

Google Data Analytics Capstone Project

Colin Wong

2023-08-28

Contents

Summary

I worked on the Google Data Analytics Capstone Project - How Does a Bike-Share Navigate Speedy Success ?. I am using R Markdown to document my process of using tools like Microsoft Excel, Google Big Query and tableau to perform data cleaning, analysis, and visualization tasks. Ultimately, I will present business recommendations based on the analysis performed.

Quick Links: [Data Source](#) |

Background

I am a junior analyst at a Bike rental company called *Cyclistic*. The company has grown a fleet of 5824 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 one. The company offers three 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.

Goal

Design marketing strategies aimed at converting casual riders into annual members.

Business task

To understand the question - “How do annual members and casual riders use Cyclistic bikes differently?” and to provide recommendations after analysing the data.

Data Source

Cyclistic's monthly historical trip data in 2022 will be used to analyse and identify trends.

Data Source:(<https://divvy-tripdata.s3.amazonaws.com/index.html>)

Microsoft Excel - Data Cleaning

I plan to use Microsoft Excel to prepare and clean the data before conducting the analysis. I have downloaded the 12 trip data files for 2022, each excel files contains trip data for one single month.

Google capstone > Cyclistic Monthly Data 2022				
Name	Date modified	Type	Size	
202201-divvy-tripdata	5/5/2023 12:19 pm	Microsoft Excel Comma Separated Values File	18,567 KB	
202202-divvy-tripdata	5/5/2023 12:19 pm	Microsoft Excel Comma Separated Values File	20,638 KB	
202203-divvy-tripdata	5/5/2023 12:20 pm	Microsoft Excel Comma Separated Values File	50,533 KB	
202204-divvy-tripdata	5/5/2023 12:20 pm	Microsoft Excel Comma Separated Values File	65,341 KB	
202205-divvy-tripdata	5/5/2023 12:21 pm	Microsoft Excel Comma Separated Values File	114,780 KB	
202206-divvy-tripdata	5/5/2023 12:11 pm	Microsoft Excel Comma Separated Values File	140,228 KB	
202207-divvy-tripdata	5/5/2023 12:23 pm	Microsoft Excel Comma Separated Values File	149,306 KB	
202208-divvy-tripdata	5/5/2023 12:23 pm	Microsoft Excel Comma Separated Values File	142,148 KB	
202209-divvy-tripdata	5/5/2023 12:25 pm	Microsoft Excel Comma Separated Values File	138,135 KB	
202210-divvy-tripdata	5/5/2023 12:26 pm	Microsoft Excel Comma Separated Values File	109,293 KB	
202211-divvy-tripdata	5/5/2023 12:26 pm	Microsoft Excel Comma Separated Values File	66,348 KB	
202212-divvy-tripdata	5/5/2023 12:27 pm	Microsoft Excel Comma Separated Values File	35,612 KB	

For each excel csv file, I have performed below action:

(As the file formatting for all files are the same, I will be using VBA macros function in Excel to automate part of the cleaning process for all the Excel files. I will first record a macro while manually performing the cleaning steps on the initial Excel file. Subsequently, I can apply the recorded macro to the remaining files to execute the same set of cleaning operations. This approach ensures consistency in the formatting and automation of the cleaning process across all Excel files.)

1. To narrow down my analysis, I will exclude unrelated columns and retain only “rideable_type,” “started_at,” “ended_at,” and “member_casual.”
2. I will generate a new column called “ride_length” by subtracting the “started_at” column from the “ended_at” column in each cell. The result will be formatted as HH:MM:SS using the format cell function.
3. Adding two new columns, “started_at_day_of_week” and “ended_at_day_of_week,”. For each cell, the formula “=weekday(cell(started_at/ended_at),1)” will be applied to determine the day of the week.
4. Adding a new column named “month” . The formula “=month(cell(started_at))” will be applied to each cell to extract the month.
5. Adding four new columns, namely “started_date,” “ended_date,” “started_time,” and “ended_time,” . Each cell will be transformed to the corresponding date/hour format from the “started_at”/“ended_at” column values, respectively.
6. Applying the replace function to substitute the values in the “started_at_day_of_week”/“ended_at_day_of_week” columns from 1-7 with Monday-Sunday. Additionally, the values in the “month” column will be replaced from 1-12 with January-December.

7. Finally, by using filtering and conditional formatting functions I will identify and eliminate errors, and use duplicate removal function to eliminate duplicated values, and subsequently save the CSV file.

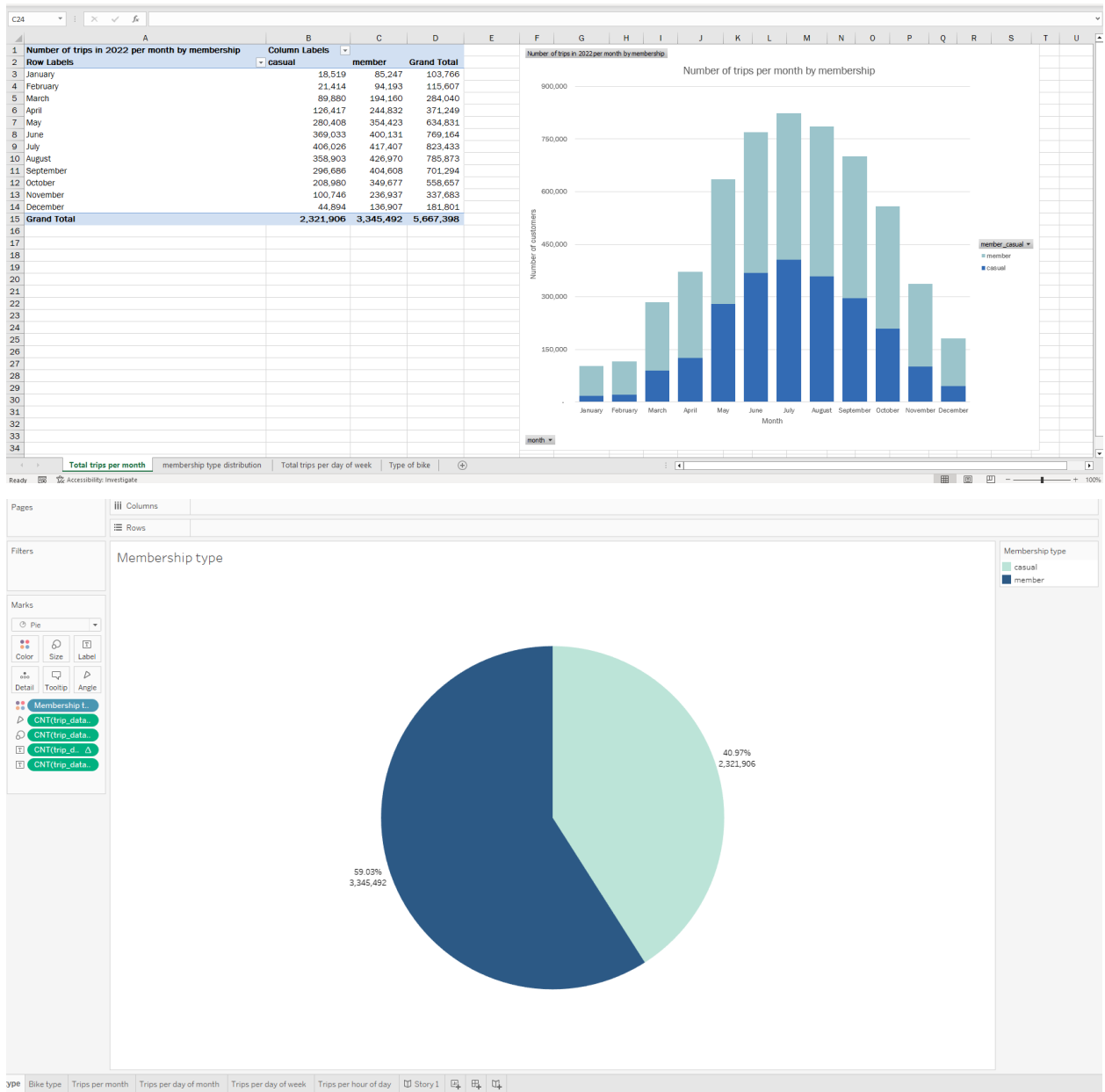
Microsoft Excel (Power Query/Pivot tables) - Data Analysis

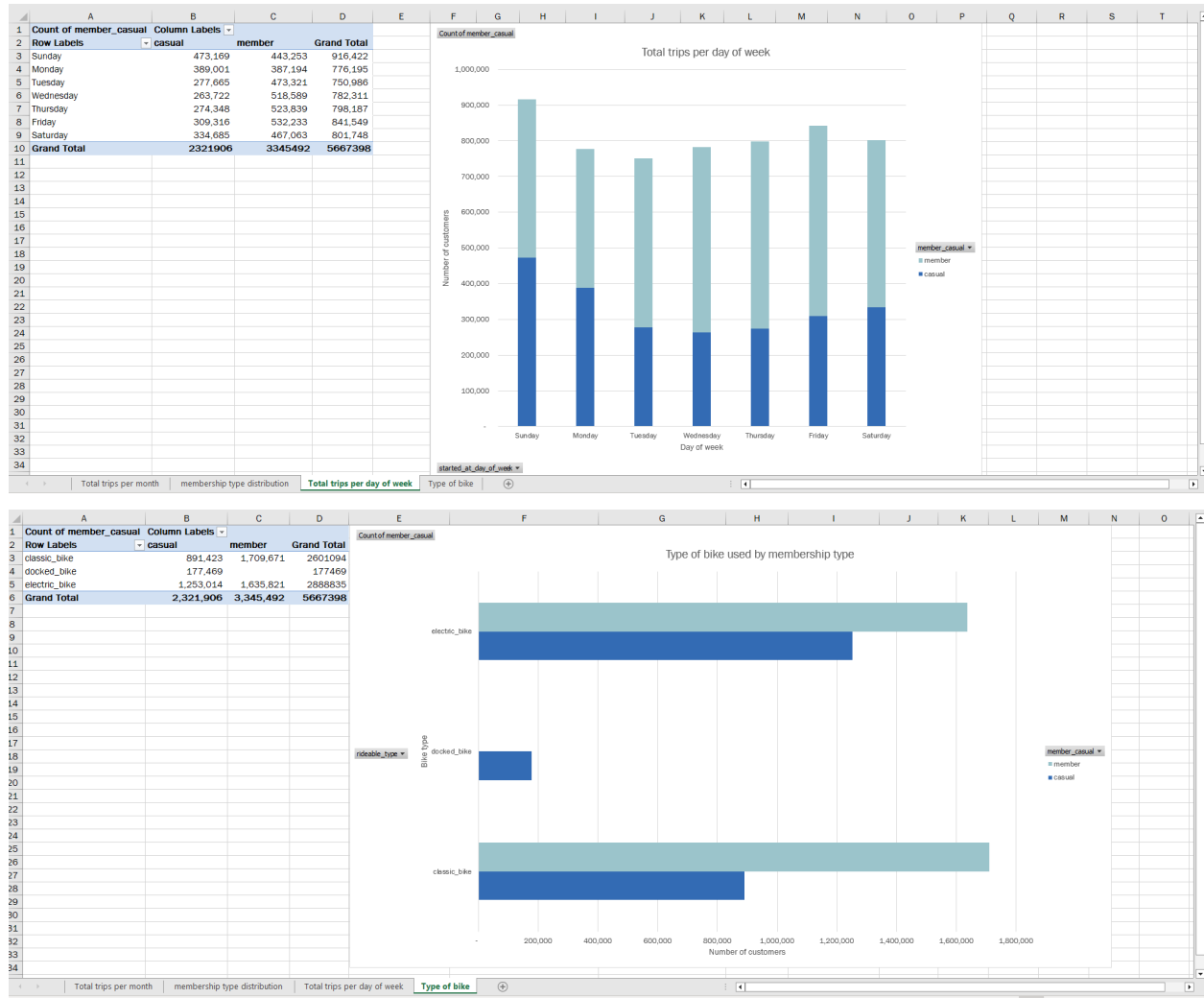
With a total of over 5 million rows, it is not feasible to merge the cleaned files into a single Excel file. Therefore, I am using Excel Power Query to combine the tables, and use pivot tables to analyse the cleaned data. With this process, I have made some initial observations of the data set and generated some ideas on what graphs I will be using in my final data visualisation dashboard.

The screenshot displays the Microsoft Excel interface with the Power Query Editor open. The editor shows a table with columns: `rideable_type`, `started_at`, `ended_at`, `member_casual`, `ride_length`, `started_at_day_of_week`, `ended_at_day_of_week`, `month`, and `started_at_date`. The table contains 12 columns and 999+ rows of data. The right-hand pane shows the 'Query Settings' for 'Cyclic Monthly Data 2022', including 'PROPERTIES' and 'APPLIED STEPS'.

Below the Power Query Editor, the Excel worksheet is visible. A PivotTable is being set up, with a dialog box titled 'PivotTable from an external source' open. The dialog box prompts the user to 'Choose Connection...' and 'Choose where you want the PivotTable to be placed'. The 'Existing Worksheet' option is selected, and the location is set to 'Sheet1!\$H\$6'. The dialog box also asks 'Choose whether you want to analyse multiple tables' and 'Add this data to the Data Model'.

The Excel ribbon is visible at the top, showing the 'PivotTable' tab. The 'Queries & Connections' pane on the right shows the list of queries, including 'Transform File from Cyclic Monthly Data 2022 [2]'.





Google BigQuery (SQL)

The Excel Power Query Table could not be directly exported to Tableau, I will be using Google cloud BigQuery to join the tables for further data visualisation processes.

I have uploaded all the cleaned csv files to BigQuery and used the UNION ALL syntax to merge all the tables from January to December.

Google Cloud | Google capstone project | Search (/) for resources, docs, products, and more

SANDBOX Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

Explorer + ADD | Q Type to search | Q Untitled 8 | RUN | SAVE | SHARE | SCHEDULE | MORE

Viewing workspace resources. SHOW STARRED ONLY

stalwart-kite-358711

External connections

trip_data

apr

aug

dec

feb

jan

jul

jun

mar

may

nov

oct

sep

trip_data_2022

```

1 SELECT *
2 FROM `stalwart-kite-358711.trip_data_jan`
3 UNION ALL
4 SELECT *
5 FROM `stalwart-kite-358711.trip_data_feb`
6 UNION ALL
7 SELECT *
8 FROM `stalwart-kite-358711.trip_data_mar`
9 UNION ALL
10 SELECT *
11 FROM `stalwart-kite-358711.trip_data_apr`
12 UNION ALL
13 SELECT *
14 FROM `stalwart-kite-358711.trip_data_may`
15 UNION ALL
16 SELECT *
17 FROM `stalwart-kite-358711.trip_data_jun`
18 UNION ALL
19 SELECT *
20 FROM `stalwart-kite-358711.trip_data_jul`
21 UNION ALL
22 SELECT *
23 FROM `stalwart-kite-358711.trip_data_aug`
24 UNION ALL
25 SELECT *
26 FROM `stalwart-kite-358711.trip_data_sep`
27 UNION ALL
28 SELECT *
29 FROM `stalwart-kite-358711.trip_data_oct`
30 UNION ALL
31 SELECT *
32 FROM `stalwart-kite-358711.trip_data_nov`
33 UNION ALL
34 SELECT *
35 FROM `stalwart-kite-358711.trip_data_dec`

```

Query results

PERSONAL HISTORY PROJECT HISTORY

SAVE RESULTS EXPLORE DATA REFRESH

Google Cloud | Google capstone project | Search (/) for resources, docs, products, and more

SANDBOX Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

Explorer + ADD | Q Type to search | Q Untitled - X | trip_data_2022 - X | QUERY | SHARE | COPY | SNAPSHOT | DELETE | EXPORT

Viewing workspace resources. SHOW STARRED ONLY

stalwart-kite-358711

External connections

trip_data

apr

aug

dec

feb

jan

jul

jun

mar

may

nov

oct

sep

trip_data_2022

SCHEMA DETAILS PREVIEW LINEAGE

Row	rideable_type	started_at	ended_at	member_casual	started_at_day_of_week	ended_at_day_of_week	month	start_date	end_date	start_hour
1	electric_bike	2022-01-01 17:55:00 UTC	2022-01-01 18:04:00 UTC	member	Sunday	Sunday	January	1	1	17
2	classic_bike	2022-01-01 02:39:00 UTC	2022-01-01 02:50:00 UTC	casual	Sunday	Sunday	January	1	1	2
3	classic_bike	2022-01-01 11:12:00 UTC	2022-01-01 11:22:00 UTC	member	Sunday	Sunday	January	1	1	11
4	electric_bike	2022-01-01 10:45:00 UTC	2022-01-01 10:46:00 UTC	member	Sunday	Sunday	January	1	1	10
5	classic_bike	2022-01-01 08:06:00 UTC	2022-01-01 08:20:00 UTC	member	Sunday	Sunday	January	1	1	8
6	electric_bike	2022-01-01 12:28:00 UTC	2022-01-01 12:26:00 UTC	member	Sunday	Sunday	January	1	1	12
7	electric_bike	2022-01-01 13:19:00 UTC	2022-01-01 13:31:00 UTC	member	Sunday	Sunday	January	1	1	13
8	dockless_bike	2022-01-01 02:19:00 UTC	2022-01-01 02:33:00 UTC	casual	Sunday	Sunday	January	1	1	2
9	electric_bike	2022-01-01 15:18:00 UTC	2022-01-01 15:21:00 UTC	member	Sunday	Sunday	January	1	1	15
10	electric_bike	2022-01-01 11:37:00 UTC	2022-01-01 12:00:00 UTC	casual	Sunday	Sunday	January	1	1	11
11	electric_bike	2022-01-01 11:56:00 UTC	2022-01-01 12:00:00 UTC	casual	Sunday	Sunday	January	1	1	11
12	electric_bike	2022-01-01 23:03:00 UTC	2022-01-01 23:17:00 UTC	member	Sunday	Sunday	January	1	1	23
13	electric_bike	2022-01-01 10:22:00 UTC	2022-01-01 10:29:00 UTC	member	Sunday	Sunday	January	1	1	10
14	classic_bike	2022-01-01 02:48:00 UTC	2022-01-01 02:52:00 UTC	member	Sunday	Sunday	January	1	1	2
15	electric_bike	2022-01-01 13:09:00 UTC	2022-01-01 13:14:00 UTC	member	Sunday	Sunday	January	1	1	13
16	dockless_bike	2022-01-01 13:32:00 UTC	2022-01-01 14:09:00 UTC	casual	Sunday	Sunday	January	1	1	13
17	electric_bike	2022-01-01 14:54:00 UTC	2022-01-01 15:04:00 UTC	casual	Sunday	Sunday	January	1	1	14
18	electric_bike	2022-01-01 13:29:00 UTC	2022-01-01 13:33:00 UTC	member	Sunday	Sunday	January	1	1	13
19	electric_bike	2022-01-01 01:36:00 UTC	2022-01-01 01:40:00 UTC	member	Sunday	Sunday	January	1	1	1
20	electric_bike	2022-01-01 02:12:00 UTC	2022-01-01 02:35:00 UTC	casual	Sunday	Sunday	January	1	1	2
21	classic_bike	2022-01-01 11:51:00 UTC	2022-01-01 12:02:00 UTC	member	Sunday	Sunday	January	1	1	11
22	classic_bike	2022-01-01 00:06:00 UTC	2022-01-01 00:09:00 UTC	member	Sunday	Sunday	January	1	1	0
23	electric_bike	2022-01-01 19:57:00 UTC	2022-01-01 01:18:00 UTC	casual	Sunday	Sunday	January	1	1	1
24	electric_bike	2022-01-01 19:50:00 UTC	2022-01-01 19:58:00 UTC	member	Sunday	Sunday	January	1	1	19
25	electric_bike	2022-01-01 00:28:00 UTC	2022-01-01 00:28:00 UTC	casual	Sunday	Sunday	January	1	1	0
26	electric_bike	2022-01-01 13:11:00 UTC	2022-01-01 13:14:00 UTC	casual	Sunday	Sunday	January	1	1	13
27	classic_bike	2022-01-01 10:46:00 UTC	2022-01-01 10:50:00 UTC	casual	Sunday	Sunday	January	1	1	10

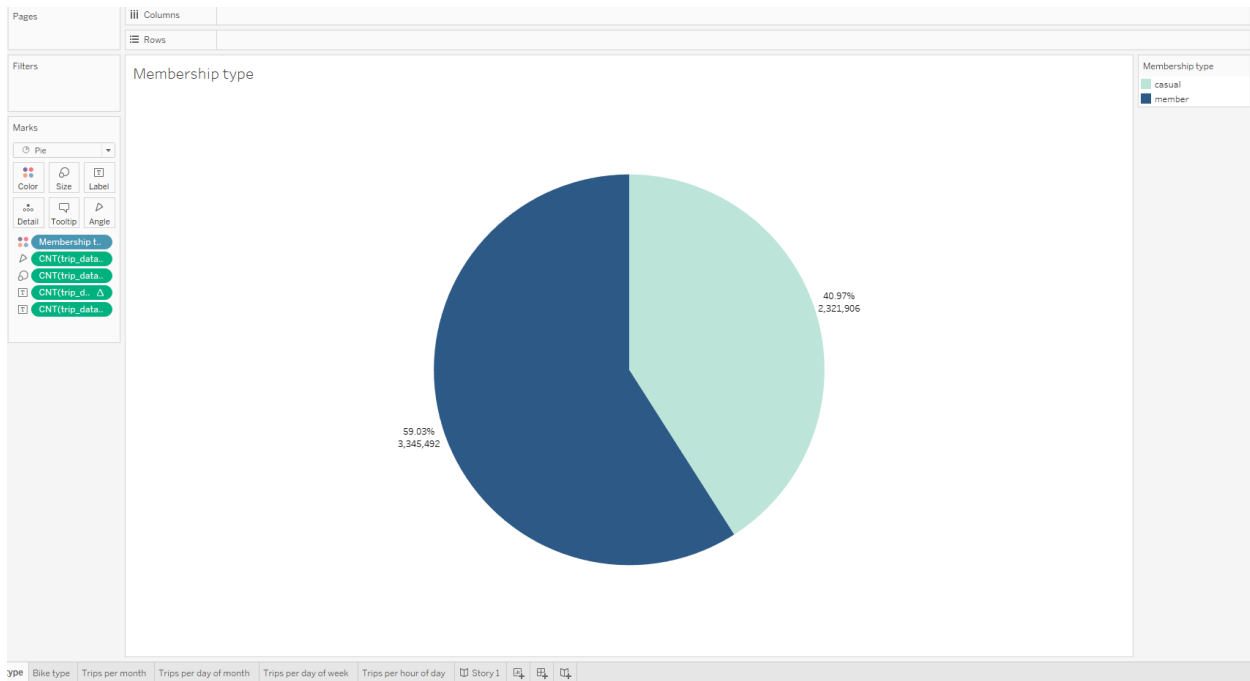
Results per page: 50 | 1 - 50 of 5667398 | REFRESH

PERSONAL HISTORY PROJECT HISTORY

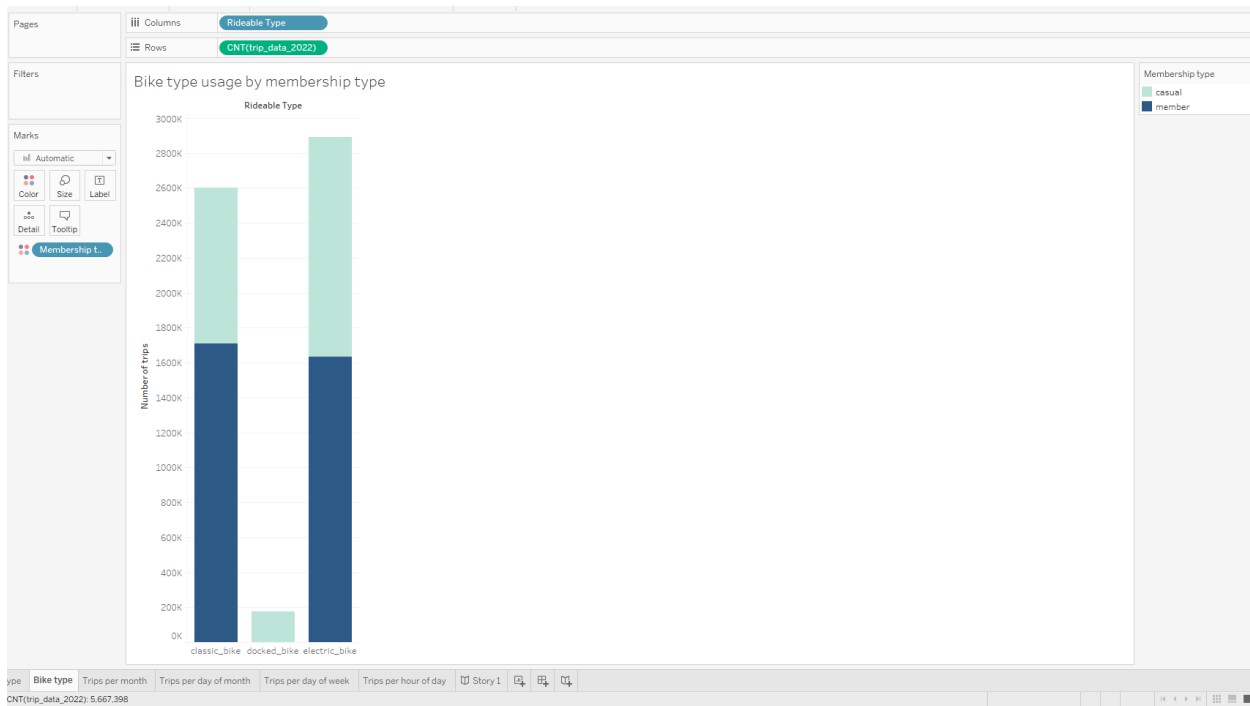
Tableau - Data Visualisation

I connected the cleaned and merged data in BigQuery to Tableau with the intention of visualising it. Similar to the excel graphs I have created, I have created visuals in Tableau to compare how annual members and casual riders use Cyclistic bikes differently.

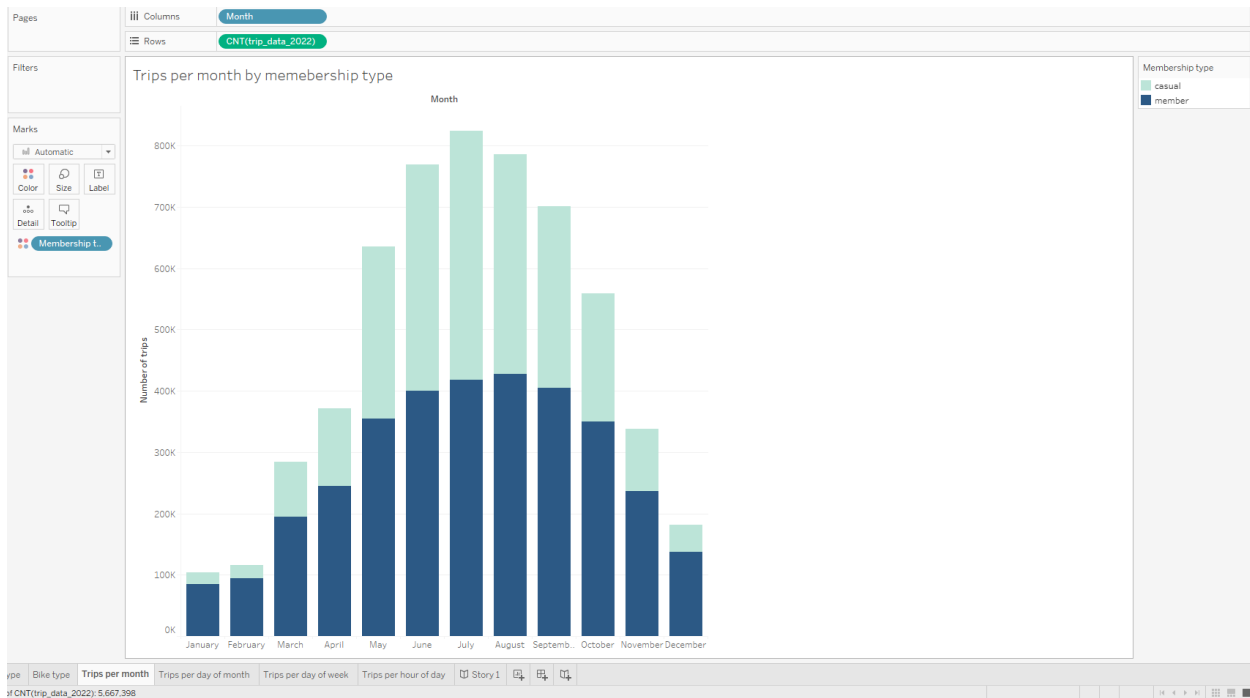
Membership type distribution



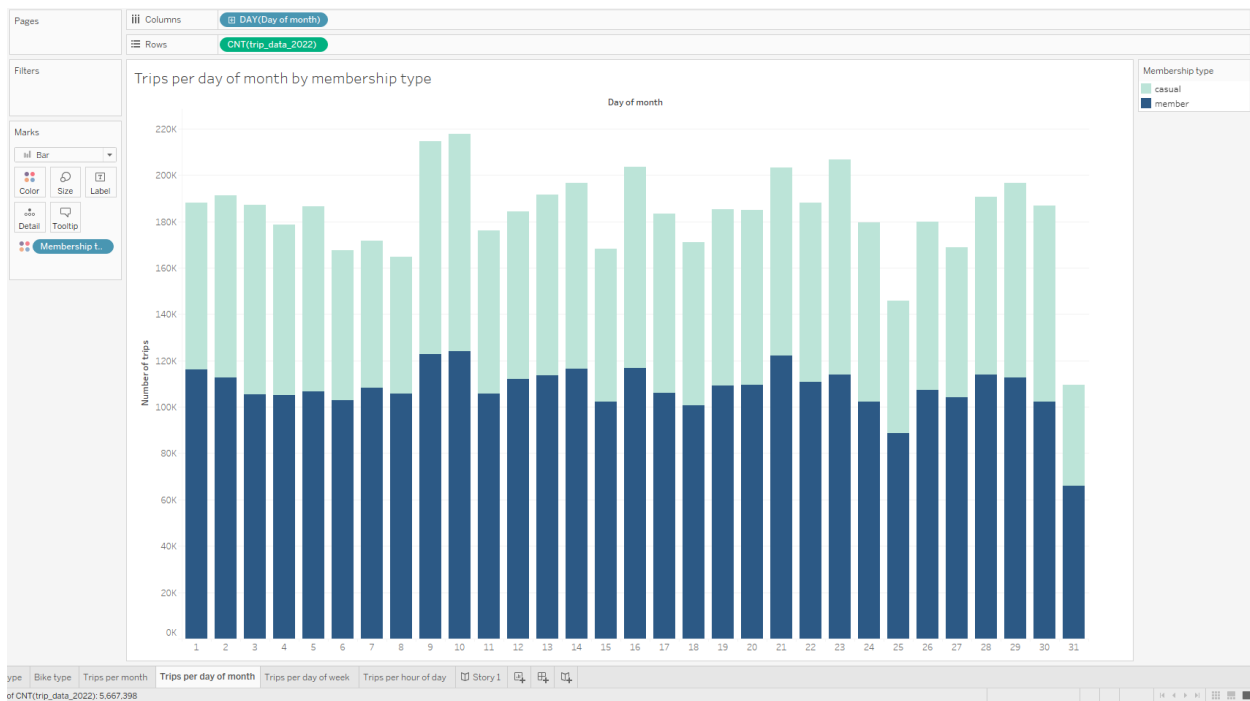
Bike type usage by membership



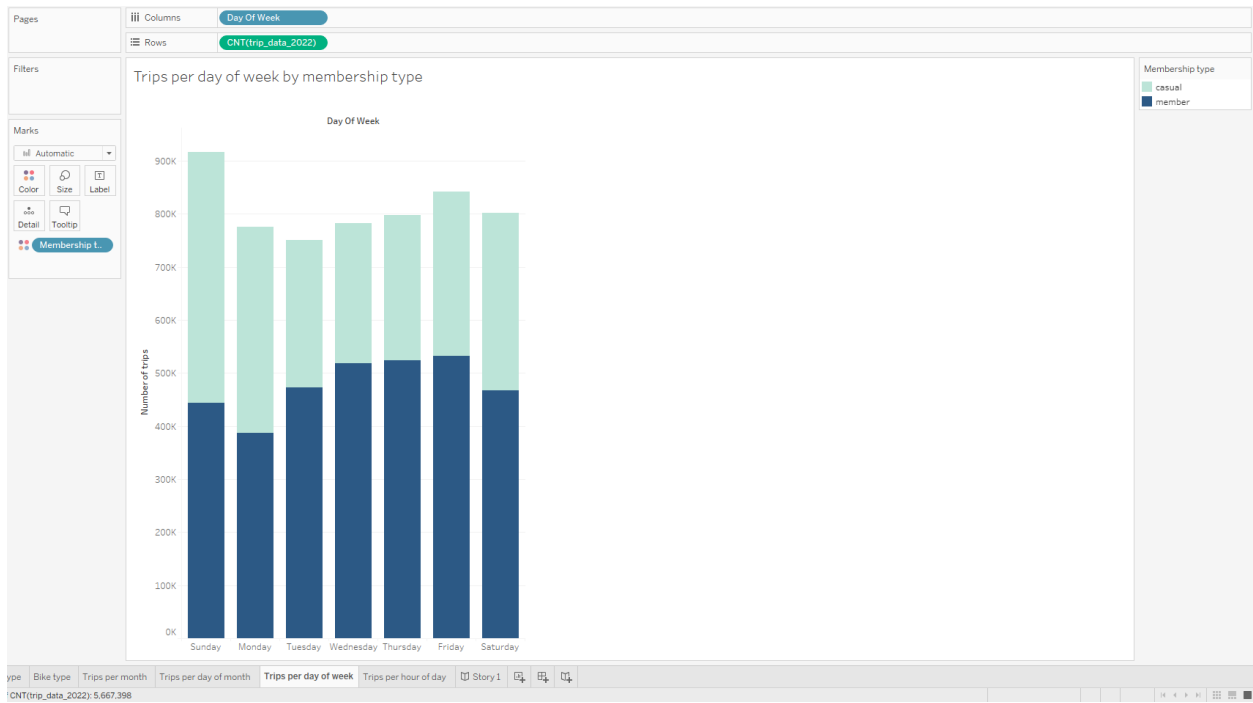
Trips per month



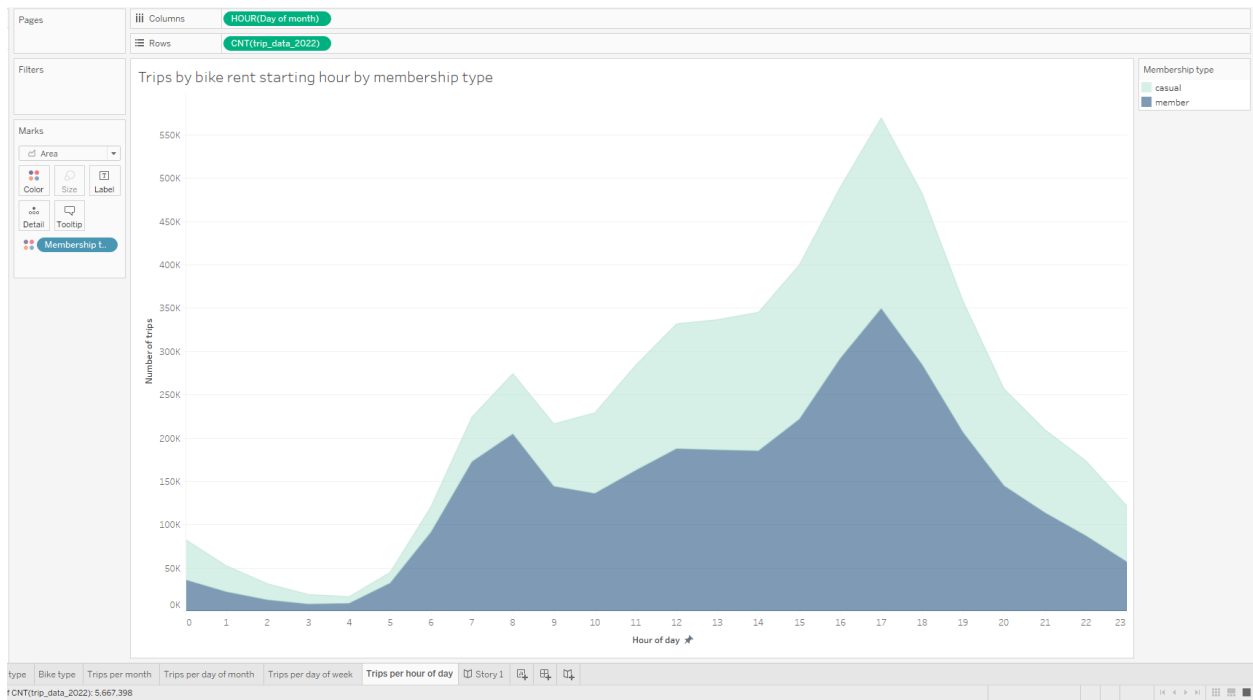
Trips per day of month



Trips per day of week



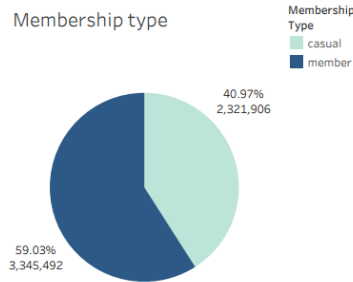
Trips per hour of day



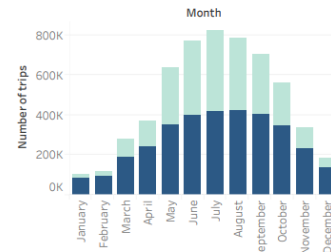
Dashboard

Cyclistic bike 2022 metrics

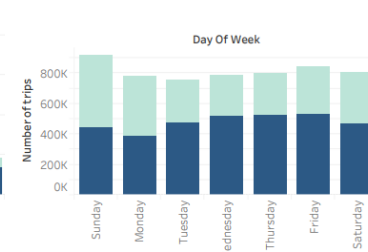
Membership type



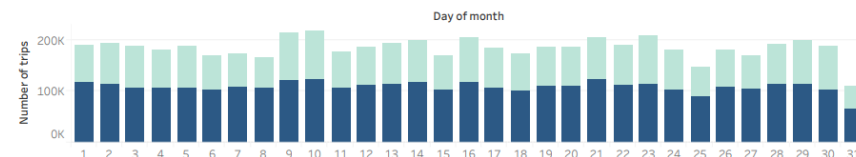
Trips per month by membership type



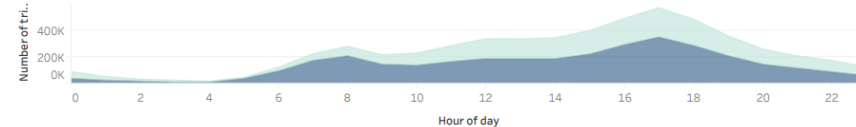
Trips per day of week by membership type



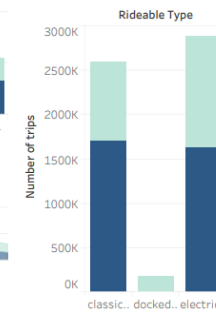
Trips per day of month by membership type



Trips by bike rent starting hour by membership type



Bike type usage by membership type



Final Summary

- Out of all trips (roughly 5.5 million in total) in 2022, 40.97% were casual users.
- The busiest months were June, July and August (Summer season) for both membership types.
- The busiest days of week were Friday, Saturday and Sunday (Weekends) for both membership types.
- Casual riders used the service the most on Sundays.
- Peak hour of day was 5 PM for both membership types. Services were most popular in the afternoon.
- The most popular bike type was electric bike.

Recommendation

Below are the recommendations for the marketing team to convert casual riders to annual members:

1. Emphasise the long term health benefits and cost saving benefits of having an annual membership over single-ride or full-day passes.
2. Run targeted advertisements on Sundays and peak months like June, July, August, when casual riders are more active and are more likely to make the conversion.
3. Offering prioritization for bicycle selection during busy hours and high demand seasons.