Case Study: Bike Share membership conversion analysis

This case study is part of the Google Data Analytics Capstone. A fictionalized company needs data driven insights to design a marketing strategy which converts a user of casual membership into a annual membership holder. This analysis is performed by following the steps of the data analysis process: **Ask, Prepare, Process, Analyze, Share, and Act.**

# **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 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, we 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.

# **Background**

Cyclistic: is a bike-share program boasting over 5,800 bicycles and 600 docking stations. What sets Cyclistic apart is its inclusivity, offering not just traditional bikes but also reclining bikes, hand tricycles, and cargo bikes, catering to people with disabilities and those who cannot use a standard two-wheeled bike. Although the majority of riders prefer traditional bikes, about 8% utilize these assistive options. While most Cyclistic users ride for leisure, around 30% use the bikes for their daily work commute.

Customers who purchase single-ride or full-day passes are classified as casual riders, whereas those who opt for annual memberships are considered Cyclistic members. Financial analysis has revealed that annual members are significantly more profitable than casual riders.

The Director of Marketing, Moreno, has outlined a clear objective: develop marketing strategies to convert casual riders into annual members. To achieve this, the marketing analyst team needs to gain a deeper understanding of the differences between annual members and casual riders, the motivations for casual riders to become members, and the potential impact of digital media on marketing efforts. Moreno and her team plan to analyze historical bike trip data to uncover relevant trends and insights.

Data Analytics Process

Step 1: Defining Business Task

The primary focus of the project assigned is to devise a marketing strategy to convert casual riders to members by identifying the different factors which will convert a casual user into an annual member of Cyclistic. To further simplify the analysis process, scope of work and stakeholder expectations must be defined.

SCOPE OF WORK:

* Acquire historical data of both types of members.
* Identify general usage patterns of casual riders and annual members
* Find factors which lead to casual riders using the product.
* Generate strategies to market the annual membership to casual riders using data-driven analysis.

STAKEHOLDERS:

1. Lily Monero, director of Marketing
2. Cyclistic executives (requires approval of recommendations)
3. Marketing analytics team

Step 2: ASK

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?

Moreno has assigned me the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

Step 3: Collection and Preparation of Data

**Data Source**

* Data is made available by Motivate International Inc. The data is [licensed](https://www.divvybikes.com/data-license-agreement) and pre-validated for reliability and fairness. It is a public dataset.
* Due to data privacy issues, data cannot be connected to the individual riders personal info and their credit card numbers.
* Data location [url](https://divvy-tripdata.s3.amazonaws.com/index.html).

**Data features**

* The data is quantitative data formatted in a structured way stored as 12 distinct files of .csv filetype.
* Every file has a months’ worth of data combining for 1 year of data.
* There are 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.

Step 4: Data transformation and Processing

Tool used: SQL query using DB browser for SQlite

Data is first inspected to find out which tools can help accurately and efficiently transform the data into something usable for analysis. The following observations are made:

* Data is split into different files which means it must be combined to perform any further data cleaning.
* It is an expansive dataset with only 12 columns but thousands of rows.

It is necessary to combine the dataset into a single table which is achieved through SQlite command line shell Eg

.import C:\Desktop\Case\_study\202305\_divvy\_tripdata.csv tripdata

This is repeated for each .csv which totalled to 12 times.

**Data Cleaning**

The steps included identifying and handling duplicate records, null values, and calculating the ride duration. The cleaned data was then exported in CSV format for further analysis. All the code is here.

1. A simple query using COUNT() function reports that 5738612 rows are present.
2. A query to calculates the number of duplicate rows based on the **ride\_id** field, which should be unique for each trip is run giving an output of 0 meaning no duplicates.
3. Identifying if Null values are present. 1402773 rows returned having NULL values in at least 1 column. All Null value rows are deleted
4. To simplify analysis process, duration per ride is needed in a separate column. A new columns is created and populated with necessary data using following code



1. A query to calculate the ride duration by subtracting the **started\_at** timestamp from the **ended\_at** timestamp and converting the result from days to minutes. The **ride\_duration** is then rounded and stored in the newly added column.



1. It is observed that 80107 rides have ride duration of less than 1 min, of which 73 rides have durations in negative. This is caused because the ride durations here are more than 24 hrs. Since this is a smaller subset of the data, it is deleted along with the rest of the data with time less than 1 min.

The data is verified to be cleaned and transformed likewise now ready for Visualization.

STEP 5: Performing Analysis and Sharing observations through Visualization

The cleaned data is imported into Tableau Public for analysis. Before beginning, final verification of data integrity is performed. Few issues in format type was discovered in lat-long data which is swiftly fixed.

**General Insights through Data aggregation**

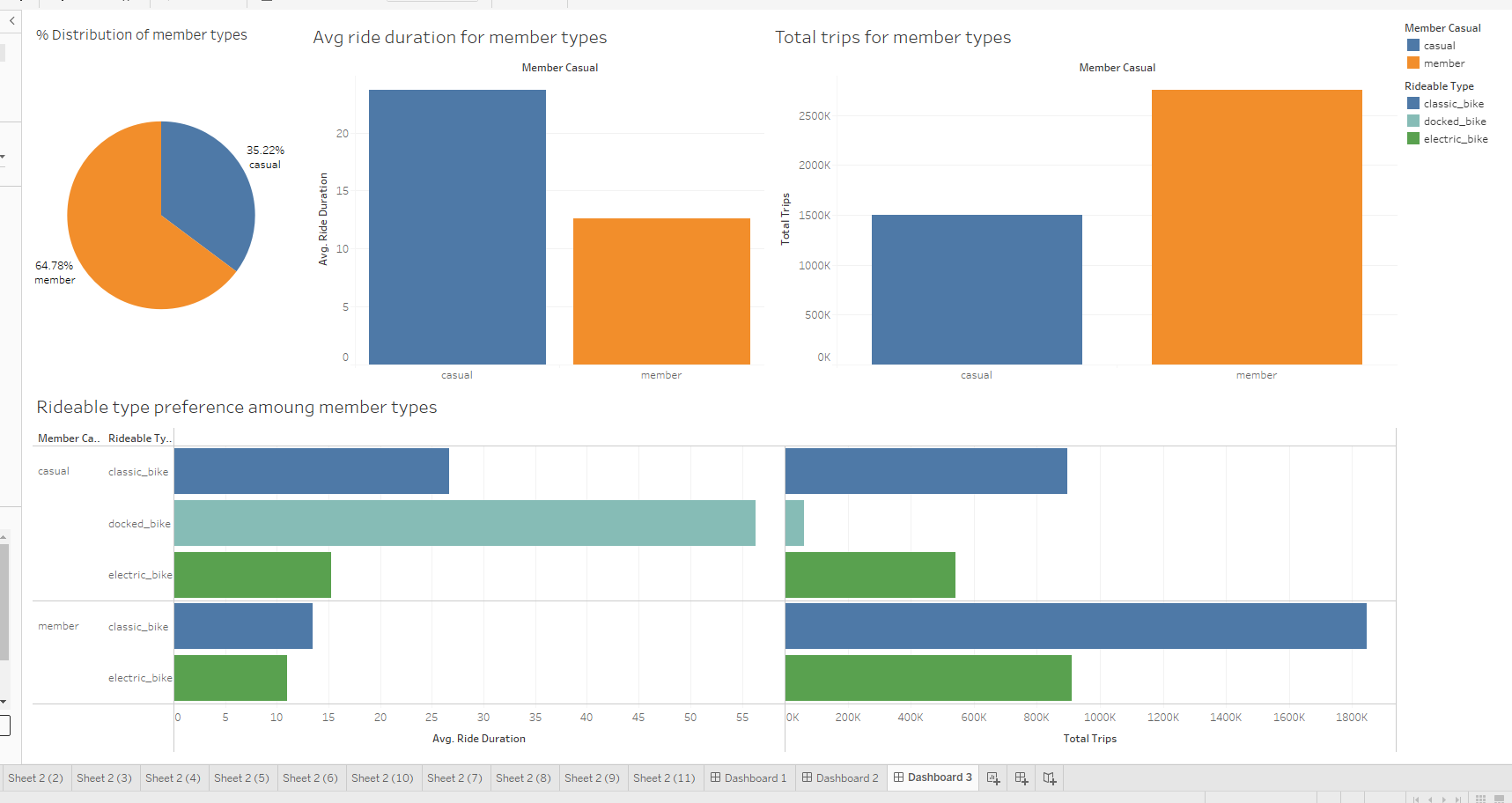
* A trend is observed during simple data scanning, most rides tend to originate and end around a general location
* Most users of Cyclistic are Annual membership holders

To answer the First question: *How do annual members and casual riders use Cyclistic bikes differently,* necessary characteristics of usage behavior and preferences needs to be identified.

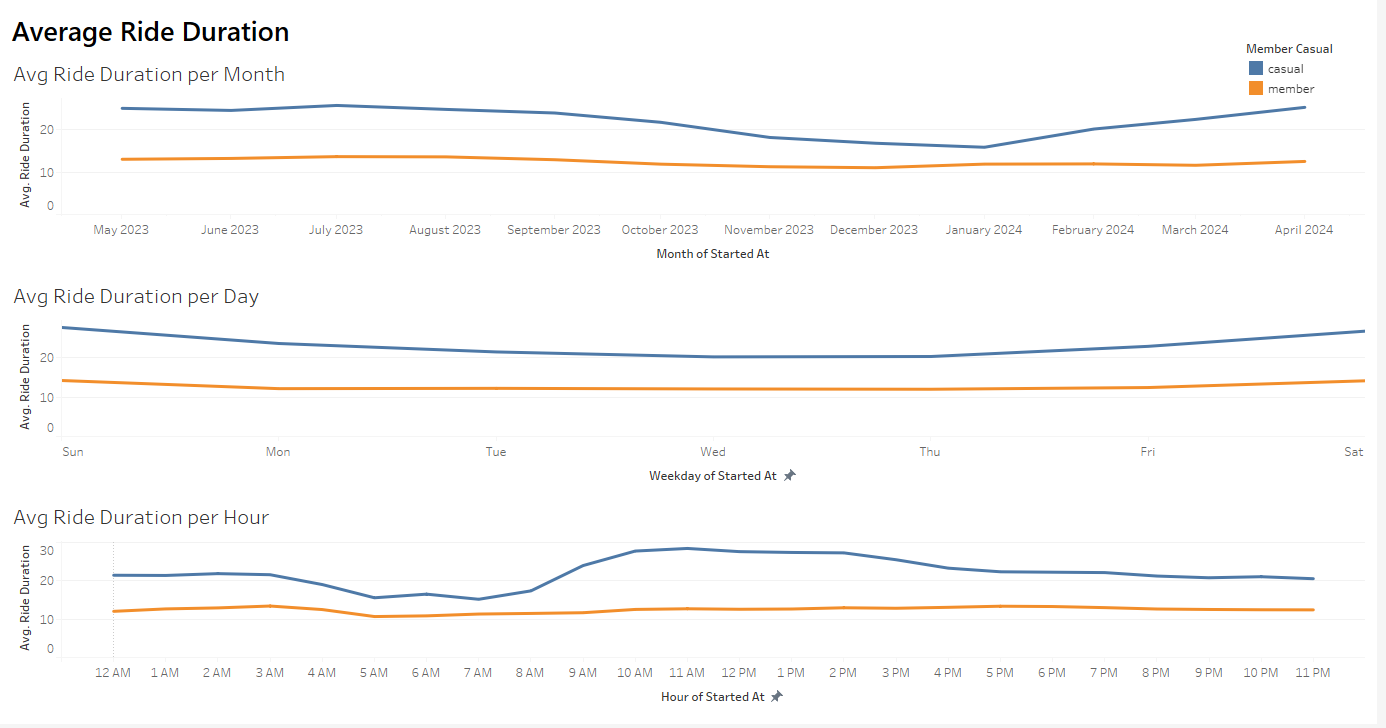
**Data Analysis**

Upon inspection of data, following observation about rideable type preference, ride duration patterns, and casual vs annual member behavior differences are spotted:

1. To begin, distribution of members types is identified. It is found that 64.78% of all riders are Members while 35.22% are Causal riders. This cements the fact that large demography of user are members.
2. No. of total trips for members are more than 25 lakhs which is in comparison 10 lakhs more than casuals riders
3. Contrasting the previous observation, the ride duration of casual riders is significantly higher ( approx. 25 min) than annual membership holders (approx. 13 min)
4. Preferences is also notable distinct among user types. If viewed from the lens of trips taken, casual prefer classic bikes the most and docked bikes the least. Members here are also seen liking classic bikes, surprisingly annual members do not use docked bike.
5. If using avg ride duration as metric, casual riders used docked bikes the most with last preference being classic bikes. As oppose to members who still preferred classic bikes but electric bikes are just slight behind by 3 min. For both users electric bike are close second.

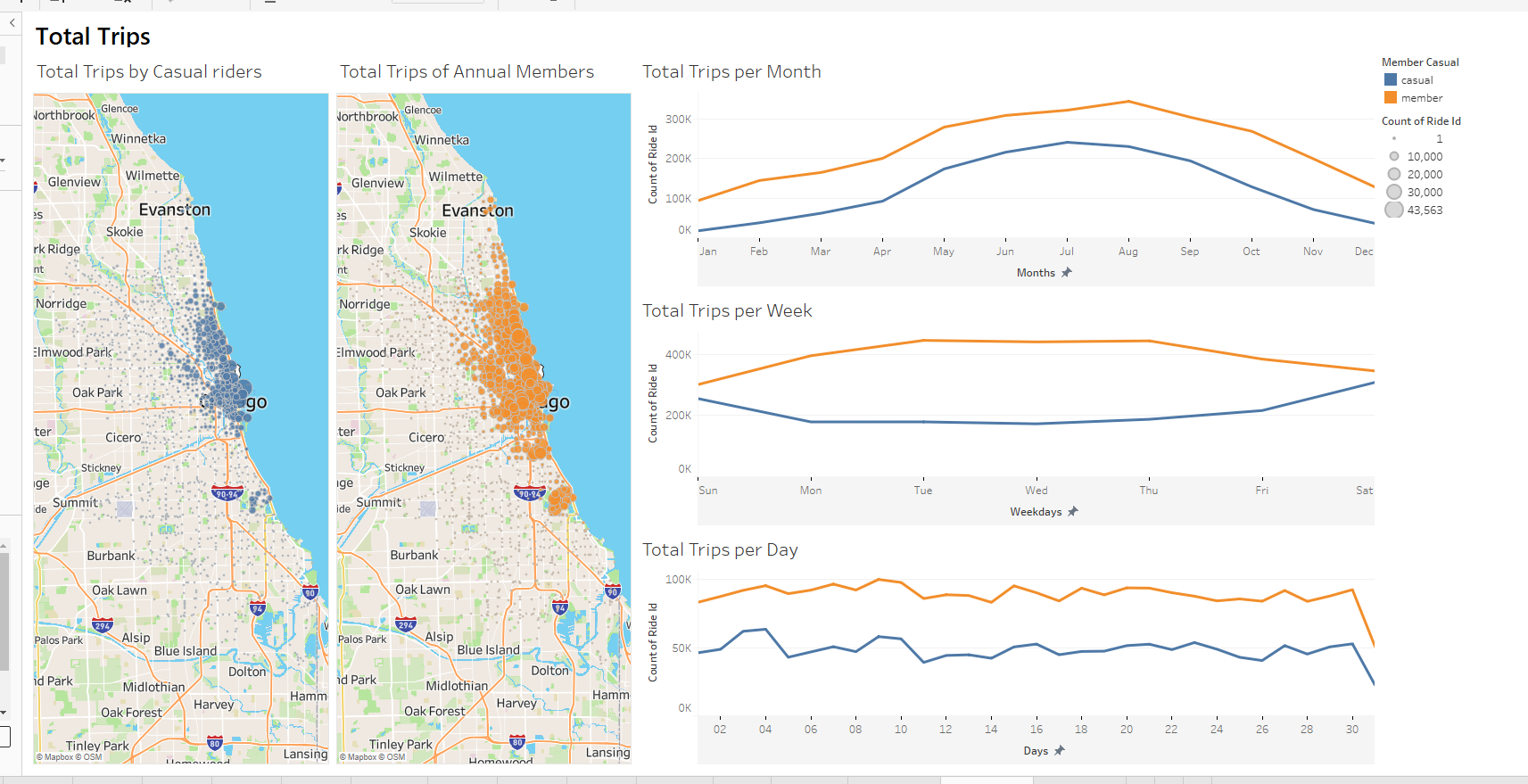


Upon looking at the Avg ride duration through a certain timespan following observation become clear:

1. Casual riders have higher avg ride duration compared to annual membership holders.
2. Monthly, causal riders show higher ride durations except for months October to January compared to annual members.
3. Average ride durations for casual riders are higher throughout the week, with both groups showing a slight decrease in duration from Sunday to Monday
4. Casual riders show higher average ride durations throughout the day, with peaks around late morning and early evening

Looking at total trips done by both users:

1. Through spatial distribution analysis, it is observed that casual rides are more concentrated towards the centre of Chicago, while annual members cover a broader area.
2. Annual members do more trips than casual riders. Both casual riders and annual members show increased trip activity during the warmer months (May to September), with annual members consistently taking more trips.
3. A higher and more consistent number of trips throughout the week for annual members, with slight peaks on weekdays, indicating regular commuting patterns. Its opposite for casual riders who show dip during weekdays showing travel during holidays.
4. The daily trip pattern shows annual members maintaining a higher number of trips each day then casual riders. Both seem to mimic same trip patterns.



**Conclusion**

The analysis shows that casual riders have longer ride durations and prefer docked bikes, while annual members use Classic bikes more consistently. Casual riders' trips peak during warmer months and weekends indicating casual riders are mostly vacation takers, whereas annual members maintain steady use throughout the week, indicating commuting patterns.

STEP 6: Act

After identifying the differences between casual riders and members, we can develop targeted marketing strategies to persuade casual riders to become members. Here are the recommendations:

1. Perform membership benefits ads and emphasize cost savings, convenience, and any other perks of annual memberships to users who have used Cyclistic services for a week consistently.
2. Target Peak Usage Times by launching promotions during warmer months and weekends to attract casual riders.
3. Electric Bike seem to be popular among casual riders for short distances. Offer trials or discounts for electric bikes availed through annual membership.
4. Personalized Communication, use data-driven insights to send tailored messages and offers.
5. Start Limited-Time promotions during peak casual usage. Use short-term discounts and offers to encourage sign-ups for annual membership.
6. Feedback Gathering: Collect and act on casual rider feedback to improve membership offerings.