### Capstone Project 2 – Cyclistic Bike share Analysis

### Tool – Power BI

### Process Phase:

#### Data Cleaning in power BI :

1. Data cleaning using power quarry editor
2. Creating reference Dimensional table in the name of Dim- Bike Rider.

During the process phase, the data cleaning steps were carried out in power BI quarry editor. The following actions were performed:

* **Checking for Duplicates:** The duplicate values were examined using built-in “Remove Duplicates” feature in the quarry editor. This helped ensure data integrity and eliminate any duplicated entries.
* **Validating Column Values:** The values in specific columns, such as rideable\_type and member\_casual, were verified to ensure consistency and accuracy. The only valid columns were retained.
* **Removing Blank Values:** The data set was checked for incomplete or blank values across all columns. Rows with missing values, particularly in columns like start\_station\_name, start\_station\_id, end\_station\_name, and end\_station\_id, were removed to ensure data completeness.
* **Removing Unwanted Columns:** Columns start\_lat, start\_lng, end\_lat, and end\_lng, which were not relevant to the analysis, were removed from each file to streamline the dataset and focus on the essential variables.
* **Renaming the columns:** Renamed the columns Ridable\_type, Started\_at and ended\_at to Bike Type, Started Time and Ended Time. Capitalized the first word.
* **Transform Column:** Extracted Date and time in to separate column in the format of “DD-MM-YR” and “HH-MM”
* **Adding the Duration Column:** A new column named “Duration” was added to calculate the duration of each ride. The value in the Duration column was obtained by subtracting the started time from the ended time.
* **Setting the Time Format:** The Duration column was formatted as “HH:MM:SS” using Time from Transform menu bar. This ensured that the ride duration was presented in a standardized time format.
* **Inserting Column:** New columns named Hours and Minutes were extracted from duration.
* **Sorting the Table:** The table was sorted in ascending order based on the date in ascending order to ensure data consistency. By cleaning data in quarry editor, the dataset was refined, inconsistencies were addressed, and the ride length information was formatted appropriately for subsequent analysis.

1. Creating reference Dim\_ Calendar Table
2. Creating Measure: A filter is creature using the column Rider\_Id.

**Building Relation Ship (Mapping)**

* The cardinality between the source table and the dimensional table Dim\_Rider was one to one.
* The cardinality between the dimensional table Dim\_Rider and Dim\_calendar was many to one.

**Dashboard Visualization:**

Number of rides by week days were calculated using bar chart.

Casual and Member portfolio was found using donut chart

Total hours of rides were calculated using bar charts.

Filters and slicer were used to easy access of data

**Insights:**

**From the Dashboard analysis, finding were interpreted and recommendations were suggested.**