

# An Analysis of the Cyclistic Bike Share Case Study with SQL

Hello!

Over the last few months, I've been working on completing the Google Data Analytics Professional Certificate on Coursera. It was quite tough and challenging but also a learning experience and so I thought I would share a documentation process.

This article will highlight my efforts to understand and extract insight from the dataset I was presented with using my knowledge of the analysis process: *Ask, Prepare, Process, Analyze, Share & Act*.



## SCENARIO

I am a junior data analyst working in 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, 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.

## 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 geotracked and locked into a network of 692

stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

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, 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 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.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order 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, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

## I. ASK

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?**

## II. PREPARE

I Downloaded the datasets from November 2021 till October 2022(202111-divvy-tripdata.zip -> 202210-divvy-tripdata.zip) using this [link](#) and validated the ROCCC (Reliability, Originality, Comprehensive, Cited, Current). I unzipped the folders and saved them into a folder "Cyclistic Bike Share Data".

As a disclaimer the data has been made available by Motivate International Inc. under this [license](#).

## III. PROCESS

Here I cleaned the data and prepared it for analysis to be sure I was gaining the right insight. The tool used for this was MS EXCEL.

- I copied the whole table and pasted it into a new sheet I named Worksheet. This worksheet is where all my cleaning was done.

- I searched for duplicates for the table and deleted any if found. N.B: The duplicates were only possible if there was more than 1 unique ride\_id.
- I hid columns E to L since they were not relevant to the cleaning.
- To validate the ride\_id data, I added a new column and used the Len function (LEN()) to determine the number of characters. I filtered the column to show only characters that were greater than or less than 16, then deleted those rows.
- I created a column to verify that the “started\_at” wasn’t greater than the “ended\_at” with a formula, I deleted all rows which were.
- I created “started\_day”, “ended\_day”, “started\_time” & “ended\_time” columns and populated them with data from the “started\_at” & “ended\_at” columns.
- I created a “ride\_length” column from the time columns to determine how long each ride lasted.
- I created a “ride\_length\_secs” column to determine how much time in seconds each ride lasted.
- I created “weekday\_started” & “weekday\_ended” columns to see what day of the week the rides started & ended. I used numbers to signify the days of the week, where 1 = Sunday, 2 = Monday, 3 = Tuesday, 4 = Wednesday, 5 = Thursday, 6 = Friday & 7 = Saturday.
- I saved this worksheet as a CSV file and applied the process to the remaining 11 datasets.

#### **IV. ANALYZE WITH MICROSOFT SQL SERVER**

I imported the 12 tables into MS SQL Server for analysis after cleaning the data. I used a union all query to combine the 12 tables. After that, I created a new Cyclicistic bike-share table and populated it with the union all query.

I used the new table to analyze the data, which led me to the following conclusions:

1. No of rides
2. Most patronized bike types
3. Average length of ride in seconds
4. Most active days
5. Max ride length
6. Top 10 stations with the most immediate cancellations
7. Stations & how many trips were started and ended at them
8. No of trips which began at different times of day

You can find the SQL codes [here](#).

#### **V. SHARE**

Unfortunately, as at the time of this article’s publishing I’ve not been able to successfully load the datasets into Tableau or Power BI for analysis & visualization. When I do get it right I will ensure to update it.

#### **VI. ACT**

Now I will answer the key question I was saddled with and give my recommendations based on the analysis to enable Cyclistic make informed decisions.

## **Summary**

. No of rides

Electric bikes had the most rides, followed by Classic bikes and then Docked bikes.

. Most patronized bike types

Members patronized both Electric & Classic bikes but not Docked bikes. This preference could be worth further analysis, it could be that the choice to have dockless bikes is a big deal for many of the users.

. Average length of ride in seconds

Members spent less time riding on average than the casuals, this could mean they travel lesser distances on average than the casuals or spend less time getting to their destinations from their start points.

. Most active days

The most active days for Casuals were the weekends, there could be a host of reasons for this, leisure, to avoid traffic, exercise etc. For Members the most active days were weekdays, this could be a primary source of commute for them around the bustling city to avoid traffic and possibly exercise at the same time.

. Max ride length

Casuals have the greater max time spent riding; this could be as a result of enjoying the ride (pleasure/leisure/exploring) or as a means of exercising.

. To find out what stations were favoured most for starting & ending trips by both user types  
The casuals favoured "Streeter Dr & Grand Ave" to start & end their trips, while the members favoured "Kingsbury St & Kinzie St" to start & end theirs.

. To find out how many trips were started at different periods of the day

There were far more day trips at 5,355,575 than there were night trips at 320,242. This would stand to reason that most users were active during the day.

## **Recommendations**

- Market membership subscription discount for docked bikes to attract more annual users to the bike type.
- Marketing campaigns should be targeted for the busiest casual rider days (weekends) and busiest hours to reach the most riders (during the day).
- Targeted premium features could be offered to persuade casual users to join as members to meet their goals for riding, mostly on weekends.
- Conduct further analysis with expansive data on Cyclistic customers to better understand their social, behavioural, and cultural preferences.