FINAL PROJECT

Azra Pehlivan
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
azra.pehlivan@metu.edu.tr

Bedirhan Salhan
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
bedirhan.salhan@metu.edu.tr

Deniz Kuzu
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
deniz.kuzu@metu.edu.tr

Melis Gündüz
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
gunduz.melis@metu.edu.tr

Yiğit Kaan Ekinci
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
ekinci.kaan@metu.edu.tr

Elif Sude Kürkçü
Statistics
Introduction to Data Processing and Visualization
elifsudekurkcu@metu.edu.tr

Pınar İlgin Güvenir
Statistics
Introduction to Data Processing and Visualization
Ankara, Türkiye
ilgin.guvenir@metu.edu.tr

ABSTRACT

1. INTRODUCTION

The transition to electric vehicles (EVs) has gained significant momentum in recent years so, understanding EV charging behavior is essential since it has a direct impact on the effectiveness and usability of charging substructure. This project focuses on using Python to visualize and analyze patterns in EV charging sessions, offering important insights into user behavior and the variables influencing charging dynamics.

The dataset used in this project contains 900 observations, each representing a unique EV charging session. It includes detailed information such as the type of vehicle being charged, battery capacity, charging station location, energy consumption, charging duration, cost, and user characteristics. Additionally, it captures environmental and temporal factors like temperature, time of day, and day of the week. These variables give us a complete picture of the circumstances surrounding EV charging. Python's data analysis and visualization libraries, including Matplotlib, Seaborn, and NumPy, are used for transforming this dataset into meaningful and interactive visual representations. These tools make it possible to investigate important topics including average energy consumption, peak charging times, and variations in charging behavior among user types or car models. The visualizations aim to identify trends and anomalies that may not be immediately apparent in raw data.

This report documents the entire process, starting from data cleaning and preprocessing to visualization and analysis. Each step is designed to ensure the accuracy and clarity of the results. With detailed visualizations,, it provides a deeper understanding of charging behavior and the underlying factors that shape it. These findings are expected to contribute to the development of more efficient, user-friendly, and sustainable EV charging solutions.

CATEGORIZING THE VARIABLES

VARIABLES	DESCRIPTION	SCALES OF MEASUREMENT
User ID	A unique identifier for each user.	Nominal
Vehicle Model	The model of the EV being charged (e.g., Tesla Model 3, Nissan Leaf).	Nominal
Battery Capacity(kWh)	The total battery capacity of the vehicle in kilowatt-hours.	Ratio
Charging Station ID	A unique identifier for the charging station used.	Nominal
Charging Station Location	The geographic location of the charging station (e.g., New York, Los Angeles)	Nominal
Charging Start Time	Timestamp indicating when the charging session began.	Interval
Charging End Time	Timestamp indicating when the charging session ended.	Interval
Energy Consumed (kWh)	Total energy consumed during the charging session, measured in kilowatt-hours.	Ratio
Charging Duration (hours)	Total time taken to charge the vehicle, measured in hours.	Ratio
Charging Rate (kW)	Average power delivery rate during the charging session, measured in kilowatts.	Ratio
Charging Cost (USD)	Total cost incurred for the charging session, measured in US dollars.	Ratio
Time of Day	The time segment when the charging occurred (e.g., Morning, Afternoon).	Nominal
Day of Week	The day of the week when the charging occurred (e.g., Monday, Tuesday).	Nominal
State of Charge (Start %)	Battery charge percentage at the start of the charging session.	Interval
State of Charge (End %)	Battery charge percentage at the end of the charging session.	Interval
Distance Driven (since last charge) (km)	Distance traveled since the last charging session, measured in kilometers.	Interval
Temperature (°C):	Ambient temperature during the charging session, measured in degrees Celsius.	Interval
Vehicle Age (years)	The age of the electric vehicle, measured in years.	Ratio
Charger Type	The type of charger used (e.g., Level 1, Level 2, DC Fast Charger).	Nominal
User Type	Classification of user based on driving habits (e.g., Commuter, Long-Distance Traveler).	Nominal

2. DATA TIDYING AND CLEANING STEPS

Column Renaming
Columns were renamed for better readability.

Removing Unnecessary Columns
The UserId column was removed as it was deemed unnecessary.

Data Preprocessing
-Whitespace Removal: Unnecessary whitespaces in the
columns VehicleModel, ChargingStationLocation, TimeOfDay, DayOfWeek, ChargerType, and UserType were removed using the strip() method.
- Handling Missing Values: Missing (NaN) values in these columns were replaced with 'Unknown' using the fillna('Unknown') method.

Unique Value Analysis
All unique values were listed using the unique() method and reviewed to detect erroneous or inconsistent values.

Data Standardization
-Mapping Dictionary: A model_mapping dictionary was created to map various misspellings or case differences into a correct and consistent format.
- Replacement: The replace() method was applied to standardize incorrect or inconsistent values based on the mapping dictionary.
- Verification: Corrections were verified by rechecking unique values using the unique() method.

Replacing Placeholder Values
Placeholder values ('Unknown') were replaced with NaN to indicate truly missing data.

Handling Missing Data
- Missing Value Analysis: Using the isna().sum() method, it was determined that:
    - 150 missing values were present in categorical variables.
    - No missing values were found in numerical variables.
    - Impact Analysis: The missing values were present across all relevant columns, confirming 150 rows with missing data (14% of the dataset).
    - Action: Rows with missing values were deleted to obtain a clean dataset, as this approach prevented distorting the analysis.

Splitting Date and Time Columns
- The ChargingStartTime and ChargingEndTime columns were split into separate date and time columns.
- Data Type Conversion:
    - Date columns were converted to datetime.date.
    - Time columns were converted to datetime.time.

Cleaning and Converting Numerical Columns
- Symbols were removed from the StateofChargeStart%, StateofChargeEnd%, and ChargingCostUSD columns.
- Data types were converted to float.

New Column Creation
- ChargeDifference%: Created to measure the difference in charge percentage.
- ChargingDurationMinutes: Created to calculate charging duration in minutes.

Outlier Detection
- Outliers were checked using the IQR (Interquartile Range) rule.

3. EXPLORATORY DATA ANAYLSIS

After the data cleaning phase, our aim is to analyze the dataset about electric vehicle (EV) charging patterns. The dataset consists of 20 columns, with 10 being quantitative and the remaining qualitative. To gain a general perspective to understand the data, we constructed a table detailing descriptive statistics for the numerical columns. In order to gain important insights regarding EV charging behavior and user characteristics, this made it easier to examine important metrics including mean, standard deviation, quartiles, minimum, and maximum values for each quantitative variable.

For quantitative data:

	Energy Consumed	Charging Duration	Charging rate	Charging Cost	Vehicle age	Battery Capacity
count	900	900	900	900	900	900
mean	43.16	4.19	25.9	27.15	4.4	63.66
std	21.22	2.13	13.79	13.92	2.85	20.29
min	5.11	0.5	2.02	2.06	0	30
25%(Q1)	25.28	2.36	14.15	15.27	2	46
50%(Q2)	43.48	4.23	26.48	27.33	4	64
75%(Q3)	60.76	5.93	37.68	39.25	7	81
max	79.86	8	49.98	50	9	99

For qualitative data:

Vehicle Model	**Frequency**
Audi e-Tron	188
Chevy Bolt	162
Hyundai Kona	176
Nissan Leaf	179
Tesla Model 3	195

Charging Station Location	**Frequency**
Chicago	174
Los Angeles	199
New York	185
San Francisco	178
Seattle	164

Time of day	**Frequency**
Morning	237
Afternoon	225
Evening	214
Night	224

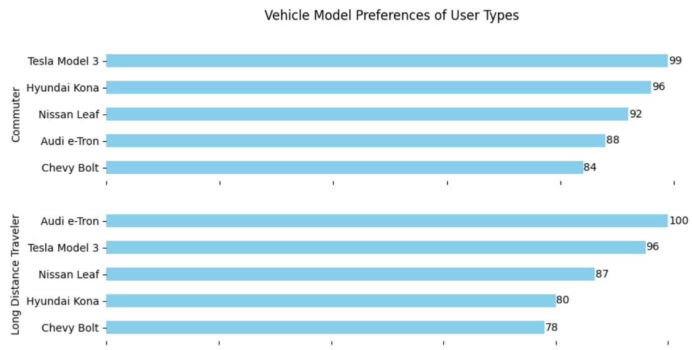
Day of week	**Frequency**
Monday	134
Tuesday	148
Wednesday	113
Thursday	129
Friday	122
Saturday	107
Sunday	147

Charger type	**Frequency**
DC Fast Charger	300
Level 1	243
Level 2	242
Level	115

User Type	**Frequency**
Commuter	459
Long Distance Traveler	441

RESEARCH QUESTIONS

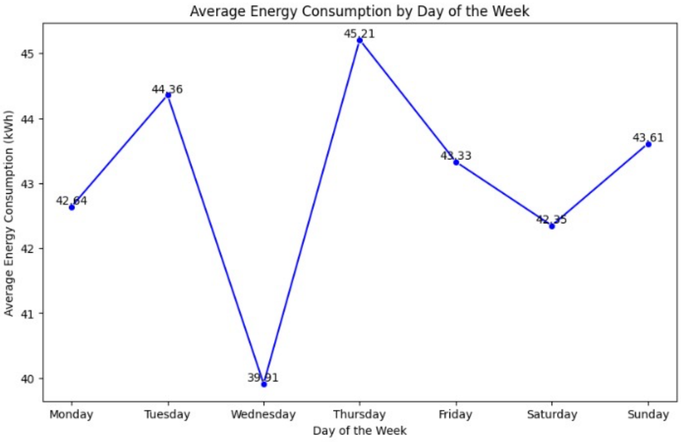
Q1) WHAT ARE THE DIFFERENCES IN VEHICLE MODEL PREFERENCES BETWEEN COMMUTERS AND LONG-DISTANCE TRAVELERS? (MELIS GÜNDÜZ)



The graph shows the preferences for different vehicle models among two types of users: commuters and long-distance travelers. There are five vehicle models in the comparison: Tesla Model 3, Hyundai Kona, Nissan Leaf, Audi e-Tron, and Chevy Bolt. Each bar represents the popularity number for the specific vehicle model among the two user types.
For commuters, the most popular model is the Tesla Model 3, with a very highest number of 99. Following that, the Hyundai Kona is also highly preferred at 96, and then the Nissan Leaf at 92. The Audi e-Tron and Chevy Bolt are less popular, with numbers of 88 and 84, respectively.

For long-distance travelers, the Audi e-Tron is the clear favorite with a perfect number with 100. Tesla Model 3 is the second most preferred vehicle with a number of 96, followed by the Nissan Leaf at 87. The Hyundai Kona and Chevy Bolt have lower popularity, with numbers of 80 and 78, respectively.

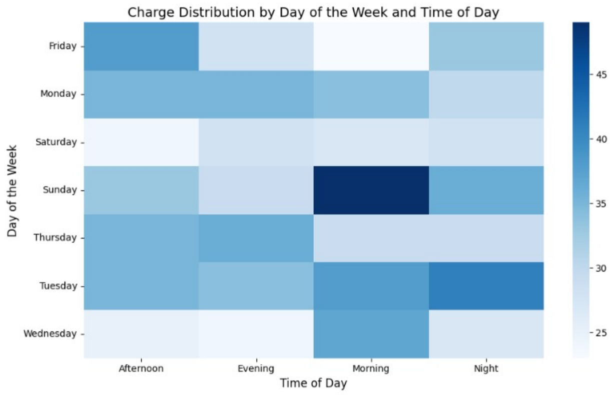
Q2) WHAT IS THE AVERAGE ENERGY CONSUMPTION BY DAYS OF WEEK? (AZRA PEHLİVAN)



The graph shows us the average energy consumption for each day of the week. It starts with Monday, the energy consumption is relatively stable, averaging around 42 kWh. The energy use on this day is consistent, without any significant drop. It reflects a normal pattern for the beginning of the week. On Tuesday, there is a peak in energy consumption, reaching approximetaly 44.36 kWh. It can suggest increased activities that may require more energy as the week progresses. On Wednesday, it drops a lot, going down to 39.91 kWh, which is the lowest point. This dip may reflect a temporary slowdown in activities. On the other hand, Thursday jumps back up again over 45 kWh., the highest level recorded during the week. After that, the consumption declines around 43.33 kWh. Saturday continues this trend, with an average falling to 42.35 kWh, one of the lowest levels observed. This reduction can be associated with reduced activities as people may engage in weekend routines. On Sunday, consumption is climbing back to 43.61 kWh. This increase may be because of the preperations for the upcoming week.

Overall, the observed trend emphasizes the importance of understanding daily energy usage patterns. It may help optimizing energy distribution and planning. By looking at the days of peak and low consumption, energy providers and consumers can improve efficiency and plan sustainability.

Q3) AT WHAT TIMES OF THE DAY DO USERS CHARGE THEIR DEVICES AND WHAT DAYS DOES THIS COINCIDE WITH? (PINAR İLGİN GÜVENİR)

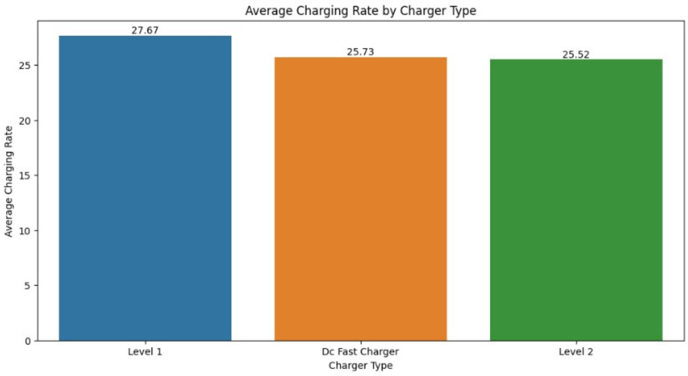


A heat map was made to analyse the electric vehicle charging habits of users by day and time of day. It is very clear that the highest charging activity takes place on Sunday morning. This can be attributed to users making weekend plans or perhaps preparing for long distance journeys. Tuesday night ranks second in the density ranking. This may indicate that users prefer charging stations during the night hours, or that they may be taking advantage of possible night tariff benefits. At the same time, this may require more efficient planning of services during night hours.

The third peak is on Tuesday and Wednesday mornings. This may be an indication of a mid-week need being met. The fact that users prefer morning hours shows that their charging needs are intensified in the morning in their daily routines. Friday afternoon can be interpreted as an important time period for preparing for the weekend. The increase in charging intensity during this period suggests that users prepare their vehicles for weekend travelling.

In general, charging activities are concentrated at certain times of the day and on certain days of the week. Sunday morning, Tuesday night and morning hours etc. This analysis provides a valuable guide for understanding user behaviour and developing charging infrastructure.

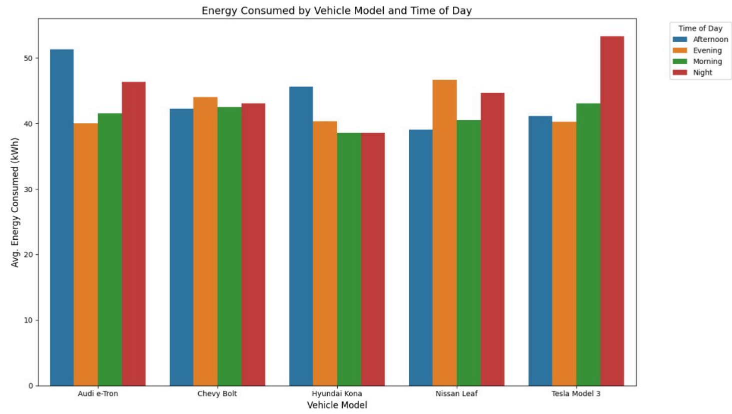
Q4) DOES THE AVERAGE OF CHARGING RATE CHANGE DEPENDING ON THE CHARGER TYPE?



(DENİZ KUZU)

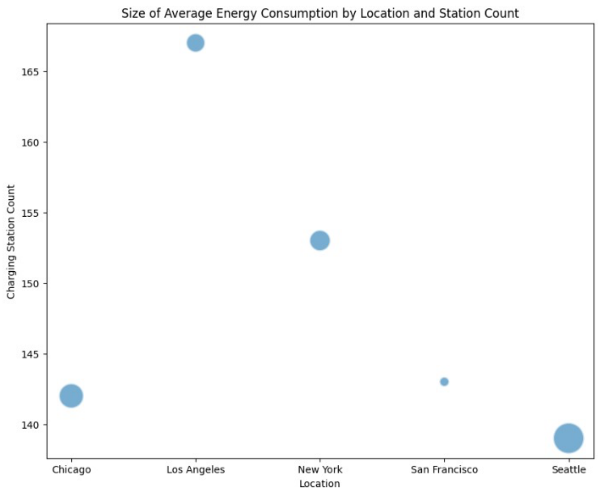
The graph presents the average charging rates for three different types of chargers: Level 1, DC Fast Charger, and Level 2. At first glance, it might seem like there are differences between the charger types, but when we examine the values closely, the differences are minimal. Level 1 chargers appear to have the highest average rate at 27.67 kW, while DC Fast Chargers, which are generally expected to be faster, come in slightly lower at 25.73 kW. Level 2 chargers are very close as well, with an average rate of 25.52 kW. This might lead us to question whether the type of charger really affects the charging rate in a significant way. In the report, we could say: "There seems to be no clear correlation between charger type and average charging rate. While there are small differences in the averages, they are not substantial enough to suggest a strong relationship. Charger type might initially appear to have an impact, but the data does not strongly support this conclusion."
This finding is somewhat unexpected because DC Fast Chargers are typically associated with higher charging speeds. However, this graph suggests that other factors, such as the battery type, charging conditions, or usage patterns, might be influencing the results. This shows that real-world data can sometimes challenge our assumptions.

Q5) HOW DOES ENERGY CONSUMPTION VARY DEPENDING ON CHARGING RATE, VEHICLE MODEL AND TIME OF DAY? (BEDİRHAN SALHAN)



The bar chart represents the energy consumption of different vehicle models at various times of the day. The chart compares the average energy consumption (kWh) of each vehicle model (e.g., Audi e-Tron, Chevy Bolt, Hyundai Kona, Nissan Leaf, and Tesla Model 3) during the morning, afternoon, evening, and night time periods. Tesla Model 3 users appears to have the highest energy consumption during the night. Chevy Bolt has a more balanced distribution of energy consumption compared to other models. There is a general increase in energy consumption during nighttime across the models. Afternoon and morning energy consumption values are generally close to each other. Some models have lower consumption during the evening. Differences in consumption between vehicle models may be related to their design and technical specifications. This chart provides insights into energy consumption behaviors and can help develop more efficient energy management strategies.

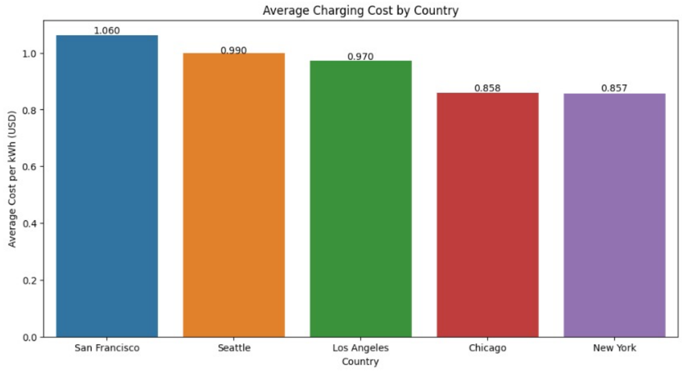
Q6) HOW MANY CHARGING STATIONS IN EACH LOCATION BASED ON AVERAGE ENERGY CONSUMPTION? (ELİF SUDE KÜRKÇÜ)



This bubble chart provides a detailed analysis of the relationship between the number of charging stations and average energy consumption in five cities: Chicago, Los Angeles, New York, San Francisco and Seattle. The x-axis represents the cities, while the y-axis shows the number of charging stations at each location. The size of the bubbles reflects the average energy consumption in each city, with larger bubbles indicating higher consumption. Among the cities, Seattle stands out with the largest bubble size, highlighting its significant energy consumption compared to other locations.

On the other hand, San Francisco has the smallest bubble, indicating minimal energy use and limited infrastructure. New York shows a moderate number of charging stations with relatively higher energy consumption, as represented by the larger bubble size. San Francisco has fewer charging stations than New York, but exhibits a smaller bubble size, indicating lower average energy use. Los Angeles, with the most charging stations, represents an average bubble size, indicating limited energy demand and infrastructure development. This visualization effectively highlights the differences in energy consumption and charging station distribution across cities, providing insights into infrastructure development and energy demand patterns.

Q7) HOW DOES THE AVERAGE ELECTRICITY PRICE PER 1 KWH IN CHARGING STATIONS VARY BY STATES? (YIĞİT KAAN EKİNCİ)



When we look at the average electricity price per 1kWh for each state, we can see that the average electricity price in San Francisco is the highest with \$1.06. San Francisco is followed by Seattle with \$0.99, Los Angeles with \$0.97, Chicago with \$0.858 and New York with \$0.857. We can see that New York has the cheapest average electricity price. The differences between these average electricity prices may be influenced by the economic situation in each states, the strength of the infrastructure support for charging station installation, the number of electric vehicle users, so the energy demand, and some other factors.

4. CONCLUSION

The analysis of EV charging behaviour revealed important insights through the research questions addressed in the project. Differences in vehicle model preferences were evident with the Tesla Model 3 favoured by commuters and the Audi e-Tron favoured by long distance travellers. Energy consumption trends varied throughout the week, with a significant drop in the middle of the week and a peak on Thursday, reflecting user routines and activities. Temporal analysis also highlighted specific periods such as Sunday mornings and Tuesday nights as times of high demand, revealing opportunities to optimise station utilisation. Geographical differences in infrastructure were significant, with cities such as Seattle showing high energy demand despite fewer stations, while Los Angeles had the most stations but moderate utilisation. Analyses of charger types found minimal differences in average charging rates, challenging assumptions about their performance. Finally, electricity prices varied across locations, with San Francisco having the highest average price and New York the lowest, emphasising the need for region-specific energy strategies. Overall, these findings provide actionable insights to improve EV infrastructure, energy efficiency and user satisfaction.

REFERENCES

https://www.datacamp.com/tutorial/guide-to-data-cleaning-in-python

https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/

GitHub Link:
https://github.com/ykaanekinci/2025_112finalproject