GOOGLE DATA ANALYTICS CAPSTONE

CYCLISTIC CASE STUDY:

HOW DOES A BIKE-SHARE NAVIGATE SPEEDY SUCCESS?

Dr Lebede Ngartera

Email: ngarteralebede12@gmail.com

https://github.com/Teraces12/Google-Data-Analytic-Certificate-How-a-Bike-Share-Navigates-Speedy-Success-

September 21st, 2023.

AGENDA

Winning Data Analytics Race with Dr Lebede Ngartera

Cyclistic

BIKE-SHARE

INTRODUCTION

As a part of the **Google Data Analytics certification course**, I was given the opportunity to conduct a case study on **Cyclistic** to analyze the data and provide insights on the users' behavior.

The aim of this case study is to provide actionable insights for the company to consider in their marketing strategy.

In order to improve its marketing strategy and drive future growth, the bike share company, **Cyclistic**, is based **in Chicago** and offers services to its users.

The company is interested in **understanding the behavior of its users** while using the services.

In this project, I will follow the steps of the data analysis process: ask, prepare, process, analyze, and share.

Winning Data Analytics Race with Dr Lebede Ngartera

Background

SCENARIO

BACKGROUND

Cyclistic operates a fleet of more than 5,800 bicycles, which can be accessed from over 600 docking stations across the city. Bikes can be borrowed from one docking station, ridden, then returned to any docking station. There are also user-friendly bike options, including electric bikes, classic bikes, and docked bikes, making Cyclistic services more inclusive to people.

Over the years, marketing campaigns have been broad and targeted a cross-section of potential users. Data analysis has shown that riders with an **annual membership** are more profitable than **casual riders**. Lily Moreno, the director of marketing, wants to implement a new marketing strategy to convert casual riders into annual members. She believes that with the right campaign, there is a very good chance of such conversions between the user types.

Lily has tasked the marketing analytics team to analyze past user data of one year to find trends and habits of Cyclistic's users to help create this marketing campaign. The marketing analyst team would like to know:

How **annual members and casual riders** differ?

Why **casual riders** would buy a membership?

How Cyclistic can use **digital media to influence casual riders to become members?**

In this context, we need to analyze the **Cyclistic historical bike trip data** to **identify trends** in the usage of bikes by **casual and member riders**.

Winning Data Analytics Race with Dr Lebede Ngartera

Business Objective

PHASE 1: ASK

BUSINESS OBJECTIVE

The aim of the case study is **to identify opportunities** for targeted marketing campaigns **to convert casual cyclists into annual members.**

This will be achieved by **analyzing cycling data** and **understanding user behavior and preferences**.

The ultimate goal is to **increase profitability** and **stimulate future growth for the company.**

Stakeholders

PHASE 1: ASK

STAKEHOLDERS

The stakeholders involved in this case study are:

1-Lily Moreno: As the **Director of Marketing at Cyclistic**, Lily is **responsible for implementing the marketing campaigns** at Cyclistic.

2-Cyclistic's Marketing Team: This team holds the **responsibility for conducting the analysis and developing the marketing strategy** based on **the insights gained**.

3-Cyclistic's Casual Riders: This group constitutes **the target audience of the marketing campaign**, with a primary focus on **analyzing their behavior and preferences**.

4-Cyclistic's Annual Members: This group is the **target for conversion from casual riders**, and their **behavior and preferences** will be **compared** to those of **casual riders during the analysis**.

5-Cyclistic's Investors and Shareholders: With a financial interest in the **company's success, investors and shareholders** may closely follow the results of the analysis and any subsequent changes to the marketing strategy.

Winning Data Analytics Race with Dr Lebede Ngartera

9

Business task:

Stakeholders's Expectations

PHASE 1: ASK

BUSINESS TASK: STAKEHOLDERS'S EXPECTATIONS

As a Data Analyst, my **duty** is to **support stakeholders' data-driven decision-making** by following these steps:

a) **Identify trends and patterns** in the usage of bikes by casual and member riders.

b) **Understand the behavior and preferences** of these two user groups.

c) **Identify opportunities for targeted marketing** campaigns **to convert casual riders into annual members**.

d) **Use insights** from the **analysis to inform the development of a new marketing strategy** for the company.

e) The **overarching goal** is to **increase profitability and drive future growth.**

Winning Data Analytics Race with Dr Lebede Ngartera

Data location

PHASE 2:
DATA PREPARATION

DATA LOCATION

1-The **data used** for this analysis was obtained from Motivate International Inc. and can be accessed via the link provided **[link](https://divvy-tripdata.s3.amazonaws.com/index.html).**

2-It includes **12 months of historical trip data from Cyclistic**, a fictitious **bike-share** company based in **Chicago**.

3-It should be noted that this **data is public** and can be **used to study** how different types of customers **use Cyclistic's bikes**.

Winning Data Analytics Race with Dr Lebede Ngartera

13

Data Organization

PHASE 2:
DATA PREPARATION

DATA ORGANIZATION

For this project, the utilized data comprises monthly CSV files spanning the past **12 months(December 2021 - November 2022)**.

These files encompass **13 columns** of information pertaining.

to ride details, including ride id, ridership type, ride time, start and end locations, and geographic coordinates, among others

The organization of the data has been structured in a **manner conducive to the analysis of trends and patterns** in **the usage of Cyclistic's bike share services.**

Winning Data Analytics Race with Dr Lebede Ngartera

Does our data ROCCC?

Are there issues with bias or credibility in this data?

PHASE 2:
DATA PREPARATION

Winning Data Analytics Race with Dr Lebede Ngartera

DOES OUR DATA ROCCC?

Does our data ROCCC(Reliable, Original, Comprehensive, Current and Cited)?

Motivate, Inc. collected the data [link](https://divvy-tripdata.s3.amazonaws.com/index.html) for this analysis directly through its management of the **Cyclistic Bike Share** program for the **City of Chicago**.

The **data is complete and consistent**, as it includes information on all trips made by users and is not a simple sample.

They are also **up-to-date**, as they are **published monthly** by the **City of Chicago**. The data is made available to the public by the City of Chicago.

Winning Data Analytics Race with Dr Lebede Ngartera

Licensing, privacy, security, and accessibility

PHASE 2:
DATA PREPARATION

Winning Data Analytics Race with Dr Lebede Ngartera

BIKE-SHARE

LICENSING, PRIVACY, SECURITY, AND ACCESSIBILITY

1-The data used for this analysis is **released under a specific license [link](https://ride.divvybikes.com/data-license-agreement)** and is made **available for use** in this analysis.

 2-To protect the privacy of users, all identifying information has been removed from the data.

3-This limitation on the data restricts **the scope of the possible analysis,** as **it is not possible to determine** whether **casual riders are repeat users or residents of the Chicago area**.

Sufficiency of data to answer business questions

PHASE 2:
DATA PREPARATION

Winning Data Analytics Race with Dr Lebede Ngartera

SUFFICIENCY OF DATA TO ANSWER BUSINESS QUESTIONS

o **The available data set is sufficient to answer the business question** concerning differences in usage between **annual members** and **occasional users.**

o **Detailed observation** of the data variables has determined that casual users generally pay for individual or daily trips, while member users tend to purchase annual passes.

o This information is **important for understanding the differences in behavior between the two groups,** and can be **used for targeted marketing campaigns**.

o Further analysis of other variables in the data, such as trip duration and location, could provide a **better understanding of the usage patterns** of annual members and casual users.

Winning Data Analytics Race with Dr Lebede Ngartera

Data difficulties

PHASE 2:
DATA PREPARATION

BIKE-SHARE

DATA DIFFICULTIES

The following difficulties were encountered during data analysis:

1) Data preparation identified **several problems**, including **duplicate records**, **missing fields**,... The large amount of **data(1.1 GB)** meant that we had to work with segments rather than trying to use the disk image functions.

2) **Data cleaning, deleting unnecessary variables** and **saving to a CSV file on the hard disk** enabled efficient data processing and analysis.

3) Specialized tools or techniques were required to work with large amounts of data.

4) Tools such as **Excel were not able to handle this amount** of data. R, SQL, and Tableau were used.

Winning Data Analytics Race with Dr Lebede Ngartera

Optimizing data analysis

PHASE 3: DATA PROCESSING

OPTIMIZING DATA ANALYSIS:

CHOOSING SQL AND RSTUDIO FOR EFFICIENT PROCESSING OF LARGE DATASETS

a) **To efficiently** prepare, process, clean, analyze and visualize the data for this project, we chose RStudio Posit Cloud as our main tool. Due to the large size of the dataset, it was not practical to use tools such as Microsoft Excel or Google Sheets to handle the volume of data. RStudio Posit cloud provided the capabilities needed to work effectively with the data and generate meaningful information.

b) Alongside RStudio Posit cloud, we also used SQL to explore our data, looking at total number of rows, distinct values, maximum, minimum or average values, and Tableau to create visualizations as part of this project. Tableau's powerful data visualization capabilities enabled us to effectively communicate the results of the analysis and highlight key trends and patterns in the data.

c) Altogether, the combination of RStudio Posit and Tableau proved to be a powerful toolbox for data preparation, processing, cleaning, analysis and visualization for this project.

Winning Data Analytics Race with Dr Lebede Ngartera

Data review

*PHASE 3:
DATA
PROCESSING*

DATA REVIEW

To better understand the data and its potential for analysis, **a review was carried out to assess variable content**, **data format and data integrity.** This initial review provided an overview of the data and identified any issues or challenges to be addressed during the preparation and analysis process. The data review consisted of

- **Verification of column names** in the 12 original files.

- **Checking** for **missing values**.

- Checking for **blank spaces**.

- **Checking duplicate records**.

- **Other data anomalies**.

Winning Data Analytics Race with Dr Lebede Ngartera

Results of the review

PHASE 3: DATA PROCESSING

RESULTS OF THE REVIEW

The results of the review found following things:

o **Duplicate** record of ID numbers.

o **Records** with missing start or end station name.

o **Records with very short or very long** ride duration.

o **Records for trips starting or ending** at an administrative station (repair or testing station).

All 12 files were combined into one data set after initial review was completed. The final data set consisted of **5709367 rows with 13 columns** of character and numeric data. This matched the number of records in all **12 monthly data files**.

Winning Data Analytics Race with Dr Lebede Ngartera

Setting up environment

PHASE 3: DATA PROCESSING

SITTING UP ENVIRONMENT


```r
153  ```{r}
154  #load packages
155  library(tidyverse)
156  library(lubridate)
157  library(janitor)
158  library(data.table)
159  library(readr)
160  library(psych)
161  library(hrbrthemes)
162  library(ggplot2)
163  library(mapview)
164  library(dplyr)
165  library(anytime)
166  library(png)
167  library(pander)
168  ```
```

The total number of records in all **12 monthly data files was calculated to be 5709367 rows and 13 columns(september_2022, october_2022, november_2022, december_2022, january_2023, february_2023, march_2022, april_2023, may_2023, june_2023, july_2023, august_2023).** This information provides an overview of the size and scope of the data, which can be helpful in planning and executing the analysis process.

Winning Data Analytics Race with Dr Lebede Ngartera

Comparison between

members and casual riders

*PHASE 4&5:
ANALYZE AND SHARE
THE DATA*

BIKE-SHARE

COMPARISON BETWEEN MEMBERS AND CASUAL RIDERS

```
277 ▾ ###Comparison between members and casual riders
278 ▾ #### Mean
279 ▾ ```{r}
280   aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = mean)
281 ▴ ```
```

Description: df [2 × 2]

all_trips_v2$member_casual <chr>	all_trips_v2$ride_length <dbl>
casual	28.98771
member	12.44616

2 rows

Pie chart

(casual member vs annual member)

PHASE 4&5: ANALYZE AND SHARE THE DATA

BIKE-SHARE

PIE CHART
(CASUAL MEMBER VS ANNUAL MEMBER)

member_casual
- casual
- member
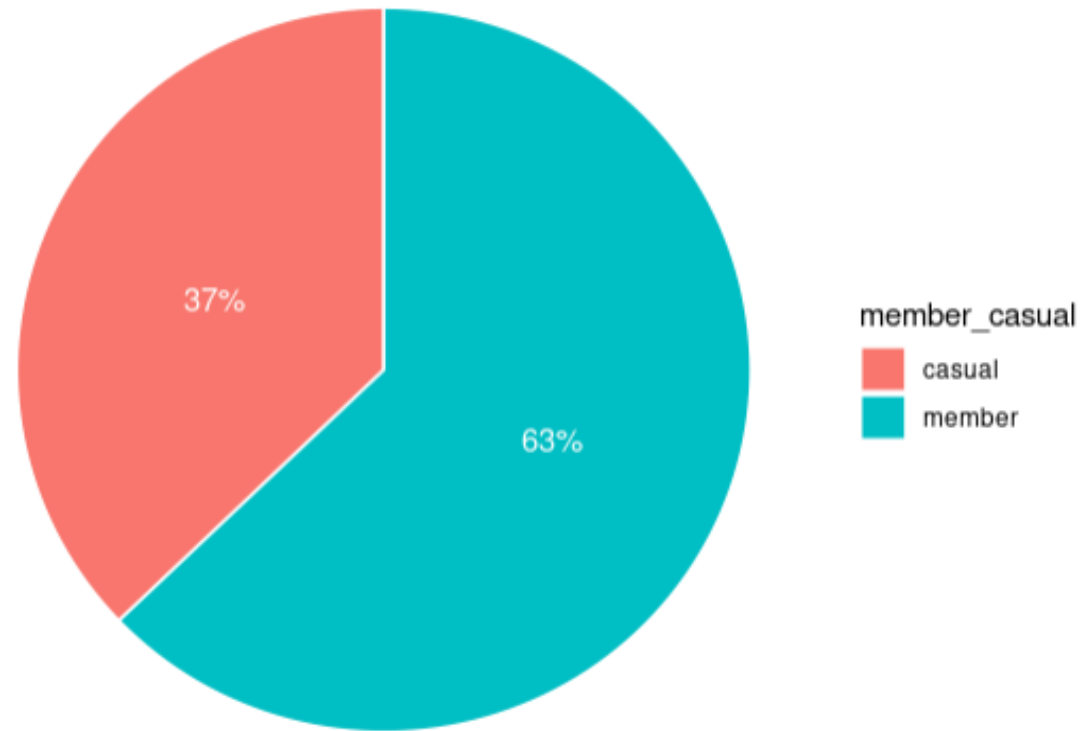
Looking at this pie chart, it shows that members take significantly more trips than casual riders.

Winning Data Analytics Race with Dr Lebede Ngartera

Customer Type vs. Day of the week

- Average

PHASE 4&5: ANALYZE AND SHARE THE DATA

BIKE-SHARE

CUSTOMER TYPE VS. DAY OF THE WEEK – AVERAGE

Average trip duration by customer type Vs. Day of the week
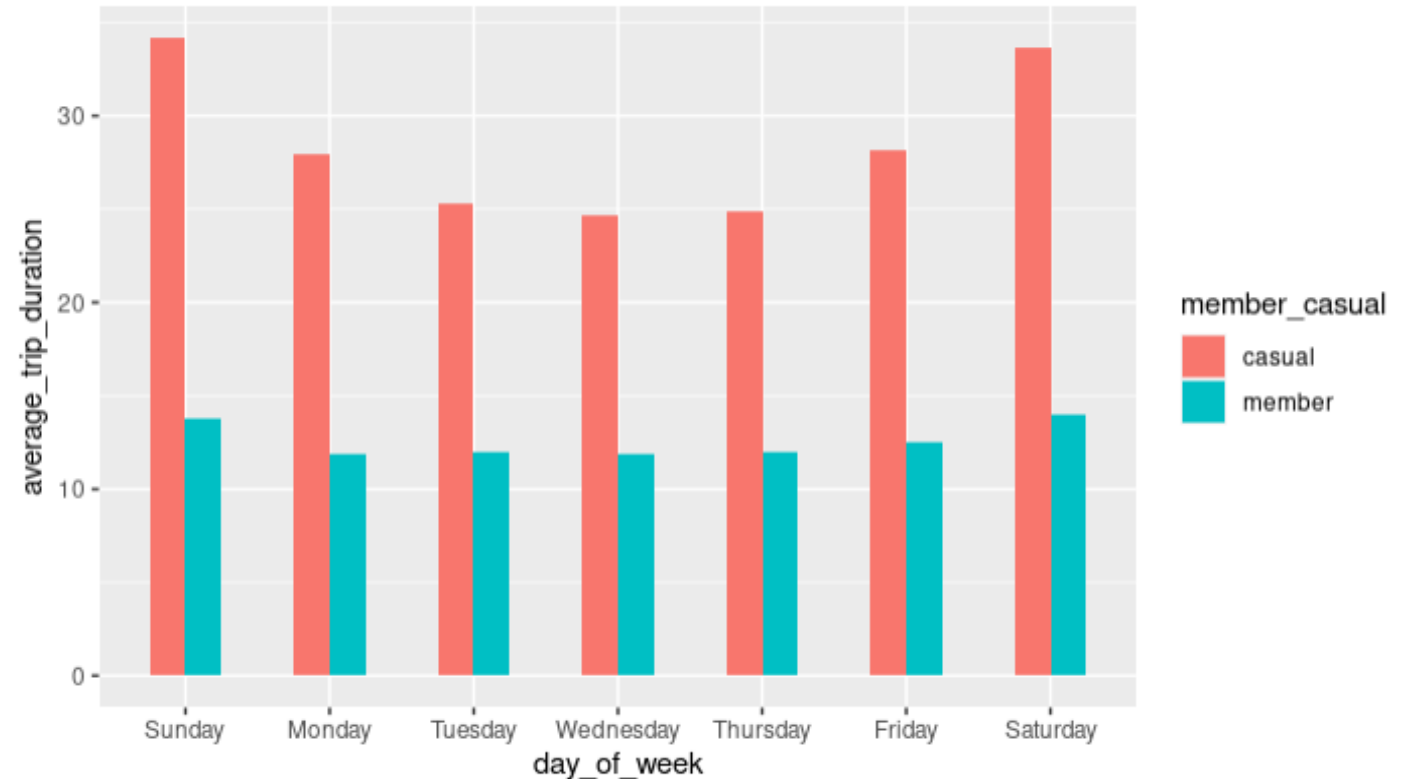
A member's typical ride lasts less than half as long as that of a casual rider. It's also interesting to see that while weekend rides tend to be longer for casual riders, regular riders tend to keep their average ride length fairly constant every day of the week. Please take note that this does not imply that casual riders necessarily cover greater distances.

Winning Data Analytics Race with Dr Lebede Ngartera

Customer Type vs. Day of the week

Count

PHASE 4&5:
ANALYZE AND SHARE
THE DATA

BIKE-SHARE

CUSTOMER TYPE VS. MONTH - COUNT

Total trips by customer type Vs. Month

Based on the data presented above, it appears that Saturdays and Sundays see the highest volume of casual riders, while weekdays are busier for members. This implies that the bikes may be primarily used for non-leisure purposes by the annual members.
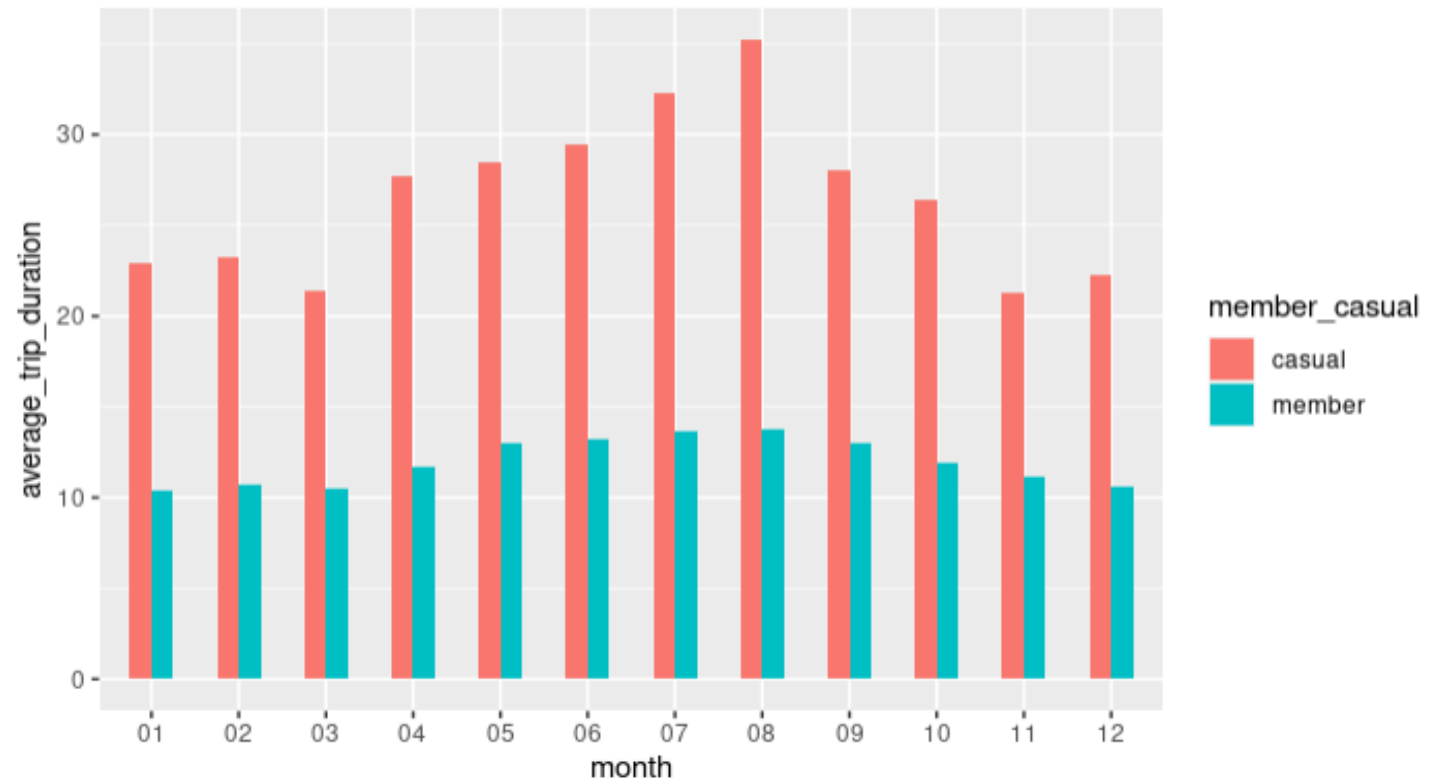
Winning Data Analytics Race with Dr Lebede Ngartera

Customer Type vs. Month

Average

PHASE 4&5: ANALYZE AND SHARE THE DATA

BIKE-SHARE

CUSTOMER TYPE VS. MONTH AVERAGE

Average trip duration by customer type Vs. Month

Over the course of a year, the typical ride time for a club member is between 10-15 minutes. There is a clear pattern when looking at casual riders, whose average trip duration varies from about 25 minutes at the end of the year to more than 35 minutes at the beginning of the year.

Winning Data Analytics Race with Dr Lebede Ngartera

Customer Type vs. Month

Count

PHASE 4&5: ANALYZE AND SHARE THE DATA

CUSTOMER TYPE VS. MONTH COUNT

Total trips by customer type Vs. Month
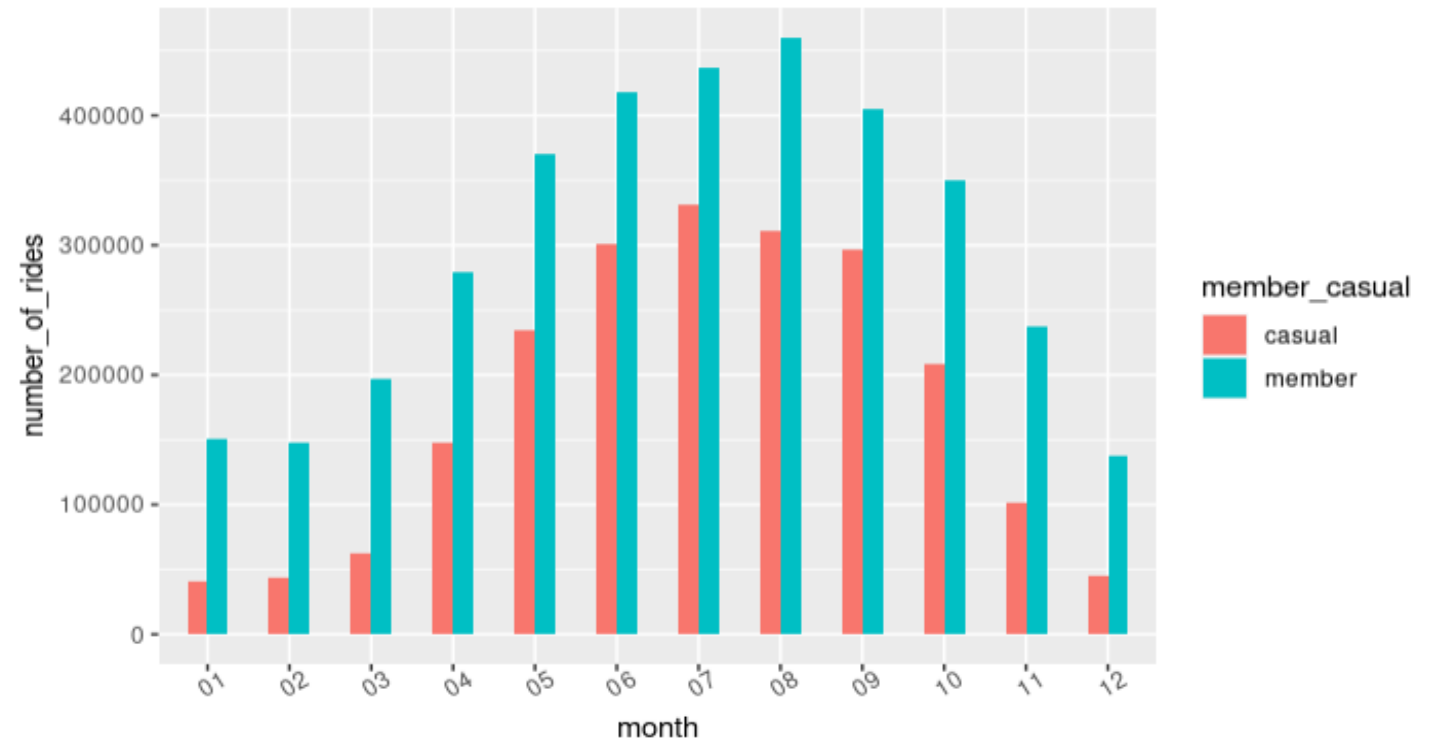
According to the graph, the busiest months of the year for members and casual riders are June, July, and August. This could be attributed to an external factor, like cold weather, that might have hindered with customer needs. Members, however, take significantly more rides than casual riders do during the rest of the year, including the winter months.

Rideable type

PHASE 4&5: ANALYZE AND SHARE THE DATA

RIDEABLE TYPE

Types of Bicycles Used
Period: 2022/09 - 2023/08
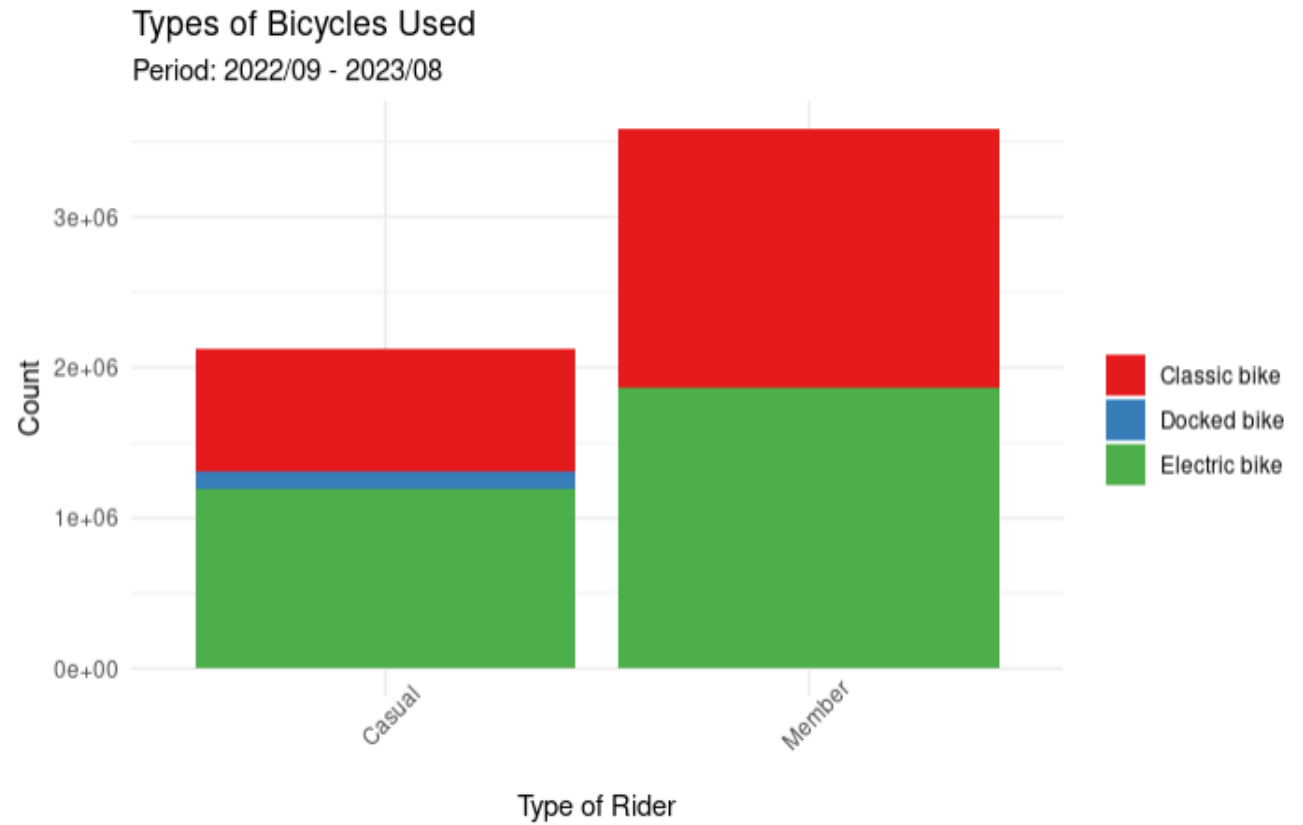
Observations:
1. Only casual riders use docked bikes
2. Both casual and annual riders use classic bikes more than electric bikes,
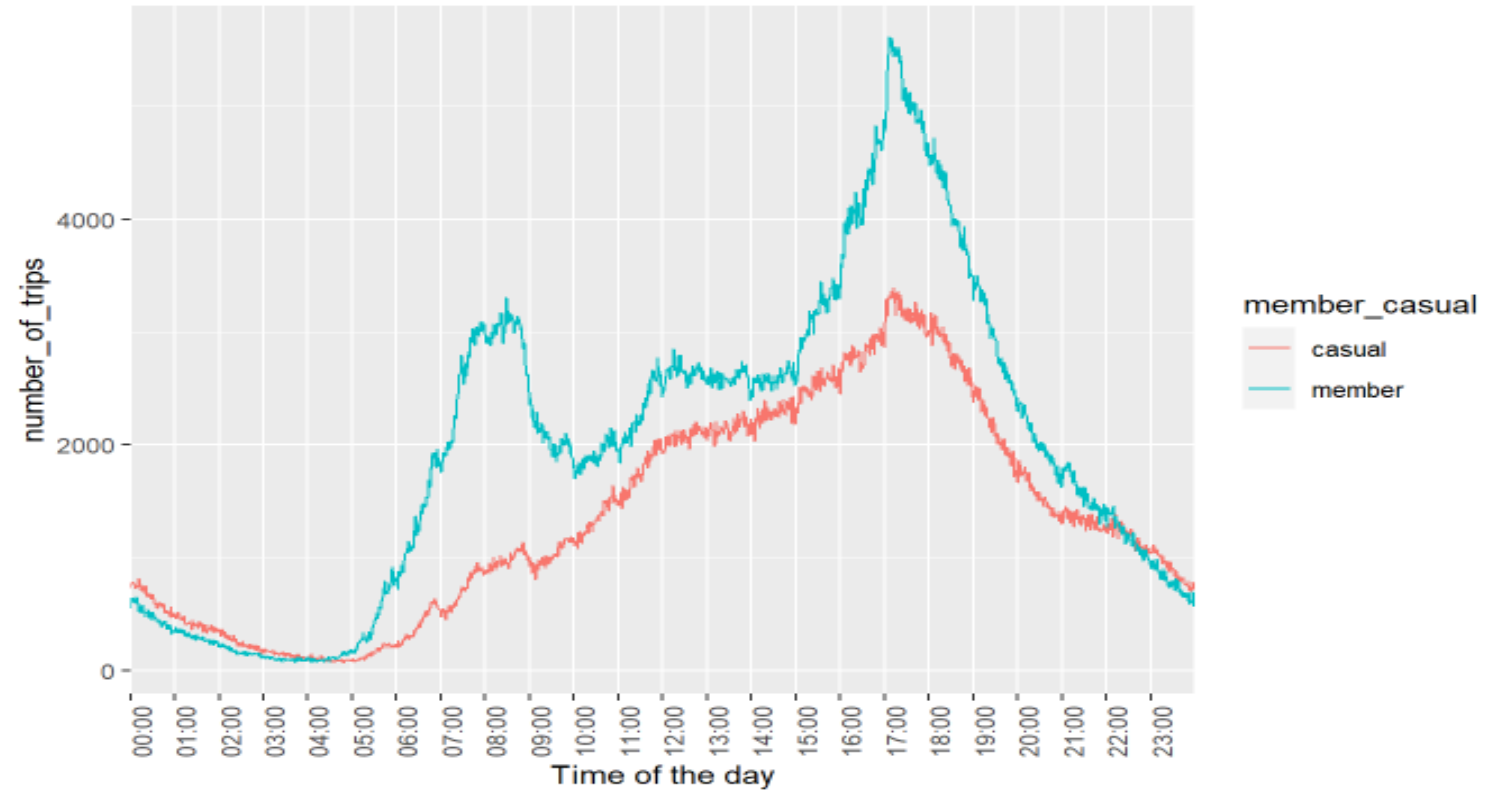3. although the difference is subtle within casual riders

Customer Type vs. Time

*PHASE 4&5:
ANALYZE AND SHARE
THE DATA*

BIKE-SHARE

CUSTOMER TYPE VS. TIME

Demand over 24 hours of a day

Members have two distinct peak demand periods: 7-9 a.m. and 5-6 p.m., the latter overlapping with the peak demand periods of occasional users. It is possible to assume that, given the high level of interest before and after office hours, the majority of members are commuters, but this hypothesis needs to be supported by more data.

Winning Data Analytics Race with Dr Lebede Ngartera

Key Takeaways

PHASE 6: ACT

KEY TAKEAWAYS

1. Casual riders made **37%** of total trips, while member riders make up **63%** during **September 2022 to August 2023**.

2. **Annual members take more trips**, while **casual riders take longer trips**.

3. **Casual customers use bikeshare services more during weekends**, while **members use them consistently over the entire week**. Both types of cyclists, but especially occasional cyclists, cycle much less during the winter months.

4. Average trip duration of casual riders is more than twice that of member riders over any given day of the week cumulatively.

5. Casual riders ride longer during the first half of the year compared to the second half, while members clock relatively similar average trip duration month over month.

6. Casual riders prefer docked bikes the most, while classic bikes are popular among members.

Winning Data Analytics Race with Dr Lebede Ngartera

Recommendations

PHASE 6: ACT

RECOMMENDATIONS

Here's an organized list of the recommendations :

a) **Offer seasonal passes for riders** who don't want to bike during the **winter**.

b) **Introduce a member-only rewards program** based on trip duration to incentivize casual riders to sign up as members and be eligible for the rewards.

c) **Offer discounted pricing during non-busy hours** so that casual riders might choose to use bikes more often and level out demand over the day.
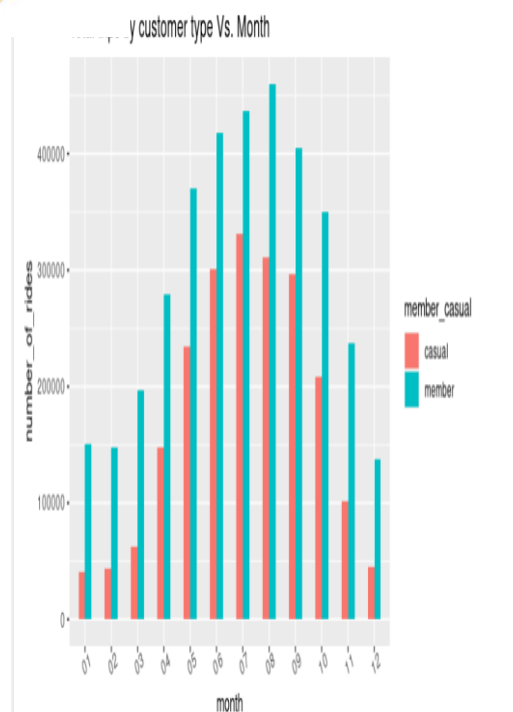
Winning Data Analytics Race with Dr Lebede Ngartera

Additional Data for Analysis

PHASE 6: ACT

ADDITIONAL DATA FOR ANALYSIS

1. **Occupation of member riders**: utilize this data to target non-members with similar occupations for membership promotion.

2. **Age and gender profile**: study the age and gender profiles to identify specific demographics that can be targeted for attracting new members.

3. **Students**: target students with tailored promotions or services, considering their specific needs and preferences.

4. **Pricing details for members and casual riders**: analyze pricing details to optimize the cost structure for casual riders or provide targeted discounts without compromising profit margins.

5. **Address/neighborhood details of members:** investigate if there are any location-specific parameters that encourage membership and use this information for targeted marketing or service improvements.

Winning Data Analytics Race with Dr Lebede Ngartera

DASHBOARD
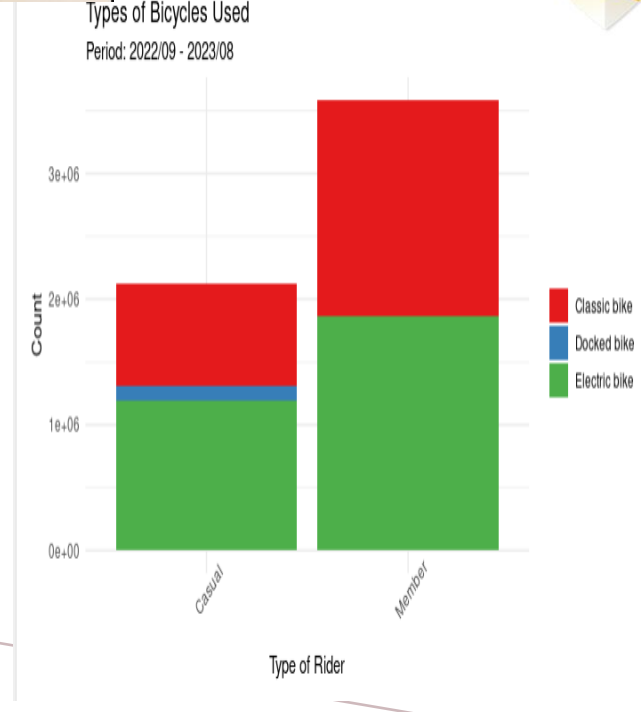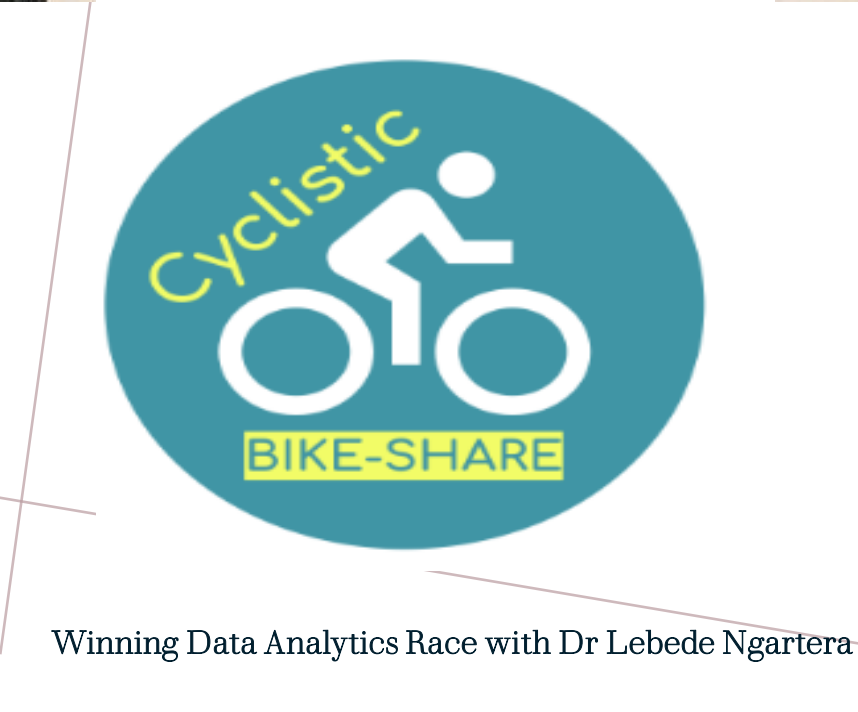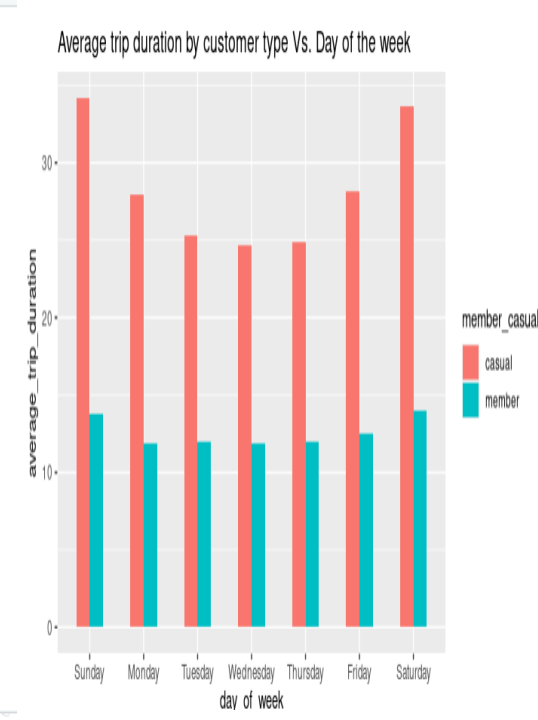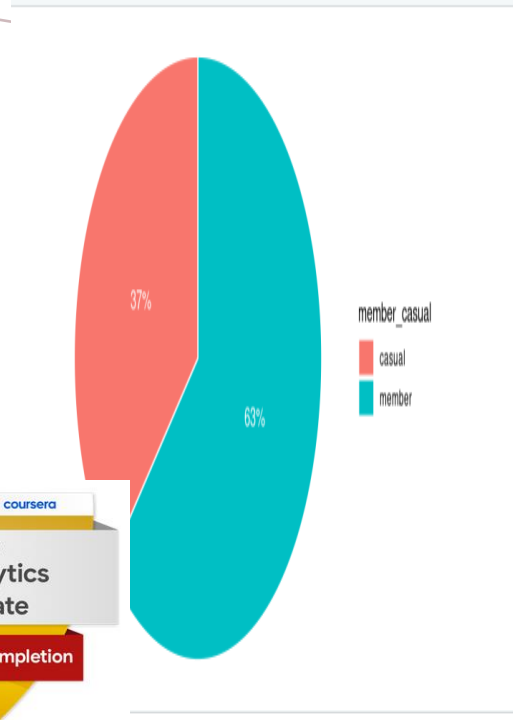
Winning Data Analytics Race with Dr Lebede Ngartera

WINNING DATA ANALYTICS RACE WITH GOOGLE DATA ANALYTIC CERTIFICATE

Dr Lebede Ngartera

THANK YOU

Presenter Name: Dr Lebede Ngartera

Email Address: ngarteralebede12@gmail.com

Website Address: https://github.com/Teraces12/Google-Data-Analytic-Certificate-How-a-Bike-Share-Navigates-Speedy-Success-