**Cyclistic**



**Case Study: How Does a Bike-Share Navigate Speedy Success?**

Introduction

Cyclistic is a Chicago based bicycle-sharing company. It provides the public access to geo-tracked bicycles with a fleet of over 5,800 bicycles that can be accessed at over 690 docking stations across the city. The bikes can be unlocked from one station and returned to any other station 24 hours a day.

Cyclistic’s business relies on it’s flexible pricing plans which gives customers the option to purchase 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. Along with traditional bicycles, Cyclistic offers reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike.

Analysts at Cyclistic have determined that annual members are significantly more profitable than casual riders. The company believes that acquiring more annual members will be key to its future growth. Using current customer data Cyclistic’s goal is to find insights on how to convert more casual riders into annual members.

Business Task

Design marketing strategies aimed at converting casual riders into annual members. Analyze the Cyclistic historical bike trip data to identify trends and answer the following questions:

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 Source

This Project uses Cyclistic’s 12-month historical trip data from year 2022 found [here](https://divvy-tripdata.s3.amazonaws.com/index.html). Datasets are first-party and strictly objective to refrain from the use of subjective biases.

(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 provide relevant and reliable data.)

The data has been made available by Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement).

Please note that data-privacy issues prohibit from using riders’ personally identifiable information. This limits data specifics such as names, addresses and credit card info to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

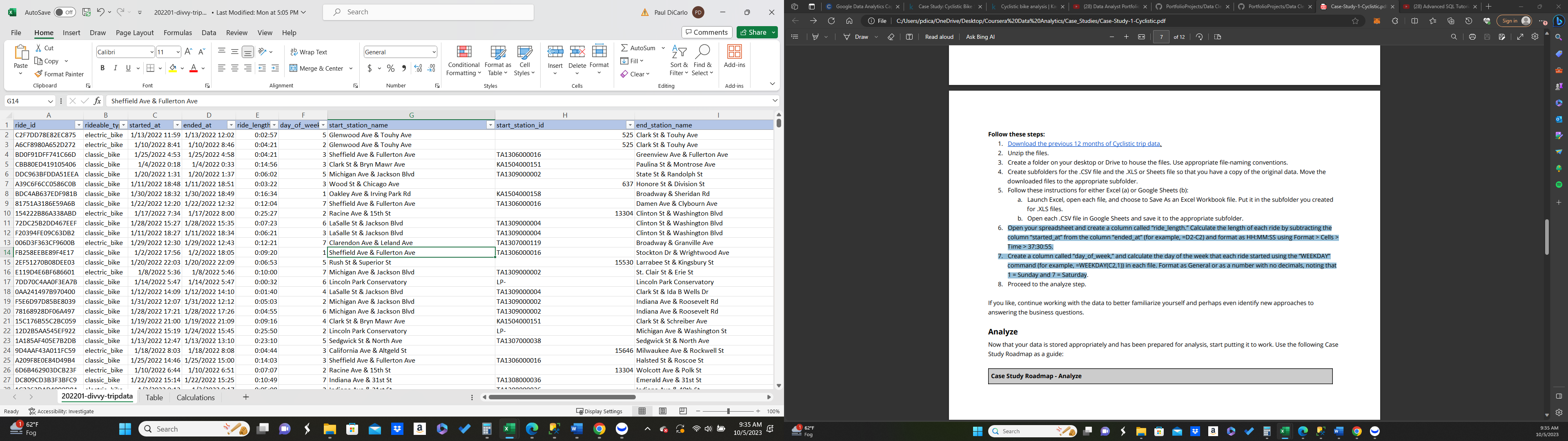
Data Process

Cyclistic’s 12-month history from year 2022 was downloaded from Motivate International Inc. in CSV format. Initial data cleaning, formatting and analysis was started in Excel to get an overall look and understanding of the data. Due to the size of the data, additional analysis was performed in SQL. For further visualizations, the findings were analyzed in Tableau.

Due to this being my first case study, analysis was limited to ride durations (ride length) and day of week. Please be advised that an updated revision may be needed to provide a more conclusive analysis.

Formatting (Excel)

Two extra columns were inserted into the spreadsheets to calculate “ride\_length” and “day\_of\_week”. The length of each ride was calculated by subtracting the column “started\_at” from the column “ended\_at” (D2-C2) and formatted as HH:MM:SS. “day\_of\_week,” calculated the day of the week that each ride started using the “WEEKDAY” command (=WEEKDAY(C2,1)) and formatted as General, noting that 1 = Sunday and 7 = Saturday.



Cleaning (Excel)

* Checked and removed any duplicates
* Checked for null values
* Checked and corrected any errors in datetime columns

For data familiarization purposes initial analysis was performed for each month using pivot tables and bar chart. Ride lengths for each month between casual riders and members stayed relatively consistent throughout the year.

|  |  |
| --- | --- |
| **Row Labels** | **Average of ride\_length** |
| **casual** | **0:32:06** |
| **member** | **0:14:00** |
| **Total Average** | **0:22:41** |

Further calculations were conducted to find the mean, max and mode.

Formulas used:

* =AVERAGE(D4:D244835)
* =MAX(D4:D244835)
* =MODE(E4:E244835)

Results for April 2022

|  |  |  |  |
| --- | --- | --- | --- |
|  | Casual |  | Member |
| mean | 0:29:32 |  | 0:11:30 |
| max | 352:02:12 |  | 24:59:58 |
| mode | 7 |  | 3 |

Analysis

Spreadsheet files were uploaded to SQL for further analysis. Please view the following link for SQL code: [PortfolioProjects/Cyclistic\_2022.sql at main · PaulDi23/PortfolioProjects (github.com)](https://github.com/PaulDi23/PortfolioProjects/blob/main/Cyclistic_2022.sql)

Table was created and monthly files were inserted per UNION ALL function.

Create TABLE Cyclistic\_2022

(

ride\_id VARCHAR(255),

rideable\_type VARCHAR(255),

started\_at VARCHAR(255),

ended\_at VARCHAR(255),

ride\_length DateTime,

day\_of\_week VARCHAR(255),

start\_station\_name VARCHAR(255),

start\_station\_id VARCHAR(255),

end\_station\_name VARCHAR(255),

end\_station\_id VARCHAR(255),

start\_lat NUMERIC,

start\_lng NUMERIC,

end\_lat NUMERIC,

end\_lng NUMERIC,

member\_casual VARCHAR(255)

);

INSERT INTO CYCLISTIC\_2022 (ride\_id, rideable\_type, started\_at, ended\_at, ride\_length, day\_of\_week,

start\_station\_name, start\_station\_id, end\_station\_name, end\_station\_id, start\_lat, start\_lng,

end\_lat, end\_lng, member\_casual)

SELECT \*

FROM PortfolioProject.dbo.Jan$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Feb$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Mar$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Apr$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.May$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Jun$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.July$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Aug$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Sep$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Oct$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Nov$

UNION ALL

SELECT \*

FROM PortfolioProject.dbo.Dec$;

Due to issues with how SQL displays and calculates time, the format for ride length was manipulated and verified back to the spreadsheets to verify and confirm time values.

------------Updated ride length format------------

-----#This was to check the ride length times matched spreadsheet data------------

SELECT CONVERT(VARCHAR(8), ride\_length, 108) AS formatted\_ride\_length

FROM cyclistic.dbo.Cyclistic\_2022

SELECT ride\_id, rideable\_type, started\_at, ended\_at, day\_of\_week, member\_casual,

CONVERT(VARCHAR(8), ride\_length, 108) AS formatted\_ride\_length

FROM cyclistic.dbo.Cyclistic\_2022

ORDER BY formatted\_ride\_length

-----------Calculated difference between start\_time and end\_time----------

SELECT DATEDIFF (MINUTE, started\_at, ended\_at) AS diff,

ride\_id

FROM PortfolioProject.dbo.Dec$

ORDER BY diff

SELECT ride\_id, started\_at, ended\_at, ride\_length, member\_casual,

DATEDIFF (MINUTE, started\_at, ended\_at)/60.00 AS time\_duration

FROM cyclistic.dbo.Cyclistic\_2022

WHERE ride\_length > '00:00:59' and started\_at < ended\_at

ORDER BY time\_duration

Max ride times, mode and rides per week per type of rider were calculated per the following. Please note that ride times under 00:59 were omitted from analysis as they may not be useful data. For mode value, casual and member riders were queried separately in the where clause.

----------#Finding max values------------------------------------------

SELECT

DATEPART(MONTH, started\_at) AS month,

member\_casual,

MAX(time\_duration) AS max\_value

FROM cyclistic.dbo.Cyclistic\_2022

GROUP BY DATEPART(YEAR, started\_at), DATEPART(MONTH, started\_at), member\_casual

ORDER BY month

-----------#Finding mode values----------------------

WITH ModeCTE AS(

SELECT

day\_of\_week AS Mode,

COUNT(\*) AS frequency,

DATEPART(MONTH, started\_at) AS month, member\_casual

FROM cyclistic.dbo.Cyclistic\_2022

WHERE member\_casual LIKE '%casual%'

GROUP BY day\_of\_week, DATEPART(MONTH, started\_at), member\_casual

),

RankedModeCTE AS(

SELECT

Mode,

frequency,

month,

member\_casual,

ROW\_NUMBER() OVER(PARTITION BY month ORDER BY frequency DESC) AS rn

FROM ModeCTE

)

SELECT

Mode,

frequency,

month,

member\_casual

FROM RankedModeCTE

WHERE rn = 1

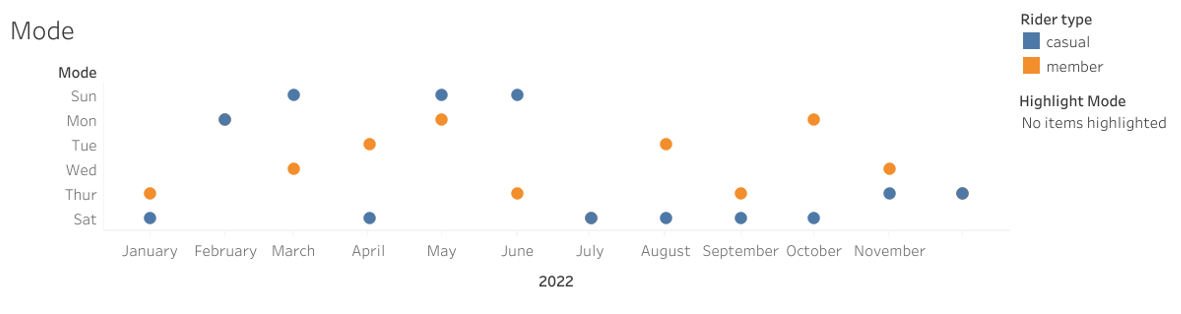
ORDER BY month

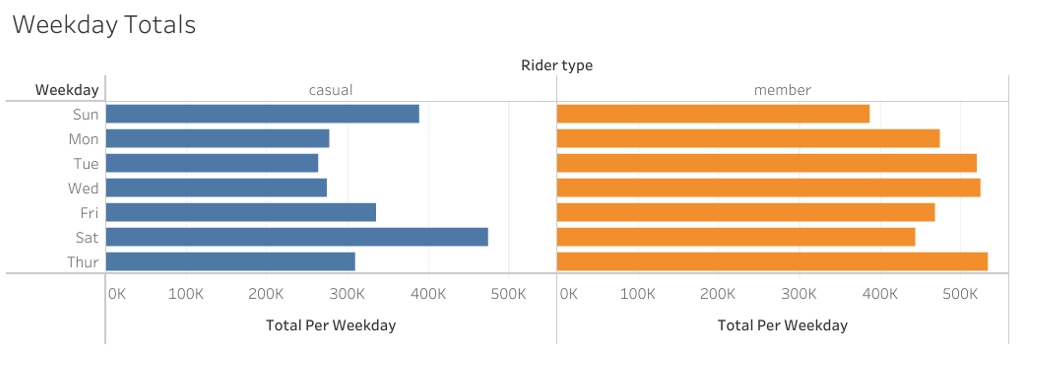
Summary of SQL Analysis

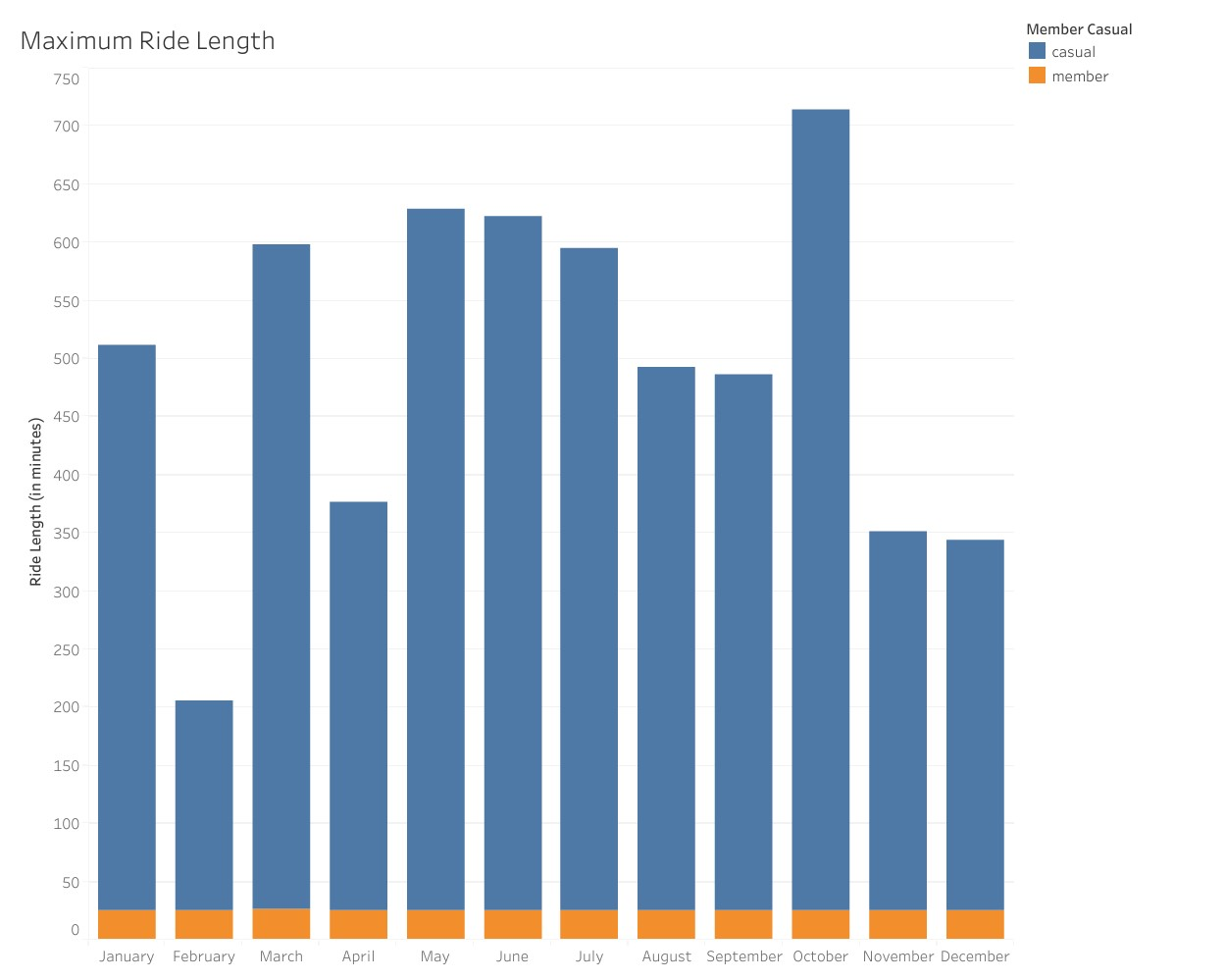
* Max ride times varied greatly between casual and member riders. On average, casual riders’ maximum rides per month were roughly 469 minutes while members were only 25. (Please note that the data received showed members with max ride times never exceeding 24.59 hours. The accuracy of the data may be flawed and questionable.)
* Modes for the day of week differed between casual and member riders as casual riders were mostly active on Saturdays and Sundays, while members were mostly active during the week.
* Broken down into day of the week, members made up most of the rides Monday through Friday, while casual riders made up slightly more rides on the weekend (sat & sun). This shows that casual riders tend to ride more on the weekends than during the week.

Please find visualizations per the following Tableau Public link: [Cyclistic\_2022 | Tableau Public](https://public.tableau.com/app/profile/paul.dicarlo/viz/Cyclistic_2022_16968196473680/Dashboard1#2)

SQL findings were uploaded to Tableau Public and the following visualizations created.







Findings & Recommendations

* The analysis shows that the behavior of casual riders utilizes Cyclistic bicycles more often on the weekend while members are more active on weekdays and for lesser periods of time. This is an indication that casual riders do not use Cyclistic for work transportation but rather for leisure, as well as indicating they may be a younger market, not yet in the work force.
* As Saturdays and Sundays are the most common days for casual riders, having special promotions for weekday rides may gather interest in casual riders to use Cyclistic more often during the week.
* Casual Riders tended to rent bicycles for longer periods of time using either single-ride or full day passes. Due to the lengthy rides of casual riders, advertising the possible cost savings by having an annual membership, may entice causal riders to obtain an annual membership.
* Targeting casual riders through social media, email and website promotions offering savings and exclusive deals for first time annual memberships.