**CYCLISTIC BIKE-SHARE CASE STUDY**

**PROJECT REPORT**

**INTRODUCTION**

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geo-tracked 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.

**PROBLEM STATEMENT**

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, Cyclistic’s director of marketing, 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 solid opportunity 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 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.

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?

**DATA ANALYSIS PROCESS**

**ASK PHASE**

The questions that would be guiding future marketing programs are:

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

2. Why would casual riders buy Cyclistic annual memberships?

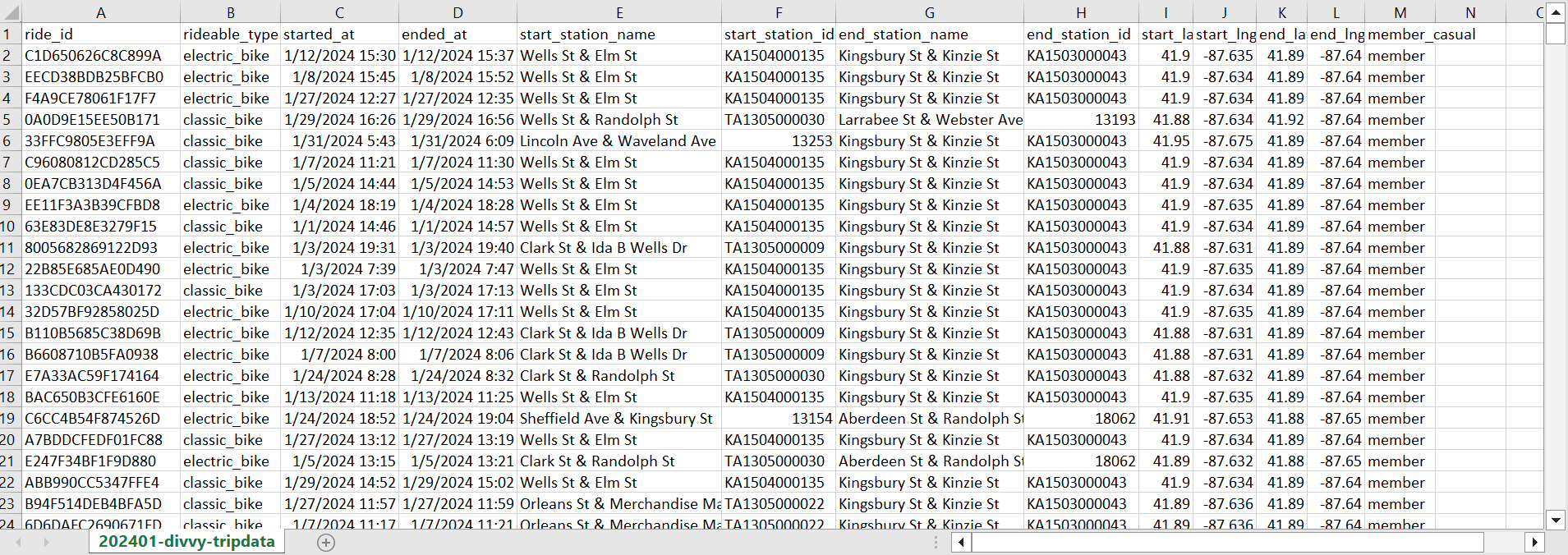
3. How can Cyclistic use digital media to influence casual riders to become members?

This project report has been specifically tasked with answering the first question, this is, how do annual members and casual riders use Cyclistic bikes differently in the last one year?

**PREPARE PHASE**

The data to be used for this project is gotten from Cyclistic’s historical trip data linked [here](https://divvy-tripdata.s3.amazonaws.com/index.html). This is a public data that can be used to explore how different customer types are using Cyclistic bikes and also used to analyze and identify trends. (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement).)

The data linked above contains trip data based on months from April, 2020 till the last month before this analysis, the data are updated after the end of every month. A snippet of how the data is organized is shown below:



Metadata:

ride\_id: Unique ID for each ride

rideable\_type: Kind of bike used for the ride

started\_at: Date and time when ride started

ended\_at: Date and time when ride ended

start\_station\_name: Station where ride started

start\_station\_id: ID of station where ride started

end\_station\_name: Station where ride ended

end\_station\_id: ID of station where ride ended

start\_lat: Latitude of station where ride started

start\_lng: Longitude of station where ride started

end\_lat: Latitude of station where ride ended

end\_lng: Longitude of station where ride ended

member\_casual: Membership type

**PROCESS PHASE**

The following steps taken to process data

1. Download the previous 12 months of trip data.

2. Unzip the files.

3. Create a folder on desktop to house the files. Use appropriate file-naming conventions.

4. Create subfolders for the .csv file and the .xlsx or Sheets file so that original copy of data is available. Move the downloaded files to the appropriate subfolder.

In Excel

5. Launch Excel, open each file, and choose to Save As an Excel Workbook file. Put it in the subfolder you created for .xlsx files.

6. Open spreadsheet and create a column called ride\_length. Calculate the length of each ride by subtracting the column started\_at from the column ended\_at (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55, then populate across each record.

7. Create a column called day\_of\_week, and calculate the day of the week that each ride started using the WEEKDAY command (for example, =WEEKDAY(C2,1)) in each file. Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday, then populate across each record.

8. Create another column called month, and input the month of each ride using the TEXT() function (for example, =TEXT(D2,”mmmm”)) and format as General, then populate across each record.

9. Proceed to remove duplicate data using Remove Duplicates, on Data tab > Data Tools > Remove Duplicates

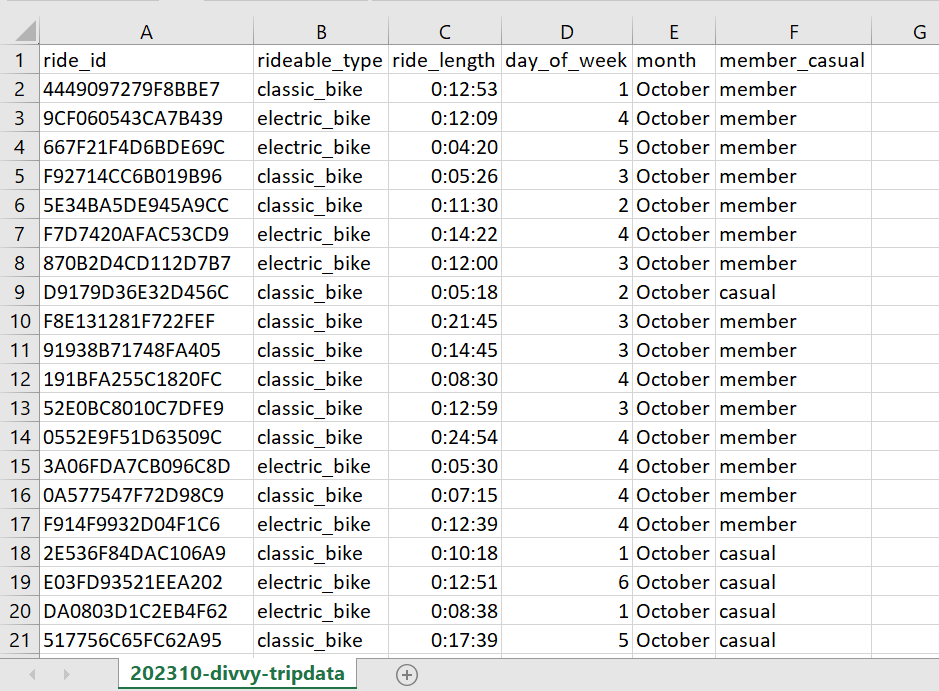
10. Remove columns that would not be needed for analysis, columns such as started\_at, ended\_at, start\_station\_name, start\_staion\_id, end\_station\_name, end\_station\_id, start\_lat, start\_lng, end\_lat, end\_lng.

11. Remove null values and unreadable text from the remaining columns using filter and sort

12. Check to confirm that each columns are in the correct format.

13. Redo from 1-12 for the other datasets

After all the steps above, a snippet of the data is shown below:



**ANALYZE PHASE**

Open one of the Excel data, then complete the following steps:

Excel (Conduct Descriptive Analysis)

1. On the opened workbook, run a few calculations to get a better sense of the data layout.

* Calculate the mean of ride\_length. For example, on cell H2, input =AVERAGE(C:C), and format as HH:MM:SS using Format > Cells > Time > 37:30:55
* Calculate the max of ride\_length. For example, on cell H3, input =MAX(C:C), and format as HH:MM:SS using Format > Cells > Time > 37:30:55
* Calculate the mode of day\_of\_week. For example, on cell H4, input =MODE(D:D)

1. Create a pivot table to quickly calculate and visualize the data.

* Calculate the average ride\_length for members and casual riders. Try rows = member\_casual; Values = Average of ride\_length.
* Calculate the average ride\_length for users by day\_of\_week. Try columns = day\_of\_week; Rows = member\_casual; Values = Average of ride\_length.
* Calculate the number of rides for users by day\_of\_week by adding Count of trip\_id to Values.

1. Open another file and perform the same descriptive analysis steps. Explore different seasons to make some initial observations.
2. Once this is done, merge them into a full-year view.

SQL (PostgreSQL)

1. Make sure the location of all 12 is easily accessible
2. Open PostgreSQL and connect to server, then create a database, follow the steps below:

* Left click Databases > Create > Databases
* Fill in the name of the database (TRIPDATA) and click save

1. Open TRIPDATA database and create tables for each data. Follow the steps below:

* Extend view of Databases > extend view of TRIPDATA > extend view of Schemas > Tables > Query Tool
* In the Query editor displayed, type in and execute the following:

CREATE TABLE IF NOT EXISTS public.tripdata01

(

ride\_id text COLLATE pg\_catalog."default",

rideable\_type text COLLATE pg\_catalog."default",

ride\_length interval,

day\_of\_week text COLLATE pg\_catalog."default",

month text COLLATE pg\_catalog."default",

member\_casual text COLLATE pg\_catalog."default"

)

TABLESPACE pg\_default;

ALTER TABLE IF EXISTS public.tripdata01

OWNER to postgres;

* Refresh the database
* This creates and formats Tripdata01
* Do this for the remaining 11 table and name accordingly, that is, tripdata02, tripdata03, tripdata04 and so on.

1. Import data into tables created in TRIPDATA database, follow the steps below:

* Right click tripdata01 > Import/Export Data
* Check Import on General tab, input file path for data to be imported or search for file
* Select comma (,) delimiter and check Header in the Options tab and click Ok
* Do this for the remaining data with January’s data as Tripdata01 and December’s data as Tripdata12 for easy identification.

1. To merge all of the month’s data to get a full-year view, Open the Query Editor and input the following:

WITH total AS

(SELECT \*

FROM tripdata01

UNION ALL

SELECT \*

FROM tripdata02

UNION ALL

SELECT \*

FROM tripdata03

UNION ALL

SELECT \*

FROM tripdata04

UNION ALL

SELECT \*

FROM tripdata05

UNION ALL

SELECT \*

FROM tripdata06

UNION ALL

SELECT \*

FROM tripdata07

UNION ALL

SELECT \*

FROM tripdata08

UNION ALL

SELECT \*

FROM tripdata09

UNION ALL

SELECT \*

FROM tripdata10

UNION ALL

SELECT \*

FROM tripdata11

UNION ALL

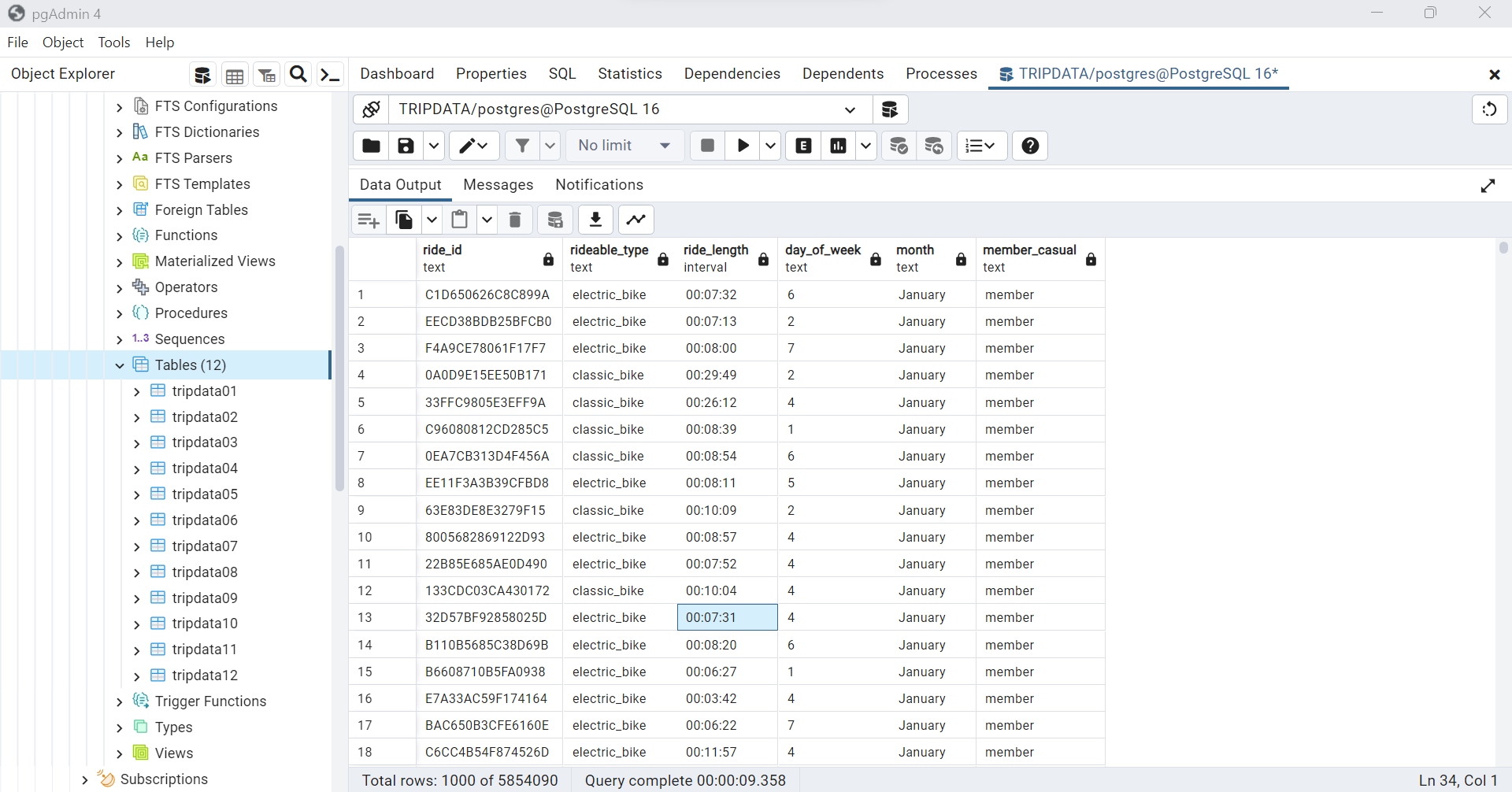
SELECT \*

FROM tripdata12)

SELECT \*

FROM total

A snippet of the merged data is shown below:



1. To create a table that summarizes all data, enter the following query in the query editor and download as CSV (named gen2):

SELECT COUNT(ride\_id) no\_of\_ride, SUM(ride\_length) total\_ride\_time,

AVG(ride\_length) ave\_ride\_time, MAX(ride\_length) max\_ride\_time,

month, member\_casual, COUNT(rideable\_type) bike\_type\_count, rideable\_type,

COUNT(day\_of\_week) day\_of\_week\_count, day\_of\_week

FROM total

GROUP BY month, member\_casual, rideable\_type,day\_of\_week

1. Also to get the average time spent on each ride grouped by membership type and months of the year, enter the following query in the query editor and download as CSV (named avg):

SELECT member\_casual, month, AVG(ride\_length) avg, SUM(ride\_length) sum

FROM total

GROUP BY member\_casual, month

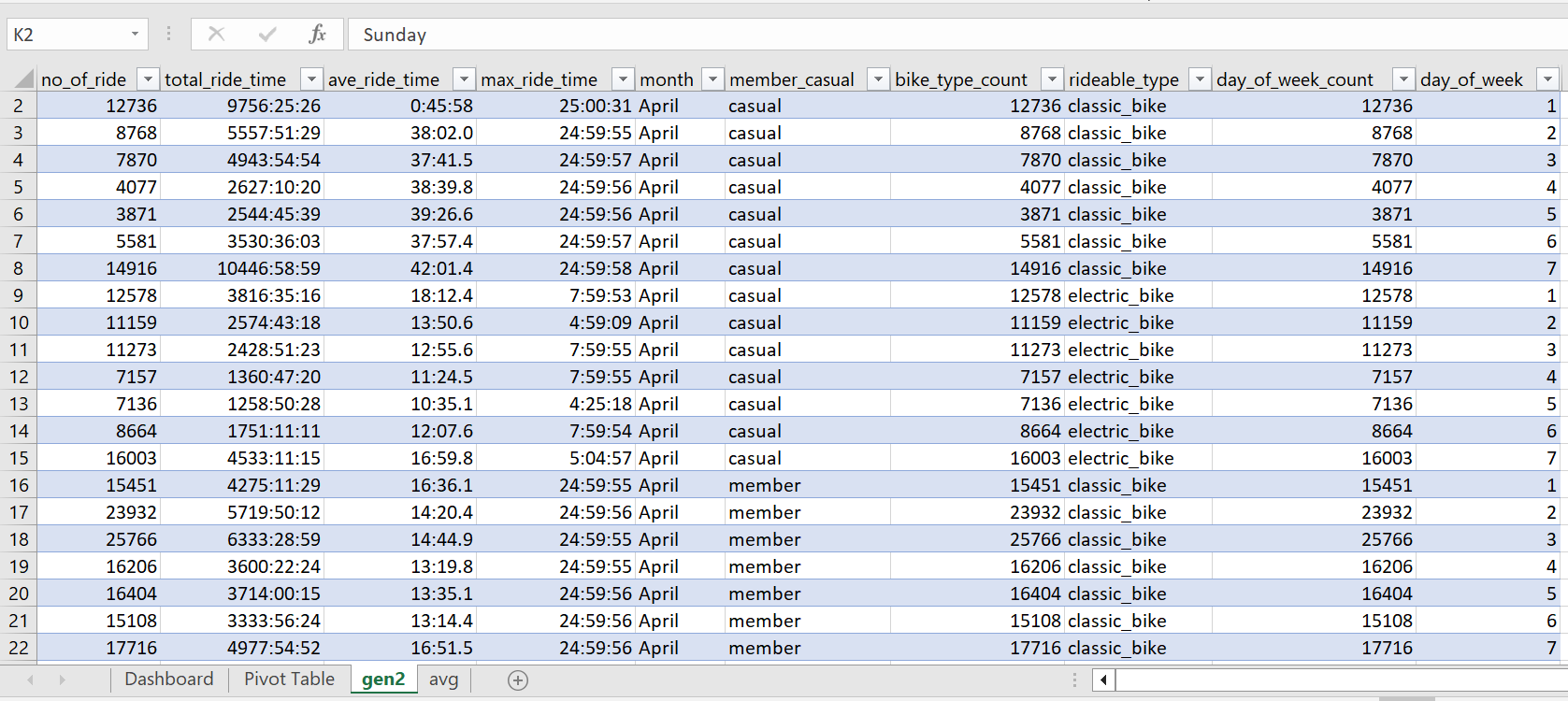
1. Lastly, get the total average time spent on ride through out the year, enter the following query in the query editor:

SELECT AVG(ride\_length)

FROM total

Excel

1. Open and resave the two CSV file as xlsx files.
2. Put both files into a single excel file but different worksheet
3. Create a table in both worksheet
4. Create two more worksheet, one for pivot tables and the other for a dashboard



Excel (Pivot Tables)

1. From gen2 sheet, create the following pivot tables:

* Calculate the total ride time for members and casual riders by the year. Try rows = month; columns = member\_casual; Values =sum of total\_ride\_time
* Calculate the total number of rides for members and casual riders by the year. Try rows = month; columns = member\_casual; Values =sum of no\_of\_ride
* Calculate the share of total ride for members and casual riders. Try rows = member\_casual; Values =sum of no\_of\_ride
* Calculate the preferred bike for members and casual riders. Try rows = rideable\_type; columns = member\_casual; Values =sum of no\_of\_ride

1. On avg sheet, create a new column (avg\_time), and populate records with 17:18, which is the total average ride time for the year
2. From avg sheet, create the following pivot table:

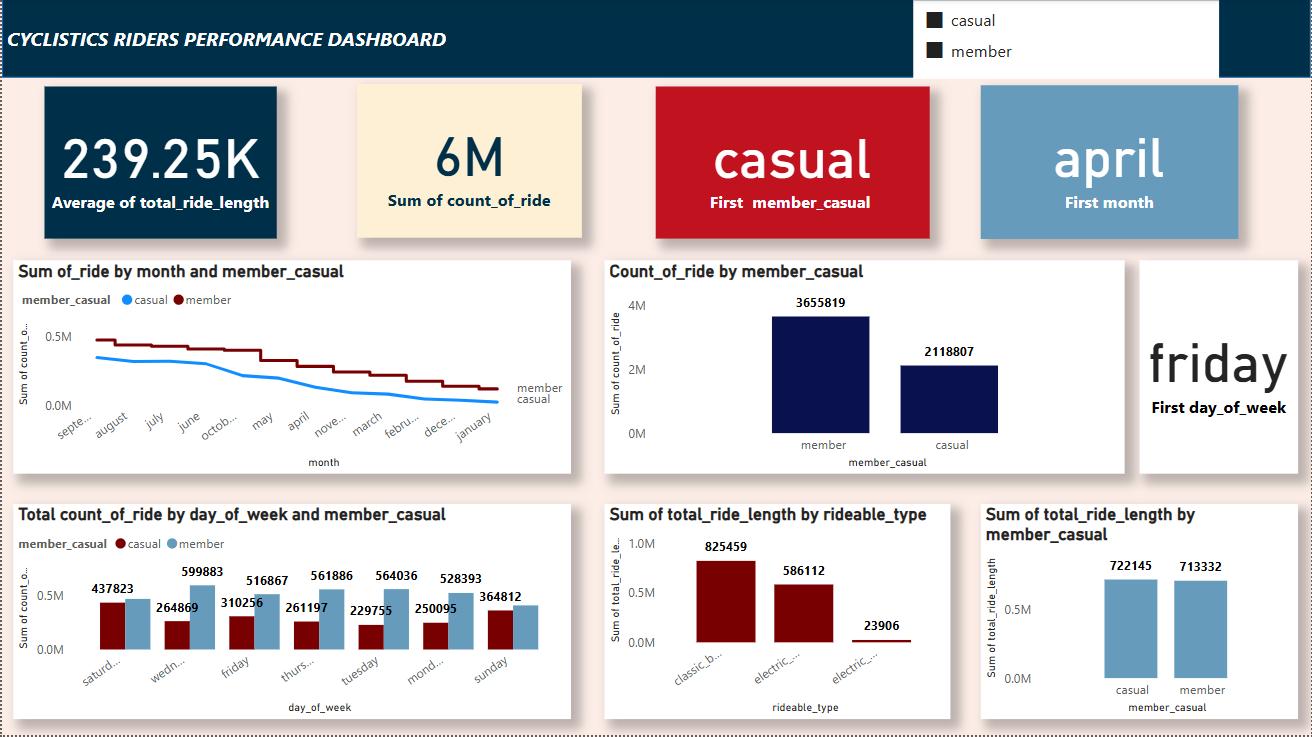
* Calculate the average ride time for members and casual riders through the year compared to the total average ride time. Try rows = month; columns = member\_casual; Values = sum of avg, min of avg\_time

Excel (Dashboard)

1. Use the pivot tables in the Pivot table sheet to create graphs accordingly
2. Design dashboard accordingly
3. Add filter for interactivity

**SHARE PHASE**

After analyzing the data, A visualization is made to support and present key findings. The visualization below shows how members and casual riders use cyclistic differently.



**OBSERVATIONS AND RECOMMENDATIONS**

1. Casual riders have the least number of rides but at average spends more time on each ride
2. Peak periods are from around May to October with casual riders taking the lead in ride time
3. Casual members use electric scooters more than members
4. There should be a discount for longer ride.
5. This discount should be within the peak period for members.
6. More advert targeted at the new electric scooters