EV Charging Trends and Predictions

INFO 511 Fall 2024
Final Project

Team: ChilePeppers
Derek Rice
Shreya Kolte
Pouya Jalali Khalilabadi

College of Information Science

Introduction & motivation

Focus Area:

- Electric Vehicle (EV) Charging Efficiency.
- Interest in exploring charging patterns and user behaviors.
- Building predictive models for charging efficiency.

Research Questions:

- How do vehicle model, user type, and starting state of charge influence the cost and duration of EV charging sessions at public stations?
- Exploring energy consumption and charging behaviors
- Building predictive models for charging efficiency

Hypotheses:

- **Vehicle Models**: Larger batteries → Longer charging times, higher costs.
- **User Profiles**: Regular users → Lower costs per session.
- **State of Charge**: Lower starting levels → More expensive sessions.

Electric Vehicle Dataset Overview

Source: Electric Vehicle Charging Patterns

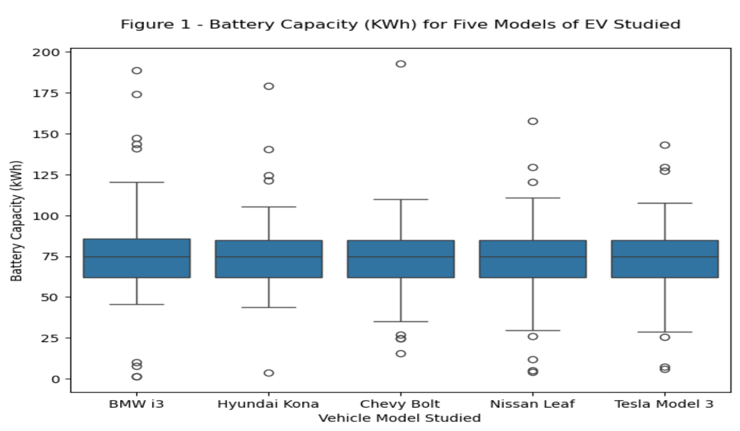
The EV Charging Dataset comprised:

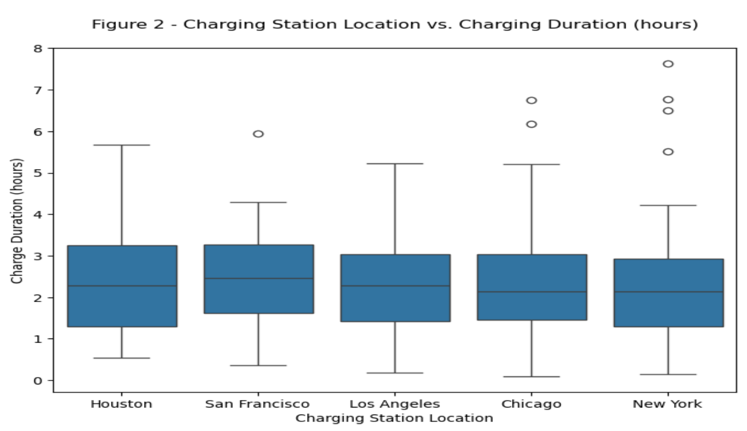
- A total of 20 variables (columns) and 1320 data points (rows)
- Each column was a charging factor studied
- Each row was a charging event

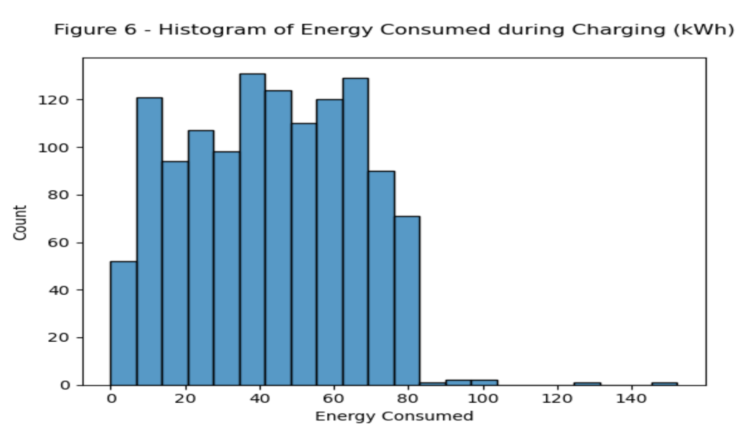
```
✓ ev_charging.info() ···

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1320 entries, 0 to 1319
Data columns (total 20 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   User ID                               1320 non-null   object
 1   Vehicle Model                         1320 non-null   object
 2   Battery Capacity (kWh)                1320 non-null   float64
 3   Charging Station ID                   1320 non-null   object
 4   Charging Station Location             1320 non-null   object
 5   Charging Start Time                   1320 non-null   object
 6   Charging End Time                     1320 non-null   object
 7   Energy Consumed (kWh)                 1254 non-null   float64
 8   Charging Duration (hours)             1320 non-null   float64
 9   Charging Rate (kW)                    1254 non-null   float64
 10  Charging Cost (USD)                   1320 non-null   float64
 11  Time of Day                           1320 non-null   object
 12  Day of Week                           1320 non-null   object
 13  State of Charge (Start %)             1320 non-null   float64
 14  State of Charge (End %)               1320 non-null   float64
 15  Distance Driven (since last charge) (km)  1254 non-null   float64
 16  Temperature (°C)                      1320 non-null   float64
 17  Vehicle Age (years)                   1320 non-null   float64
 18  Charger Type                          1320 non-null   object
 19  User Type                             1320 non-null   object
dtypes: float64(10), object(10)
memory usage: 206.4+ KB
```

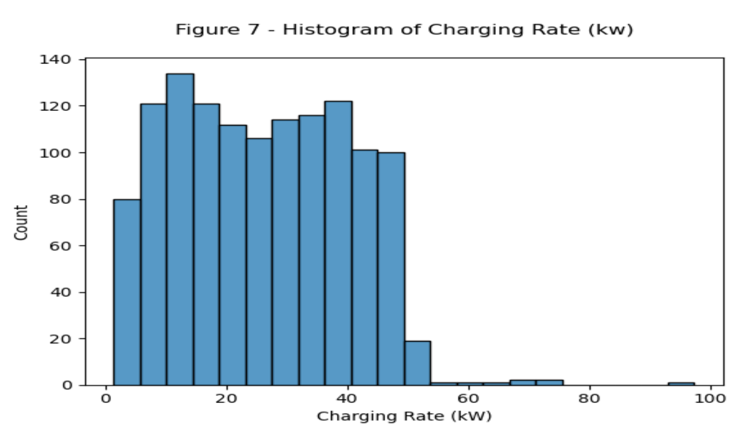
EV charging dataset info() summary. The dataset included two data types, object and float64. One-hot-encoding was employed to include object data types in the predictive model building effort.

Exploratory Data Analysis









Modeling

```
Table 5 - Second Iteration, Statistically Significant Factors, Multivariable Regression Analysis
                            OLS Regression Results
==============================================================================
Dep. Variable:     Charging Cost (USD)   R-squared:                       0.041
Model:                             OLS   Adj. R-squared:                  0.021
Method:                  Least Squares   F-statistic:                     2.035
Date:                 Sun, 15 Dec 2024   Prob (F-statistic):            0.00277
Time:                         14:12:32   Log-Likelihood:                -4271.4
No. Observations:                 1131   AIC:                             8591.
Df Residuals:                     1107   BIC:                             8712.
Df Model:                           23
Covariance Type:             nonrobust
==================================================================================================================
                                           coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                   27.1726      2.102     12.930      0.000      23.049      31.296
State of Charge (End %)                 -0.0405      0.019     -2.123      0.034      -0.078      -0.003
Temperature (°C)                         0.0467      0.022      2.145      0.032       0.004       0.089
Vehicle Model_Chevy Bolt                 0.7493      1.027      0.730      0.466      -1.265       2.764
Vehicle Model_Hyundai Kona               2.6782      1.027      2.607      0.009       0.662       4.694
Vehicle Model_Nissan Leaf                2.6001      1.013      2.568      0.010       0.613       4.587
Vehicle Model_Tesla Model 3              1.0828      0.998      1.085      0.278      -0.875       3.041
Charging Station Location_Houston       -1.8020      1.031     -1.749      0.081      -3.824       0.220
Charging Station Location_Los Angeles   -0.6514      1.012     -0.644      0.520      -2.637       1.334
Charging Station Location_New York      -1.2522      1.040     -1.205      0.229      -3.292       0.788
...
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

Multivariable Regression Analysis (R2 adjusted = 2.1%)

```
Fitting 5 folds for each of 25 candidates, totalling 125 fits
===== RBF Kernel: Model Summary =====
Best Hyperparameters: {'C': 0.1, 'gamma': 'scale', 'kernel': 'rbf'}

Performance Metrics:
Mean Squared Error (MSE): 126.26
R² Score: -0.03
Adjusted R² Score: -0.19

Best Estimator:
SVR(C=0.1)
```

Support Vector Machine (SVM) kernel = RBF with R^2 adjusted = -.19

```
Fitting 5 folds for each of 32 candidates, totalling 160 fits
===== Kernel = Poly: Model Summary =====
Best Hyperparameters:
{'C': 0.1, 'degree': 2, 'gamma': 'scale', 'kernel': 'poly'}

Performance Metrics:
Mean Squared Error (MSE): 125.40
R² Score: -0.02
Adjusted R² Score: -0.18

Best Estimator:
SVR(C=0.1, degree=2, kernel='poly')
```

Support Vector Machine (SVM) kernel = Polynomial with R^2 adjusted = -.18

Results & Discussion

Key Findings:

- **Exploratory Data Analysis:** *Battery capacity was consistent across models; energy consumption followed a normal distribution. No clear visual trends for charging cost vs. time of day or location.*

- **Statistical Analysis (ANOVA):** *Vehicle model and charger type significantly influenced charging costs ($p < 0.05$). Charging location had no significant impact. Higher temperatures slightly increased costs (linear regression $p < 0.05$).*

- **Predictive Modeling:** *Regression and SVM models performed poorly, explaining only ~2% of the variance (R^2-adjusted ~0.021).*

Conclusions & Future Works

Conclusions

- The analysis identified statistically significant relationships between vehicle model, charger type, and charging cost, as well as a linear relationship between charging temperature and cost. However, multivariable regression and SVM models struggled to explain variance, with adjusted R-squared values below 5%.
- Graphical and statistical reviews provided insights into the dataset's structure but highlighted limitations in predicting charging costs due to potential multicollinearity, noise, or missing key factors.

Future Works

- Explore more advanced machine learning models, such as random forests or neural networks, to better capture nonlinear relationships and interactions in the dataset.
- Incorporate additional data points, such as user demographics or environmental factors, to improve the model's explanatory power.
- Investigate potential data cleaning and preprocessing enhancements, including feature engineering and addressing outliers, to reduce noise and improve prediction accuracy.

Thank you!

Questions?