Case Study: How Does a Bike-Share Navigate Speedy Success

Scenario:
You are a data analyst working in the marketing analyst team at Cyclistic, a bike-share company. The director of the marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

About Cyclistic Operations:
In 2016, Cyclistic launched a successful bike-share offering. It has a fleet of 5824 bicycles those are geo tracked and locked into a network of 692 stations across a state. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Stage 1: Ask

A Clear Statement of the business task
How do annual members and casual riders use Cyclistic bikes differently?

Primary Stakeholder:
Cyclistic Executive Team (This team will decide whether to approve the recommended marketing program

Secondary Stakeholders:
Marketing Director (Responsible for the development of the campaigns)
Analytics Team (responsible for collecting, analysing and reporting)

Required Deliverables:
1. A clear statement of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendation based on your analysis

Stage 2: Prepare

Data Collection:
Data is collected from the Cyclistic's internal operational database. This is an .csv extract.
This is internal (First Party) and Structured Data.
Data is collected from 01 – Jan – 2023 to 30 – Sep – 2023 for the analysis. (9 months or 3 quarters of data)
To collect all these data from the operational data base appropriate data exchange agreement is signed.
Data will be received in nine (9) separate files, one file for each month.

Data Integrity:

To ensure data integrity we have to check the Credibility, Ethics, Privacy and Access

- **Credibility of Data Collected**
    - **R**eliability: Accurate and not biases as it is scoured from the internal (first party) automated operational system.
    - **O**riginal: Yes, original data source is within the reach and accessibility.
    - **C**omprehensible: As the data is sourced from the operation system, it is not fully comprehensive for Data Analytics hence transformation is required to greater extent.
    - **C**urrent: Data is current as per the standard timeline set for the analysis.
    - **C**ited: Data source is known an vetted and managed by the organization itself.

- **Ethics of Data**
    - **E**thics: The data is classified as Restricted. The data shouldn't be shared other than the approved stakeholders with in the organization. Operation system data is owned by the Global Operations Head. The data set received for this analysis is owned by the Data & Analytics Head. All the aspects of Transaction Transparency, Consent, Privacy, Access and Classification are agreed between the owner and exchange agreement has been signed.

Collected Data (.csv file):

Month	Collected File
Jan – 2023	202301-divvy-tripdata.csv
Feb – 2023	202302-divvy-tripdata.csv
Mar – 2023	202303-divvy-tripdata.csv
Apr – 2023	202304-divvy-tripdata.csv
May – 2023	202305-divvy-tripdata.csv
Jun – 2023	202306-divvy-tripdata.csv
July – 2023	202307-divvy-tripdata.csv
Aug – 2023	202308-divvy-tripdata.csv
Sep – 2023	202309-divvy-tripdata.csv

Stage 3: Process

A Data Landing Area is designed and implemented with regards to the data privacy and accessibility protocols. 9 data files are received in the landing area.

To process those data, **PowerBI** is chosen as per the organization.

Step 1: Extracting and Loading
- All the 9 raw data files are loaded in PowerBI for processing
- The PowerBI Data Analysis file is saved as "CaseStudy_Cyclistic_Oct_2023.pbix"

Step 2: Transformations
- While loading the data PowerBI automatically **Promoted Headers** and Detected the Attribute Type and **Changed Attribute Type**
- Reviewing column names in order to align them for merging into a single workable dataset
- Renaming the remaining columns of each dataset to match the first
- Reviewing the Attribute Types before merging
- Append all the files into one workable dataset and named it as "Cyclistic_TripData_Jan23_to_Sep23"
- Rename columns as per standard / preference as required for analysis in the merged table
- Remove duplicate and null values
- Make sure User_Type and Bike_Type are limited to defined reference values each
- Separate the 'start' timestamp column
- Rearrange the columns as preferred, remove the unused columns (Start_Latitude, Start_Longitude, End_Latitude, End_Longitude) for the analysis
- Add new column as Ride_Length (HH:MM:SS) = End_Time – Start_Time
- Add new column as Day_of_Week = WEEKDAY (Start_Time), This column will contain values 1 to 7, 1 = Sunday and 7 = Saturday
- Remove bad data for ride_length_sec – deleting any trip >24 hours or <0 sec

Stage 4: Analyze

1. Calculate mean of Ride_Length (Creating a measure named as Mean_Ride_Length)
2. Calculate max Ride_Length (Creating a measure named as Max_Ride_Length)
3. Calculate the mode of Day_of_Week (Creating a measure named as Mode_Day_of_Week)
4. Calculate the average ride length for "members" & "casual riders"
   a. Creating a measure named as Avg_Ride_Length_Mbr for "members"
   Avg_Ride_Length_Mbr=CALCULATE(AVERAGE(Cyclistic_TripData_Jan23_to_Sep23[Ride_Length]),FILTER(Cyclistic_TripData_Jan23_to_Sep23,Cyclistic_TripData_Jan23_to_Sep23[User_Type]="member"))
   b. Creating a measure named as Avg_Ride_Length_CR for "Casual Riders"
   Avg_Ride_Length_CR=CALCULATE(AVERAGE(Cyclistic_TripData_Jan23_to_Sep23[Ride_Length]),FILTER(Cyclistic_TripData_Jan23_to_Sep23,Cyclistic_TripData_Jan23_to_Sep23[User_Type]="casual"))

5. Calculate the average Ride_Length for users by Day_of_Week.
6. Calculate the number of rides for users by day_of_week by adding count of trip_id to values

Stage 5: Share
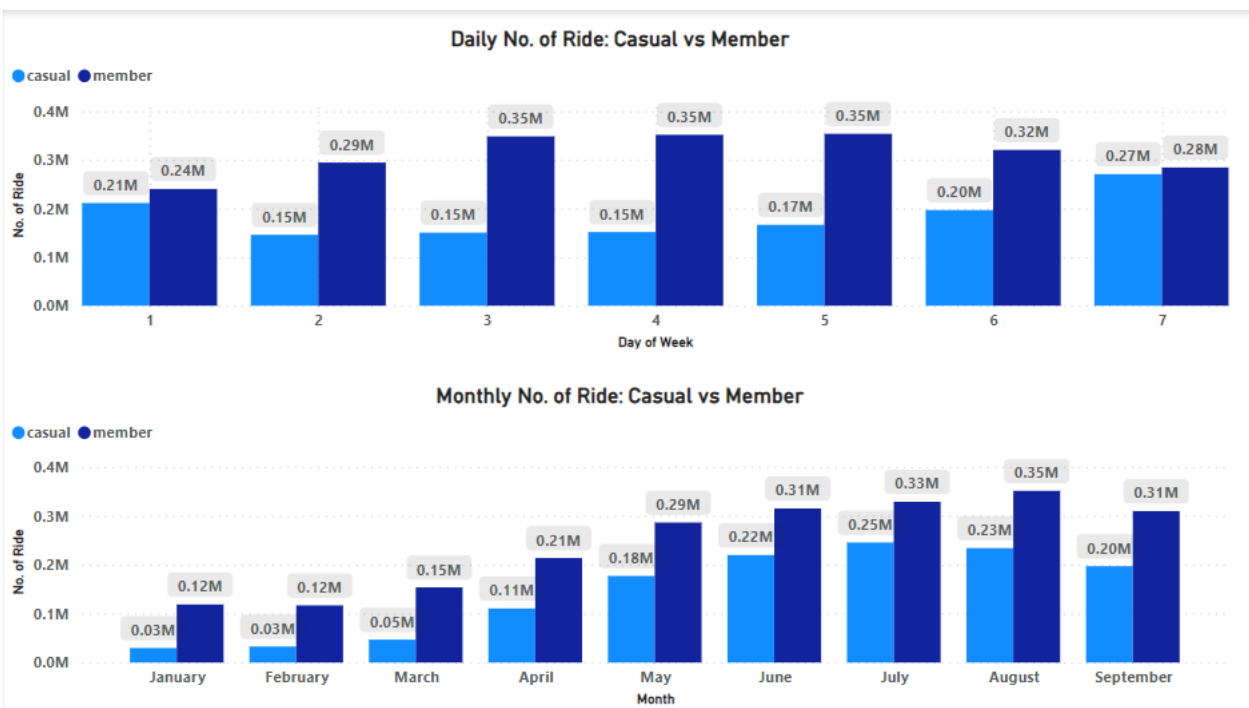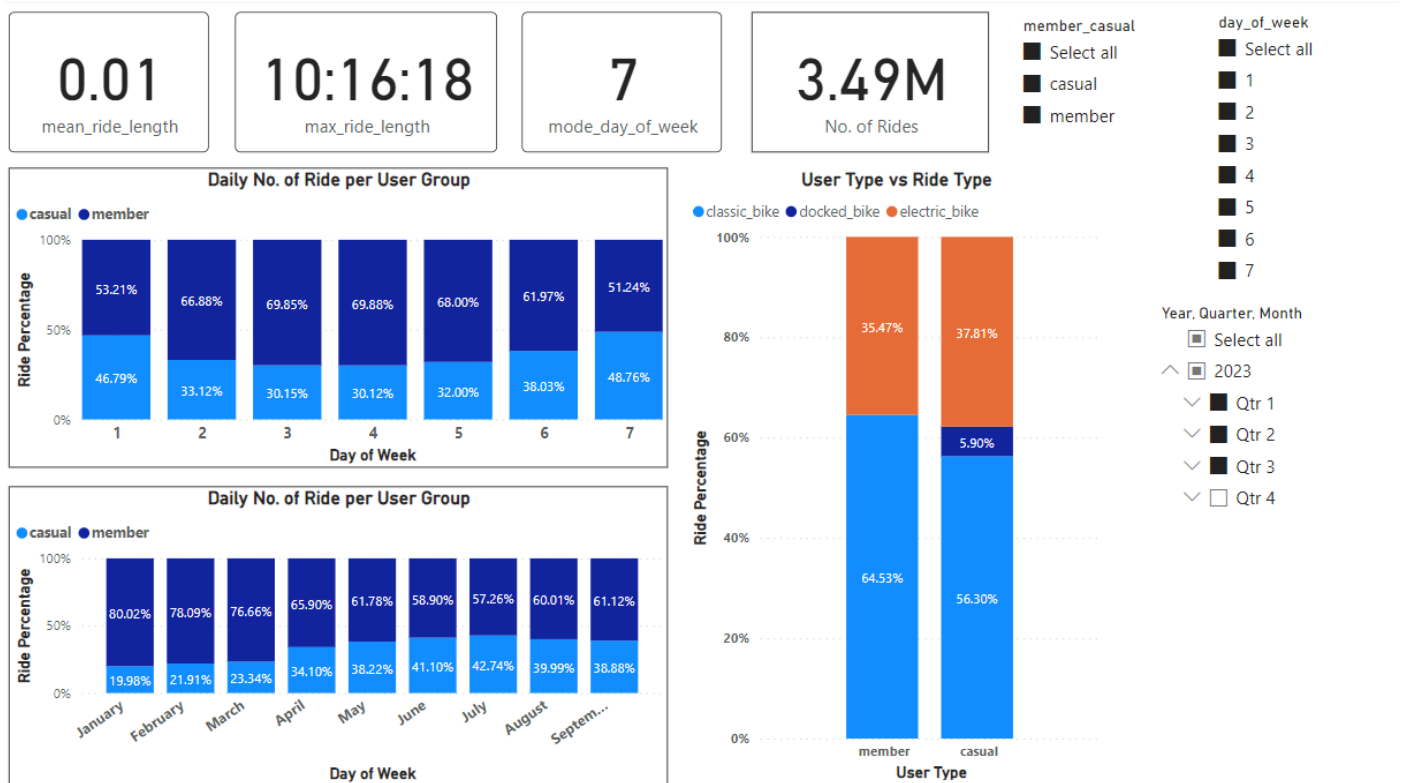
- Members use wither classic bike or electric bike. Casual users use all three types of bikes, however docked bike is used significantly less.
- Casual users taking a ride increases as the months progresses until July then it tries to reduce gradually. This might be a seasonality. As the winter is approaching the casual user ride decreasing. Similar pattern can be observed in case of Member users however at higher number of rides.
- On Monday and Sunday, the casual user ride is at the higher end, however in the week days it is lower.

Stage 6: Act

- Having "Member" as users will definitely provide stability to the business operations, however "Casual" users can't be ignored as they also have significant contribution to the business operations.
- To retain existing "Members" and to increase the no. of "Members" company can provide some extra benefits, this might attract new "Members" either from the "Casual" user group or totally new user.
- Instead of focusing on converting casual members to "Members" company should focus on leveraging them by introducing weekly passes and slight increasing the price of one-way / both way daily passes especially during the peek months (April to August).