Cyclistic Case Study

Scenario

I’m currently working as a junior data analyst on the marketing team at Cyclistic, a bike-share company based in Chicago. The director of marketing has identified that the company’s future growth hinges on increasing the number of annual memberships. To help with this, my team is focused on understanding how casual riders and annual members use the bikes differently. The goal is to use these insights to develop a new marketing strategy that can encourage more casual riders to become annual members.

Before I can move forward with any marketing initiatives, I need to present my findings to Cyclistic’s executives. They need to see data-backed recommendations, which means I must deliver clear insights supported by professional data visualizations that show the value of converting casual riders into long-term members.

Characters and teams

● Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use the bikes to commute to work each day.

● Lily Moreno: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

● Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. I joined this team six months ago and have been busy learning about Cyclistic’s mission and business goals—as Ill as how you, as a junior data analyst, can help Cyclistic achieve them.

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

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

Business Task

The business task is to analyze Cyclistic's historical bike trip data to identify key differences in usage patterns between casual riders and annual members. The goal is to use these insights to inform marketing strategies that encourage casual riders to convert to annual memberships. This analysis should result in data-backed recommendations, along with professional data visualizations, to illustrate the potential value of converting more casual riders into long-term members for Cyclistic's executive team.

Key Stakeholders:

- **Lily Moreno (Director of Marketing)** – Leads the marketing strategy and campaigns for Cyclistic.
- **Cyclistic Marketing Analytics Team** –Works to dig into the data and uncover insights that guide the marketing decisions.
- **Cyclistic Executive Team** – Will either approve or reject the marketing proposals.
- **Casual Riders** – The people Cyclistic are hoping to convert into annual members.
- **Annual Members** – Loyal customers who already use the service long-term.
- **Cyclistic Finance Team** – Responsible for ensuring that the marketing strategy makes financial sense.
- **Cyclistic Operations Team** – Manage the logistics of the bike-share program.

Data & Licensing

Using Cyclistic's historical trip data to analyze and identify trends. Here is a download of the previous 12 months of Cyclistic trip data here. (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable me to answer the business questions. The data has been made available by Motivate International Inc. under this license.) This is public data that can be used to explore how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit using riders' personally identifiable information. This means that I won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

Company Research

This project will be using Divvy as the source of data and information so I will collect and state any noteworthy information found on the website that may affect the data here.

Current Pricing Structure:

Single ride:
- Classic bike: $1 unlock+ $0.18/min
- Electric bike: $1 unlock+ $0.44/min

Day Pass: $18.10/day
- Classic bike: 3hrs free, then $0.18/min
- Electric bike: free unlocks + $0.44/min

Divvy: $143.90/yr
- Classic bike: 45 mins free, then $0.18/min
- Electric bike: free unlocks + $0.18/min
- Bike Angels**

Lyft Pink: $199/yr
- Classic bike: 3hrs free, then $0.18/min
- Electric bike: free unlocks + $0.18/min
- Bike Angels
- Rideshare benefits**

Divvy for Everyone (D4E):

Together, Divvy, The Chicago Department of Transportation (CDOT) and the City of Evanston are expanding access to bikeshare through $5 annual memberships for qualifying residents. No annual commitment required.

- Classic bikes: Unlimited free unlocks, $0.05/minute after 45 minutes
- Ebikes: Unlimited free unlocks, $0.05/minute

Bike for Business:

Divvy partners with businesses to create plans that give employees an affordable, fun, and healthy way to commute.

**information on Bike Angels (rewards program to incentivize even distribution of bikes across stations).

**Rideshare benefits in the lyft pink plan include discount on lyft and grubhub full info here.

**Pricing structure also includes scooters, the data does not include the scooters so they will be excluded from the project.

If you're on an ebike you can also park at one of over 600 public racks or posts for free. Check the app to find racks nearby. Parking an ebike or scooter at any other legal public location is an extra $2.40 for non-members and $1.20 for Divvy members.

(*Ebikes do not have to be at specific stations for either start or end of trips but classic bikes need to be returned to stations*)

On the website it also states that bikes kept over 24 hours will may charge the rider $250 as a lost or stolen bike fee

The Data

Each trip is anonymized and includes:

- Trip start day and time
- Trip end day and time
- Trip start station
- Trip end station
- Rider type (Member, Single Ride, and Day Pass)

The data has been processed to remove trips that are taken by staff as they service and inspect the system; and any trips that were below 60 seconds in length (*potentially false starts or users trying to re-dock a bike to ensure it was secure*).

Data constraints

- Trips must be longer than 60 seconds or 1 minute, and less than 24hrs or one day. Any data outside of those constraints will be removed
- Single ride and day pass will be considered casual because I am trying to convert them into annual members.
- Customer information is removed for privacy.

- Data used will be historical data ranging from September 2023 to August 2024 and is organized in CSV format separated by month. The 12 CSV files will be stored on my local drive for this project.
- Electric bikes may not have a start or end location.
- Data is reliable, current at the time of this project and updated every month, free of bias, anonymized, and owned by the city of Chicago

For reliability and accuracy, I removed incomplete and irregular data points. Here's why this step was important:

- **Incomplete Records**: Some entries lack critical details, such as the start or end locations for classic bikes, which are necessary because these bikes must dock at specific stations. Keeping such records could have introduced errors, so they were excluded to maintain the integrity of the data.
- **Irregular Data**: Trips that lasted under a minute (*likely due to docking issues, maintenance or errors*) or over 24 hours (*possibly involving lost or stolen bikes*) were outliers that didn't represent normal usage. Including them could have distorted key metrics, such as average trip duration.

Once only accurate and relevant data remains, I am able to base the analysis on a trustworthy dataset, leading to actionable and meaningful insights.

Data Breakdown

Data used for analysis spans 12 months from September 2023 to August 2024 and contains data on all riders using the service separated by month. Columns included in data are as follows:

- Ride_id: unique string for each trip, and is the primary key contains 16 characters
- Ridable_type:type of bike used for trip, all values are either electric_bike or classic_bike
- Started_at: date/time trip started at
- Ended_at: date/time the trip ended
- Start_station_name: name of the station that the bike was taken from at the start of a trip.
- Start_station_id: unique string for specific docking station used to take the bike away at the start of a trip.
- End_station_name: name of the station that the bike was docked away at the end of a trip.
- end_station_id: unique string for a specific docking station used to dock the bike away after the end of a trip.
- start_lat: the start of a trip's latitude.
- start_lng: the start of a trip's longitude.
- end_lat: the end of a trip's longitude.

- end_lng: the end of a trip's latitude.
- member_casual: the only values are member or casual and indicate membership status for the trip.

Excel Exploration/Cleaning

Converted CSV files into XLSX do begin data exploration and cleaning in excel:

Created new columns:

- ride_length by subtracting the started_at column from the ended_at column and formatted into time in HH:MM:SS notation
- day_of_week by using the weekday function with 1-7(sunday-saturday) format using the started_at column, formatted as general

Deleted columns/rows(done in order) :

- start_station_id/end_station_id: these columns have various inconsistencies in length and some have strings while others are only number and wouldnt provide any extra utility to the project
- Using excel remove duplicate function on ride_id, none found
- Using filters removing any rows without start location or end location for any trips used with classic bikes because classic bikes need to be returned/picked up at a station. Rows removed:7,990
- Using ride_length column to remove any trips that return a negative value because negative time trips aren't possible. Rows removed: 288
- Using the ride_length column to remove rows that have trips under 1 minute and over 24 hrs (under one minute is for maintenance or false starts and over 24 hrs is likely stolen). Rows removed:129,539
- deleted any ride_id's that are outside of the 16 characters, used LEN() function to check for IDs over or under 16 characters. Rows removed:724

Reformated:

- start_lat/end_lat & start_lng/end_lng: some values came up as strings so I created a new column using the VALUE() function to transform the cells into value and pasted the values back in.
- From June-August 2024 the columns for started_at and ended_at have different formatting changed it to match the other tables which is yyyy-mm-dd hh:mm
- Some ID's were only numerical and came up as exponent values so I made sure they were formatted to show all 16 characters

Imported Into SQL

Once imported into SQL I combined all the cleaned tables into one file table named combined_tripdata. I then removed any rows which had a NULL ride ID to make sure no extra/blank rows were brought in from excel other than the data. That left me with 5,561,052 rows/trips.

SQL Query

```
-- Step 1: Combine all monthly trip data tables into one table for analysis --
DROP TABLE IF EXISTS combined_tripdata;

-- The combined table will include data from January 2023 to August 2024
SELECT *
INTO combined_tripdata
FROM (
   SELECT * FROM [GoogleCapstoneA].[dbo].['202408-divvy-tripdata$'] -- August 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202407-divvy-tripdata$'] -- July 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202406-divvy-tripdata$'] -- June 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202405-divvy-tripdata$'] -- May 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202404-divvy-tripdata$'] -- April 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202403-divvy-tripdata$'] -- March 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202402-divvy-tripdata$'] -- February 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202401-divvy-tripdata$'] -- January 2024 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202312-divvy-tripdata$'] -- December 2023 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202311-divvy-tripdata$'] -- November 2023 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202310-divvy-tripdata$'] -- October 2023 data
   UNION
   SELECT * FROM [GoogleCapstoneA].[dbo].['202309-divvy-tripdata$'] -- September 2023 data
) combined_tripdata;

-- Step 2: Remove incomplete records where 'ride_id' is NULL --
```

```
DELETE FROM combined_tripdata WHERE ride_id IS NULL;

-- Step 3: Separate the data by membership type --
-- Members: Users with subscriptions (expected: only rows with 'member_casual' = 'member').
SELECT * FROM combined_tripdata WHERE member_casual = 'member'; -- Query for
members

-- Casual: Users without subscriptions (expected: only rows with 'member_casual' = 'casual').
SELECT * FROM combined_tripdata WHERE member_casual = 'casual'; -- Query for casual
riders

-- Step 4: Identify the longest rides for members and casual riders --
-- For members, create a table to store the ride ID, max ride length, day of the week, and user
type.
DROP TABLE IF EXISTS ride_length_member;
SELECT *
INTO ride_length_member
FROM (
   SELECT
      ride_id,
      MAX(CAST(ride_length AS TIME)) AS ride_length, -- Convert to TIME format for
readability
      day_of_week, -- Day when the ride occurred
      member_casual
   FROM combined_tripdata
   WHERE member_casual = 'member'
   GROUP BY ride_length, member_casual, day_of_week, ride_id
) ride_length_member;

-- View the longest rides for members, sorted in descending order of ride length.
SELECT * FROM ride_length_member ORDER BY ride_length DESC;

-- Repeat the same analysis for casual riders.
DROP TABLE IF EXISTS ride_length_casual;
SELECT *
INTO ride_length_casual
FROM (
   SELECT
      ride_id,
      MAX(CAST(ride_length AS TIME)) AS ride_length,
      day_of_week,
      member_casual
   FROM combined_tripdata
   WHERE member_casual = 'casual'
```

```
    GROUP BY ride_length, member_casual, day_of_week, ride_id
) ride_length_casual;

-- View the longest rides for casual riders, sorted in descending order of ride length.
SELECT * FROM ride_length_casual ORDER BY ride_length DESC;

-- Step 5: Calculate the average ride length for members and casual riders --
-- Average ride time for members (result: a single value representing the average).
DROP TABLE IF EXISTS average_ride_member;
SELECT
    CAST(AVG(CAST(ride_length AS FLOAT)) AS TIME) AS avg_ride_member
INTO average_ride_member
FROM combined_tripdata WHERE member_casual = 'member';

-- Average ride time for casual riders (result: a single value representing the average).
DROP TABLE IF EXISTS average_ride_casual;
SELECT
    CAST(AVG(CAST(ride_length AS FLOAT)) AS TIME) AS avg_ride_casual
INTO average_ride_casual
FROM combined_tripdata WHERE member_casual = 'casual';

-- Step 6: Determine the most popular days for rides by type --
-- Total ride counts by day of the week for all riders.
DROP TABLE IF EXISTS rides_by_day;
SELECT
    day_of_week,
    COUNT(day_of_week) AS rider_day_of_week -- Count of rides for each day
INTO rides_by_day
FROM combined_tripdata
GROUP BY day_of_week;

-- Similar breakdowns for members and casual riders.
DROP TABLE IF EXISTS rides_by_day_member, rides_by_day_casual;
SELECT
    day_of_week,
    COUNT(day_of_week) AS count_of_days
INTO rides_by_day_member
FROM combined_tripdata
WHERE member_casual = 'member'
GROUP BY day_of_week;

SELECT
    day_of_week,
    COUNT(day_of_week) AS count_of_days
```

```
INTO rides_by_day_casual
FROM combined_tripdata
WHERE member_casual = 'casual'
GROUP BY day_of_week;

-- Step 7: Analyze preferences for bike types (classic vs electric) --
DROP TABLE IF EXISTS classic_bike_member, electric_bike_member, classic_bike_casual,
electric_bike_casual;

-- Classic bikes for members.
SELECT
   COUNT(ride_id) AS classic_bike_member
INTO classic_bike_member
FROM combined_tripdata
WHERE member_casual = 'member' AND rideable_type = 'classic_bike';

-- Electric bikes for members.
SELECT
   COUNT(ride_id) AS electric_bike_member
INTO electric_bike_member
FROM combined_tripdata
WHERE member_casual = 'member' AND rideable_type = 'electric_bike';

-- Classic bikes for casual riders.
SELECT
   COUNT(ride_id) AS classic_bike_casual
INTO classic_bike_casual
FROM combined_tripdata
WHERE member_casual = 'casual' AND rideable_type = 'classic_bike';

-- Electric bikes for casual riders.
SELECT
   COUNT(ride_id) AS electric_bike_casual
INTO electric_bike_casual
FROM combined_tripdata
WHERE member_casual = 'casual' AND rideable_type = 'electric_bike';

-- Step 8: Examine average ride lengths and starting hours by day of the week --
DROP TABLE IF EXISTS avg_by_day_member, avg_by_day_casual, start_hour_member,
start_hour_casual;

-- Average ride lengths for members by day of the week.
SELECT
   CAST(AVG(CAST(ride_length AS FLOAT)) AS TIME) AS avg_ride_length,
```

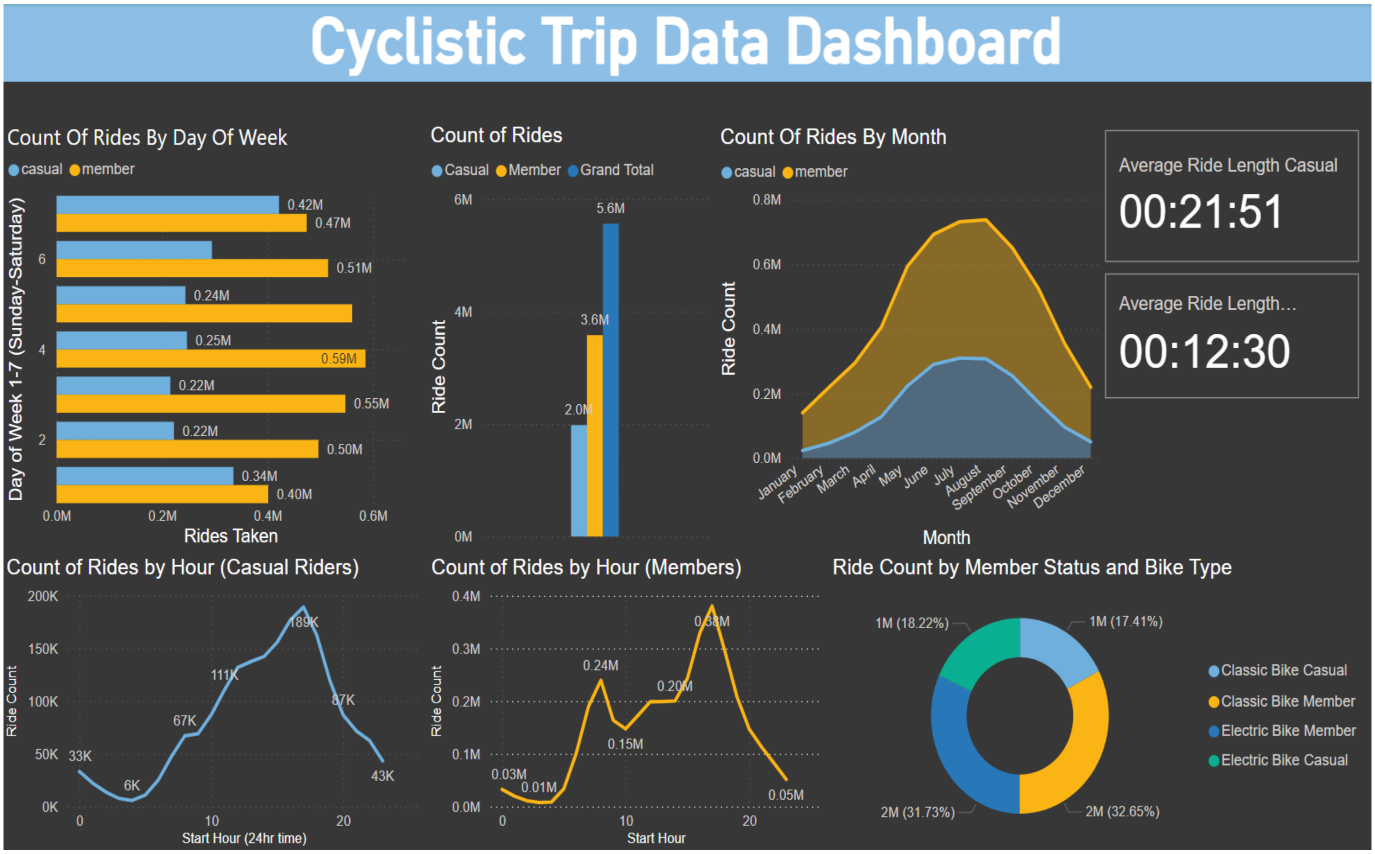
```
    day_of_week
INTO avg_by_day_member
FROM combined_tripdata
WHERE member_casual = 'member'
GROUP BY day_of_week;

-- Average ride lengths for casual riders by day of the week.
SELECT
    CAST(AVG(CAST(ride_length AS FLOAT)) AS TIME) AS avg_ride_length,
    day_of_week
INTO avg_by_day_casual
FROM combined_tripdata
WHERE member_casual = 'casual'
GROUP BY day_of_week;

-- Extract ride start hours for members.
SELECT
    ride_id,
    member_casual,
    DATEPART(HOUR, CAST(started_at AS TIME)) AS start_hour
INTO start_hour_member
FROM combined_tripdata
WHERE member_casual = 'member';

-- Extract ride start hours for casual riders.
SELECT
    ride_id,
    member_casual,
    DATEPART(HOUR, CAST(started_at AS TIME)) AS start_hour
INTO start_hour_casual
FROM combined_tripdata
WHERE member_casual = 'casual';
"""
```

Dashboard Created In Power BI



Insights

Analysis of Cyclistic Rider Behavior

The data highlights distinct differences between casual riders and annual members, providing valuable insights into usage patterns and potential strategies for converting casual riders into members.

Key Insights:

1. **Rider Behavior by Day of Week:**
   - Casual riders predominantly use the service on weekends, with peak activity on Saturdays and Sundays.
   - Annual members show consistent usage across all days, with a slight increase during weekdays, likely tied to commuting patterns.
   - **Implication:** Casual riders treat biking as a recreational activity, while members rely on it for daily transportation.

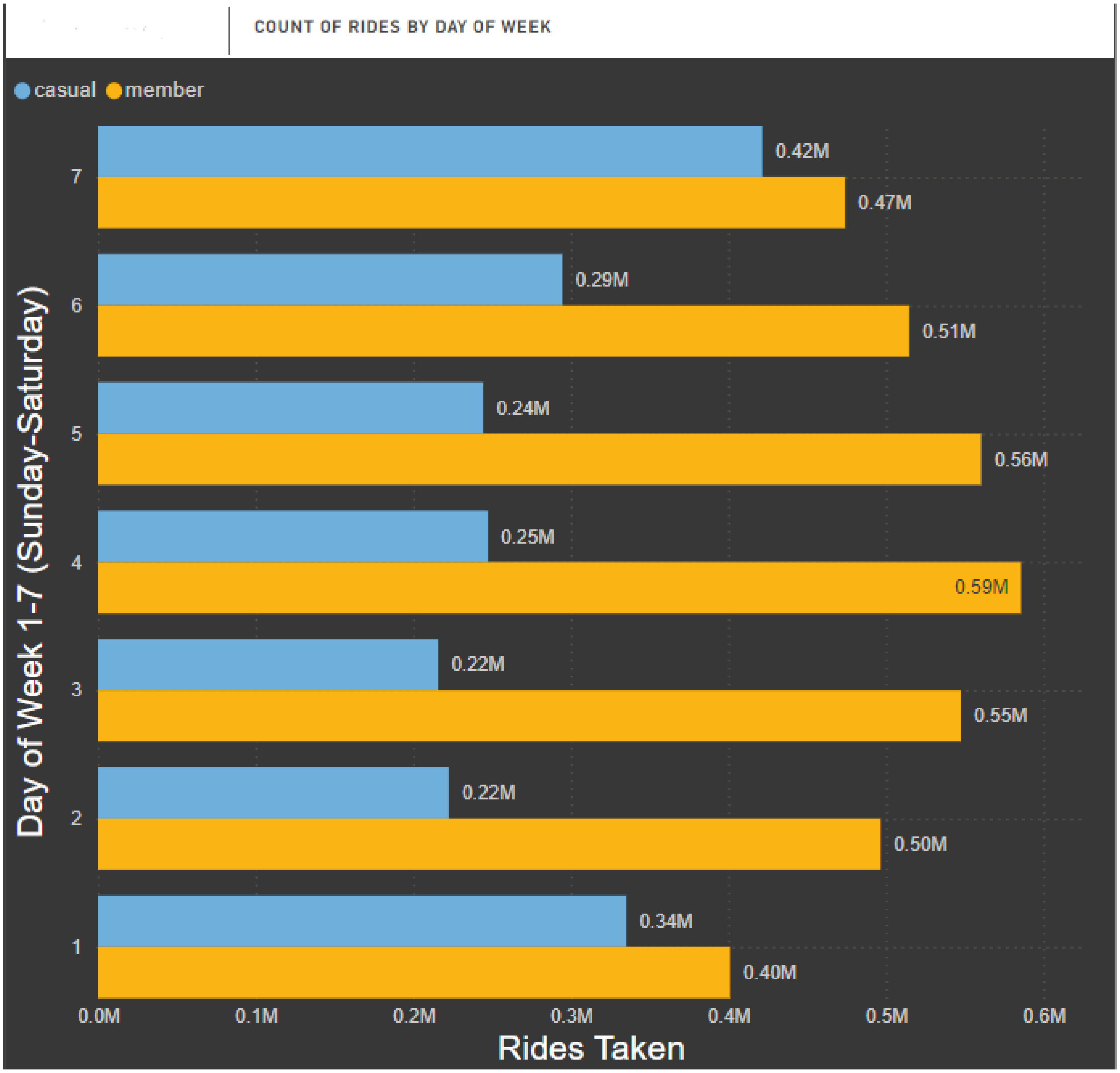


Figure 1: Count of rides by day of the week for casual riders and members

2. **Ride Length:**
    - Casual riders average 21 minutes per trip—nearly double the 12-minute average of members
    - **Implication:** Casual riders prefer longer, leisure-oriented trips, while members focus on shorter, utilitarian rides.
3. **Usage by Hour:**
    - Casual riders are most active between **11 AM and 2 PM**, aligning with leisure hours.
    - Members have distinct peaks during **8–9 AM** and **5–6 PM**, corresponding to commuting times.

- **Implication:** Marketing strategies should address these different usage windows to better engage each group.

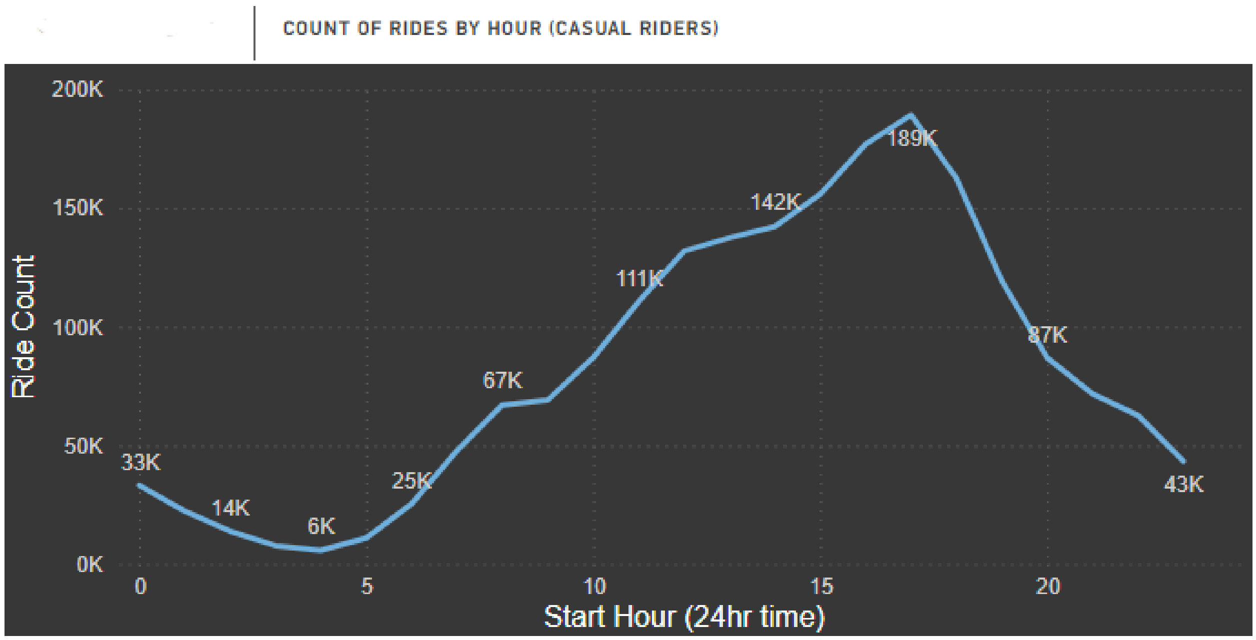


Figure 3: Count of rides by hour for casual riders

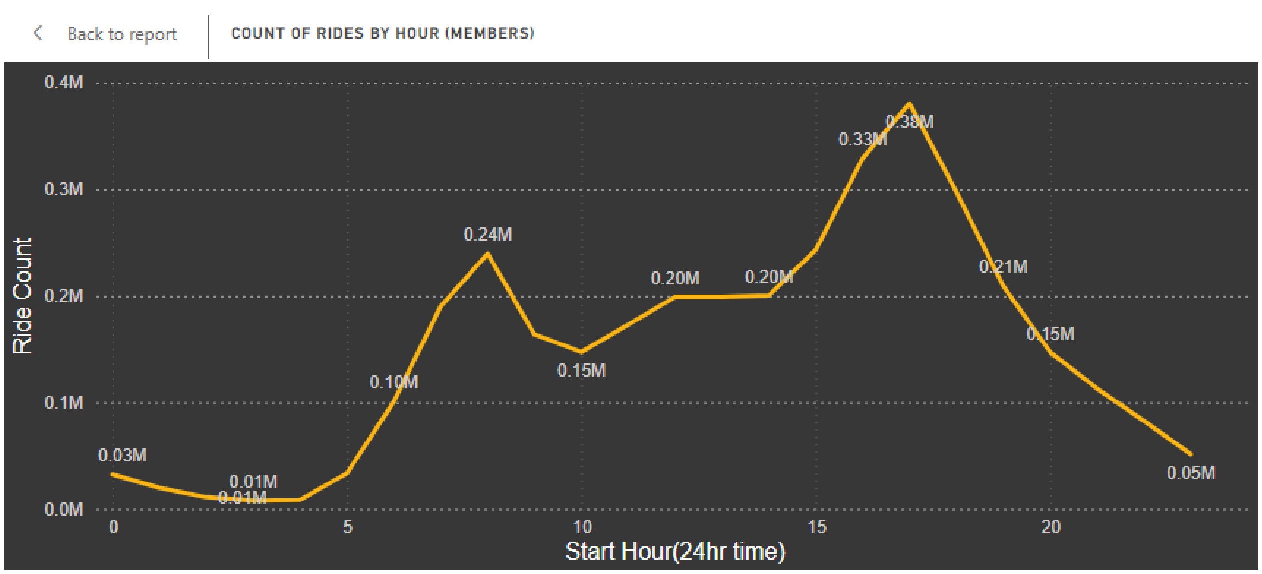


Figure 4: Count of rides by hour for members

4. **Bike Type Preferences:**
    - **Classic Bikes Dominate**: Combined, classic bikes account for nearly **64.38%** of rides, highlighting their overall popularity across all rider types.

- **Electric Bike Adoption**: While electric bikes are used less frequently overall (**35.63%**), their adoption is relatively balanced between casual and member riders.
- **Implication:** while classic bikes are more popular among members, promoting electric bikes may have higher member signup rates

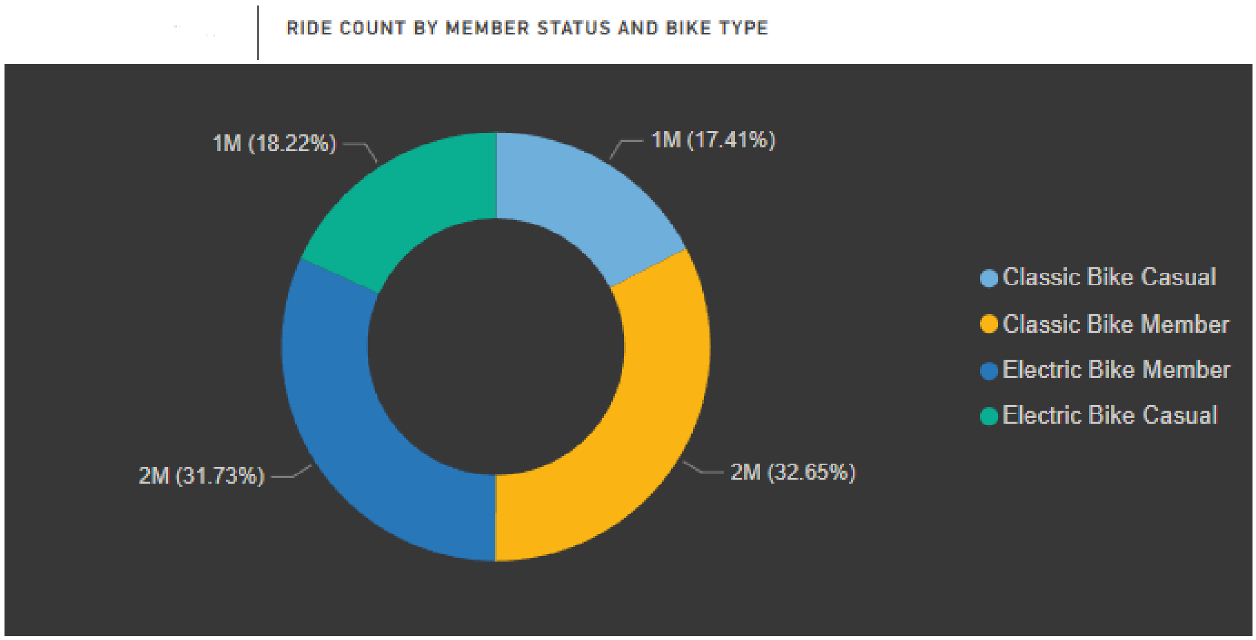


Figure 4: Usage of electric vs. classic bikes by casual riders and members.

5. **Seasonality:**
    - Ridership for both groups peaks during summer (June–August) and declines in winter.
    - **Implication:** Seasonal strategies are essential to maintain ridership during off-peak months.

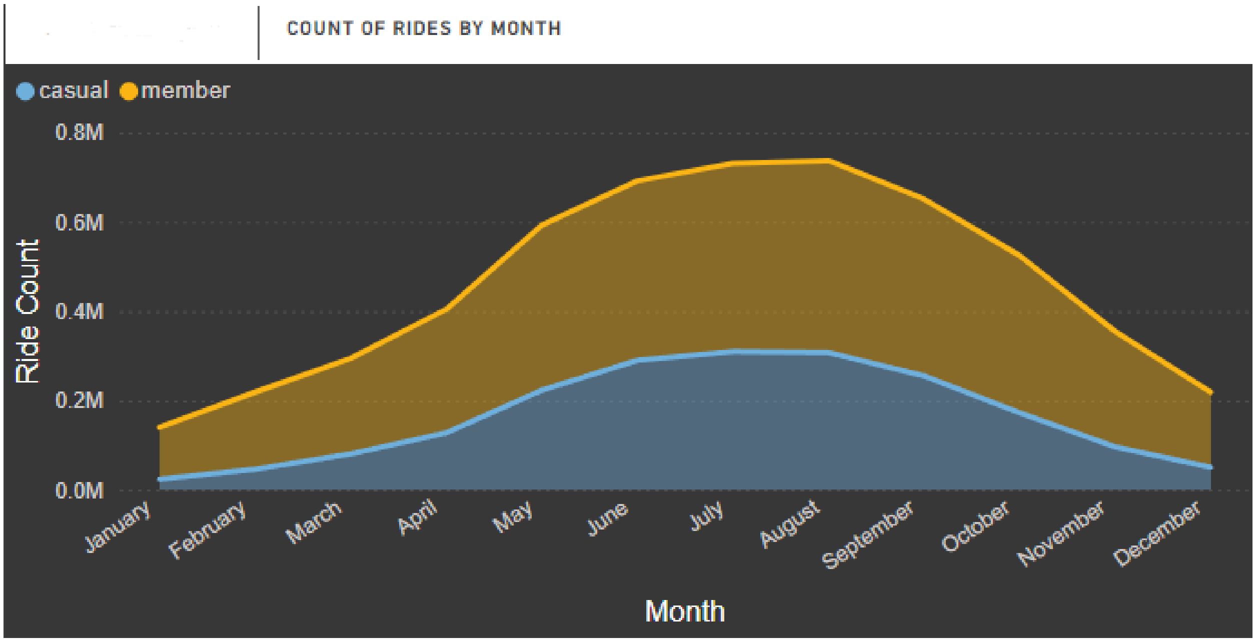


Figure 6: Monthly ride counts for casual riders and members.

Final Recommendations & Prioritization

To drive Cyclistic's growth by converting casual riders into long-term members, the following three strategies are proposed. These strategies are grounded in insights drawn from rider behavior and usage patterns, and are designed to be effective at increasing membership conversion.

1. Flexible "Pay-as-You-Go" Membership Plan

- **Rationale**:
  - **Cost Sensitivity**: A flexible, low-cost monthly plan provides a lower barrier to entry while encouraging frequent usage, making it more accessible to a larger segment of casual riders.
  - **Behavioral Alignment**: Casual riders often make infrequent trips but tend to prefer longer rides. By offering lower per-minute rates and free unlocks, this plan directly addresses their usage patterns, making it more cost-effective and attractive.
  - **Long-Term Scalability**: As casual riders increase their usage, they may recognize the value of an annual membership, providing a natural progression from pay-as-you-go to full membership.
- **Prioritization**:
  - **Impact**: High – Directly addresses cost barriers for casual riders, and aligns with their usage behaviors.

    - **Feasibility**: High – integrates with the current pricing structure with minimal technical changes.
    - **Cost**: Medium –initial revenue per trip may decrease, the long-term potential for membership upgrades may offset this.
    - **Priority**: **#1** – This presents the most scalable and immediate opportunity to increase membership conversions.

2. Membership Trial Plan ("Test the Ride")

- **Rationale**:
    - **Low-Risk Commitment**: Offering casual riders a risk-free trial period allows them to experience the benefits of membership, such as free rides and discounted electric bike rates, without the upfront commitment of an annual plan.
    - **Conversion Focus**: Providing a credit toward full membership after the trial period incentivizes conversion, while demonstrating the tangible benefits of being a member.
    - **Trust and Engagement**: The trial period builds trust and engagement, giving riders the opportunity to experience membership first-hand, addressing potential skepticism regarding its value.
- **Prioritization**:
    - **Impact**: High – This strategy lowers the risk for casual riders and effectively introduces them to membership, increasing the likelihood of conversion.
    - **Feasibility**: High – Requires minimal adjustments to existing systems and promotional campaigns, making it easy to implement.
    - **Cost**: Low – Trials can be marketed with minimal upfront costs, relying on the existing user base and promotional channels.
    - **Priority**: **#2** – A highly effective, low-cost strategy with potential for driving membership conversions.

3. Weekend Pass with Electric Bike Perks

- **Rationale**:
    - **Targets Peak Usage**: Casual riders tend to use the service more on weekends, particularly for leisure trips. A weekend-specific membership plan aligns with this pattern.
    - **Electric Bike Incentives**: Many casual riders prefer longer trips, and electric bikes are particularly suited to this behavior. Offering perks like free unlocks or discounted rates for electric bikes highlights the value of the weekend pass.
    - **Increased Usage**: By encouraging repeat purchases through bundled promotions (e.g., "Earn a free weekend pass after three purchases"), this strategy helps increase engagement and incentivizes long-term membership conversion.
- **Prioritization**:
    - **Impact**: Medium – This strategy targets casual riders' peak usage times but may not have as broad an appeal as the other strategies.

    - **Feasibility**: Medium – this plan requires some updates to pricing and promotional materials, it can be implemented using existing infrastructure and data.
    - **Cost**: Medium – Offering electric bike perks may reduce short-term revenue, but increased ride volume and membership conversions will likely offset the cost.
    - **Priority**: **#3** – A focused, data-driven approach that maximizes existing weekend ridership trends, but with more limited potential for broad application.

Why These Recommendations Stand Out

- **Broad Appeal**: These strategies directly address key financial and behavioral factors that influence casual riders' decision-making, as revealed through data analysis of rider habits.
- **Scalability & Profitability**: Each recommendation is designed to scale effectively over time, with a clear path to increased ridership and membership conversion, all while ensuring Cyclistic remains profitable.
- **Ease of Implementation**: These recommendations integrate seamlessly with Cyclistic's existing systems, minimizing disruption and leveraging existing data and promotional frameworks.

Conclusion

These three recommendations offer a robust, multi-pronged approach to converting casual riders into long-term members. By focusing on addressing cost sensitivity, providing low-risk entry points, and leveraging peak usage periods, Cyclistic can increase membership conversions and grow sustainably in the long term.