Maximising Annual Memberships- A Cyclistic Bike-Share Case Study

A Google Data Analytics Capstone Project

Presenting to you by:

Sahaja Reddy Sathi

sahajareddy162@gmail.com

April 2025- May 2025

Introduction:

The Cyclistic Bike-Share Case Study is a capstone project completed as part of the **Google Data Analytics Professional Certificate**. The objective of this project is to analyze 12 months of bike-share data to uncover user behavior trends and provide actionable insights to help **Cyclistic convert casual riders into annual members**.

By following the six-step data analysis process — **Ask, Prepare, Process, Analyze, Share, and Act** — I performed data cleaning in Excel, conducted in-depth SQL analysis, and visualized key findings in Power BI. The final recommendations are aimed at supporting Cyclistic's marketing strategy through data-driven decision-making.

About the Company:

Cyclistic is a Chicago-based bike-share program with over **5,800 bikes** and **600+ docking stations**, offering inclusive options like reclining bikes and hand tricycles for riders with disabilities. While most users ride for leisure, around **30%** commute to work daily. Since launching in 2016, Cyclistic has expanded rapidly, supporting flexible usage with **single-ride, full-day, and annual membership plans**.

The company has found that **annual members are more profitable** than casual riders. To drive sustainable growth, Cyclistic aims to **convert casual riders into annual members** by analyzing historical usage data and designing targeted marketing strategies.

➢ Business Task:

Cyclistic's marketing team wants to **increase the number of annual memberships**.

To achieve this, we need to **analyze how casual riders and annual members use Cyclistic bikes differently**.

Insights from this analysis will help design effective marketing strategies aimed at **converting casual riders into annual members**.

➢ Key Stakeholders:

Cyclistic Marketing Team — will use insights to plan marketing campaigns.

Cyclistic Executive Team — will approve budgets and strategic actions.

Cyclistic Operations/Product Team (optional) — may adjust bike stations or service based on findings.

➢ Key Business Questions:

- How do annual members and casual riders differ in their bike usage patterns?
- When do casual riders ride the most (days, times, seasons)?
- What types of bikes are most used by members vs casual riders?
- How long are the rides on average for members vs casual riders?

- Data Source: Divvy Bike Share (operated by Lyft), Chicago Open Data Portal

Link: https://divvy-tripdata.s3.amazonaws.com/index.html

Time Frame: 12 months– January 2024 to December 2024

File Format: Monthly CSVs (converted to Excel for cleaning); each file contains 13 columns

Key Fields: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, member_casual.

Data Credibility (ROCCC Check):

- Reliable: Official Divvy/Lyft trip data
- Original: Based on real user trips
- Comprehensive: Covers ride duration, time, station, and user type
- Current: Full-year dataset (2024)
- Cited: Publicly cited and hosted under open data license

Licensing, privacy & Access:

- License: Open for non-commercial/educational use
- Privacy: No personal info included (Only trip metadata)
- Security: Files handled locally with no online sharing
- Accessibility: Cleaned & stored in standardized (.csv) for analysis.

Key Tasks Completed:

✓ Downloaded all 12 months of 2024 data from the official site.

✓ Assessed credibility using ROCCC framework

Tools used:

Excel: For cleaning and formatting monthly files

Power shell: Merged all 12 files (~4.2M rows)

Data cleaning & Transformation:

Removed blanks using filter option and handled duplicates using " Remove duplicates" function in Excel.

Reformatted key columns to consistent data types which affected columns- ride id, station names and IDs, member_casual.

Created ride_length= ended_at - started_at

Verified timestamp consistency (started_at, ended_at, ride_length)

Added day_of_week using Excel's "weekday()" function and formatted to numeric.

Merging CSV files using Power Shell:

Placed all 12 monthly CSV files in a single folder (e.g: Cyclistic Cleaned Data)

Ensured all files have the same columns

Right-click inside the folder -> "Open Power Shell window"

Paste and Run the below Power Shell Command:

```
Get-Content (Get-ChildItem -Filter *.csv | Select-Object -First 1) | Set-Content combined_2024_data.csv
Get-ChildItem -Filter *.csv | Select-Object -Skip 1 | ForEach-Object { Get-Content $_ | Select-Object -Skip 1 } >>
combined_2024_data.csv
```

A single merged file: "combined_2024_data.csv" with no duplicate headers is ready for analysis

Importing dataset into My SQL using command prompt:

Followed the below steps to import the dataset into My SQL:

Press “windows + S”, type cmd

Right-click command prompt -> Select run as administrator

Paste and run the below command to import the dataset into My SQL:

```
mysql -u root -p --local-infile=1 -e "LOAD DATA LOCAL INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/combined_2024_data.csv' INTO TABLE cyclistic.bike_trips FIELDS TERMINATED BY ',' ENCLOSED BY '\"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;"
```

Data Integrity & Documentation:

Original files preserved and edits done on duplicates

Verified all 15 columns before merging

Imported dataset into MySQL using command prompt

Documented all steps and issues for future audits/interviews.

Key completed Tasks:

- ✓ Cleaned missing and duplicate data
- ✓ Standardized formats
- ✓ Created new calculated fields
- ✓ Merged datasets
- ✓ Imported dataset.
- ✓ Documented full cleaning and transformation process.

Data Organization & Formatting:

Data from 12 months (2024) was combined into a single dataset using the command prompt.

Imported into **MySQL** after cleaning, formatting, and calculating key metrics (ride_length, day_of_week) in Excel.

Structured table created with clearly defined fields to support SQL-based analysis.

Data Cleaning & Preparation:

Null values in essential columns were checked and excluded.

Records with **negative ride durations** (due to invalid timestamps) were removed to ensure data integrity.

Time fields were properly formatted; ride_length was analyzed in **minutes** using SQL.

Key Discoveries & Insights:

Casual riders have a **higher average ride duration** of 38% than members, suggesting leisure-based usage.

Members account for the **highest number of total rides**, especially on weekdays, indicating routine or commuting behavior.

Sundays show the **highest ride count** across both user types, but more so among members.

Ride patterns indicate **different usage behaviors**: members are frequent users; casuals ride less often but for longer durations.

Trends and Relationships:

User Type vs. Ride Duration: Casual > Member (in ride length)

User Type vs. Frequency: Member > Casual (in ride count)

Day of Week Impact: Weekends drive high usage for casuals; weekdays are dominant for members.

These trends help tailor marketing strategies and operational decisions.

Conclusion:

The analysis reveals that **casual riders enjoy longer, likely leisure-focused trips**, while **members ride more frequently**, possibly for daily commutes. These behavioral patterns align with Cyclistic's business goal: **converting casual riders into loyal members** to increase revenue and engagement. Strategic offers and promotions based on ride timing and patterns can drive this transition effectively.

Total Number of Rows:

```
SELECT COUNT(*) AS cleaned_rows FROM bike_trips;
```

Output:

	total_rows
▶	3543675

Insight: After data cleaning, the dataset contains **~3.5 million** valid ride records.

Average Ride Length by Rider Type:

```
SELECT
  member_casual,
  ROUND(AVG(TIME_TO_SEC(ride_length)/60), 2) AS avg_ride_length_minutes
FROM bike_trips
GROUP BY member_casual;
```

Output:

	member_casual	avg_ride_length_minutes
▶	member	10.71
	casual	15.95

Insight: Casual riders have longer ride durations than annual members.

Indicates members may use bikes for quick commutes, while casual riders use them for leisure

Rides by Day of Week:

```
SELECT
    day_of_week,
    member_casual,
    COUNT(*) AS total_rides
FROM bike_trips
GROUP BY day_of_week, member_casual
ORDER BY total_rides DESC;
```

Output:

day_of_week	member_casual	total_rides
7	member	1558633
7	casual	906202
4	member	151211
3	member	143712
5	member	142199
2	member	132486
6	member	121754
1	member	90517
1	casual	63624
6	casual	52582
2	casual	48170
4	casual	46260
3	casual	43254
5	casual	43071

Insights:

Members ride consistently throughout the week, with the highest usage on Sundays.

Casual riders mostly ride on weekends, especially Sundays, and less during Weekdays.

This suggests members use bikes for commuting, while casuals use them for leisure.

Rideable type by Rider Type:

```
SELECT
    rideable_type,
    member_casual,
    COUNT(*) AS count
FROM bike_trips
GROUP BY rideable_type, member_casual;
```

Output:

rideable_type	member_casual	total_rides
classic_bike	member	1522513
electric_bike	member	798811
classic_bike	casual	744729
electric_bike	casual	437410
electric_scooter	casual	21024
electric_scooter	member	19188

Insights:

Classic bikes are preferred by members.

Casual riders often use docked or electric bikes, which may be easier or more accessible for short-term use.

CODE:

```
-- 1. Create the database
CREATE DATABASE IF NOT EXISTS cyclistic;
USE cyclistic;
-- 2. Create the table with data types
CREATE TABLE IF NOT EXISTS bike_trips (
    ride_id VARCHAR(255),
    rideable_type VARCHAR(50),
    started_at TIME,
    ended_at TIME,
    start_station_name VARCHAR(255),
    start_station_id VARCHAR(255),
    end_station_name VARCHAR(255),
    end_station_id VARCHAR(255),
    start_lat FLOAT,
    start_lng FLOAT,
    end_lat FLOAT,
    end_lng FLOAT,
    member_casual VARCHAR(50),
    ride_length TIME,
    day_of_week VARCHAR(10));
```

```
-- 3. (Assumes data is already imported)
-- 4. Total row count
SELECT COUNT(*) AS total_rows FROM
bike_trips;
--5. Check for NULLs in key columns
SELECT * FROM bike_trips
WHERE ride_id IS NULL
   OR rideable_type IS NULL
   OR started_at IS NULL
   OR ended_at IS NULL
   OR member_casual IS NULL;
-- 6. Disable safe update modeSET
SQL_SAFE_UPDATES = 0;
-- 7. Delete rows with negative
ride_length (data quality check)
DELETE FROM bike_trips
WHERE TIME_TO_SEC(ride_length) < 0;
-- 8. Re-enable safe update mode
SET SQL_SAFE_UPDATES = 1;
```

```
-- 9. Verify cleaned data
SELECT COUNT(*) AS cleaned_rows FROM
bike_trips;
-- 10. Avg ride length (in minutes if
formatted correctly) by user type
SELECT member_casual,
ROUND(AVG(TIME_TO_SEC(ride_length)/60), 2)
AS avg_ride_length_minutes
FROM bike_trips
GROUP BY member_casual;
-- 11. Rides per day of the week by user
type
SELECT
    day_of_week,
    member_casual,
    COUNT(*) AS total_rides
FROM bike_trips
GROUP BY day_of_week, member_casual
ORDER BY total_rides DESC;
```

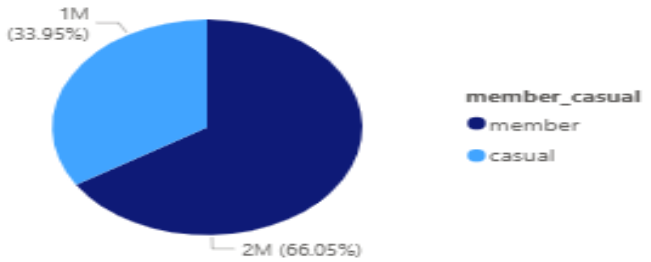
```
--12. Rideable type usage by user
type
SELECT
rideable_type,
member_casual,
COUNT(*) AS total_rides
FROM bike_trips
GROUP BY rideable_type,
member_casual;
```

Ask | Prepare | Process | Analyze | **Share** | Act

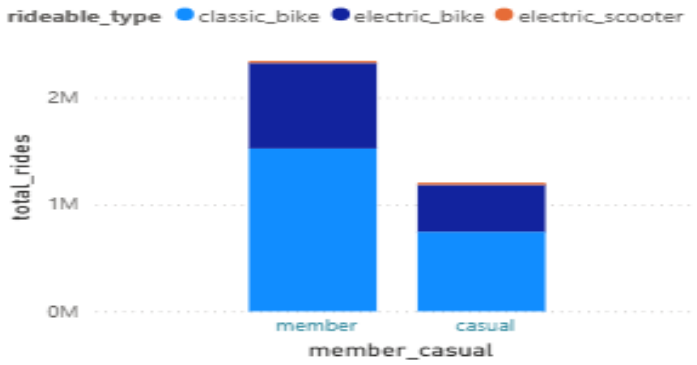
Summary of Rides

day_of_week	casual	member	Total
1	63624	90517	**154141**
2	48170	132486	**180656**
3	43254	143712	**186966**
4	46260	151211	**197471**
5	43071	142199	**185270**
6	52582	121754	**174336**
7	906202	1558633	**2464835**
Total	**1203163**	**2340512**	**3543675**

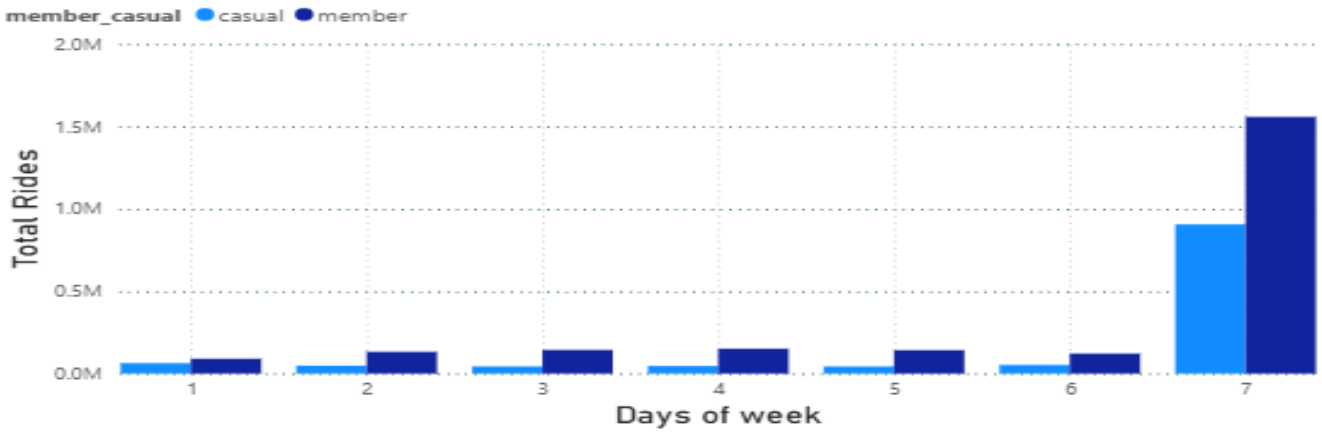
Total Rides by annual members and casual members



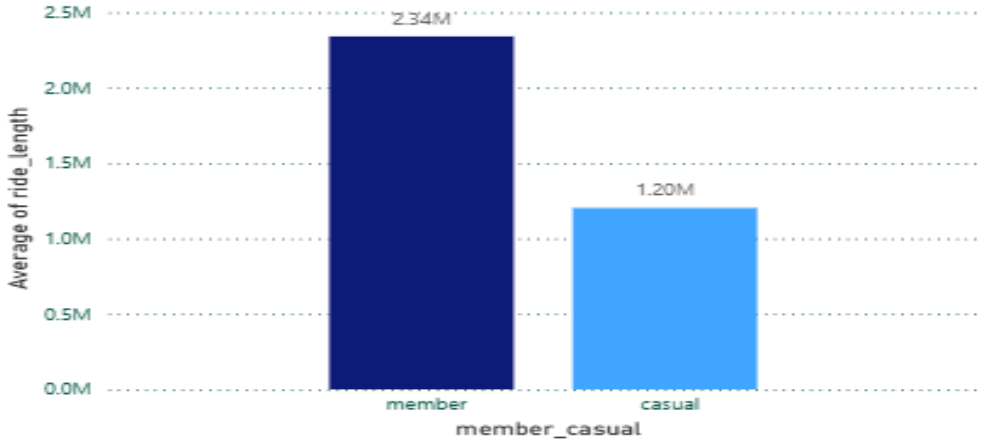
Rideable type by User



Most rides happen on weekends by casual users



Average ride duration by Rider Type



Insights:

Casual riders take **longer rides** on average than **member riders**.

Members ride more frequently on **weekdays**, while **casual riders** are more active on **weekends**.

The most popular ride types differ between casual and member users.

Despite **members being more in number**, casual riders generate **higher ride volume** — a potential missed revenue opportunity.

Steps followed to generate a Power BI Report:

Data Connection in Power BI:

Imported the MysQL data into Power BI using the correct server and database name

Loaded the bike_trips table

Created a new measure: total_rides= COUNT(bike_trips[ride_id]) in model view.

Visualization in Power BI:

Created 4 meaningful visuals:

Bar chart showing total rides by member_casual.

Bar chart comparing average ride length by user type.

Stacked column chart showing rides by day of week and user type.

Pie chart showing rideable usage by user type.

Customization:

Used tooltips, labels, and legends for better readability.

Applied color coding and formatting to differentiate member and casual users.

Conclusion:

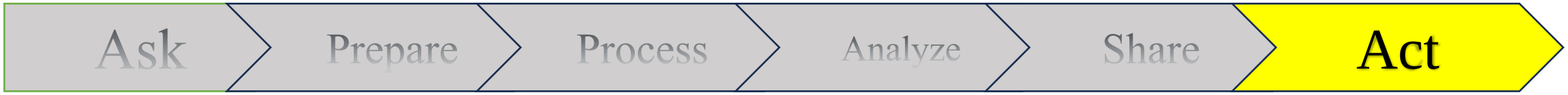
These insights show clear behavioral differences between casual and member riders. Cyclistic can use this data to:

Target casual users for conversion into members by offering weekday or commuter benefits.

Optimize bike availability based on usage trends by day and ride type.

Plan marketing strategies around weekdays for casual users.

.



Key Findings:

- **Casual riders take longer trips** on average than annual members.
- **Casual riders ride more on weekends**, especially Saturdays and Sundays.
- **Annual members ride more frequently**, but mainly on **weekdays**, indicating regular usage patterns (like commuting).
- **Rideable type preferences** show both user types often choose classic bikes, but casual riders slightly favor docked bikes more.
- The overall number of rides by **casual users is higher**, but **members are more consistent across the week**.

Recommendations:

- **Create targeted weekend promotions** for casual users to encourage them to subscribe as annual members.
- **Offer loyalty benefits or discounts** for frequent weekday rides to reward consistent member behavior.
- **Improve marketing campaigns** around casual rider behaviors (e.g., ads promoting membership during peak casual usage times).
- **Educate casual riders on member benefits** like cost savings and flexibility to drive conversion.
- **Analyze station usage data further** to optimize bike/station placement according to demand patterns.

Key Learnings

1. End-to-End Data Analysis Workflow: Learned to execute a complete data analysis lifecycle using the ASK → PREPARE → PROCESS → ANALYZE → SHARE → ACT framework.
2. Real-World Data Cleaning: Faced challenges like missing values, incorrect timestamps, and negative durations—learned how to clean and prepare large datasets efficiently using Excel and SQL.
3. SQL Proficiency: Strengthened skills in writing queries for filtering, aggregating, and extracting insights—especially grouping, date-based filtering, and joining logic.
4. Handling Large Datasets: Understood how to combine and manage large files (887 MB+), and optimized import methods via command prompt and MySQL Workbench.
5. Insight Generation:

Discovered patterns such as:

- Casual riders take longer rides than members.
- Member riders prefer weekdays, while casual riders are more active on weekends .

6. Power BI Visualization Skills:

Created meaningful, interactive dashboards with:

- Bar charts
- Average ride duration visuals
- Comparison visuals for rider types
- Custom color coding and tooltips

7. Storytelling with Data: Learned to interpret data visually and communicate findings effectively for business decision-making.
8. Business Impact Focus:

Aligned analysis with the company's goal: converting casual riders into annual members by understanding user behavior.

9. Documentation & Presentation: Developed the skill to document analysis clearly, build a professional PowerPoint, and prepare for interviews by being ready to explain each decision.

8 Courses

Foundations: Data, Data, Everywhere

Ask Questions to Make Data-Driven Decisions

Prepare Data for Exploration

Process Data from Dirty to Clean

Analyze Data to Answer Questions

Share Data Through the Art of Visualization

Data Analysis with R Programming

Google Data Analytics Capstone: Complete a Case Study

Apr 9, 2025

Sahaja Reddy Sathi

has successfully completed the online, non-credit Professional Certificate

Google Data Analytics

Those who earn the Google Data Analytics Professional Certificate have completed eight courses, developed by Google, that include hands-on, practice-based assessments and are designed to prepare them for introductory-level roles in Data Analytics. They are competent in tools and platforms including spreadsheets, SQL, Tableau, and R. They know how to prepare, process, analyze, and share data for thoughtful action.

Amanda Brophy
Global Director of Google Career Certificates

The online specialization named in this certificate may draw on material from courses taught on-campus, but the included courses are not equivalent to on-campus courses. Participation in this online specialization does not constitute enrollment at this university. This certificate does not confer a University grade, course credit or degree, and it does not verify the identity of the learner.

Verify this certificate at:
https://coursera.org/verify/professional-cert/WK03UGC4962X

Thank you