GOOGLE DATA ANALYTICS CAPSTONE PROJECT

Author: Ezennia Divine Onyedikachi

Date: 30th October, 2025

In [115…

```
# Authenticate google account
from google.colab import auth
auth.authenticate_user()
print('Authenticated successfully!')
```

```
Authenticated successfully!
```

In [116…

```
# Import BigQuery client library
from google.cloud import bigquery
project_id = 'axiomatic-set-467921-a9'
Client = bigquery.Client(project='axiomatic-set-467921-a9')
print('Connected to BigQuery')
```

```
Connected to BigQuery
```

Google Data Analytics Capstone Project : Cyclistic Bike-Share Case Study

1. Introduction

Business Scenerio:
The Cyclistic bike-share Company aims to maximize annual membership by understanding how casual riders and annual members use Cyclistic bikes differently.

The analysis explores user behavior using trip data from **September 2024** to **August 2025**. It focuses on identifying patterns in ride duration, timing and usage frequency that can help develop strategies to convert casual riders into annual members.

Data Source:
Publicly available Cyclistic trip data (12-month dataset, CSV format) was obtained from the official **Motivate International Inc** datasets.
The data contains ride details such as start time, end time, bike type, start station, end station and user type.

Tools Used:

- BigQuery (data cleaning and analysis)
- Tableau (data visualization)
- Google Colab (documentation)

Business Task:

- Identify key differences in user behavior, ride patterns and usage frequency between casual riders and annual members.
- Suggest actionable insights to improve membership conversion and overall ridership.

2. Data Preparation

Steps Taken:

- Importation and Merging:
  Imported and merged monthly datasets from the Cyclistic public data source using 'UNION DISTINCT' to form a complete view of unique 1,975,968 ride records.

In [117…

```
# Combine monthly ride data into one dataframe
query = """
--Dataset_1
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2024_11`
UNION DISTINCT --Remove duplicate records

--Dataset_2
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2024_12`
UNION DISTINCT
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2025_01`
UNION DISTINCT
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2025_02`
UNION DISTINCT
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2025_03`
UNION DISTINCT
SELECT
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2025_04`
UNION DISTINCT
SELECT
```

```
    *
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2025_05`
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[117…

	ride_id	rideable_type	started_at	ended_at	start
0	6EF370384EC1315C	classic_bike	2025-01-20 00:46:32.227000+00:00	2025-01-21 01:46:24.693000+00:00	Wa
1	F1929422FF660FA3	classic_bike	2025-01-02 03:37:24.125000+00:00	2025-01-03 04:37:11.571000+00:00	Dea
2	3FD5ECA0AD06FD4A	classic_bike	2025-01-02 03:36:26.031000+00:00	2025-01-03 04:36:20.799000+00:00	Dea
3	93D1F736499C97BB	classic_bike	2025-01-21 14:59:33.471000+00:00	2025-01-22 15:59:25.370000+00:00	Day
4	4037BBE97936009A	classic_bike	2025-01-04 18:19:09.983000+00:00	2025-01-04 19:02:45.623000+00:00	Cun

- Cleaning and Transformation:
  Cleaned data by removing leading and trailing spaces in station name columns using `TRIM()`, replacing 18.4% of the records of start id and it's station name column and 19.2% of the records of end id and it's station name column, all missing with 'unknown' using `IFNULL()` so as to keep all ride records for analysis and correcting started at and ended at columns using `LEAST()` and `GREATEST()` to ensure durations are non-negative.

In [118…

```
# prepare data into a cleaned ride dataset
query = """
SELECT
    ride_id,
    rideable_type,
--Ensure that the start time and end time values are earlier and later respectiv
    LEAST(started_at, ended_at) AS started_at,
    GREATEST(started_at, ended_at) AS ended_at,
--Remove whitespaces and replace null station names and ids
     IFNULL(TRIM(start_station_name), 'unknown station') AS start_station_name,
     IFNULL(start_station_id, 'unknown id') AS start_station_id,
     IFNULL(TRIM(end_station_name), 'unknown station') AS end_station_name,
     IFNULL(end_station_id, 'unknown id') AS end_station_id,
--keep all other columns previously had
     start_lat,
     start_lng,
     end_lat,
     end_lng,
     member_casual

FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_2024_2025`
"""
```

```
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[118…

	ride_id	rideable_type	started_at	ended_at	start_
0	578AB72DAA395AEC	electric_bike	2024-11-04 15:19:33.048000+00:00	2024-11-04 15:19:46.887000+00:00	u
1	066AC387CB3A8BB1	electric_bike	2025-04-26 07:51:11.495000+00:00	2025-04-26 08:04:37.559000+00:00	u
2	169980F1A6DF7D66	electric_bike	2024-11-26 19:16:46.490000+00:00	2024-11-26 19:20:11.820000+00:00	
3	1450891B85F9427A	electric_bike	2025-05-07 14:18:18.135000+00:00	2025-05-07 14:27:30.011000+00:00	Cum
4	E8385DCCE2CA318B	classic_bike	2025-05-06 21:40:34.339000+00:00	2025-05-06 22:17:25.297000+00:00	Cum

- Table Enrichment:
  Enriched dataset by filtering out incomplete record of **October [2024]** to ensure fair monthly comparison and creating derived fields;
  - Trip Duration(sec)
  - hour of day(0-23) started at & ended at
  - time of day started at & ended at
  - day of week started at & ended at
  - week part (weekend/weekday) started at & ended at
  - month started at & ended at

In [119…

```
# Create an enriched summary dataset
query = """
 SELECT
    ride_id,
    member_casual,
    rideable_type,
    start_station_name,
    start_station_id,
    end_station_name,
    end_station_id,
    started_at,
    ended_at,
    trip_duration_sec,
    hour_of_day_started,
    hour_of_day_ended,
    time_of_day_started,
    time_of_day_ended,
    day_of_week_started,
    day_of_week_ended,
    CASE
      WHEN day_of_week_started IN ('Saturday', 'Sunday') THEN 'Weekend'
      ELSE 'Weekday'
```

```
      END AS week_part_started,
    CASE
      WHEN day_of_week_ended IN ('Saturday', 'Sunday') THEN 'Weekend'
      ELSE 'Weekday'
      END AS week_part_ended,
    month_started,
    month_ended,
    start_lat,
    start_lng,
    end_lat,
    end_lng
  FROM
    (
      SELECT
        TIMESTAMP_DIFF(ended_at, started_at, SECOND) AS trip_duration_sec,
      --Pull the hour of day here
        EXTRACT(HOUR FROM started_at) AS hour_of_day_started,
        EXTRACT(HOUR FROM ended_at) AS hour_of_day_ended,
      --Categorize it into time-of-day segments
        CASE
        WHEN EXTRACT(HOUR FROM started_at) BETWEEN 5 AND 11 THEN 'Morning'
        WHEN EXTRACT(HOUR FROM started_at) BETWEEN 12 AND 16 THEN 'Afternoon'
        WHEN EXTRACT(HOUR FROM started_at) BETWEEN 17 AND 20 THEN 'Evening'
        ELSE 'Night'
        END AS time_of_day_started,
        CASE
        WHEN EXTRACT(HOUR FROM ended_at) BETWEEN 5 AND 11 THEN 'Morning'
        WHEN EXTRACT(HOUR FROM ended_at) BETWEEN 12 AND 16 THEN 'Afternoon'
        WHEN EXTRACT(HOUR FROM ended_at) BETWEEN 17 AND 20 THEN 'Evening'
        ELSE 'Night'
        END AS time_of_day_ended,
      --Pull the day of week here
        FORMAT_DATE('%A', DATE(started_at)) AS day_of_week_started,
        FORMAT_DATE('%A', DATE(ended_at)) AS day_of_week_ended,
      --Pull month here
        FORMAT_DATE('%B', DATE(started_at)) AS month_started,
        FORMAT_DATE('%B', DATE(ended_at)) AS month_ended,
        ride_id,
        member_casual,
        rideable_type,
        start_station_name,
        start_station_id,
        end_station_name,
        end_station_id,
        started_at,
        ended_at,
        start_lat,
        start_lng,
        end_lat,
        end_lng
      FROM
        `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_cleaned`
    )
  WHERE
    month_started <> 'October'
    AND month_ended <> 'October'
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[119…

	ride_id	member_casual	rideable_type	start_station_name	start_station_i
0	8B6F30800D5FB70B	casual	electric_bike	Public Rack - Cumberland & Catherine	1182.(
1	FEE6B0792C653AD0	casual	classic_bike	Cumberland Ave & Catherine Ave	2446
2	B4622270AD45E617	member	electric_bike	Orange Ave & Addison St	2132(
3	46F0186D29981A21	member	electric_bike	unknown station	unknown ic
4	97A1F45ECDA14426	member	electric_bike	unknown station	unknown ic

5 rows × 24 columns

Data Limitations:

- TOOL NOTE:
  Due to BigQuery sandbox's 100MB file upload limit, trip data for **September-October [2024]** and **June-September [2025]** could not be uploaded. As a result, the monthlly trend analysis covers only the available months between **November[2024] and May[2025]**. Insights should therefore be interpreted with awareness of these missing months.

- DATA NOTE:
  The dataframe contained partial data for **October [2024]**. To ensure fair monthly comparisons, **October** was excluded from final analysis and visualization.This decision preserves the integrity of trend-based insights.

- DATA VALUE - "*UNKNOWN*" NOTE:
  During data cleaning, approximately 18.4% and 19.2% of the fields - start and end stations were missing and replaced with "*unknown*". Interestingly, this category accounted for the highest ride volume, highlighting significant data quality issues in ride distribution. Subsequent analyses exclude "*unknown*" values to ensure that insights about ride-based demand reflect actual ride patterns.

If granted access to a full BigQuery Cloud Storage Integration, I would re-run the analysis including the missing months to validate seasonal trends.

Verification:
Confirmed data cleaning efforts was well executed and validated data integrity by checking for nulls, outliers and logical errors.

3. Data Exploration

Objective:
To perform light data exploration using visuals and basic descriptive analysis to understand general usage patterns.

During exploration, several preliminary analyses were conducted to identify trends in ride duration and frequency. While not all led to final insights, they informed findings and have been summarized below for completeness.

3.1 Total cycling record

To understand the dataset as a whole, I examined the total ride records and distinct bike types to help find patterns before comparing.

In [120…

```
# Count all ride observations
query = """
SELECT
  COUNT(ride_id) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[120…

	total_ride_count
0	1975922

In [121…

```
# Return all unique ride types
query = """
SELECT
  DISTINCT rideable_type
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[121…

	rideable_type
0	electric_bike
1	classic_bike

The analyses result shows that the dataset contains **1,975,922 ride records** with two types of bikes - **electric** and **classic**. This helps establish the scale before comparing usage patterns between member types.

3.2 Total ride count by user type

To identify which user most frequently take ride, I analyzed total ride count taken by each user.

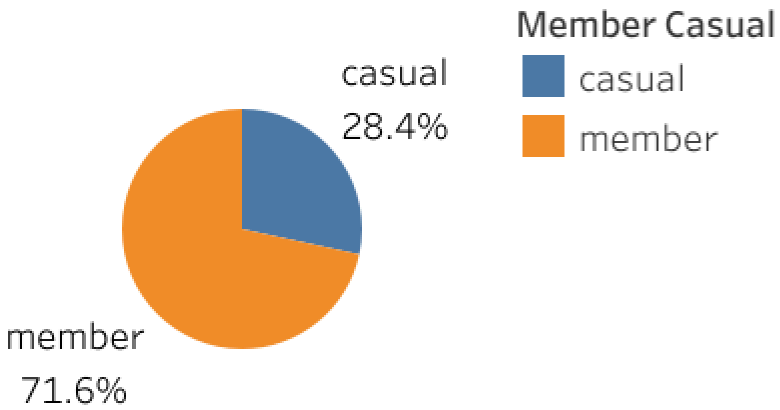
In [122…

```
# Count total ride per user type
query = """
SELECT
  member_casual,
  COUNT(ride_id) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual
ORDER BY
  COUNT(ride_id) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[122…

	member_casual	total_ride_count
0	member	1414862
1	casual	561060

Ride Count Distribution between Members and Casual Riders



The pie chart above illustrates the proportion of ride count contributed by each user type represented by distinct colors. It shows that member accounts for the largest share representing 71.6% while casual contributes 28.4%, suggesting members to be the major dominants of Cylistic bike-share ride volume.

3.3 Total and average trip duration by user type

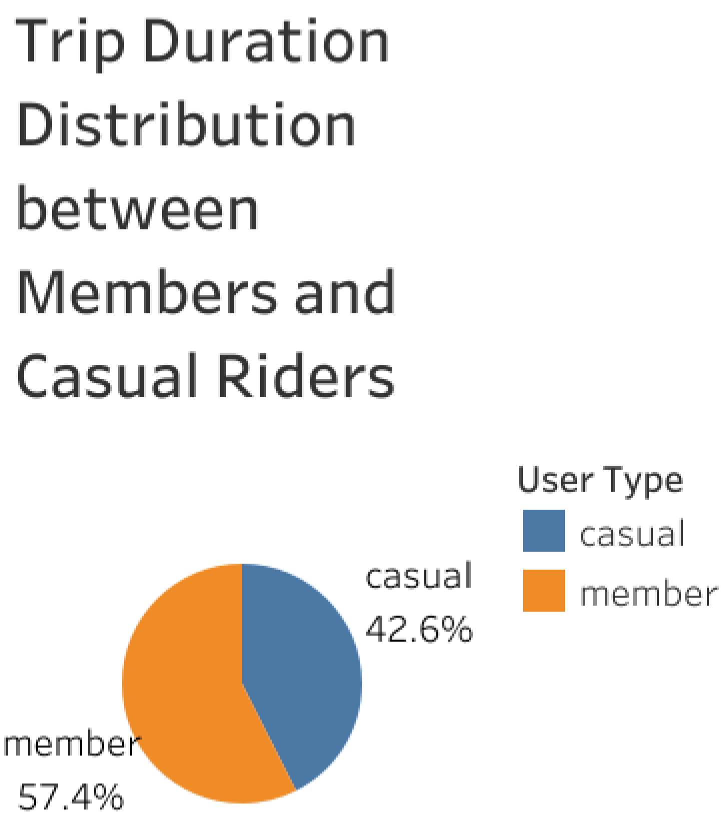
To determine which user engage in longer trip duration, I examined each user average trip duration and volume.

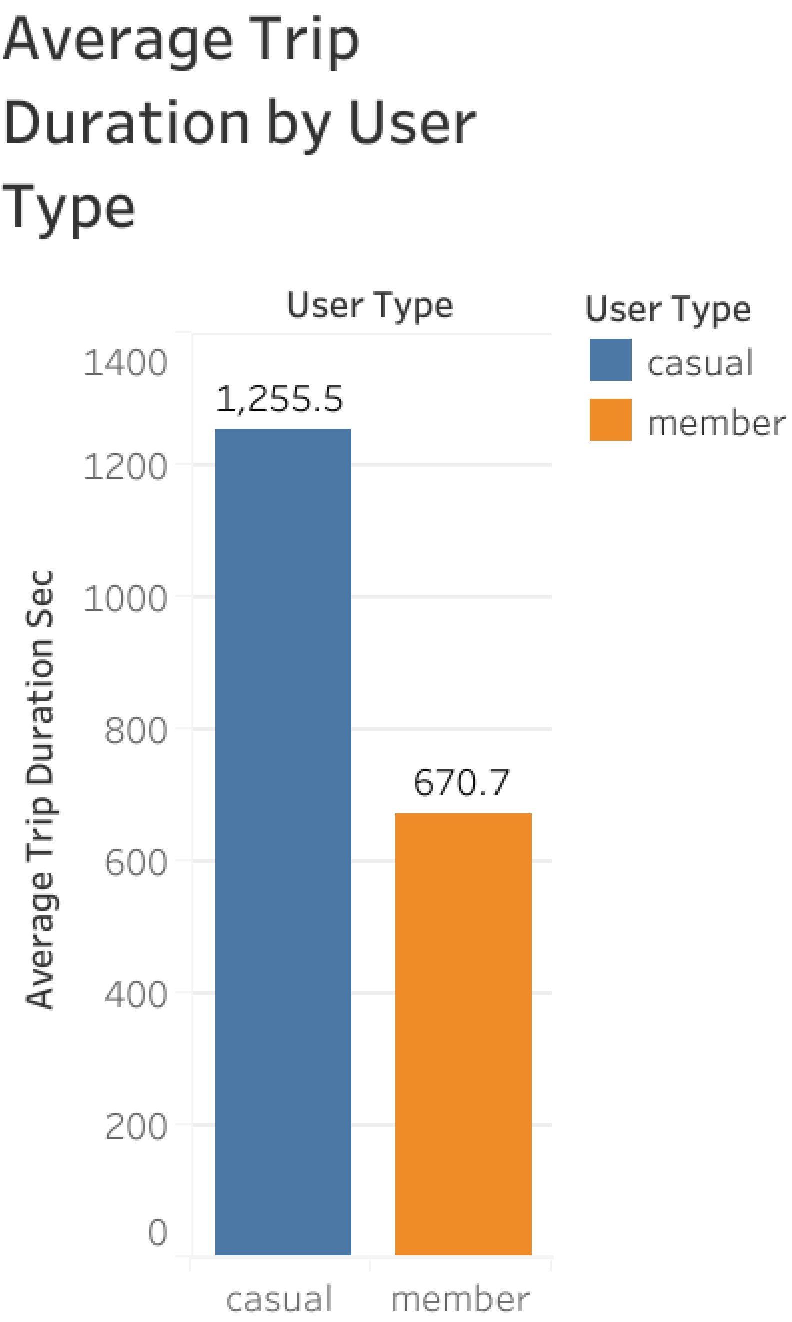
In [123…

```
# Calculate the min, max, average and total trip duration per user type
query = """
SELECT
    member_casual,
    SUM(trip_duration_sec) AS total_trip_duration_sec,
    ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec,
    MAX(trip_duration_sec) AS max_trip_duration_sec
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual
ORDER BY
    SUM(trip_duration_sec) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[123…

	member_casual	total_trip_duration_sec	avg_trip_duration_sec	max_trip_duration_sec
0	member	948592002	670.45	89997
1	casual	703425749	1253.74	93595





The pie chart and bar graph above represents the total and average trip duration across user type respectively. It highlights that member accounts for the largest proportion of trip duration, overall due to higher ride demand constituting 57.4% while casual 42.6%, however, casual riders have longer average trip duration, with most rides lasting about **2x** longer than that of members. This implies that members utilize the service for short, frequent commutes, while casual riders engage in longer, occasional leisure-oriented trips.

3.4 Total ride count by bike type

To discover the bike used most frequently, I explored total ride count for each bike type.

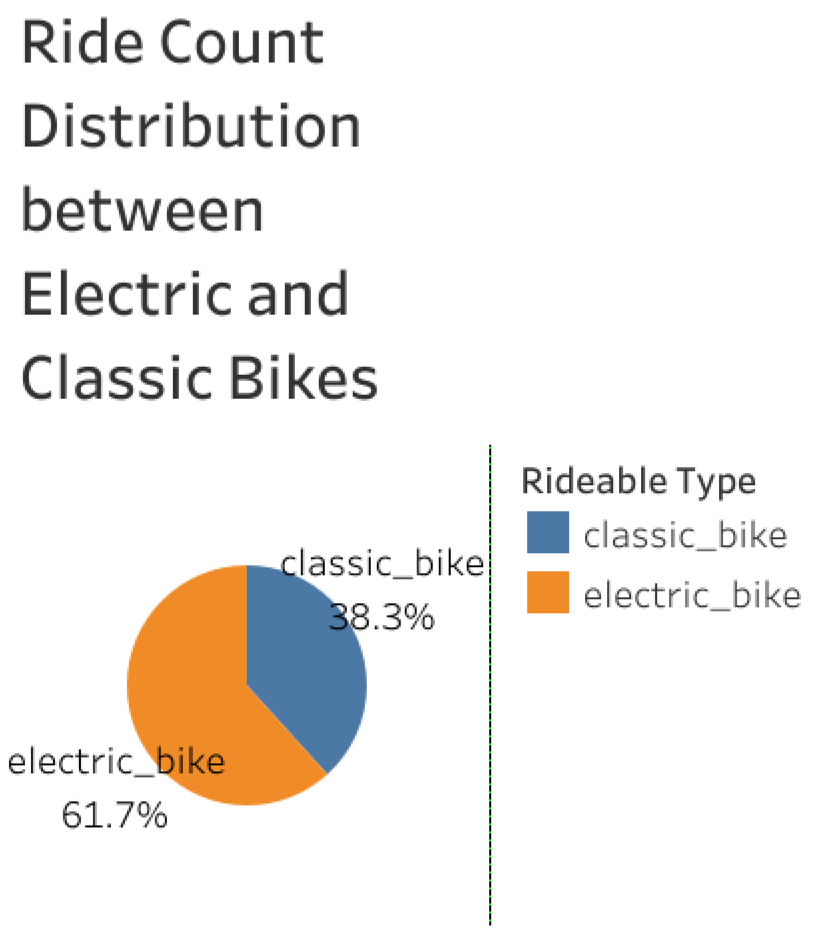
In [124…

```
# Count total ride across each bike type
query = """
SELECT
```

```
  rideable_type,
  COUNT(*) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  rideable_type
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[124...

	rideable_type	total_ride_count
0	electric_bike	1219180
1	classic_bike	756742



The pie chart illustrates the distribution of ride count among bike types, represented with distinct colors. Electric bike constitutes the largest proportion amounting to 61.7% of total ride count while classic bike accumulates the remaining 38.3%, affirming the major demand for electric bike among Cyclistic users.

3.5 Total trip duration by bike type

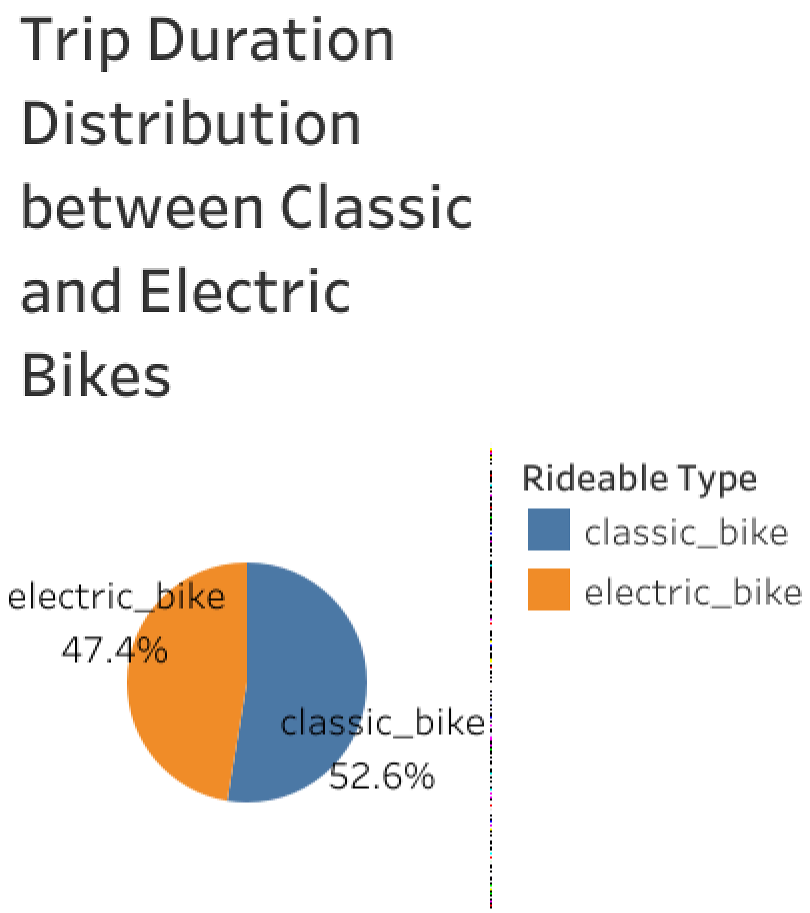
To identify the bike used to take longer trip duration, I compared total trip duration across bike types.

In [125…

```
# Calculate total trip duration across each bike type
query = """
SELECT
  rideable_type,
  SUM(trip_duration_sec) AS total_trip_duration_sec
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  rideable_type
ORDER BY
  SUM(trip_duration_sec) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[125…

	rideable_type	total_trip_duration_sec
0	classic_bike	869702867
1	electric_bike	782314884



The trip duration distribution pie chart shows that despite classic bike records lower ride demand, yet it represents the largest share of total trip duration constituting 52.6% while electric bike contributing 47.4%. Among users, this suggests the occasional use of classic bike for longer distance travels whereas shorter frequent commutes using electric bike.

4. Key Analytical Findings

Objective:
To highlight the most significant findings derived from the Cyclistic trip data analysis.

After exploring several temporal patterns, I narrowed down to key insights.

Insight 1 - Temporal Station Trend

To identify the most common start station and end station where rides begin and end most frequently by users, I analyzed ride count by start and end station by user type.

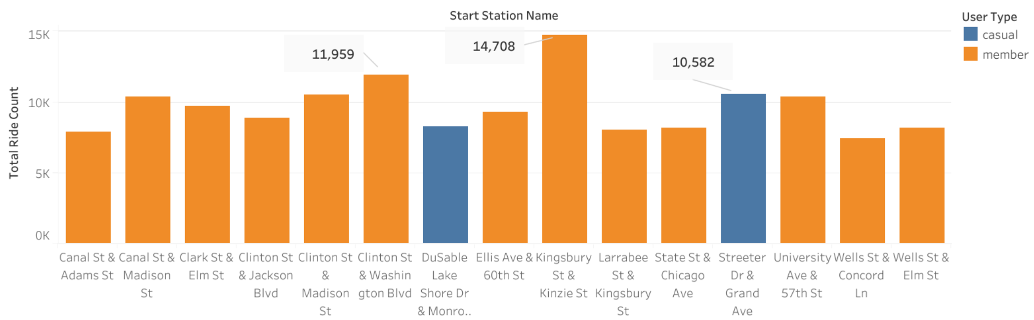
In [126…

```
# Count total ride by start station by user type
query = """
SELECT
  member_casual,
  start_station_name,
  COUNT(*) AS total_ride_count
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
WHERE start_station_name <> 'unknown station'
GROUP BY
  member_casual,
  start_station_name
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[126…

	member_casual	start_station_name	total_ride_count
0	member	Kingsbury St & Kinzie St	14708
1	member	Clinton St & Washington Blvd	11959
2	casual	Streeter Dr & Grand Ave	10582
3	member	Clinton St & Madison St	10533
4	member	University Ave & 57th St	10419

Total Ride Count by Start Station by User Type



In [127…

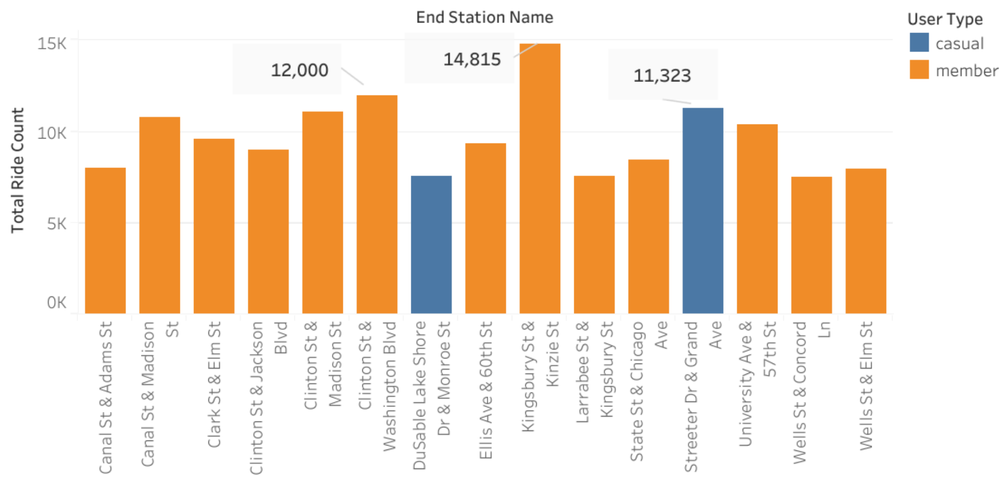
```
# Count total ride by end station by user type
query = """
SELECT
  member_casual,
  end_station_name,
```

```
  COUNT(*) AS total_ride_count
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
WHERE end_station_name <> 'unknown station'
GROUP BY
  member_casual,
  end_station_name
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

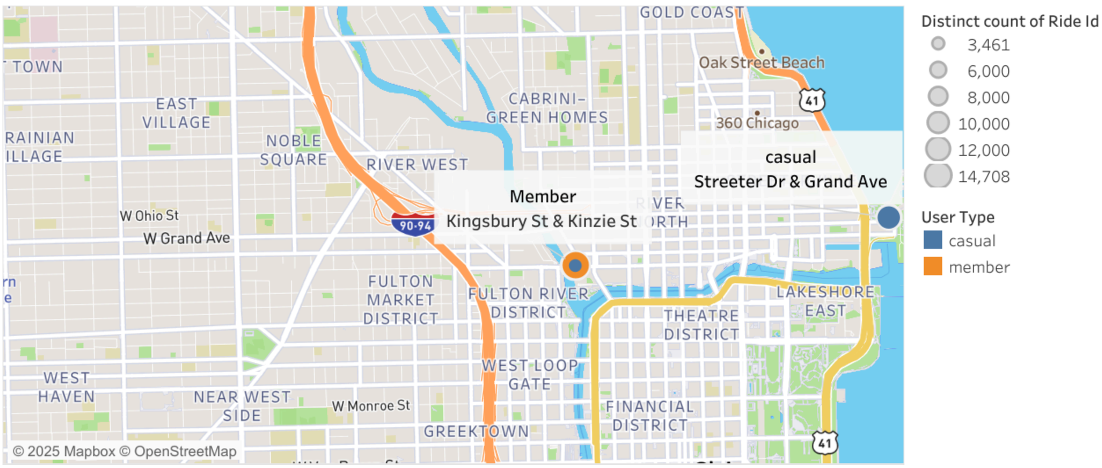
Out[127…

	member_casual	end_station_name	total_ride_count
0	member	Kingsbury St & Kinzie St	14815
1	member	Clinton St & Washington Blvd	12000
2	casual	Streeter Dr & Grand Ave	11323
3	member	Clinton St & Madison St	11125
4	member	Canal St & Madison St	10814

Total Ride Count by End Station by User Type



Geographic Representation of Stations



The bar charts display the total ride count across top 15 start and end stations per user type, with each bar representing start and end stations, distinguished by users using color contrast. Overall, Kingsbury St & Kinzie St, Clinton St & Washington Blvd and Streeter Dr & Grand Ave accounts for the top 3 stations with highest ride demand, highlighting Kingsbury St & Kinzie St as the most popular station among members and Streeter Dr & Grand Ave among casual riders. The accompanying map visualizes their geographical distribution with each station represented by unique colored points, sized proportionally to it's ride volume. Revealing that the start and end point high ride concentration in Kingsbury St & Kinzie St being an urban corridor, reflects most members use around business, academic and commercial institutes while Streeter Dr & Grand Ave being a recreational outdoor, indicates most casual riders use around social, tourist and recreational activity zones.

Insight 2 - Principal Station-to-Station Pair Structure

To determine the major route station, through which most rides frequently begin and end by users, I explored ride count across start and end station per user type.

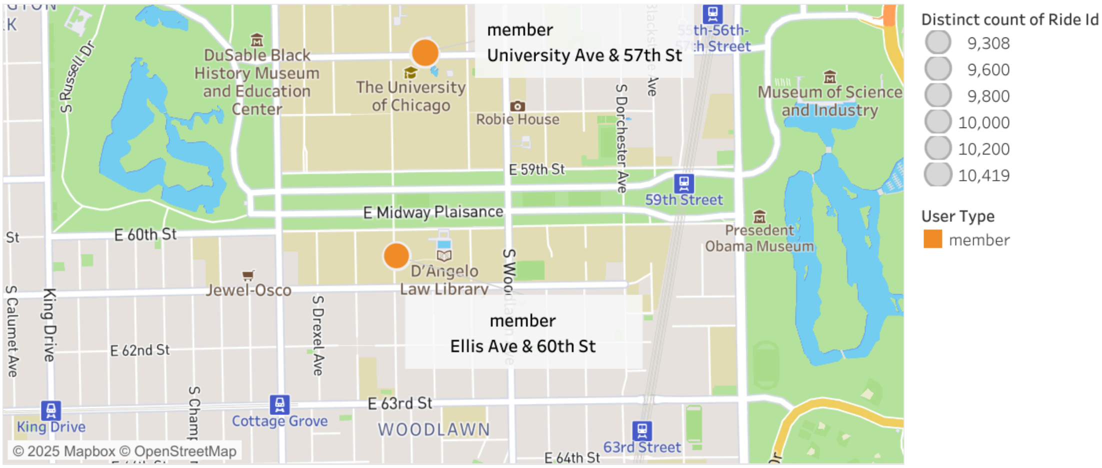
In [128…

```
# Count total ride taken across station-to-station pairs by distinct users
query = """
SELECT
  member_casual,
  start_station_name,
  end_station_name,
  COUNT(*) AS total_ride_count
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
WHERE start_station_name <> 'unknown station'
AND end_station_name <> 'unknown station'
GROUP BY
  member_casual,
  start_station_name,
  end_station_name
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[128…

	member_casual	start_station_name	end_station_name	total_ride_count
0	member	University Ave & 57th St	Ellis Ave & 60th St	2545
1	member	Ellis Ave & 60th St	University Ave & 57th St	2519
2	member	Ellis Ave & 60th St	Ellis Ave & 55th St	1996
3	casual	DuSable Lake Shore Dr & Monroe St	DuSable Lake Shore Dr & Monroe St	1882
4	member	Ellis Ave & 55th St	Ellis Ave & 60th St	1845

Geographic Examination of Stations



The analysis compares total ride count across specific start - end station per user type. Overall, members engage in higher ride frequency, University Ave & 57th St - Ellis Ave & 60th St records the highest ride route taken by member and DuSable Lake Shore Dr & Monroe St - DuSable Lake Shore Dr & Monroe St amounts to the highest among casual riders contributing 0.1% of total ride count respectively. Further spatial examination revealed that the positions of the major route among members are of close proximity and concentrated in business and educational districts reflecting short commuting behavior, in contrast casual riders are heavily centred in recreational hub, reaffirming their long leisure-oriented pattern usage.

Insight 3 - Station Trip Duration Variations

To uncover the most popular route station, where most users who take longer trip duration tend to begin and culminate, I examined average trip duration by start and end station among users.

In [129…

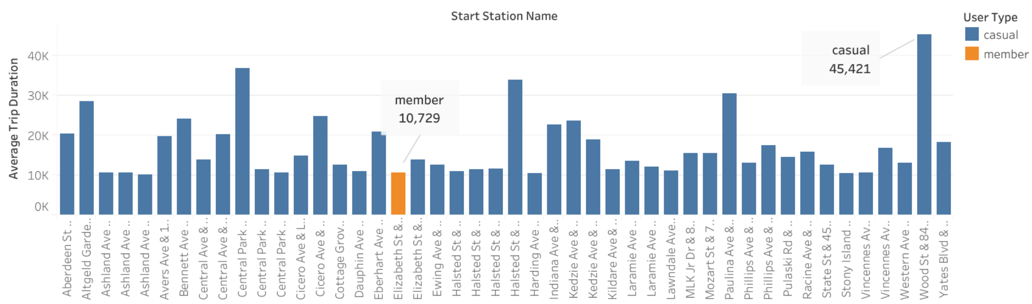
```
# Calculate average trip duration across station-to-station pairs taken by disti
query = """
SELECT
  member_casual,
  start_station_name,
  end_station_name,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
WHERE
  start_station_name <> 'unknown station'
  AND end_station_name <> 'unknown station'
GROUP BY
  member_casual,
  start_station_name,
  end_station_name
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[129…

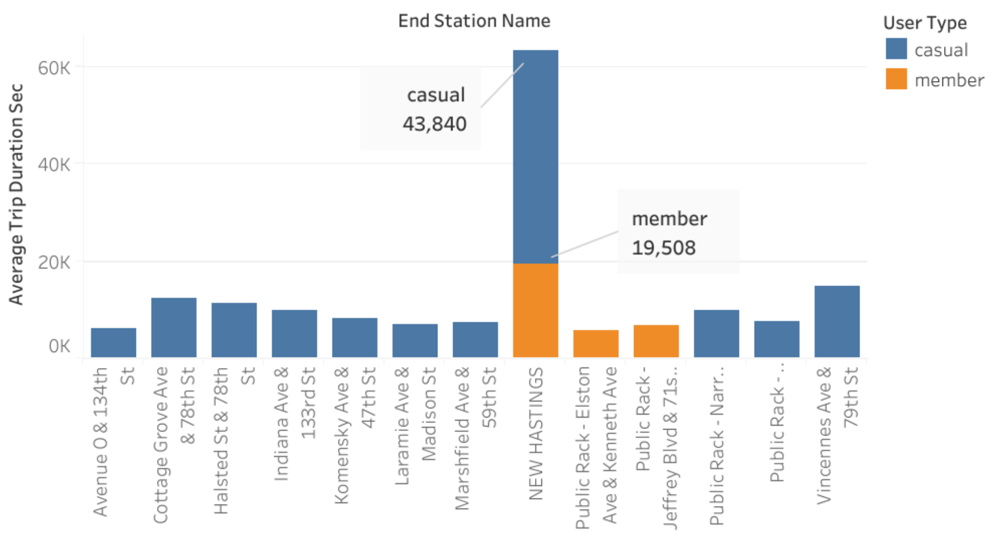
	member_casual	start_station_name	end_station_name	avg_trip_duration_sec
0	casual	Shedd Aquarium	Vincennes Ave & 79th St	89693.0
1	casual	Karlov Ave & Armitage Ave	Humboldt Blvd & Armitage Ave	86024.0
2	casual	Malcolm X College	Michigan Ave & 14th St	86017.0
3	member	Ogden Ave & Roosevelt Rd	Western Ave & Division St	85114.0
4	member	Christiana Ave & Lawrence Ave	Leavitt St & North Ave	83922.0

The analysis depicts the comparison of average trip duration across specific route station taken by users. In general, casual riders account for longer trip duration, highlighting unusually longer trip duration across certain route stations for both users. Deeper analysis revealed that this top station-to-station pairs contributed only 1 recorded ride each and thus were not considered statistically significant for insight generation, suggesting these were outliers rather than representative ride pattern. Consequently, futher interpretation focused on station-level pattern.

Comparison of Average Trip Duration by Start Station between Members and Casual Riders



Comparison of Average Trip Duration by End Station between Members and Casual Riders



The bar charts above illustrate the top average trip duration by start and end station per user type. In general, particularly among casual riders, Wood St & 84th St records the top start station with trip duration above **45,400sec** whereas among members, Elizabeth St & 47th St represents the most popular start station with less than **10,800sec**. Conversely, in general, NEW HASTING accounts for the top end station with the highest average trip duration across both users, with most rides of it's engaging casual riders exceeding **2x** that of members. The implies that the longer trip duration among casual riders are centered around recreational hubs, while members shorter trip length reflects commutes around business, educational and residential areas.

Insight 4 - Regular Bike Type Trend

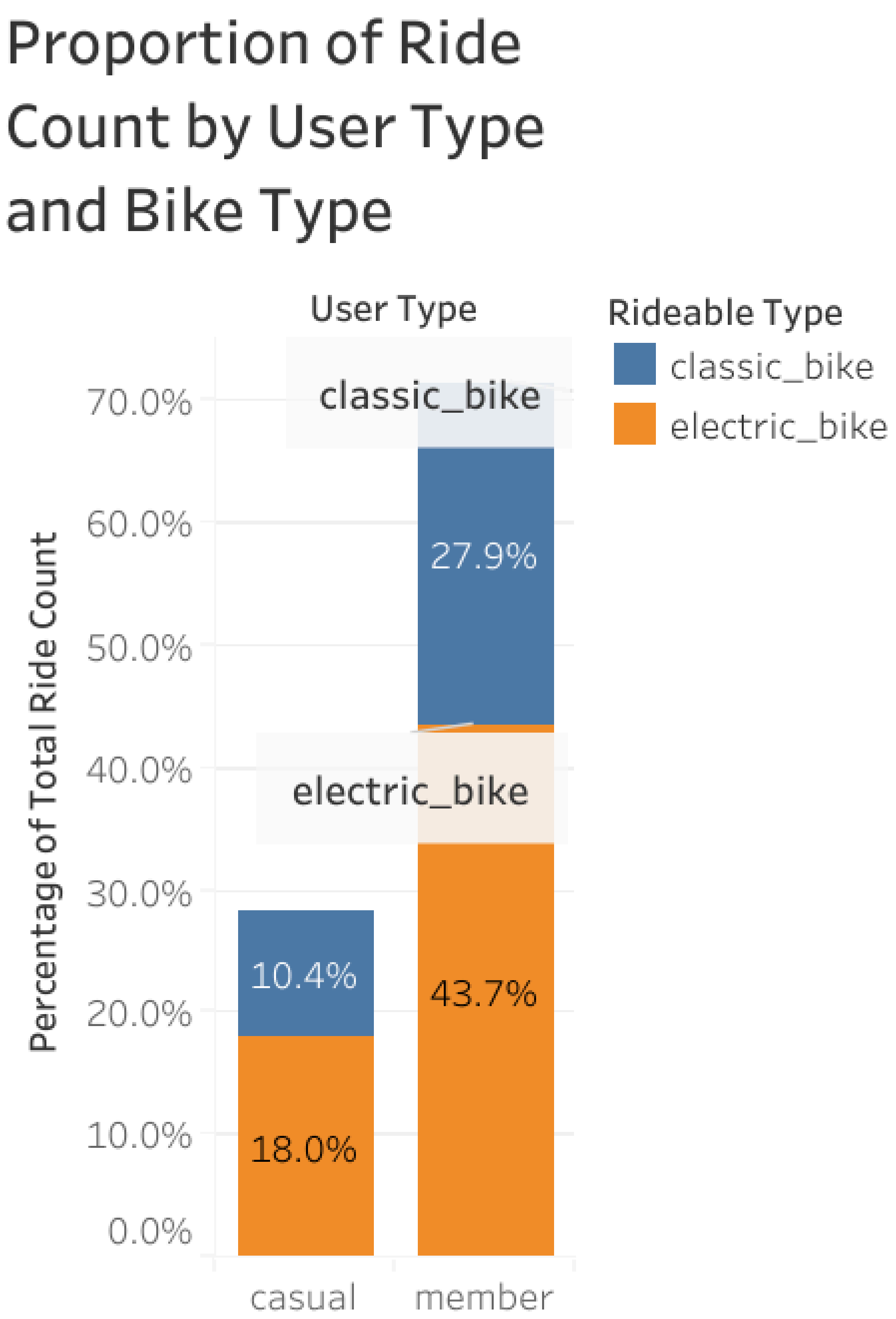
To understand which bike type distinct users utilize most frequently, I analyzed total ride frequency per bike type across users.

In [130…

```
# Count total ride per bike type across users
query = """
SELECT
  member_casual,
  rideable_type,
  COUNT(*) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  rideable_type
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[130…

	member_casual	rideable_type	total_ride_count
0	member	electric_bike	862793
1	member	classic_bike	552069
2	casual	electric_bike	356387
3	casual	classic_bike	204673



The above stacked column chart displays the percentage proportion of ride volume per user type and bike type. Across both users, electric bike accumulates the largest proportion recording 43.7% and 18.0% of the total ride volume, among member and casual riders respectively whereas classic bike accounts for 27.9% among member and 10.4% among casual riders. This reaffirms electric bike to be the utmost propellant of Cyclistic bike-share ride volume.

Insight 5 - Bike Type Trip Duration Contrast

To determine the bike used distinctly to engage in longer trip duration, I examined each user total and average trip duration per bike type.

In [131…

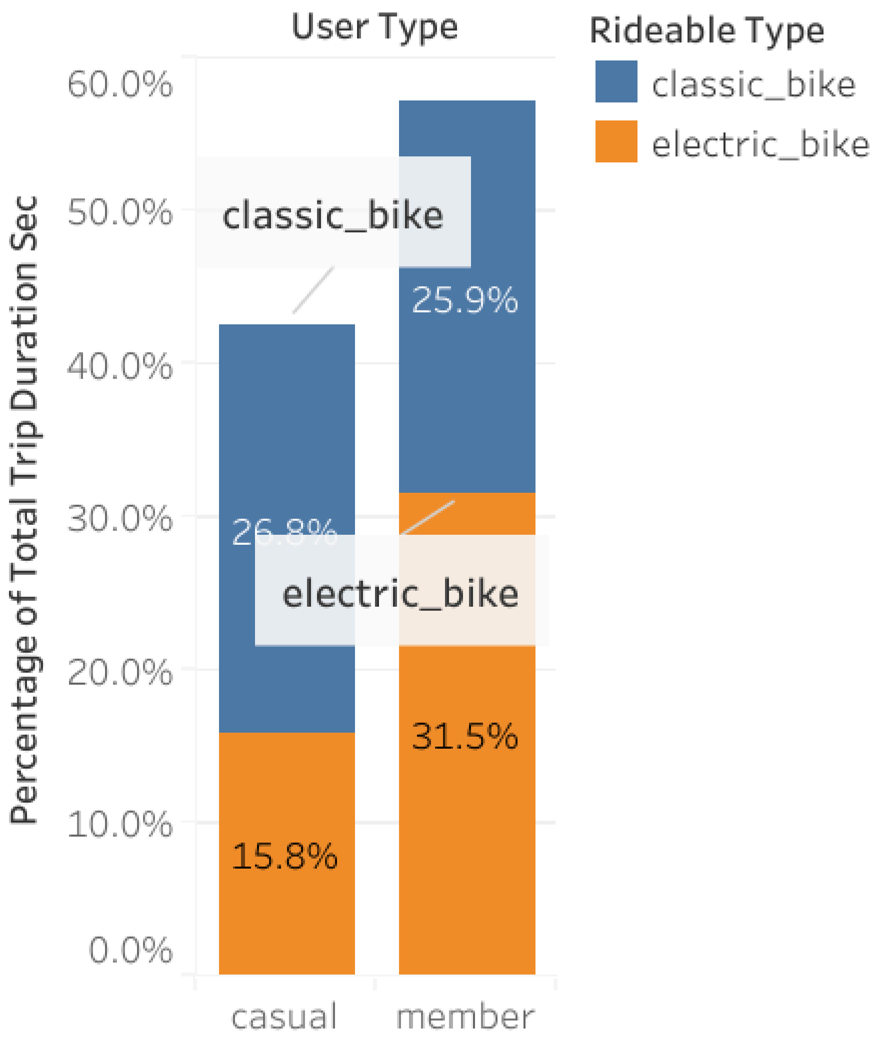
```
# Calculate the total and average trip duration per bike type across users
query = """
SELECT
  member_casual,
```

```
  rideable_type,
  SUM(trip_duration_sec) AS total_trip_duration_sec,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  rideable_type
ORDER BY
  SUM(trip_duration_sec) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

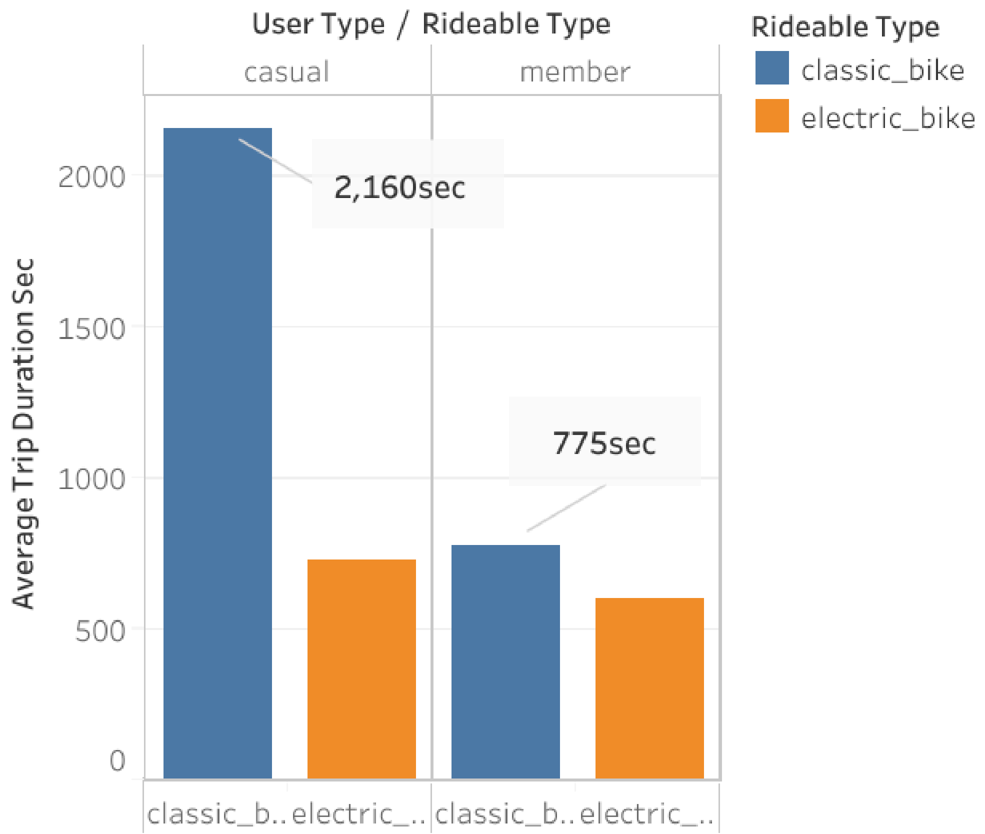
Out[131…

	member_casual	rideable_type	total_trip_duration_sec	avg_trip_duration_sec
0	member	electric_bike	520931891	603.77
1	casual	classic_bike	442042756	2159.75
2	member	classic_bike	427660111	774.65
3	casual	electric_bike	261382993	733.42

Proportion of Trip Duration by User Type and Bike Type



Comparison of Average Trip Duration by Bike Type between Members and Casual Riders



The column charts hightlight the total and average trip duration by user type per ride type. At the volumetric level, electric bike accounts for the the largest proportion among members, possibly as a result of it's high ride frequency engagement, contributing 31.5% however, despite the high ride volume associated with electric bike among casual riders, classic bike records a higher trip duration volume amounting to 26.8% of total trip duration recorded. Conversely, the behavioural relationship reveals that among both users, classic bike associates with higher average trip duration depicting a significant value above **2100sec** among casual riders and among members, below **800sec**. This reflects the leisure-oriented engagement on classic bike around social-centric zones by casual riders while commute-specific operation around business districts among members.

Insight 6 - Peak Ride Hours Trend

To uncover hourly commuting patterns, I examined the distribution of ride count across start and end hours.

In [132…

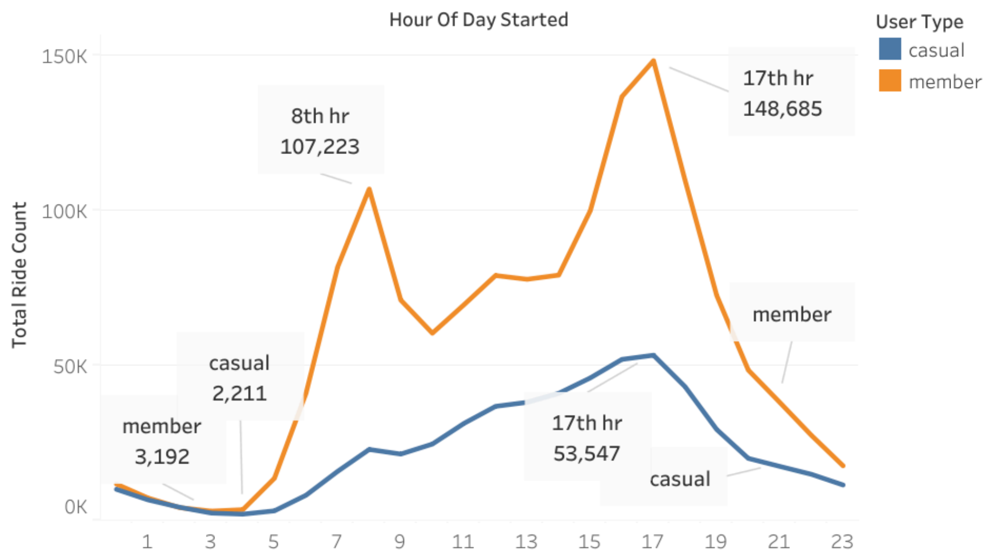
```
# Count total ride across start-to-end hour pairs of each user
query = """
SELECT
  member_casual,
  hour_of_day_started,
```

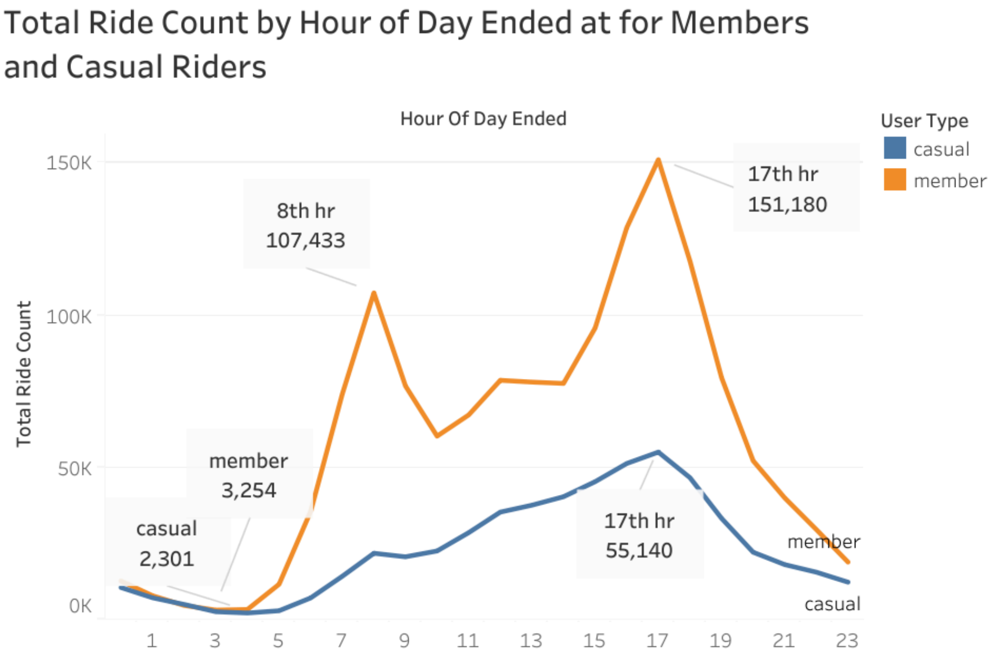
```
  hour_of_day_ended,
  COUNT(*) AS total_ride_count
FROM
 `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  hour_of_day_started,
  hour_of_day_ended
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[132…

	member_casual	hour_of_day_started	hour_of_day_ended	total_ride_count
0	member	17	17	123576
1	member	16	16	109278
2	member	18	18	93000
3	member	8	8	91435
4	member	15	15	80671

Total Ride Count by Hour of Day Started at for Members and Casual Riders





The line graphs display the ride volume across start and end hours represented in a 24 hour period scale **[0 = 12:00a.m , 12 = 12:00p.m]**, with distinct colored lines distinguishing changes across both users. During start hours, member's ride activity is lower between 2:00a.m and 4:00a.m recording a least ride volume **3192** at 3:00a.m, demand begins to rise sharply after 5:00a.m, highlighting a peak period spread from 7:00a.m before reaching a peak **[107,223]** at 8:00a.m in the morning and another spread from 3:00p.m before attaining peak **[148,685]** at 5:00p.m in the evening before declining gradually whereas across it's end hours, ride ended is lower between 2:00a.m and 4:00a.m noting a least volume **[3254]** at 3:00a.m, increases sharply after 5:00a.m, illustrating a peak spread from 7:00a.m before reaching a peak **[107,433]** at 8:00a.m in the morning and another peak spread spread from 3:00p.m before earning a peak **[151,180]** at 5:00p.m in the evening before declining gradually. Conversely, among casual riders, ride activity is lower between 2:00a.m and 5:00a.m highlighting a least value **[2211]** at 4:00a.m, ride frequency begins to increase relatively after 6:00a.m in the morning, indicating a peak spread from 12:00p.m before reaching a peak **[53,547]** at 5:00p.m in the evening before declining gradually while end hour, trends lower between 2:00a.m and 5:00a.m recording a least value **[2301]** at 4:00a.m, increases uniformly after 6:00a.m in the morning, highlighting a peak spread from 12:00p.m before reaching a peak **[55,140]** at 5:00p.m in the evening before declining gradually. This indicates a traditional work-related commuting pattern (8:00a.m and 5:00p.m) for members whereas a leisure-oriented pattern (after 5:00a.m and peak at 5:00pm) for casual riders.

Insight 7 - Hourly Trip Duration Differences

To understand hourly behavioral patterns, I explored average trip duration across start and end hour.

In [133…

```
# Calculate the average trip duration across start-to-end hour pairs for each us
query = """
SELECT
  member_casual,
```

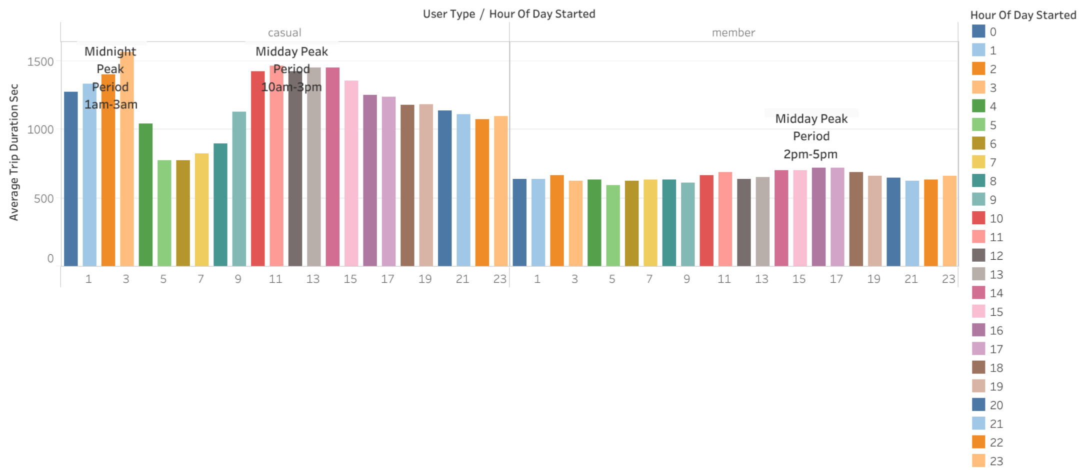
```
  hour_of_day_started,
  hour_of_day_ended,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
 `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  hour_of_day_started,
  hour_of_day_ended
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

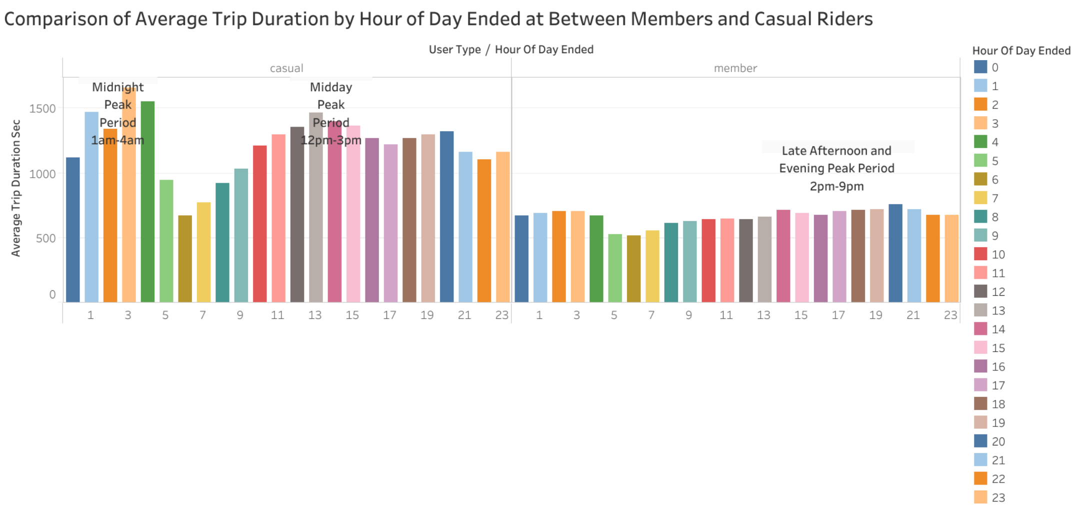
Out[133…

	member_casual	hour_of_day_started	hour_of_day_ended	avg_trip_duration_sec
0	casual	19	18	84763.0
1	casual	15	14	84529.0
2	member	16	15	83795.0
3	casual	20	19	83575.0
4	casual	16	15	83359.8

Although start-end pair analysis was performed, the results were too granular to offer clear insight. Therefore, the summary visuals of start and end patterns were used to represent user behavior more effectively.

Comparison of Average Trip Duration by Hour of Day Started at Between Members and Casual Riders





The bar graphs display the comparison of average trip duration between members and casual riders with each distinct colored bars, representing different hours of day when rides begin and end. In general, casual riders begin and end rides of longer duration, casual trips which begin at midnight periods(1:00a.m - 3:00a.m) and midday periods(10:00a.m - 3:00p.m) and end at midnight periods(1:00a.m - 4:00a.m) and midday periods(12:00p.m - 3:00p.m) associates with significant longer trip duration exceeding **1,330sec**. On the other hand, members' trip durations are evenly distributed across all hours, however, highlighting peak spread between 2:00p.m and 5:00p.m across start hours and between 2:00p.m and 9:00p.m across end hours, lasting under **760sec**. This demonstrates longer recreational-focused operation at midnight and midday periods among casual riders while members' steadier but longer ride length during late afternoon and evening hours.

Insight 8 - Prime Time of Day Pattern

To understand day time commuting trends, I inspected the shifts in ride frequency across time of day.

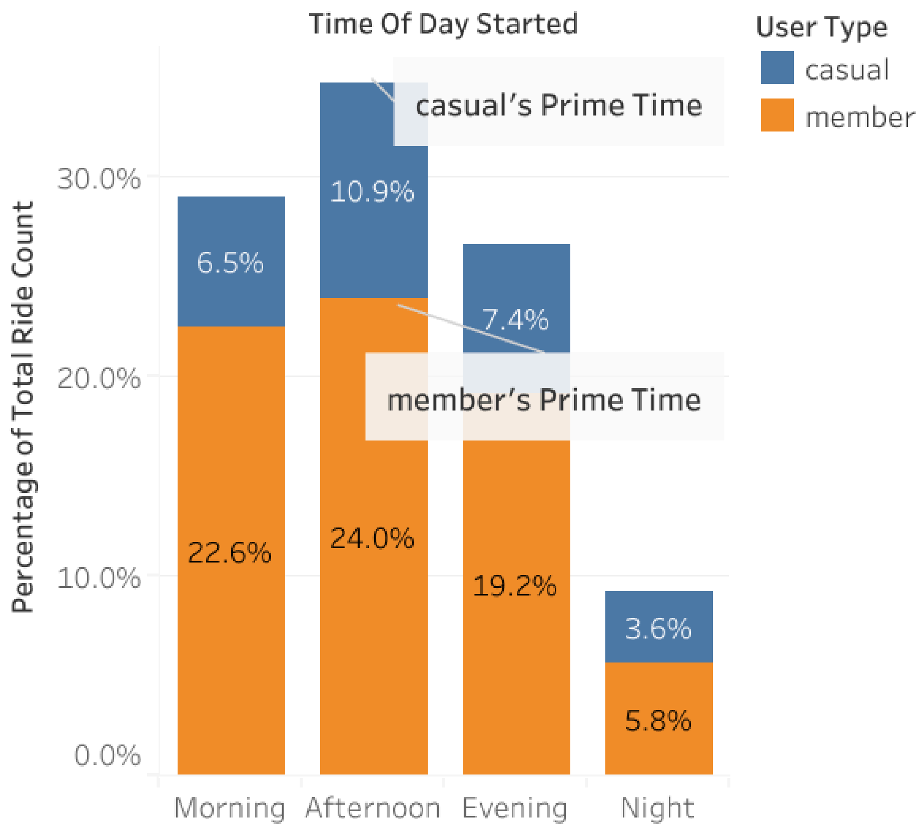
In [134…

```
# Count total ride across start-to-end day time pairs of each user
query = """
  SELECT
  member_casual,
  time_of_day_started,
  time_of_day_ended,
  COUNT(*) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  time_of_day_started,
  time_of_day_ended
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

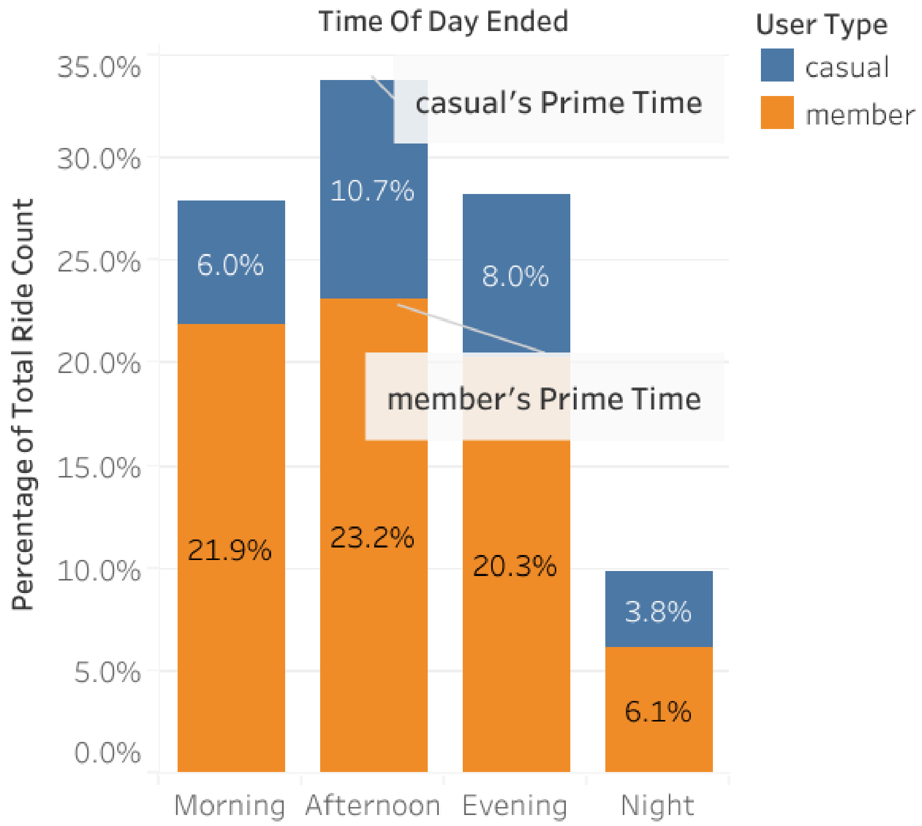
Out[134...

	member_casual	time_of_day_started	time_of_day_ended	total_ride_count
0	member	Afternoon	Afternoon	445862
1	member	Morning	Morning	432723
2	member	Evening	Evening	372756
3	casual	Afternoon	Afternoon	199628
4	casual	Evening	Evening	142190

Proportion of Ride Count by Time of Day Started at and User Type



Proportion of Ride Count by Time of Day Ended at and User Type



The stacked column charts above show the percentage proportion of ride volume across time of the day ride begins and ends, contributed by users represented by distinct colors. Overall, rides taken within afternoon window contribute the highest percentage of ride volume across both users while at night, the least. Among members, rides which begin and end at afternoon period constitutes the highest share, 24.0% and 23.2% of total ride counts across start and end period respectively. In contrast, casual riders accounts for 10.9% and 10.7% of total ride volume across start and end afternoon timeframe respectively. This implies that both users are most active in the afternoon, but members dominate overall ride activity during this period.

Insight 9 - DayTime Trip Duration Variation

To determine daytime behavioral structures, I analyzed start and end time average trip duration.

In [135…

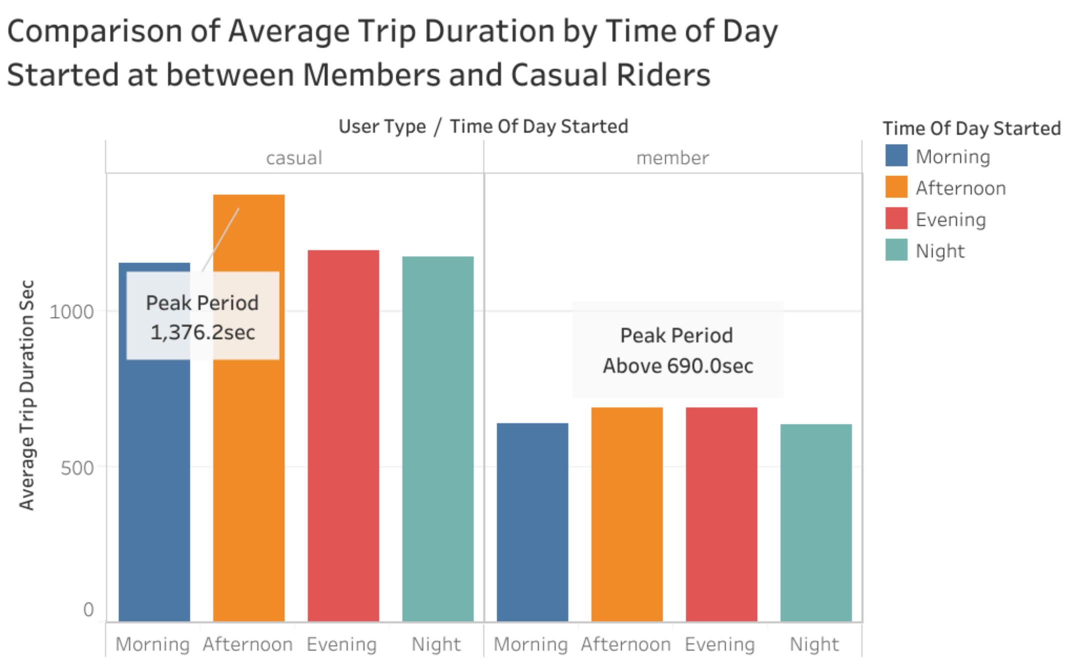
```
# Calculate the average trip duration across start-to-end daytime pairs for each
query = """
  SELECT
  member_casual,
  time_of_day_started,
  time_of_day_ended,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
```

```
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  time_of_day_started,
  time_of_day_ended
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

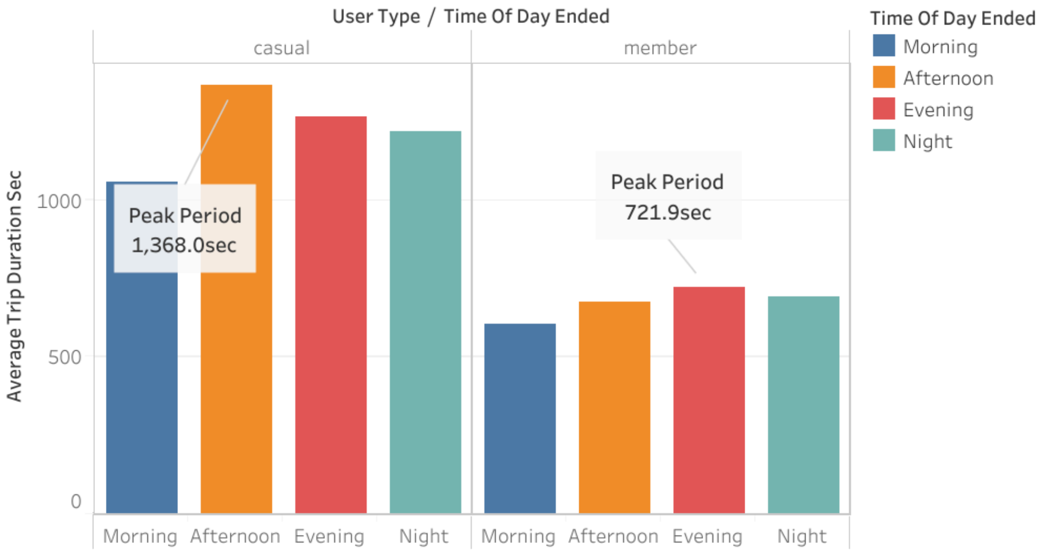
Out[135…

	member_casual	time_of_day_started	time_of_day_ended	avg_trip_duration_sec
0	casual	Evening	Afternoon	71153.82
1	member	Evening	Afternoon	70623.24
2	casual	Night	Evening	68236.90
3	casual	Afternoon	Morning	66028.33
4	member	Afternoon	Morning	65180.15

While the start and end daytime visuals below reveal that afternoon periods are generally most enjoyed among casual riders, the start-end pair combinations show few afternoon-afternoon ride record. This may result from data sparsity or overlaps across different time windows.



Comparison of Average Trip Duration by Time of Day Ended at between Members and Casual Riders



The column charts illustrate the comparison of average trip duration between members and casual riders, with each unique colored bars representing specific start and end time of the day. In general, members engage in shorter ride length, which appears as a uniform spread across start hour, however, highlight a peak at both afternoon and evening periods when longer ride begins and a peak at evening when longer ride ends. Conversely, casual riders take longer ride length, which also appears evenly distributed across start hour, however,it's trip duration realized from rides which begin and culminate at afternoon window is approximately **2x** that of members. This reaffirms the general longer pleasure-driven operation especially at afternoon window among casual riders while shorter commute-purpose-driven rides typically taken by members, with longer trip occuring at afternoon and evening periods.

Insight 10 - Rush day of Week Trend

To discover day of week travelling patterns, I investigated the ride demand across day of week.

In [136…

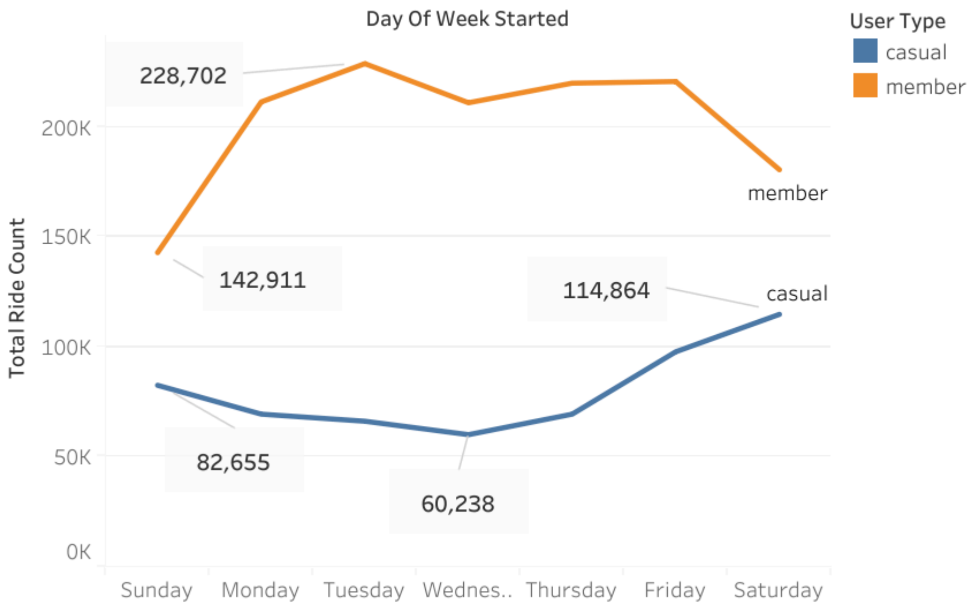
```
# Count total ride across start-to-end daily pairs of users
query = """
SELECT
  member_casual,
  day_of_week_started,
  day_of_week_ended,
  COUNT(*) AS total_ride_count
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  day_of_week_started,
  day_of_week_ended
ORDER BY
  COUNT(*) DESC
"""
```

```
df_user = Client.query(query).to_dataframe()
df_user.head()
```

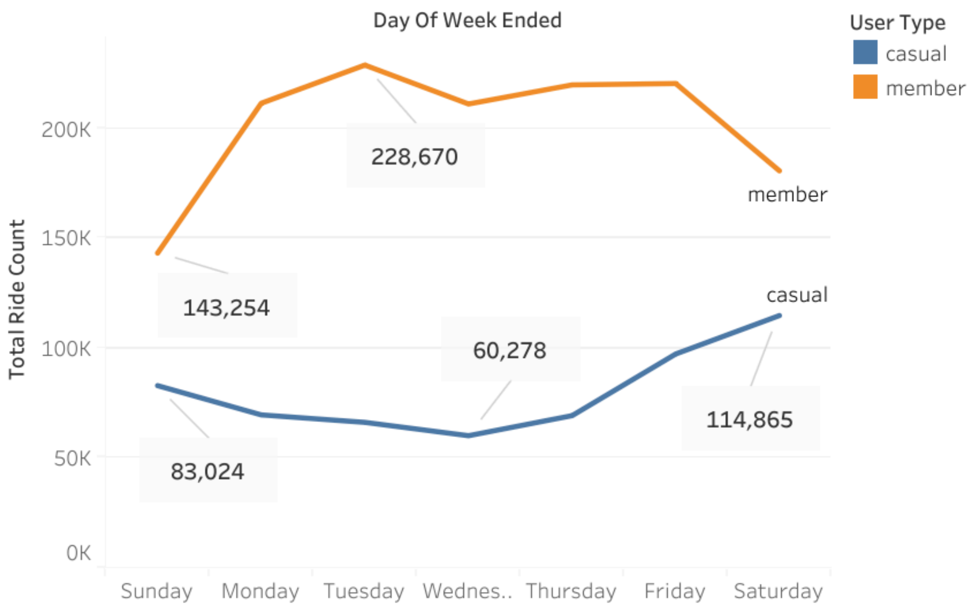
Out[136…

	member_casual	day_of_week_started	day_of_week_ended	total_ride_count
0	member	Tuesday	Tuesday	228376
1	member	Friday	Friday	219872
2	member	Thursday	Thursday	219412
3	member	Monday	Monday	211031
4	member	Wednesday	Wednesday	210672

Total Ride Count by Day of Week Started at for Members and Casual Riders



Total Ride Count by Day of Week Ended at for Members and Casual Riders



The line charts illustrates the shifts in ride counts across start and end day of the week. Across start hour among member, ride frequency is low **[142,911]** on Sunday, increases sharply on Monday, before reaching a peak **[228,702]** on Tuesday, remained constant before declining gradually after Friday whereas ride ended is equally low **[143,254]** on Sunday, increases steeply on Monday, before attaining a peak **[228,670]** on Tuesday, remained constant before declining gradually after Friday. On the other hand, casual rider's activity starts at an initial peak **[82,655]** on Sunday, before declining gradually, reaching a trough **[60,238]** on Wednesday, increases rapidly after Thursday, before concluding a final peak **[114,864]** on Saturday likewise, it's ride activity ended is initially peak **[83,024]** on Sunday, before declining gradually, earning a trough **[60,278]** on Wednesday, rises rapidly after Thursday, before attaining a final peak **[114,865]** on Saturday. This reflects work-associated commute pattern (Monday - Friday), peaking midweek among members while amid casual riders, a leisure-centric travel trend associated with peak leisure time (Friday - Sunday).

Insight 11 - Daily Trip Duration Distinctions

To identify daily behavioral configurations, I compared trip duration across day of week.

In [137…

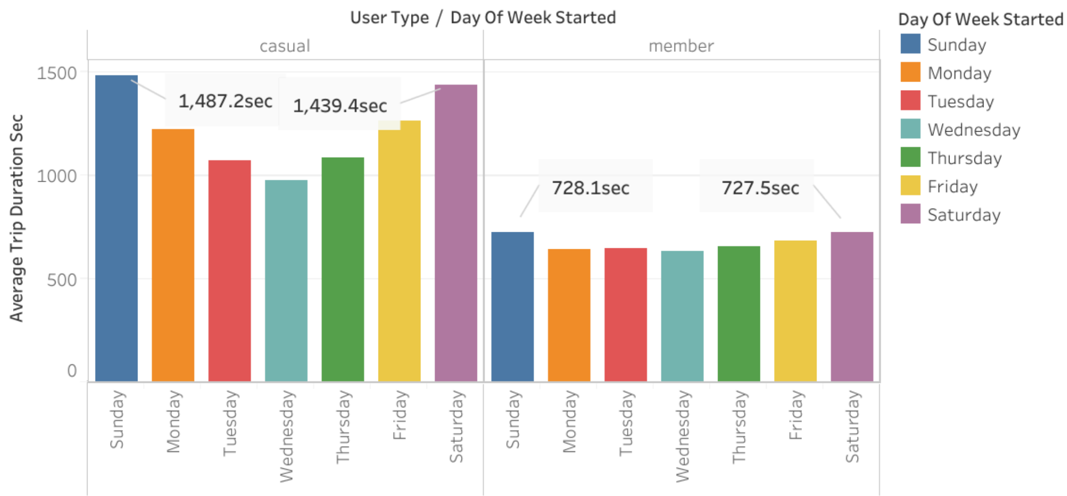
```
# Calcualate average trip duration across start-to-end daily pairs taken by user
query = """
SELECT
  member_casual,
  day_of_week_started,
  day_of_week_ended,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  day_of_week_started,
  day_of_week_ended
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[137…

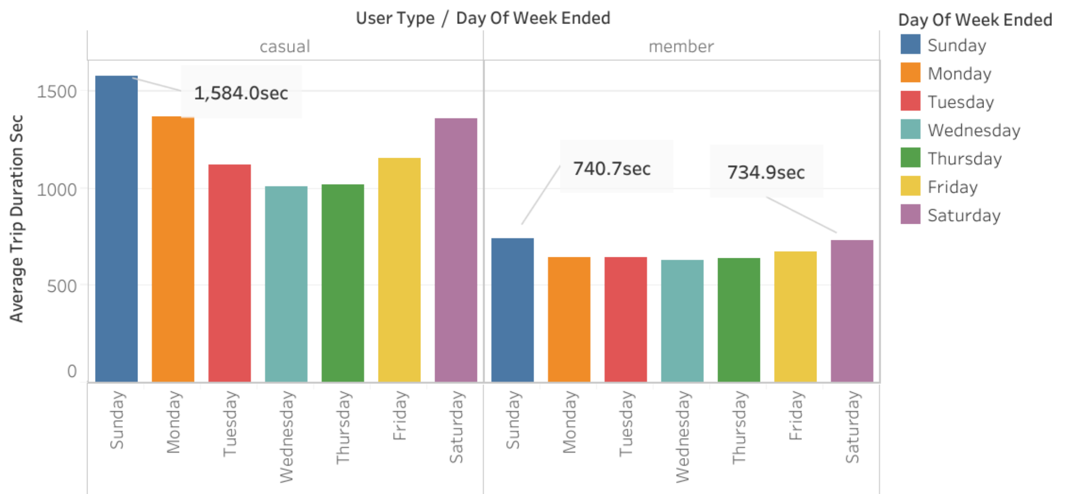
	member_casual	day_of_week_started	day_of_week_ended	avg_trip_duration_sec
0	casual	Friday	Sunday	89994.40
1	casual	Monday	Wednesday	89993.67
2	member	Wednesday	Friday	89993.00
3	member	Thursday	Saturday	89992.50
4	casual	Saturday	Monday	89991.22

Although start-end pair analysis was performed, the results were too granular to offer clear insight. Therefore, the summary visuals of start and end patterns were used to represent user behavior more effectively.

Comparison of Average Trip Duration by Day of Week Started at between Members and Casual Riders



Comparison of Average Trip Duration by Day of Week Ended at between Members and Casual Riders



The column charts show the comparison of average trip duration between members and casual riders across start and end day of week represented by unique colored bars. In general, casual riders take longer ride length, however, highlights that rides which begin between Sunday and Saturday and culminate on Sunday are heavily associated with significant higher average trip durations, lasting above **1,400sec**, in contrast, amid members, trip durations are shorter and relatively distributed across both start and end day of week, nevertheless, illustrates longer trip duration of above **720sec** among rides which begin and end within Saturday - Sunday timeframe. This reveals that members take short-distance commutes which are balanced through the week, with slightly longer rides on Sunday and Saturday whereas casual riders take longer relaxational-oriented ride which often begins on Saturday and Sunday and end especially on sunday.

Insight 12 - Highest Demand Week Part Trend

To understand weekly travel changes, I examined the spread of ride activity across week parts.

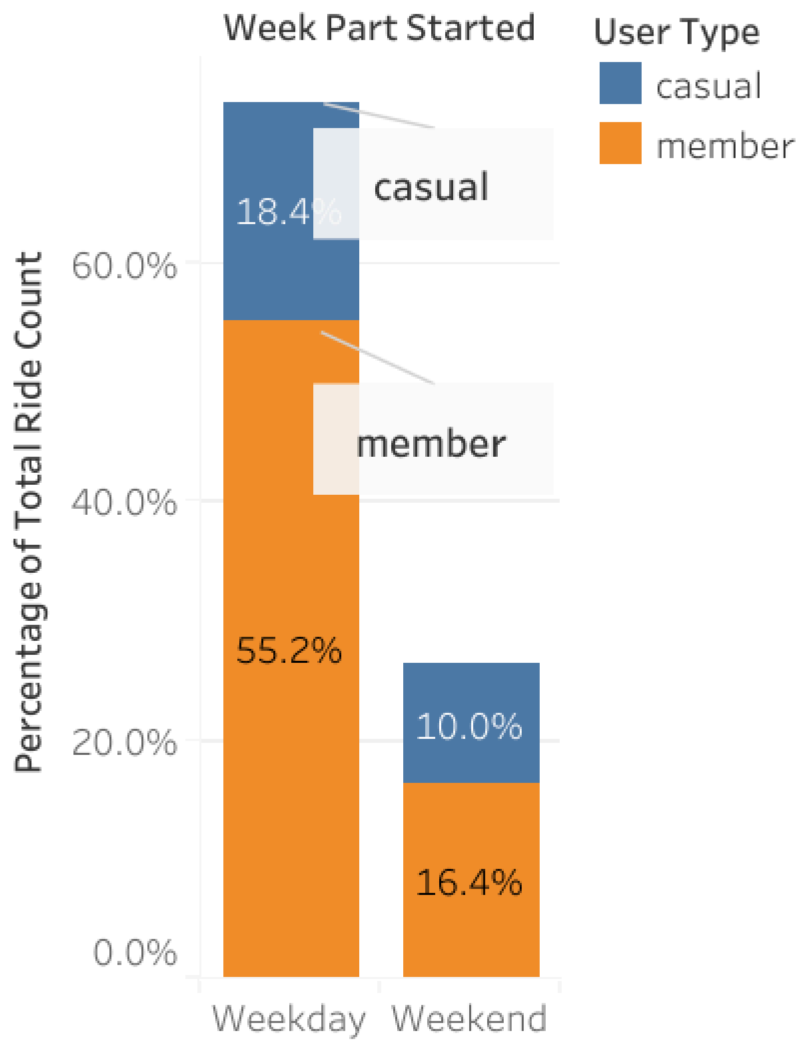
In [138…

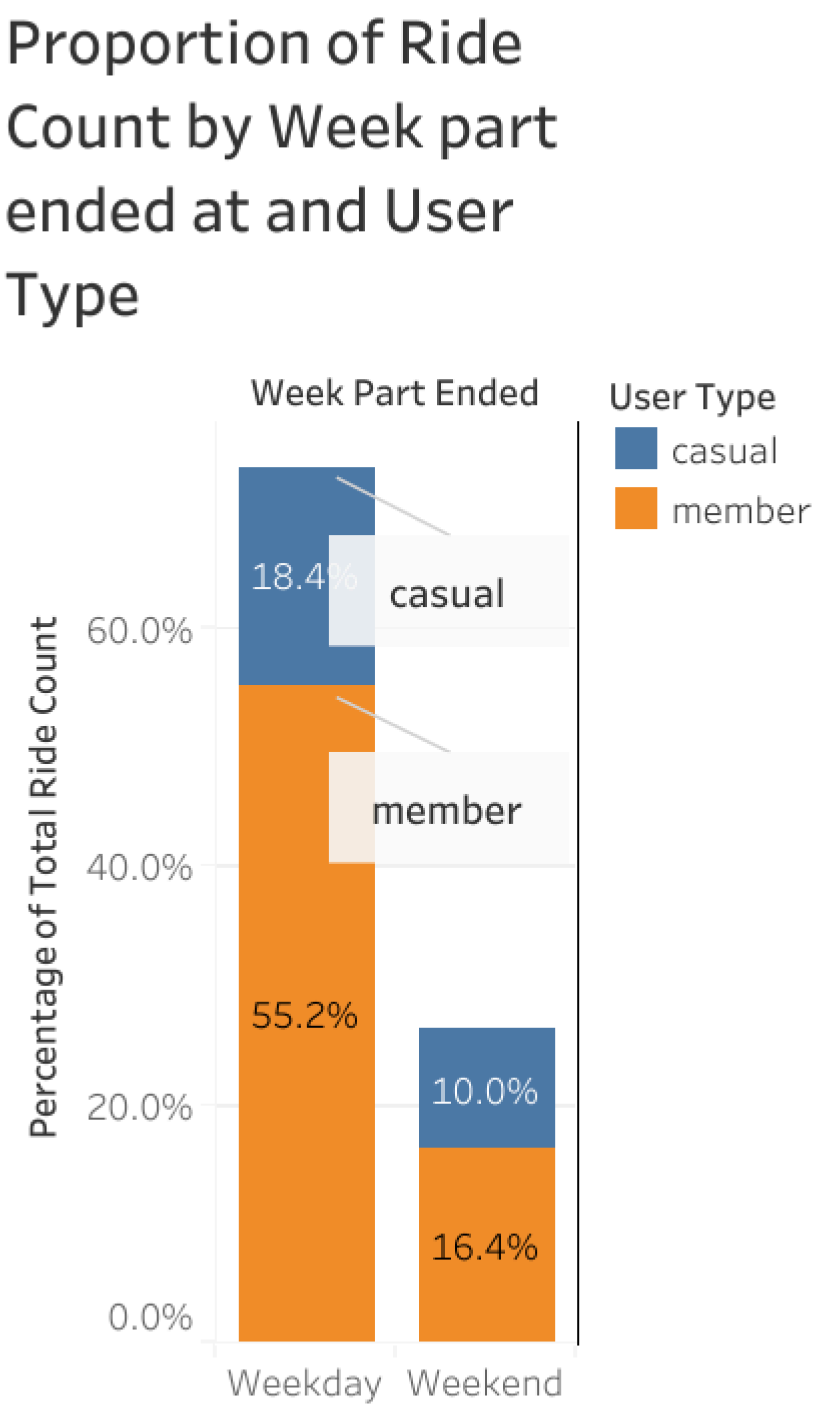
```python
# Count total ride across start-to-end weekpart pairs of each user
query = """
SELECT
  member_casual,
  week_part_started,
  week_part_ended,
  COUNT(*) AS total_ride_count
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  week_part_started,
  week_part_ended
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[138…

	member_casual	week_part_started	week_part_ended	total_ride_count
0	member	Weekday	Weekday	1090675
1	casual	Weekday	Weekday	362584
2	member	Weekend	Weekend	323163
3	casual	Weekend	Weekend	196932
4	casual	Weekday	Weekend	957

Proportion of Ride Count by Week part Started at and User Type





The stacked column charts above display the percentage proportion of ride volume across start and end week parts contributed by users, represented by unique colored bins. Overall, ride activity of both users majorly dominates weekday periods likely due to large day composition than weekend, with members accumulating the largest proportion, 55.2% of total start and end ride count while casual riders constituting 18.4%. This suggests both users are most engaged during weekday, but members influence general ride engagement during this period.

Insight 13 - Weekly Trip Duration Contrast

To uncover weekly behavioral shifts, I explored weekly trip duration across users.

In [139…

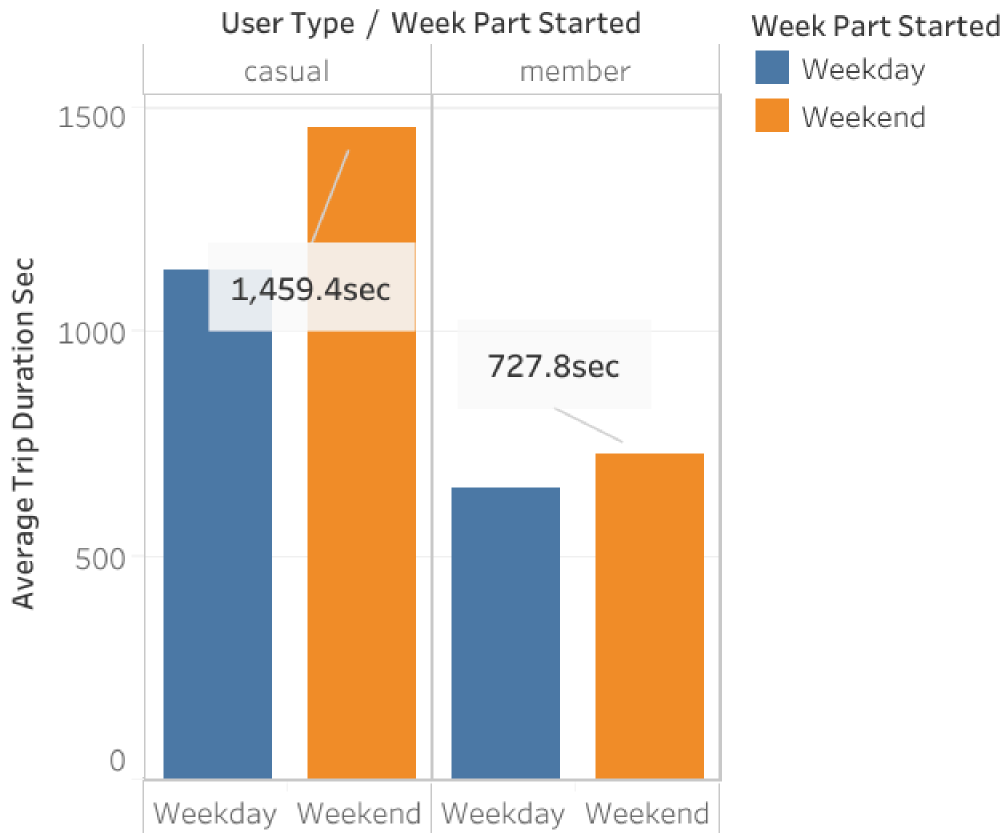
```
# Calculate average trip duration across start-to-end weekpart pairs for each us
query = """
SELECT
  member_casual,
  week_part_started,
  week_part_ended,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
    `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  week_part_started,
  week_part_ended
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

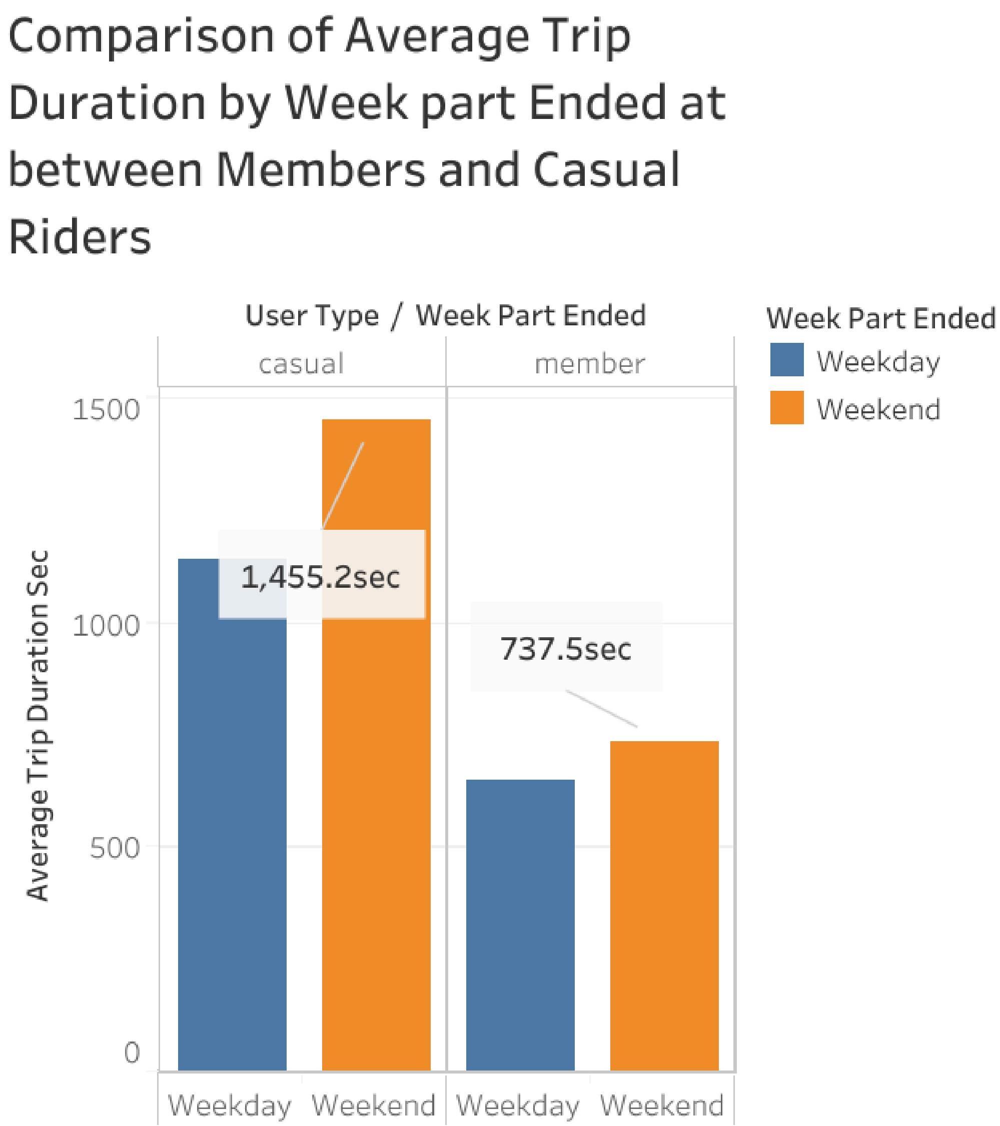
Out[139…

	member_casual	week_part_started	week_part_ended	avg_trip_duration_sec
0	casual	Weekend	Weekday	49363.95
1	casual	Weekday	Weekend	29969.86
2	member	Weekend	Weekday	16752.23
3	member	Weekday	Weekend	11690.76
4	casual	Weekend	Weekend	1316.63

While the start and end week part visuals below reveal that weekend periods are generally most enjoyed among both users, the start-end pair combinations show few weekend-weekend ride observation. This may result from data sparsity or overlaps across different time windows.

Comparison of Average Trip Duration by Week part Started at between Members and Casual Riders





The column charts highlight the comparison of average trip duration between members and casual riders across start and end week parts distinguished by color contrast. Overall, both users engage in longer trip duration across start-to-end weekend window, however, rides associated with casual riders accounts for longer trip duration, exceeding **1,450sec** whereas under **740sec** among members, reflecting significant longer recreational-oriented activity among casual riders during weekend while amid members, a shorter travel-centric usage at same period.

Insight 14 - Busiest Month Pattern

To discover monthly commuting patterns, I inspected the ride demand across start and end months.

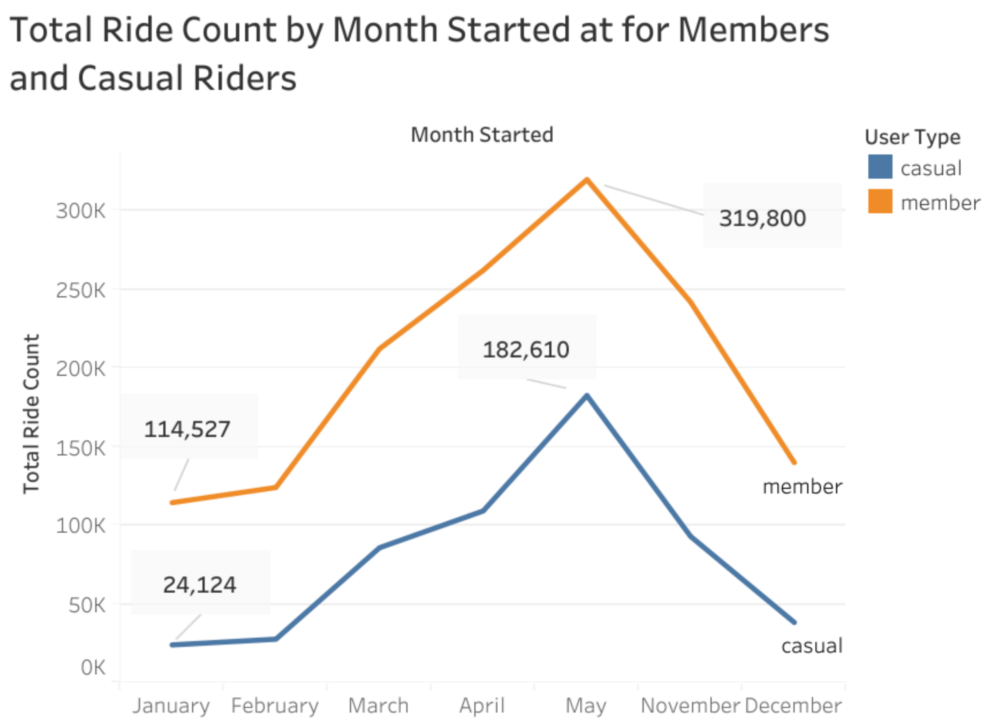
In [140…

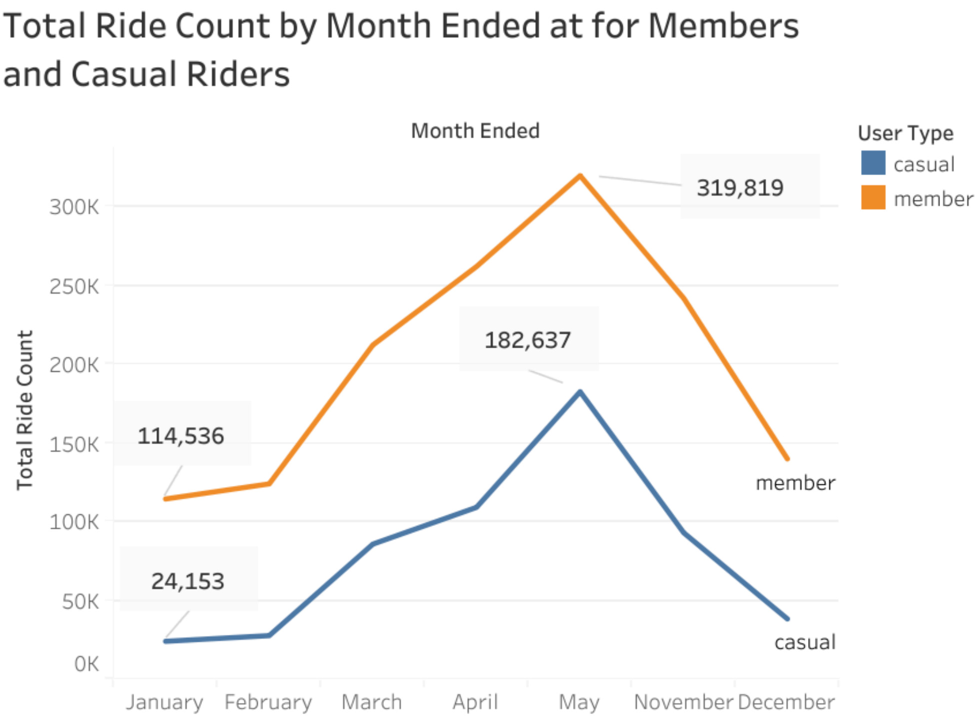
```
# Count total ride across start-to-end month pairs for each user
query = """
SELECT
  member_casual,
  month_started,
  month_ended,
  COUNT(*) AS total_ride_count
FROM
```

```
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  month_started,
  month_ended
ORDER BY
  COUNT(*) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[140…

	member_casual	month_started	month_ended	total_ride_count
0	member	May	May	319800
1	member	April	April	262118
2	member	November	November	241974
3	member	March	March	212261
4	casual	May	May	182610





The line charts above illustrate the changes in ride demand across start and end months. Ride demand among members is low **[114,527]** in January, increases rapidly after February, highlighting a peak period spread from March, before reaching a peak **[319,800]** in May possibly as a result of spring seasonal influence, before declining gradually likewise across end months, ride count is low **[114,536]** in January, rises sharply after February, revealing a peak period spread from March, before attaining a peak **[319,819]** in May, then decline gradually through November. On the other hand, across start month, ride volume in January is low **[24,124]**, begins to increase gradually after February, before peaking **[182,610]** in May, followed by a gradual decline, similarly, rides ended are low **[24,153]** in January, increases steadily after February, before it peaks **[182,637]** in May, followed by a gradual decline through November. This reaffirms the general energetic ride activity engagement during spring season by both users, however, members dominate the overall ride volume throughout the season.

Insight 15 - Monthly Trip Duration Differences

To identify monthly behavioral trends, I explored average trip duration across start and end months.

In [141…

```
# Calculate the average trip duration across start-to-end month pairs for each u
query = """
SELECT
  member_casual,
  month_started,
  month_ended,
  ROUND(AVG(trip_duration_sec), 2) AS avg_trip_duration_sec
FROM
  `axiomatic-set-467921-a9.Cyclistic_Bikes.Trip_Data_summary`
GROUP BY
  member_casual,
  month_started,
  month_ended
```

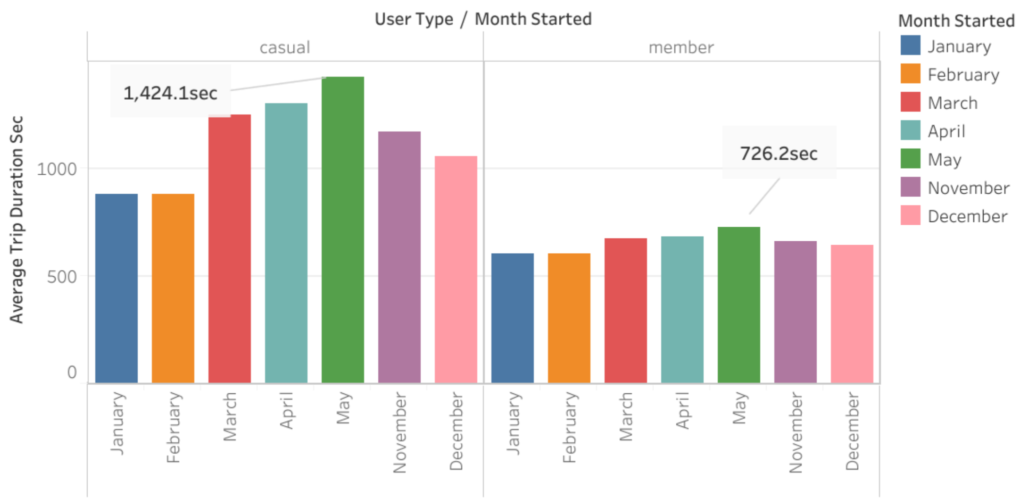
```
ORDER BY
  ROUND(AVG(trip_duration_sec), 2) DESC
"""
df_user = Client.query(query).to_dataframe()
df_user.head()
```

Out[141…

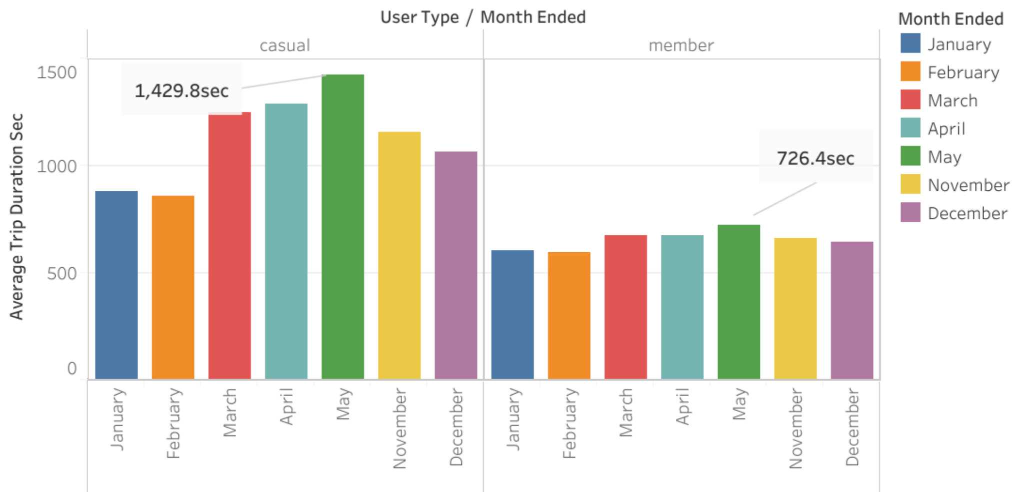
	member_casual	month_started	month_ended	avg_trip_duration_sec
0	casual	November	December	71211.43
1	casual	April	May	39904.04
2	casual	February	March	39581.27
3	member	February	March	18738.00
4	member	November	December	16434.83

While the start and end month visuals below reveal that May periods are generally most enjoyed among both users, the start-end pair combinations show few May-May ride observation. This may result from data sparsity or overlaps across different time windows.

Comparison of Average Trip Duration by Month Started at between Members and Casual Riders



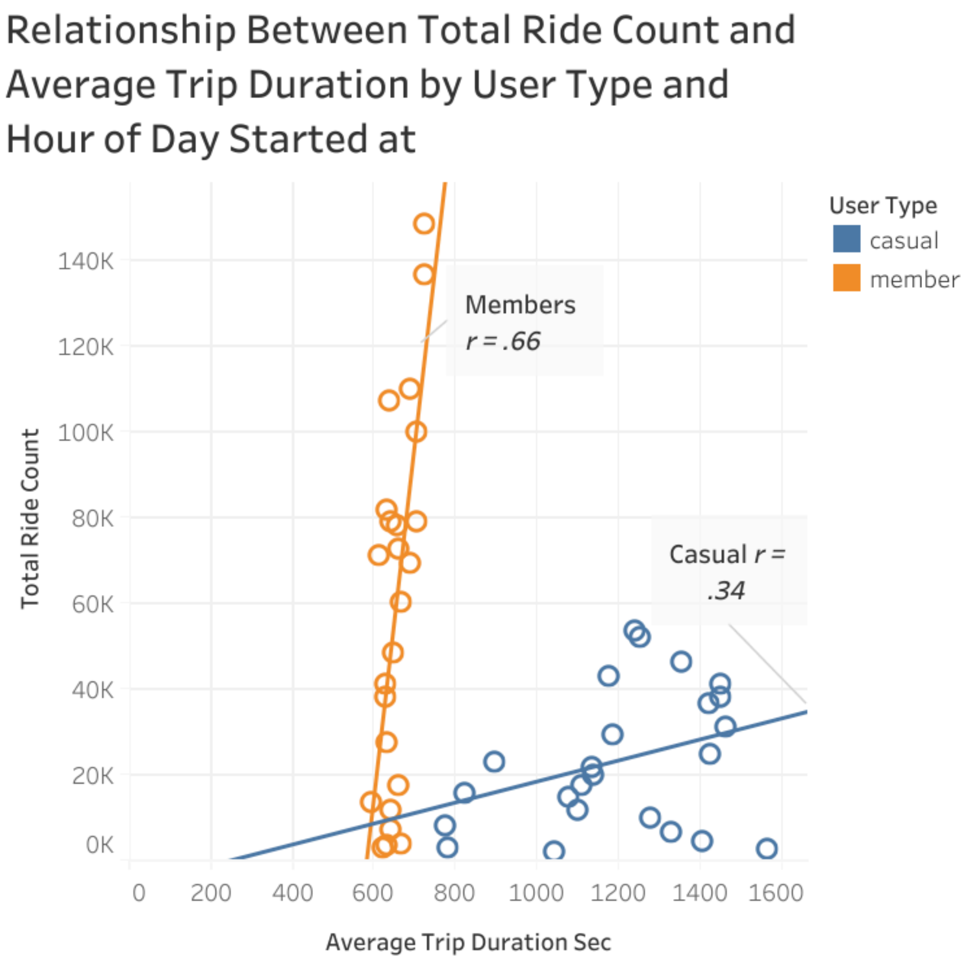
Comparison of Average Trip Duration by Month Ended at between Members and Casual Riders



The column charts above show the comparison of average trip duration between members and casual riders across start and end months distinguished by unique colored columns. In general, casual riders engage in longer trip duration than members, illustrating that among members, ride durations are shorter and uniformly dispersed across both start and end month, however, highlight longer length across rides which begin and end particularly in May, conversely, amid casual riders, rides which often operate within March - May window are more consistent, indicating May trips to be longer, futhermore, approximately **2x** the duration of members. This reflects shorter and stable ride length across month, with slightly longer rides in May among members whereas demonstrates a prime utilization of March - May seasonal offerings for longer pleasure-driven rides among casual riders - especially in May.

Insight 16 - Start Hour Ride Count and Trip Duration Relationship

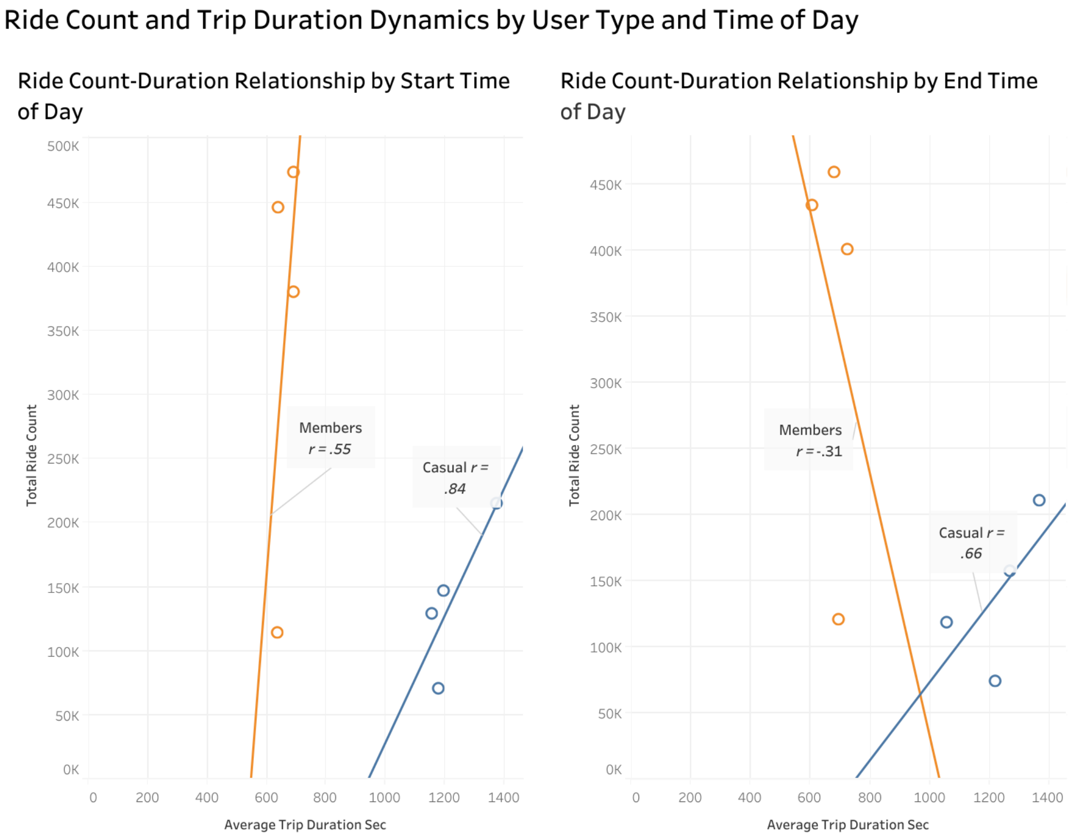
To understand start hour relationship trends, I explored the relationship between total ride count and average trip duration across hours when ride begins.



The scatterplot above illustrates the relationship between start hourly ride count and average trip duration, with each distinct colored point representing distinct user's hour of the day, plotted according to it's corresponding values on both metrics. The distribution shows a moderate positive relationship - as ride count increases, average trip duration also tends to moderately increase among members, in contrast, a weak positive relationship - no clear correlation is observed amid casual riders. Among members, high-demand start hours (8:00a.m and 3:00p.m-6:00p.m) are characterized by slightly longer trips, while longer durations occur during early morning (3:00a.m) and early midday (11:00a.m, 1:00p.m and 2:00p.m) hours, when ride activity is slightly high. This suggests across start hours, members show steady engagement, with longer usage translating into slightly longer rides during key commuting hours, whereas casual riders' weaker relationship suggests more irregular, leisure-driven usage patterns.

Insight 17 - DayTime Ride Count and Trip Duration Association

To discover the daytime association patterns, I compared how ride count interrelates with trip duration across start and end time of the day.

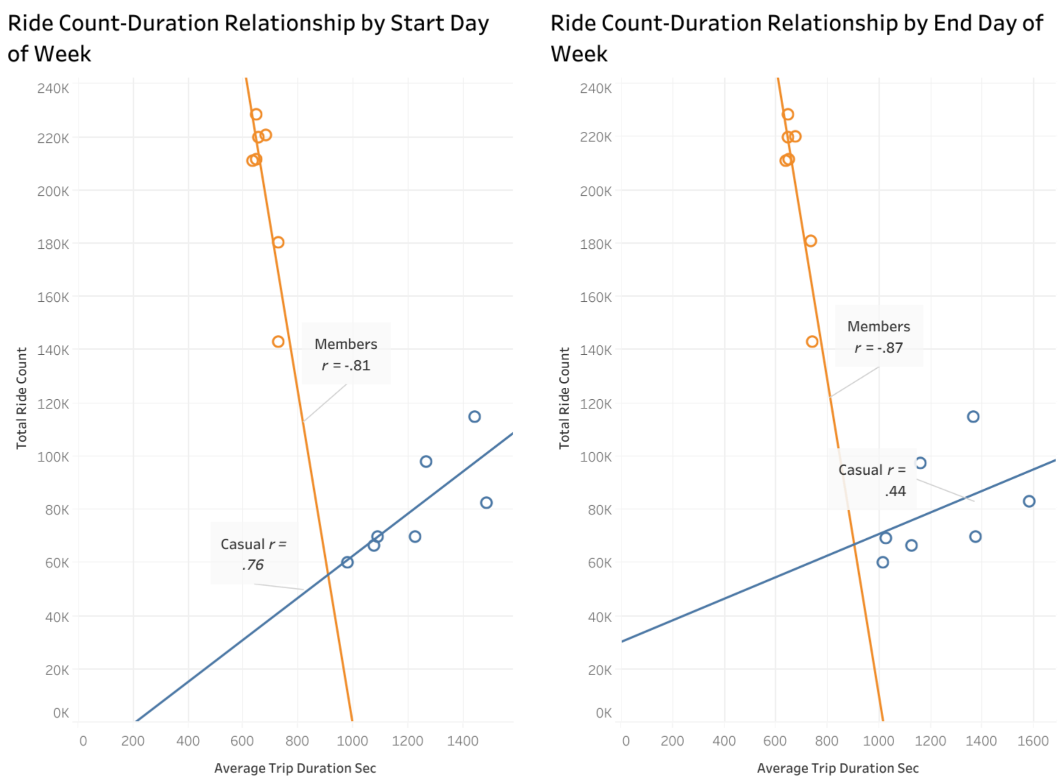


The figures above visualizes the relationship between ride demand and average trip duration, with each unique colored point representing specific user's time of day in which rides begin and end. The plotted points highlight among casual riders, a strong positive correlation across start day time while a moderate positive correlation across end day time, in contrast, among member, a moderate positive correlation across start day time while a moderate negative correlation across end day time. This indicates that among casual riders, longer trips often occur during (afternon) period of higher ride activity, especially at the start of rides while among members, ride beginning during busy period(morning and afternoon) do not necessarily end with longer durations, implying casual riders' behavior is activity-driven while members' rides are time-efficient.

Insight 18 - Daily Ride Count and Trip Duration Dependency

To identify daily connection configuration, I examined the connection between daily ride demand and length.

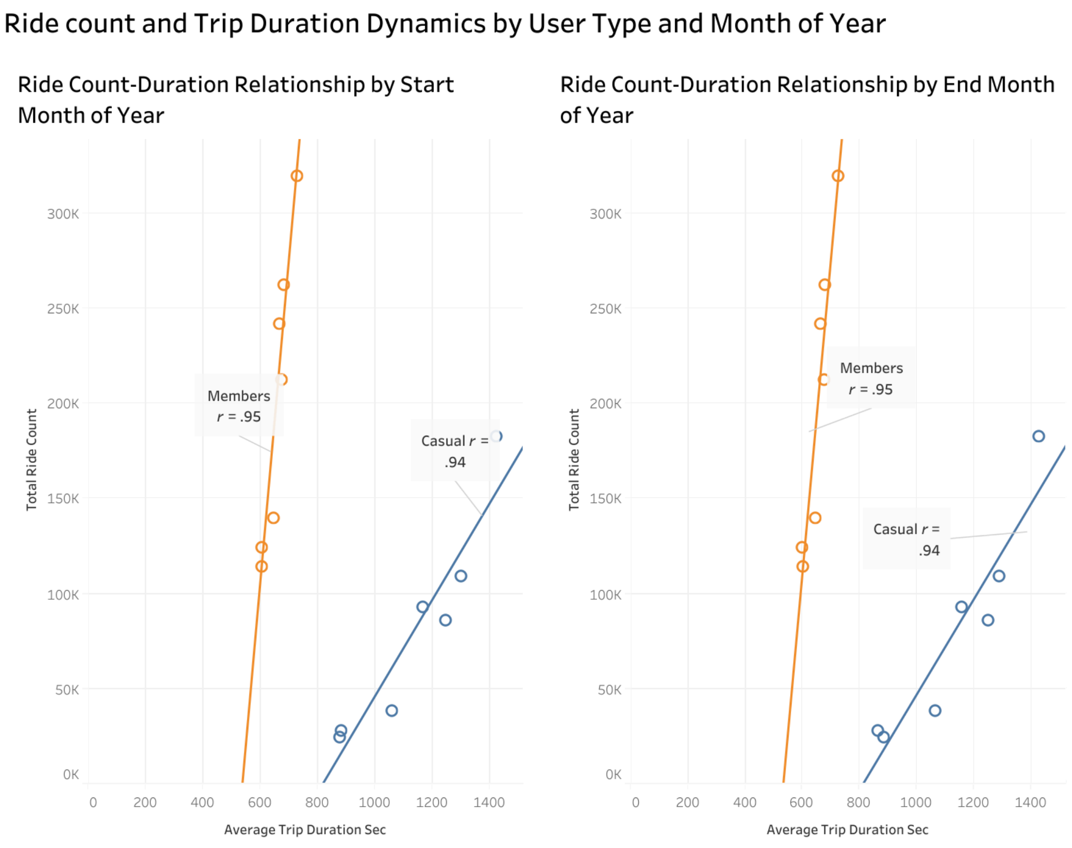
Ride Count and Trip duration Dynamics by User Type and Day of Week



The scatterplots above highlight the relationship between ride volume and average trip duration, with each distinct colored point representing specific user's start and end day of week in which rides begin and end. Among members, a strong inverse relationship is observed across both start and end day of week - higher ride counts are associated with shorter trip duration, on the other hand, amid casual riders, a strong positive correlation is observed across start day of week while a moderate positive correlation across end. This illustrates that during peak period (Tuesday and Friday), members tend to make shorter, more frequent trips, reflecting routine commuting behavior whereas, among casual riders, periods (Saturday-Sunday) with longer trip duration also record higher ride activity especially at the begining of rides - indicating leisure-focused usage patterns.

Insight 19 - Monthly Ride Count and Trip Duration Correlation

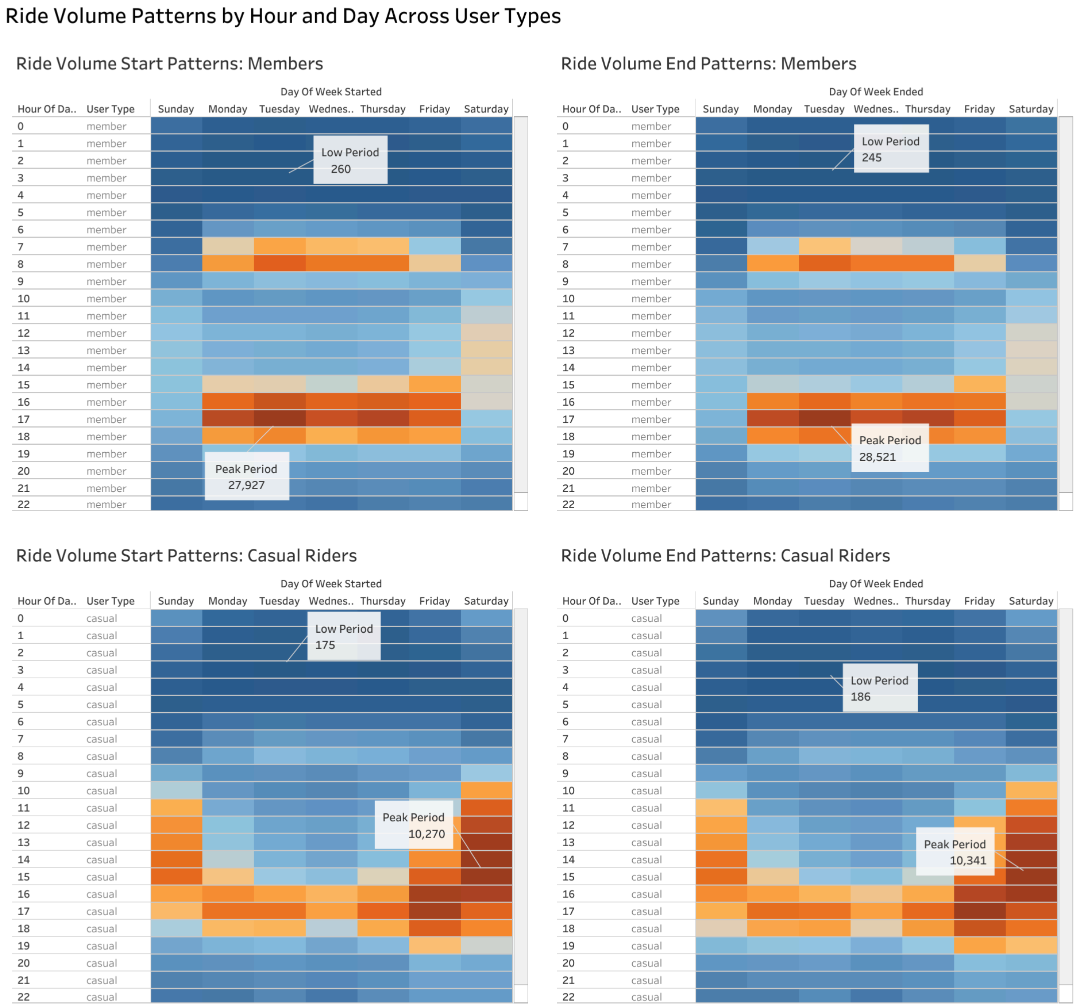
To determine monthly interrelation structure, I investigated ride frequency and travel distance relationship across start and end months.



The figures above display the connection between monthly ride volume and average trip duration, with each distinct colored point, highlighting definite user's month. Overall, the plotted points reveals a strong positive relationship across both start and end month and both users - as ride volume increases, average trip duration also tends to rise across the months. High ride activity month (May) are dominated by longer duration among both users, suggesting higher engagement and longer rides during peak months.

Insight 20 - Weekly Ride Patterns

To understand weekly travel trends, I visualized weekly ride distribution across start and end hours of the day.

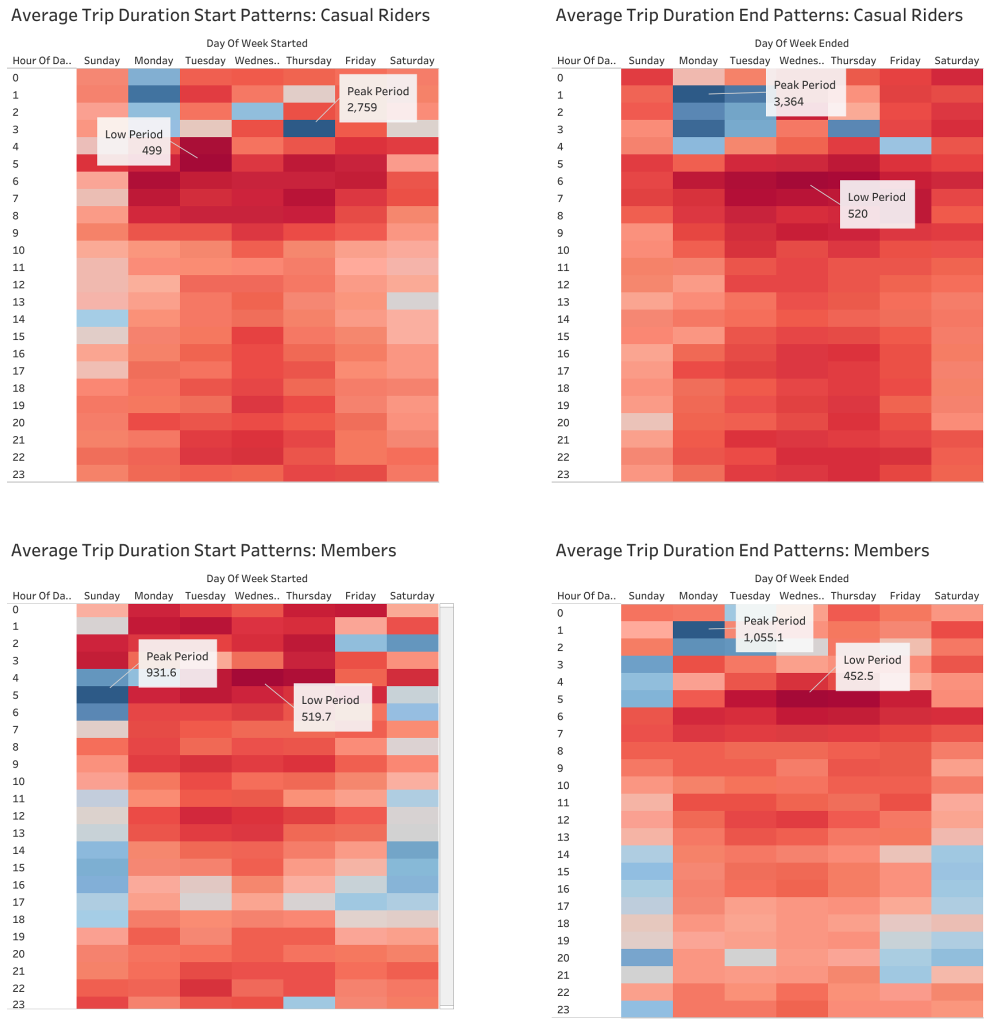


The heatmap visualizes the distribution of ride demand across different hours of the day and days of the week. Each cell represents a unique combination of hour and day in which rides begin and end, with color intensity depicting ride count magnitude, futhermore, darker orange shades correspond to higher ride counts, while darker blue shades correspond to lower ride demand. The plot reveals high concentration of member's ride activities during morning (7:00a.m and 8:00a.m) and evening (3:00p.m - 6:00p.m) on weekday across both start and end day of week and it's corresponding hours whereas low ride frequency is observed between 12:00a.m and 4:00a.m throughout the week. Conversely, amid casual riders, high ride volume clusters densely during early afternoon and evening periods (10:00a.m - 7:00p.m) through the days of the week while low counts are concentrated during early morning periods(12:00a.m - 5:00a.m) across both start and end day of weeek. This demonstrates weekday commutes among members during morning and evening hours while leisure-related usage during afternoon and evening periods across the week among casual riders.

Insight 21 - Weekly Trip Duration Trends

To uncover weekly ride length variations, I visually illustrated weekly ride length share across start and end hours of the day.

Average Trip Duration Patterns by Hour and Day Across User Types



The heatmaps above highlight the variation of average trip duration across specific hours of the day and days of the week. Each cell displays a definite combination of hour and day in which rides begin and culminate, color intensity indicates ride volume magnitude, furthermore, warmer blue shades represent higher ride counts, whereas warmer red shades indicate lower ride demand. Among members, longer ride lengths spread across hours varying between 2:00a.m - 6:00a.m in the morning and throughout early afternoon to evening (11:00p.m - 6:00p.m), across Friday - Sunday, however, highlight clusters of low duration during early morning (12:00a.m - 6:00a.m) on weekday across start of rides whereas across the end, various cluster of high average duration is observed between Friday and Tuesday during early morning (12:00a.m - 6:00a.m) and throughout the rest hours from midday (2:00p.m - 11:00p.m) while low ride length clusters between 5:00a.m and 7:00a.m on weekday. On the other hand, across start rides, high ride activities cluster during 12:00a.m and 3:00a.m on Monday, Wednesday and Thursday while low trip lengths cluster during 4:00a.m and 8:00a.m on weekday, in contrast, ride ended indicate longer trip duration between 1:00a.m and 4:00a.m on Monday, Tuesday, Thursday and Friday while shorter length during 5:00a.m and 9:00a.m on weekday among casual riders. This implies that members typically use the service for balanced weekday commuting and weekend leisure from morning through evening, while casual riders engage longer in late-night hours on weekday, suggesting social or recreational driven activities.

5. Summary and Recommendations

Summary

This section summarizes the key takeaways derived from the analytical findings and highlights what the insights collectively reveal about user behavior and business performance.

Objective Recap

The primary goal of this analysis was to understand ride usage behavior among **members** and **casual riders**, identify key differences in their riding patterns and provide actionable insights that can help the business increase membership conversion and operational efficiency.

Key Analytical Summaries

- **Riding Patterns:**
  Members typically engage in weekday rides during commuting hours whereas casual riders are more active during weekend at afternoon and evening hour, suggesting leisure-oriented usage.

- **Time-Based Ride Duration Trends:**
  Casual riders have significantly longer average trip duration on weekdays notably at late night period while extended trip occur during midnight and midday weekend periods, particularly in the spring months indicating recreational or exploratory rides in contrast members take short, steadier ride across all months, with slightly longer duration all through weekend hours and May period.

- **Station Usage:**
  Streeter Dr & Grand Ave, located near tourist attractions, parks and entertainment centers accounts for the high-demand station for casual riders, while members primarily start and end at **Kingsbury St & Kinzie St** situated around residential, business and educational districts. **University Ave & 57th St - Ellis Ave & 60th St** highlights member's common ride route whereas casual riders major ride route taken indicates - **Dusable Lake Shore Dr & Monroe St - Dusable Lake Shore Dr & Monroe St**. **Elizabeth St & 47th St** and **Wood St & 84th St** accounts for the station where longer duration rides begin among members and casual riders resectively however, both users ended longer rides in **New Hastings**.

- **Bike Preference Patterns:**
  Both members and casual riders operate electrc bike during short regular ride trips while classic bike is utilized for longer trip duration, suggesting leisure-specific engagement.

- **Temporal Trends:**
  Across both users, peak ridership occurs in warmer months, particularly during spring season, highlighting the influence of seasonality on ride demand.

- **User Engagement Insights:**
  Members exhibit predictable, routine-based ride behavior, whereas casual riders demonstrate spontaneous, experience-driven usage patterns.

Overall Interpretation

Overall, the analysis reveals two distinct customer segments:

1. **Members** - primarily cummuters who rely on the service for daily travel.
2. **Casual Riders** - mostly leisure-oriented user whose behaviors are influenced by time, weather and location attractiveness.

These behavioral distinctions provide clear opportunities for targeted marketing, operational planning and product design.

Recommedations

Based on the findings, the following data-driven recommendations are proposed to support business growth, operational optimization and membership conversion:

---

1. Marketing and Membership Conversion

- Launch **targeted membership campaigns** in **Streeter Dr & Grand Ave** where casual riders often begin trips, to attract frequent casual riders.

- Introduce **flexible membership plans** (weekend or 3-day plan) to appeal to the casual segment.

- Initiate **route-based offers** to casual riders who take **Dusable Lake Shore Dr & Monroe St - Dusable Lake Shore Dr & Monroe St** ride route.

- Employ **advertisement bill boards** near recreational hubs and high-leisure areas.

- Tailor **targeted membership benefits** to casual riders in **Wood St & 84th St**, where longer trip duration rides often begin.

- Offer **referral programs** to encourage casual riders to convert to members.

- Deploy **promotional banners** at all mentioned stations to encourage casual riders to explore membership plans.

2. Operational Optimization

- Strengthen operational planning at **New Hastings**, where longer trip duration end, to ensure ride availability, reduce wait time and improve turnaround efficiency.

- Use ride duration data to optimize **maintenance scheduling** and **redistribution logistics**.
- Rebalance **electric bikes** toward popular high demand station and **classic bikes** toward stations associated with longer duration in alignment with respective time-demand to meet casual riders' need.

3. Customer Engagement and Retention

- Send **personalized in-app notifications** highlighting potential savings for casual riders who frequently use the sevice.
- Introduce **ride streaks or rewards** to increase ride frequency and build brand attachment.
- Implement **user satisfaction surveys** to gather qualitative insights on casual rider's motivations and barriers to memberships.

4. Strategic and Long-term Actions

- Patner with **parks, museum, local events malls and tourism boards** to promote seasonal campaigns tied to casual rider activity.
- Continously monitor start-end flows to identify **new high-traffic corridors** for potential station expansion.
- Explore the feasibility of **night operations** in areas with consistent late-night ridership.
- Periodically evaluate **rider behavior over time** to measure the impact of implemented recommendations and refine business strategies.

Conclusion

This analysis provides a strong foundation for understanding customer segmentation and optimizating service delivery.
By leveraging these insights, the company can **increase membership conversion**, **enhance rider satisfaction** and **achieve sustainable operational efficiency**.

In [146…

```
import os
import requests
import re

# This script finds Google Drive links and downloads them as local files
def fix_drive_images(notebook_path):
    with open(notebook_path, 'r') as f:
        content = f.read()

    # Finds the IDs from my Google Drive links
    links = re.findall(r'id=([a-zA-Z0-9_-]+)', content)

    for file_id in set(links):
        url = f'https://drive.google.com/uc?export=download&id={file_id}'
```

```python
        img_name = f'{file_id}.png'
        # Downloads the image to Colab
        with open(img_name, 'wb') as f:
            f.write(requests.get(url).content)
        # Updates the notebook to point to the local file instead of the link
        content = content.replace(f'https://drive.google.com/uc?export=view&id={
        content = content.replace(f'https://drive.google.com/uc?export=download&

    with open('Fixed_Notebook.ipynb', 'w') as f:
        f.write(content)
    print("Success! Your images are now localized.")

# Replace this with my actual path
fix_drive_images('/content/drive/MyDrive/Colab Notebooks/Google_Data_Analytics_C
```

```
Success! Your images are now localized.
```

In [147…

```python
# This uses the notebook that ALREADY has the working images
!jupyter nbconvert --to html --embed-images Fixed_Notebook.ipynb --output 'Cycli
```

```
[NbConvertApp] Converting notebook Fixed_Notebook.ipynb to html
[NbConvertApp] Writing 5200519 bytes to Cyclistic_Bike_Share_Technical_Analysis_N
ov2024_May2025.html
```