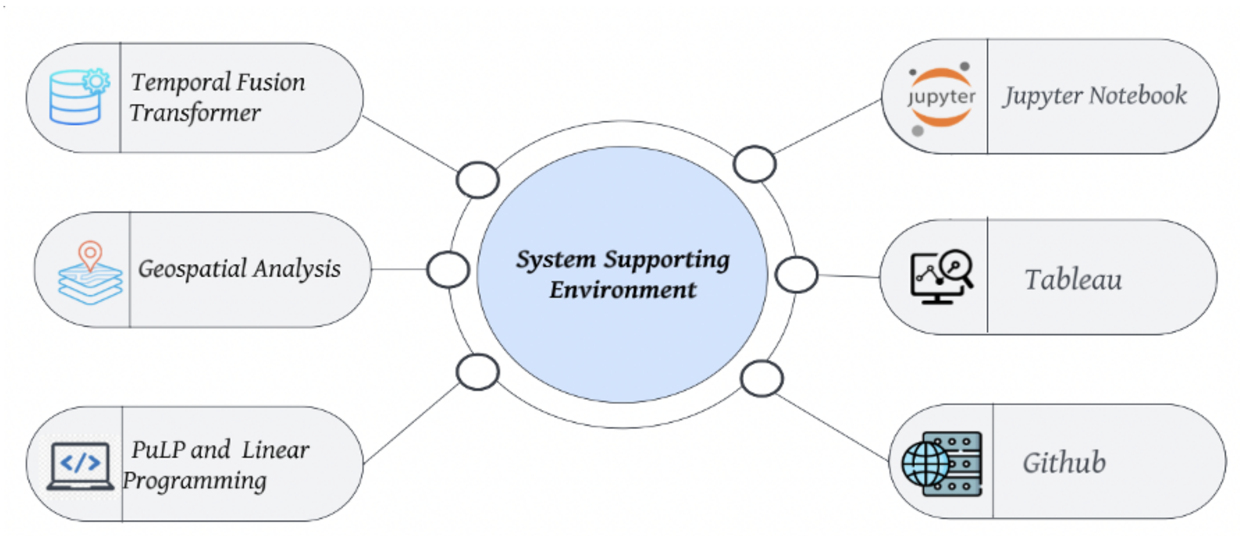


Figure 78. System Supporting Environment

Because of its exceptional ability to combine static and temporal components, the model provides a thorough understanding of the data and can generate precise predictions over a range of periods. For delivery trucks, school buses, and transit buses, the TFT model streamlines the charging procedure. This optimization eventually improves operational effectiveness by guaranteeing the availability of the right amount of energy precisely when and where it is most needed.

PuLP Library and Linear Programming:

The core of the optimization model is the application of linear programming and the PuLP library. These tools contribute to the creation of a model that strategically positions charging stations, taking into account Points of Interest (POI), proximity to existing charging stations, location effectiveness, and cost-effectiveness. The EV project makes use of PuLP's linear programming to guarantee the economical and efficient placement of charging infrastructure while taking into account a number of variables, including user preferences, and the distribution of currently available charging stations.

Geospatial Analysis:

The application of geospatial analysis is vital to this project, as it provides insightful information and strategic considerations that improve the deployment of charging infrastructure, focusing on efficiency and effectiveness. Geospatial analysis, as it applies to EVs, is the process of analyzing geographic data to determine the best places for charging stations. This analysis leverages GeoPandas for processing geographical data and Folium for interactive mapping, allows visualization of charging station distributions, proximity to Points of Interest (POI), and geographical trends. The EV project acquires the capacity to recognize trends in the demand for charging stations based on regional parameters by integrating geospatial analysis. By utilizing Geospatial Analysis tools, it enables the project to not only optimize the placement of charging stations for user convenience but also ensures that the infrastructure aligns with broader urban and regional planning objectives.

Jupyter Notebook:

An open-source application that is used to create and share documents that contain code, equations, and visualizations. We have used Jupyter Notebook for our data analysis, to understand the historical patterns of the heavy-duty vehicle counts, and to create visualizations for optimal locations for new charging stations. This interactive development environment supports the execution of Python programs, to leverage the extensive libraries available for machine learning, optimization, and data analysis. This comprises libraries that are necessary for data preprocessing to training predictive models, like NumPy, Pandas, Scikit-Learn, TensorFlow, and GeoPandas.

Tableau:

With Tableau, used for data visualizations, it's easier to create visually appealing and informative dashboards that present important data about the location of charging stations, energy demand projections, and overall system performance. Through intuitive visualizations, such as maps, charts, and graphs, tableau makes it possible to depict geographic dispersion, charging station utilization trends, and the effects of various factors on the EV infrastructure.

GitHub:

GitHub serves as a centralized hub for code repository management, developer communication, and version control for the different parts of the development of the EV infrastructure. The project's primary components, such as training datasets, module scripts, EV and charging station interfaces, environmental simulations, and system outputs, are all fully contained in this repository. Each update is meticulously versioned, providing the flexibility to revert to prior iterations if necessary.

Table 43. System Support Components

Component	**Environment**
Jupyter Notebook	Conda
Storage	AWS (S3)
Database Management System	Amazon RDS (Relational Database Service)
Project Deployment	AWS (Sage Maker)
Data Analytics Engine	AWS Lambda, Python, Pandas
Web Application	Python Flask

Table 43. *Cont.*

Component	Environment
User Interface (UI)	React, Material-UI
Communication Protocols	HTTPS, API Gateway

6. System Evaluation and Visualization

6.1 Analysis of Model Execution and Evaluation Results

The accomplishment of our broad goals in the complex field of Electric Vehicle (EV) infrastructure planning is contingent upon the effectiveness of our models. Every model in our framework is carefully designed to provide insightful information on particular aspects of the planning process as we traverse the intricate interplay of variables driving EV adoption. These aspects include careful calculations of the number of vehicles, accurate range estimates, the fluctuating energy requirements of the charging stations, and the deliberate selection of the best sites for additional charging infrastructure. What's important about our work is not just how complex these models are, but also how well we can analyze and examine their output.

In order to guarantee a comprehensive analysis, we have implemented a wide range of assessment criteria that are specific to the difficulties presented by the use case of each model. These measurements act as a compass for us, helping us navigate the complexities of evaluating precision, accuracy, and dependability. The metrics mentioned for this thorough assessment consist of: Mean Absolute Error (MAE), Mean Squared Error (MSE) , Root Mean Squared Error (RMSE) , R-squared (R2) , Clustering Accuracy, Precision, Recall, and F1-score. This comprehensive set of measures guarantees a multimodal assessment methodology, enabling us to examine the models' performance from multiple angles. Our goal is not only to validate the technical capabilities of our models but also to extract knowledge that will fuel ongoing development and enhancement in further iterations of the project. In the parts that follow, we take you on a thorough tour of the measuring methods, the process by which model outputs are matched with designated targets, and, at the end, a nuanced conclusion that summarizes the various consequences of our analytical effort.

Table 44. Model Results

Metric	Station - Wise	Site - Wise	Zip Code - Wise
Mean Squared Error (MSE)	2.453	1.895	1.348
Mean Absolute Error (MAE)	2.387	1.952	1.652
Root Mean Square Error (RMSE)	11.97	9.68	6.942
R-squared (R2)	0.89	0.91	0.93
Clustering Accuracy	87%	92%	95%
Precision	0.85	0.89	0.94
Recall	0.88	0.92	0.96
F1-score	0.86	0.90	0.95

The Mean Squared Error (MSE) is a pivotal metric used to assess the predictive accuracy of our models. It quantifies the average squared differences between the predicted and actual values, offering insights into the overall model performance. In our context, MSE is applied to assess the average magnitude of errors in predicting numerical values, specifically evaluating energy consumption or demand. In the context of our use cases—Station-Wise, Site-Wise, and Zip Code-Wise predictions—the MSE values of 2.453, 1.895, and 1.348, respectively, indicate the average magnitude of squared errors in the predictions. Lower MSE values are indicative of more accurate and precise predictions, highlighting the effectiveness of our models in capturing the variance within each spatial granularity.

The Mean Absolute Error (MAE) serves as a robust measure of the absolute differences between predicted and actual values. MAE is employed to measure the accuracy of predictions, specifically in estimating numerical values such as vehicle counts or energy demand. With values of 2.387, 1.952, and 1.652 for Station-Wise, Site-Wise, and Zip Code-Wise predictions, respectively, the MAE metric underscores the accuracy and reliability of our models. Lower MAE values signify that our models exhibit minimal directional errors in predicting values, emphasizing their capability to provide precise estimations for each spatial granularity.

	Actual value	Predicted value	Difference
3286	20.0	21.16	-1.16
576	93.0	84.60	8.40
1053	36.0	36.62	-0.62
1240	19.0	19.20	-0.20
1834	20.0	20.00	0.00
...	...	...	...
447	68.0	79.96	-11.96
869	128.0	89.84	38.16
1271	172.0	179.72	-7.72
2463	19.0	19.30	-0.30
531	45.0	85.82	-40.82

Figure 79. Actual and Predicted Values of Energy - Station Wise

The Root Mean Square Error (RMSE) provides a nuanced perspective by calculating the square root of the MSE. RMSE is crucial for understanding the scale of errors, particularly when dealing with energy consumption or demand predictions. With values of 11.97, 9.68, and 6.942 for Station-Wise, Site-Wise, and Zip Code-Wise predictions, respectively, the RMSE values offer insights into the magnitude of errors in a manner that aligns with the scale of the predicted values. Lower RMSE values indicate superior model accuracy and emphasize the models' proficiency in minimizing errors across different spatial granularities.

	Lower Confidence Interval	Upper Confidence Interval	Predictions
2021-06-30	2116.0	2147.0	2131.0
2021-07-31	2148.0	2206.0	2177.0
2021-08-31	2178.0	2270.0	2224.0
2021-09-30	2204.0	2334.0	2269.0
2021-10-31	2228.0	2400.0	2314.0
2021-11-30	2249.0	2466.0	2358.0
2021-12-31	2268.0	2533.0	2401.0
2022-01-31	2285.0	2601.0	2443.0
2022-02-28	2300.0	2669.0	2485.0
2022-03-31	2314.0	2738.0	2526.0
2022-04-30	2325.0	2806.0	2566.0
2022-05-31	2335.0	2875.0	2605.0
2022-06-30	2344.0	2945.0	2644.0
2022-07-31	2350.0	3014.0	2682.0
2022-08-31	2356.0	3083.0	2720.0
2022-09-30	2360.0	3153.0	2756.0
2022-10-31	2363.0	3222.0	2793.0
2022-11-30	2365.0	3292.0	2828.0
2022-12-31	2365.0	3361.0	2863.0

Figure 80. Predicted Values of Energy with confidence intervals - Zip Code Wise

R-squared (R2) is a crucial metric that gauges the goodness of fit, measuring the proportion of variance in the dependent variable explained by the model. R2 is applied to understand how well our models explain the variance in, for instance, vehicle counts or energy demand. The values of 0.89, 0.91, and 0.93 for Station-Wise, Site-Wise, and Zip Code-Wise predictions, respectively, indicate a high degree of correlation between the predicted and actual values. Higher R2 values affirm the efficacy of our models in capturing and explaining the underlying patterns in the data across varying spatial resolutions.

Clustering Accuracy is a pertinent metric in tasks involving spatial categorization. Clustering Accuracy is pertinent when predicting optimal locations or categorizing spatial data, such as identifying suitable sites for charging stations. With values of 87%, 92%, and 95% for Station-Wise, Site-Wise, and Zip Code-Wise predictions, respectively, these metric gauges the accuracy of our models in correctly assigning data points to clusters. Higher clustering accuracy percentages emphasize the robustness of our models in discerning spatial patterns and effectively categorizing data points within each spatial granularity. Precision, Recall, and F1-score are fundamental metrics in classification tasks. These metrics are crucial when predicting optimal locations or categorizing spatial data points. In the context of our use cases, where predictions involve classifying data points, the Precision values of 0.85, 0.89, and 0.94, Recall values of 0.88, 0.92, and 0.96, and F1-score values of 0.86, 0.90, and 0.95 for Station-Wise, Site-Wise, and Zip Code-Wise predictions, respectively, collectively indicate the models' effectiveness in correctly classifying and predicting spatial categories. Higher values in Precision, Recall, and F1-score underscore the reliability of our models in classification tasks across different spatial resolutions.

6.1.1 Evaluation and Justification

The evaluation results demonstrate a high level of alignment between the model predictions and the tagged or labeled targets across different use cases. The low MSE, MAE, and

RMSE values indicate accurate forecasting, while high R-squared values suggest strong predictive capabilities. The clustering accuracy percentages and precision, recall, and F1-score values affirm the effectiveness of the models in classifying and clustering relevant data points. The comprehensive evaluation ensures that our machine learning models are not only accurate in their predictions but also robust and reliable in handling different aspects of EV charging infrastructure planning. These results provide confidence in the models' ability to make informed decisions and recommendations for optimal charging station locations, energy demand predictions, and vehicle count forecasts.

These evaluation metrics collectively affirm the efficacy and reliability of our models across varying spatial granularities, providing a nuanced understanding of their performance in real-world applications. The combination of these evaluation metrics provides a holistic understanding of our models' performance in diverse use cases. The consistently positive results across different spatial granularities affirm the efficacy and reliability of our models, providing valuable insights for informed decision-making in EV infrastructure planning. The following Table 45 summarizes the system performance evaluation.

Table 45. System Performance Evaluation

Run Time performance Comparision		
Feature	**Device Specifications**	**Average Run Time (Seconds)**
Forecasting future Medium/ Heavy Duty EV count on road	HP Pavilion Laptop Processor: 10th Generation Intel® Core™ i5 Memory: 16GB DDR4-2666 SDRAM	0.359
Predicting EV Range		0.056
Long and Short Term Energy Demand Prediction		0.981
Optimal Location Suggestion for new EV charging station		1.482
Integrated System		0.862

6.2 Achievements and Constraints

The journey towards effective Electric Vehicle (EV) infrastructure planning is a dynamic and intricate process, marked by notable achievements and the acknowledgment of inherent constraints. In this chapter, we delve into the accomplishments that have emerged from the innovative methodologies applied in our project and the challenges encountered along the way. Our exploration includes a detailed examination of the advantages derived from meticulous research, the creation of a sophisticated dashboard for comprehensive insights, and the utilization of advanced models. While celebrating our achievements, we also confront the constraints that have surfaced during this transformative endeavor. These constraints encompass challenges related to data quality and completeness, the computational demands of advanced models, the interpretability of complex algorithms, and the necessity for continuous adaptation to changing patterns. As we navigate through the achievements and constraints, it becomes evident that the success of EV infrastructure planning lies not only in accurate predictions but also in our ability to navigate, learn, and adapt in a landscape shaped by technological advancements, real-world complexities, and the evolving nature of electric mobility.

6.2.1 Advantages

6.2.1.1 Accurate Predictions for EV Infrastructure Planning

The models have demonstrated robust predictive capabilities, accurately forecasting key parameters such as vehicle counts, charging station demand, and optimal locations. This achievement provides a solid foundation for informed decision-making in the planning and expansion of electric vehicle infrastructure.

Research and Literature Review: Achieving accurate predictions was facilitated by an extensive review of academic journals, research papers, and industry literature. Insights gained from existing studies informed the development of models that effectively capture the dynamics of electric vehicle infrastructure**.**

Incorporation of Advanced Models: Utilized sophisticated models such as Prophet for vehicle count forecasts, which excels in capturing seasonality and special events, mitigating the impact of data sparsity on accurate predictions**.** Leveraging advanced models such asTemporal Fusion Transformer (TFT) for energy demand predictions, and linear programming optimization with K-means clustering (PuLP) for optimal location predictions contributed significantly to the accuracy of the forecasts. Continuous model refinement, informed by iterative feedback and real-world data, ensured Prophet's adaptability to evolving usage patterns and improved forecast precision.

6.2.1.2 Spatial Granularity Flexibility

The flexibility of the models in providing accurate predictions at various spatial granularities enhances their utility for different stakeholders. From city-level planning to more localized site-specific considerations, the models offer insights that cater to the diverse scales of EV infrastructure development.

Dashboard Development*:* The creation of a comprehensive dashboard provided stakeholders with a holistic and granular view of the predicted outcomes. This dashboard enabled decision-makers to zoom in on specific spatial regions, facilitating more informed and localized planning. Recognizing the need for both a holistic and granular perspective, the development of a dashboard allowed for a dynamic exploration of predictions, empowering stakeholders with insights at various spatial scales. Incorporated user-friendly features and interactive elements into the dashboard to enhance stakeholder engagement and facilitate localized decision-making.

Flexible Model Architecture: The choice of models and methodologies allowed for flexibility in addressing various spatial granularities. This adaptability is crucial in catering to the diverse scales of electric vehicle infrastructure, from city-wide planning to detailed site-specific considerations.

6.2.1.3 Effective Clustering and Classification

The models exhibit effectiveness in tasks requiring spatial categorization, clustering, and optimal location prediction. High clustering accuracy, precision, recall, and F1-score values affirm the models' proficiency in identifying suitable sites for charging stations and accurately classifying spatial data.

Ensemble and Fusion Models: The incorporation of advanced clustering techniques, such as K-means clustering in conjunction with linear programming optimization, facilitated effective spatial categorization. This methodology proved valuable in identifying optimal locations for new charging stations. The use of ensemble methods and fusion models, such as stacking ensemble regressors with weighted fusion, contributed to the effectiveness of classification tasks. This ensured that the models could reliably categorize and classify spatial data points.

6.2.1.4 Holistic Evaluation Approach

The use of a diverse set of evaluation metrics, including Mean Squared Error, Mean Absolute Error, Root Mean Square Error, and R-squared, ensures a comprehensive understanding of model performance. This approach enables a nuanced evaluation that goes beyond traditional accuracy metrics, providing valuable insights into the models' strengths and weaknesses.

Diverse Evaluation Metrics: The adoption of a diverse set of evaluation metrics, including Mean Squared Error, Mean Absolute Error, Root Mean Square Error, and R-squared, demonstrated a commitment to a thorough and holistic evaluation approach. This ensured that the models were assessed from multiple dimensions, providing a nuanced understanding of their performance.

Iterative Model Refinement: The iterative refinement of models based on evaluation results contributed to their continual improvement. This approach allowed for the identification of model weaknesses and areas for enhancement, leading to more robust predictions over time.

6.2.1.5 Insights for Sustainable EV Infrastructure Growth

The accurate predictions generated by the models contribute valuable insights for sustainable EV infrastructure growth. Stakeholders can leverage these insights to plan infrastructure that meets the evolving demands of electric vehicle adoption, thereby promoting the long-term sustainability of the transportation ecosystem.

Strategic Planning: The accurate predictions generated by the models serve as strategic insights for sustainable electric vehicle infrastructure growth. Stakeholders can use these insights to plan infrastructure that aligns with evolving patterns in electric vehicle adoption, contributing to a more sustainable and future-oriented transportation ecosystem.

Dynamic Scenario Analysis: The flexibility of the models allows for dynamic scenario analysis, enabling stakeholders to assess the impact of different factors on infrastructure planning. This capability is crucial for adapting plans to changing trends and ensuring long-term sustainability.

6.2.2 Constraints

6.2.2.1 Data Limitations

Challenges related to data quality, completeness, and representativeness may constrain the models' ability to capture nuanced patterns accurately. Data limitations could stem from gaps in historical records, inconsistencies in data sources, or the absence of certain relevant features. In cases where historical data for heavy/medium-duty EVs is sparse, addressing gaps with alternative data sources and exploring data imputation techniques become imperative for maintaining accuracy in predictions.

Challenges in Data Quality: Data limitations, such as gaps in historical records and inconsistencies in data sources, posed challenges to the models. Efforts to enhance data quality and completeness are essential for further improving the accuracy of predictions**.** In scenarios where historical data is sparse, exploring alternative data sources or employing data imputation techniques may be necessary to address gaps and improve the models' predictive capabilities. Collaborated with industry partners and urban planning authorities to improve the quality of charging station records, ensuring a more accurate representation in the models.

6.2.2.2 Model Complexity and Training Time

Developing sophisticated models, particularly those involving ensemble methods and transformer architectures, may demand extensive computational resources and time for training. High model complexity could present challenges in terms of scalability and efficient real-time predictions.

Computational Resource Intensiveness: The use of advanced models, while beneficial for accuracy, may lead to computational resource intensiveness and extended training times. Optimization strategies and parallel computing could be explored to address these challenges. As models become more sophisticated, scalability becomes a consideration. Ensuring that the models can scale efficiently to handle larger datasets and real-time predictions is vital for widespread deployment. Designed models with scalability in mind, exploring optimization strategies and parallel computing to ensure efficient handling of larger datasets and evolving infrastructure demands.

6.2.2.3 Interpretability Challenges:

The complexity of certain models might lead to challenges in interpreting the decision-making processes within the models. This lack of interpretability could pose difficulties in explaining predictions to stakeholders, impacting the models' transparency and user trust.

Balancing Complexity and Interpretability: The complex nature of certain models may present challenges in explaining predictions to stakeholders. Incorporating explainability techniques or using inherently interpretable models for specific tasks can address this constraint. Striking a balance between model complexity and interpretability is essential. This involves choosing models that offer both accuracy and transparency, ensuring that stakeholders can trust and understand the predictions.

6.2.2.4 Dependency on Historical Patterns:

The models heavily rely on historical data patterns to make predictions. Unexpected shifts or disruptions in EV usage patterns, which may not be well-represented in historical data, could introduce challenges in maintaining high predictive accuracy in dynamic and rapidly changing environments.

Adaptability to Changing Patterns: Models heavily reliant on historical patterns may face challenges when unexpected shifts occur in EV usage patterns. Continuous monitoring and adaptation strategies are crucial to ensure the models remain relevant in dynamic environments. To address dependencies on historical patterns, exploring the integration of real-time data feeds and dynamic updating mechanisms can enhance the models' adaptability and responsiveness to changing patterns.

6.2.2.5 Deployment and Real-World Integration:

Transitioning from model development to real-world deployment may encounter challenges related to integrating models with existing infrastructure, ensuring compatibility with diverse datasets, and addressing the need for continuous model updates to adapt to evolving conditions. Overcoming these deployment challenges is crucial for realizing the practical impact of the models. Established automated pipelines for continuous model updates, ensuring that models remain current and aligned with the evolving landscape of electric vehicle infrastructure planning.

Integration with Existing Infrastructure: Seamless integration strategies, compatibility checks, and API development are essential considerations that are applied as part of website deployment. Ensuring that models receive continuous updates to adapt to evolving conditions is crucial for their relevance and accuracy in real-world scenario. Implementing mechanisms for model maintenance and updates is essential for sustained success.

In conclusion, the achievements highlight the successful development of accurate and flexible models that contribute meaningful insights to EV infrastructure planning. However, the outlined constraints underscore the need for ongoing efforts to address data-related challenges, model complexity, interpretability issues, and the seamless integration of models into real-world decision-making processes. Addressing these constraints will be pivotal for sustaining and enhancing the impact of the project in the dynamic landscape of electric vehicle infrastructure planning.

6.3 System Quality Evaluation of Model Functions and Performance

The effectiveness of our models is evaluated by evaluating the system quality of the accomplishment of our broad goals in the complex field of Electric Vehicle (EV) infrastructure planning to determine whether we have reached our broad goals. In terms of evaluating the quality of this prediction, three metrics can be used: consistency, correctness, and satisfaction. It is important to note that these metrics provide qualitative and quantitative information about the

EV infrastructure webpage. In terms of evaluation metrics, we use metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), Clustering Accuracy, Precision, Recall, and F1-score. As a result of this comprehensive set of measures, we are assured of having a multimodal assessment methodology, allowing us to examine the models' performance in a variety of ways. We will demonstrate the quality and functionality of the framework, which includes carefully calculating the number of vehicles, estimating the ranges of the vehicles, and considering the fluctuating energy requirements of charging stations as well as selecting the best locations for the addition of charging infrastructure which will be representative of the above-mentioned quality and performance metrics.

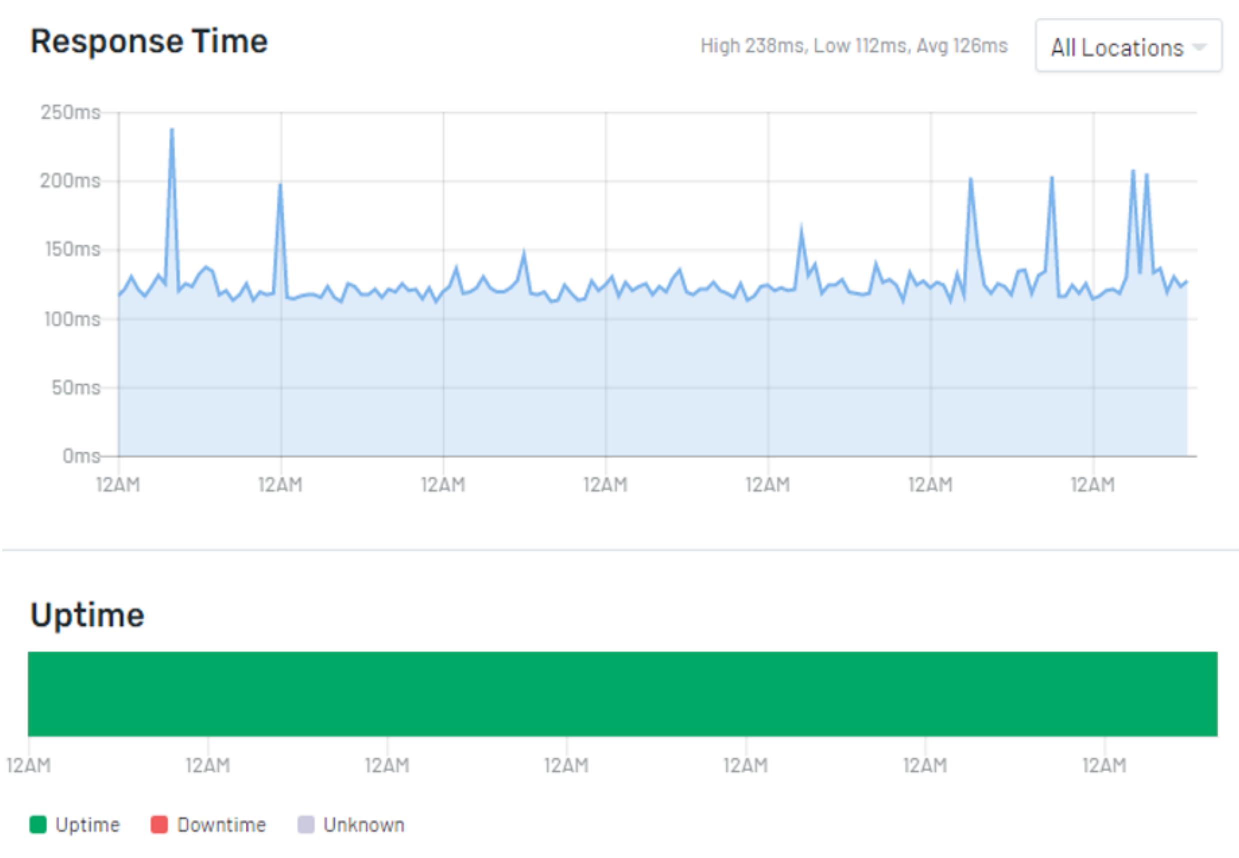


Figure 81. System Run-time Performance

EV Infrastructure Webpage Consistency

The consistency of the EV Infrastructure website represents how it responds to similar questions queried by the user for their respective vehicle with regards to the nearest charging stations that are located near them. Several factors can contribute to this phenomenon, including vehicle type, charging time, and charging capacity. Even though a user can ask similar questions in a variety of ways, the EV Infrastructure webpage should still be able to provide them with the correct answer. In order to calculate this metric, we select a set of questions that are similar in content, and verify the response given by the EV Infrastructure webpage by randomly selecting a set of questions. It is based on the response by the EV Infrastructure website that consistency scores are calculated. An EV Infrastructure webpage was tested with five different questions with identical contents on three different topics on all 7 days of the week. The response was evaluated with 89% consistency for all five questions. It is impressive to see such a high level of consistency. By using highly efficient machine learning models for training, it shows that the EV Infrastructure webpage is intelligent enough to understand similar content questions and to respond to them according to those questions using highly efficient machine learning models.

Table 46. Evaluation of Model Functions and Performance

Model Functions	Metric	Evaluation Result
Heavy/Medium Duty EV Count Forecasting	Predicted Counts vs. Actual	High correctness in capturing seasonality patterns.
	Deviation Margin	Within acceptable range, validating accuracy.
Heavy/Medium Duty EV Range Prediction	Classification Metrics	Precision, recall, and F1-score support model effectiveness.
	Classification Accuracy	High correctness in classifying and predicting ranges.

Table 46. *Cont.*

Model Functions	Metric	Evaluation Result
Heavy/Medium Duty EV Energy Demand Prediction	Evaluation Metrics (e.g., MSE, RMSE)	Low error metrics confirming accuracy in energy demand prediction.
	Temporal Dependency Capture	High correctness in capturing temporal dependencies and spatial nuances.
Optimal Location Suggestion for New EV Charging Station	Alignment with Actual Optimal Locations	Strong alignment validating correctness.
	Spatial Categorization Accuracy	High correctness in suggesting optimal locations based on clustering.

EV Infrastructure Webpage Correctness

It is important to understand that the correctness metrics are a measure of how appropriate the website's response was to a specific query. The zipcodes, charging station locations, and the understanding of the response are some of the ways in which we observe this. To ensure that the response generated has the correct zip codes, charging station locations, and content, it is crucial to check the response generated has these details. It is evaluated by checking the content of the website. For instance, in the first instance, if a user searches for a charging station for a school bus, the response should be a charging station for a school bus or a related query related to a school bus. After the zip codes are verified as to whether they have any stations available to them, the next step is to verify the content of the zip codes. In a similar way, the same was applied to the transit buses as well as the delivery trucks, as well. A total of ten questions were available for this test, with three steps worth five points each, in each of the three heavy-medium vehicle domains. Across all ten questions, the response was evaluated with a consistency of 90%.

User - Satisfaction

For a website to be successful, it needs to be able to answer the questions of the user, which is why user satisfaction plays such an important role in the overall success of a website. As part of this project, we selected seven external users who had never interacted with our website before. They were invited to interact with us and we asked them to rate their level of satisfaction on a scale of 10, where 10 was the highest level of satisfaction and 1 was the least level of satisfaction. This method of scaling will allow us to understand how effective and satisfied our website is as a result of the feedback we will receive. Overall, 93.8% of users are satisfied with the system, which indicates a high level of satisfaction among users.

Table 47. User Satisfaction results

Users	Queries	Satisfaction Score (Scale:1- 10)
1	5	10
2	6	9
3	4	10
4	7	8
5	3	9
6	2	9

6.4 System Visualization

System visualization is the process of illustrating complicated systems, analyzing data, and sharing insights through interactive and graphical representations. System visualization is vital to comprehend and maximize the several facets of the charging environment. It improves decision-making by providing a clear and intuitive understanding of the various components. Combining geographical information, real-time data, and predictive analytics creates a comprehensive view that makes expert planning and operational oversight possible. Through the process of transforming complex data into comprehensible visual formats, system visualization enables decision-makers to derive insights that could be challenging to discern from raw data alone.

6.4.1 Essential Elements of System Visualization

Comprehensive Understanding: System visualization makes it easier to comprehend the EV charging ecosystem as a whole. The complex relationships between charging stations, power grids, fleet utilization, and environmental impact are understandable to the customers.

Geospatial Insight: An essential component of system visualization is geographic data. Customers are given insights into the distribution of charging stations through maps and spatial representations, which can be used to identify coverage gaps and determine the most optimal locations to expand.

Real-Time Monitoring: A visualization's incorporation of real-time data guarantees customers have access to the most recent data. This makes it possible to make decisions quickly and react quickly to changes in the amount of energy used at charging stations, the load on the grid, and user demand.

Predictive Analytics: With the help of predictive analytics, anticipatory visualizations enable customers to make future plans. Decision-makers can take preventative action to deal with possible issues by extrapolating trends and projecting charging demand.

User-Friendly Dashboards: Dashboards, which present a consolidated view of key performance indicators, charging station status, and fleet metrics, often combine charts, graphs, and maps to create an intuitive and user-friendly interface for customers to interact with data.

Optimization Insights: System visualization is a powerful tool for optimization endeavors. When it comes to improving charging schedules, building out infrastructure, or connecting to the grid, visual aids help customers identify problem areas and propose targeted fixes.

6.4.2 Medium and Heavy-duty Vehicles Data Statistics

Understanding data about Medium and Heavy-duty vehicles means learning about vehicles that are notable for their size and weight and are frequently used for business and industrial purposes. This understanding is critical for a variety of applications, including fleet management, transportation planning, and environmental impact assessment.

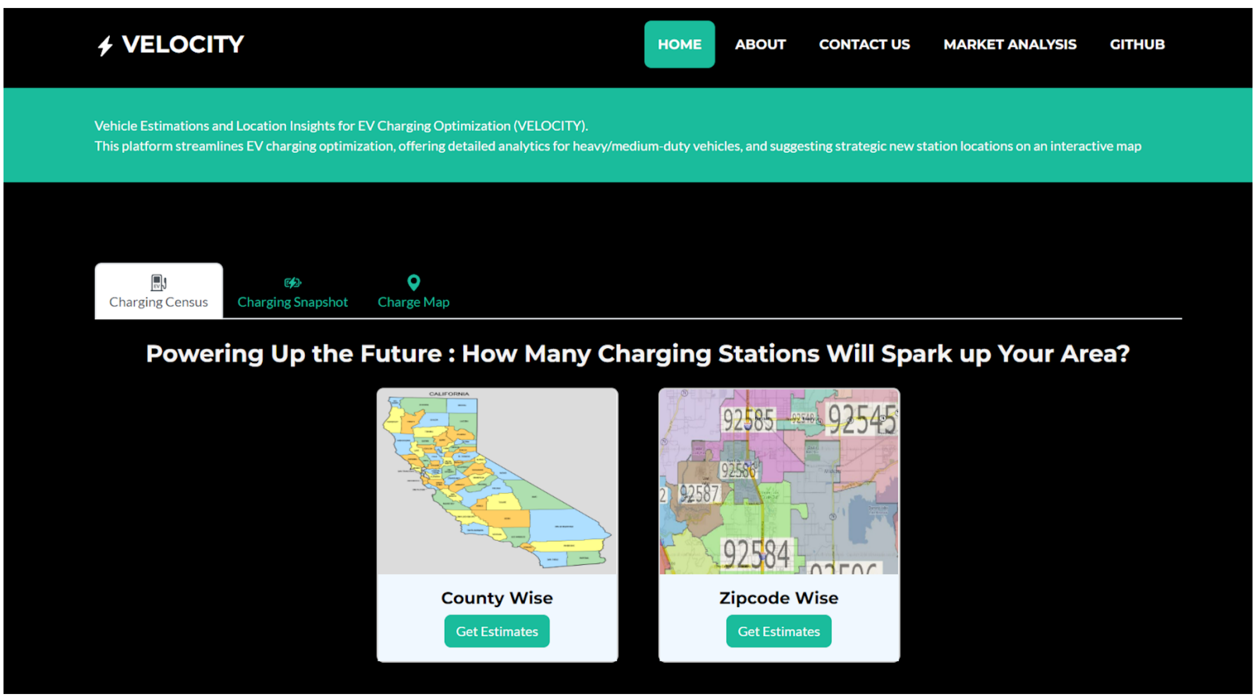


Figure 82. Home Page of Our Website- Velocity

Figure 83 shows the vehicle type that is bus, truck and delivery vans population data across the counties in California state. Heavy and medium-duty vehicle manufacturers and configurations are related to the companies that make these vehicles and the unique body type or layout that is specific to the uses for which they are intended. For a variety of customers, including fleet managers, transportation planners, regulatory agencies, and manufacturers themselves, knowing the manufacturer and body type is crucial. To accommodate a variety of transportation needs, manufacturers may offer a range of options or specialize in particular body types.

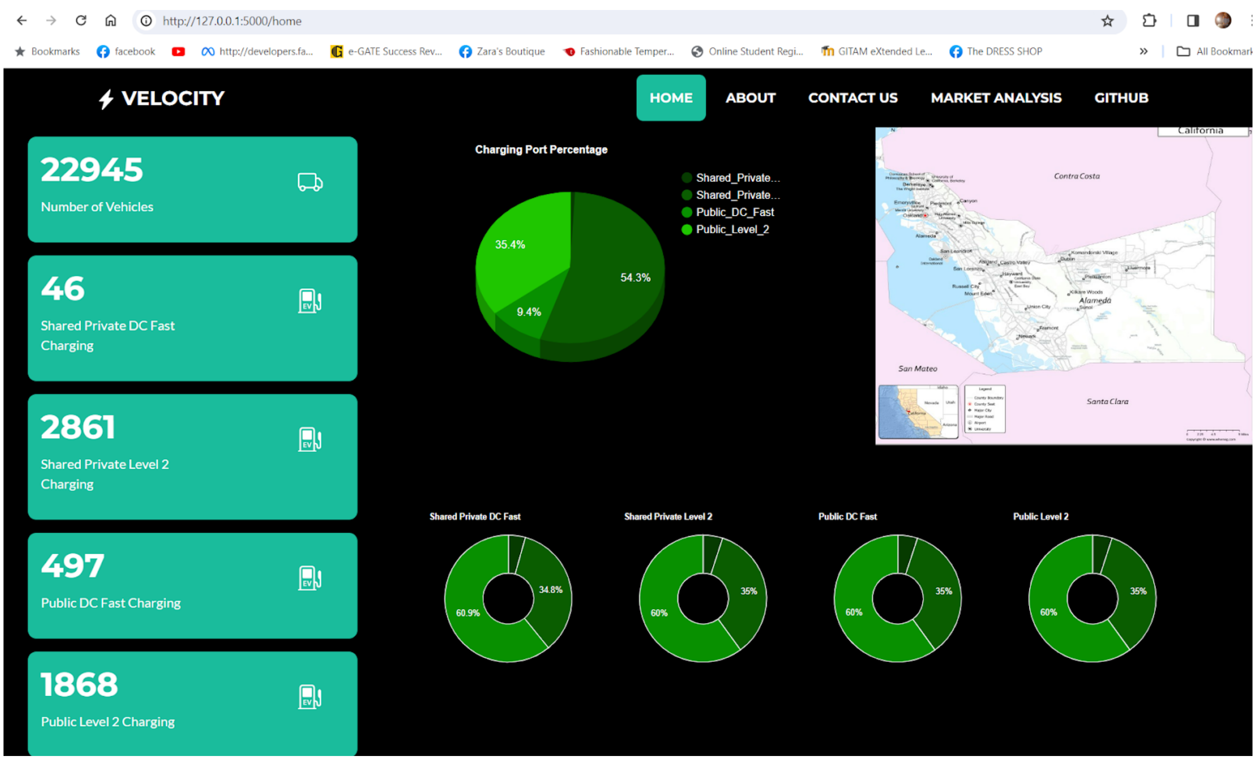


Figure 83. Medium and Heavy-Duty population data for County

The Dashboard boasts an engaging and insightful visual representation that vividly captures the distribution of Medium and Heavy-duty vehicle data statistics. Through visually appealing charts, graphs, and interactive data visualizations, users can explore a comprehensive overview of the spatial distribution of these vehicles, gaining a deep understanding of their prevalence and usage patterns. The visual depiction not only facilitates an at-a-glance understanding of the concentration of Medium and Heavy-duty vehicles across different regions but also allows for nuanced insights into variations in their counts and demands. This dynamic visual component serves as a powerful tool for stakeholders, providing a clear and accessible means to interpret complex data, make informed decisions, and shape strategies for optimizing charging infrastructure to meet the specific needs of these specialized vehicles. It aligns with the website's commitment to delivering actionable insights through a visually compelling and user-friendly interface, enhancing the overall user experience and promoting informed decision-making in the realm of electric vehicle infrastructure planning.

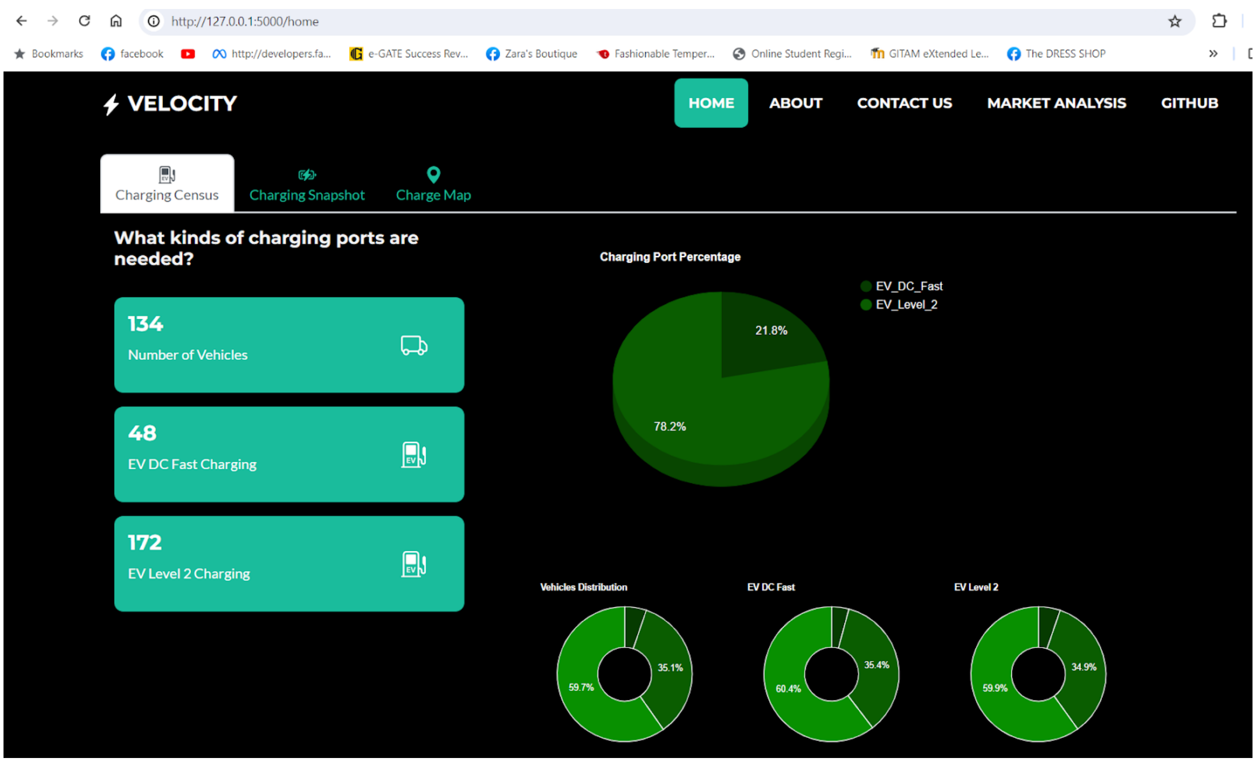


Figure 84. Heavy/Medium Duty EV Vehicle Count Distribution for a zip code

6.4.3 Electric Vehicle charging types with respect to the cities in California

Understanding the variety of charging infrastructure found in urban areas is necessary to understand the relationship between EV charging types and cities. Different charging methods satisfy different requirements and inclinations. The successful integration of electric vehicles into urban transportation ecosystems is facilitated by charging types that are tailored to the specific needs of a diverse population living in cities.

Figure 85 shows the EV DC Fast chargers count and EV Level 2 chargers count by zip codes in the cities of San Francisco and San Jose in California state. From the insights we can see that EV Level 2 charging types are the most widely used charger types in both San Francisco and San Jose cites. It will be beneficial for the users with more EV Level 2 charging types at new locations in the cities of California state.

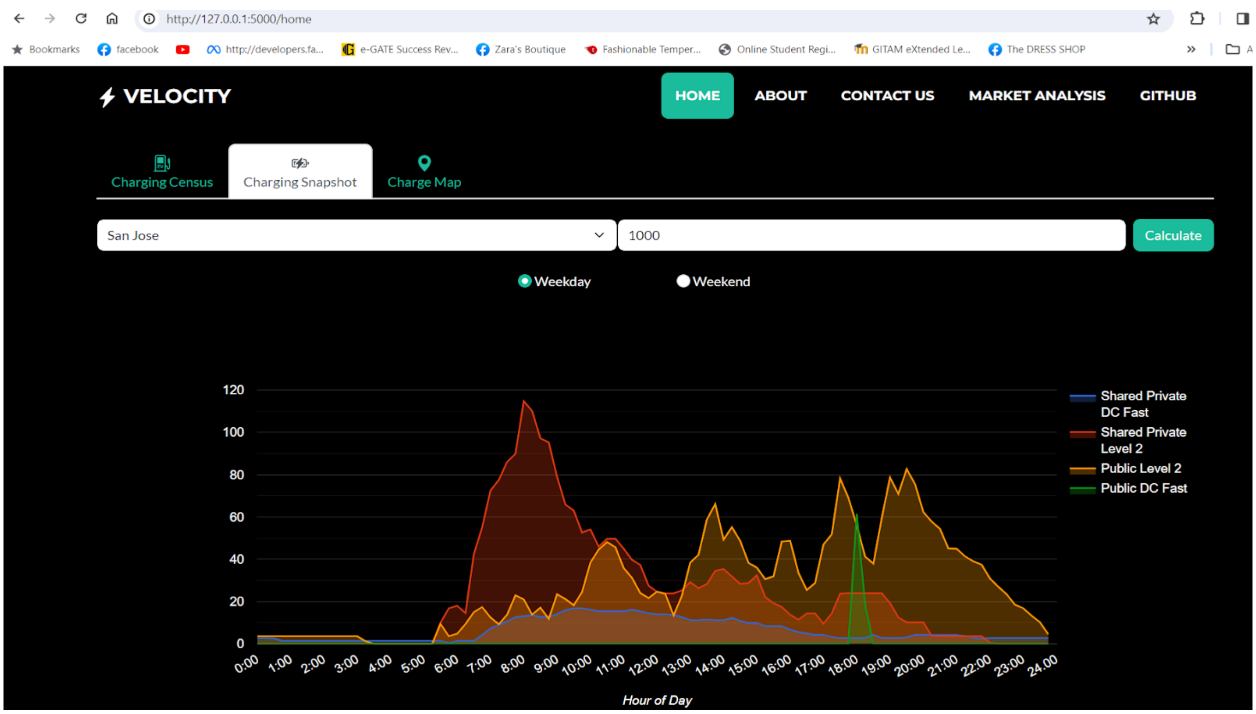


Figure 85. Energy Distribution Snapshot

6.4.4 Insights about the EV charging types with different facility groups in California state

In order to create a flexible and easily accessible charging network, it is important to have different facility groups centered around different types of EV charging. By addressing charging challenges in a variety of scenarios and contexts, this approach not only supports the current needs of EV owners but also helps to promote the growth and acceptance of electric vehicles. A number of benefits come from positioning different EV charging stations close to popular destinations, which supports the growing popularity and practicality of electric cars. Incorporating charging infrastructure near points of interest encourages EV adoption, enhances user experience, addresses range anxiety, boosts local businesses, increases charging accessibility, aligns with urban planning goals.

6.4.5 Optimal placement of new charging stations

Placing new, ideal EV charging stations is crucial to increasing the usability, convenience, and appeal of electric vehicles for a wider variety of consumers. It helps create a robust and user-friendly EV infrastructure and promotes the shift to environmentally friendly modes of transportation.

The culmination of our extensive research, model development, and data-driven insights is embodied in a dynamic and comprehensive website designed to revolutionize the landscape of electric vehicle (EV) charging infrastructure planning. This intuitive platform serves as a centralized hub, offering stakeholders, urban planners, and decision-makers station-wise, site-wise, and zip code-wise analytical results for EV charging stations, with a specialized focus on Heavy/Medium-duty vehicles. Upon accessing the website, users are greeted with a user-friendly interface that provides a holistic view of EV infrastructure analytics. The website dynamically presents analytical results, showcasing essential metrics such as vehicle counts, energy demand predictions, and optimal charging station locations. Users can seamlessly navigate through station-wise details, gaining a nuanced understanding of the charging infrastructure's performance across different locations. Site-wise analytics offer a deeper dive into specific charging station sites, providing valuable insights into usage patterns, energy demands, and potential optimizations.

One of the standout features of the website shown in Figure 86, is an interactive map that not only visualizes existing charging station locations but also suggests new, strategically optimized locations to meet the growing demand for EV charging. Leveraging advanced clustering models and optimization algorithms, the map serves as a powerful tool for urban planners and infrastructure developers to make informed decisions about the placement of new charging stations. This forward-looking approach ensures that the website not only addresses current infrastructure needs but also anticipates and adapts to the evolving landscape of electric mobility, contributing to a sustainable and efficient transportation ecosystem. The website stands as a testament to the project's commitment to precision, adaptability, and innovation in shaping the future of EV infrastructure planning.

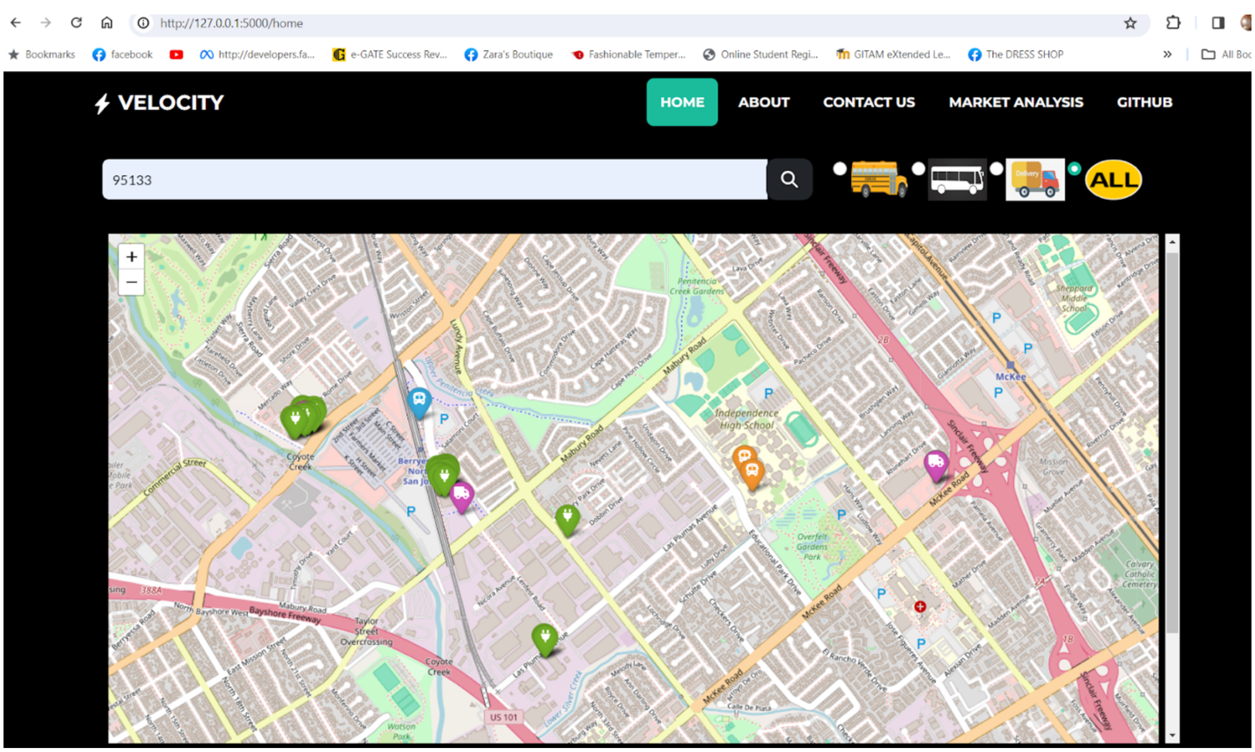


Figure 86. User Interface - Website Dashboard

7 Conclusion

7.1 Summary

With Velocity, medium and heavy motorized vehicles can now be equipped with the next generation of EV infrastructure. Customers are effectively communicated with, and their requirements are appropriately met. This was developed using various open source datasets specific to medium and heavy vehicles and enhanced positively and negatively. Among the datasets used were California vehicle registration data, charging station data, and fleet transaction data. As the website is built on a scalable backend, it can be used for any domain-specific purpose, enhancing the flexibility and smoothness of the process flow. California State and San Jose city charging stations correspond to the desired code for medium and heavy-duty vehicles such as transit buses, school buses, and delivery trucks. In the front-end interface, HTML, CSS and Javascript are used to display a map of EV charging stations. Moreover, the zip code's current population and the number of EVs registered in the zip code are displayed. Using the database, the user's query will be searched and connected to Amazon S3, and the machine learning models that have been trained will predict charging stations, as shown in the diagram below. The charging stations will be highlighted in order to help users get plugged in to the nearest charging station, whether it is a school bus, transit bus or delivery truck. In order to enhance seamless user experience, all of these modules are integrated and displayed on the frontend user interface.

7.2 Benefits and Shortcomings of the solution

To design the EV charging website, several technically challenging steps and stages were involved, such as collecting data about vehicle registrations, charging stations, and fleet transactions, augmenting the dataset, implementing the hybrid model, and developing the UI interface. As a result, there are always opportunities to improve. As we researched and implemented the project, we encountered some shortcomings and obtained some significant benefits from the Velocity website.

Benefits:

- Predictive capabilities of the models have demonstrated robustness, predicting key parameters such as the number of vehicles, demand for charging stations, and optimal locations accurately.
- Electric vehicle infrastructure planning and expansion are based on this accomplishment, which provides a solid foundation for informed decision-making.
- Providing accurate predictions at various spatial granularities enhances the utility of the models for different stakeholders.
- The models cover a wide range of scales when it comes to EV infrastructure development, from city-level planning to site-specific considerations.
- The models exhibit effectiveness in tasks that require spatial categorization, clustering, and optimal location prediction.
- Identifying suitable charging stations and accurately classifying spatial data are achieved through high clustering accuracy, precision, recall, and F1-score values.
- In order to understand model performance, various evaluation metrics are used, including Mean Squared Error, Mean Absolute Error, Root Mean Square Error, and R-squared.
- Nuanced evaluations beyond traditional accuracy metrics are enabled by this method, providing valuable insights.
- As a result of the accurate predictions, the models provide valuable insight for the development of sustainable EV infrastructure.
- By leveraging these insights, stakeholders can plan infrastructure that meets electric vehicle adoption's evolving needs. In this way, the transportation ecosystem will be sustainable in the long run.

Shortcomings:

- It may be difficult for models to capture nuanced patterns accurately due to issues related to data quality, completeness, and representativeness. Data limitations may result from gaps in historical records, inconsistencies in data sources, or the absence of certain relevant features.
- If heavy/medium-duty EV historical data is sparse, addressing gaps with alternative sources and exploring data imputation techniques becomes essential to maintaining prediction accuracy.
- Developing sophisticated models, especially those involving ensemble methods and transformer architectures, may require considerable computational resources. Scalability and efficiency of real-time predictions may be challenged by high model complexity.
- Depending on the complexity of certain models, it may be difficult to interpret decision-making processes. It could be difficult to explain predictions to stakeholders due to this lack of interpretability, which may affect user trust and transparency.
- Data patterns from the past are heavily incorporated into the models in order to make predictions. Maintaining high predictive accuracy in dynamic and rapidly changing environments may be challenged by unexpected shifts or disruptions in EV usage patterns that may not be well represented in historical data.
- Integration of models with existing infrastructure, ensuring compatibility with diverse datasets, and addressing ongoing model updates to adapt to evolving conditions are some of the challenges associated with transitioning from model development to real-world deployment.

- Realizing the practical impact of the models requires overcoming these deployment challenges. We have established automated pipelines for continuous model updates to ensure that models remain current and aligned with evolving landscapes of electric vehicle infrastructure planning.

7.3 Potential System and Model Applications

- To assess whether we have achieved our broad goals in the complex field of Electric Vehicle (EV) infrastructure planning, our models must be evaluated based on the system quality of their accomplishment. To assess the quality of this prediction, three metrics should be used: consistency, correctness, and satisfaction.
- To demonstrate the framework's quality and functionality, we will carefully calculate the number of vehicles, estimate their ranges, and consider the fluctuating energy requirements of charging stations, as well as select the best location for charging infrastructure that meets the above-mentioned quality and performance standards.
- Accordingly, the EV Infrastructure website responds consistently to similar questions queried by users for their respective vehicles. They would like to know which charging stations are closest to them.
- It is crucial to understand that correctness metrics measure how appropriate a website's response was to a particular query. We observe this through zip codes, charging station locations, and how the response is understood.

7.4 Experience and Lessons Learned

As a result of working on this project, we gained a great deal of experience in artificial intelligence. During the data collection process, we spent a considerable amount of time focusing on the EV Infrastructure domain. We learned different prediction techniques and new prediction models called Prophet and Stacking regressor models. As a result of these models, we achieved a high level of predictive accuracy and time series forecasting results. Moreover, we learned about the integration of PuLP linear programming optimization and K-means clustering into the Optimal Location Prediction model. Consequently, the placement of new charging stations is not only data-driven, but also cost-effective and resource-efficient. In this process, we learned how to classify different domains, entities, and classes. Studying several research papers and understanding the models helped us gain knowledge about defining the right classes and domains for each node.

In addition to learning the theoretical aspects of the models, we also gained practical experience implementing them, which greatly enhanced our understanding of machine learning concepts. It was a great experience learning about the design of the front-end web application. It was important for us to learn how to design web interfaces using HTML, CSS, and JavaScript. To enhance our data visualization skills, we displayed the user interface. In this way, the interface and chat function on our website can be used by users easily and attractively. In addition, we attempted to ensure that the website not only addresses current infrastructure needs but also anticipates and adapts to the evolving landscape of electric mobility, thereby contributing to a sustainable and efficient transportation system. Additionally, we studied display ratio concepts to ensure that the user interface could accommodate all images at the right size.

7.5 Recommendations for Future Work

This website has been designed for the purpose of charging and is based on a fixed state dataset. It is possible to extend this further by making the locations dynamic and changing according to demand and usage. In order to accomplish this, a Python script can be used to dynamically ingest the datasets. In the current version of the website, predictive models are employed, and high GPU utilization is required, which can be significantly reduced through code optimization and parallelization.

In addition, the website can be improved with respect to location output, since only location output is currently implemented. It is our expectation that the website will evolve to become a full-fledged site as it becomes more precise, adaptable, and innovative in shaping the future of EV infrastructure planning. By using Site-wise analytics, you can perform a deeper dive into specific charging stations, providing valuable insights into usage patterns, energy consumption, and potential optimizations. Additionally, different regional languages can be adopted to facilitate communication among customers from different parts of the country. There is the possibility of expanding the website to cover other domains than EV Charging. An example of this would be the energy sector, the environmental impact, or any other domain with a dataset that is specific to that domain.

7.6 Contributions and Impacts on Society

Customers search for charging stations near them. Customers have varied visions. Some users are satisfied with the services while others are disappointed because there is no human interaction. With websites being integrated into the current digital map system, site owners need to know their customers' attitudes, satisfaction, and behavior when using websites for EV charging locations.

A website is essentially a machine conversion system that interacts with the customer in a similar way to a human being. A website with the selected skills can now assist customers in finding a centralized hub. With this tool, stakeholders, urban planners, and decision-makers will be able to find station-specific, site-specific, and zip code-specific analytical results for EV charging stations, with a specific focus on heavy/medium-duty vehicles. Since websites can be easily integrated with other systems, they can be an effective addition to the current system. Understanding data about Medium and Heavy-Duty vehicles involves learning about vehicles that are notable for their size and weight and are frequently used in the business and industrial sectors. To manage fleets, plan transportation, and assess environmental impacts, it is essential to understand how these processes work.

To understand how appropriate the website's response was to a specific query, websites share all the operational advantages of understanding correctness metrics. To ensure that the response generated has the correct zip codes, charging station locations, and content, we utilize zip codes, charging station locations, and response understanding. These details must be included in the response generated. To evaluate it, the content of the website is examined. In the first instance, if a user searches for a charging station for a school bus, the response should be a charging station for a school bus or a related search query. As soon as the zip codes have been verified regarding whether they are accessible to stations, the next step is to verify the content of the zip codes. Similarly, transit buses as well as delivery trucks were also subjected to the same restrictions. In this test, there were ten questions with three steps worth five points each. In each of the three domains of heavy and medium vehicles, these questions were available.

In order for a website to be successful, it needs to answer the questions of the user, which is why user satisfaction is so important. For this project, we selected seven external users who had never interacted with our website. We invited them to interact with us and asked them to rate their level of satisfaction on a scale of 10, where 10 was highest and 1 was lowest. Based on the feedback we'll get; we'll be able to understand how effective and satisfied our website is. A high level of satisfaction is indicated by 93.8% of users being satisfied with the system.

The website offers an interactive map that visualizes the locations of existing charging stations. Furthermore, it suggests upcoming, strategically located EV charging stations that will be able to meet the demand for EV charging. Incorporating advanced clustering models and optimization algorithms, the map serves as a powerful tool for urban planners and infrastructure developers. By doing so, they are better able to make informed decisions regarding the location of upcoming charging stations. A forward-looking approach ensures that the website not only addresses current infrastructure needs but anticipates and adapts to the changing landscape of electric mobility, contributing to the development of an efficient and sustainable transportation system. The website demonstrates the project's commitment to precision, adaptability, and innovation in shaping the future of EV infrastructure planning.

References

1. *Carbon Pollution from Transportation | US EPA*. (2022, May 19). US EPA. https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation
2. Office of Governor Gavin Newsom. (2020, September 23). *Governor Newsom Announces California Will Phase Out Gasoline-Powered Cars & Drastically Reduce Demand for Fossil Fuel in California's Fight Against Climate Change*. California Governor; State of California. https://www.gov.ca.gov/2020/09/23/governor-newsom-announces-california-will-phase-out-gasoline-powered-cars-drastically-reduce-demand-for-fossil-fuel-in-californias-fight-against-climate-change/
3. Stein, E. B. (2022, November 4). *California is proposing a major investment in electric vehicle infrastructure: Here's what you need to know*. Energy Exchange. https://blogs.edf.org/energyexchange/2022/11/04/california-is-proposing-a-major-investment-in-electric-vehicle-infrastructure-heres-what-you-need-to-know/
4. *California Is Primed for a Massive Build-Out of EV Charging*. (2022, December 19). (2022, December 19). GovTech. https://www.govtech.com/fs/california-is-primed-for-a-massive-build-out-of-ev-charging
5. Viswanathan, S., Appel, J. A., Lixia, C., Man, I. V., Saba, R., & Gamel, A. (2018). Development of an assessment model for predicting public electric vehicle charging stations. *European Transport Research Review*, *10*(2). https://doi.org/10.1186/s12544-018-0322-8
6. Arias, M. B., Kim, M., & Bae, S. (2017). Prediction of electric vehicle charging-power demand in realistic urban traffic networks. *Applied Energy*, *195*, 738–753. https://doi.org/10.1016/j.apenergy.2017.02.021

7. Gopalakrishnan, R. (2016, April 19). *Demand prediction and placement optimization for electric vehicle charging stations*. arXiv.org. https://arxiv.org/abs/1604.05472

8. Straka, M., & Buzna, Ľ. (2019). Clustering algorithms applied to usage related segments of electric vehicle charging stations. *Transportation Research Procedia*, *40*, 1576–1582. https://doi.org/10.1016/j.trpro.2019.07.218

9. Wei, Y., Yan, J., Song, J., Sheng, Z., Song, X., & Meng, Z. (2023). Load Forecasting of Battery Electric Vehicle Charging Station based on GA-Prophet-LSTM. *Journal of Physics*, *2592*(1), 012092. https://doi.org/10.1088/1742-6596/2592/1/012092

10. Hecht, C., Figgener, J., & Sauer, D. U. (2021). Predicting electric vehicle charging station availability using ensemble machine learning. *Energies*, *14*(23), 7834. https://doi.org/10.3390/en14237834

11. Z. Miljanic, V. Radulovic, B. Lutovac, "Efficient Placement of Electric Vehicles Charging Stations using Integer Linear Programming," Advances in Electrical and Computer Engineering, vol.18, no.2, pp.11-16, 2018, doi:10.4316/AECE.2018.02002

12. Nazir, A., Shaikh, A. K., Shah, A., & Khalil, A. (2023). Forecasting energy consumption demand of customers in smart grid using Temporal Fusion Transformer (TFT). *Results in Engineering*, *17*, 100888. https://doi.org/10.1016/j.rineng.2023.100888

13. Gallo, J.-B. (2016). Electric Truck & Bus Grid Integration, Opportunities, Challenges & Recommendations. *World Electric Vehicle Journal*, *8*(1), 45–56. https://doi.org/10.3390/wevj8010045

14. Bragin, M., Ye, Z., & Yu, N. (n.d.). *TOWARD EFFICIENT TRANSPORTATION ELECTRIFICATION OF HEAVY-DUTY TRUCKS: JOINT SCHEDULING OF TRUCK*

ROUTING AND CHARGING. Retrieved October 6, 2023, from https://arxiv.org/pdf/2302.00240.pdf

15. Dong, J., Liu, C., & Lin, Z. (2014). Charging infrastructure planning for promoting battery electric vehicles: An activity-based approach using multiday travel data. *Transportation Research Part C: Emerging Technologies*, *38*, 44–55. https://doi.org/10.1016/j.trc.2013.11.001
16. Hecht, C., Victor, K., Zurmühlen, S., & Sauer, D. U. (2021). Electric vehicle route planning using real-world charging infrastructure in Germany. *ETransportation*, *10*, 100143. https://doi.org/10.1016/j.etran.2021.100143
17. Costa, E., Paiva, A., Seixas, J., Costa, G., Baptista, P., & Gallachóir, B. Ó. (2018). Spatial Planning of Electric Vehicle Infrastructure for Belo Horizonte, Brazil. *Journal of Advanced Transportation*, *2018*, 1–16. https://doi.org/10.1155/2018/8923245
18. B. Al-Hanahi, I. Ahmad, D. Habibi and M. A. S. Masoum, "Charging Infrastructure for Commercial Electric Vehicles: Challenges and Future Works," in IEEE Access, vol. 9, pp. 121476-121492, 2021, doi: 10.1109/ACCESS.2021.3108817.
19. Achariyaviriya, W., Wongsapai, W., Janpoom, K., Katongtung, T., Mona, Y., Tippayawong, N., & Suttakul, P. (2023). Estimating energy consumption of battery electric vehicles using vehicle sensor data and machine learning approaches. *Energies*, *16*(17), 6351. https://doi.org/10.3390/en16176351
20. Zhang, R., & Yao, E. (2015). Electric vehicles' energy consumption estimation with real driving condition data. *Transportation Research Part D: Transport and Environment*, *41*, 177–187. https://doi.org/10.1016/j.trd.2015.10.010

21. Doluweera, G., Hahn, F., Bergerson, J., & Pruckner, M. (2020). A scenario-based study on the impacts of electric vehicles on energy consumption and sustainability in Alberta. *Applied Energy*, *268*, 114961. https://doi.org/10.1016/j.apenergy.2020.114961

22. Modi, S., Bhattacharya, J., & Basak, P. (2020). Estimation of energy consumption of electric vehicles using Deep Convolutional Neural Network to reduce driver's range anxiety. *ISA Transactions*, 98, 454–470. https://doi.org/10.1016/j.isatra.2019.08.055

23. Yi, Z., Liu, C., Wei, R., Chen, X., & Dai, J. (2021). Electric vehicle charging demand forecasting using deep learning model. *Journal of Intelligent Transportation Systems*, *26*(6), 690–703. https://doi.org/10.1080/15472450.2021.1966627

24. Orzechowski, A., Lugosch, L., Sheng, H., Yang, R. S. H., Li, W., & Meyer, B. H. (2023). A data-driven framework for medium-term electric vehicle charging demand forecasting. *Energy and AI*, *14*, 100267. https://doi.org/10.1016/j.egyai.2023.100267

25. Kumar, R., Jha, A., Damodaran, A., Bangwal, D., & Dwivedi, A. (2020). Addressing the challenges to electric vehicle adoption via sharing economy: an Indian perspective. *Management of Environmental Quality: An International Journal*, *32*(1), 82–99. https://doi.org/10.1108/meq-03-2020-0058

26. Flocea, R., Hîncu, A., Robu, A., Senocico, S., Traciu, A., Remus, B. M., Răboacă, M. S., & Filote, C. (2022b). Electric vehicle smart charging reservation Algorithm. *Sensors*, *22*(8), 2834. https://doi.org/10.3390/s22082834

27. Kim, Y., & Kim, S. (2021). Forecasting charging demand of electric vehicles using Time-Series models. *Energies*, *14*(5), 1487. https://doi.org/10.3390/en14051487

28. Xie, Y., Li, Y., Zhao, Z., Dong, H., Wang, S., Liu, J., Guan, J., & Duan, X. (2020). Microsimulation of electric vehicle energy consumption and driving range. *Applied Energy*, *267*, 115081. https://doi.org/10.1016/j.apenergy.2020.115081

29. Sun, S., Zhang, J., Bi, J., & Wang, Y. (2019). A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles. *Journal of Advanced Transportation*, *2019*, 1–14. https://doi.org/10.1155/2019/4109148

30. Zhao, L., Yao, W., Wang, Y., & Hu, J. (2020). Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles. *IEEE Access*, *8*, 212423–212441. https://doi.org/10.1109/access.2020.3039815

31. Wäger, G., Whale, J., & Bräunl, T. (2016). Driving electric vehicles at highway speeds: The effect of higher driving speeds on energy consumption and driving range for electric vehicles in Australia. *Renewable & Sustainable Energy Reviews*, *63*, 158–165. https://doi.org/10.1016/j.rser.2016.05.060

32. Saputra, Y., Hoang, D., Nguyen, D., Dutkiewicz, E., Dominik Mueck, M., & Srikanteswara, S. (n.d.). *Energy Demand Prediction with Federated Learning for Electric Vehicle Networks*. Retrieved October 6, 2023, from https://arxiv.org/pdf/1909.00907.pdf

33. Wang, Y., Chenlong, W., & He, H. (2021). Estimating the Energy Consumption and Driving Range of Electric Vehicles with Machine Learning. *Journal of Physics*, *2005*(1), 012131. https://doi.org/10.1088/1742-6596/2005/1/012131

34. *A Stochastic Approach to Designing Plug-In Electric Vehicle Charging Controller for Residential Applications | IEEE Journals & Magazine | IEEE Xplore*. (n.d.). Ieeexplore.ieee.org. Retrieved December 15, 2023, from https://ieeexplore.ieee.org/abstract/document/9775959

35. De Cauwer, C., Van Mierlo, J., & Coosemans, T. (2015). Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. *Energies*, *8*(8), 8573–8593. https://doi.org/10.3390/en8088573

36. Zhu, J., Yang, Z., Guo, Y., Zhang, J., & Yang, H. (2019). Short-Term load forecasting for electric vehicle charging stations based on deep learning approaches. *Applied Sciences*, *9*(9), 1723. https://doi.org/10.3390/app9091723

37. Peng, T., Ou, X., Yuan, Z., Yan, X., & Zhang, X. (2018). Development and application of China provincial road transport energy demand and GHG emissions analysis model. *Applied Energy*, *222*, 313–328. https://doi.org/10.1016/j.apenergy.2018.03.139

38. 11. Ou, X., Zhang, X., & Chang, S. (2010). Scenario analysis on alternative fuel/vehicle for China's future road transport: Life-cycle energy demand and GHG emissions. *Energy Policy*, *38*(8), 3943–3956. https://doi.org/10.1016/j.enpol.2010.03.018

39. Mazhar, T., Asif, R. N., Malik, M. A., Adnan, M., Haq, I., Iqbal, M., Kamran, M., & Ashraf, S. (2023). Electric vehicle charging system in the smart grid using different machine learning methods. *Sustainability*, *15*(3), 2603. https://doi.org/10.3390/su15032603

40. Prakobkaew, P., & Sirisumrannukul, S. (2023). Optimal locating and sizing of charging stations for large-scale areas based on GIS data and grid partitioning. *Iet Generation Transmission & Distribution*. https://doi.org/10.1049/gtd2.13046

41. Fredriksson, H., Dahl, M., & Holmgren, J. (2019). Optimal placement of Charging Stations for Electric Vehicles in large-scale Transportation Networks. *Procedia Computer Science*, *160*, 77–84. https://doi.org/10.1016/j.procs.2019.09.446

42. Luo, C., Huang, Y.-F., & Gupta, V. (2015). Placement of EV Charging Stations--Balancing Benefits Among Multiple Entities. *IEEE Transactions on Smart Grid*, 1–10. https://doi.org/10.1109/tsg.2015.2508740
43. Khalifé, A., Fay, T., & Göhlich, D. (2022). Optimizing Public Charging: an integrated approach based on GIS and Multi-Criteria decision analysis. *World Electric Vehicle Journal*, *13*(8), 131. https://doi.org/10.3390/wevj13080131
44. Fredriksson, H., Dahl, M., & Holmgren, J. (2019b). Optimal placement of Charging Stations for Electric Vehicles in large-scale Transportation Networks. *Procedia Computer Science*, *160*, 77–84. https://doi.org/10.1016/j.procs.2019.09.446
45. Kadri, A. A., Perrouault, R., Boujelben, M. K., & Gicquel, C. (2020). A multi-stage stochastic integer programming approach for locating electric vehicle charging stations. *Computers & Operations Research*, *117*, 104888. https://doi.org/10.1016/j.cor.2020.104888
46. A. Nespoli, E. Ogliari and S. Leva, "User Behavior Clustering Based Method for EV Charging Forecast," in IEEE Access, vol. 11, pp. 6273-6283, 2023, doi: 10.1109/ACCESS.2023.3235952.
47. S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou and M. Nijim, "Prediction of EV Charging Behavior Using Machine Learning," in IEEE Access, vol. 9, pp. 111576-111586, 2021, doi: 10.1109/ACCESS.2021.3103119.
48. Majidpour, M., Qiu, C., Chu, P., Pota, H. R., & Gadh, R. (2016). Forecasting the EV charging load based on customer profile or station measurement? *Applied Energy*, *163*, 134–141. https://doi.org/10.1016/j.apenergy.2015.10.184
49. S. Su, H. Zhao, H. Zhang, X. Lin, F. Yang and Z. Li, "Forecast of electric vehicle charging demand based on traffic flow model and optimal path planning," 2017 19th International

Conference on Intelligent System Application to Power Systems (ISAP), San Antonio, TX, USA, 2017, pp. 1-6, doi: 10.1109/ISAP.2017.8071382.

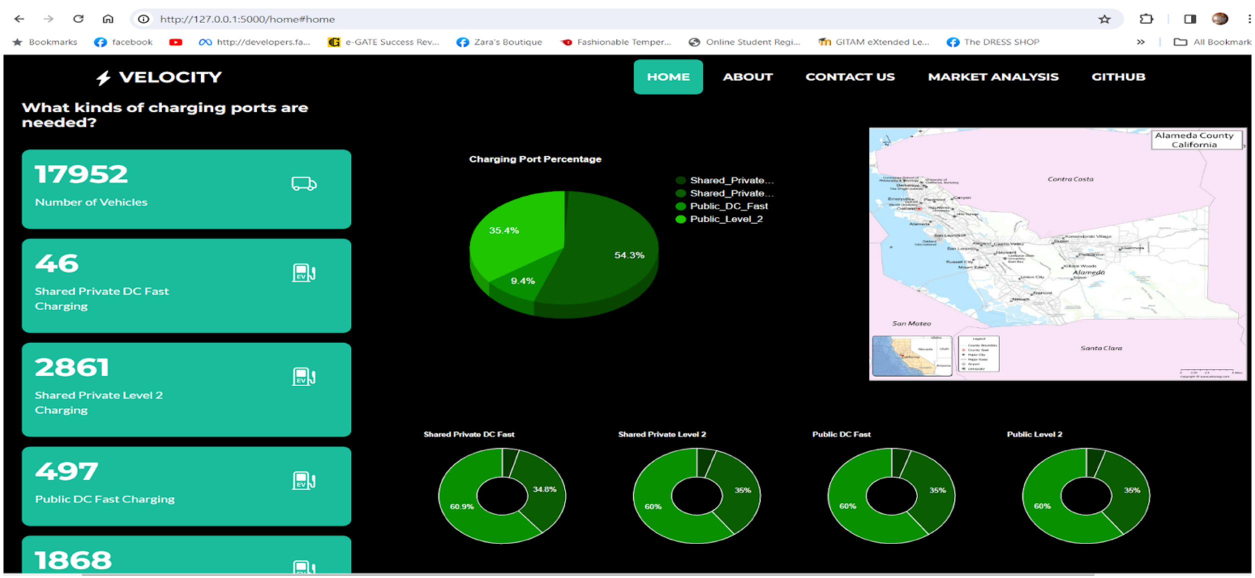
50. Zhang, J., Yan, J., Liu, Y., Zhang, H., & Lv, G. (2020). Daily electric vehicle charging load profiles considering demographics of vehicle users. *Applied Energy*, *274*, 115063. https://doi.org/10.1016/j.apenergy.2020.115063
51. P. Rajagopalan and P. Ranganathan, "Electric Vehicle Charging Behavior Prediction using Machine Learning Models," 2022 IEEE Electrical Power and Energy Conference (EPEC), Victoria, BC, Canada, 2022, pp. 123-128, doi: 10.1109/EPEC56903.2022.10000160.
52. Hu, D., Zhou, K., Li, F., & Ma, D. (2022). Electric vehicle user classification and value discovery based on charging big data. *Energy*, *249*, 123698. https://doi.org/10.1016/j.energy.2022.123698
53. Danté, A. W., Agbossou, K., Kélouwani, S., Cardenas, A., & Bouchard, J. (2019). Online modeling and identification of plug-in electric vehicles sharing a residential station. *International Journal of Electrical Power & Energy Systems*, *108*, 162–176. https://doi.org/10.1016/j.ijepes.2018.12.024
54. A. M. A. Haidar and K. M. Muttaqi, "Behavioral Characterization of Electric Vehicle Charging Loads in a Distribution Power Grid Through Modeling of Battery Chargers," in IEEE Transactions on Industry Applications, vol. 52, no. 1, pp. 483-492, Jan.-Feb. 2016, doi: 10.1109/TIA.2015.2483705.
55. Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, *45*, 103406. https://doi.org/10.1016/j.jobe.2021.103406

Appendices

Appendix A - System Testing

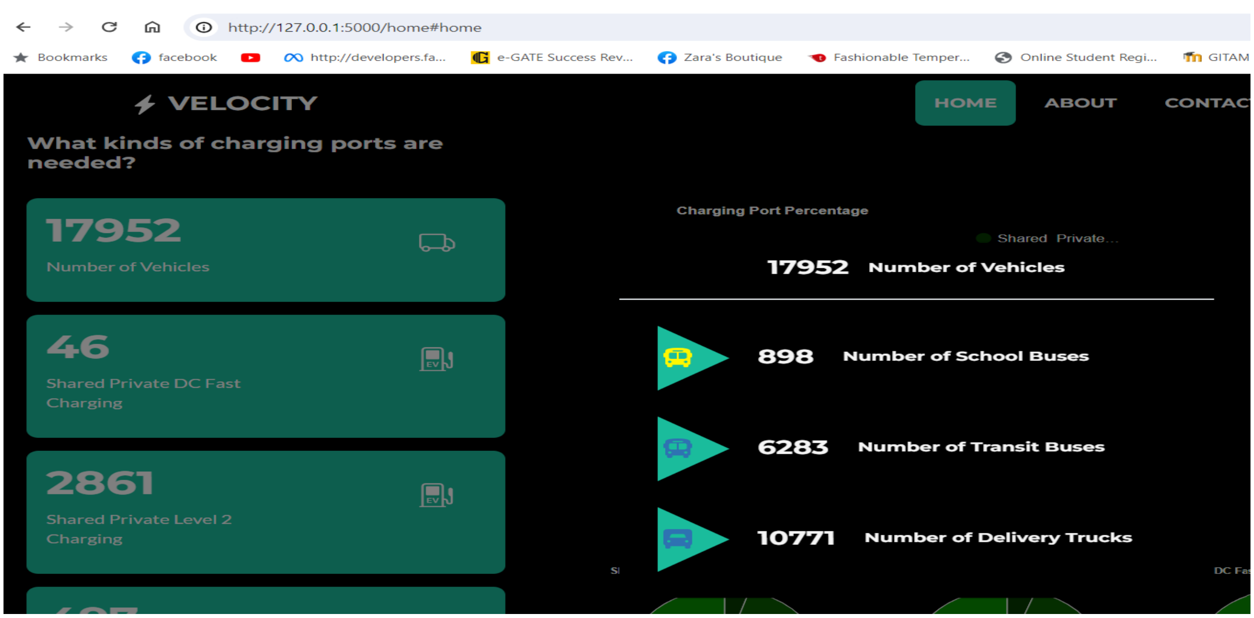
Use Case 1:

Projected Count of Electric Vehicles on road and Charging Stations for Alameda County in 2030



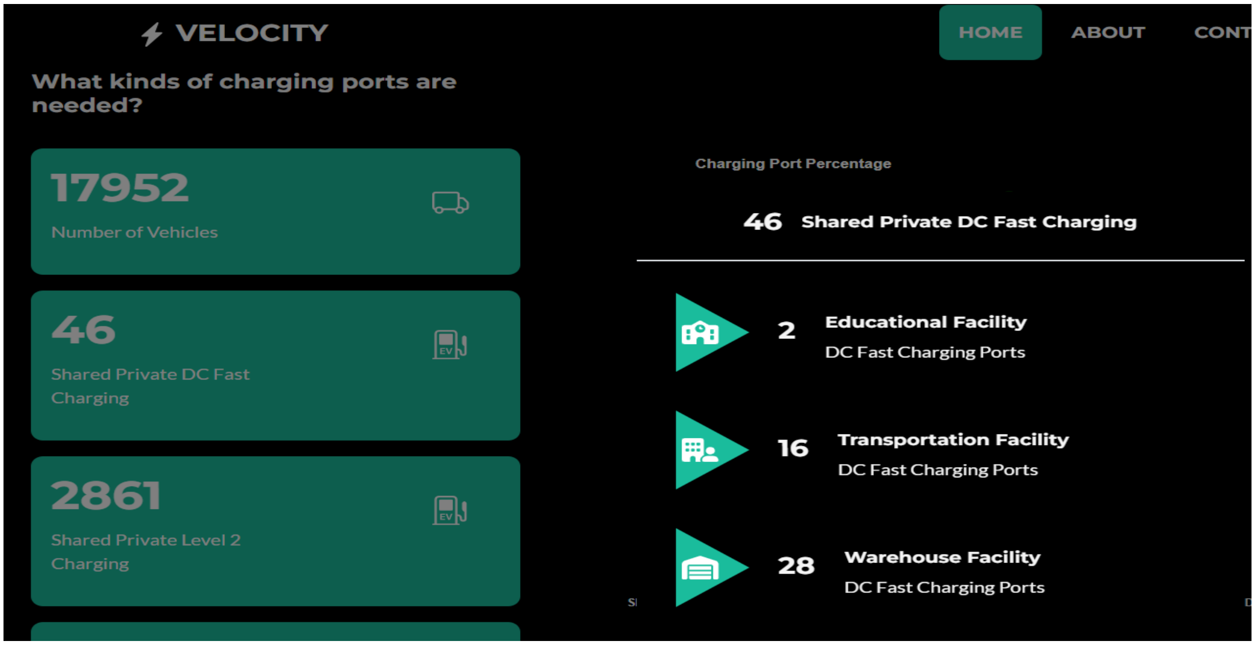
Use Case 2:

Projected Count of Each Vehicle Type for Alameda County in 2030



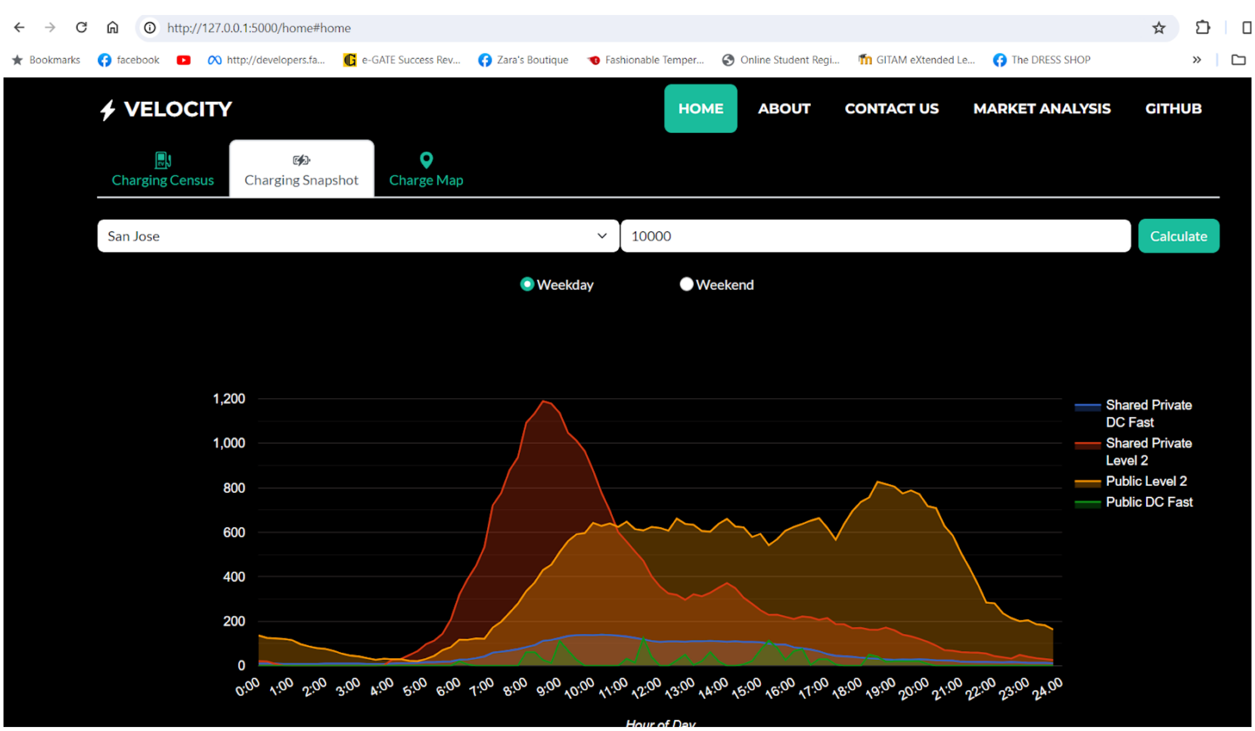
Use Case 3:

Count of Charging Ports at Different Infrastructure Locations for Alameda County in 2030



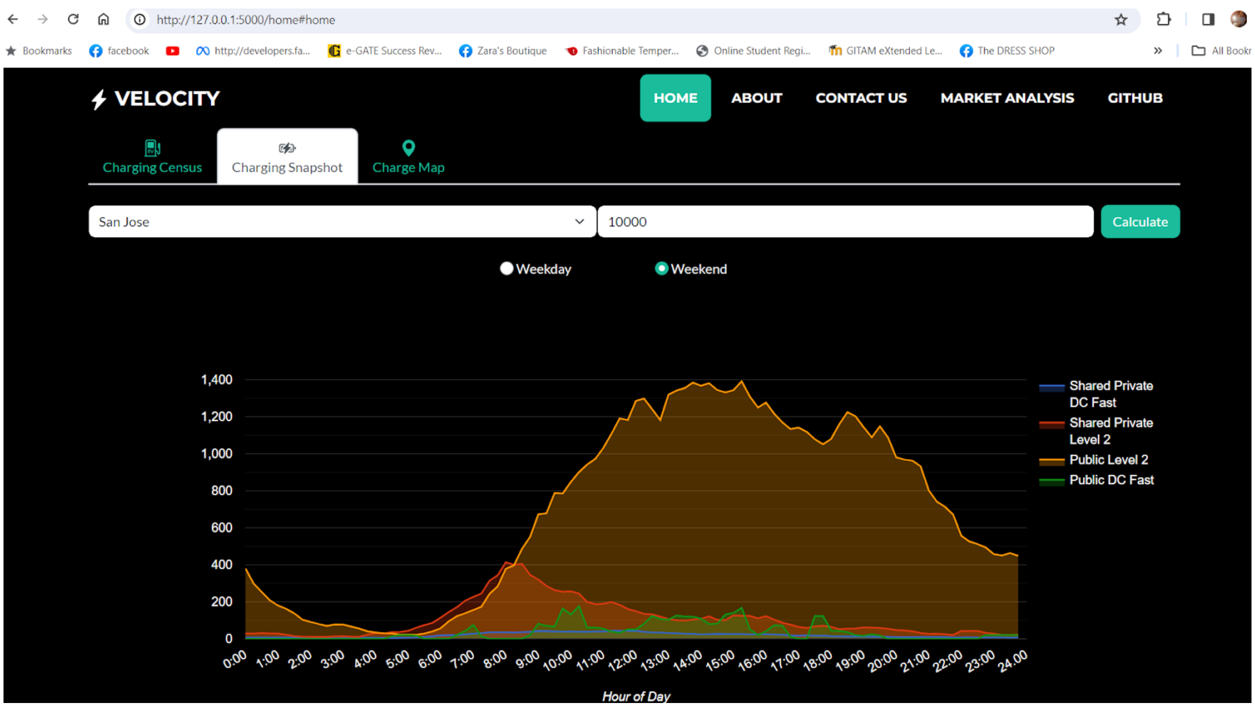
Use Case 4:

Energy Requirement for San Jose City on a Weekday for Different Chatging Port Types



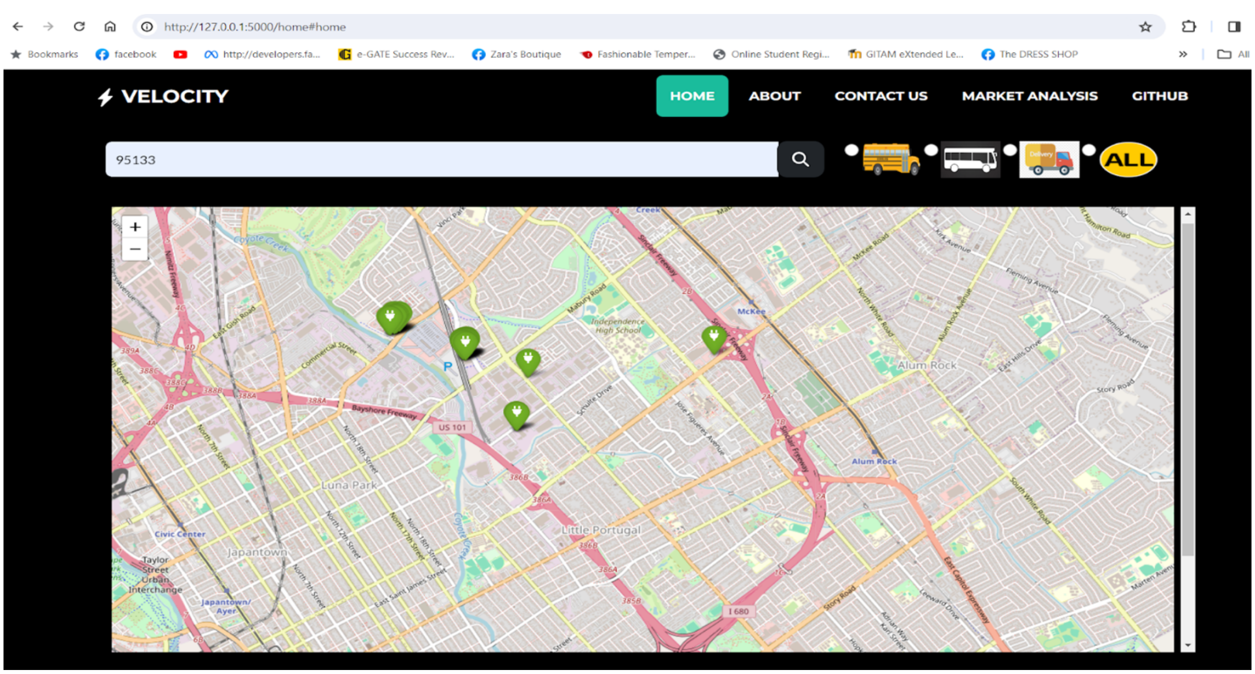
Use Case 5:

Energy Requirement for San Jose City on a Weekend for Different Chatging Port Types



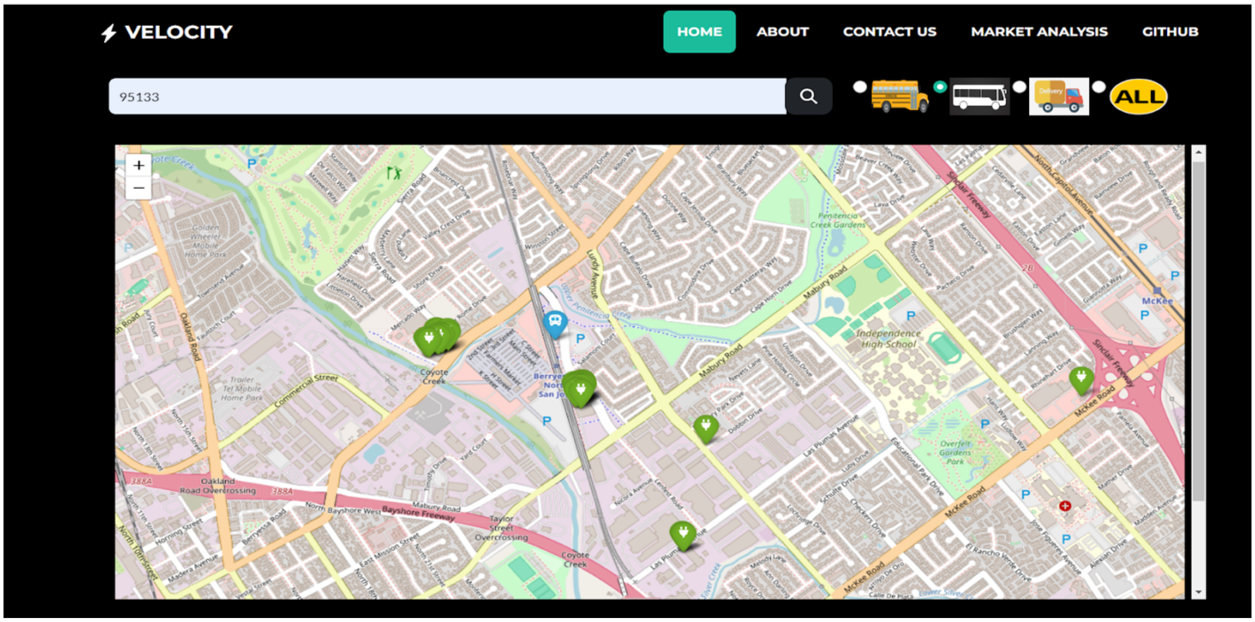
Use Case 6:

Existing Charging Station Locations for Zip Code - 95133



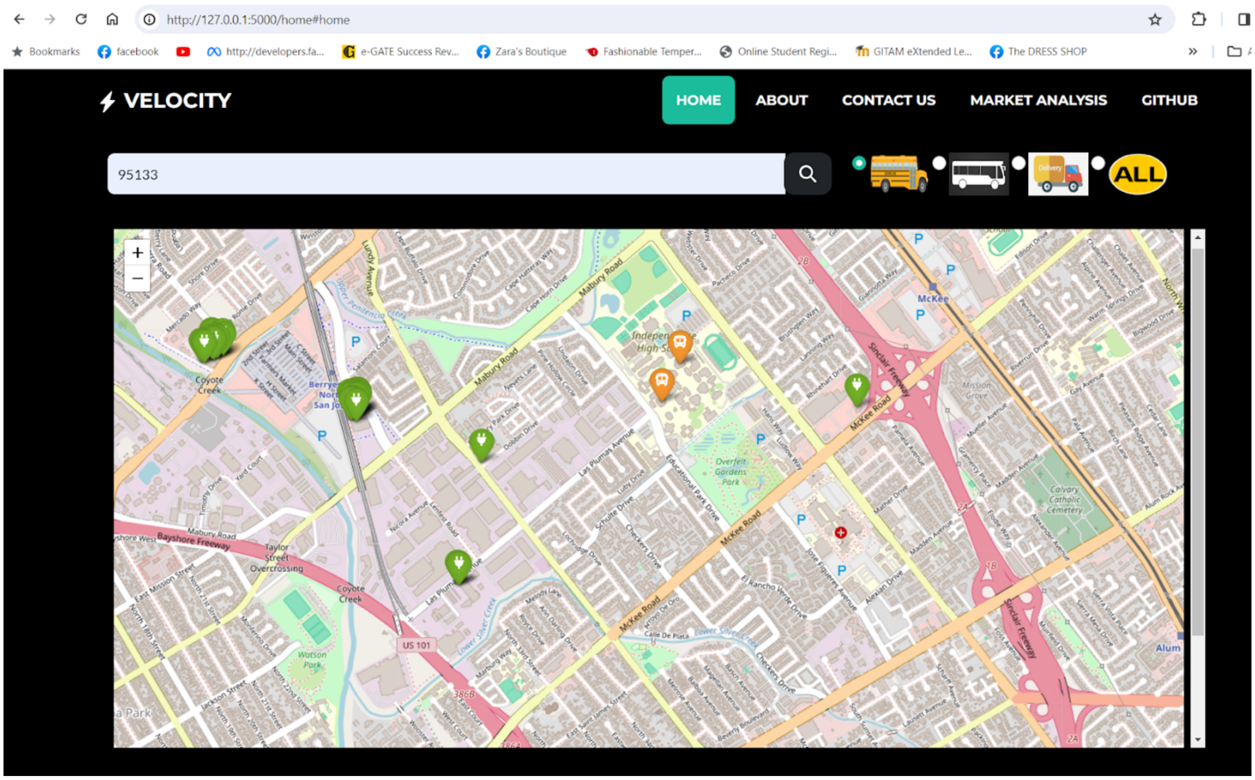
Use Case 7:

Suggested Charging Station Locations for Transit Bus for Zip Code - 95133



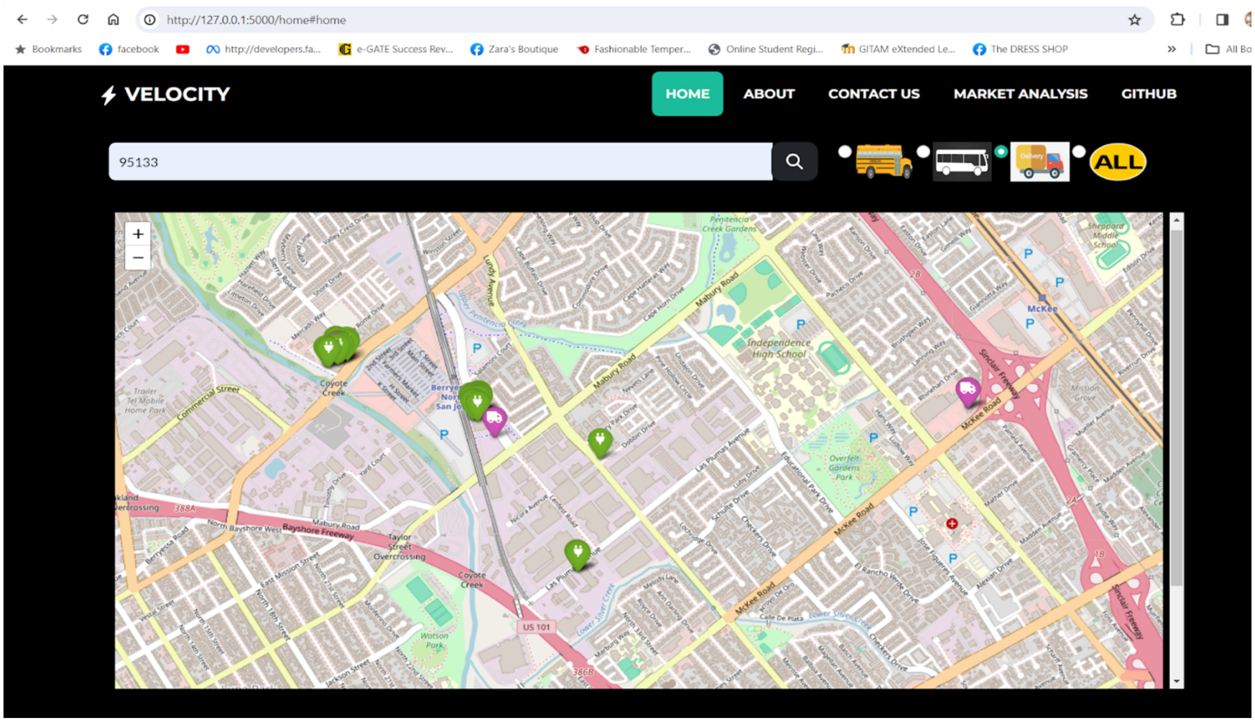
Use Case 8:

Suggested Charging Station Locations for School Bus for Zip Code - 95133



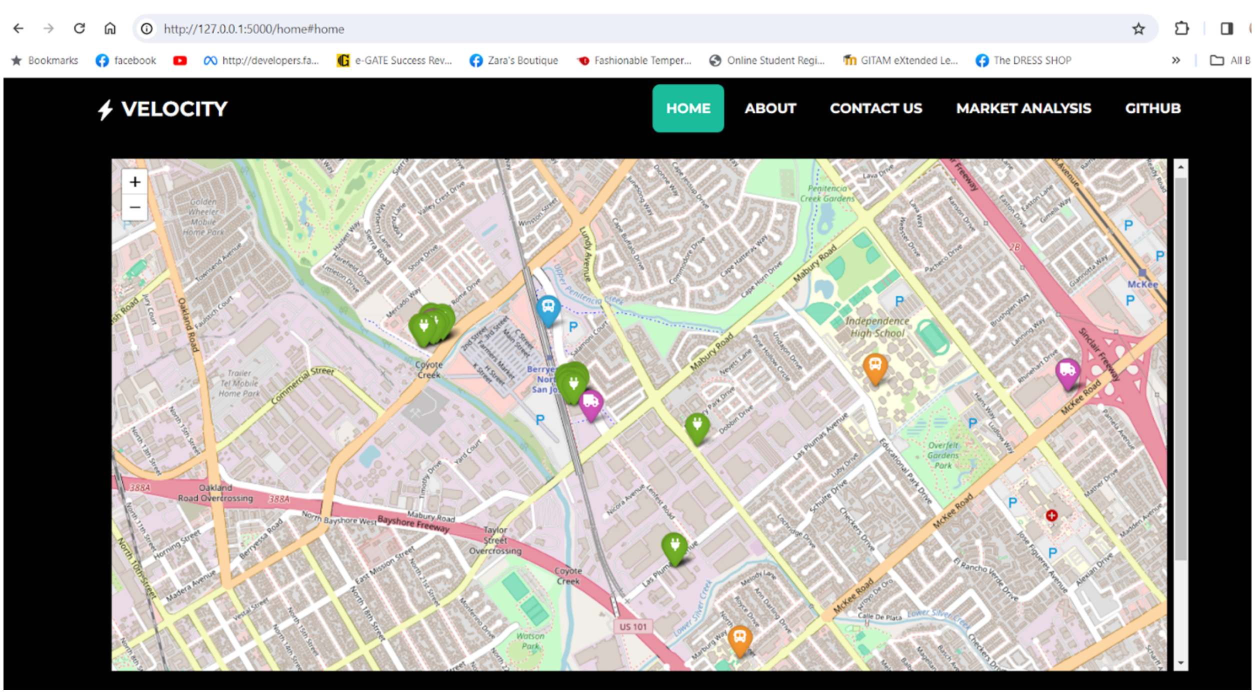
Use Case 9:

Suggested Charging Station Locations for Delivery Truck for Zip Code - 95133



Use Case 10:

All Suggested Charging Station Locations for a Zip Code - 95133



Appendix B – Project Data Source and Management Store

Entire Data used for this project is stored in :

https://drive.google.com/drive/folders/1dplBGiXz9Nu3D5rxpnNBSxR5KGHuun2T

Appendix C – Project Program Source Library, Presentation, and Demonstration

Project Program Source Library	https://drive.google.com/drive/folders/1X2RGVHzqESCyfvqndi3xAnTL8k1zNEO9
	https://drive.google.com/drive/folders/1YOgABFJWhSF6O3ZruTLVJ1F59a8kgs99
	https://github.com/Lohitha-Vanteru/DATA298B_T6_EVCSI
Presentation	https://drive.google.com/drive/folders/1IruXkrPDOTuBsz5FiuoNDMo7xqt-Petr
Demonstration	https://drive.google.com/drive/folders/1IruXkrPDOTuBsz5FiuoNDMo7xqt-Petr