

Google Data Analytics Certificate

Case Study: How does a bike-share navigate speedy success?

Part 1. Introduction

Scenario

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

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. Most riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use the bikes to commute to work each day.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members.

Three questions will guide the future marketing program:

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?

I was assigned the first question: How do annual members and casual riders use Cyclistic bikes differently?

Part 2. Problem

Business task: What are the differences between annual members and casual riders in the use of Cyclistic bikes and how we can use this to convert casual riders to annual members?

Key stakeholders: Lily Moreno (my manager and director of marketing) and Cyclistics executive analytics team that will decide whether to approve the recommended marketing program.

Part 3. Solution

Data Sources

The data has been made available by Motivate International Inc under this [license](#). The data has collected by Lyft Bikes and Scooters, LLC that operates the City of Chicago's Divvy bicycle sharing service. For the next analysis, I downloaded data for the past 12 months of trip data (from 2024 June to 2025 May). I saved data using consistent file naming conventions. Additionally, I made a backup of data before I started to process data. Data has been provided as .csv files.

The Data Manipulation and Cleaning Process

Note: If you want to complete this Case study using spreadsheet, launch Excel application on your computer. Some spreadsheets have a size more than 100MB. Google Sheets can't manage this volume of data. After you launch Excel program go to the Data tab in the Toolbar. Select From Text/CSV icon in the Get Data menu in the left corner of the toolbar. Be sure that COMMA is selected as the delimiter and press load. It is very crucial. If you open the csv file directly you can break the data types and position of your data in your document.

During the Data Cleaning process, the next steps were completed:

1. Improving the View of spreadsheet: apply bold to column names, autofit column names, freeze first rows and columns
2. Searching for outliers (sort and filter each column)
3. Seeking for duplicates or blank fields (sort and filter each column, remove duplicates for ride_id)
4. Creating ride_length column using function =D2-C2. The results were formatted as HH:MM:SS (Format > Cells > Time > 37:30:55)
5. Creating a day_of_week column using formula =WEEKDAY(C2). Transforming Format of the day_of_week column cells to number with decimals equals 0 (zero)
6. Replacing the position of member_casual column ahead because we need this column during our analysis.

Analysis and Creating Summary File

The Analysis can be done using spreadsheet programs. But it is very hard to analyze more than 700,000 rows in spreadsheets. Additionally, my older laptop couldn't handle it. So, the analysis phase was performed using SQL (in my case using Google Bigquery). To analyze a file that has a size more than 100MB in Bigquery, first you need to upload it to Google Cloud.

To create a more suitable table for analysis (that contains only necessary columns) the following query was used. The **EXTRACT()** function with DAYOFWEEK convert the date object and return it as a day of week with Sunday as '1' and Monday as '7'.

```
SELECT
    ride_id,
    rideable_type,
    `started_at`,
    `ended_at`,
    (`ended_at` - `started_at`) AS ride_length,
    EXTRACT(DAYOFWEEK from `started_at`) AS day_of_week,
    member_casual
FROM
    arcane-dolphin-462114-v2.Case_Study.june_2024
```

Then with **MAX function** I calculate the maximum of ride length.

```
SELECT
    MAX(ride_length) AS max_ride_length
FROM
    arcane-dolphin-462114-v2.Case_Study.June
```

By adding the WHERE clause with member_casual = 'casual' for casual users and member_casual = 'member' for members I identified the maximum of ride length by user type. This element was used in all necessary situations throughout the analysis.

```
SELECT
    MAX(ride_length) AS max_ride_length
FROM
    arcane-dolphin-462114-v2.Case_Study.June
WHERE
    member_casual = 'casual'
```

Similarly, the average ride length was calculated using **AVG()** function.

```
SELECT
    AVG(ride_length) AS avg_ride_length
FROM
    arcane-dolphin-462114-v2.Case_Study.June
WHERE
    member_casual = 'casual'
```

To calculate **the number of rides per day of week** by user type I used the following query. The next query can also work without HAVING clause. But we can use HAVING clause to select all weekdays starting from specific day (for Monday HAVING > 1, for Tuesday HAVING > 2 and so on).

```
SELECT
    day_of_week,
    COUNT(ride_id) AS rides_per_day
FROM
    arcane-dolphin-462114-v2.Case_Study.June
WHERE
```

```

    member_casual = 'casual'
GROUP BY
    day_of_week
HAVING
    day_of_week > 0

```

Then I identified the **average ride length by day of week** with the help of the next query.

```

SELECT
    day_of_week,
    AVG(ride_length) AS avg_ride_length_per_day
FROM
    arcane-dolphin-462114-v2.Case_Study.June
WHERE
    member_casual = 'casual'
GROUP BY
    day_of_week

```

To identify **the total number of rides by user type** and **the total ride length by user type** the following queries were used.

Total number of rides by user type:

```

SELECT
    member_casual,
    COUNT(ride_id) AS number_of_rides
FROM
    arcane-dolphin-462114-v2.Case_Study.June
GROUP BY
    member_casual

```

Total ride_length per users:

```

SELECT
    member_casual,
    SUM(ride_length) AS total_ride_length
FROM
    arcane-dolphin-462114-v2.Case_Study.June
GROUP BY
    member_casual

```

Finally, I reviewed the **bike type preferences by user type**. For this purpose, the following query was used:

```

SELECT
    rideable_type,
    COUNT(rideable_type) AS number_of_bikes,
    member_casual
FROM
    arcane-dolphin-462114-v2.Case_Study.July
GROUP BY
    member_casual,
    rideable_type

```

I analysed the data for the certain month to identify trends for the specific month. We can merge data from different months using UNION ALL operator. Then we can conduct the analysis in the scale of the year. The following query merge data for three spring months – March, April and May.

```
SELECT
    ride_id,
    rideable_type,
    `started_at`,
    `ended_at`,
    (`ended_at` - `started_at`) AS ride_length,
    EXTRACT(DAYOFWEEK from `started_at`) AS day_of_week,
    member_casual
FROM
    arcane-dolphin-462114-v2.Case_Study.March
UNION ALL
SELECT
    ride_id,
    rideable_type,
    `started_at`,
    `ended_at`,
    (`ended_at` - `started_at`) AS ride_length,
    EXTRACT(DAYOFWEEK from `started_at`) AS day_of_week,
    member_casual
FROM
    arcane-dolphin-462114-v2.Case_Study.April
UNION ALL
SELECT
    ride_id,
    rideable_type,
    `started_at`,
    `ended_at`,
    (`ended_at` - `started_at`) AS ride_length,
    EXTRACT(DAYOFWEEK from `started_at`) AS day_of_week,
    member_casual
FROM
    arcane-dolphin-462114-v2.Case_Study.May
```

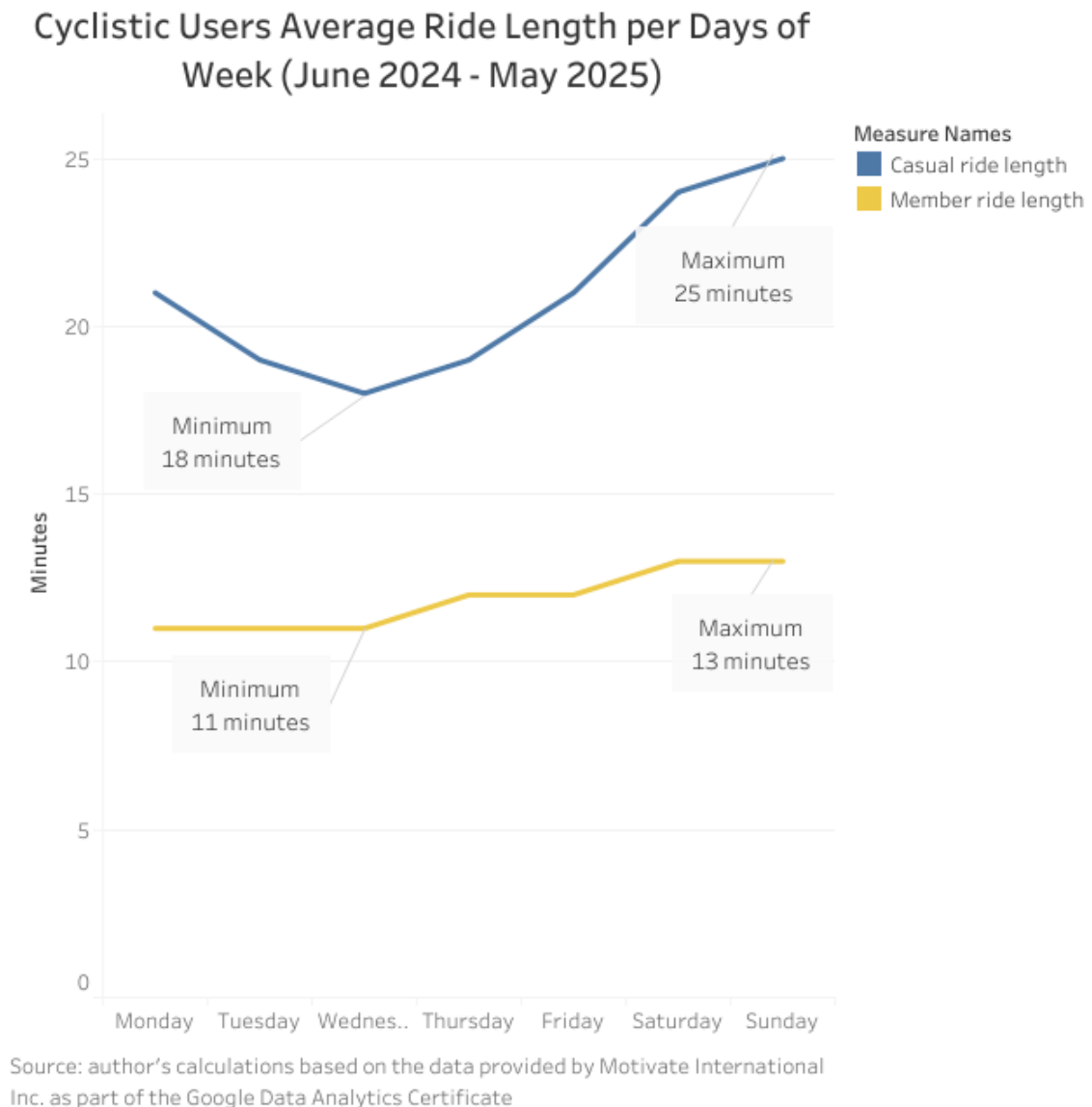
We can continue this merging process by adding another month data with UNION ALL operator. But I collect the most insightful data for each month in the Summary file. Then I used the Summary file to review figures for the year.

Part 4. Conclusions

Based on the data provided in the Summary file I identified some insights. To better demonstrate these insights, I created some data visualizations. The Tableau platform was used to create these data visualizations.

During our Google Data Analytics Certificate Program **Kevin Hartman** (one of our instructors and former Chief Analytics Strategist at Google) pointed our attention to one important moment. The executive team doesn't have too much time. In this situation the best option is to show your most important insights using one visualization. Our goal is to demonstrate the main difference

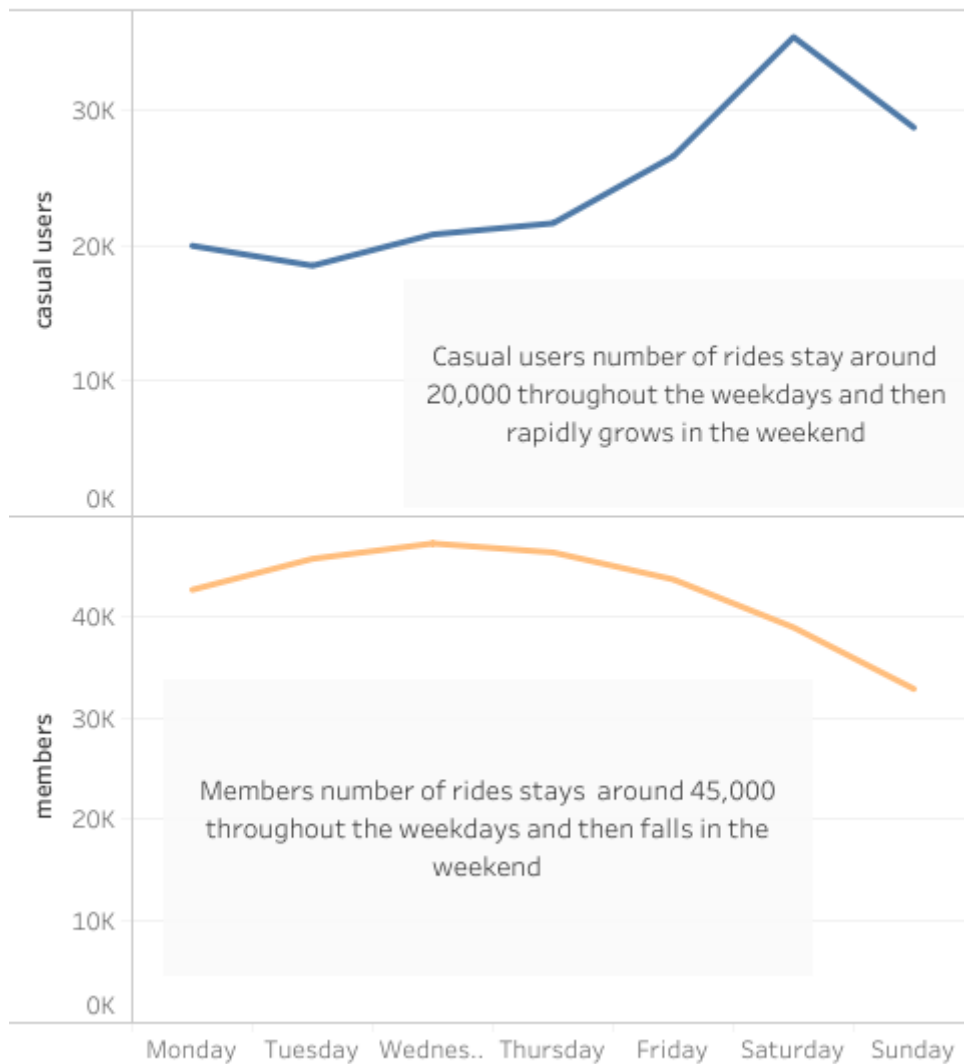
between Cyclistic casual users and members. In my opinion the best candidate for completing this task is the visualization that shows **Cyclistic users' average ride length per day of week** (look at the following graph).



According to this graph we can draw some conclusions. **First**, casual users' average ride length for each day of week is longer than members average ride length. Casual users' average ride length jumps from 18 to 25 minutes throughout the week. While members average ride length is located between 11-13 minutes. **Second**, members demonstrate more stable riding habits throughout the weekdays. While casual users' average ride length increases in the weekend. **Finally**, average ride length for casual riders equals about 22 minutes. This value for members is about 12 minutes and for overall users is 15 minutes (see more details in the Summary file).

After demonstrating the main insights, we can support our ideas using other visualizations. For example, to confirm our **thesis about users' riding preferences** we can analyze the next visualization. The following graph shows the number of rides per day of week for user type.

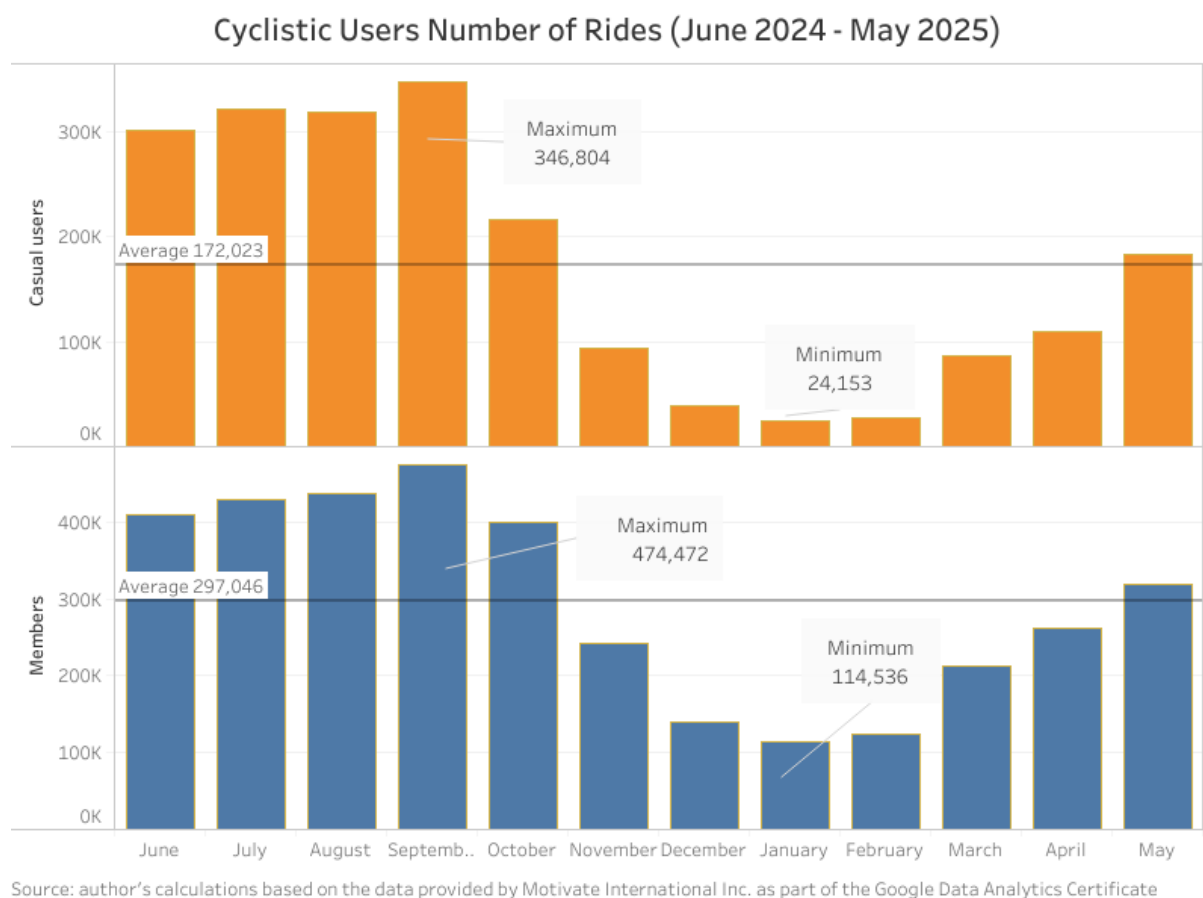
Cyclistic Users Number of Rides per Days of Week (June 2024 - May 2025)



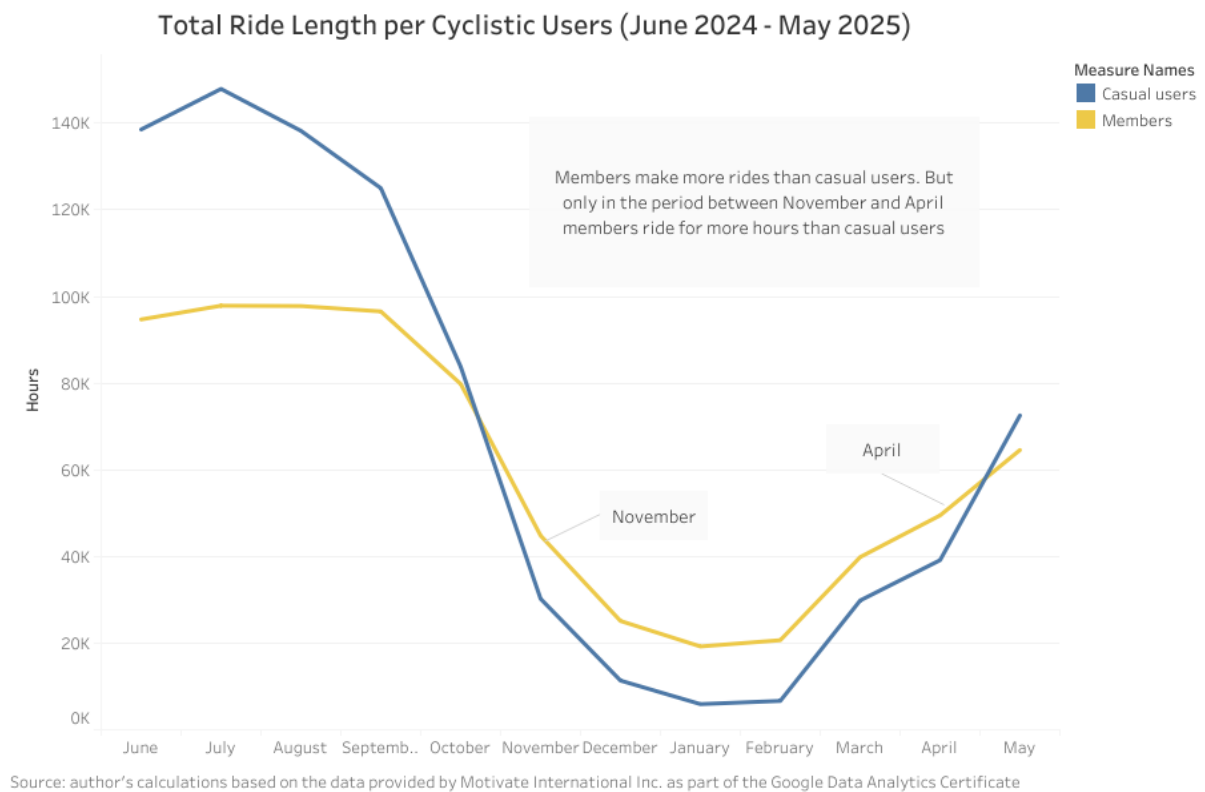
Source: author's calculations based on the data provided by Motivate International Inc. as part of the Google Data Analytics Certificate

According to this graph casual users' number of rides stay around 20 thousand most of the weekdays. Then on Friday this indicator rises to 26,6 thousand number of rides, hits the peak (35,5 thousand rides) on Saturday and goes down to 28,8 thousand on Sunday. While members' number of rides stay around 45 thousand throughout the weekdays and then falls in the weekend. The minimal number of rides for members is 32,9 thousand on Sunday. The difference between average and maximum number of rides per day of week for casual users is 77,5%, while for members it is only 26,9%. These figures can be explained by the fact that about 30% of Cyclistic users ride a bike to work (look at the Scenario, on page 1). It is very likely that a part of Cyclistic members use bikes to commute to work. Generally, people don't work on Sundays and in many on Saturdays too. So, this decrease in the number of member rides is obvious. The increase in the number of casual users' rides is also not surprising. The majority of Cyclistic users ride bikes to spend leisure time (again look at the Scenario, on page 1). Because of no work at the weekend there is more leisure time that can be used for other matters, such as cycling. Days with the fewest trips are Sunday and Monday. So, these days can be selected for carrying out technical maintenance.

The next difference between Cyclistic users is **the total number of rides by users**. The next bar graph shows the number of Cyclistic users' rides during the year. We can see that the number of Cyclistic users' rides changes throughout the year. The main reason for these changes is associated with the revolution of the seasons. The weather in Chicago from November to March is usually cold and snowy. This type of weather is not very suitable for bike riding. According to the graph we can conclude that the number of members' rides is greater than the number of casual users' rides. On average 297 thousand rides are made by members and 172 thousand by casual riders during the year. The number of rides exceeds average values in the period from May to October and does not achieve average values in the period from November to April. September is the month with the maximum number of rides and January is the month with the minimum number of rides for both groups (members and casual users).



The last graph shows **the total ride length** for members and casual riders throughout the year. During the period from May to October casual users ride more hours on a bike than members. This fact confirms again the importance of casual riders and their potential for future company growth. It is caused by the average ride length of casual users. The average ride length for casual users is almost twice as much as the average ride length for members (look at page 6). Due to this the total ride length for casual riders is more than the same indicator for members. Even though the total number of rides made by casual users is less than members' rides. Only in the period from November to April, the members' total ride length is greater than the casual users' total ride length. It is caused by the fact that the number of casual users' rides in this period is too small. And even a longer average ride length can't compensate for this drop in the number of rides.



During the analysis of **the maximum ride length**, we did not find any significant differences between casual users and members. Throughout the year the maximum ride length for both types of riders stays around 25 hours. Only in March the maximum ride length for casual users exceeds the same value for members by an hour (about 26 hours for casual users and 25 hours for members). During the rest of the year the maximum ride length for both types of users is almost identical (see more details in the Summary file).

There are no differences in **bike type preferences** for casual users and members. Both casual users and members gave a more preference to electric bikes rather than classic bikes (see more details in the Summary file).

Summing up all the above we can conclude that **the main differences between annual members and casual riders in the use of Cyclistic bikes** consist of the following:

1. Casual users' average ride length is longer than members average ride length. Casual users' average ride equals about 22 minutes. This value for members is about 12 minutes and for overall users is 15 minutes.
2. Casual users give a greater preference to bike rides on the weekend. While members demonstrate more stable riding habits throughout the weekdays. In the weekend both the average ride length and the number of rides for casual riders significantly go up. While the changes for members are less dramatic.
3. The total number of rides made by casual users is less than the same value for members. On average during the year casual users make around 172 thousand casual rides per month. While members make around 297 thousand rides for the same period. But these values change depending on the weather conditions.

5. Next Steps

I completed the task given to me by my manager Lily Moreno. Three questions will guide the future marketing program:

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?

I was assigned the first question: How do annual members and casual riders use Cyclistic bikes differently?

I think I found all main differences that exist between Cyclistic casual users and members based on the data provided to me in this Case Study. I think that **the main indicator that can be useful during the future marketing campaign is the difference between casual users' and members' average ride length**. By showing how many dollars on each trip and overall casual users can save we can convince them to buy annual membership.

But deep down I feel that my curiosity is unsatisfied. I think **if I had more data I could give more recommendations about the organization of the future marketing campaign**. For example, if I was provided by data **rider_id**, that was individual for each Cyclistic user, I can give more insights. In current conditions we can only identify the maximum or the average ride length, the total number of rides and the number of rides per day of the week, the average ride length per day of the week. But with **rider_id** we can give more individual information. If we have individual for each user **rider_id** we can calculate the total number of Cyclistic users. Additionally, we can say how many of them are members and casual riders. We can use **COUNT DISTINCT** for this purpose:

SELECT

member_casual,

COUNT (DISTINCT rider_id) AS number_of_users

FROM

Table name

GROUP BY

member_casual

If we had data about the number of users, we can calculate on average how many trips make each user. And then we can identify the total length of average user's rides for week or month by multiplying the average number of rides by the average trip length (22 minutes for casual users and 12 minutes for members).

Without these figures my analysis result seems to be abstract for me. If we have an approximate price of bike-share services, it can help us to give more realistic insights. For this task we can check the prices for [Divvy](#) bike share that operates in Chicago. Divvy suggests single rides by price 1\$ unlocking fee + 0.18\$ for each minute of ride. While annual membership costs around 144\$.

Now we can use these figures to demonstrate our approach to the new marketing campaign. As a result of our analysis, it is known that casual riders' average ride length is 22 minutes. So, if we use the real prices the average casual user trip costs around 5\$ ($1\$ + (22 * 0.18) = 4.96$). In this situation if a casual user makes 30 average rides it will cost 150\$ (more than annual membership price). Using these figures we can start a potential marketing campaign.

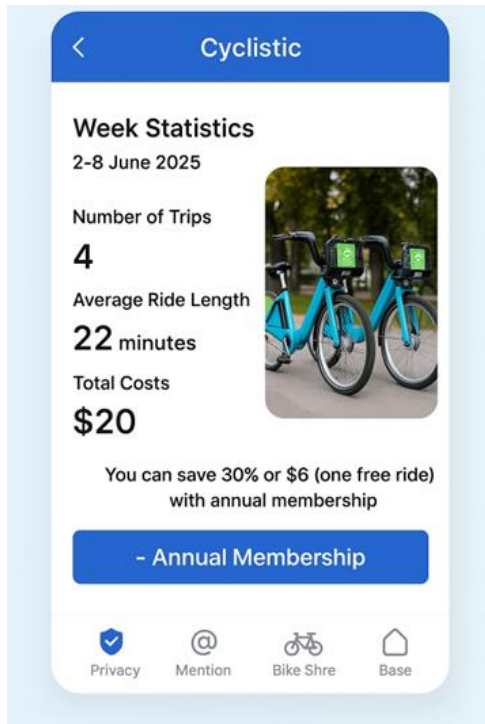
I think that **the best digital media type for deploying the future marketing campaign is Cyclistic mobile application**. There are many types of [digital media](#). But the mobile application can let us use a more individual approach. For example, at the end of each week we can provide individual statistics for each user that include the number of rides, the average ride length, how many dollars it costs and how many dollars the user can save by using an annual membership.

During the time I studied this Data Analytics Program, I took another course. This course is "AI Essentials" developed by Google. In this course I have learned basic knowledge and developed basic skills in using AI. So, it will be a sin not to use my AI skills to realize my ideas in practice. I gave the AI tool the **following prompt**:

Create a page of mobile app

1. You can use this image as an example (I gave an image as a reference)
2. Mobile app language is English
3. Our company is a bike share company
4. The company's name is "Cyclistic"
5. The information is
 - "Week statistics
 - 2-8 June 2025
 - Number of trips - 4
 - Average ride length- 22 minutes
 - Total costs- 20\$
 - You can save 30% or 6\$ (one free ride) with annual membership"

This is **the result**:



I think that mobile app notifications are the best way to influence casual users. Because **users buy subscriptions on an individual basis. And we need to approach them individually.** For example, if Netflix analysts calculated that an average user without a subscription loses 40\$ each year. Does this information influence me? I don't think. I am not a fan of watching movies. So, I don't watch too many films to justify the purchase of Netflix subscription. So, we need to work individually with each user. The mobile app of Cyclistic can generate individual statistics for each user based on their account and suggests individual offers.

Additionally to [motivate users to buy](#) annual subscriptions Cyclistic can suggest a discount on the first purchase of annual membership or let pay for a membership in installments. I bought my first Coursera annual subscription in installments. I made this in my bank mobile app. But if Coursera suggested this method, I made this purchase much earlier. Some people can't or don't like to pay a large sum at once.

Note: I guarantee that this Case Study will be completed by me (**Rufat Aghagasimov**) without using AI tool. I only use AI to demonstrate my skills to use it and only for creating content of the image demonstrated on the this page (page 12).