Data Collection and Preprocessing Phase

Date	18 March 2025
Team ID	LTVIP2025TMID26729
Project Title	Visualization Tool for Electric Vehicle Charge And Range Analysis in Tableau
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
Data Overview	A comprehensive visualization tool for electric vehicle (EV) charge and range analysis depends on a multifaceted dataset. This data broadly falls into four categories: static vehicle specifications, dynamic charging infrastructure data, environmental and geographical information, and optional user usage data. Vehicle specifications, including battery capacity, rated range, and charging capabilities, provide the baseline for performance comparisons. Charging infrastructure data, such as station locations, charger types, and real-time availability, is crucial for route planning and charging optimization. Environmental and geographical data, encompassing temperature, terrain, and traffic, reveal the impact of external factors on EV performance.
Data Cleaning	Geocoding addresses to obtain precise latitude and longitude coordinates for charging stations is crucial for accurate mapping. Furthermore, data validation checks should be implemented to identify and correct outliers or errors, ensuring the dataset's integrity. Finally, consistent formatting of vehicle models, charging network names, and weather conditions will enhance the clarity and usability of the visualizations. By meticulously cleaning the data, the visualization tool can provide reliable and insightful analyses of EV charge and range performance.
Data Transformation	Data transformation is a vital phase in preparing data for an EV charge and range visualization tool, enabling it to deliver accurate and meaningful insights. This process involves converting raw data into a format suitable for analysis and visualization. Initially, data aggregation might be necessary, such as calculating average charging times or summarizing charging station usage patterns. Feature engineering plays a key role, creating new variables like "range per kilowatt-hour" to analyses vehicle efficiency or "distance to nearest charging station" to assess infrastructure accessibility. Normalization or standardization of numerical data, such as battery capacity or charging speeds, ensures that variables are on a comparable scale.

Data Type Conversion	Data type conversion is a fundamental step in preparing data for an electric vehicle (EV) charge and range analysis visualization tool, ensuring that all data is in the correct format for accurate processing and visualization. This process involves changing the data type of various columns to match their intended use. For instance, numerical data, such as battery capacity or charging speeds, must be converted to numeric types like integers or floats to enable mathematical operations and charting. Date and time data, like charging session timestamps or weather observation times, must be converted to date/time formats to allow for time-series analysis and chronological visualizations. Geographical coordinates, typically stored as text, need to be converted to numeric types for accurate mapping and spatial calculations.
Column Splitting and Merging	column merging combines multiple columns into a single column, which can be useful for creating unique identifiers, such as merging manufacturer and model columns into a "vehicle identifier" column, or concatenating latitude and longitude into a "location" column. This streamlines data management and enables the creation of composite keys for relational data. Both techniques are crucial for restructuring data to better suit the visualization tool's requirements, improving data clarity, and enabling more effective analysis of EV charge and range parameters.
Data Modeling	A relational model is often preferred, organizing data into distinct tables such as "Vehicles," "Charging Stations," "Environmental Data," and potentially "Usage Data," with clearly defined relationships between them. The "Vehicles" table would contain static vehicle specifications, while the "Charging Stations" table would hold infrastructure details. The "Environmental Data" table would store weather and terrain information, and "Usage Data" (if available) would capture driving and charging habits. Primary and foreign keys are used to establish relationships, enabling joins and queries across tables.
Save Processed Data	The processed data, having undergone cleaning, transformation, and modelling, should be saved in a structured and accessible format. Common formats include CSV files for portability, Parquet files for efficient storage and retrieval of large datasets, or relational database tables for structured querying. Cloud-based storage solutions, like AWS S3 or Google Cloud Storage, provide scalability and accessibility for large datasets. Regular backups should be implemented to prevent data loss and ensure data integrity.