**Cyclistic Bike-Share Analysis**

**By Oluwadamilola Ojo**

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

ScenarioYou are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of 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. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

**Mission**

To design marketing strategies aimed at converting casual riders into annual members

**The ASK Phase**

How do annual members and casual riders use Cyclistic bikes differently?

1. The business task: To understand how casual riders and annual members use Cyclistic bikes differently.
2. Data sources used: The data has been made available by Motivate International Inc. under this license <https://ride.divvybikes.com/data-license-agreement>
3. Key Stakeholders: The director of marketing and “my manager”, Cyclistic marketing analytics team and Cyclist executive team

**The Prepare Phase**

1. Data location: The dataset used is a public dataset on Amazon AWS
2. Data Organization: It is organized in monthly format
3. Credibility and bias issues: The dataset is Reliable, Original, Comprehensive, Current, and Cited
4. The data was collected by the company itself and saved in csv format. It has several columns. Some may not be needed for the analysis; hence we keep information of users private.
5. For this analysis I used dataset from May 2021 to April 2022
6. Downloaded files
7. Created folders and sub folders
8. Unzipped, opened, renamed the files as cyc\_month\_year e.g. cyc\_05\_2021 and saved files as .xls

**The PROCESS Phase**

1. I used Excel to clean the data and used R for the analysis because the files are too large for Excel
2. I sorted and filtered to remove duplicate, leading and trailing spaces. This was repeated in all the 12 months’ dataset.
3. I deleted columns like *start\_lat, start\_lng, end\_lat, end\_lng* were deleted because they are not needed for the analysis
4. I renamed column *member\_casual* to *ride\_plan for* better unnderstanding
5. I created a new column **ride\_length** to find the time difference between the end and start time.
6. I created a column **day\_of\_week** to calculate the day of week that each ride started
7. I filtered and deleted where ride\_length value less than or equals 00:00:00. Quite a lot in hundreds were found
8. I saved the cleaned files in .xls
9. This reduced the size of the files

**The ANALYZE Phase**

**The Key Objectives are to**:

1. Aggregate the data so it’s useful and accessible.
2. Organize and format the data.
3. Perform calculations.
4. Identify trends and relationships.

For this phase, I switched from Excel to R because combining all the months into a single worksheet might be impossible or take too long to process.

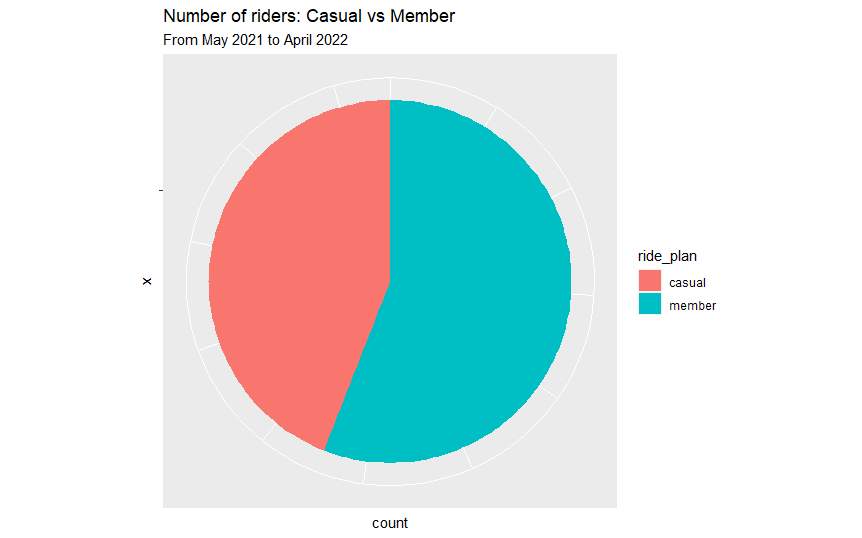
1. Installed and loaded packages- tidyverse, lubridate, ggplot2
2. Imported the files into folder
3. Combined the 12 months datasets into a single dataframe- cyclist\_year
4. Did some statistical summary(mean, median, maximum, minimum) on the dataset
5. Aggregate analysis-
   1. Average ride length
   2. number of rides group by day\_of\_week and ride\_plan
   3. number of rides group by rideable\_type and ride\_plan
   4. number of rides for ALL users
   5. average by ride group by rideable\_type and ride\_plan

**The SHARE Phase**

Task:

To create visualizations to share findings

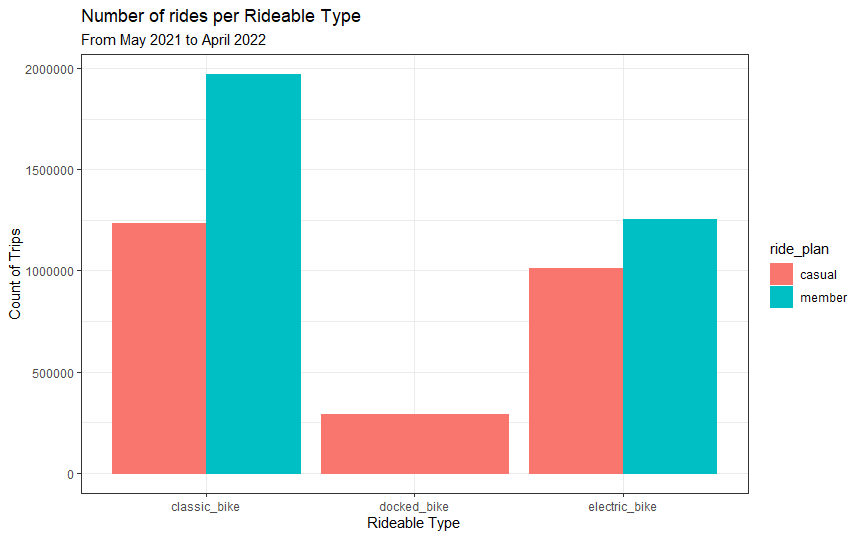
1. **Number of riders: Casual vs Member**



**Observation 1**

Member riders (3220897) used the Cyclistic bikes more than the Casual riders (2536002) from month May 2021 to April 2022

1. **Number of Rides Per Rideable\_Type**



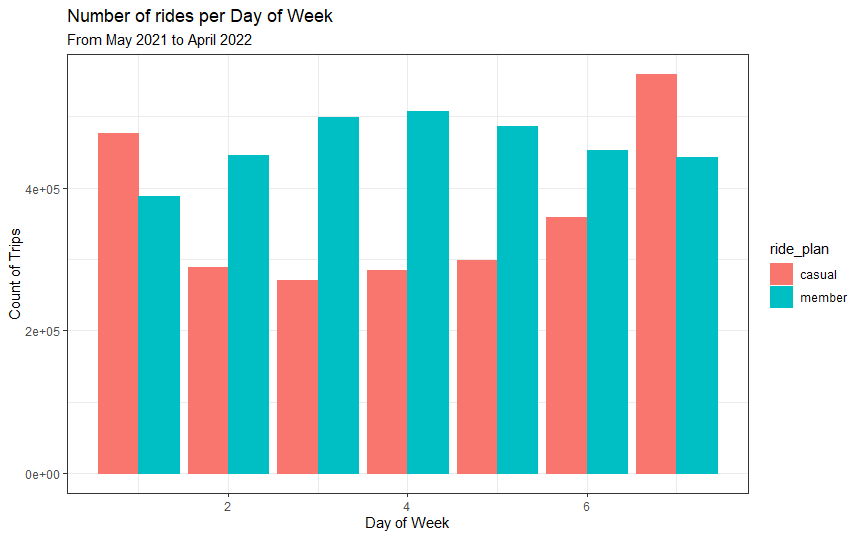
**Observation 2**

* 1. Member users did not use Docked\_bike for the Month May 2021 to April 2022
  2. Member users used the Classic and Electric Bikes more than Casual users for the Month May 2021 to April 2022

**Recommendation**

Cyclistic may invest:

1. less in Docked bike
2. more in Classic Bikes, followed by Electric Bikes
3. **Number of Rides by Day of week**



**Observation 3**

1. Casual users used the bikes more on Sunday and Saturday
2. Member users consistently use the bikes on all days. Tuesday and Wednesday seemed to be more consistent

**Recommendation**

Incentives and promotional offers can be offered to Casual users that use the bikes more on weekends.

**Future Work**

Deeper analysis can still be done to gain more insights into the data

Link to data used: