

Cylistic Case Study

By Daniel Timms

Contents

Step 1 - Ask	3
Business Task	3
Key Stakeholders	3
Step 2 - Prepare	3
Data Location	3
Data Organization	3
Data Credibility	3
Licensing, Privacy, Security, and Accessibility	3
Notes	4
Step 3 – Process	5
Tool	5
Notes	5
Data Cleaning	5
Step 4 – Analyze.....	8
Notes	8
Queries	8
Step 5 – Share	13
Notes	13
Charts and Findings	13
Step 6 – Act	16
Recommendation	16
Self-Reflection	16

This is the first case study that I have done. I also had to complete it within 2 days due to some time constraints. If there's anything that I can improve on, do reach out!

Context

The director of marketing believes the company's future success depends on maximizing the number of annual memberships.

Your team wants to understand how casual riders and annual members use Cyclistic bikes differently.

Your team will design a new marketing strategy to convert casual riders into annual members.

Moreno has assigned you the first question to answer: *How do annual members and casual riders use Cyclistic bikes differently?*

Step 1 - Ask

Business Task

To understand how casual riders and annual members use Cyclistic bikes differently. This is to design a new marketing strategy to convert casual riders into annual members.

Key Stakeholders

Lily Moreno: The Director of Marketing & my manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program.

Cyclistic Executive Team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

Step 2 - Prepare

Data Location

I'll be using data from <https://divvy-tripdata.s3.amazonaws.com/index.html>. I'll be focusing on the data from 2023.

Data Organization

The data is organized each month every year in a zip file containing a CSV file.

Data Credibility

The data **IS** credible. It is: -

- Reliable : The data is accurate and complete.
- Original : The data is validated.
- Comprehensive : The data contains critical information needed.
- Current : The data is up to date.
- Cited : The data is cited from a credible primary source.

Licensing, Privacy, Security, and Accessibility

The data used is made available by Motive International Inc. Their data license agreement can be viewed here: <https://divvybikes.com/data-license-agreement>.

Notes

- 1 Downloaded Jan23 to Dec23 Files. Saved the CSV files into a folder.
- 2 I combined the CSV files into one with Power Query: <https://www.ablebits.com/office-addins-blog/merge-multiple-csv-files-excel/>
- 3 I found that the dataset contained 14 columns and 5,719,877 rows. I made sure there were no duplicate data in the ride_id column as it is unique.
Select Column > Select Home > Conditional Formatting > Highlight Cells Rules > Duplicate Values
Data > Remove Duplicates > Select Column
- 4 I then deleted entries that had blank rows for start_station_id followed by end_station_id.
- 5 I then selected the columns that I felt were important for analysis and moved it to a new sheet: Ride_ID, Rideable_Types, Member_Casual.
- 6 At this point, I realized that I should use BigQuery instead of Excel due to the large number of rows.

Step 3 – Process

Tool

I've decided to use BigQuery due to the large number of rows.

Notes

1. I created a new project in BigQuery and named it Cyclistic Case Study.
2. In the new project, I created a dataset which I named Jan23.
3. In the dataset Jan23, I created and uploaded 12 tables. (Jan23 to Dec23)
4. Due to the limitation of my free BigQuery account, I was unable to upload May to October files. So, I decided to only use 6 months' worth of data: Jan, Feb, March, April, November, December.
5. Once I uploaded all the files, I merged them into one table using UNION.
6. Once merged, I found that there were rows with missing values. So, I cleaned them up by creating a new table "6MonthDataCLEANED".
7. I then checked for duplicate ride_id data.
8. Next, I thought of the data I wanted to share: Casual Riders vs Annual Riders, Rides by bike types, Ride duration, Rides a day.

Data Cleaning

Merging data into one table

```
CREATE TABLE `cyclistic-case-study-416806.Jan23.6MonthData` AS
SELECT * FROM `cyclistic-case-study-416806.Jan23.Jan23`
UNION ALL
SELECT * FROM `cyclistic-case-study-416806.Jan23.Feb23`
UNION ALL
SELECT * FROM `cyclistic-case-study-416806.Jan23.Mar23`
UNION ALL
SELECT * FROM `cyclistic-case-study-416806.Jan23.Apr23`
UNION ALL
SELECT * FROM `cyclistic-case-study-416806.Jan23.Nov23`
UNION ALL
SELECT * FROM `cyclistic-case-study-416806.Jan23.Dec23`
```

Result

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH			
Row	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id		
1	893685201A727C27	electric_bike	2023-02-14 10:43:35 UTC	2023-02-14 10:54:36 UTC	null	null	null	null		
2	ED7FEA75B71F23C7	electric_bike	2023-03-09 20:18:19 UTC	2023-03-09 21:07:43 UTC	null	null	Public Rack - Washtenaw Ave & ...	930		
3	B71E0BA8E1C83EC	electric_bike	2023-02-07 14:36:38 UTC	2023-02-07 14:41:59 UTC	null	null	null	null		
4	7B577C617059B23F	electric_bike	2023-11-18 22:27:14 UTC	2023-11-18 22:30:03 UTC	null	null	null	null		
5	3EB210AC680C36B2	electric_bike	2023-04-10 13:25:11 UTC	2023-04-10 13:39:24 UTC	null	null	Public Rack - McCormick Blvd ...	488		
6	6A12A397063DAB9A	electric_bike	2023-11-10 16:15:48 UTC	2023-11-10 16:20:37 UTC	null	null	null	null		
7	C269778E0CF033A8	electric_bike	2023-11-13 18:54:31 UTC	2023-11-13 18:57:59 UTC	null	null	null	null		
8	71173099CD1D217F	electric_bike	2023-12-28 22:01:43 UTC	2023-12-28 22:17:15 UTC	null	null	null	null		
9	2287EFA3E8013A7	electric_bike	2023-12-03 12:42:05 UTC	2023-12-03 13:04:06 UTC	null	null	null	null		

Results per page: 501 - 50 of 165265

Removing missing values

```
CREATE TABLE `cyclistic-case-study-416806.Jan23.6MonthDataCLEANED` AS
SELECT * FROM `cyclistic-case-study-416806.Jan23.6MonthData`
WHERE NOT (start_station_id IS NULL OR end_station_name IS NULL OR end_station_id IS NULL)
```

Result

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH		
Row	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	
1	79054A74E860E22E	electric_bike	2023-12-05 14:29:25 UTC	2023-12-05 14:58:35 UTC	Menard Ave & Division St	305	Damen Ave & Chicago Ave	13132	
2	94B27028A5FDA350	electric_bike	2023-04-21 08:39:42 UTC	2023-04-21 08:40:03 UTC	Major Ave & Bloomingdale Ave	308	Major Ave & Bloomingdale Ave	308	
3	98EE0C476A7F1058	electric_bike	2023-11-21 11:20:06 UTC	2023-11-21 11:27:41 UTC	Narragansett & McLean	309	Menard Ave & North Ave	376	
4	454B78377849BF41	electric_bike	2023-04-14 18:27:33 UTC	2023-04-14 18:27:53 UTC	Lockwood Ave & Wrightwood A...	312	Lockwood Ave & Wrightwood A...	312	
5	B48E0CB2DEE13C3A	electric_bike	2023-04-02 14:02:42 UTC	2023-04-02 14:17:04 UTC	Long Ave & Belden Ave	313	Public Rack - Cicero Ave & Wrig...	547	
6	5B3190B49FB57681	electric_bike	2023-12-09 21:07:00 UTC	2023-12-09 21:08:38 UTC	Long Ave & Belden Ave	313	Long Ave & Belden Ave	313	
7	BC2527898A2613C	electric_bike	2023-12-19 21:39:22 UTC	2023-12-19 21:42:48 UTC	Narragansett & Wrightwood	314	Public Rack - Austin Ave & Wrig...	565	
8	253F619BA27BA955	electric_bike	2023-12-18 21:41:13 UTC	2023-12-18 21:44:14 UTC	Narragansett & Wrightwood	314	Public Rack - Austin Ave & Wrig...	565	
9	8D8B1012173C93E1	electric_bike	2023-04-23 22:23:29 UTC	2023-04-23 22:37:44 UTC	Narragansett & Wrightwood	314	Narragansett & Wrightwood	314	

Results per page: 501 - 50 of 1264429

Check for duplicates

```
SELECT ride_id, COUNT(*)
FROM `cyclectic-case-study-416806.Jan23.6MonthDataCLEANED`
GROUP BY ride_id
HAVING COUNT(*) > 1
```

Result

Query results

JOB INFORMATION

RESULTS

CHART

JSON

EXECUTION DETAILS

EXECUTION GRAPH

SAVE RESULTS

EXPLORE DATA

There is no data to display.

Step 4 – Analyze

Notes

1. I decided to create a new table called “6MonthDataDateTime”. The table was going to contain extracted month/time values.
2. At this point, I realized that there were certain columns that I didn’t need to create.
3. I decided to find the min, avg, & max for difference in minutes.
4. I found that the shortest ride value for negative. This meant that there were some trips that lasted more than a day. I decided to investigate.
5. I then decided to create a new table “6MonthDataDateTimeCleaned”.
6. Next, I decided to calculate the number of rides each day separated by member or casual.
7. Next, I decided to get the rides by bike type.
8. Next, I decided to get the average length of trip separated by member or casual.
9. Finally, I decided to get the number of members vs casuals.
10. After every query, I exported it into a CSV file.

Queries

Creating New Table

```
CREATE TABLE `cyclistic-case-study-416806.Jan23.6MonthDataDateTime` AS
SELECT rideable_type,
EXTRACT (MONTH FROM started_at) AS month_start,
EXTRACT (TIME FROM started_at) AS time_start,
EXTRACT (DAYOFWEEK FROM started_at) AS day,
EXTRACT (MONTH FROM ended_at) AS month_end,
EXTRACT (TIME FROM ended_at) AS time_end,
TIMESTAMP_DIFF(ended_at, started_at, second) AS difference_sec,
TIMESTAMP_DIFF(ended_at, started_at, minute) AS difference_min, start_station_name,
end_station_name, member_casual,
FROM `cyclistic-case-study-416806.Jan23.6MonthDataCLEANED`
```

Result

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH				
Row	rideable_type	month_start	time_start	day	month_end	time_end	difference_sec	difference_min	start_station_name	end_station_name	
1	classic_bike	1	17:54:20	6	1	17:58:36	256	4	Ellis Ave & 60th St	University Ave & 57th St	
2	classic_bike	1	20:14:25	5	1	20:18:41	256	4	Ellis Ave & 60th St	Ellis Ave & 55th St	
3	electric_bike	1	10:03:59	4	1	10:08:15	256	4	N Green St & W Lake St	Desplaines St & Jackson Blvd	
4	classic_bike	1	15:07:49	4	1	15:12:05	256	4	Wolcott (Ravenswood) Ave & ...	Clark St & Montrose Ave	
5	electric_bike	1	13:39:30	4	1	13:43:46	256	4	Clifton Ave & Armitage Ave	Larrabee St & Armitage Ave	
6	classic_bike	1	22:17:08	7	1	22:21:24	256	4	Ogden Ave & Race Ave	Noble St & Milwaukee Ave	
7	electric_bike	1	17:27:26	3	1	17:31:42	256	4	Clark St & Grace St	Pine Grove Ave & Irving Park Rd	
8	classic_bike	1	17:39:11	7	1	17:43:27	256	4	Sedgwick St & North Ave	Halsted St & Willow St	
9	electric_bike	1	11:36:57	4	1	11:41:13	256	4	Peoria St & Jackson Blvd	Franklin St & Adams St (Temp)	

Results per page: 501 - 50 of 1264429<>>

Shortest Ride, Average Ride, Longest Ride

```
SELECT MIN(difference_min) as Shortest, AVG(difference_min) as Average,  
MAX(difference_min) as Highest  
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTime`
```

Result

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	Shortest ▾	Average ▾	Highest ▾	
1	-54	12.38940581084...	2349	

Negative Rides & Rides more than a day

```
SELECT COUNT (*) as rides_longer_than_day  
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTime`  
WHERE difference_min > 1440
```

```
SELECT COUNT (*) as rides_negative  
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTime`  
WHERE difference_min < 0
```

Result

Query results	
JOB INFORMATION	
Row	rides_longer_than_day
1	34

Query results	
JOB INFORMATION	
Row	rides_negative
1	31

Creating the Table 6MonthDataDateTimeCleaned

```
CREATE TABLE `cyclistic-case-study-416806.Jan23.6MonthDataDateTimeCleaned` AS
SELECT *
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTime`
WHERE difference_min > 0 AND difference_min < 1440
```

Rides each day

```
SELECT COUNT(rideable_type) as rides,
CASE WHEN day = 1 THEN 'Sunday'
      WHEN day = 2 THEN 'Monday'
      WHEN day = 3 THEN 'Tuesday'
      WHEN day = 4 THEN 'Wednesday'
      WHEN day = 5 THEN 'Thursday'
      WHEN day = 6 THEN 'Friday'
      WHEN day = 7 THEN 'Saturday' END AS day_of_the_week, member_casual,
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTimeCleaned`
GROUP BY member_casual, day
ORDER BY member_casual
```

Result

Row	rides	day_of_the_week	member_casual
1	47595	Sunday	casual
2	34323	Monday	casual
3	39368	Tuesday	casual
4	44187	Wednesday	casual
5	48925	Thursday	casual
6	48073	Friday	casual
7	59796	Saturday	casual
8	94186	Sunday	member
9	126441	Monday	member
10	147691	Tuesday	member
11	156204	Wednesday	member
12	154223	Thursday	member
13	128693	Friday	member
14	104282	Saturday	member

Rides by bike type

```
SELECT rideable_type, COUNT(rideable_type) as num
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTimeCleaned`
GROUP BY rideable_type
```

```
SELECT rideable_type, COUNT(rideable_type) as num, member_casual
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTimeCleaned`
GROUP BY rideable_type, member_casual
```

Result

Row	rideable_type	num
1	classic_bike	725389
2	electric_bike	493306
3	docked_bike	15292

Row	rideable_type	num	member_casual
1	classic_bike	567536	member
2	electric_bike	149122	casual
3	electric_bike	344184	member
4	classic_bike	157853	casual
5	docked_bike	15292	casual

Average Trip Length

```
SELECT CASE
  WHEN day = 1 THEN 'Sunday'
  WHEN day = 2 THEN 'Monday'
  WHEN day = 3 THEN 'Tuesday'
  WHEN day = 4 THEN 'Wednesday'
  WHEN day = 5 THEN 'Thursday'
  WHEN day = 6 THEN 'Friday'
  WHEN day = 7 THEN 'Saturday' END AS day_of_the_week, ROUND(AVG(difference_min),2)
AS avg_ride_length, member_casual
FROM `cyclistic-case-study-416806.Jan23.6MonthDataDateTimeCleaned`
GROUP BY member_casual, day
ORDER BY member_casual, day
```

Result

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	day_of_the_week ▼	avg_ride_length ▼	member_casual ▼		
1	Sunday	22.26	casual		
2	Monday	16.7	casual		
3	Tuesday	16.01	casual		
4	Wednesday	16.24	casual		
5	Thursday	16.91	casual		
6	Friday	18.3	casual		
7	Saturday	22.07	casual		
8	Sunday	11.67	member		
9	Monday	9.97	member		
10	Tuesday	10.15	member		
11	Wednesday	10.23	member		
12	Thursday	10.33	member		
13	Friday	10.34	member		
14	Saturday	11.77	member		

Member vs Casual

```
SELECT COUNT(rideable_type) AS rides, member_casual
FROM `cyclictic-case-study-416806.Jan23.6MonthDataDateTimeCleaned`
GROUP BY member_casual
```

Result

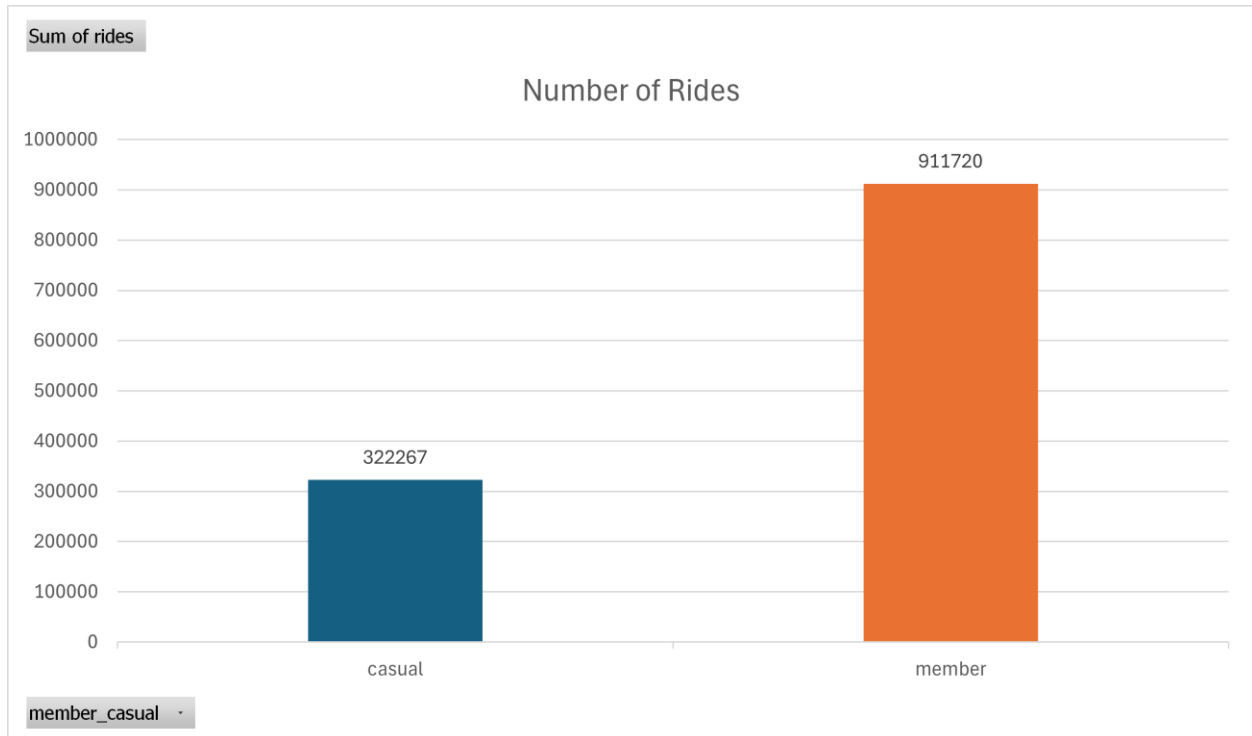
Row	rides ▼	member_casual ▼
1	911720	member
2	322267	casual

Step 5 – Share

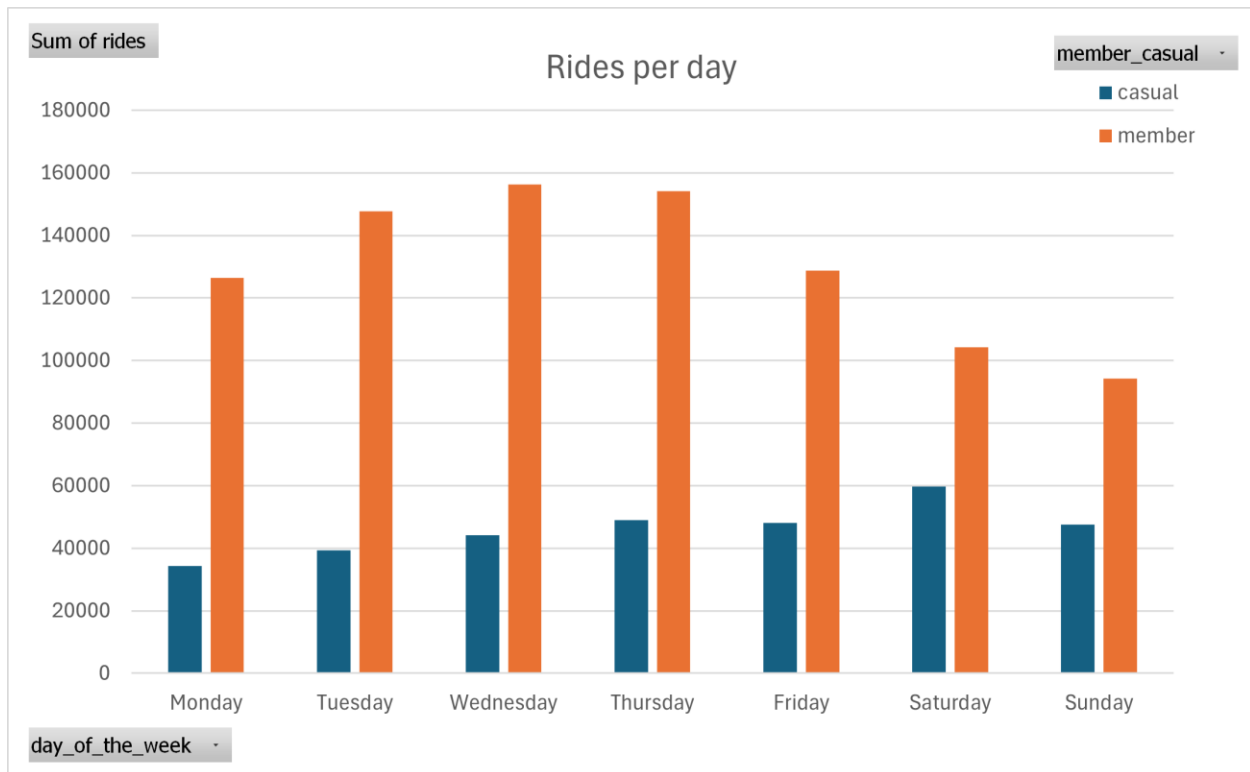
Notes

1. For visualizations, I decided to use Excel as I was having some issues with RCloud. I will use R in my next case study!

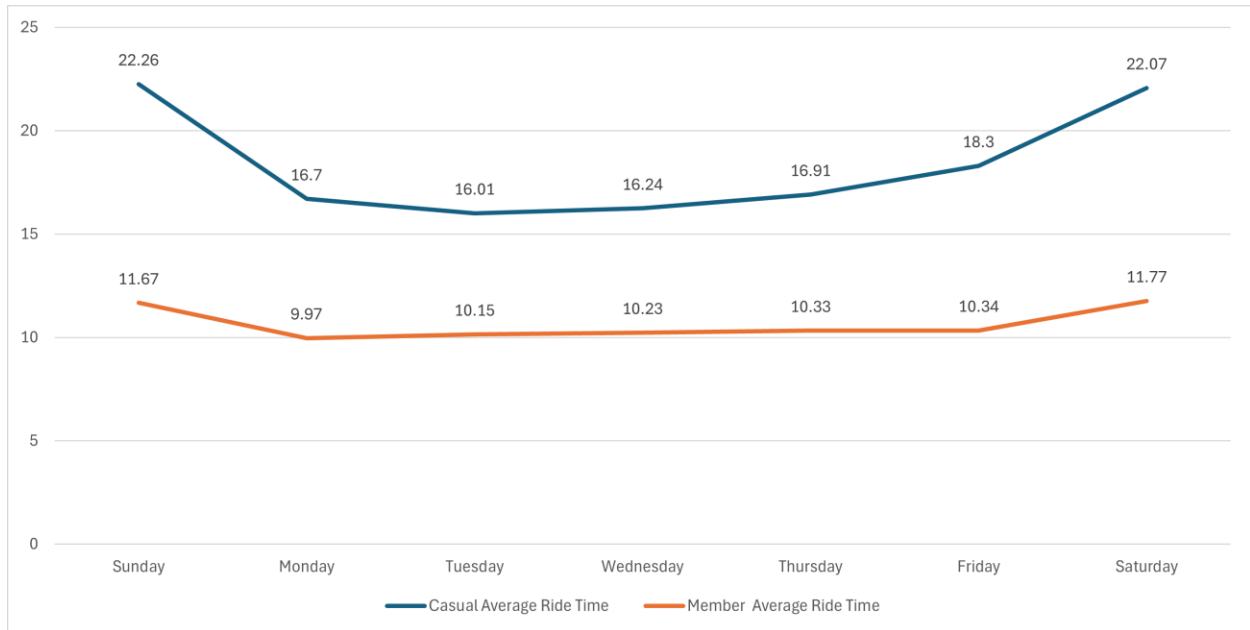
Charts and Findings



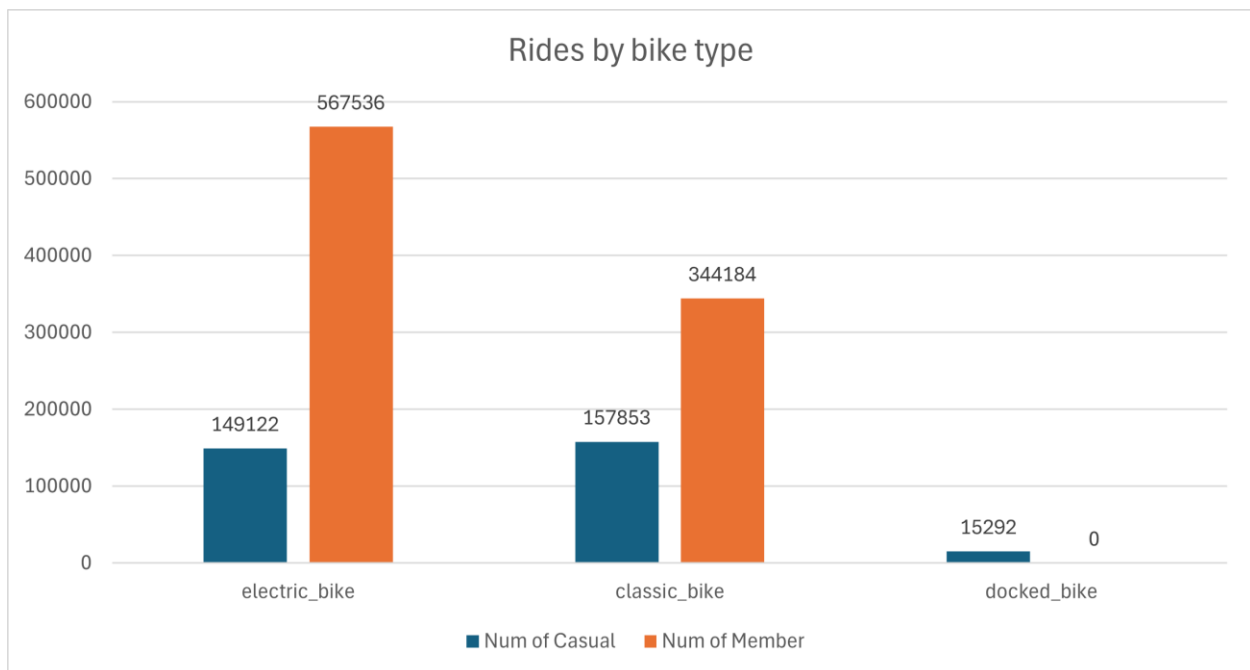
I found that more rides were taken by members. Out of the 1,233,987 rides in those 6 months, only 26% were taken by casuals.



I then found out that casuals were more likely to take rides towards the end of the week. Saturdays in particular were the busiest days for casuals with almost 60,000 rides. On the other hand, members seem to be the busiest on weekdays. From this I can hypothesize that members are most likely riding to get to work whereas casuals use the bikes for leisure.



The average ride time chart further proves my hypothesis as casuals seem to spend more time riding the bikes compared to members. It's also shown that casuals spend more time riding on the weekends compared to weekdays.



Finally, electric bikes seem to be the more popular option for members compared to the classic bike. On the other hand, casuals seem to not mind both electric and classic bikes. Docked bikes, however, seem to be the least popular option for both groups.

Step 6 – Act

Recommendation

1. Target Weekend Riders with Membership Offers: Since casual riders are more active on weekends, tailor promotional offers to encourage them to sign up for membership. Highlight the benefits of membership for weekend leisure activities, such as unlimited rides or discounted rates for weekend use.
2. Create Engaging Content: Develop marketing campaigns and content that showcase the experiences and benefits of being a member. Use testimonials or stories from existing members who use bikes for commuting, leisure, or fitness, illustrating how membership enhances their lifestyle.

Self-Reflection

- When analyzing it I should've looked at the routes taken as it would've been helpful.
- I should stop rushing through my work and think it through.
- When analyzing, I should've looked at the time of rides.