Introduction

Motivation

The growing adoption of electric vehicles has increased pressure on charging infrastructure, yet station utilization remains uneven—some sites face congestion while others are underused, affecting revenue and customer experience. This project applies a data-driven approach to forecast demand, identify inefficiencies, and recommend operational improvements.

Objective

To develop an accurate forecasting model to predict charging demand for individual EV charging stations or station clusters.

Dataset

The primary sources of data for our project were as below:

1. EV charging session data from 2021 to 2025 from NYC Municipal garages, providing detailed records of charging activity across various municipal locations.
2. The geographical context, garage location details—including address, latitude, and longitude—were collected through Google search.
3. Weather information such as temperature, humidity, and conditions was retrieved using the Visual Crossing Weather API, allowing the integration of environmental factors into the analysis

Together, these sources form a comprehensive dataset for understanding and modelling EV charging demand patterns.

Data Preprocessing

1. Cleaned and standardized text fields
2. Reindexed data to create a complete hourly time grid
3. Created a feature for time since last charging session
4. Added calendar-based features (day of week, hour of day, month, weekend, etc.)
5. Engineered weather-derived features (is_hot, is_snowing, humidity bands, etc.)
6. Aggregated data from session-level to hourly-level
7. Generated time-series features such as lags and rolling windows

Exploratory Data Analysis (EDA)

Univariate analysis was performed for all numerical and categorical variables, and correlations were computed for numerical features. The plots and exploratory analysis revealed several key insights:

1. Duration-related and energy-related variables show strong right skew, indicating many short or average charging sessions and only a few very long ones.
2. While most locations receive steady footfall, a few stations remain significantly underutilized.
3. Users generally prefer short charging sessions rather than fully charging their vehicles.
4. Afternoon hours, followed by mornings, emerged as the most preferred periods for charging activity.

Modelling Approach

We followed couple of approaches for getting the desired output.

Approach 1: Time-Series Forecasting using CATBoost

1. Built hourly forecasting models for each charging location.
2. Model 1 predicts **number of charging sessions per hour**.
3. Model 2 predicts **hourly energy consumption**.
4. Incorporated historical usage, time-based features, and weather-driven variables to capture seasonality and external effects.

Approach 2: Time-Series Forecasting using XGBoost

1. Ensemble technique that builds decision trees sequentially
2. Model splits data based on conditions (e.g. "is it a weekday?", is temp > 90 degree"")
3. Each iteration predicts the residuals of previous model, incrementally reducing bias.
4. Model predicts the expected load, energy in kwh which is directly used for predicting price using a deterministic calculation.

Evaluation Strategy

Approach 1: CATBoost

1. Evaluate model performance on future (unseen) time period
2. We used time-based validation, not random splitting

3. Measure accuracy separately for 2 target variables.
    a. Metrices used: MAE and RMSE

Approach 2: XGBoost

1. Use the last 20% of data for testing to respect temporal order
2. Metrics used: MAE and RMSE

Key Observations

Approach 1: CATBoost

1. EV charging demand is highly seasonal and predictable
    a. There is a strong daily cycle and a consistent weekly pattern.
    b. Autocorrelation metrics(ACF/PACF) proves this
2. Session forecast from this model are highly accurate.
    a. MAE ≈ 0.30 sessions per hour
    b. RMSE ≈ 0.69 sessions per hour
    c. This can be used for staffing, fleet planning, and operational scheduling.
3. A subset of charging stations has zero or near-zero activity.
    a. These can be considered for relocation

To summarise:
EV charging demand across locations follows a stable, highly predictable pattern driven mainly by daily/weekly seasonality. The forecasting model captures this structure with high accuracy, enabling reliable 1-month operational planning, capacity management, and station-level optimization.

Future Scope

1. Improve time series forecasting by considering more attributes like holidays, downtime etc.
2. Develop a real time dashboard for hourly predictions.
3. Integrate dynamic pricing analysis for revenue optimization

Individual Contribution

Rahul Kumar

1. Consolidated data from multiple raw sources into a unified, clean, and standardized dataset.

2. Ensured consistency across locations, weather inputs, and session-level attributes
3. Worked collaboratively on EDA for gathering key insights
4. Engineered temporal features (lags, rolling averages, seasonality indicators)
5. Built and validated forecasting models for hourly sessions and energy consumption
6. Performed time-based train/validation splits and model evaluation (MAE, RMSE)

Abhishek Malik

1. Worked collaboratively on EDA, data cleanup and normalization.
2. Feature engineering to generate predictive features from temporal, spatial, and user history data.
3. Trained XGBoost model using Time Series Split and evaluated performance.
4. Model evaluation (MAE, RMSE)
5. Formulated deterministic Dynamic pricing logic using the model's load prediction.