Cyclistic Bike-Share Case Study How Annual Members and Casual Riders Use Cyclistic Bikes Differently

Name: Abdalrhman Dabor

Date: 12/6/2025

1. Business Task

- **Statement of the Business Task**
  As a junior data analyst on the Cyclistic marketing analytics team, I have been tasked to answer the following business question:
- ***how casual riders and annual members use Cyclistic bikes differently?***

The goal is to get insights based on the riders data, which will help my team to design a new marketing strategy to convert casual riders into annual members.

- **Stakeholders**

  - Lily Moreno (Director of Marketing)

  - Cyclistic Marketing Analytics Team

  - Cyclistic Executive Team.

2. Data sources

- **Data Used**

  I used data that represent 1 year time frame to make the analyzing process more accurate. The data that was used:

- 202410-divvy-tripdata
- 202411-divvy-tripdata
- 202412-divvy-tripdata
- 202501-divvy-tripdata
- 202502-divvy-tripdata
- 202503-divvy-tripdata
- 202504-divvy-tripdata
- 202505-divvy-tripdata
- 202506-divvy-tripdata
- 202507-divvy-tripdata
- 202508-divvy-tripdata
- 202509-divvy-tripdata
- 202510-divvy-tripdata

These datasets are public trip records provided by the City of Chicago and Motivate International Inc. under the [Divvy Data License Agreement](https://www.divvybikes.com/data-license-agreement) (open data for non-commercial use). Note that the files are named "Divvy" instead of "Cyclistic" because Cyclistic is a fictional company in this case study, but the data is equivalent and suitable for analysis.

- **Data Organization and Integrity (ROCCC Evaluation):**

  - ***Reliable***: Sourced from official Divvy records (City of Chicago), with consistent structure across files.

  - ***Original***: First-hand trip logs from the bike-share system.

  - ***Comprehensive***: Covers 5–6 million trips, including start/end stations, bike types, and user types (no personally identifiable information for privacy).

  - ***Current***: Latest available data up to October 2025.

  - ***Cited***: All files from https://divvy-tripdata.s3.amazonaws.com/index.html.

- **Privacy and Accessibility Notes**

  - No personally identifiable information

  - Data is freely accessible.

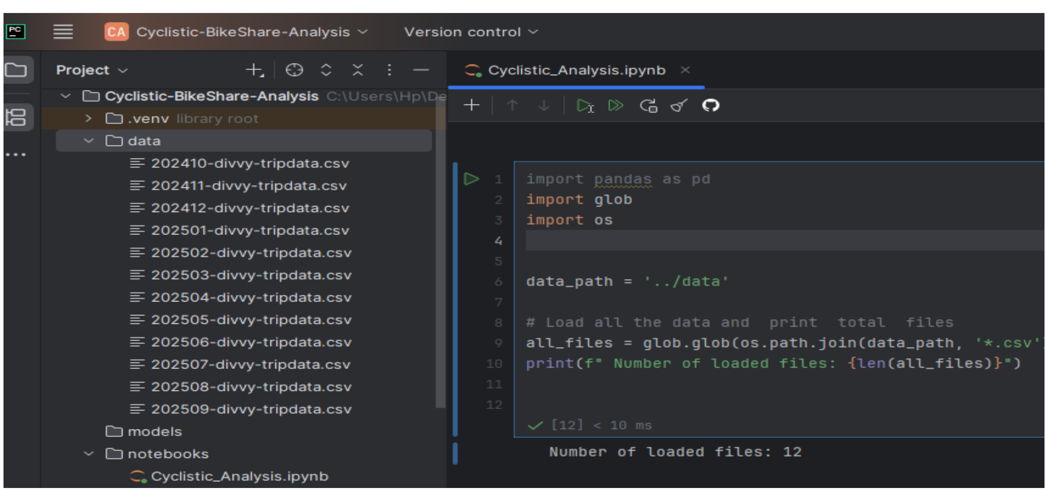


Figure 1: Showing 12 CSV files loaded

3. Data Cleaning & Preparation

- **Tools Used**

  I used pandas - Python Data Analysis Library in Jupyter Notebook

- **Initial Data Overview**

  The data contains 5,539,521 rows, which represent the total trips over a year.

  There are also 13 Columns: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, member_casual.

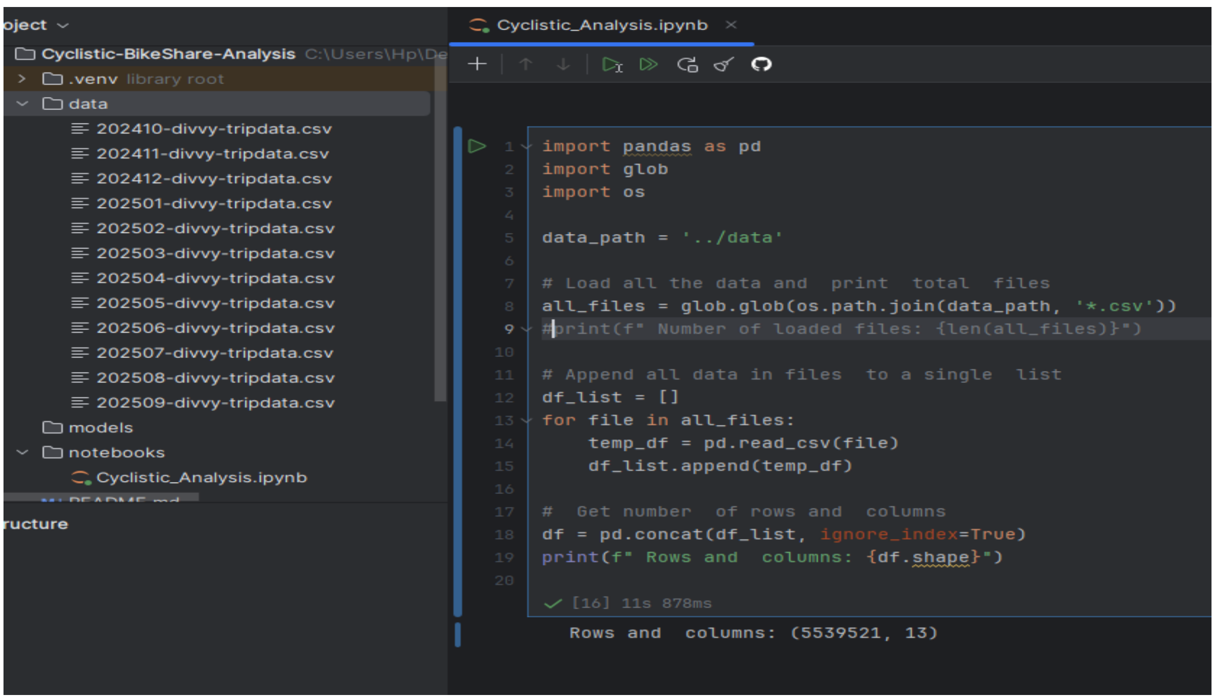


Figure 2: Successful merge all data (5539521 columns, 13 rows)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5539521 entries, 0 to 5539520
Data columns (total 13 columns):
 #   Column              Dtype
---  ------              -----
 0   ride_id             object
 1   rideable_type       object
 2   started_at          object
 3   ended_at            object
 4   start_station_name  object
 5   start_station_id    object
 6   end_station_name    object
 7   end_station_id      object
 8   start_lat           float64
 9   start_lng           float64
 10  end_lat             float64
 11  end_lng             float64
 12  member_casual       object
dtypes: float64(4), object(9)
memory usage: 549.4+ MB
None
```

Figure 3: Columns and their data types

Some rows have outliers as follow:

- started_at & ended_at are strings not datetime
- Many null values
- Some trips' duration is seconds or days

- **Cleaning Steps**

Cleaning Steps Performed:

1. Converted "started_at" and "ended_at" columns to a datetime format

2. Calculated ride duration in minutes in a new column: "ride_length_min"

3. Exclude rides shorter than 1 minute and longer than 24 hours

4. Added "day_of_week", "month", and "hour" columns to make analysis easier.

```python
df['started_at'] = pd.to_datetime(df['started_at'])
df['ended_at'] = pd.to_datetime(df['ended_at'])

print(df.info())
```

```
✓ [33] 30s 76ms
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5539521 entries, 0 to 5539520
Data columns (total 13 columns):
 #   Column              Dtype
---  ------              -----
 0   ride_id             object
 1   rideable_type       object
 2   started_at          datetime64[ns]
 3   ended_at            datetime64[ns]
 4   start_station_name  object
 5   start_station_id    object
 6   end_station_name    object
 7   end_station_id      object
 8   start_lat           float64
 9   start_lng           float64
 10  end_lat             float64
 11  end_lng             float64
 12  member_casual       object
dtypes: datetime64[ns](2), float64(4), object(7)
memory usage: 549.4+ MB
None
```

Figure 4: Converted 'started_at' & 'ended_at' columns to datetime

```python
# Exclude rides shorter than 1 minute and longer than 24 hours
validTrips = (df['duration_minutes'] > 1) & (df['duration_minutes'] < 1440)
df_filtered = df[validTrips]

print(f"Count original rows: {len(df)}")
print(f"Count Rows after applying the filter: {len(df_filtered)}")
print(f"Count deleted rows: {len(df) - len(df_filtered)}")
```

```
✓ [38] 33s 699ms

Count original rows: 5539521
Count Rows after applying the filter: 5397554
Count deleted rows: 141967
```

Figure 5: Results after applying the filter (Exclude rides shorter than 1 minute and longer than 24 hours)

```python
# Adding 3 new rows day_of_week, month, hour
df['day_of_week'] = df['started_at'].dt.day_name()
df['month']       = df['started_at'].dt.month_name()
df['hour']        = df['started_at'].dt.hour

# ordering the days
day_order = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday']
df['day_of_week'] = pd.Categorical(df['day_of_week'], categories=day_order, ordered=True)

# Print the new columns
print("The new columns: ")
print(df[['ride_duration_mins', 'day_of_week', 'month', 'hour']].head(10))
```

```
✓ [43] 20s 929ms

The new columns:
   ride_duration_mins day_of_week    month  hour
0            6.874200      Monday  October     3
1            5.426067      Sunday  October    19
2            7.230283      Sunday  October    23
3           11.974250      Monday  October     2
4            1.619617      Sunday  October    19
5            7.077433      Monday  October     6
6            4.770633      Monday  October     9
7            3.173750      Friday  October    21
8            1.950767    Saturday  October    11
9            0.349867    Saturday  October    12
```

Figure 6: Display the first 10 rows of the new columns.

- **Result**

  After applying the filter, 141,967 rides removed and the final dataset = 5,397,554 rides

- ## Handling Missing Values

  I found some missing fields in start/end station names, and I decided to keep the rows since they still contain valid duration and user type information, which are still useful in the analysis.

4. Analysis & Key Findings

– Total Rides by User Type

We have 5,397,554 total rides:

3,479,797 (64.5%) of them are Annual Members.

1,917,757 (35.5%) of them are Casual Riders.

Insight: Annual members dominate overall usage.

– Average Ride Duration

The overall average trip duration is: 14.9 minutes

for Casual Riders: 20.2 minutes

for Annual Members: 12.0 minutes

Insight: Casual riders take longer trips

– Ride Patterns by Day of Week

Annual Members ride Patterns by Day of Week:

On Tuesday, there are 568165 trips = 16.33% of the total Annual members' trips.

On Thursday, there are 548303 trips = 15.76% of the total Annual members' trips

On Wednesday, there are 532789 trips = 15.31% of the total Annual members' trips

Casual Members ride Patterns by Day of Week:

On Saturday, there are 391845 trips = 20.43% of the total casuals' trips.

On Sunday, there are 323850 trips = 16.89% of the total casuals' trips.

On Friday, there are 301154 trips = 15.70% of the total casuals' trips.

Insight**:** Annual Members ride more during weekdays, Casual Members ride more during the weekends

– Ride Patterns by Hour

Annual members' ride patterns by hours:

17 o'clock, there are 371627 trips = 10.68% of the total annual members' rides

16 o'clock, there are 327242 trips = 9.40% of the total annual members' rides

18 o'clock, there are 287662 trips = 8.27% of the total annual members' rides

Casual members' ride patterns by hours:

17 o'clock, there are 182728 trips = 9.53% of the total casual members' rides

16 o'clock, there are 169503 trips = 8.84% of the total casual members' rides

18 o'clock, there are 156409 trips = 8.16% of the total casual members' rides

Insight**:** Peaks at 7–9 AM & 4–6 PM for members, Casual: afternoon/evening usage

– Bike Type Preference

Annual members' preferred bike type:

electric_bike = 2125579 trips (61.08%)

classic_bike = 1354218 trips (38.92%)

Casual members' preferred bike type:

electric_bike = 1200469 trips (62.6%)

classic_bike = 717288 trips (37.4%)

***Insight*:** Electric bikes are more preferred

– Seasonal Trends

Seasonal Trends for Annual members:

In August, there are 443130 trips = 12.73%

In September, there are 440939 trips = 12.67%

In July, there are 430394 trips = 12.37%

Seasonal Trends for Casual Riders:

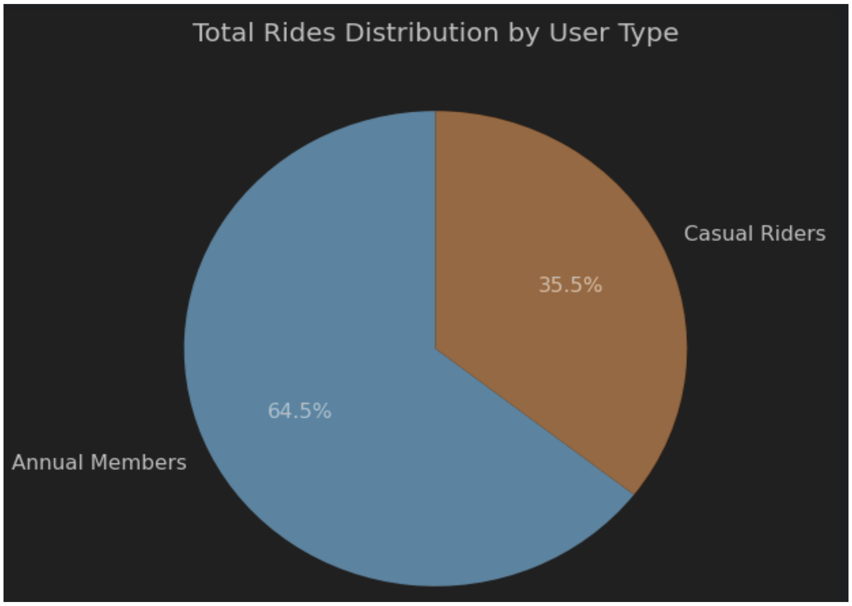
In August, there are 323533 trips = 16.87%

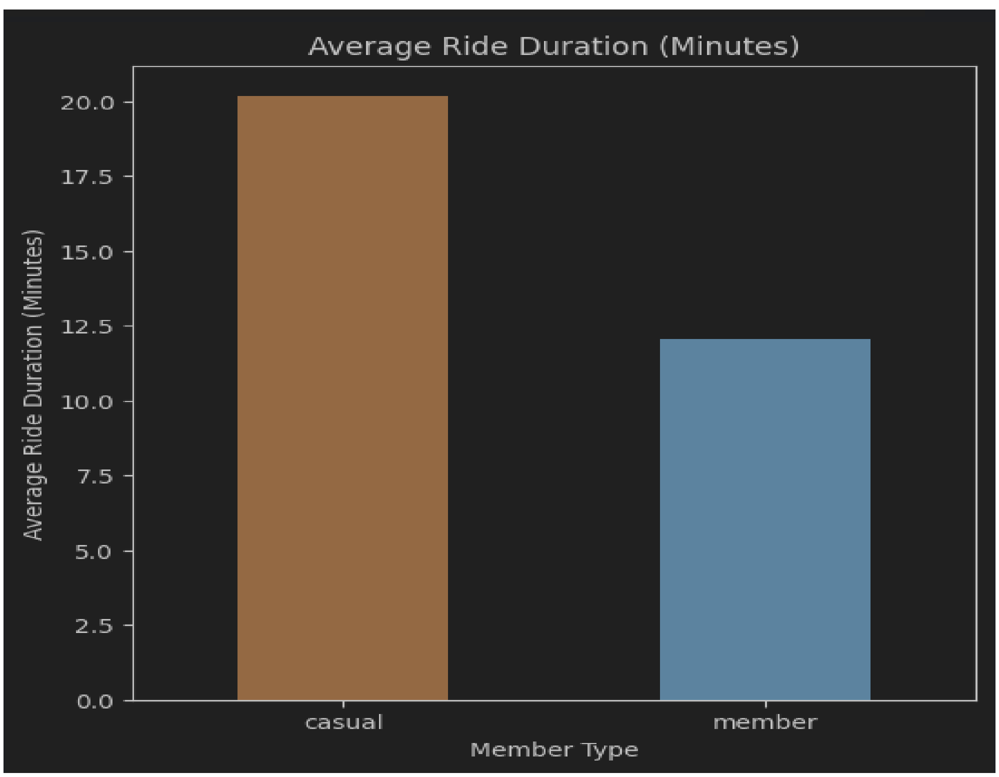
In July, there are 308446 trips = 16.08%

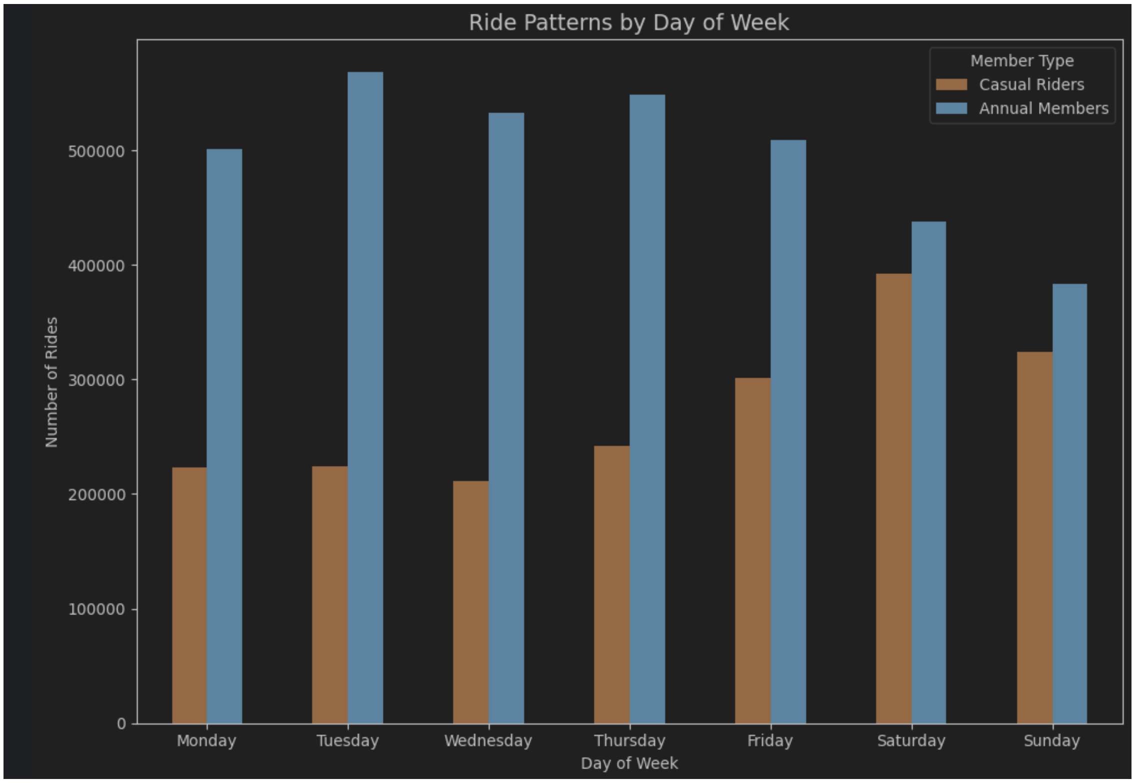
In June, there are 278702 trips = 14.53%

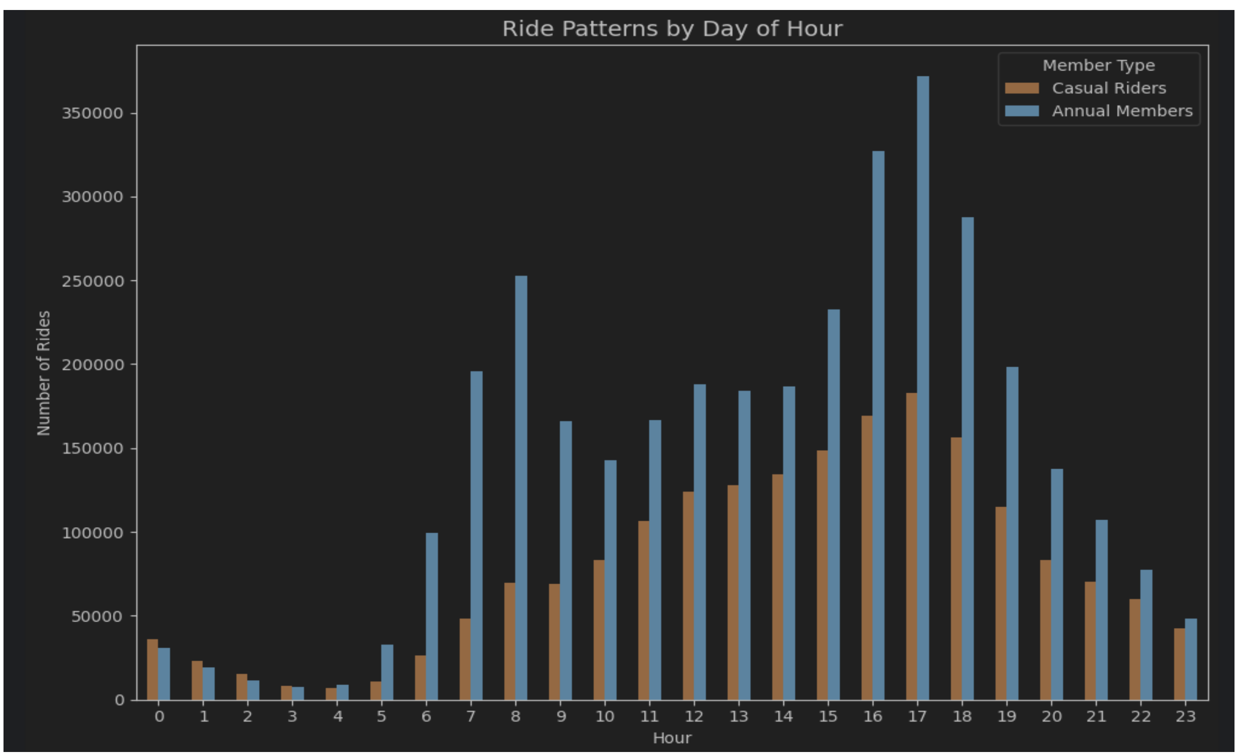
***Insight*:** Peak months: Jul–Sep (summer), Lowest: Dec–Feb (winter). Weather is the biggest driver!

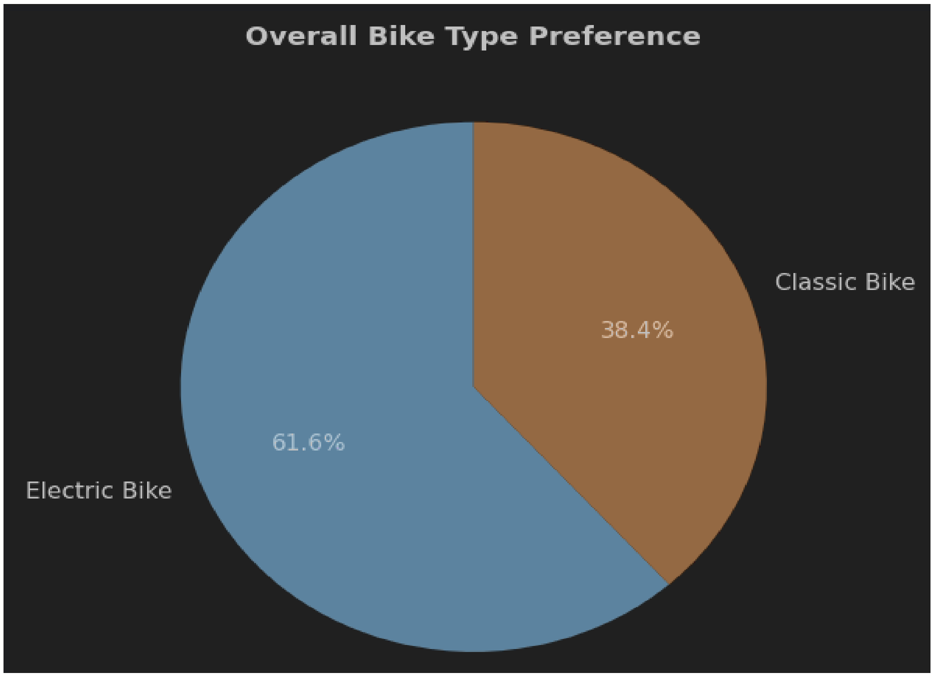
- Visualizations

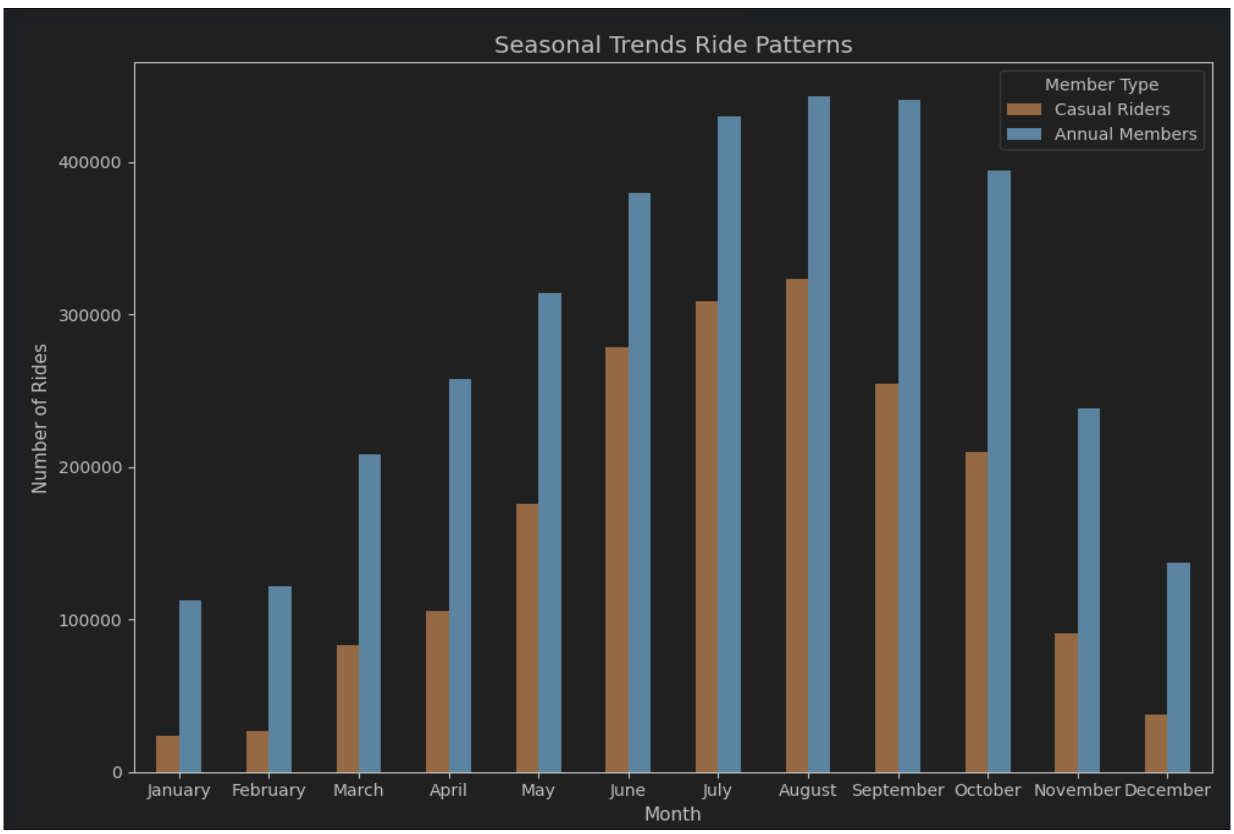





Ride Patterns by Day of Week
Member Type
Casual Riders
Annual Members
Number of Rides
0
100000
200000
300000
400000
500000
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week


Ride Patterns by Day of Hour
Member Type
Casual Riders
Annual Members
Number of Rides
0
50000
100000
150000
200000
250000
300000
350000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour


Overall Bike Type Preference
Classic Bike
38.4%
61.6%
Electric Bike


Seasonal Trends Ride Patterns
Member Type
Casual Riders
Annual Members
Number of Rides
400000
300000
200000
100000
0
January
February
March
April
May
June
July
August
September
October
November
December
Month

5. Recommendations

1- Weekends discount

Casual peak rides are during the weekends; this can be an opportunity convince them to upgrade to an annual membership by giving them X% discount for rides per weekend, with a maximum of X days per month.

2- Free rides in during summer.

As July to August (summer time) is the peak period for casual rides. Then, we can make a campaign to convince them to upgrade to an annual membership by giving them X free rides during the summer if they upgrade to an annual membership.

3- Electric bikes + Peak hours discount for the first month

As the preferred bike type is electric bikes, and the peak hours are between 12 and 18. Then, we can offer them a discount of X% on electric bike rides during the peak hours for the first month if did buy a membership.

6. Conclusion

I found that members use Cyclistic bikes differently: members for daily commuting, casuals for leisure on weekends and in summer.

In order to target casual riders to buy a membership. Then, we need to offer them discounts/free rides during weekends and summer