**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2025 |
| Team ID | LTVIP2025TMID26725 |
| Project Title | visualization tool for electric vehicle charge and range analysis |
| 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 | The datasets contain information on EV charging stations, electric vehicles in India, affordable electric cars in Germany/UK, and detailed EV specifications. |
| Data Cleaning | -Checked for missing values in all datasets.  - Some missing prices in the "Cheapest Electric Cars" dataset.  - Standardized inconsistent column names (e.g., "Range\_Km" vs. "Range"). |
| Data Transformation | -Converted price formats (Lakhs to INR, Euro to INR for consistency).  - Split "Subtitle" in "Cheapest Electric Cars" to extract battery capacity.  - Sorted vehicles by range, price, and efficiency. |
| Data Type Conversion | -Converted price columns to numeric format.  - Standardized range values (removed "Km/Full Charge" for consistency). |
| Column Splitting and Merging | Split "Subtitle" in the Cheapest Electric Cars dataset to get battery size separately.  - Merged relevant datasets on vehicle name to create a combined dataset for analysis. |
| Data Modeling | Defined relationships:  - EV Charging Stations linked by region and location.  - EV Specifications combined across datasets using Car Model as the key |
| Save Processed Data | Stored cleaned and transformed data for analysis, ensuring uniform formats and merged datasets where needed. |