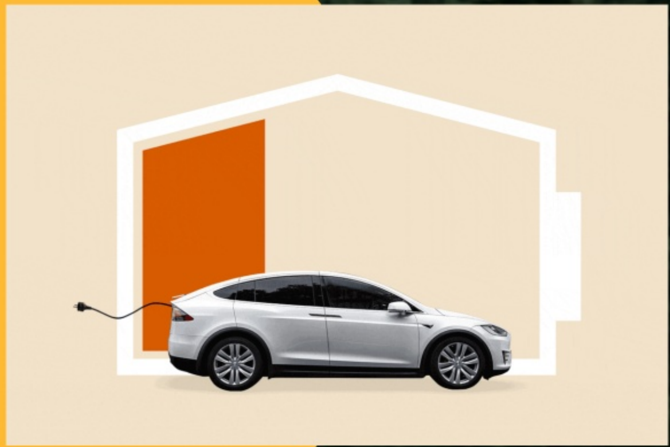
Data-Driven Analysis and Planning of EV Charging Infrastructure for a Smart City

- Introduction
- Requirements
- Organization
- Literature survey
- Data Statistics
- Data Preparation
- Analytics
- Modelling
- Prototypes
- Developments in the Industry

Presented by : Team 7

Megha Gupta
Venkatasai Rao Dheekonda
Sai Sankeerth Gangasani
Sajit Valiya Kizhakke

Introduction



Project Background & Executive Summary

Targeted Problem & Project Deliverables

Existing Technologies

Overview of the Issue

California, a national leader in EV adoption, faces hurdles in establishing an efficient charging system. To tackle this, the project 'Data-Driven Analysis and Planning of EV Charging Infrastructure for a Smart City' will assess current infrastructure, highlight gaps, and devise a detailed plan for improvement.

Grasping charging demand, user behavior, and distribution is vital for matching escalating demand. If successful, this research could drive broader EV acceptance, aligning with California's sustainability goals.

This project presents a crucial opportunity to not only optimize infrastructure but also to accelerate the state's transition towards green mobility.

Targeted Problems:

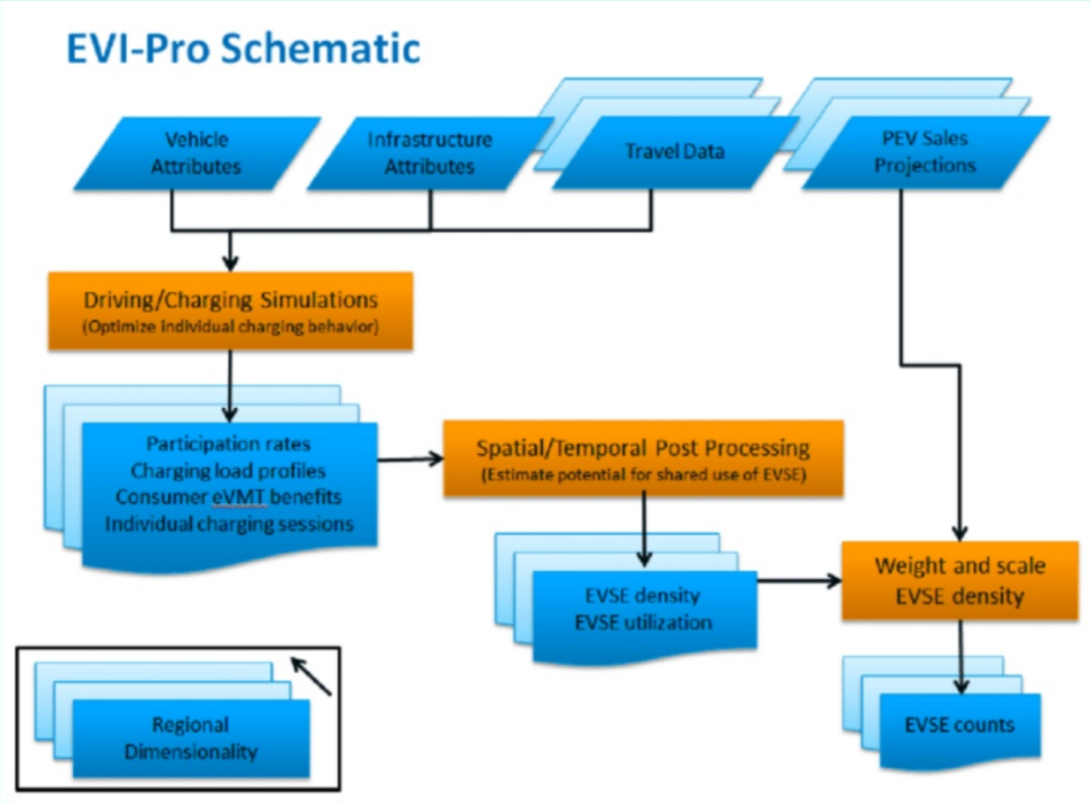
- EV integration hampered by underdeveloped, inefficient charging infrastructure.
- Research targets EV charging infrastructure insufficiency in California cities.
- Data-driven approach to analyze and optimize existing infrastructure.
- Aim to create reliable, accessible, efficient EV charging system.

Project Deliverables:

- Comprehensive understanding of current state, predictive models of future demand, strategic infrastructure enhancement recommendations.
- Research insights to inform actionable plan for infrastructure improvement.
- Plan tailored to stakeholders' unique needs, supporting effective EV use and California's sustainability goals

Existing Technologies

EVI-Pro Lite : By US Department of Energy



ChargeHub.com

ChargeHub.com is a website and online platform that provides information and services related to electric vehicle (EV) charging.

ChargePoint

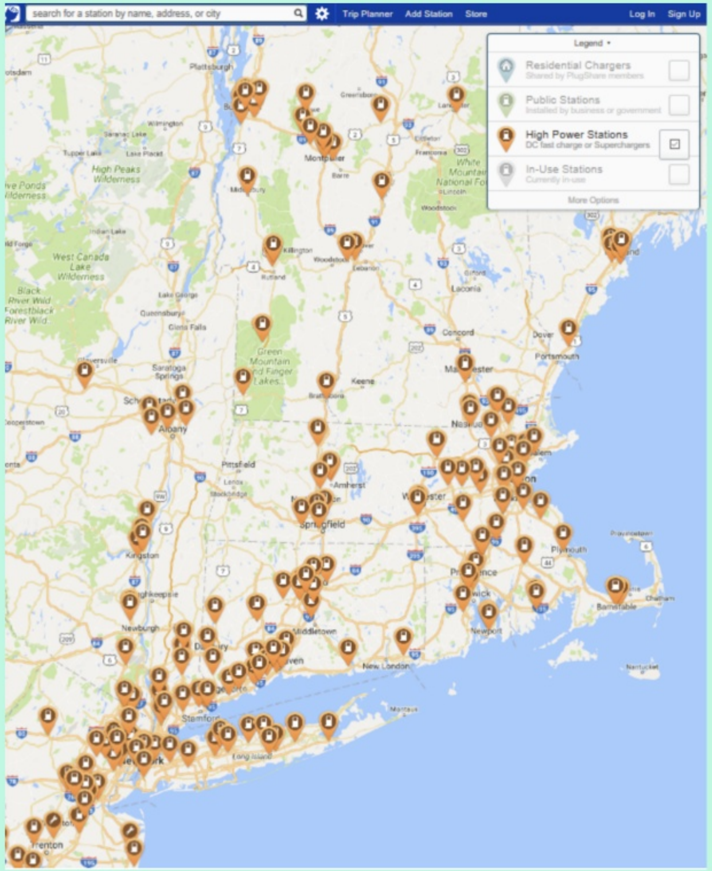
ChargePoint offers a comprehensive range of charging solutions for various applications, including residential, workplace, commercial, and public charging.

EVgo

EVgo operates an extensive network of fast-charging stations designed to support the growing demand for EV charging across the country

EV Hub

The objective of EV Hub is to bring a data-driven approach to policymaking around transportation electrification and accelerate market growth.



Comparison of Existing Current Technologies

Feature	EVI-Pro Lite	ChargeHub.com	ChargePoint	EVgo	EVhub
Charging Station Locator	Yes	Yes	Yes	Yes	Yes
Charging Station Reviews and Ratings	No	Yes	Yes	Yes	Yes
Charging Network Information	No	Yes	Yes	Yes	Yes
EV Charging Resources	No	Yes	Yes	Yes	Yes
Charging Cost Calculator	No	Yes	Yes	Yes	Yes
Mobile App	No	Yes	Yes	Yes	Yes
Network Size and Coverage	Moderate	Wide coverage	Wide coverage	Wide coverage	Not specified
Membership and Payment Options	Not specified	Varies by charging network	Varies by charging network	Varies by charging network	Not specified
Additional Services	Not specified	Not specified	EV driver community, driver support	EV driver support, pricing plans	Not specified
User Interface and Experience	Not specified	User-friendly and intuitive	User-friendly and intuitive	User-friendly and intuitive	Not specified

Project Requirements

Required Resources

Functional

Data

AI-powered

Functional Requirements

The system's functional requirements include processing large datasets, learning from historical data, and gathering real-time data for EV charging infrastructure optimization. It monitors infrastructure performance, optimizes charging processes considering demand and energy availability, and provides actionable insights for infrastructure expansion. Essential functionalities are handling extensive data, learning from past data, real-time data collection, process optimization, performance tracking, and insightful recommendations.

AI-powered Requirements:

This system uses machine learning to forecast EV charging demand and streamline services. It employs predictive analytics for proactive maintenance, and optimizes charging schedules with AI, considering traffic, energy, and user preferences.

Notably, it adapts to user inputs, offering personalized recommendations. Other features include anomaly detection, data visualization, demand forecasting, energy management, user behavior analysis, automated reporting, and real-time decision-making.

This project showcases AI's potential in enhancing efficiency and user experience in smart city EV infrastructure.

DATA Requirements:

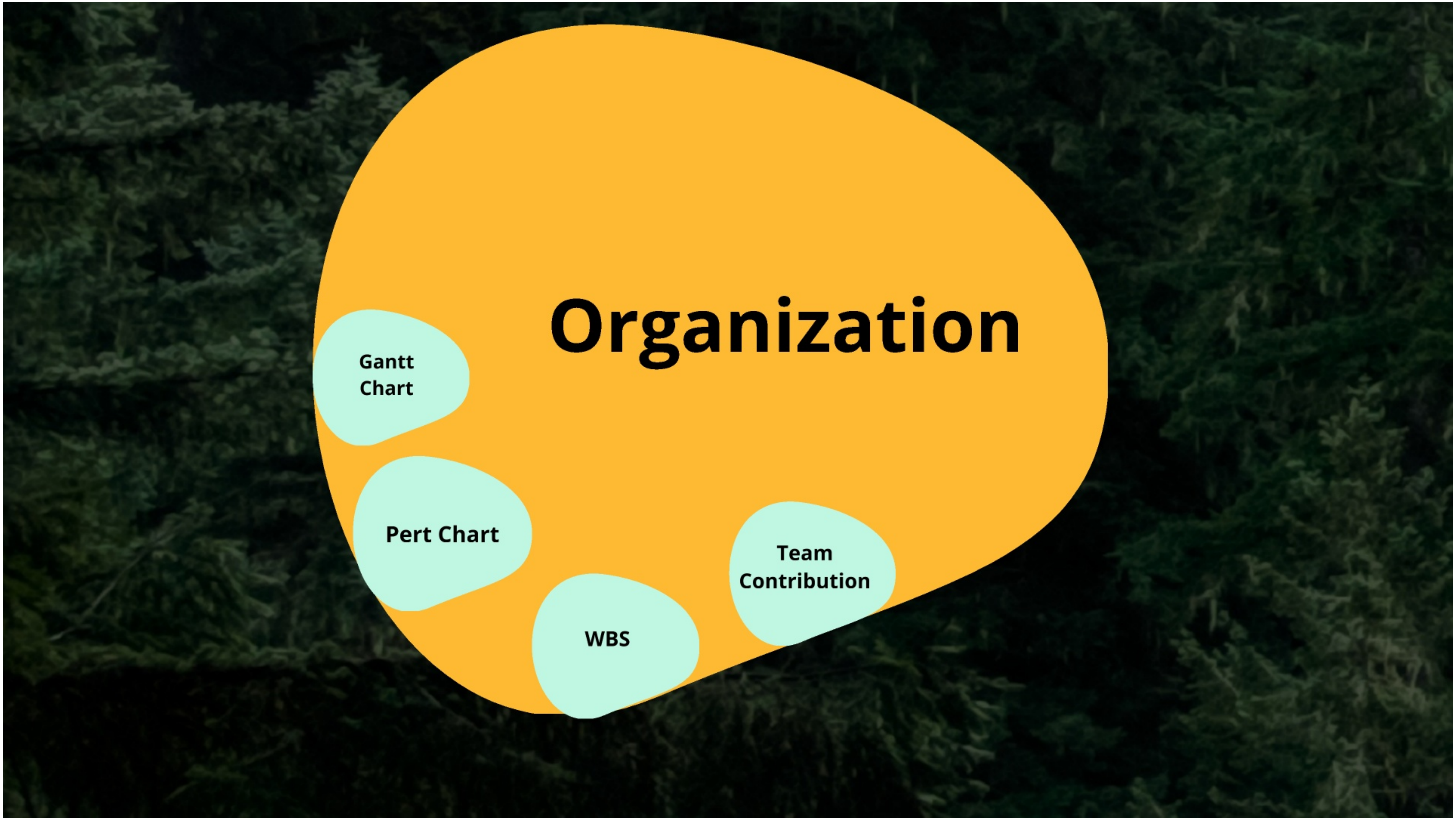
- The project demands real-time data on EV charging station performance, usage, and accessibility. It requires historical trends in EV charging use, and environmental data like weather and traffic affecting charging demand. Information about energy supply, particularly renewables, and details on EV distribution, location, and charging stations in the smart city is crucial. User data, including charging preferences and driving styles, is also needed.

- To propose optimal charging locations, the system requires data on existing EV infrastructure, including type, quantity, location, and availability of charging stations. Population information, including household and business counts and EV driver numbers, is essential. Traffic data helps identify areas with high volumes needing more stations.

- Demographic data predicts the type and level of EV adoption, while energy data identifies areas with renewable power sources for charging stations. Land use data locates available land or parking spaces, and economic data provides insight into industries that could benefit from EV infrastructure.

Required Resources

Function	Resource Type	Resource
Local Machine	Hardware	64 - bit machine or higher
Data Storage	Hardware	AWS S3
Cloud Service Management Tool	Software	AWS CLI
Training, Testing, and Deployment	Software	AWS SageMaker
ML Frameworks	Software	Scikit-learn
Visualization Tool	Software	Tableau


Organization
Gantt Chart
Pert Chart
WBS
Team Contribution

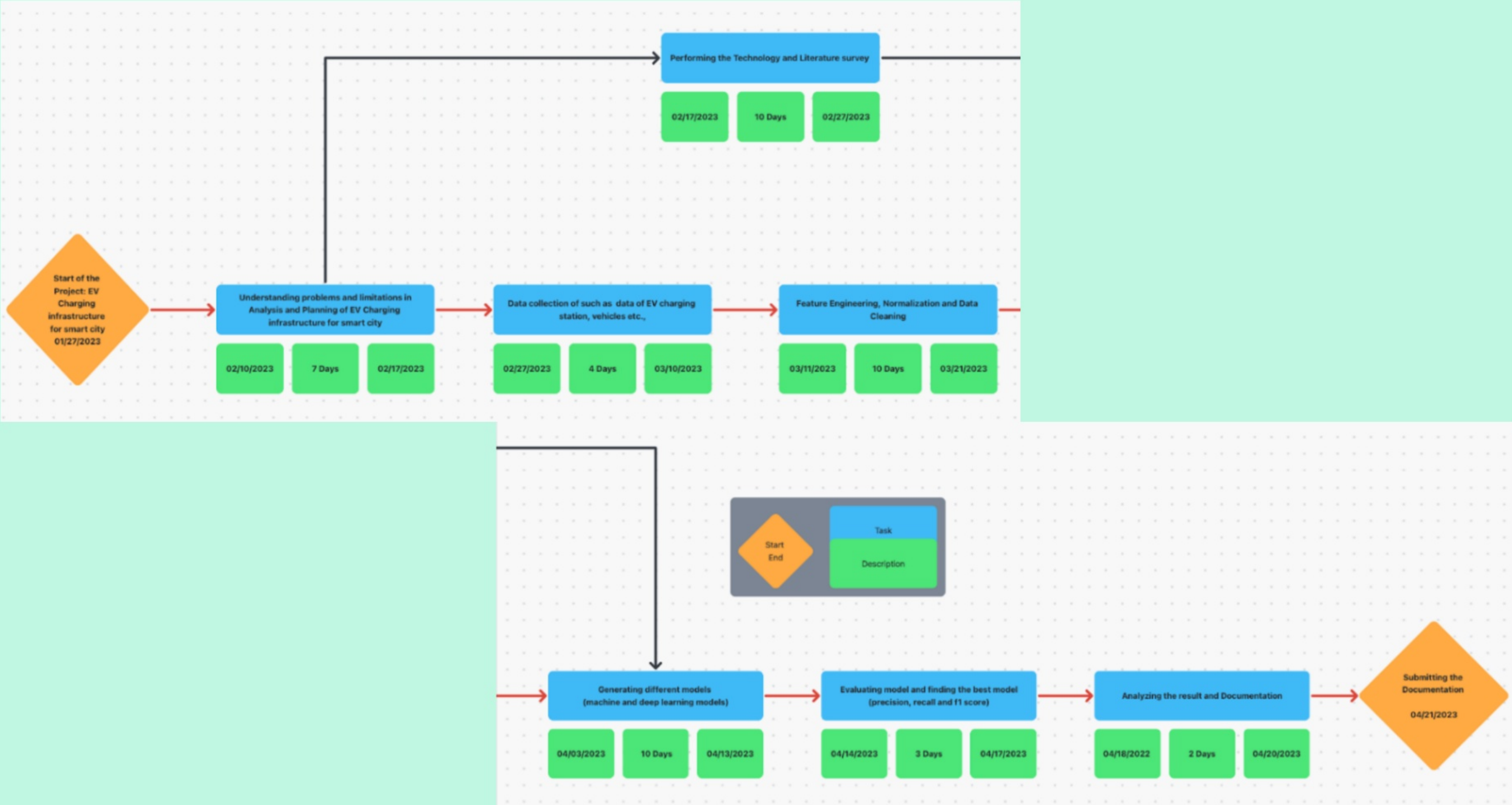
Team Organization

Task	Team Member
Project Understanding	Megha, Venkatasai, Sankeerth, Sajit
Data Understanding	Venkatasai, Sankeerth
Data Collection	Megha, Sajit
Data Preparation	Megha, Sajit
Modelling	Megha, Venkatasai, Sankeerth, Sajit
Evaluation & Results	Venkatasai, Sankeerth
Documentation	Megha, Venkatasai, Sankeerth,Sajit

WBS

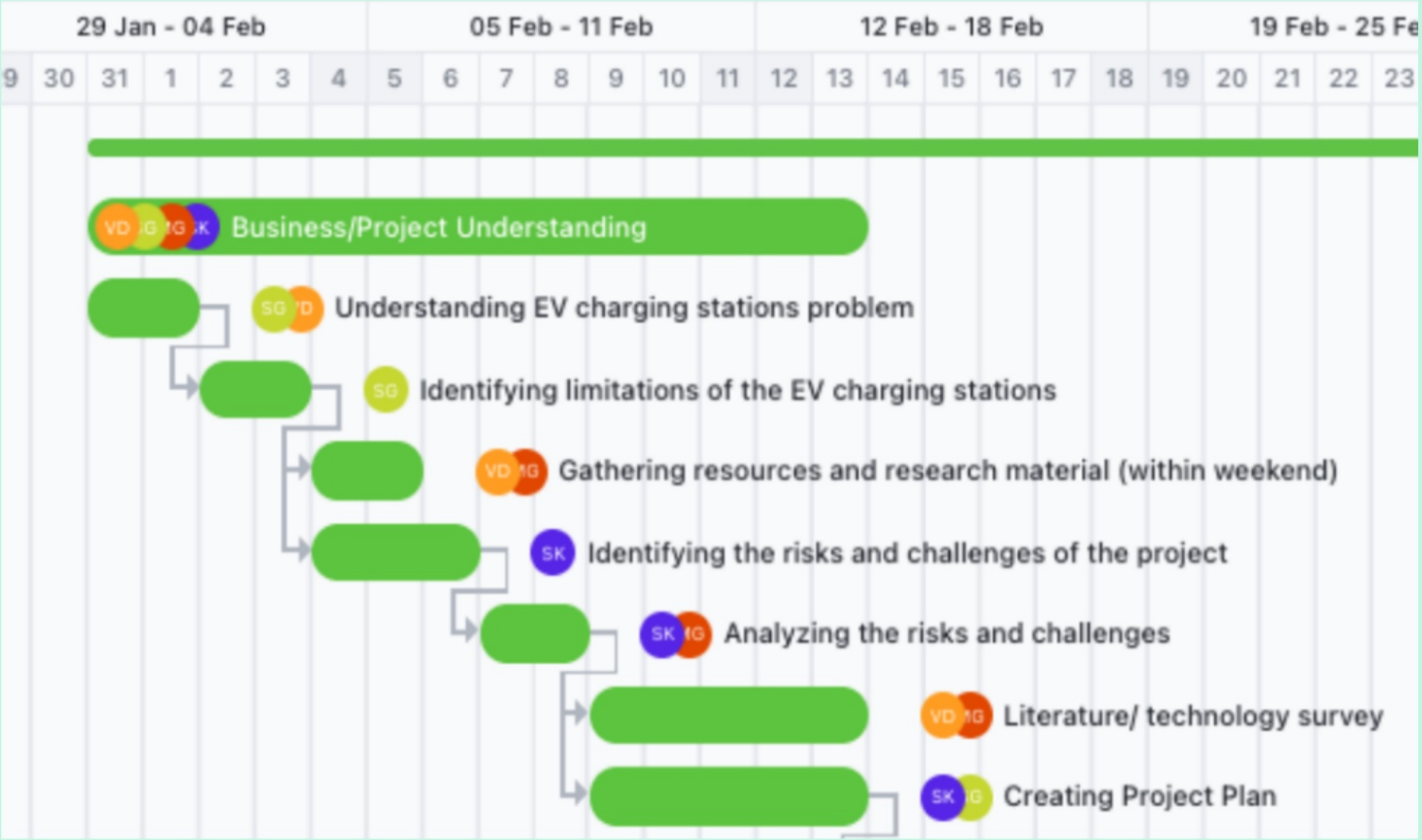
Data-Driven Analysis and Planning of EV Charging Infrastructure for a Smart City
Business/Project Understanding
Understanding EV Charging infrastructure problem
Identifying limitations of the EV Charging infrastructure problem
Gathering resources and research material
Identifying the risks and challenges of the project
Perform literature/ technology survey
Creating project plan
Data Understanding
Data collection from DMV, government sites etc.
Performing EDA on various datasets
Feature engineering on the data
Data Preparation
Normalization of the data
Data Cleaning
Removing outliers in the data
Data Visualization
Identifying Gaps
Merge the datasets
Modelling
Split data for training, testing and validation
Creating machine and Deep learning model to classify images
Addition of layers and hyperparameter tuning
Fit the model on trainng and validation datasets
Evaluation
Evaluate the model for test data distribution
Perform Cross Validation
Model evaluation using precision, recall and f1 score
Analyze confusion matrix of all models
Finding out the best-performing model
Deployment
Deploy the best performing neural network model
Ensure that the old model can be rolled back if necessary
Analyzing the results
Analyze the distribution of model predictions
Preparation of final project report

Pert Chart

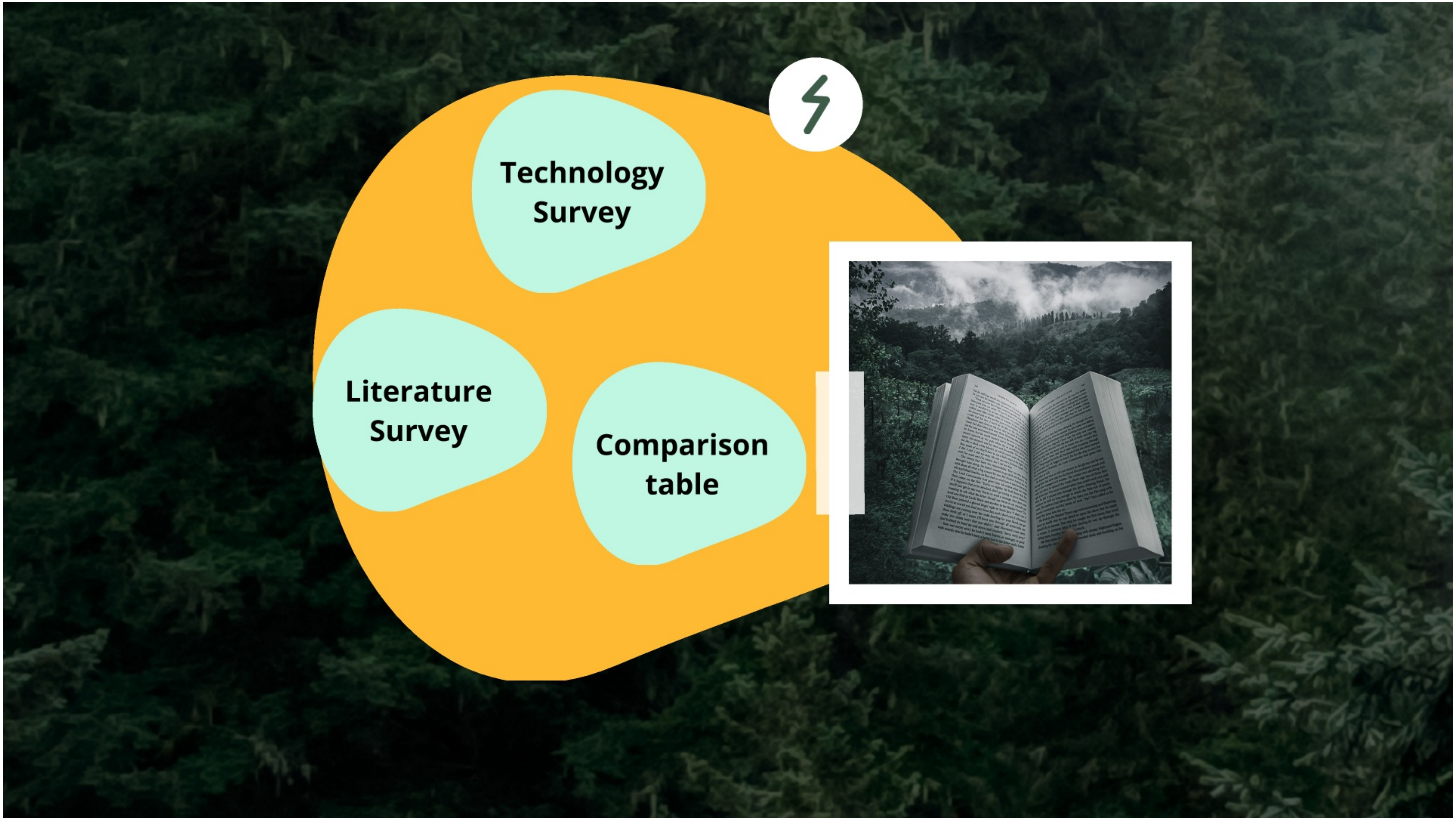


Start of the Project: EV Charging infrastructure for smart city 01/27/2023
Understanding problems and limitations in Analysis and Planning of EV Charging infrastructure for smart city
02/10/2023
7 Days
02/17/2023
Performing the Technology and Literature survey
02/17/2023
10 Days
02/27/2023
Data collection of such as data of EV charging station, vehicles etc.,
02/27/2023
4 Days
03/10/2023
Feature Engineering, Normalization and Data Cleaning
03/11/2023
10 Days
03/21/2023
Start
End
Task
Description
Generating different models (machine and deep learning models)
04/03/2023
10 Days
04/13/2023
Evaluating model and finding the best model (precision, recall and f1 score)
04/14/2023
3 Days
04/17/2023
Analyzing the result and Documentation
04/18/2022
2 Days
04/20/2023
Submitting the Documentation
04/21/2023

Gantt Chart



29 Jan - 04 Feb
05 Feb - 11 Feb
12 Feb - 18 Feb
9 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
VD
Business/Project Understanding
SG
Understanding EV charging stations problem
SG
Identifying limitations of the EV charging stations
VD
Gathering resources and research material (within weekend)
SK
Identifying the risks and challenges of the project
SK
Analyzing the risks and challenges
VD
Literature/ technology survey
SK
Creating Project Plan


Technology Survey
Literature Survey
Comparison table

Literature Survey

Study	Models Used	Results
Ran Wei et al. (2022)	Seq2Seq	Various performance metrics show that Seq2Seq outperforms other models
Tianyu Hu et al. (2021)	Machine Theory of Mind (MToM)	Identify a comprehensive and accurate method for forecasting EV charging session power requirements
Zhile Yang et al. (2019)	Long Short-Term Memory (LSTM)	Forecast EV charging station load
Huikung Yang et al. (2019)	Deep Neural Network (DNN), a Recurrent Neural Network (RNN), LSTM and Gated Recurrent Unit (GRU)	GRU model outperformed the other three models
Juncheng Zhu et al. (2019)	Long Short-Term Memory (LSTM)	Compared to traditional artificial neural network models, the long-short-term memory approach reduces forecasting errors by over 30%.

Technology Survey

Study	Models Used	Results
Karakose et al. (2021)	Genetic Algorithm (GA) based on Graph Theory	Optimized set of EV charging station locations. Sustainable transportation infrastructure planning, promoting widespread EV adoption, reduced travel time
Bayhan et al. (2020)	Integer Linear Programming (ILP)	Optimal locations for EV charging stations with maximum capacity and coverage. Efficient infrastructure planning, increased EV adoption, improved charging accessibility, cost savings
Yenchamchalit et al. (2018)	Particle Swarm Optimization (PSO)	Optimal locations and sizes for EV charging stations. Efficient infrastructure planning, increased EV adoption, cost savings, improved grid reliability
Zheng et al. (2016)	Bi-level programming model	Optimal locations for EV charging facilities considering traffic equilibrium. Efficient infrastructure planning, reduced congestion, cost savings, increased EV adoption, accessibility

Comparisons of Literature Survey

Author Name	Methodology	Objectives	Techniques	Key Considerations	Benefits	Limitations
Luo et al. (2022)	Genetic Algorithm (GA)	EV charging station location optimization using a cost model	Cost model, demand points, candidate locations, fitness evaluation, selection, crossover, mutation	Cost model, demand points, candidate locations, fitness evaluation, selection, crossover, mutation	Efficient infrastructure planning, cost savings, increased EV adoption, improved charging accessibility	Reliance on predefined termination criterion
A.A. Ahmadi (2009)	Robust optimization model and discrete event simulation	Improve reliability and accessibility of EV charging infrastructure and reduce environmental impact	Discrete event simulation and robust optimization	Uncertain traffic flows and existing power grid networks, utilization and service rates of proposed charging stations	Improved reliability and accessibility of EV charging infrastructure, consideration of traffic and grid uncertainties, optimization of charging capacity utilization	Reliance on assumptions made during modeling and simulation
Z. Ren et al (2016)	Development of a metro area EV charging station management system (CSMS)	Address challenges of managing EV charging stations due to multiple brands	Application development to create a CSMS application to monitor and control charging stations	Design of the system connected to multiple charging stations in three cities in the urban area	Tool for monitoring and controlling charging stations, addressing challenges of managing multiple brands in a central system	Not mentioned
Zheng et al. (2016)	Bi-objective Mixed Integer Linear	Minimized uncovered charge and installation cost	Route selection policies, driver's reserve battery preference, Monte Carlo simulation, mixed integer linear programming model	Vehicles structure, driving range, sockets count of rechargeable vehicles per day, estimating vehicle flow	Consideration of route selection policies and driver's reserve battery preference, reduction of uncovered charge demand, minimization of installation cost	Reliance on assumptions made during modeling
Yenchamchalit et al. (2018)	Development of an end-user cloud app that recommends EV charging stations based on a cloud-based and customer-oriented approach	Provide a useful and accurate recommender for EV drivers to find nearby charging stations that meet their requirements	Random Forest Algorithms (RFAs) for finding nearby stations, Linear Search Algorithms (LSAs) for filtering stations, Haversine formulas	Focus on technical aspects such as machine learning algorithms and data mining techniques	Improved user experience for EV drivers, increased adoption of electric vehicles	Lack of discussion on limitations or challenges faced during development

Data
Statistics
Dataset
Samples
Data
Data
Collection

Data Collection

Dataset	Dataset Source	Purpose (Issue Addressed)
Boulder EV Load	https://open-data.bouldercolorado.gov/datasets/39288b03f8d54b39848a2df9f1c5fca2_0/explore	Charging Behavior, Charging Duration
City of Palo Alto EV Charging Data	https://data.cityofpaloalto.org/dataviews/257812/electric-vehicle-charging-station-usage-july-2011-dec-2020/	Charging Behavior, Charging Demand
ACN Data	https://ev.caltech.edu/dataset	Infrastructure and Recommendation
Alternative Fuels Data Center	https://afdc.energy.gov/data/categories/afvs-and-hevs	Charging Behavior, Charging Duration, Charging Demand

Dataset Samples

	A	B	C	D	E	F	G	H	I	J	K	M
1	Station Name	Street Address	Intersection I	City	ZIP	St	Groups	EV Network	Latitude	Longitu	EV Connecto	Types
13164	PF CA SM ALA	2000 Alameda de las Pulgas		San Mateo	94403	E	Public - Cre	POWERFLEX	37.54471	-122.3213		
13672	PF CA SD 15231Ave	15231 Avenue of Science		San Diego	92128	E	Public	POWERFLEX	32.99231	-117.0805		
13673	PF CA SD KaiserElCajon	1630 E Main St		El Cajon	92021	E	Public	POWERFLEX	32.8079	-116.9226		
13674	PF CA SD Kaiser8080Pkwy	8080 Parkway Dr		La Mesa	91942	E	Public	POWERFLEX	32.77551	-117.0236		
13675	PF CA SD 15073Ave	15073 Avenue Of Science		San Diego	92128	E	Public	POWERFLEX	32.99133	-117.0803		
13676	PF CA PT 1800BSP	1800 S McDowell Blvd		Petaluma	94954	E	Public	POWERFLEX	38.23051	-122.598		
13677	PF CA PT 5341BSP	5341 Old Redwood Hwy N		Petaluma	94954	E	Public	POWERFLEX	38.2779	-122.6696		
13678	PF CA PT 1420BSP	1420 N McDowell Blvd		Petaluma	94954	E	Public	POWERFLEX	38.27677	-122.6711		
13679	PF CA SM 3552nd	355 2nd Ave		San Mateo	94401	E	Public	POWERFLEX	37.56703	-122.3232		
13739	Boulevard Transit Center	The Boulevard Transit Ce	W Jackman Stre	Lancaster	93534	E	Private - Fl	WAVE	34.70148	-118.137		
13740	Palmdale Transportation Center	39000 Clock Tower Plaza	Transportation C	Palmdale	93550	E	Private	WAVE	34.59147	-118.1211		
13741	Owen Memorial Park	43063 10th St W	W Ave K 8 and S	Lancaster	93534	E	Private	WAVE	34.66674	-118.1486		
13742	South Valley Health Center	38359 40th St E	E Palmdale Blvd	Palmdale	93552	E	Private - Fl	WAVE	34.57896	-118.0568		
13743	Bay Area Rapid Transit Center	18666 San Miguel Dr	N California Blvd	Walnut Cree	94596	E	Private - Fl	WAVE	37.9058	-122.0683		
13744	Universal Studio Tour	100 Universal City Plaza	100 Universal Ci	Los Angeles	91608	E	Private	WAVE	34.13904	-118.3527		
13745	San Pedro Catalina Terminal	San Pedro Catalina Term	N Harbor Blvd ar	Los Angeles	90731	E	Private - Fl	WAVE	33.75166	-118.2753		

	V	W	X	Y	Z	AA	AB	AC	
1	EV Network	EV Network Web	Geocode St	Latitude	Longitude	Date Last Co	ID	Updated At	Owner
2	Non-Networked		GPS	34.24831915	-118.3879714	1/10/2023	1517	2023-02-15 22:4	LG
3	Non-Networked		200-8	34.052542	-118.448504	1/10/2023	1519	2023-02-15 22:4	LG
4	Non-Networked		GPS	34.040539	-118.271387	1/10/2023	1523	2023-02-14 15:5	P
5	Non-Networked		GPS	34.059133	-118.248589	1/10/2023	1525	2023-02-15 22:4	LG
6	Non-Networked		GPS	33.759802	-118.096665	1/10/2023	1531	2023-02-15 22:4	LG
7	Non-Networked		200-8	33.770508	-118.265628	1/10/2023	1552	2023-02-15 22:4	LG

Station_Name	Address	City	State_Province	Zip_Postal_Code	Start_Date___Time	Start_Time_
BOULDER / BASELINE ST1	900 Baseline Rd	Boulder	Colorado	80,302	1/19/2018, 11:29 PM	MDT
BOULDER / BASELINE ST1	900 Baseline Rd	Boulder	Colorado	80,302	2/9/2018, 8:57 AM	MDT
BOULDER / JUNCTION ST1	2280 Junction Pl	Boulder	Colorado	80,301	1/1/2018, 9:49 AM	MDT
BOULDER / REC CENTER ...	1360 Gillaspie Dr	Boulder	Colorado	80,305	1/25/2018, 10:58 AM	MDT
BOULDER / FACILITIES ST1	1745 14th street	Boulder	Colorado	80,302	2/5/2018, 9:01 AM	MDT
BOULDER / ATRIUM ST1	1770 13th St	Boulder	Colorado	80,302	1/31/2018, 12:45 AM	MDT
BOULDER / BASELINE ST1	900 Baseline Rd	Boulder	Colorado	80,302	1/20/2018, 3:42 AM	MDT
BOULDER / JUNCTION ST1	2280 Junction Pl	Boulder	Colorado	80,301	1/2/2018, 12:52 AM	MDT
COMM VITALITY / 1000WA...	900 Walnut St	Boulder	Colorado	80,302	2/9/2018, 9:53 AM	MDT
BOULDER / REC CENTER ...	1360 Gillaspie Dr	Boulder	Colorado	80,305	1/25/2018, 10:59 AM	MDT

Station Name	MAC Address	Org Name	Start Date	Start Time Zone	End Date	End Time Zone	Transaction Date (Pacific Time)	Total Duration (hh:mm:ss)	Charging Time (hh:mm:ss)	Energy (kWh)	GHG Savings (kg)
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/29/2011 20:17	PDT	7/29/2011 23:20	PDT	7/29/2011 23:20	3:03:32	1:54:03	6.249457	2.625
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/30/2011 0:00	PDT	7/30/2011 0:02	PDT	7/30/2011 0:02	0:02:06	0:01:54	0.106588	0.045
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/30/2011 8:16	PDT	7/30/2011 12:34	PDT	7/30/2011 12:34	4:17:32	4:17:28	14.951777	6.28
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/30/2011 14:51	PDT	7/30/2011 16:55	PDT	7/30/2011 16:55	2:03:24	2:02:58	7.159643	3.007
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/30/2011 18:51	PDT	7/30/2011 20:03	PDT	7/30/2011 20:03	1:11:24	0:43:54	1.957765	0.822
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/30/2011 21:38	PDT	7/31/2011 1:30	PDT	7/31/2011 1:30	3:52:13	1:30:58	4.80288	2.017
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/31/2011 4:33	PDT	7/31/2011 10:40	PDT	7/31/2011 10:40	6:06:19	4:56:47	17.171463	7.212
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/31/2011 12:25	PDT	7/31/2011 13:35	PDT	7/31/2011 13:35	1:09:54	1:09:49	3.799148	1.596
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	7/31/2011 17:41	PDT	7/31/2011 22:31	PDT	7/31/2011 22:31	4:49:46	4:41:16	16.238552	6.82
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/1/2011 19:11	PDT	8/1/2011 22:34	PDT	8/1/2011 22:34	3:23:34	1:35:08	5.041244	2.117
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/2/2011 12:26	PDT	8/2/2011 13:27	PDT	8/2/2011 13:27	1:00:50	1:00:43	3.516131	1.477
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/2/2011 14:00	PDT	8/2/2011 15:13	PDT	8/2/2011 15:13	1:12:43	1:12:33	4.197217	1.763
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/2/2011 19:03	PDT	8/2/2011 21:30	PDT	8/2/2011 21:30	2:26:46	2:26:43	8.497075	3.569
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/3/2011 5:35	PDT	8/3/2011 9:15	PDT	8/3/2011 9:15	3:40:53	3:40:37	12.837713	5.392
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/3/2011 8:18	PDT	8/3/2011 10:02	PDT	8/3/2011 10:02	1:44:32	1:40:14	4.620635	1.941
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/3/2011 9:38	PDT	8/3/2011 10:27	PDT	8/3/2011 10:27	0:48:46	0:48:19	2.340949	0.983
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/3/2011 14:24	PDT	8/3/2011 15:27	PDT	8/3/2011 15:27	1:02:45	1:02:33	3.013084	1.265
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/3/2011 22:33	PDT	8/4/2011 8:32	PDT	8/4/2011 8:32	9:59:07	3:09:13	8.832448	3.71
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/4/2011 11:57	PDT	8/4/2011 15:37	PDT	8/4/2011 15:37	3:40:26	3:40:18	10.598673	4.451
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/4/2011 17:45	PDT	8/4/2011 18:19	PDT	8/4/2011 18:19	0:34:29	0:34:23	2.003678	0.842
PALO ALTO CA / HAMILTON #2	000D:6F00:009E:D39E	City of Palo Alto	8/4/2011 20:08	PDT	8/4/2011 21:44	PDT	8/4/2011 21:44	1:35:44	1:23:19	3.627027	1.523
PALO ALTO CA / HAMILTON #1	000D:6F00:015A:9D76	City of Palo Alto	8/4/2011 20:13	PDT	8/4/2011 23:11	PDT	8/4/2011 23:11	2:58:33	2:57:27	8.74036	3.671

Data Statistics

Statistics for Boulder EV Load Data

Dataset	RAW Data Statis	Preprocessed Data Statistics
Format	CSV	CSV
Count of Features used	16	16
Columns with Nulls	0	0
Duplicate rows	0	0
Total Rows	24081	24081

Statistics for Palo Alto Data

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	CSV	CSV
Count of Features used	33	11
Columns with Nulls	4	0
Duplicate rows	0	0
Total Rows	9999	9999

Statistics for ACN Data

CalTech

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	JSON	CSV
Count of Features used	13	10
Columns with Nulls	3	0
Duplicate rows	0	0
Total Rows	29128	29128

JPL

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	JSON	CSV
Count of Features used	13	10
Columns with Nulls	3	0
Duplicate rows	0	0
Total Rows	29724	29724

Office

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	JSON	CSV
Count of Features used	13	10
Columns with Nulls	3	0
Duplicate rows	0	0
Total Rows	1481	1481

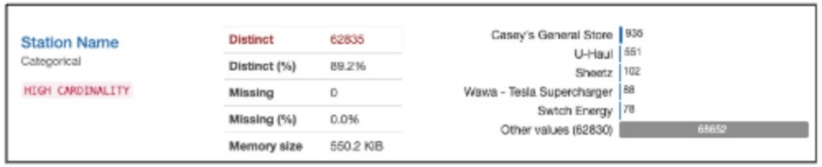
ACN- Merged Data

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	JSON	CSV
Count of Features used	13	10
Columns with Nulls	3	0
Duplicate rows	0	0
Total Rows	1481	1481

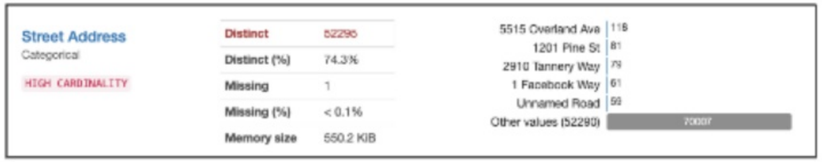
Statistics for Alternative Fuels Data

Dataset	RAW Data Statistics	Preprocessed Data Statistics
Format	CSV	CSV
Count of Features used	65	35
Columns with Nulls	30	0
Duplicate rows	0	0
Total Rows	8584	8584

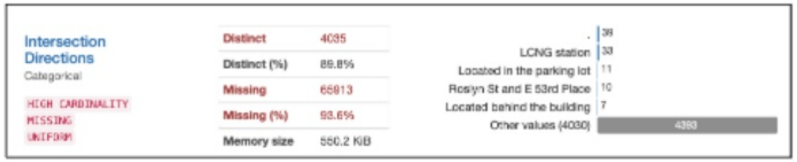
Dataset Characteristics (Fuel Type)

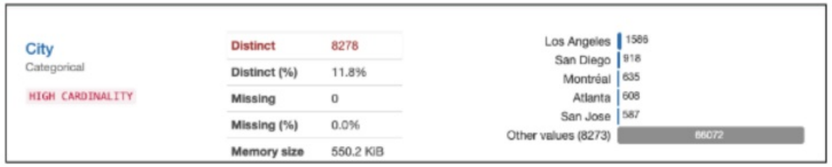


Dataset Characteristics (Station Name)

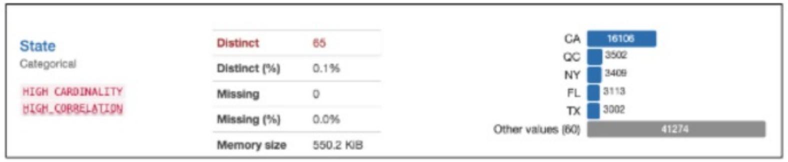


Dataset Characteristics (street Address)

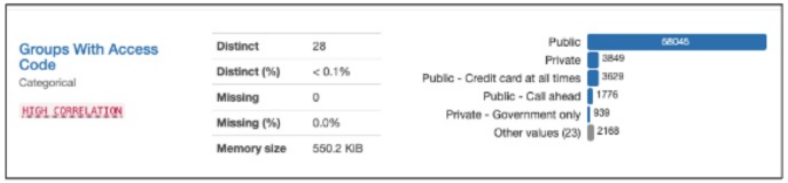




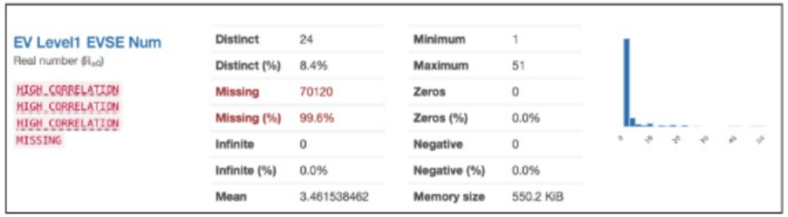
Dataset Characteristics (city)



Dataset Characteristics (state)



Dataset Characteristics (Group with access code)

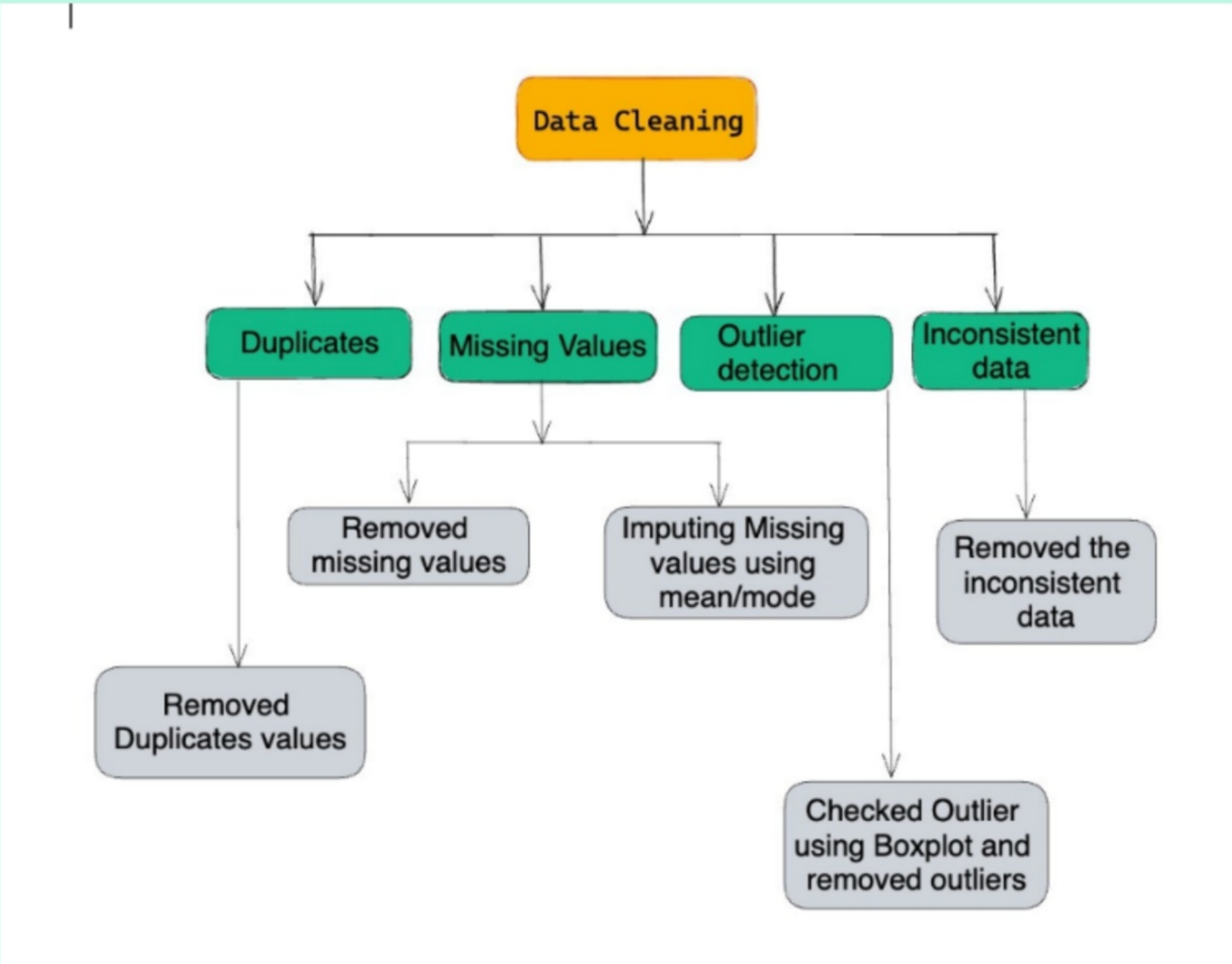


Dataset Characteristics (EV Level 1)

Data Preparation

- Dataset Splitting
- Data transformation
- Data Pre-processing

Data Pre-Processing



Handling NULL values and Zero Variance Columns

```
df_elec_filtered=df_elec[['station_name','latitude', 'longitude', 'city', 'zip', 'ev_dc_fast_num',
        'ev_level1_evse_num', 'ev_level2_evse_num', 'ev_network']].copy(deep=True)
```

```
from sklearn.cluster import KMeans
```

```
df_elec_filtered.isna().sum()
```

```
station_name            0
latitude                0
longitude               0
city                    0
zip                     0
ev_dc_fast_num      14211
ev_level1_evse_num  16056
ev_level2_evse_num   1698
ev_network              0
dtype: int64
```

```
df_elec_filtered['ev_dc_fast_num'] = df_elec_filtered['ev_dc_fast_num'].fillna(df_elec_filtered['ev_dc_fast_num'].median())
df_elec_filtered['ev_level1_evse_num'] = df_elec_filtered['ev_level1_evse_num'].fillna(df_elec_filtered['ev_level1_evse_num'].median())
df_elec_filtered['ev_level2_evse_num'] = df_elec_filtered['ev_level2_evse_num'].fillna(df_elec_filtered['ev_level2_evse_num'].median())
```

```
# Data preprocessing
data = data.drop(columns=[ 'MAC_Address', 'Start_Date', 'Start_Time_Zone',
                          'End_Date', 'End_Time_Zone', 'Transaction_Date_(Pacific_Time)',
                          'GHG_Savings_(kg)', 'Gasoline_Savings_(gallons)', 'Port_Type',
                          'Plug_Type' , 'Address_1',
                          'Ended_By', 'Plug_In_Event_Id', 'Driver_Postal_Code', 'User_ID'])
data['Total_Duration_(hh:mm:ss)'] = pd.to_timedelta(data['Total_Duration_(hh:mm:ss)']).dt.total_seconds()
data['Charging_Time_(hh:mm:ss)'] = pd.to_timedelta(data['Charging_Time_(hh:mm:ss)']).dt.total_seconds()
data['Energy_(kWh)'] = pd.to_numeric(data['Energy_(kWh)'])
```

```
list_null_cols= []
for i in data.columns:
    if data[i].isna().sum() == data.shape[0]:
        list_null_cols.append(i)
```

```
list_single_cols = []
for i in data.columns:
    if data[i].nunique() == 1:
        list_single_cols.append(i)
```

```
data.drop(list_single_cols+list_null_cols,axis=1,inplace=True)
```

Data Transformation

```
# Feature engineering
data['Transaction_Date'] = pd.to_datetime(data['Transaction_Date_(Pacific_Time)'])
data['transaction_time'] = data['Transaction_Date'].apply(lambda x: x.timestamp())
data['hour'] = data['Transaction_Date'].dt.hour
data['dayofweek'] = data['Transaction_Date'].dt.dayofweek
data['month'] = data['Transaction_Date'].dt.month
data['season'] = (data['Transaction_Date'].dt.month % 12 + 3)//3
data['is_weekend'] = np.where(data['dayofweek'] >= 5, 1, 0)
```

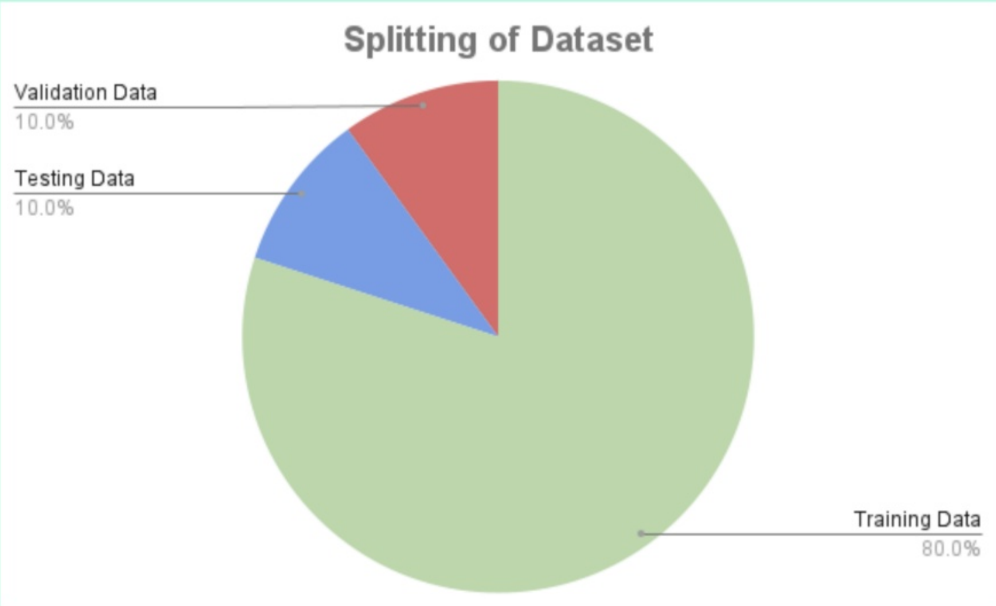
```
data['week_number'] = ((data['Transaction_Date'] - data['Transaction_Date'].min()) / np.timedelta64(1, 'W')).astype(int)
```

```
data.columns=data.columns.str.replace(' ','_')
```

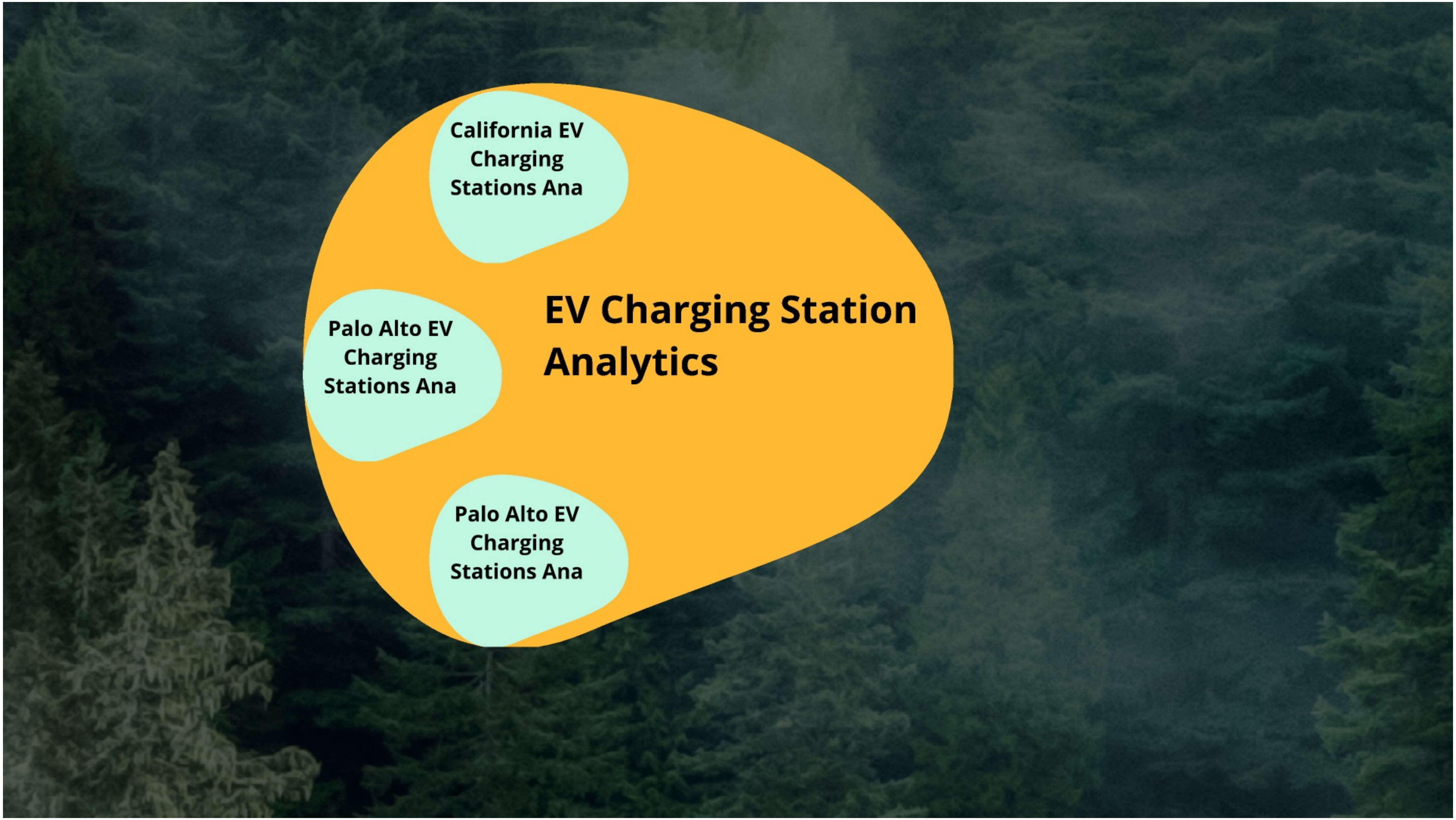
Data Transformation

```
# Import the OneHotEncoder class
from sklearn.preprocessing import OneHotEncoder

# Create an instance of the OneHotEncoder
encoder = OneHotEncoder(handle_unknown='ignore')

# Fit the encoder on the 'Plug_Type' column
encoder.fit(X_train[['Station_Name']])

# Transform the 'Plug_Type' column using the encoder
X_train_encoded = pd.DataFrame(encoder.transform(X_train[['Station_Name']]).toarray(), columns=encoder.get_feature_names_out(['Station_Name']))
X_train_encoded.index = X_train.index

# Concatenate the encoded 'Plug_Type' column with the rest of the X_train data
X_train = pd.concat([X_train.drop(columns=['Station_Name']), X_train_encoded], axis=1)

# Repeat the same process for X_test
X_test_encoded = pd.DataFrame(encoder.transform(X_test[['Station_Name']]).toarray(), columns=encoder.get_feature_names_out(['Station_Name']))
X_test_encoded.index = X_test.index
X_test = pd.concat([X_test.drop(columns=['Station_Name']), X_test_encoded], axis=1)
```

Dataset Splitting



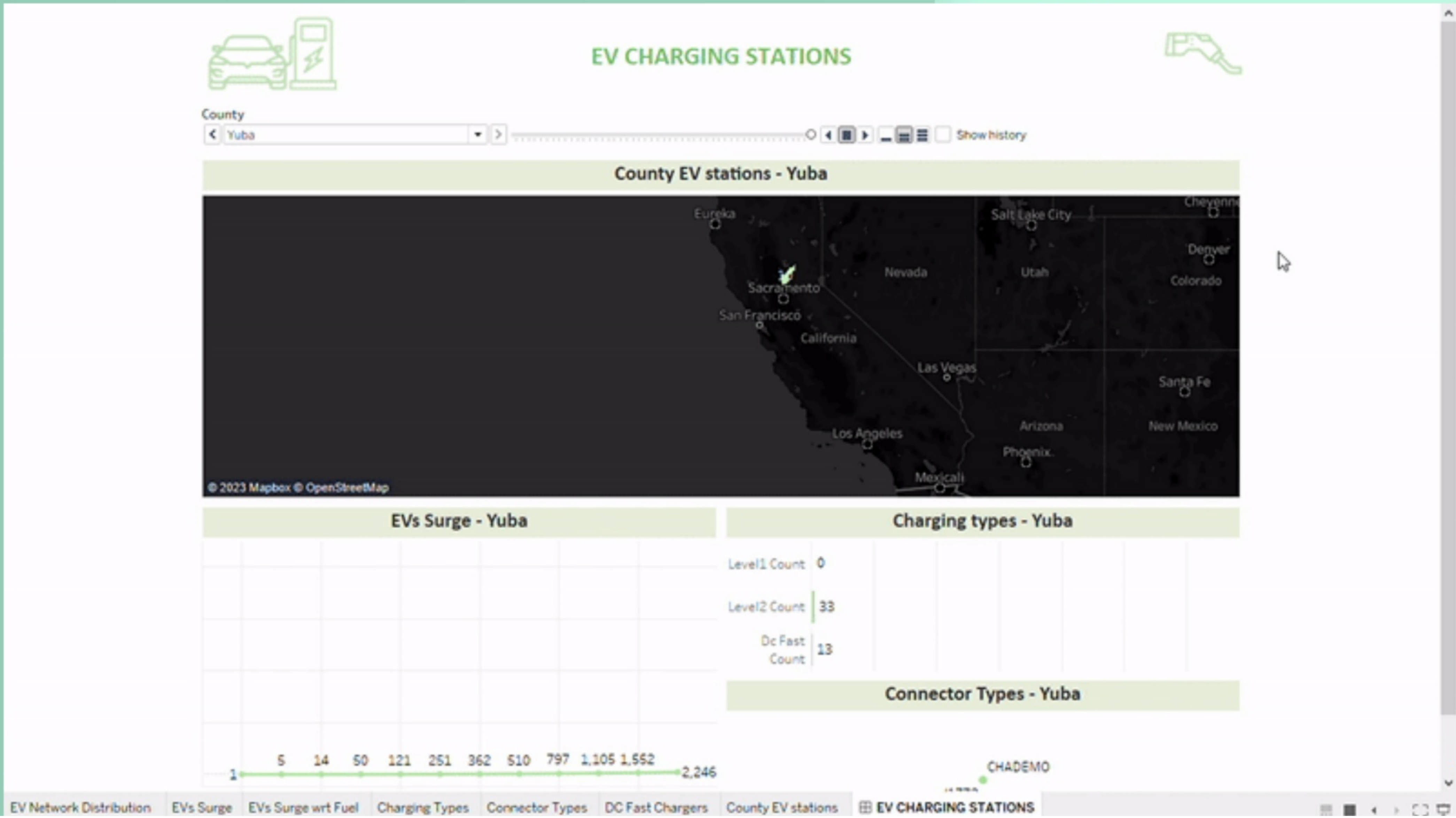

```
# Define the train-test split ratio
train_ratio = 0.8
train_size = int(len(X) * train_ratio)
```


California EV Charging Stations Ana
Palo Alto EV Charging Stations Ana
EV Charging Station Analytics
Palo Alto EV Charging Stations Ana

Evaluation Metrics (Ca Ev Charging)

- We implemented a data driven approach to analyze the existing EV charging stations.
- Various metrics like charging duration, energy consumed, ev network distribution within California are studied and visualized.
- Further insights were also achieved from analyzing the ev sales, charger types etc.

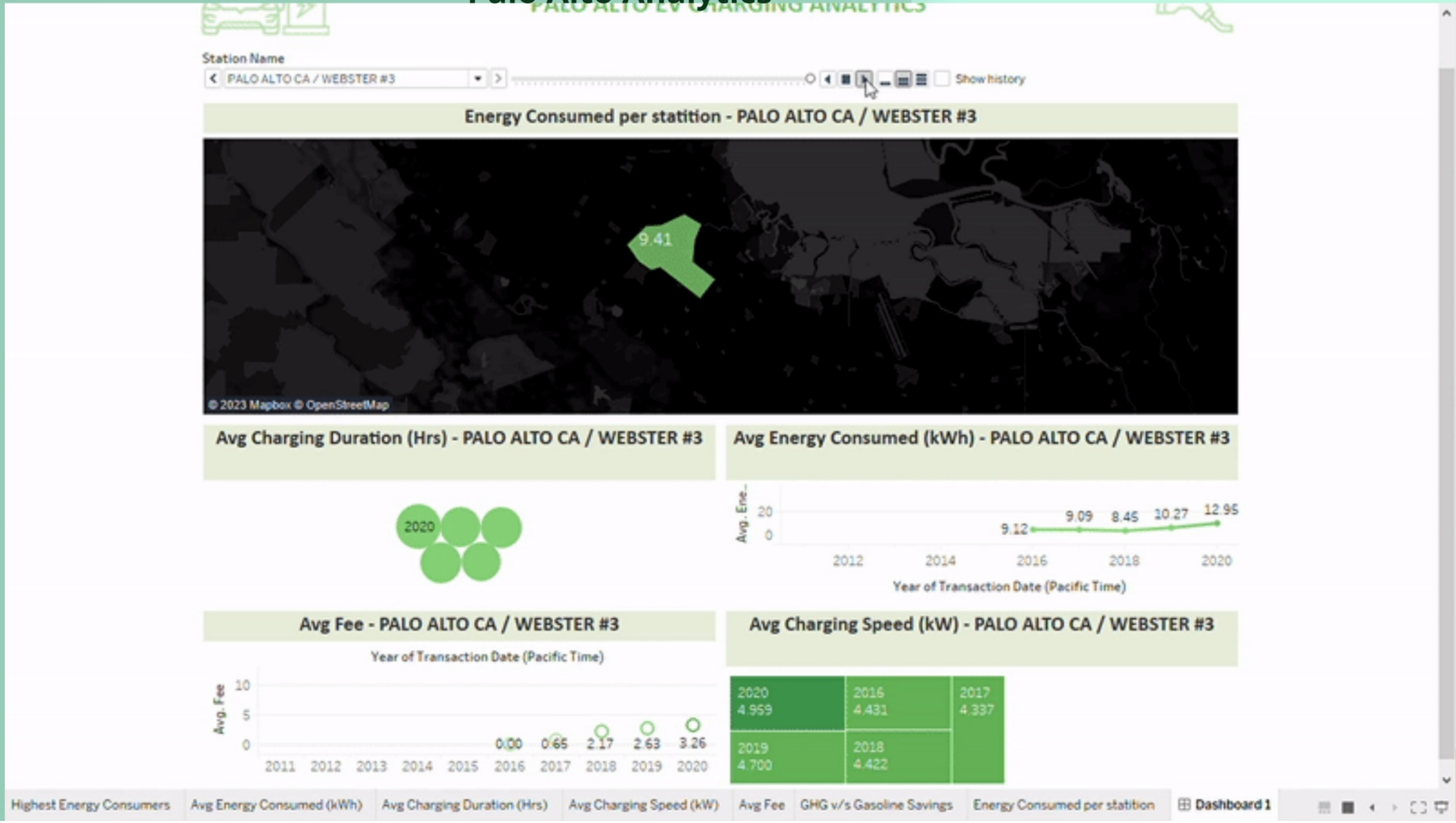
Tableau Dashboard - Ca EV Charging



Evaluation Metrics (Palo Alto)

- From The previous analysis, Palo Alto was one the top performing cities in California. We wanted to get a clearer picture by just analyzing the statistics of Palo Alto EV Charging stations.
- Three major stations were addressed for analyzing purposes.
- Various metrics like charging duration, energy consumed are studied and visualized.

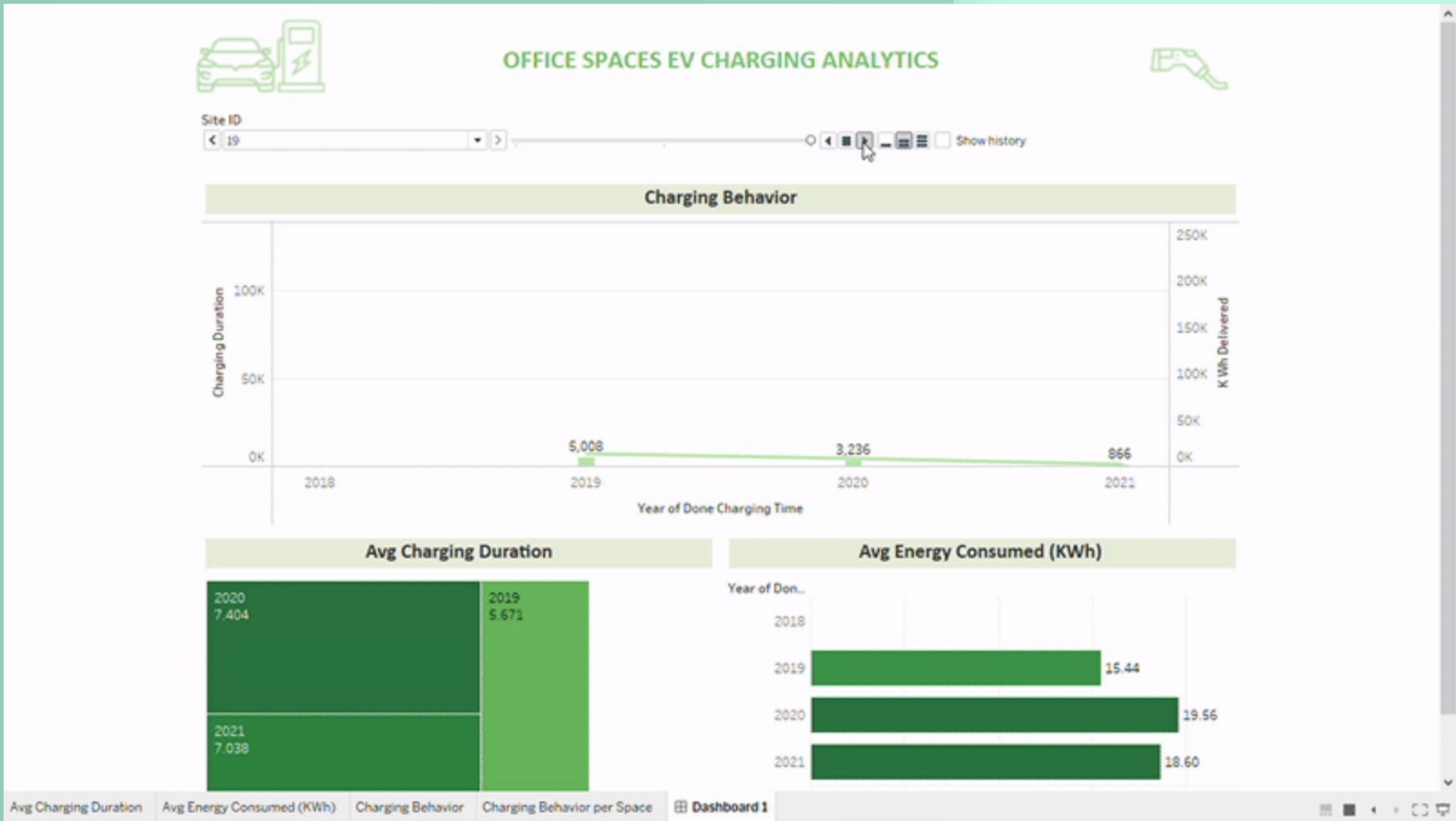
Palo Alto Analytics

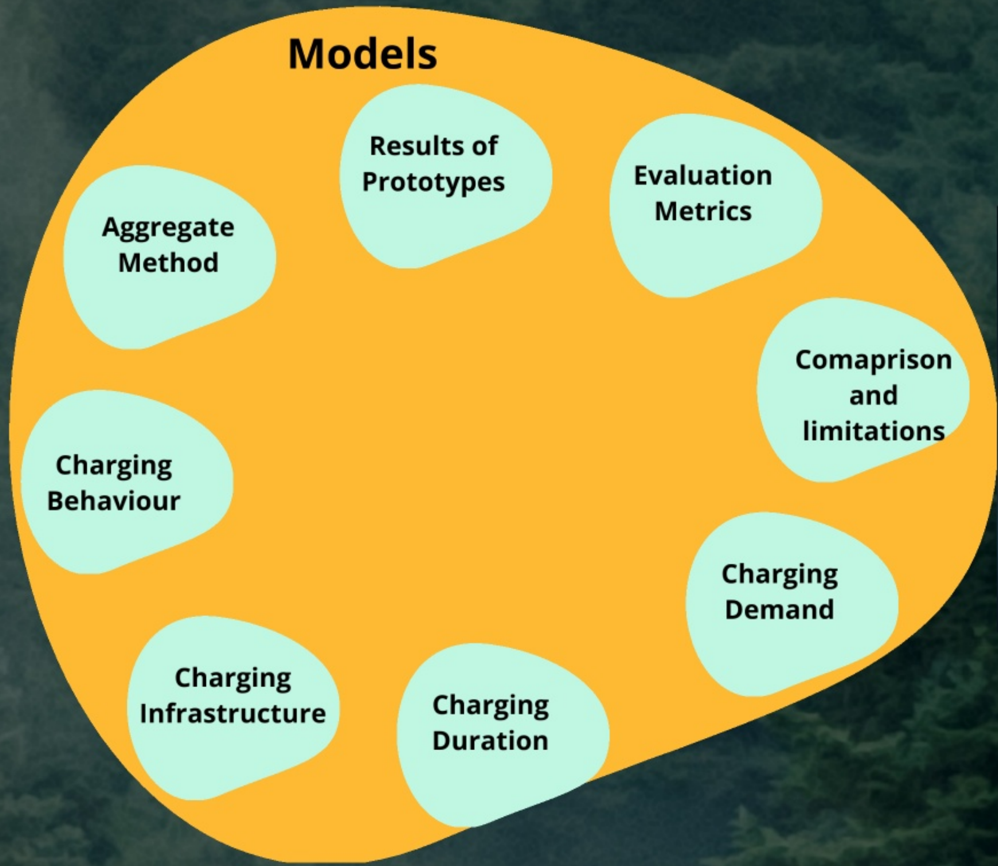


Evaluation Metrics (Palo Alto EV Station)

- From the previous analysis, we noticed that office spaces were one of the most sought out facility types to provide as well as be utlized for charging EVs.
- Charging consumption, Charging duration and the trend in charging behavior were addressed and visualized in this analysis.

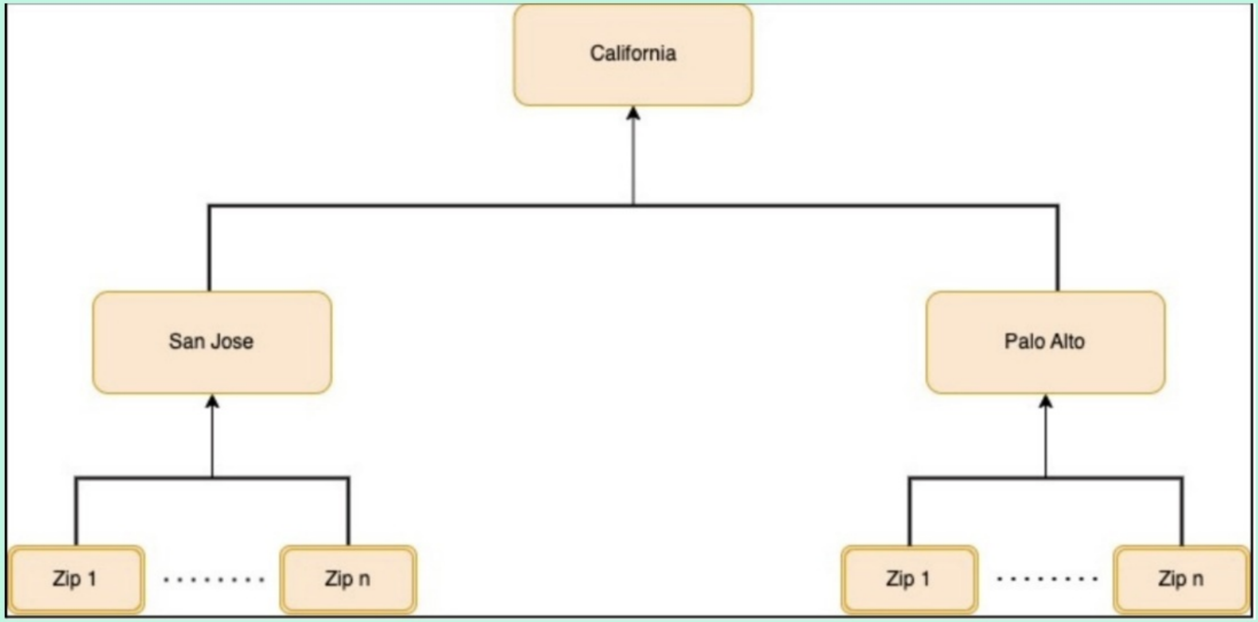
Tableau Dashboard (Palo Alto EV Station)

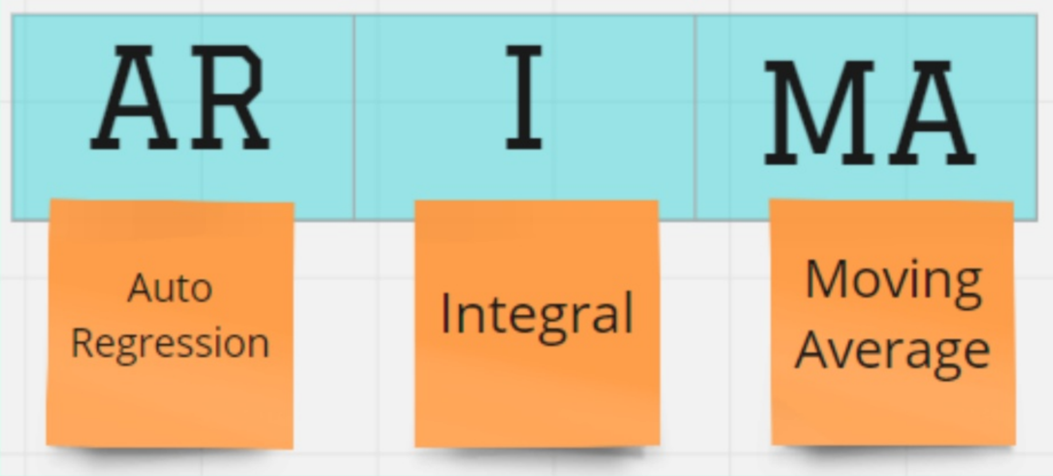



Models
Results of Prototypes
Evaluation Metrics
Aggregate Method
Comaprison and limitations
Charging Behaviour
Charging Demand
Charging Infrastructure
Charging Duration

Aggregate Method:

The bottom-up approach, also known as the aggregate method, involves predicting charging time for lower-level regions (e.g., zip codes) and aggregating those predictions to higher-level regions (e.g., cities, counties, states). This approach utilizes specific models for each level of aggregation, such as SARIMA or LSTM, and combines predictions through aggregation methods (e.g., mean, median). It allows for spatial hierarchy considerations and provides estimates of charging time for higher-level regions based on underlying predictions at lower levels.





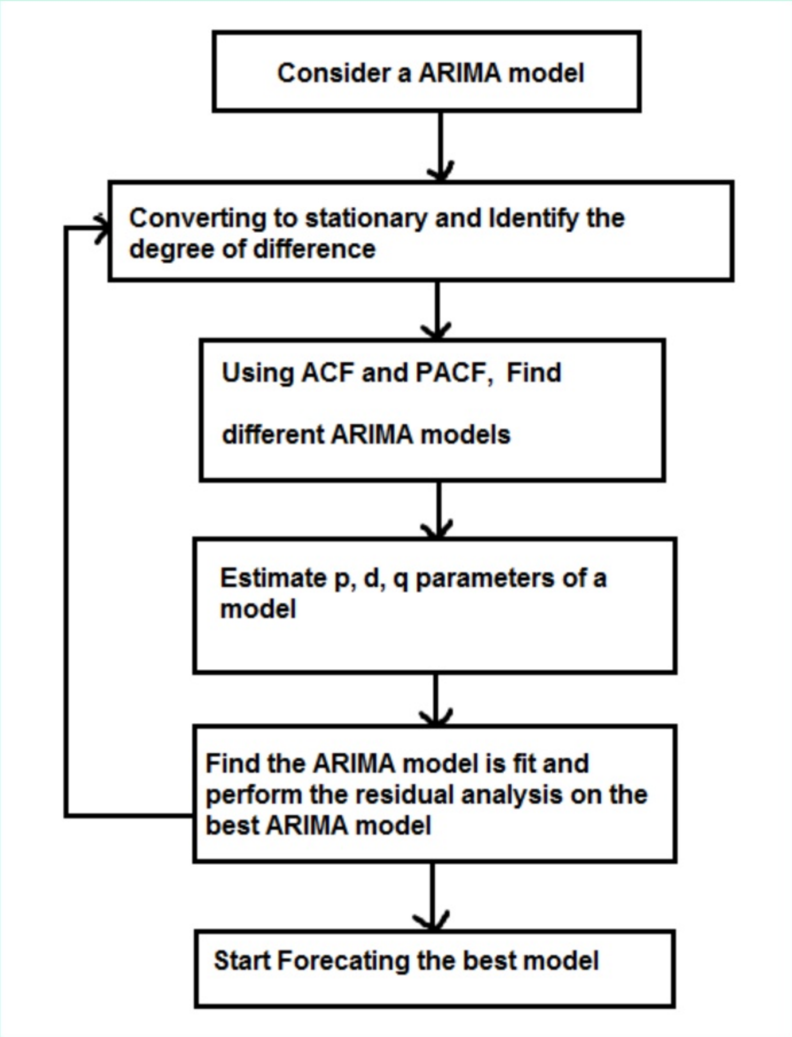
Charging Behaviour:

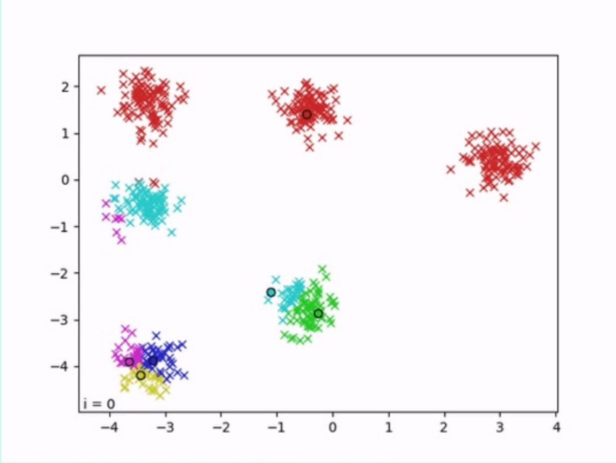
Problem Statement: Understanding the charging behavior of electric vehicle owners to guide the design and management of charging infrastructure, involving the analysis of charging trends, demographics, and other influencing factors.

The ARIMA and LSTM models were chosen due to their capacity to handle time series data effectively, with ARIMA managing data with a trend and LSTM capturing long-term dependencies in data.

ARIMA is a forecasting technique that combines autoregressive, integrated, and moving average models, while LSTM is a type of recurrent neural network that can learn and remember over long sequences.

ARIMA Model Architecture





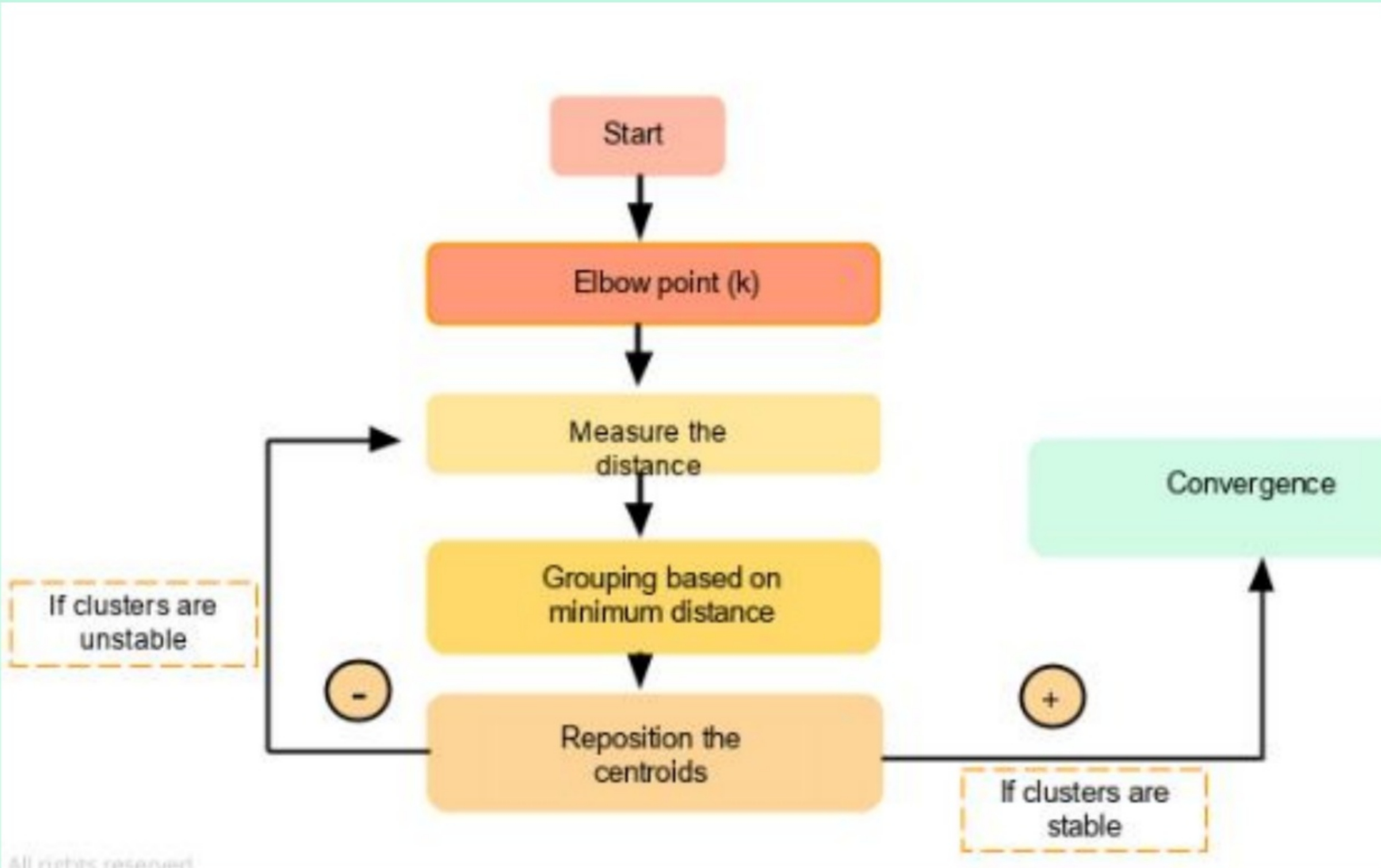
Charging Infrastructure:

Establishing the optimal locations for new electric vehicle charging stations based on the grouping of vehicle owners by charging behavior using enhanced K-means clustering.

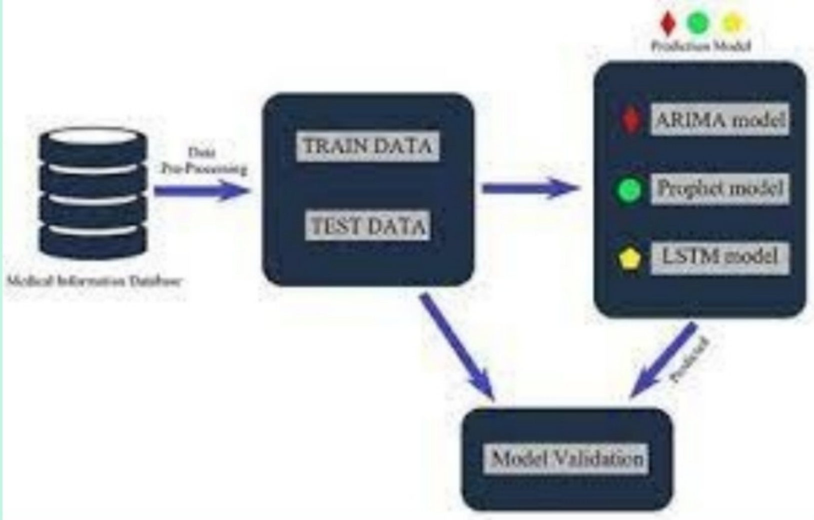
Enhanced K-means clustering was chosen for its capability to handle large datasets and incorporate additional factors for more accurate clustering, making it more effective in managing outliers than traditional K-means.

Enhanced K-means is a modified version of the K-means clustering algorithm, a popular unsupervised learning technique used to find groups in data. It incorporates additional factors and techniques to improve the clustering results.

Enhanced K means Clustering Architecture



Charging Duration:

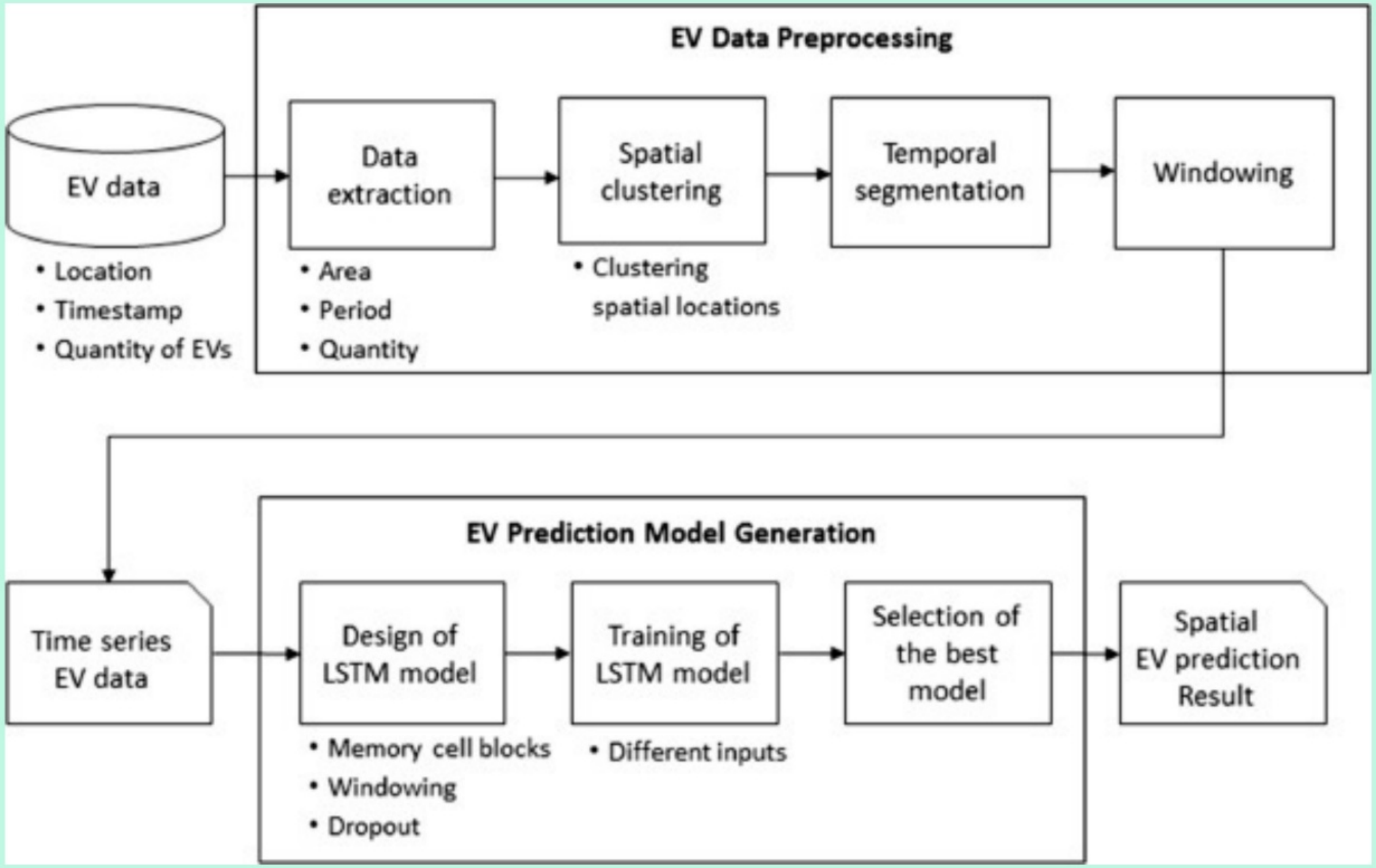


Determining the ideal charging time for electric vehicles, balancing between sufficient charge for the journey and minimizing time at the charging station, through analyzing battery capacity, driving habits, and charging rates.

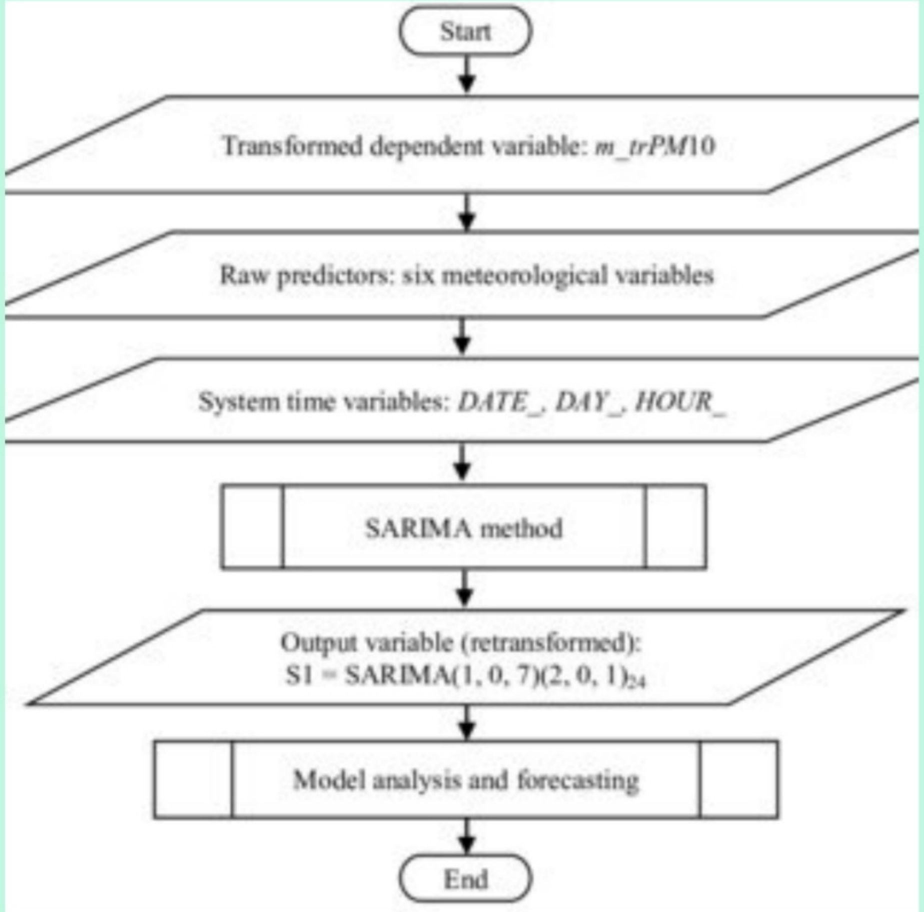
The hybrid model of SARIMA, Prophet, and LSTM was used for its ability to handle various data patterns and combine the strengths of these individual models for a more accurate prediction.

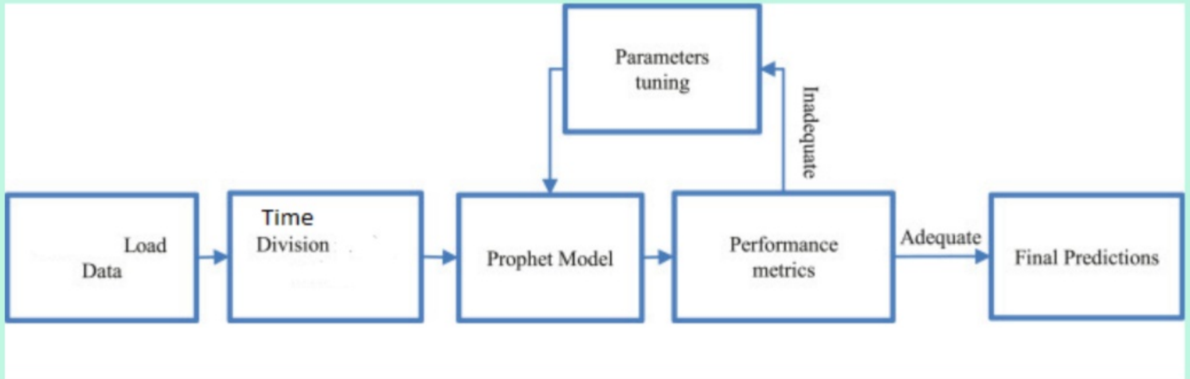
SARIMA is an extension of ARIMA that handles seasonal variations, Prophet is a procedure for forecasting time series data based on an additive model, and LSTM is a recurrent neural network model that can learn dependencies in long sequences.

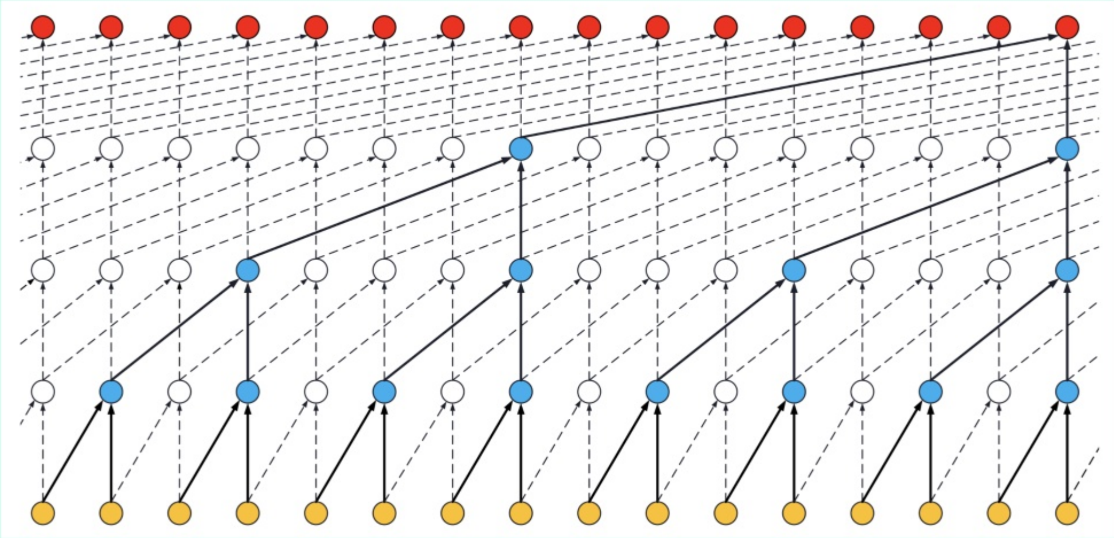
LSTM Model Architecture



Model Architectures: SARIMA, Prophet







Charging Demand:

Predicting the demand for electric vehicle charging at different times and locations to ensure sufficient infrastructure and avoid grid congestion, based on driving habits, charging patterns, and other influencing variables.

The Temporal Convolutional Network (TCN) and LSTM with Attention were chosen due to their ability to capture both local and long-term patterns, and weigh the importance of different inputs.

TCN is a type of convolutional neural network for sequential data that outperforms RNNs for longer sequences, while LSTM with Attention is an extension of the LSTM model that includes an attention mechanism, which allows it to weigh the importance of different inputs in the sequence.

Comparison & Limitations

Charging Behavior

ARIMA model & LSTM

These models use historical charging data for future behavior prediction. ARIMA analyzes and forecasts time series data with trends, while LSTM captures long-term dependencies in data effectively. Their limitations include ARIMA's requirement for manual parameter configuration and LSTM's computational intensity and risk of overfitting.

Charging Duration

Hybrid Model (SARIMA+Prophet+LSTM using Aggregate method)

This method combines SARIMA, Prophet, and LSTM models to forecast charging duration, aggregating results for increased accuracy. The approach harnesses the strengths of each model, reducing the impact of individual model weaknesses. It's more computationally expensive and performance-dependent on each model.

Charging Demand

Temporal Convolutional Network (TCN) + Long Short-Term Memory with Attention (LSTM-Attention)

TCN and LSTM-Attention models predict future charging demand, focusing on local and long-term patterns. LSTM-Attention weighs the importance of different inputs. These models are computationally heavy, and the attention mechanism may overemphasize less relevant data.

EV location Recommendation

Enhanced Kmeans Clustering,This technique groups EV owners based on charging behavior, recommending optimal new station locations. It effectively handles large datasets and outliers, and incorporates additional factors for improved clustering. It still requires specified cluster numbers and assumes clusters are spherical with equal variance.

Evaluation Metrics

Different metrics which we used for Evaluation

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

- Root Mean Square Error (RMSE): The square root of the average of squared differences between prediction and actual observation.
- Mean Absolute Error (MAE): The average of the absolute differences between prediction and actual observation.
- Mean Squared Error (MSE): The average of the squares of the differences between prediction and actual observation.
- R2 Score (Coefficient of Determination): A statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model.

Divide by the total number of data points

Predicted output value

Actual output value

$$MAE = \boxed{\frac{1}{n}} \underset{\text{Sum of}}{\Sigma} \underbrace{\left| \boxed{y} - \boxed{\hat{y}} \right|}_{\text{The absolute value of the residual}}$$

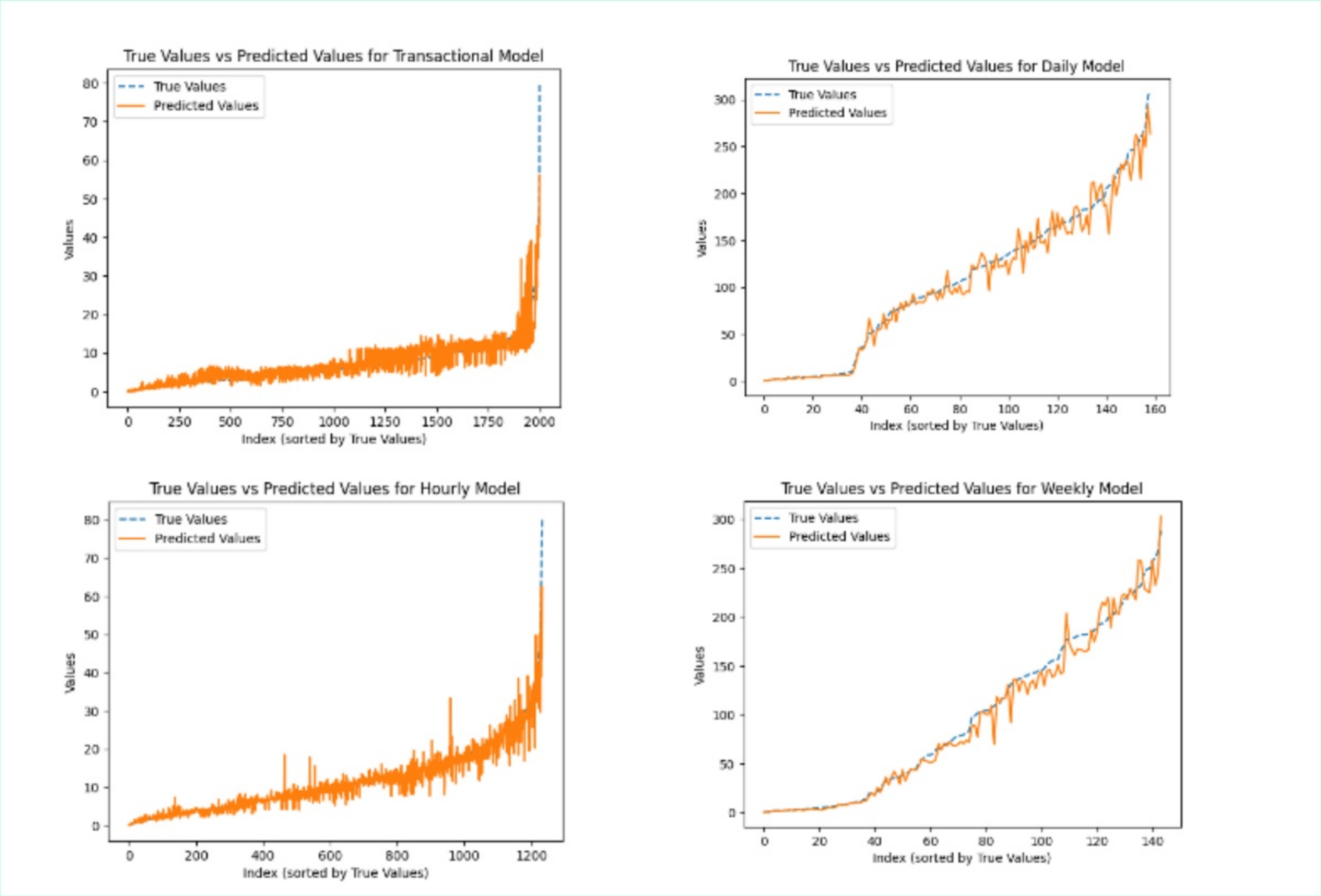
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

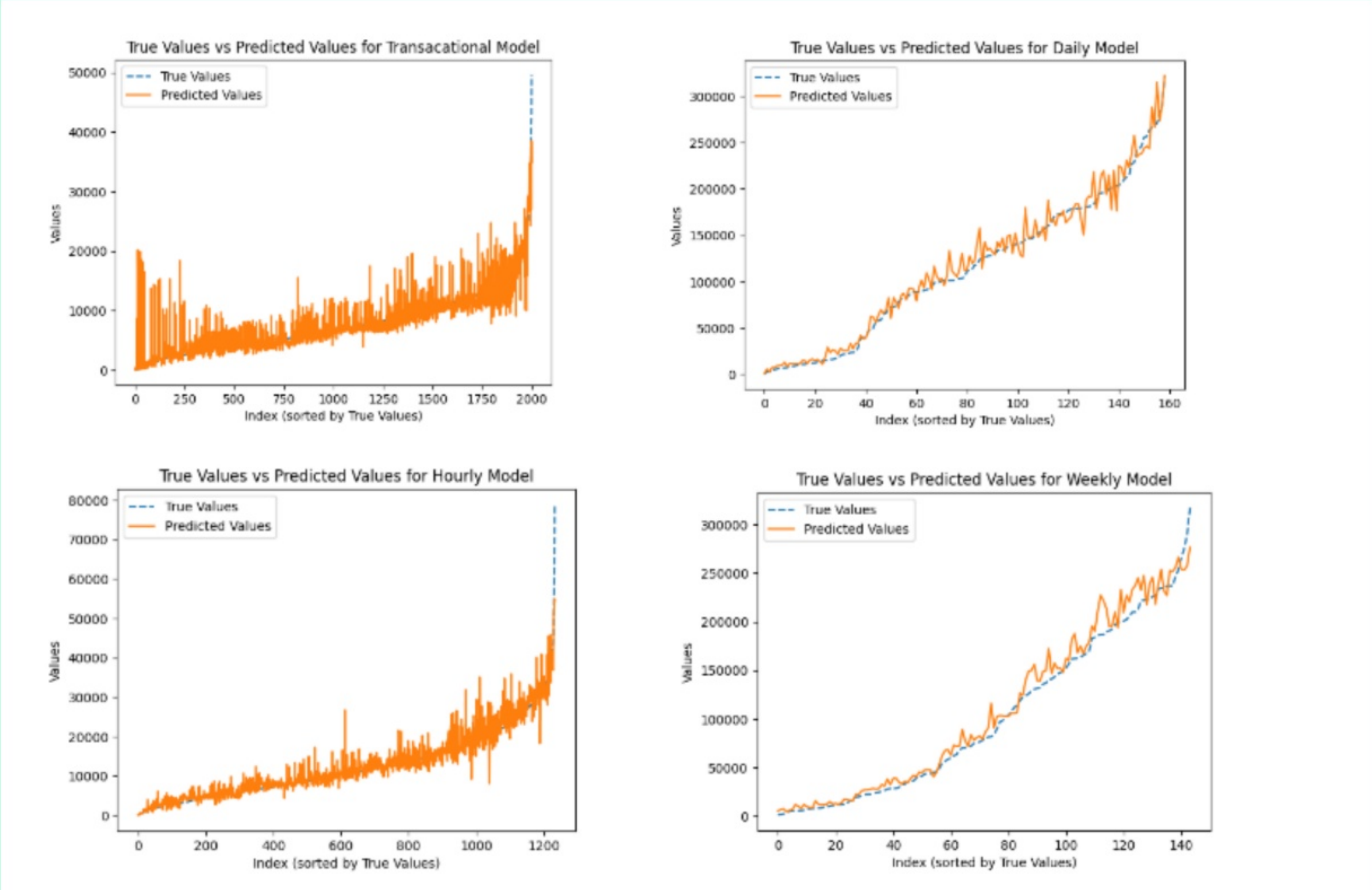
Results

Metrics/Dataset	Charging Duration Test	Charging Duration Train	Charging Behaviour Test	Charging Behaviour Train	Charging Demand Test	Charging Demand Train
MAE	1496	344	14	34	1.6763	0.3305
MSE	6934946	385770	69	38	6.9063	0.3879
RMSE	2633	621	26	6	2.6280	0.6228
R2	0.738	0.984	0.767	0.965	0.8029	0.9819

Charging Demand



Charging Duration




Recommendation
Duration, Demand and Behavior

Recommended Locations

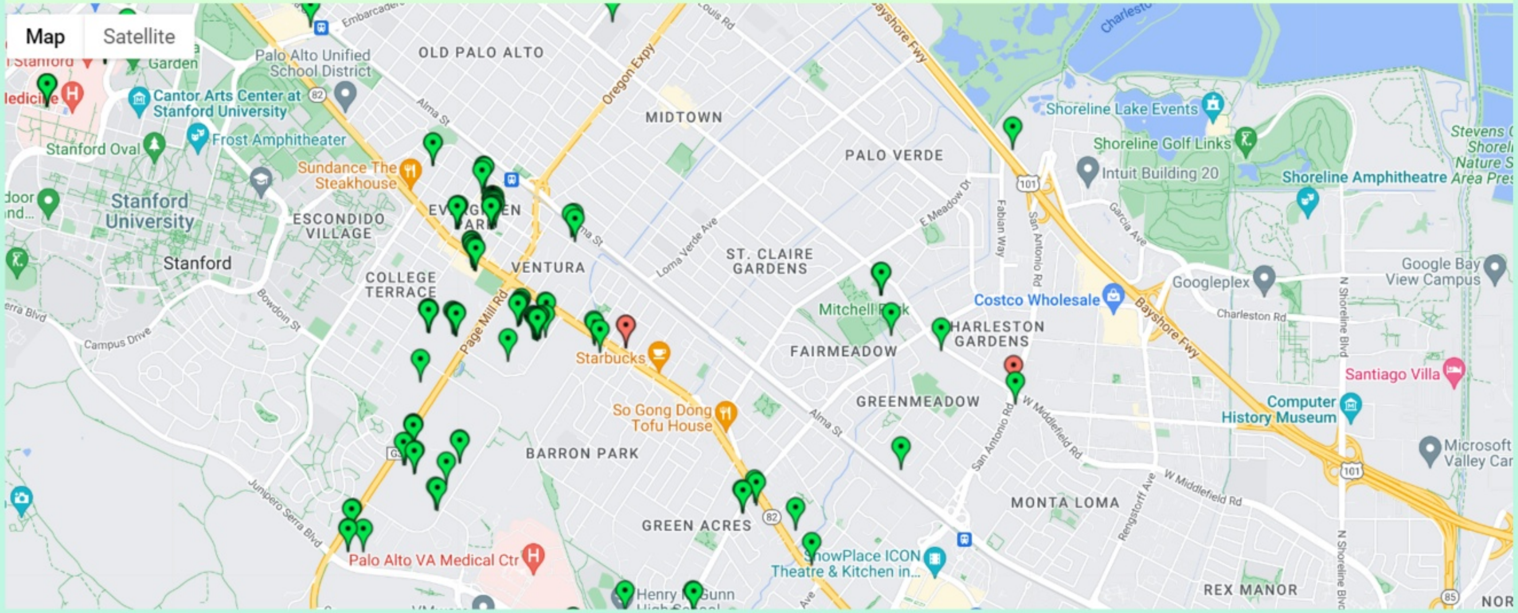
Filter By Fuel Type

All

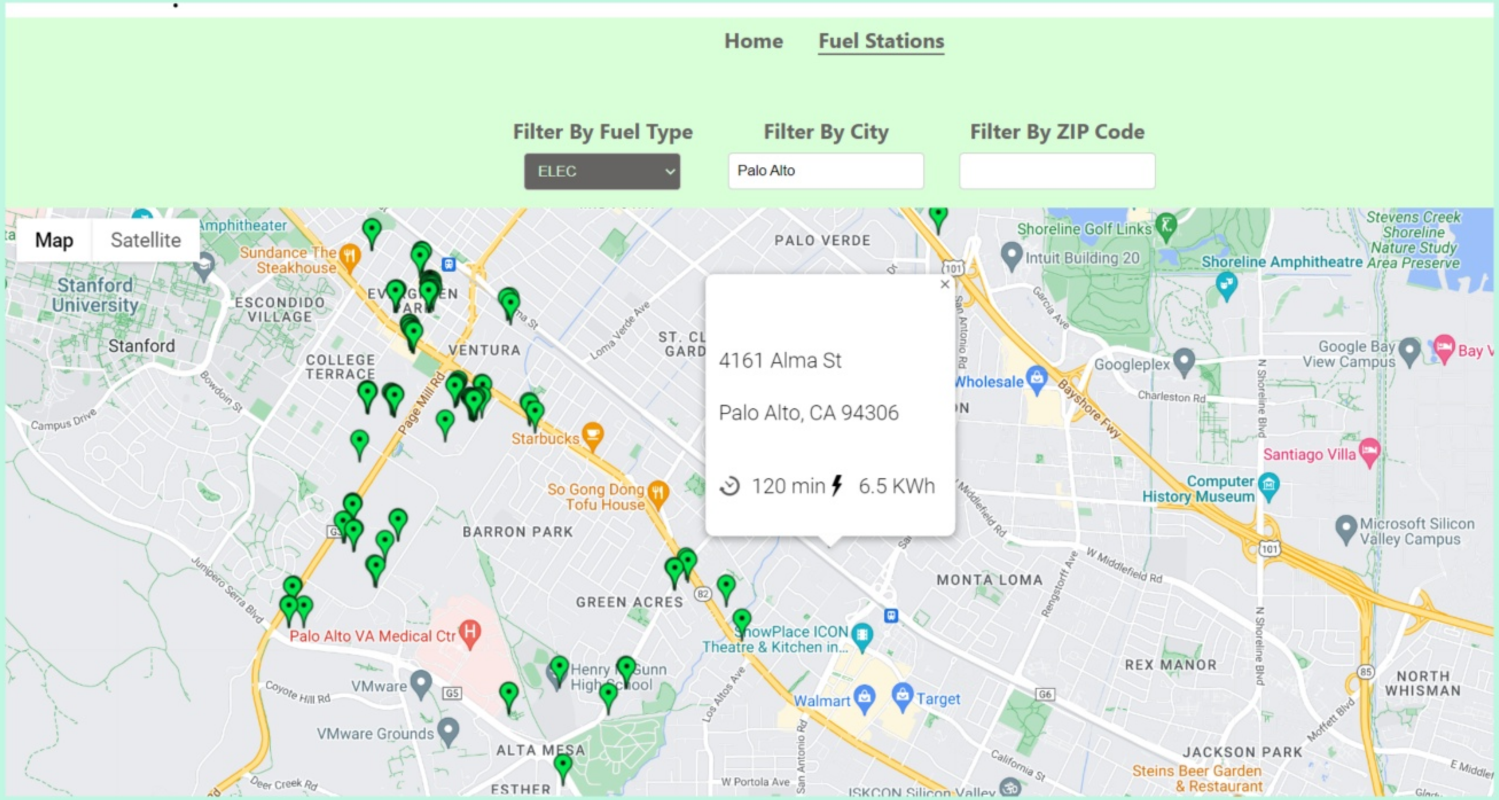
Filter By City

Palo Alto

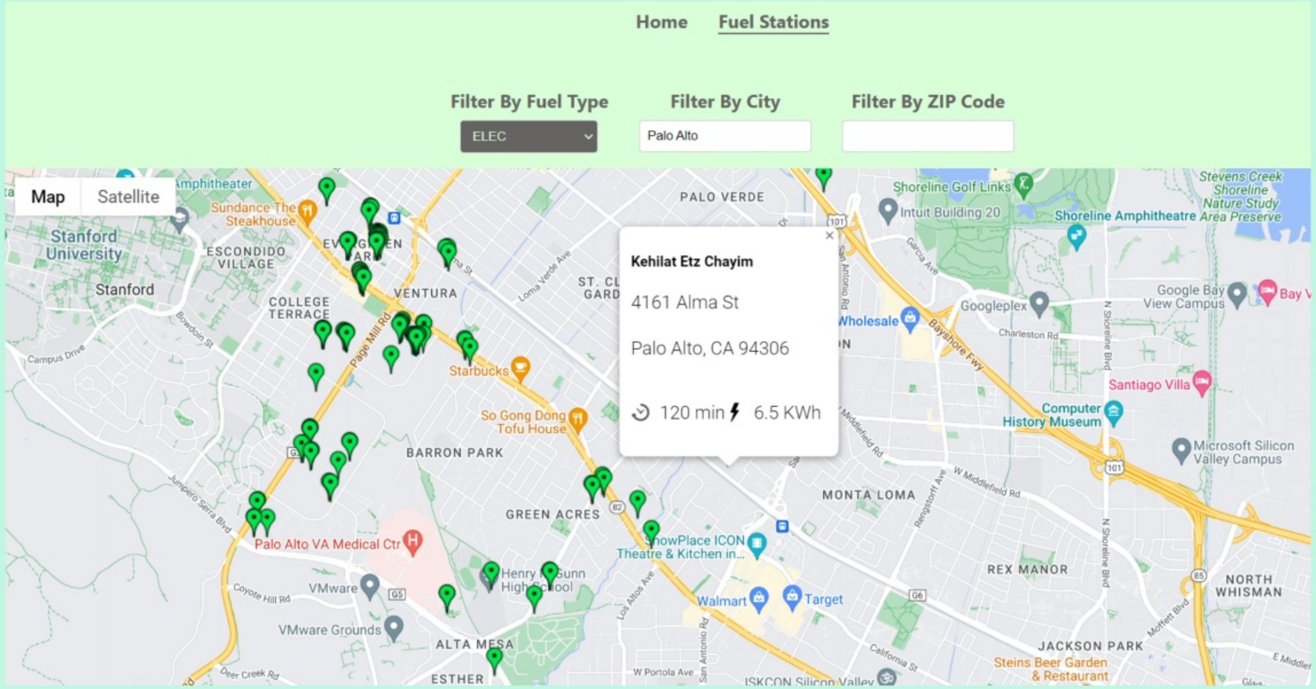
Filter By ZIP Code



Tooltip for Recommended Location



Duration, Demand & Behaviour



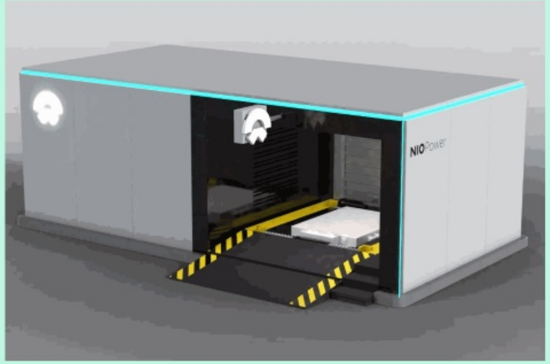
Comaprison table

Development in the Industry

Developments in the Industry

Battery Swap Tech

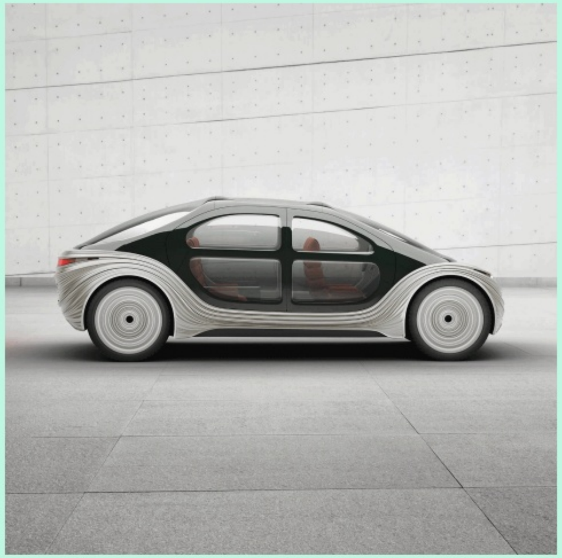
- Battery swapping replaces depleted EV batteries with fully charged ones, reducing charging time and range anxiety.
- NIO and Ample are leading companies that have automated the battery swapping process.
- Battery swapping improves EV ownership experience and enables better battery management.
- This technology extends the life of EVs by optimizing battery cycling and performance.



Zeekr's Battery Tech

- The battery is made by CATL, the world's largest battery maker.
- It has a capacity of 100 kWh and can charge from 10% to 80% in 15 minutes.
- It is expected to be used in the Zeekr 009, a new electric SUV from Geely.
- It is made using a new type of lithium iron phosphate (LFP) battery chemistry that is more energy dense and can charge faster than traditional lithium-ion batteries.

Chinese and Taiwan New EV Developments



- China leads the global EV market with domestic manufacturers like BYD and NIO driving innovation.
- Strong government policies in China have greatly supported EV adoption and market growth.
- Taiwan excels in EV component manufacturing, especially in battery technology, with major suppliers like Foxconn.
- Taiwan's commitment to sustainability is reflected in their efforts to establish an efficient EV ecosystem and contribute to the global EV market.

Rivian Charging Network

- Rivian is aiming to develop 3,500 charging stations across the United States by 2025. The company has invested $2 billion in its charging network, which will include a mix of DC fast chargers and Level 2 chargers.

- The DC fast chargers will be able to charge Rivian vehicles from 10% to 80% in about 20 minutes. The Level 2 chargers will be slower, but they will still be able to charge Rivian vehicles overnight.

- Rivian's charging network will be open to all electric vehicles, but Rivian owners will have priority access. The company is also working on a subscription program that will give Rivian owners unlimited access to its charging network.

Tesla Commercial Network

- Tesla is planning to launch a commercial fleet of electric semi-trucks in 2023 with a range of 500+ miles per single charge.
- The semi-trucks will be powered by Tesla's new 4680 battery cells, which are expected to offer a significant range improvement over the company's current battery cells.
- Tesla is also planning to launch a commercial fleet of electric delivery vans in 2023 with a range of 250 miles per single charge
- Pepsico and Frito Lay are already using Tesla Semis and plans to add upto 100 semi vehciles by the end of 2023.

Comparison of EV Providers

Company name	Features
Chinese and Taiwan New EV Developments	Strong support in China and Taiwan excels in battery manufacturing
Battery Swap	Implemented by NIO, Ample, Reduces charging time and range anxiety
Zeekr's Battery Tech	CATL batteries, 100 kWh, 10%-80% in 15 minutes
Rivian Charging Network	3,500 stations by 2025, DC fast chargers, Level 2, 10%-80% in about 20 minutes
Tesla Commercial Network	Tesla 4680 battery cells, 500+ miles per charge, 500+ miles per charge


THANK YOU