**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 16 March 2024 |
| Team ID | 17 |
| Project Title | VoyageVista: Illuminating Insights from Uber Expeditionary Analytics Using 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 | * The dataset contains 1,156 rows and 7 columns. * **Columns** include START\_DATE\*, END\_DATE\*, CATEGORY\*, START\*, STOP\*, MILES\*, and PURPOSE\*. * Purpose: Analyse Uber ride data, covering trip details, ride purpose, and distance metrics. * Missing values and duplicates were identified in the dataset. |
| Data Cleaning | **Missing Values**:   * **END\_DATE**\*, **CATEGORY**\*, **START**\*, and **STOP**\*: 1 missing value each; handled by removing the incomplete row. * **PURPOSE**\*: Significant missing entries (503 rows). Replaced missing values with "Unknown" for better analysis.   **Duplicates**:   * Identified and removed duplicate entries to ensure data integrity. |
| Data Transformation | * **Filtering**: Filtered data to include only valid entries with complete trip details. * **Sorting**: Sorted rides by **START\_DATE**\* to establish chronological order. * **Calculated Columns**: Added a column for trip duration by subtracting START\_DATE\* from END\_DATE\*. |
| Data Type Conversion | Converted **START\_DATE**\* and **END\_DATE**\* columns to datetime format for accurate time-based analysis. |
| Column Splitting and Merging | **Column Splitting**:   * Extracted separate date and time components from **START\_DATE**\* and **END\_DATE**\* for detailed temporal analysis. |
| Data Modeling | * Defined relationships between **CATEGORY**\*, **PURPOSE**\*, and **Route** for analysis. * Created measures such as total miles by category, average trip duration, and number of rides per purpose. |
| Save Processed Data | Saved the cleaned and processed data as a CSV file for future use and visualization. |