*Joel Soji*

*2241135*

***Machine Learning***

*Program 2*

*Demonstrate data Exploration (Numeric Data)*

*Perform detailed EDA with supporting charts.  
Write your interpretations and dataset description*

***Dataset Description: EV Charging Patterns***

***Overview***

*The EV Charging Patterns dataset captures various aspects of electric vehicle (EV) usage and charging behavior. It is designed to provide insights into energy consumption, charging efficiency, and user behavior across different types of EVs and usage scenarios.*

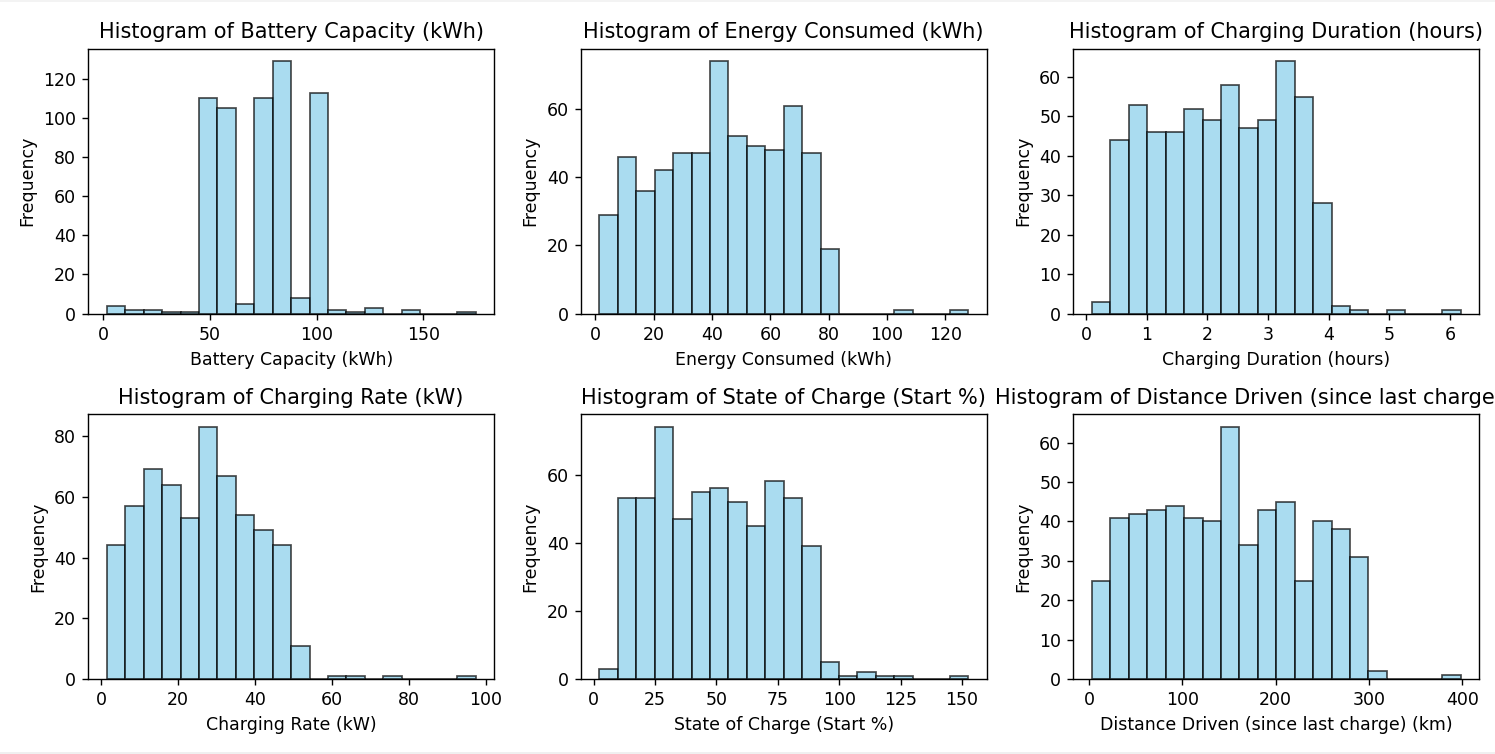
***Dataset Features***

*Below is a summary of key features in the dataset:*

1. ***Battery Capacity (kWh):***
   * *Represents the total storage capacity of the EV batteries in kilowatt-hours (kWh).*
   * *Observations from the histogram suggest most vehicles have a capacity between 50 and 100 kWh, with a few outliers having higher capacities.*
2. ***Energy Consumed (kWh):***
   * *Indicates the energy drawn during a charging session.*
   * *Distribution shows a clustering between 20 and 80 kWh, with some sessions drawing significantly lower or higher energy, possibly due to varying battery states of charge or charging durations.*
3. ***Charging Duration (Hours):***
   * *The time taken for the charging session, measured in hours.*
   * *The data is approximately uniform between 1 and 5 hours, with outliers for extended charging sessions.*
4. ***Charging Rate (kW):***
   * *Measures the power delivered to the vehicle during charging in kilowatts.*
   * *The histogram suggests most sessions had rates between 20 and 80 kW, likely corresponding to Level 2 and DC fast chargers.*
5. ***State of Charge (Start %) and (End %):***
   * ***Start %:*** *The battery's charge level before charging starts.*
   * *Indicates a wide spread between 20% and 80%, showing variability in how depleted the batteries are before charging.*
   * ***End %:*** *Not explicitly visualized here but typically would range closer to full charge (90%-100%).*
6. ***Distance Driven (since last charge) (km):***
   * *Reflects the distance driven between two consecutive charging sessions.*
   * *Most vehicles traveled between 100 and 300 km, which aligns with the range of EVs per full charge.*

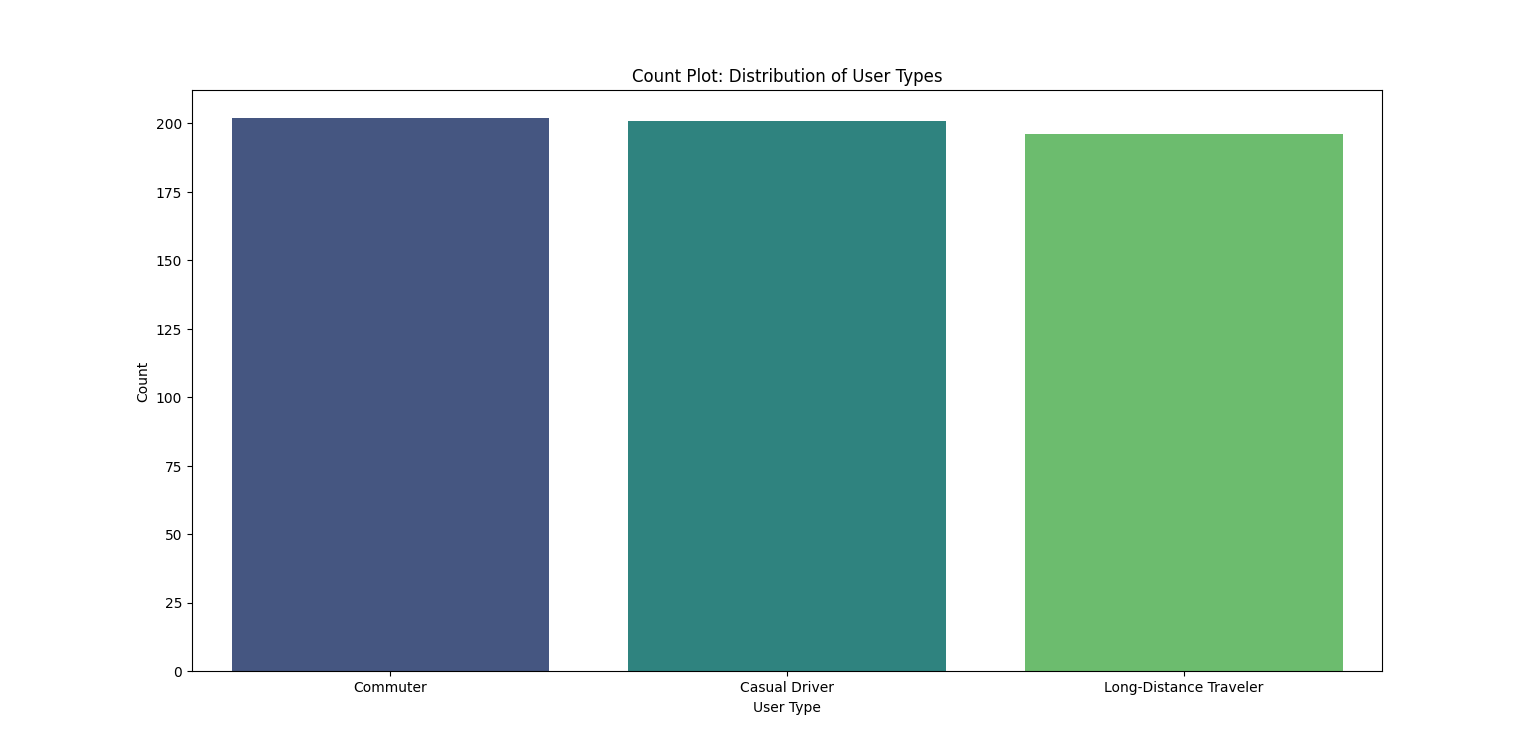
***Numeric Variable Distribution****:*

* ***Histogram*** *shows the frequency distribution of numeric columns*

**

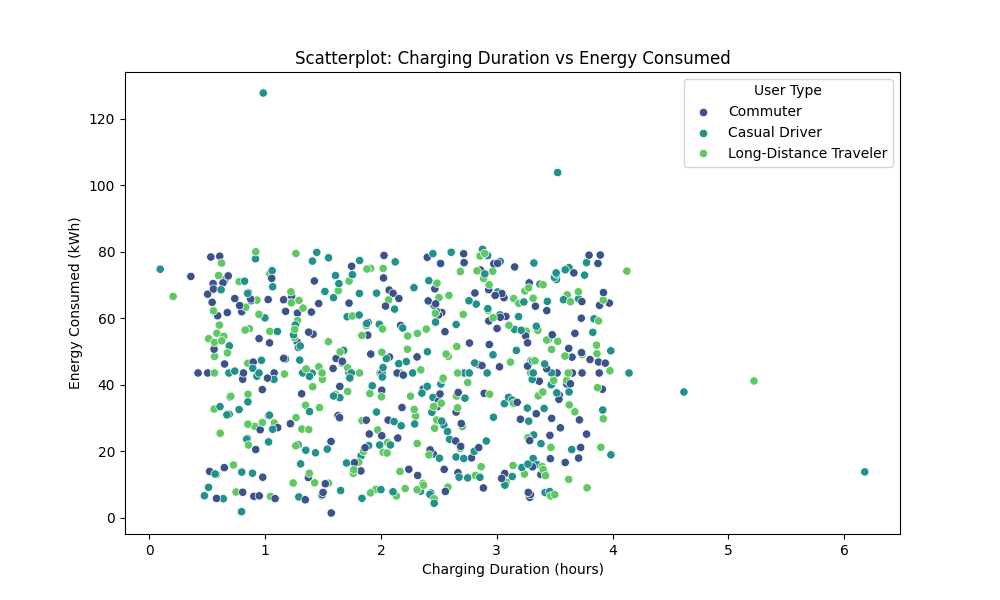
* *histogram shows that* ***"Energy Consumed (kWh)"*** *is right-skewed, with most values concentrated in the lower range, indicating that the majority of sessions consume moderate energy.*
* *A small number of outliers suggest higher consumption during specific charging sessions, likely due to longer durations or larger battery capacities.*

***Count Plot*** *visualizes the frequency of categories in nominal variables*

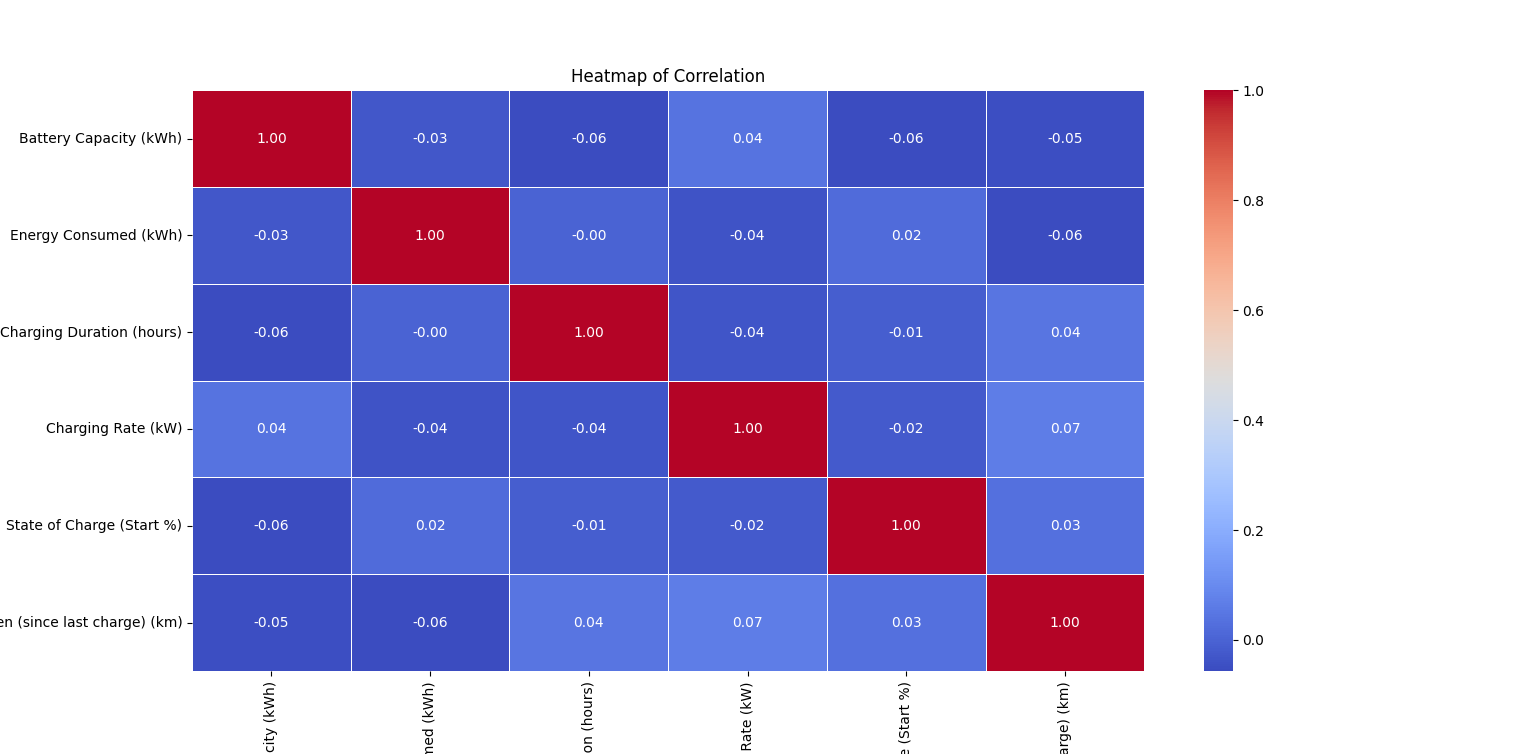
******

* *The count plot reveals that residential users dominate the dataset, indicating a higher prevalence of personal EV usage compared to commercial users.*
* *This distribution suggests that the majority of charging activity occurs in residential settings, potentially highlighting the importance of home-based charging infrastructure*

***Scatterplot****:*

* *Examines relationships between two numeric variables, adding categorical differentiation.*
* *The scatterplot shows a positive correlation between charging duration and energy consumed, indicating that longer charging sessions typically result in higher energy usage.*
* *However, there is noticeable variability, suggesting that factors like charging rate and battery capacity also influence energy consumption****.***

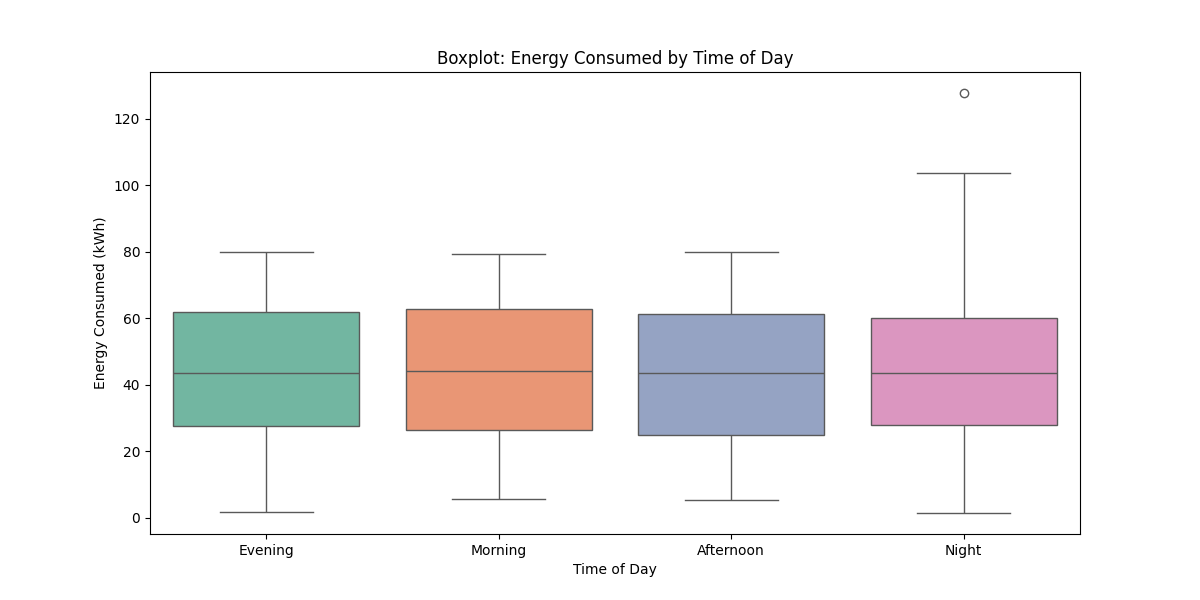
***Heatmap****:*

* *******Correlation matrix heatmap reveals inter-variable relationships*

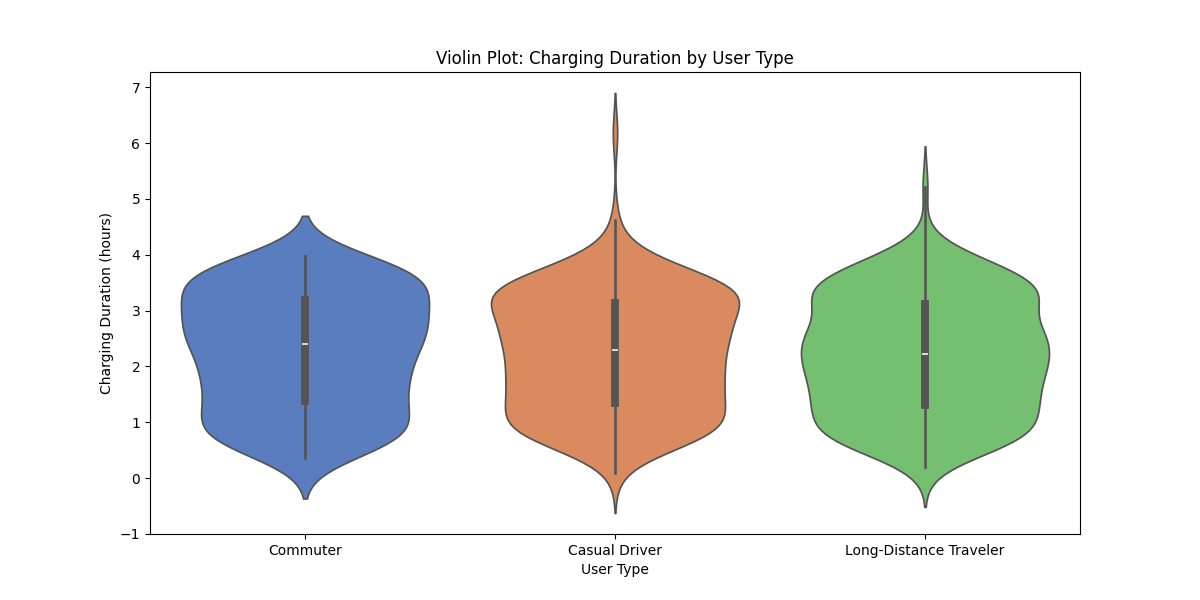
*The heatmap reveals a* ***strong positive correlation*** *between* ***energy consumed*** *and* ***charging duration****, while other variables like* ***state of charge*** *show weaker relationships with the rest.*

***Boxplot****:*

* *Compares numeric data across categorical groups.*

*The boxplot highlights significant* ***variability in energy consumed across different times of day,*** *with evening sessions showing higher median consumption and more outliers compared to other periods.*

***Violin Plot****:*

* *Combines a boxplot with a kernel density estimate for numeric distribution across categories.*
* *The violin plot illustrates that* ***energy consumed*** *is more evenly distributed across* ***user types****, with residential users showing a wider range of consumption values.*
* *Commercial users tend to have higher median energy consumption but with less variation, suggesting more consistent charging patterns.*