# Cyclistic Case Study

In this case study, I analyze data from the last 12 months from a Chicago based rental bike company in order to identify how their customers use their services. The main tools I used in this case study are Excel, SQLite and Tableau for the data visualization.

## Scenario

A Chicago based bike rental company asked its analytics team to analyze data from the past 12 months and find differences between its two types of customers, Members and casual riders. Members are customers that purchase an annual membership while casual riders only buy single ride tickets or daily passes. Cyclist financial analyst have concluded that members are more profitable than casual riders, although the pricing flexibility helps Cyclistic attract more customers the director of Marketing believes that maximizing annual members is the key for future growth.

The marketing analytics team wants to understand how casual riders and members use the company services differently. The primary stakeholders for this project is the director of Marketing and the Cyclistic executive team.

## Business task

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?
For this specific case study, I was assigned to answer the first task. *How do annual members and casual riders use Cyclistic bikes differently?*

## Data Source

I will be using Cyclistic's historical data to analyze and identify trends from July 2022 and June 2023 which can be downloaded from here. The data have been made available by Motivate International Inc. under this license.

This is public data that you can use to explore how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit you from using riders' personally identifiable information. This means that you won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

## Data Organization

There are 12 files here each one for the appropriate month of the year we are going to be analyzing. Each file contains the same 13 columns named ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng and member_casual.

## Data manipulation and data cleaning

After downloading and inspecting each of the 12 datasets I created some new columns in Excel. These columns were ride_length which is a column that subtracts the time a ride ended by the time a ride started. I also created separate columns for year, month, day, start_time, end_time, date and day_of_week, data that all were already given in the dataset I just extracted them in separate columns so its easier to use them in my analysis.

After creating all the appropriate columns, I saved my work as the appropriate month name that the data represented. After saving my new datasets I decided SQLite was the best tool available for me to continue this project.

The first thing I did is I combined the monthly tables into appropriate quarters, named q3_2022, q4_2022, q1_2023, q2_2023 and then combined the 4 quarters into a table named past_12_months.

```
1   CREATE TABLE q4_2022 AS
2   SELECT * FROM october
3   UNION
4   SELECT * FROM november
5   UNION
6   SELECT * FROM december;
```

```
1   CREATE TABLE q1_2023 AS
2   SELECT * FROM january
3   UNION
4   SELECT * FROM february
5   UNION
6   SELECT * FROM march;
```

```
1   CREATE TABLE q3_2022 AS
2   SELECT * FROM july
3   UNION
4   SELECT * FROM august
5   UNION
6   SELECT * FROM september;
```

```
CREATE TABLE q2_2023
SELECT * FROM april
UNION
SELECT * FROM may
UNION
SELECT * FROM june;
```

```
1   CREATE TABLE past_12_months AS
2   SELECT * FROM q3_2022
3   UNION
4   SELECT * FROM q4_2022
5   UNION
6   SELECT * FROM q1_2023
7   UNION
8   SELECT * FROM q2_2023;
```

## Data Analysis

After creating new quarter tables and combining whole of my data into one table I started using queries in order to get tables relevant to my analysis so I can later visualize them with Tableau.

Firstly, I queried a table that counts the number of total trips on all quarters, and the past 12 months table in order to see the number of rides made by each type of customer.

```
1   SELECT
2       count(*) AS total_trips,
3       (SELECT COUNT (member_casual) FROM q3_2022 WHERE member_casual='member') AS total_trips_member,
4       (SELECT COUNT (member_casual) FROM q3_2022 WHERE member_casual='casual') AS total_trips_casual
5   FROM q3_2022;
```

| | total_trips | total_trips_member | total_trips_casual |
|---|---|---|---|
| 1 | 2310759 | 1249083 | 1061676 |

Then I used the day_of_week column I created in order to see what day of the week most of the customers prefer to use the rental bikes.

```
1   select day_of_week, count(ride_id)
2   from past_12_months
3   GROUP BY day_of_week
4   ORDER BY count(ride_id) DESC;
```
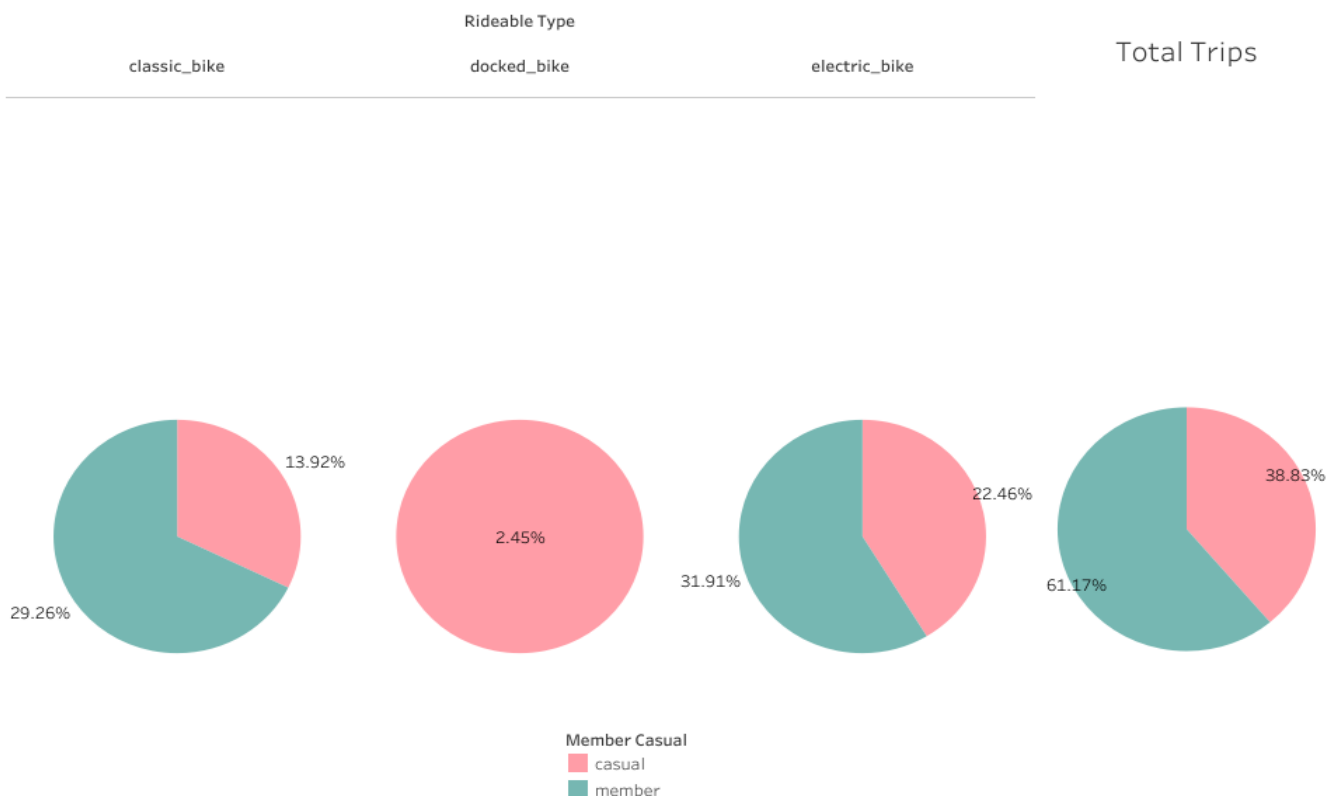
|   | day_of_week | count(ride_id) |
|---|-------------|----------------|
| 1 | Saturday    | 924388 |
| 2 | Friday      | 865994 |
| 3 | Thursday    | 863962 |
| 4 | Wednesday   | 848891 |
| 5 | Tuesday     | 806146 |
| 6 | Sunday      | 739351 |
| 7 | Monday      | 730712 |

## Data Visualization

After creating the appropriate csv files I needed for my visualization I went into Tableau Public loaded them in and started the visualization process. During the data visualization process I used the charts in order to answer the question I was assigned.

First and foremost, I wanted to see the number of each customer and what type of bike they used. For that visualization I chose I pie chart with all 3 possible bike types.

Ride type per customer type

Rideable Type

| classic_bike | docked_bike | electric_bike | Total Trips |



13.92%

2.45%

22.46%

38.83%

29.26%

31.91%

61.17%

Member Casual
casual
member

The members make 61% of the customer base with half of them preferring classic bikes and the other half preferring electric bikes. While on the other hand casual riders make up 39% of the customer base and most of them tend to prefer the electric bike instead of the classic bike.

Next the number of trips per month, and trips per day are visualized for each member type.
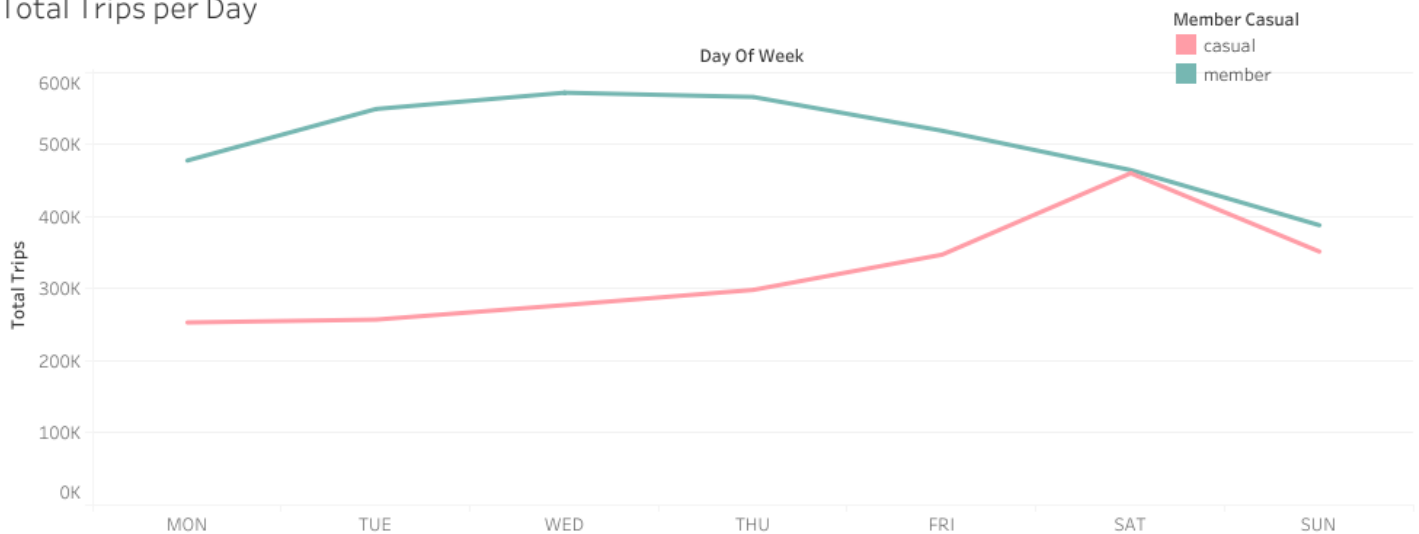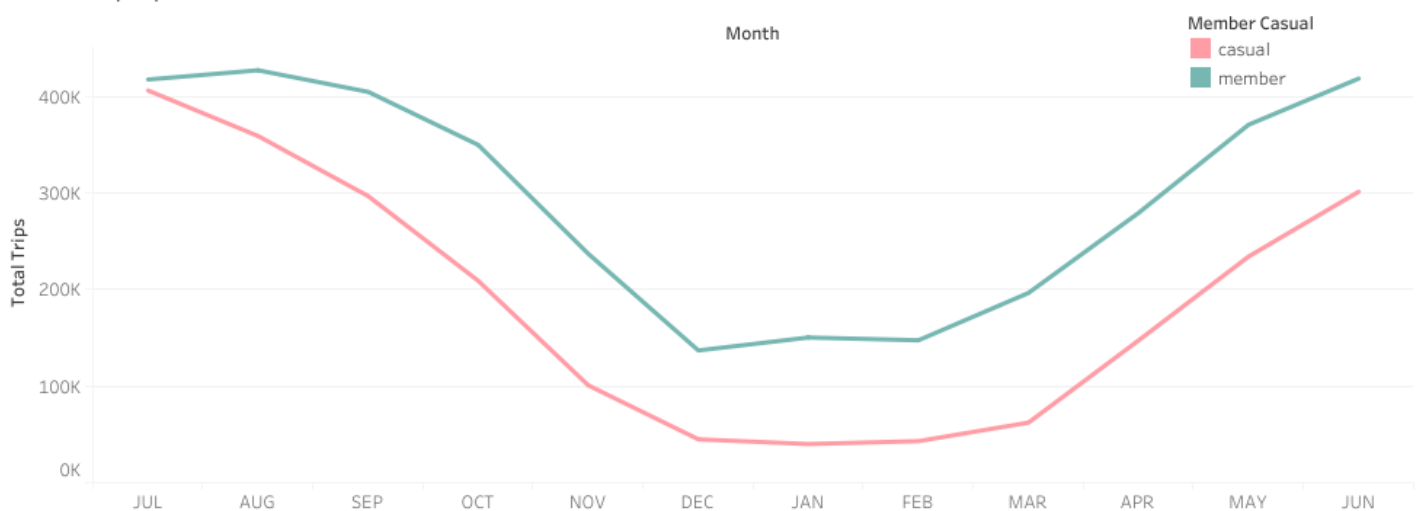


From this visualization as far as days of the week are concerned, we can see that members prefer to rent bikes during the week and we observe a slight decrease during the weekend while on the other hand most casual members tend to use the bikes during the weekend. And now for the monthly usage we can see that both set of customers prefer to use our bikes during the spring and summer months more than any other season of the year.

From the first visualization we can conclude that most member use our services possible for commute to and from work/class during the weekdays, while members use our services during their leisure time on a weekend. And as far as the months we can see that when the weather is probably warmer that more customers tend to use our bikes than when the weather is colder during the winter.

For our final visualization we will compare the average ride duration of each of the two customer types.

## Avg Ride Per Month
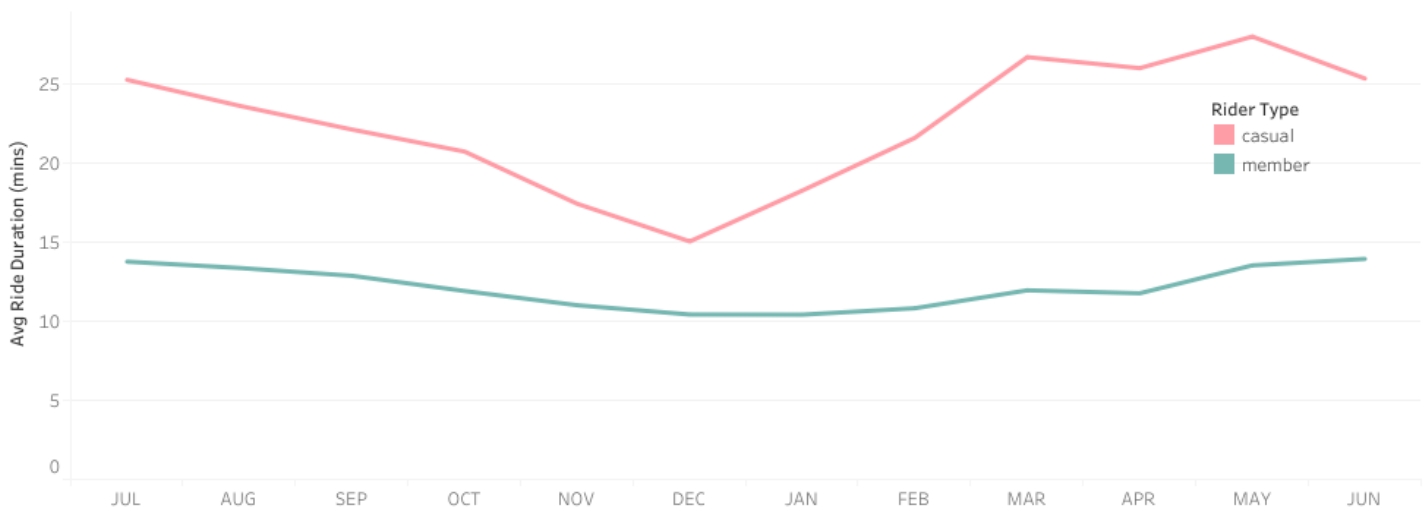


## Avg Ride per Day of Week
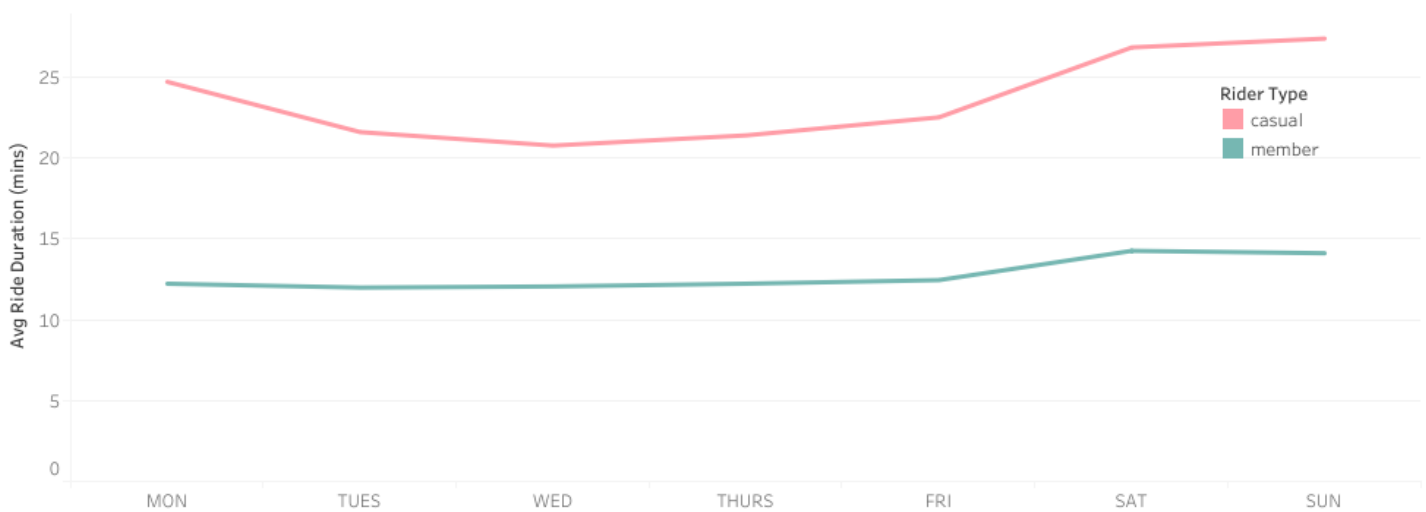


As we can see casual members tend to have way bigger duration timers every month of the year with a slight decrease during the 4th quarter of the year while members tend to have almost the same average duration for the whole entire year without a massive drop in ride duration.

And now for the duration during each day we can see that during the week members tend to have the same ride duration every day but there's a slight increase during the weekend and for casual riders we can see that Saturday and Sunday have the longest rides.

From that dashboard we can conclude that members indeed have the same ride duration every day, as in commute from and to work, while casual riders tend to have different ride durations each day.

## Summary

To summarize from the visualizations, we discovered that casual riders travel almost twice as long as the members but in a much smaller frequency than the annual members. We can also conclude that they make longer journeys during the weekend and the spring and summer months which we can deduce that it might be for recreational reasons instead of daily commute to and from work/class etc.

# Recommendations

After discovering the differences between annual members and casual riders there might be some marketing strategies that could be used to persuade casual riders to become annual members.

1. A marketing campaign should be created just before spring in which a discounted annual membership could be offered for the "busy" season.
2. Since casual riders tend to use the bikes during weekends a special annual membership that only applies for weekends could be offered.
3. Casual riders tend to use the bikes longer than annual members so an annual membership with some kind of discount or a duration incentive bonus could be offered which is something that might even push annual members to use the service for longer.