Cyclistic Bike-Share Report

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are tracked and locked into a network of stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system at any time.

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**.

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, the Marketing Director 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, she believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Business Goal

The director of marketing has set a clear goal:
Design marketing strategies aimed at converting casual riders into annual members.

To do that, however, the marketing analyst team needs to better understand:

- How annual members and casual riders differ.
- Why casual riders would buy a membership.
- How digital media could affect their marketing tactics.
- The Cyclistic historical bike trip data to identify trends.

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?

The Marketing Director has assigned me the first question to answer:

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

Data Sources

- I collected the Cyclistic Trip data of **2022**, which is 12 datasets, where each contains trip data for each month over the year. Each of them contains 13 columns (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, member_casual).

- For each month's dataset, the records are above 100,000, with a total of more than 5.5 million records.
- I empathized that there is no personally identifiable information, and it is relevant to the task at hand.

Data Processing & Cleaning

I used in this stage Excel to perform the cleaning process for the 12 datasets, and I performed the following steps:

1. I made a copy of the raw CSV files and placed them in a separate folder from the folder where I worked on the XLSX files.
2. Renamed the column names to more simple names (Ride ID, Bike Type, Start Time, End time, Start Station Name, Start Station ID, End Station Name, End Station ID, Start Latitude, Start Longitude, End Latitude, End Longitude, and Client Type).
3. Checked for the blanks through data and found blanks in the start and end stations, I handled this issue by replacing them with the value "N/A" and did the same for the start and end latitude and longitude.
4. Created a column to calculate the Ride duration in minutes by subtracting the end time from the start time, which enabled me to catch the data errors of negative numbers.
5. Formatted the columns data and ensured consistency over the 12 datasets.

Data Analysis stage

1. Loaded the whole folder into Power BI, to be treated as one dataset and called it "Cyclistic Bike-Share 2022" to perform calculations over the whole year as one dataset.
2. Created a "Calendar lookup" table and connected it with the "Cyclistic Bike-Share 2022" data table through the "Start Date" field that is extracted from the "Start Time" field with the "Date" field.
3. Removed latitude and longitude columns as we are operating in the same city and can't be used to create a map visual.
4. Added a new column "Day of Week", assuming that the week starts on Monday in both tables.
5. Added a column called "Weekdays" to show the weekdays (1~5) and weekends (6~7), which will be used further to show their frequency with clients and bikes.
6. Added two calculated columns, one to show the Ride Duration in minutes "Ride Duration (Min), and the other to show the Ride Duration in hours "Ride Duration (Hrs).
7. Created the following metrics to be used in the data visualization and to figure out the business task later:
- Total Rides = 5.5 M
- Total Duration (Min) = 108.2 M & (Hours) = 1.8 M
- Average Duration (Min) = 20 minutes
- Max Duration (Hours) = 690 hours
- Total member rides and categorizing it into the three Bike types and Weekdays
- Total Casual rides and categorizing it into the three Bike types and Weekdays
- Total Member Duration (Min) and segmenting it with the three Bike types and Weekdays
- Total Casual Duration (Min) and segmenting it with the three Bike types and Weekdays

Client Type		Total Rides	Total Duration (Min)	Average Duration (Min)	Max Duration (Hours)
⊟ casual		2,244,619	66,570,713	30	690
	Weekday	1,407,396	38,275,692	27	597
	Weekend	837,223	28,295,021	34	690
⊟ member		3,278,036	41,653,804	13	26
	Weekday	2,465,229	30,200,205	12	25
	Weekend	812,807	11,453,599	14	26
Total		5,522,655	108,224,517	20	690

Data visualization

 I asked myself 3 important questions:

1. What Type of Data am I working with?
2. What do I want to communicate?
3. Who is the End User and what do they need?

Sticking with those 3 questions in mind helped me to dig through data to target what I needed and present them properly.

- I am working with categorical and Time-series data
- I want to communicate data by showing some comparisons and compositions through visuals like bar charts, line charts, and Donut charts.
- This will help the Marketing Director get what he aims for to help him figure out how annual members and casual riders use Cyclistic bikes differently.

I used Power BI to visualize my data findings and used the following visuals to present them:

1. Cards that show high-level information such as total rides, total duration, Average and max duration.
2. Bar charts to compare the total rides and total duration for member and casual clients.
3. Bar charts to compare the total rides and total duration for the usage of the 3 bike categories.
4. Donut charts to show how clients use the 3 bikes differently.
5. Donut charts to show how clients use our bikes through weekdays and weekends and used as a **custom tooltip** that pops up when you hover over the bar charts related to clients and bike types.
6. Area chart to show the total rides and total duration over the year to depict where it peaks & and slopes downward, and see if there is a correlation between the overall rides and ride duration.
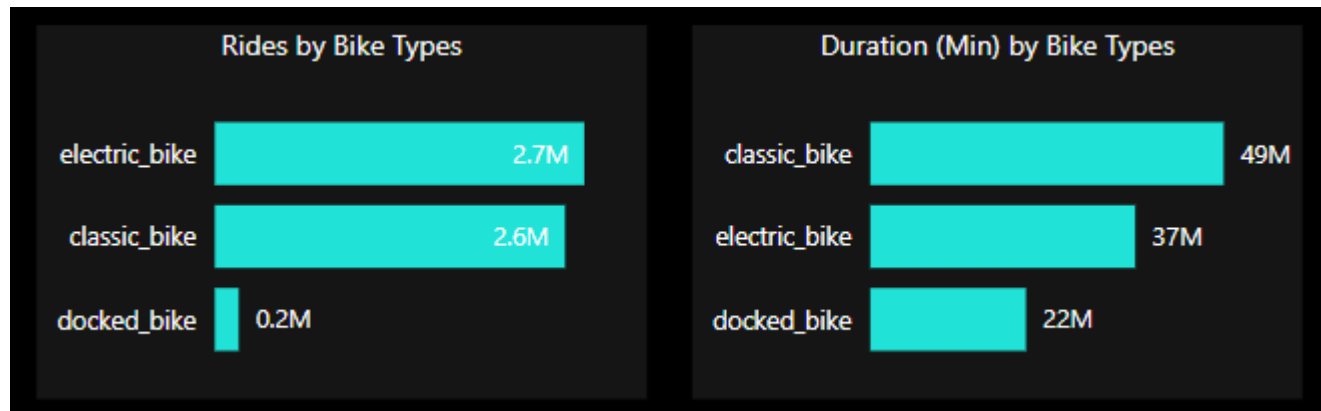
Data Findings & Insights

- We have 5.5 million rides, with 108.2 minute duration, and an average of 20 minutes per ride.
- The maximum ride reached up to 690 hours and that was for a casual client.

- I found that member clients have total rides more than casual clients, although casual clients have longer ride duration than member clients.
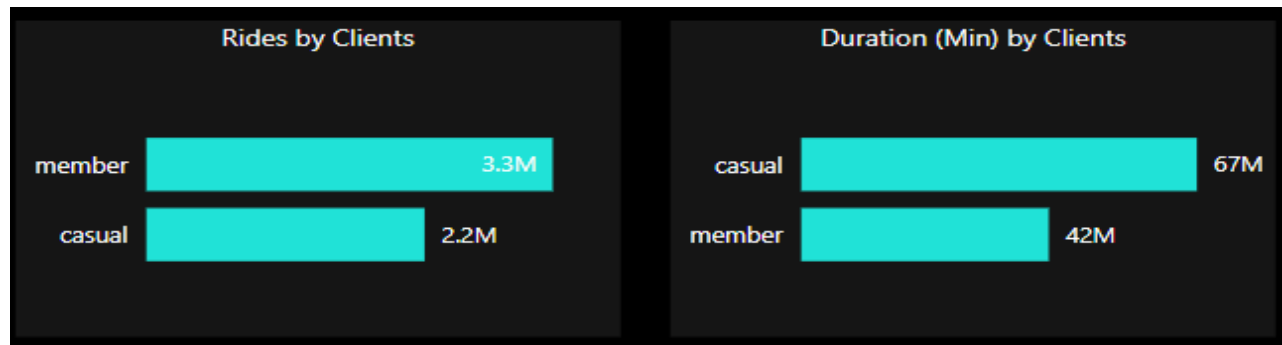

- Clients use bikes on weekdays more often than on weekends, in both total rides and time duration, which means we need to make sure that we have a proper number of bikes on weekdays, so they find them easily.

- The reason they use bikes on weekdays might be to go to school, work, etc.

- We may need to promote an offer to use our bikes on weekends.

- I embedded it as a custom tooltip to get more detailed information on this with different types of clients and bikes.
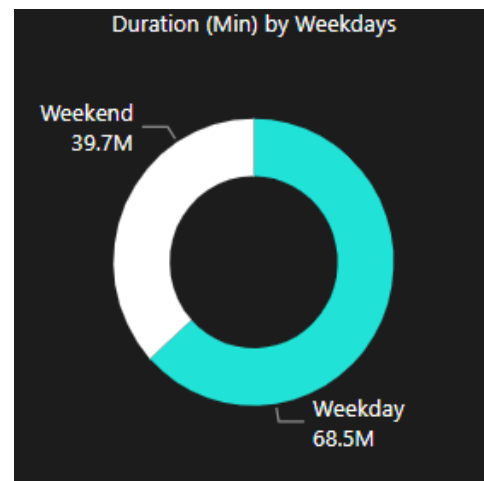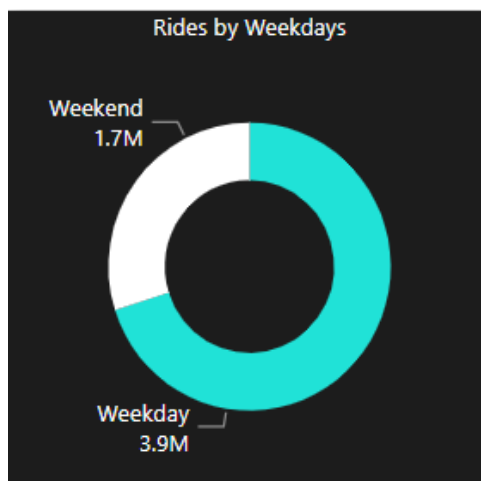
- For the overall rides, the most used bike type is the electric with a little more usage than the classic bike.
- The least used bike is the docked bike with a huge difference from the other 2 categories.
- For the overall duration, the most used bike is the classic followed by the electric followed by the docked bike.
- All bike categories are preferred to be used on weekdays than on weekends.

Rides by Bike Types		Duration (Min) by Bike Types	
electric_bike	2.7M	classic_bike	49M
classic_bike	2.6M	electric_bike	37M
docked_bike	0.2M	docked_bike	22M

- Member clients prefer to use only the classic and electric bikes respectively.

Member Rides per Bike

electric_bike
1.6M

classic_bike
1.7M

Member Duration per Bike

electric_bike
17.9M
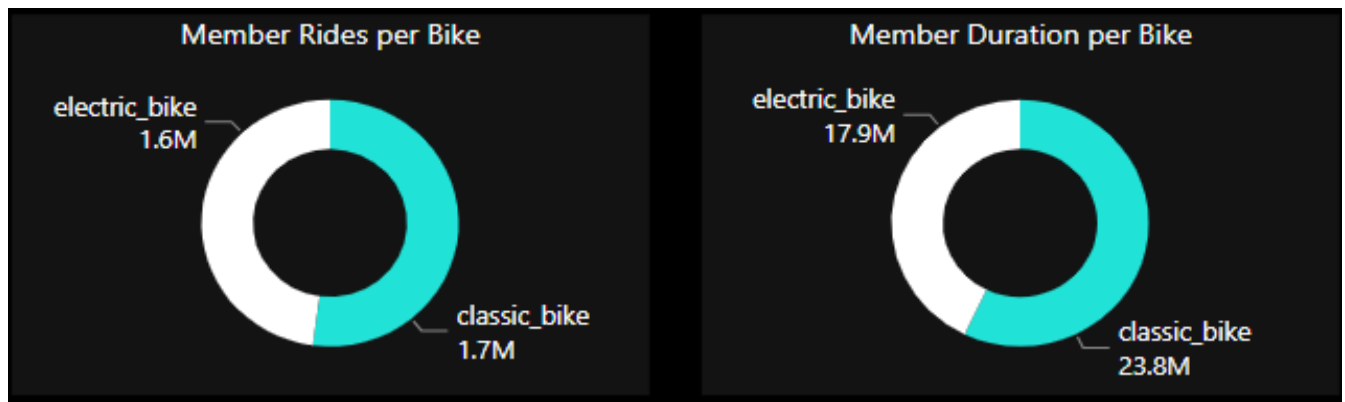
classic_bike
23.8M

- Casual clients prefer to use electric bikes in total rides but they use classic bikes for longer in the overall time duration.
- The docked bike comes as the second longest-used bike in time duration for casual clients more than the usage of electric bikes.
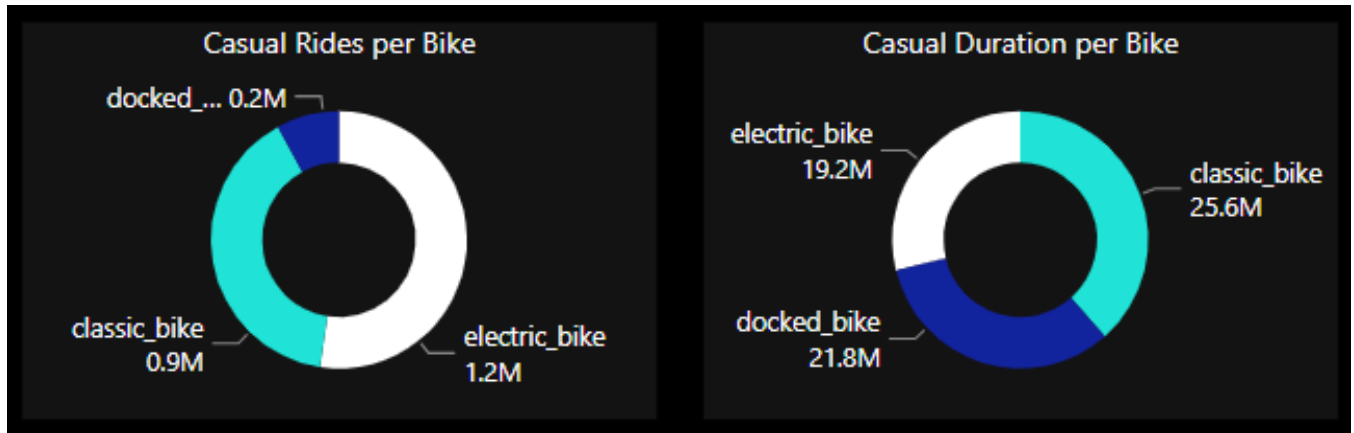
- The line graph over the year shows that it starts to increase from March towards the highest point in July, then goes down till it reaches the end of the year. This might show that bike-share is affected by seasonal changes, and we can promote offers in the seasons where there is a downward slope starting from August.
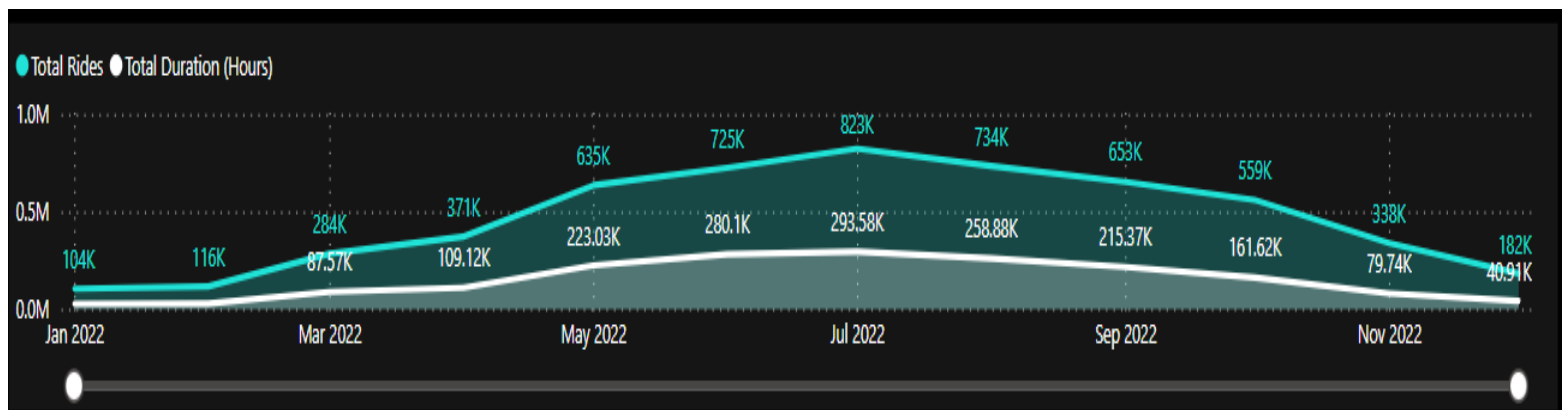
Limitations

- After coming through all these findings, I wish if had more data that supports my insights so we can understand our clients' behavior.

- I think we need data that shows us deeper aspects about our clients like:

  - Age

  - Education level

  - Income level

  - Marital status

  - Parent or not

That kind of data will help us interpret and understand more about our clients' behavior to build a more clear and more trustworthy strategy to guide our business.

Recommendation

Based on these insights, I can summarize my recommendation in the following points:

1. We need to promote offers for annual membership for the casual clients for the electric and classic bikes that might push more members to the list.

2. Special offers for the docked bike for both member and casual clients and mentioning its advantages to the environment over the classic one, and how it is useful to maintain our bodies sporty and healthy.

3. We need to run a marketing campaign to encounter customer attrition starting from August and increase our offers gradually towards December.

4. We can also include some offers for weekend bike-share services to gain more frequency over this part of the week.

I think these actionable insights can help:

- gain more customers and increase their retention.

- increase the usage of a bike type like the docked bike.

- overcome the curve recession over the seasonal changes.