**BUSINESS TASK:**

Moreno, the Director of Marketing, seeks to encourage casual riders to purchase memberships to enhance the profitability of Cyclistic bike-share. As a Junior Data Analyst, my objective is to analyze the available data to elucidate the distinctions between casual riders and members, advocate for the benefits of membership to casual riders, and provide insights on how the marketing team can effectively leverage digital media to influence purchasing decisions. This analysis will serve to support the marketing team in their promotional efforts, demonstrate the advantages of membership to casual riders, and ultimately contribute to increased profits for the company.

**PREPARE**

For this project, I will utilize the data made available at https://divvy-tripdata.s3.amazonaws.com/index.html. This data is publicly accessible for such purposes. It is important to note that the data does not contain any individual’s personal information, thereby ensuring that privacy is maintained.

The data has been organized separately for each month across multiple years. In line with the recommendations from the Google Data Analytics certification, I will be utilizing the past twelve months of data. I have downloaded each file, covering the period from December 2023 to November 2024, to my local drive in CSV format, which offers broad compatibility with various analytical tools.

**PROCESS**

To conduct a comprehensive analysis of one year's worth of data, it is necessary to consolidate all relevant tables, clean the data, and transform it for optimal usability. This refined dataset will subsequently be employed for analysis using SQL. The data cleaning process was executed utilizing Microsoft Excel. Throughout this process, numerous empty cells were identified, and 'NULL' values were inserted in their place. Additionally, the datetime format was standardized to yyyy-mm-dd h:mm to enhance compatibility with SQL requirements. The cleaned dataset was meticulously verified to ensure that all entries were appropriately cleaned and transformed, with no empty cells or duplicate records based on the ride\_id, thereby adhering to the required format. All tables maintained a consistent schema. The cleaning process is outlined as follows:

1. Downloaded data from the site in csv format
2. Opened each file with Microsoft EXCEL
3. Added filter to the table
4. Searched blank data and replaced it with ‘NULL’
5. Changed date format in started\_at and Ended\_at to make it suitable for analysis in SQL
6. Checked for duplicate data based on ride\_id

Following the cleaning and transformation of the data, it became necessary to consolidate it into a single sheet. However, due to the substantial volume of data, it was impractical to combine everything in one sheet. To integrate the data for the entire year, I utilized SQL. In order to import the data, it is essential to create both a database and a table with the following schema:

*A screenshot of a computer program

Description automatically generated*

All CSV files were subsequently imported into the table labeled 'cycledata' on an individual basis. All data was successfully loaded. To ensure accuracy, we executed an additional SQL query for verification purposes.

*A close up of words

Description automatically generated*

And confirmed the total number of tows.

*A screenshot of a computer

Description automatically generated*

To calculate the duration added new column to the table named ride\_length and calculated the difference in start time and end time. I used sql code as under:

A close-up of words

Description automatically generated

SQL does not take difference in DATETIME format, so changed the data type to INT.



Later the data was not calculating in safe mode, to change used following code:



Finally added the calculated column in the table:

A close up of blue text

Description automatically generated

A screenshot of a computer

Description automatically generated

Added weekday column to the table.

A close up of a number

Description automatically generated

A screenshot of a computer

Description automatically generated

**ANALYSIS**

**Mean of ride\_length**



A screenshot of a computer

Description automatically generated

**Description:** The average ride length for the trip was 17 minutes.

**Maximum of ride\_length**

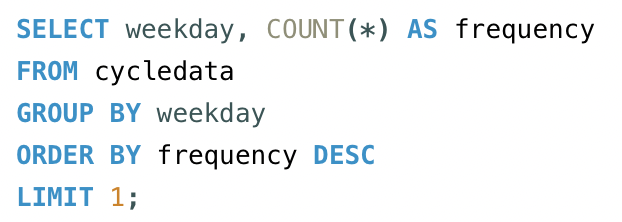


A screenshot of a calculator

Description automatically generated

**Description:** The maximum ride length for all the trips was 1560 minutes.

**Calculating mode for weekday**



A screenshot of a calendar

Description automatically generated

**Description:** Most of the trips were taken on Saturday.

**Average ride length as per membership type**

A close-up of a computer code

Description automatically generated

A screenshot of a phone

Description automatically generated

**Description:** Casual users utilized bicycles for an average duration of 24 minutes, in contrast to members, who averaged 12 minutes of usage. This indicates that, on average, casual users engage with bicycles more frequently than members.

**Average ride length as per weekday**

A white background with black text

Description automatically generated

A screenshot of a grid

Description automatically generated

**Description:** According to the data collected on weekly usage patterns, users predominantly engaged with the platform during weekends, averaging approximately 20 minutes of use.

**Monthly total number of rides**

A close-up of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

**Description:** The utilization of bicycles was significantly greater during the months of May, June, July, August, and September compared to the remaining months of the year.

**Exported data without NULL fields in CSV from SQL**

A screen shot of a computer

Description automatically generated

The exported CSV was utilized for data visualization using Tableau.

**DATA VISUALIZATION**

A graph of a bar

Description automatically generated

**Description:** The average ride length was higher for classic bikes than for electric bikes.

A graph of blue bars

Description automatically generated

**Description:** Considering weekdays, the average usage of bikes was more on Friday followed by Wednesday. This is the average of all types of bikes and user types.

A graph of different colored bars

Description automatically generated

**Description:** The data presented in the graph indicates that casual users exhibit a higher preference for classic bikes in comparison to electric bikes. In contrast, members demonstrate a similar level of usage for both types of bikes, with their average usage patterns remaining relatively consistent throughout the week. Notably, casual users predominantly favor classic bikes, particularly on Fridays, while their engagement with electric bikes remains comparatively lower.

**Recommendation:** The data indicates that casual users engage with the service more frequently than members, on average. Furthermore, it appears that casual users require additional benefits during weekends. Presenting a comparative analysis of costs to casual users and members may enhance marketing efforts.

**Further scope:** It is advisable to examine the map to identify stations that are particularly advantageous for casual users. While the location can be determined using longitude and latitude coordinates, this method can be time-intensive. In contrast, obtaining precise addresses that include city, state, and country would facilitate a more efficient and comprehensive analysis of the locations.