Cyclistic Bike Share Data Analysis

Introduction

This is a data analytics capstone project, carried out in fulfillment of the requirement for the Google data analytics professional certificate course, Course 8, module 2.

Project overview

Cyclistic is a bike-share company that enables people to rent bikes for their mobility needs and return them at various locations.

The finance analyst of the Cyclistic aims to maximize the number of annual members which she believes would impact future growth, by converting casual bike users to Members.

Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She also notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Note: Members are riders with annual subscription, while casual riders use a single-ride or full day passes.

My role:

In this project, I assumed the role of a junior data analyst in the marketing team to uncover insights from their company’s data, necessary to influence their business objective and offer recommendations to optimize work flow.

Business objective:

Design an effective marketing strategy targeted at converting causal riders to annual members.

Business Question:

The Cyclistic bike share marketing needs to understand the following in order to carryout an effective marketing program in future using the following questions:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

Data Source:

The dataset of the previous 12 months (January 2023- December 2023) was accessible and contained in an index file called divvy trip data index file, from which I downloaded and extracted the datasets as csv file.

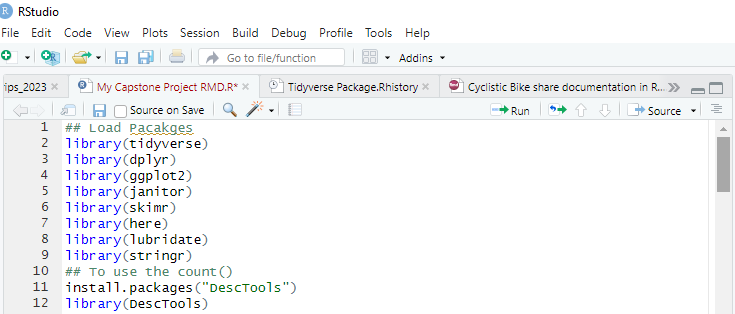
Tools

* Rstudio : Data wrangling, manipulation, analysis and visualization.
* Tableau: Dashboard to visualize insights.

Data Preparation and observation

I worked with the most recent annual dataset for the year 2023, which was available for use. The dataset was first uploaded to google sheet but after figuring out that some months contained very large number of datasets, I decided to use the RStudio desktop application.

I uploaded the dataset into the Global environment in R, and to have access to the full functionalities of R, I installed the necessary packages for my analysis and loaded them.



Using the str() function, I inspected the schema of each dataset and observed that all of them contains 13 variables, labeled:

* ride\_id
* member\_casual (which is the user type)
* rideable\_type (available bike type)
* started\_at
* ended\_at
* start\_station\_id
* end\_station\_id
* start\_station\_name
* end\_station\_name
* start\_lat
* end\_lat
* start\_lng
* end\_lng

Note that columns with the label: start\_station\_name and end\_station\_name were hidden for most of the months and so was not used in my analysis and my outcome not influenced by them.

Process

Data cleaning

1. Merge dataset:

After loading the dataset for 2023 into my R studio global environment, I observed the consistency in the column names and arrangement before proceeding to write a code chunk:

* Divvy\_tripdata\_2023merged<-bind\_row(January\_divvytrip\_2023 – Dec\_divvytrip\_2023)

to merge them into one dataframe. This way I could save time, stay organized and focus better. After merging, I had a total of 57,198,77 obs. and 13 variables

1. Data Cleaning and Manipulation:

After successfully merging the datasets, I created four new columns to the merged dataframe:

1. trip\_duration(in minutes)
2. time\_in\_hours
3. month
4. day\_of\_week

After which I used the clean\_names() to make the title of the variables consistent,

* clean\_names(divvytrips\_2023Merged)

1. I went over the business task and the dataset I was working with, to ensure that I was in line with goal and my dataset, comprehensive.

Then, I proceeded to delete rows with “null” values in the trip duration column, and rows where trip\_duration equals 0sec. I ran the code chunk

* divvytrips\_2023 <- divvytrips\_2023Merged[!(divvytrips\_2023Merged$tripduration <= 0),]

This cleaning process reduced my dataset to 5718608 obs. and 17 columns and also created a new dataframe, divvytrips\_2023, which I used through out my analysis.

With these processes completed, my datasets was ready for analysis.

ANALYZE

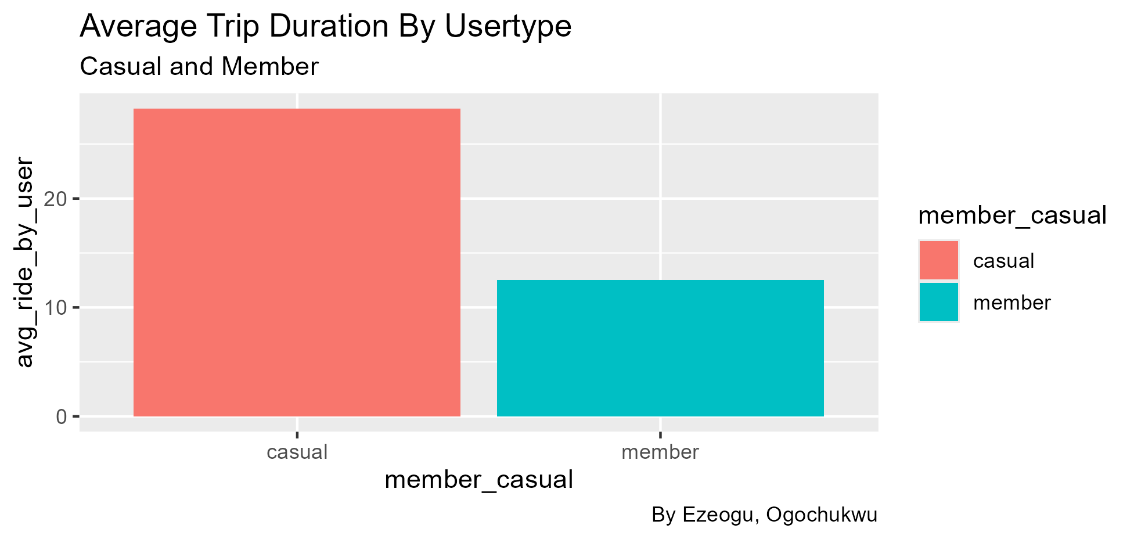
The following exploratory analysis was carried out on the dataset to gradually gain insight into Cyclistic bike data for the year 2023:

1. Total number of rides: The total number of rides = 57,186,08 rides
2. Average trip duration: The average trip duration = 18.189 mins
3. Maximum trip duration/longest ride length = 98489.07mins
4. Mode trips by day = Saturday

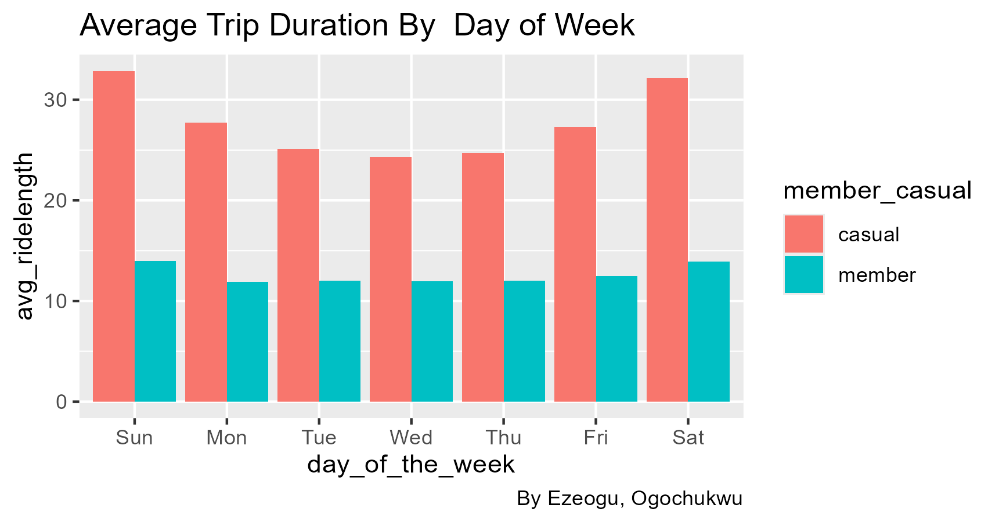
Next, I performed more specific analysis tailored to the mode of bike usage between annual members of Cyclistic bike share and casual riders. I also developed charts in R to better communicate my derived insights.

1. Average trip duration by user type (member and causal):

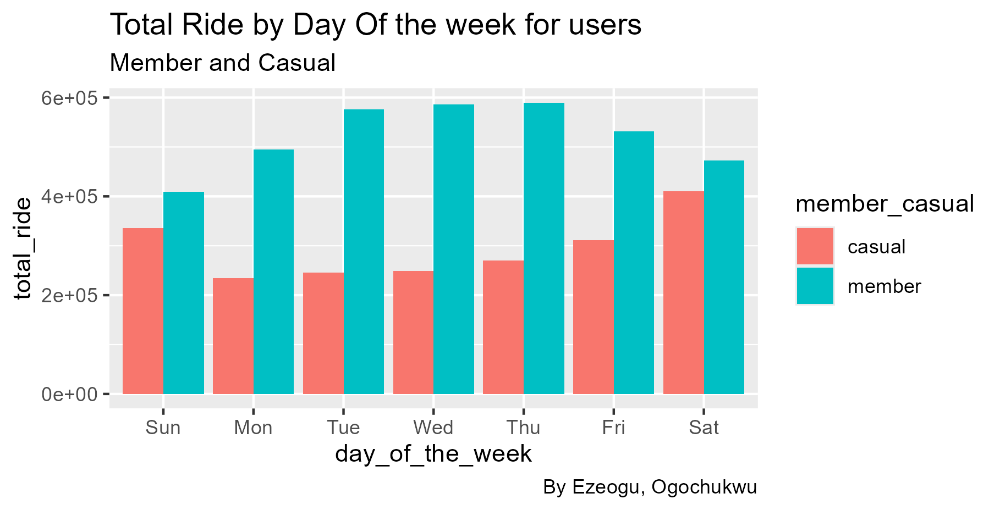
In the chart, it can be seen that casual riders had the longest average trip duration of 28.26 mins compared to annual members which has 12.530 mins.



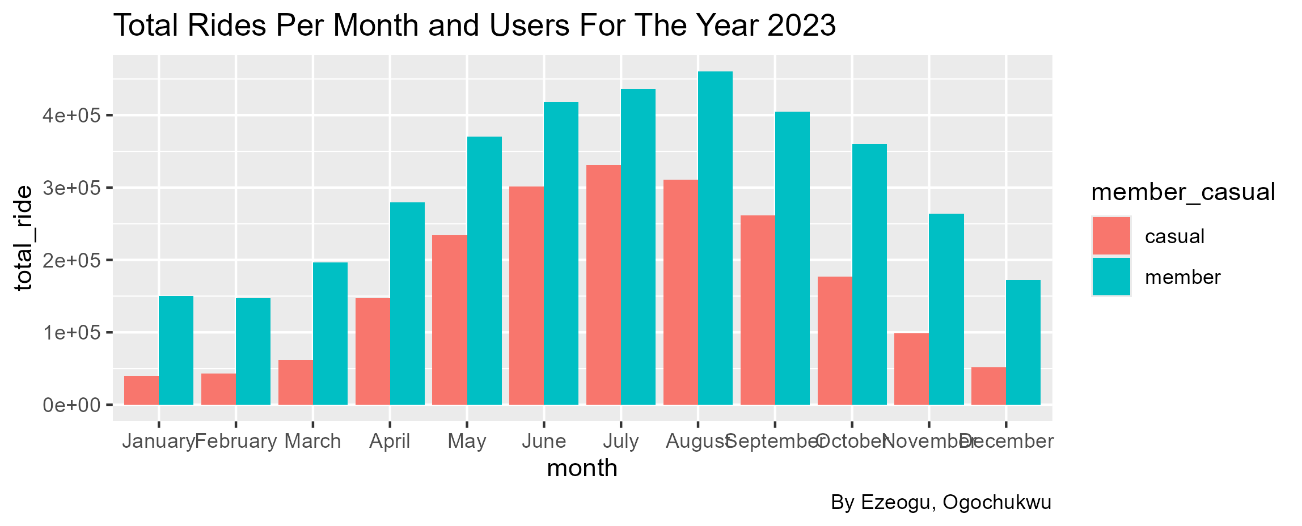
1. Average trip duration by user type by day of the week: The chart clearly illustrates the average usages pattern of both user types by day of the week in the year 2023. It can be seen that casual riders had the longest trip duration for each day with members having relatively the same trip length for each week throughout the year.

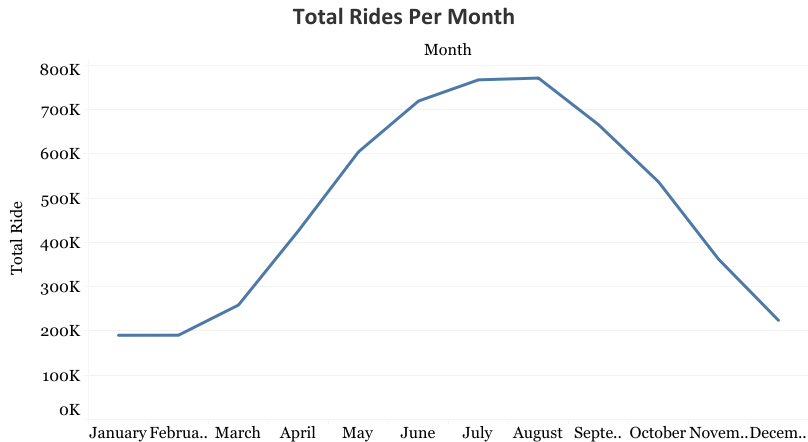


1. Total rides by day of the week: The chart shows the total ride count, measuring up to 60,000,00. It can be seen that members used the bike the most throughout the week. And when compared to the average trip duration by day of week, members make use of the bikes for so many times but for shorter distances compared to casual riders.

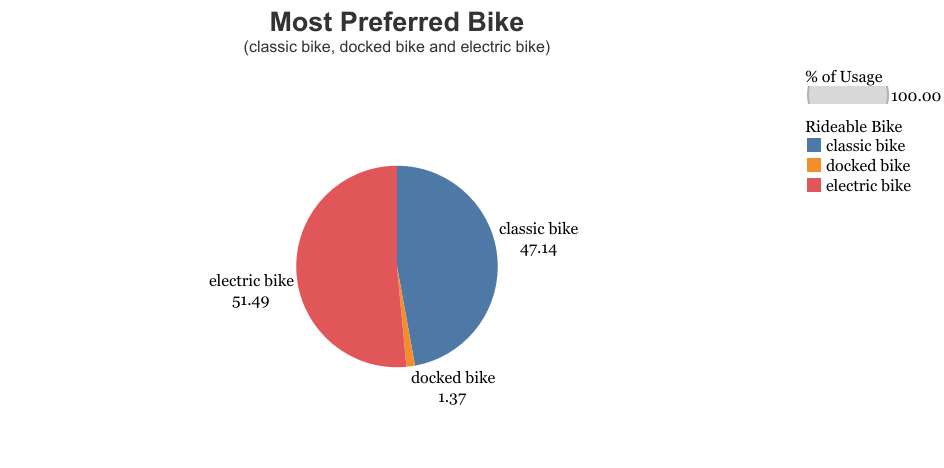


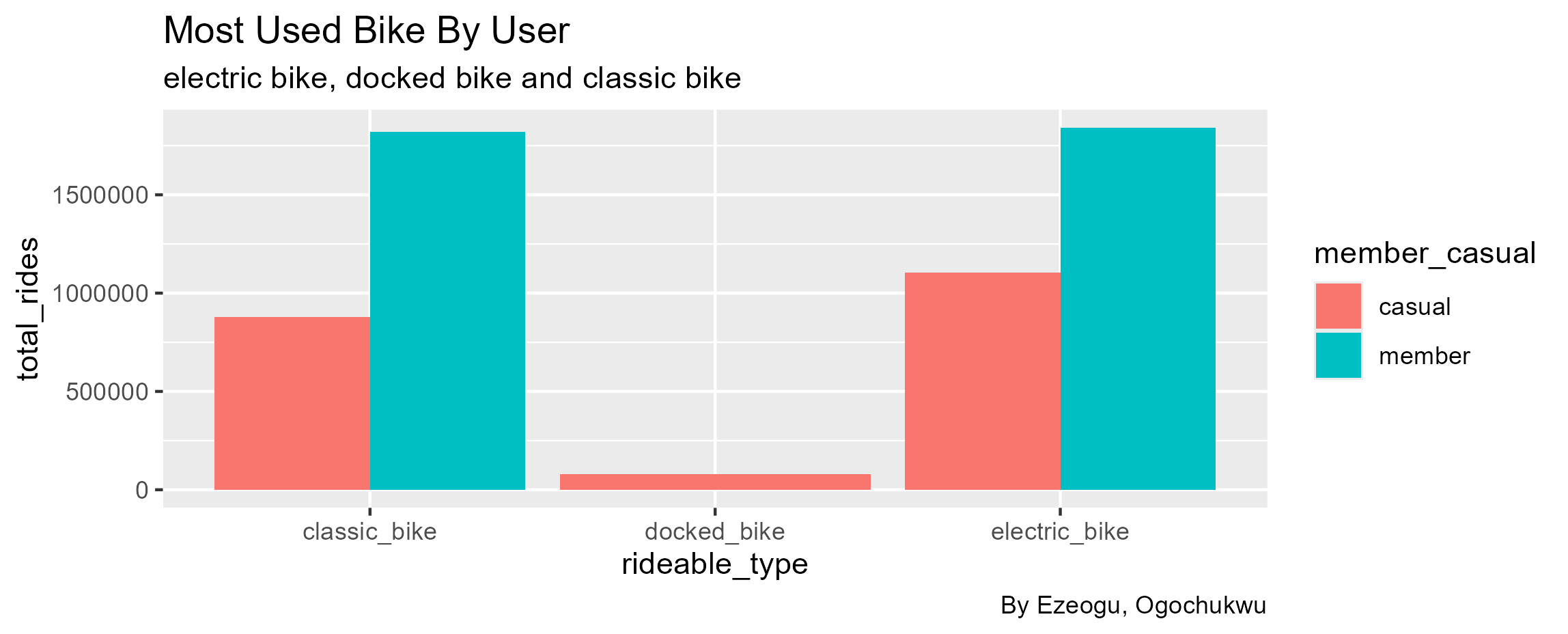
1. Total of rides by user type per month: Measuring up to 40,000,05, the chart gives a clear illustration of the peak months for both user types. It shows that members had the highest number of rides, with the month of August being the peak.



The line chart also gives a clear illustration of the peak months alongside the decrease.

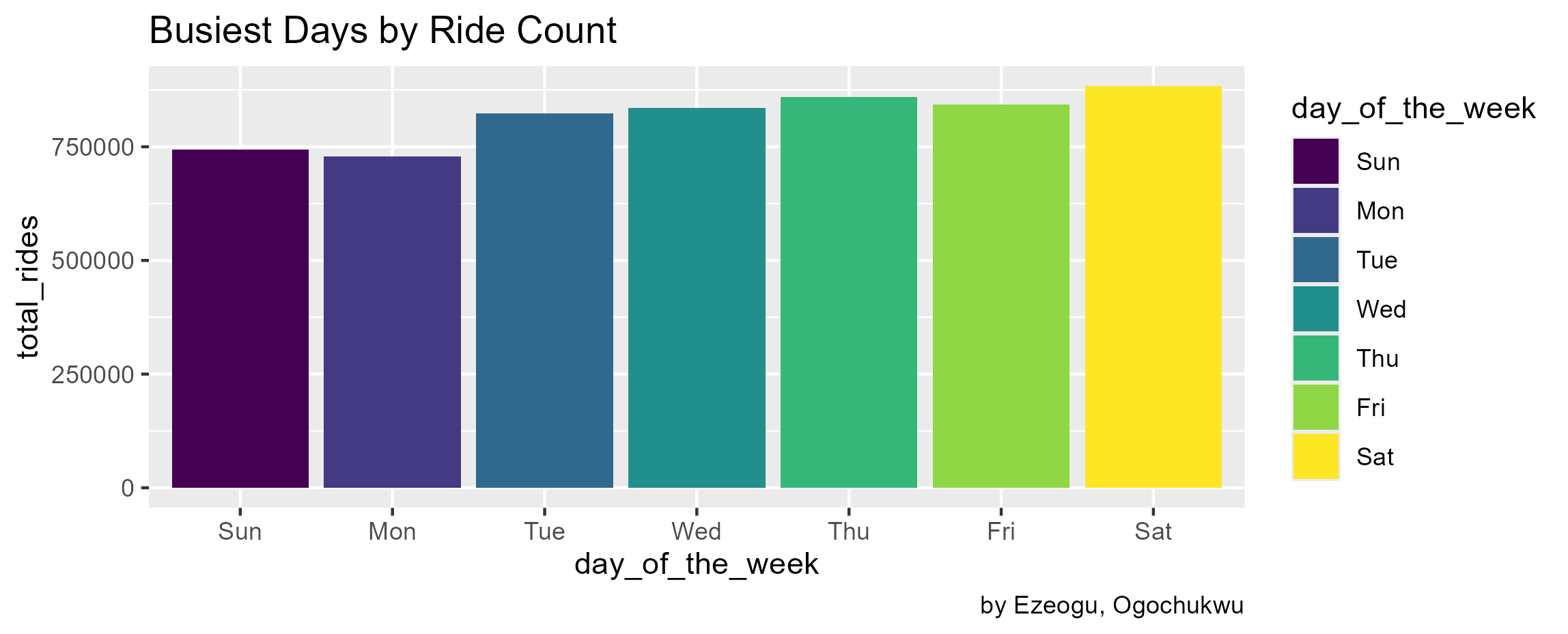
1. Preferred bike for both user type: Cyclistic bike share company provides her users with three rideable bike options, namely: Docked bike, Electric bike, Classic bike.



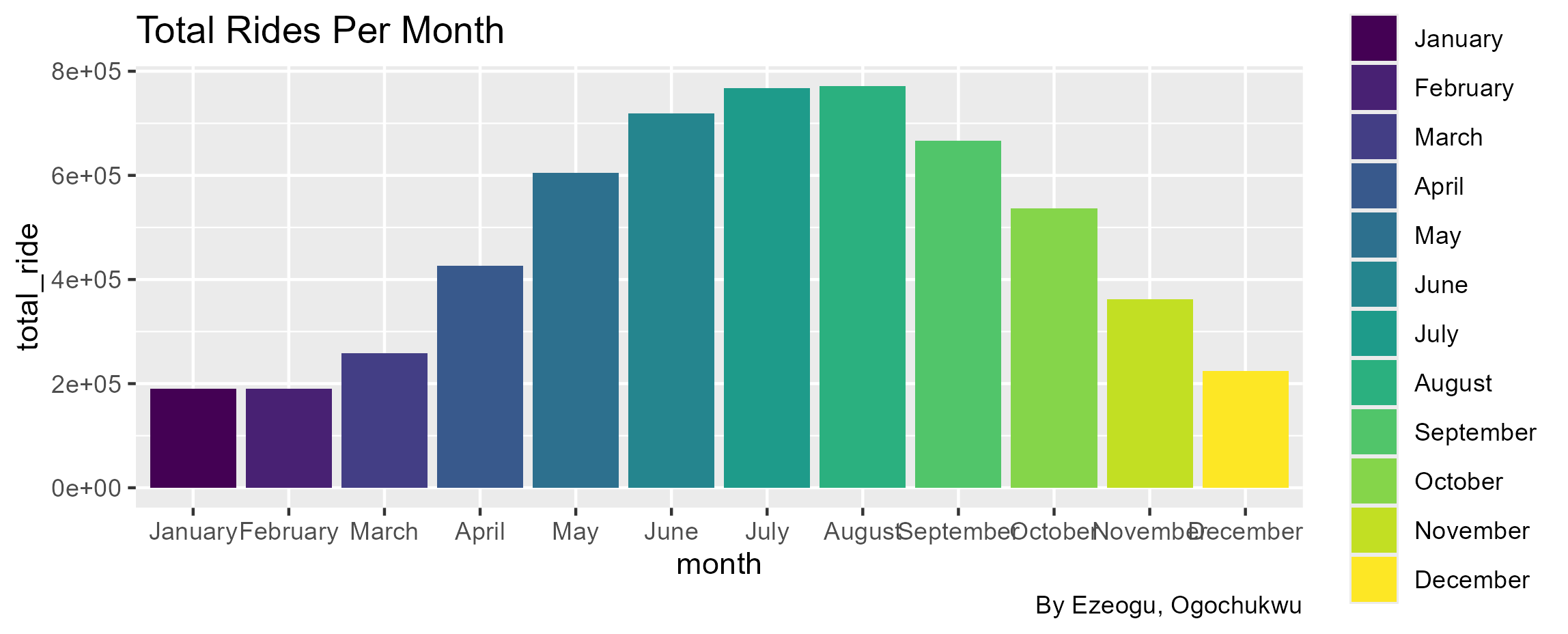


Further analysis with charts:

In the year 2023, Cyclistic bike had more riders on Saturdays and less on Mondays and Sundays.



Busiest Months of the year by ride count: Cyclistic bike usage increases in the Summer season, around the months of June, July and August. And decreases towards Winter season up till Fall when it peaks again in the month of May.



Share:

The insights derived from the Cyclistic bike data are summarized thus:

1. Casual riders trust Cyclistic bike for long trips and so makes up to 69.28% of the total average trip duration with Annual members taking up 30.72%.
2. Annual members covered 64% of the total bike rides in the year 2023, compared to Casual riders.
3. Saturday and the summer season have the highest number of bike rides, which is possibly influenced by the weekend spree and the summer holidays respectively.
4. Annual member has a relatively similar ride length and so do not use Cyclistic bikes for longer durations as would Casual riders.

Recommendations

1. Annual members are typically stable users of Cyclistic bikes, as they use the bike for their basic mobility needs, and make up a higher percentage of the total bike usage compared to Casual riders. Casual riders are seasonal riders who make use of Cyclistic bike for recreational activities, tours among others.

However, based on the average trip duration of both user types, Casual riders make up the longest average trip durations for the year with around 70% of the total average ride length.

1. In order to effectively convert casual riders to annual members, Cyclistic should strategically target their digital adverts on weekends, especially on Saturdays. And also, during the Summer season where there is a noticeable increase in bike usage by causal riders. The rise in bike usage by casual riders during this season can be attributed to the summer holidays.
2. To make the digital advert yield successful conversion of casual riders and the general public, I recommend that Cyclistic offers new subscribers discounts on their subscription plans as a summer holiday bonus. This bonus should last for a period of time within the holiday season to attract a large number of new subscribers and convert casual riders. After which, they can curate adverts for Saturdays to make their normal subscription plans known to both casual riders and the public at large.are