# Google Data Analytics Capstone Project: Cyclistic Bikeshare Case Study

## **Introduction**

The cyclistic bike-share analysis case study is a Capstone Project of the Google Data Analytics Certificate course. I have followed the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act to answer the business question of a hypothetical scenario. I have used SQL to process and analyze the data and Tableau to visualize it.

## **Scenario**

Cyclists is a bike-share company in Chicago. The company offers a fleet of 5,824 tracked bicycles and is locked into a network of 692 stations across Chicago. Cyclists have divided its customers into two types:

* **Members**, customers who purchase annual memberships
* **Casuals**, customers who purchase single-ride or full-day passes

The marketing director believes the company’s future success depends on maximizing the number of annual memberships. Therefore, our team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, our team will design a new marketing strategy to convert casual riders into annual members.

## **Ask**

### Business Task

Create a marketing strategy to convert casual riders to members.

### Analysis Questions

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 below analysis is done to answer the first question: How do annual members and casual riders use Cyclistic bikes differently?

## **Prepare**

### Data Source

In this project, Cyclistic’s historical trip data is analyzed to identify trends from Jan 2023 to Dec 2023. The data can be downloaded from [divvy\_tripdata](https://divvy-tripdata.s3.amazonaws.com/index.html) and has been made available by Motivate International Inc. under this [license](https://divvybikes.com/data-license-agreement). The data is appropriate to explore how different customer types are using Cyclistic bikes. But note that data privacy issues prohibit using riders’ personally identifiable information, so connecting 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 is not possible.

### Data Organization

12 CSV files for each month of the year 2023 were downloaded and saved in the folder named ‘divvy\_tripdata’ in the local system. The names of the extracted files were changed from YYYYMM-divvy-tripdata to tripdata\_MM\_2023. Each file includes 13 columns as below:

1. ride\_id – Unique ride id
2. rideable\_type- Types of bikes
3. started\_at- start time
4. ended\_at- end time
5. start\_station\_name- start station name
6. start\_station\_id- start station id
7. end\_station\_name- end station name
8. end\_station\_id- end station id
9. start\_lat- start station latitude
10. start\_lng- start station longitude
11. end\_lat- end station latitude
12. end\_lng – end station longitude
13. member\_casual- type of customer (member or casual)

## **Process**

SQL DB browser was used as an SQL tool for processing the data as it was not possible to handle large amounts of data in Microsoft Excel.

The following steps were used in combining the 12 files into one table:

* Created a new database in the SQL DB browser
* Imported all 12 CSV files as tables in the SQL DB browser
* Used the SQL code to create a new table ‘combined\_2023’ by combining all 12 tables.
* The new table includes 5,719,877 rows and 13 columns

The following steps are used in further exploring, cleaning, and manipulating the data (SQL query here)

* Checked the column name and their data typede\_id
* Column rideable\_type has 3 unique values: classic, docked, and electric
* Column member\_casual has 2 unique values: member and casual
* There were a few missing values found for station name columns however, ride\_id, rideable\_type, started\_at, ended\_at, and member\_casual columns have no missing values so none of the rows were deleted
* There were no duplicates found as all the ride\_id values were unique
* Added 3 new columns- ride\_length, ride\_day and month
* Trips with ride\_duration 0 and negative values were deleted
* A total of 88,950 rows were deleted, so the total rows or total number of trips was 5,630,927

## **Analyze**

The following calculations were done as a part of the descriptive analysis and to compare how member riders and casual riders used Cyclists bike-sharing rides differently in SQL (SQL codes here).

* Total number of rides taken
* Average of the ride \_length
* Maximum value of the ride \_length
* Mode of the ride\_day
* Number of rides by customer type
* Average of the ride \_length by customer type
* Number of rides by bike type
* Number of rides taken by day of the week
* Number of rides taken by month
* Number of rides taken by hour of the day

## **Share**

The data was visualized in Tableau ([Tableau link here](https://public.tableau.com/app/profile/neetika7566/viz/Book1_17048255362360/Dashboard4)). Some of the key findings are mentioned below with visuals to support them.

1. **Number and rides and average of ride length by customer type**

The member took more rides constituting **64%** of the total rides compared to the casual riders comprising only **36%** of the total rides. However, the average duration of the rides was more for casual riders than for member riders.

A graph of a user

Description automatically generated with medium confidence

1. **Number of rides by bike type**

Both types of customers prefer classic and electric bikes constituting **51%** and **47%** of the total bike rides compared to docked bikes which are tried by very few (~1%) casual riders only.

A screenshot of a graph

Description automatically generated

1. **Number of rides taken by day of the weeks**

Members took maximum rides on Tuesday, Wednesday, and Thursday whereas casual riders used it mostly on weekends.

A graph of blue and orange bars

Description automatically generated

1. **Number of rides taken by hour of the day**

Member riders used bikes at office peak hours (8 am & 5 pm) and casual riders usage increased during the daytime, peaking at 5 pm

A graph showing the time and the time

Description automatically generated with medium confidence

1. **Number of rides taken by month**

Both types of riders typically use Cyclists bikes during the month of summer in June, July, and August

A graph of different colored bars

Description automatically generated

## **Act**

### Recommendations

1. **As casual customers are renting bikes for longer durations so, Cyclistic should consider offering some incentives for members who are using bikes for a longer time which can motivate casual customers to register as members.**
2. **Cyclistic should consider offering monthly or seasonal membership as most of the casual riders are renting bikes during the summer months.**
3. **Cyclistic should focus its marketing campaign on tourists as most of the casual riders rent bikes on weekends and also use them during the daytime which shows that they are not using bikes for office commute but for leisure. So, targeting tourists would be helpful and can encourage casual riders who are tourists to register for a monthly or seasonal membership.**