Cyclistic Bike Share Case Study using Spreadsheet, SQL and Power BI

How does a bike-share navigate speedy success?

Introduction

In this case study, I analyze historical data from a Chicago based bike-share company in order to identify trends in how their customers use bikes differently. In order to answer the key business questions, I will follow the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act.

The main tools I used are Microsoft Excel, SQL and PowerBI.

Here are the **Quick links:**

Data Source: divvy_tripdata

SQL Queries:

1. **Prepare Data**
2. **Process Data**

Data Visualizations: PowerBI

A more in-depth breakdown of the case study scenario is included below, followed by my full report.

Scenario

As a junior data analyst on the marketing analyst team at Cyclistic, a bike-share company in Chicago, my role is to help uncover insights that drive business growth. The Director of Marketing believes that the key to the company's future success lies in increasing the number of annual memberships. To support this goal, my team is focused on analyzing how casual riders and annual members differ in their usage of Cyclistic bikes.

By identifying key trends and patterns, we aim to develop an effective marketing strategy that encourages casual riders to become annual members. However, before implementing any strategy, our recommendations must be reviewed and approved by Cyclistic's executives. To ensure a strong case, we will support our findings with data-driven insights and professional visualizations that clearly communicate the value of our proposed approach.

Background

Cyclistic is a bike-share program that operates over 5,800 bicycles across 600 docking stations. What sets Cyclistic apart is its commitment to inclusivity, offering not just traditional bikes but also reclining bikes, hand tricycles, and cargo bikes for riders with disabilities or those unable to use standard two-wheeled bikes. While most users prefer traditional bikes, about 8% opt for assistive options. The majority of riders use Cyclistic for leisure, though 30% rely on it for commuting to work.

So far, Cyclistic's marketing strategy has focused on building brand awareness and appealing to a broad customer base. A key factor in this approach has been the flexibility of its pricing plans, which include single-ride passes, full-day passes, and annual memberships. Customers using single-ride or full-day passes are classified as casual riders, while those who purchase annual memberships are considered Cyclistic members.

Financial analysis has shown that annual members are significantly more profitable than casual riders. While the flexible pricing model attracts a wide range of customers, Moreno, the Director of Marketing, believes that converting casual riders into annual members will be essential for Cyclistic's long-term growth. Instead of focusing on attracting brand-new customers, Moreno sees a strong opportunity to convert existing casual riders into loyal members, as they are already familiar with and actively using the service.

To achieve this goal, the marketing analytics team needs a deeper understanding of how casual riders and annual members differ, why casual riders might be motivated to buy a membership, and how digital marketing strategies could influence their decisions. Analyzing Cyclistic's historical bike trip data will help uncover key trends that can shape targeted marketing efforts to drive membership conversions.

Phase 1. Ask: Defining the Problem

Business Task

Devise marketing strategies to convert casual riders to members.

Key Stakeholders:

- **Director of Marketing:** Needs actionable insights to shape marketing campaigns.
- **Executive Team:** Requires data-backed recommendations to approve funding.
- **Marketing Analytics Team:** Supports your analysis and implements findings.

Analysis Questions

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

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

Phase-2. Prepare: Gather and Organize Data

Data Source

I used Cyclistic's historical trip data to analyze and identify trends from Jan 2024 to Dec 2024 which can be downloaded from divvy_tripdata. 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 from using riders' personally identifiable information. This means that we 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.

Data Organization

There are 12 files with naming convention of YYYYMM-divvy-tripdata and each file includes information for one month, such as the ride id, bike type, start time, end time, start station, end station, start location, end location, and whether the rider is a member or not. The corresponding column names are 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 and member_casual.

Gather and Organize Data

Steps:

1. Download Data: Obtain the last 12 months of Cyclistic trip data from the provided [data source].

2. **Organize Files:**

   - Unzip the downloaded files.
   - Store data files in a structured folder (e.g., raw data in one subfolder, processed data in another).

3. **Inspect Data:**

   - Open CSV files primarily in Excel to understand structure and identify key columns like ride_id, started_at, ended_at, member_casual.

4. **Verify Credibility:**

   - Ensure the data is accurate and follows Cyclistic's privacy and licensing guidelines.

Description of the Source Data

The table below shows the all column names and their data types. The ride_id column is our primary key.

Row	column_name	data_type
1	ride_id	STRING
2	rideable_type	STRING
3	started_at	TIMESTAMP
4	ended_at	TIMESTAMP
5	start_station_name	STRING
6	start_station_id	STRING
7	end_station_name	STRING
8	end_station_id	STRING
9	start_lat	FLOAT64
10	start_lng	FLOAT64
11	end_lat	FLOAT64
12	end_lng	FLOAT64
13	member_casual	STRING

Combining the Data Tables

I combined all the 12 months data by writing the following SQL query. I mainly used "UNION ALL" in between all the 12 months of data to combine all the separate tables into one comprehensive table. Here, I also carefully selected the necessary columns by using "SELECT".

```
SELECT
ride_id,
rideable_type,
started_at,
ended_at,
start_station_name,
end_station_name,
member_casual
FROM dbo.[202401-divvy-tripdata]
UNION ALL

SELECT
ride_id,
rideable_type,
started_at,
ended_at,
start_station_name,
end_station_name,
member_casual
FROM dbo.[202402-divvy-tripdata]
UNION ALL
```

Phase-3. Process: Cleaning and Transforming Data

Cleaning and Preparing Data for Analysis

🔷 I chose SQL Server Management Studio (SSMS) as my primary tool for data cleaning and transformation. SSMS provides robust SQL querying capabilities, allowing efficient data manipulation, validation, and cleaning. It enables handling large datasets, ensuring data integrity and consistency while optimizing performance.

🔷I ensured data integrity by performing:

- Removing duplicate rows to avoid skewed analysis.
- Checking for missing values and handling them appropriately.
- Validating data types to ensure consistency across columns.
- Verifying primary and foreign keys to maintain relational integrity.

🔷 I documented every SQL query and transformation step, including:

- Initial data issues found (e.g., duplicates, null values).
- The exact SQL queries used to clean the data.

- Before and after summaries (e.g., how many nulls were removed).
- Final validation steps to confirm data integrity.

Key Tasks Performed

✅ **Checked the data for errors** – Identified duplicates, null values.
✅ **Chose SQL Server (SSMS) for cleaning** – Due to its efficiency in handling large datasets.
✅ **Transformed the data** – Using SQL queries to clean, filter, and format it properly.
✅ **Documented the cleaning process** – SQL scripts and summaries recorded for reference.

Data Cleaning Documentation

Summary of Data Cleaning Steps Taken:

- **Identify and Remove Null values**

```
SELECT *
FROM dbo.cyclistic_bike
WHERE
ride_id IS NULL
OR rideable_type IS NULL
OR started_at IS NULL
OR ended_at  IS NULL
OR start_station_name IS NULL
OR end_station_name IS NULL
OR member_casual IS NULL;

DELETE FROM dbo.cyclistic_bike
WHERE
ride_id IS NULL
OR rideable_type IS NULL
OR started_at IS NULL
OR ended_at  IS NULL
OR start_station_name IS NULL
OR end_station_name IS NULL
OR member_casual IS NULL;
```

- **Identify and remove duplicate values**

```
SELECT *,
COUNT(*) AS duplicatecount
FROM cyclistic_bike
GROUP BY
ride_id,
rideable_type,
started_at,
ended_at,
start_station_name,
end_station_name,
member_casual
HAVING COUNT(*) >1;
```

Finally, I saved the cleaned data in a CSV file.

Phase-4. Analyze: Identifying the Trends

1. To analyze, perform necessary calculations, and get a better sense of the data layout, I imported the CSV file in PowerBI. I did necessary adjustments using the Power Query Editor i.e added columns like Ride Length, Day of the Week, Day Name, Month, Month Name.

Ride Length	Ride Length(Min)	Day of Week	Day Name	Month	Month Name
0.00:07:32	7.53	5	Friday	1	January
0.00:07:13	7.22	1	Monday	1	January
0.00:08:00	8	6	Saturday	1	January
0.00:29:49	29.82	1	Monday	1	January
0.00:26:12	26.2	3	Wednesday	1	January
0.00:08:39	8.65	0	Sunday	1	January
0.00:08:54	8.9	5	Friday	1	January
0.00:08:11	8.18	4	Thursday	1	January
0.00:10:09	10.15	1	Monday	1	January
0.00:08:57	8.95	3	Wednesday	1	January

2. Then, I closed and loaded the data and performed these calculations.

- Mean of ride_length
- Max ride_length
- Mode of day_of_week

17.32
Average of Ride Length(Min)

1.56K
Max of Ride Length(Min)

Wednesday
Mode Day of Week

To find the mode of the day I used the following DAX Query:

```
Mode_Day_of_Week =
VAR ModeTable =
    TOPN(1,
        SUMMARIZE(
            'Raw Data',
            'Raw Data'[Day Name],
            "Count", 'Raw Data'[Day Name]
        ),
        [Count], DESC
    )
RETURN
    MAXX(ModeTable, 'Raw Data'[Day Name])
```

3. Then I created some tables to quickly calculate and visualize the data.

- Average ride_length for members and casual riders

Membership Type	Average of Ride Length(Min)
casual	25.15
member	12.77
Total	**17.32**

- Average ride_length for users by day_of_week.

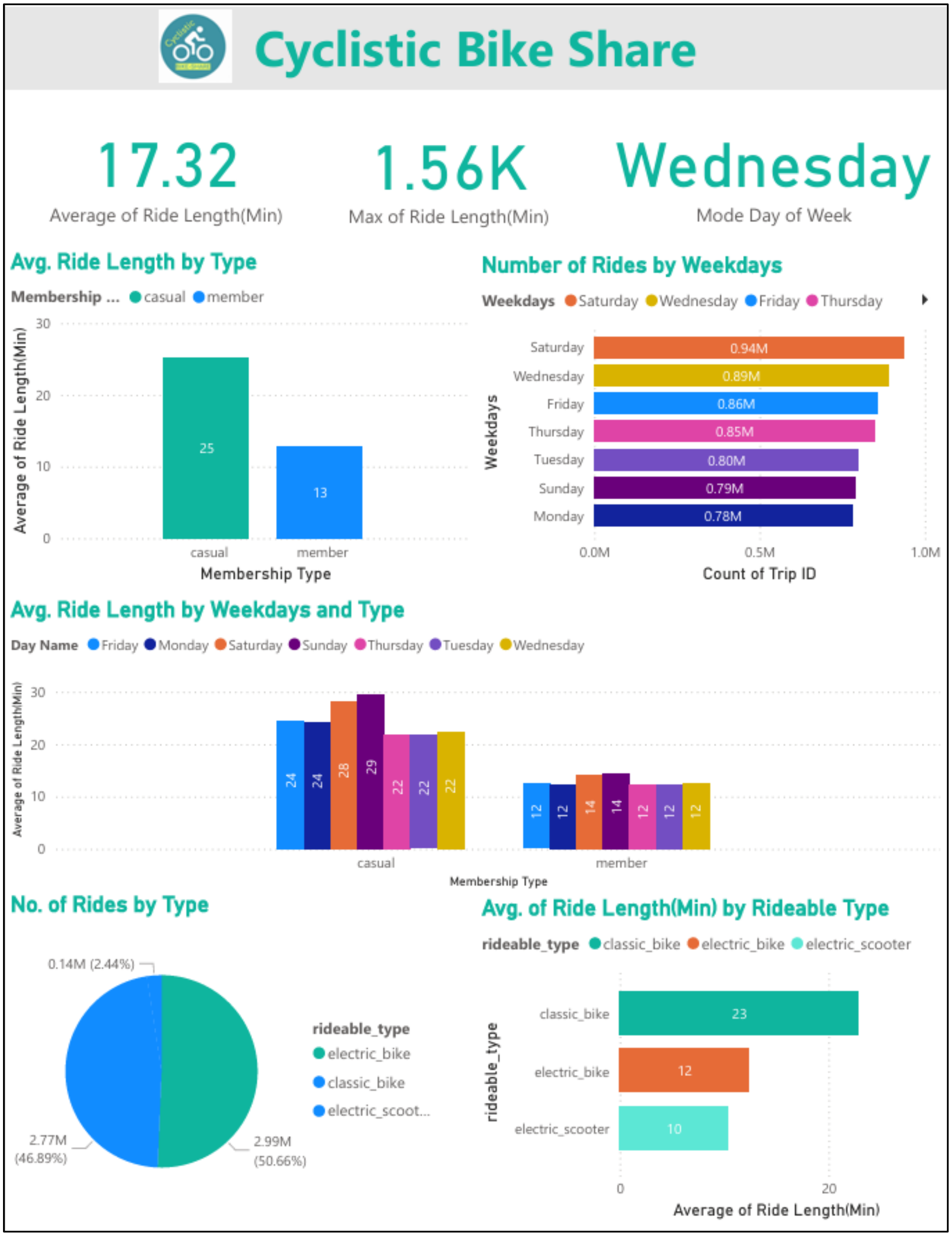
Membership Type	Average of Ride Length(Min)	Day Name
casual	29.42	Sunday
casual	28.19	Saturday
casual	24.53	Friday
casual	24.14	Monday
casual	22.28	Wednesday
casual	21.91	Thursday
casual	21.51	Tuesday
member	14.27	Sunday
member	14.07	Saturday
member	12.45	Wednesday
member	12.42	Friday
member	12.28	Tuesday
member	12.26	Thursday
member	12.21	Monday
Total	**17.32**	

- Number of rides for users by day_of_week

Weekdays	Count of ride_id
Saturday	925097
Wednesday	879625
Friday	841477
Thursday	835692
Tuesday	803077
Monday	788188
Sunday	787201
Total	**5860357**

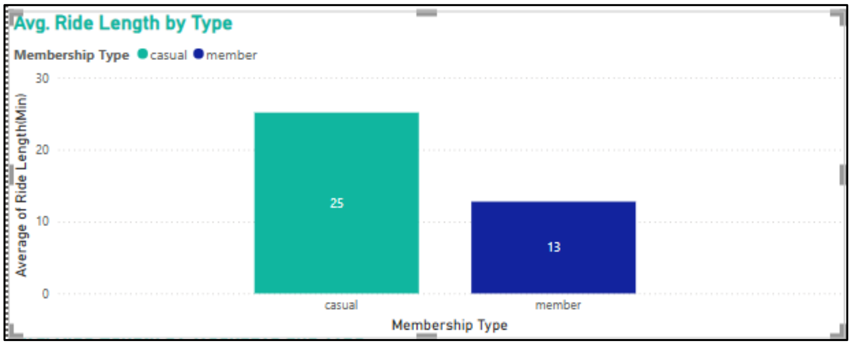
Phase-5. Share: Insights & Visualizing Report

PowerBI Dashboard:

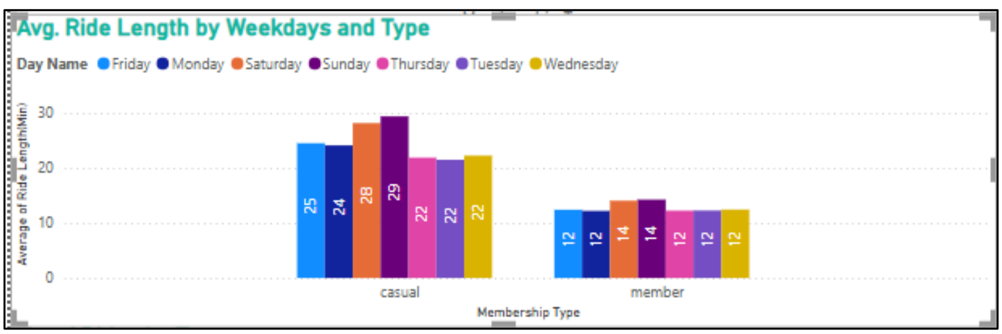


These visualizations clearly show significant differences between **casual riders and annual members** in their usage patterns:

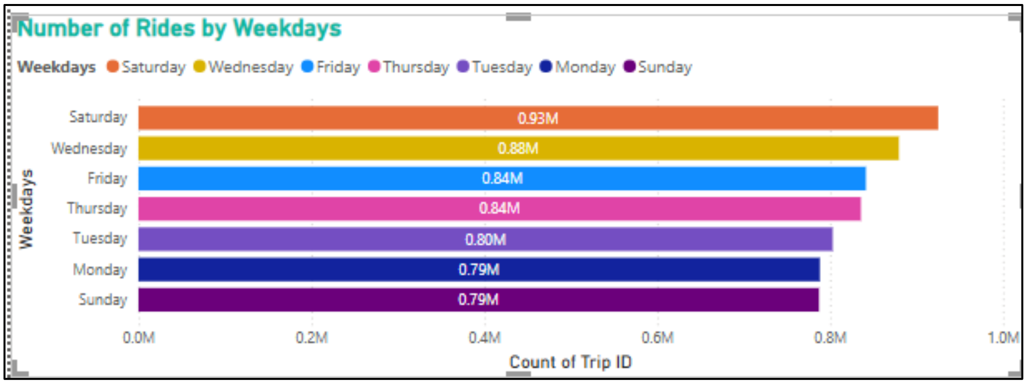
- **Ride Duration**:
    - Casual riders take longer trips, averaging **25 minutes** per ride.
    - Members have a significantly shorter average ride duration of **13 minutes**.
    - This suggests casual riders use bikes more for **leisure and exploration**, while members use them for **short, frequent commutes**.



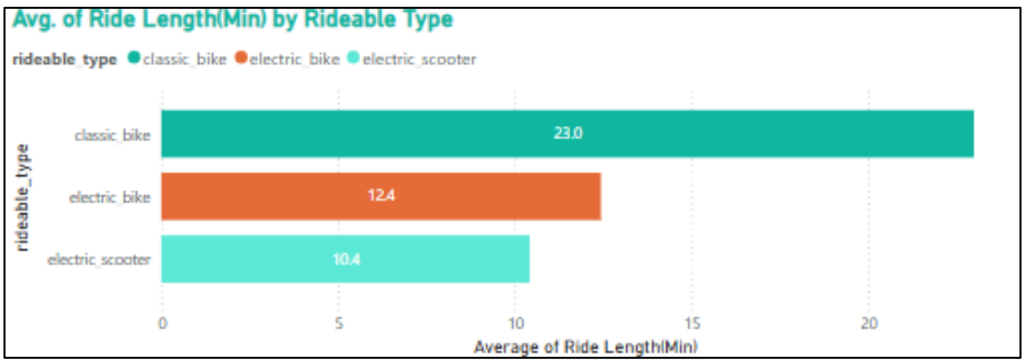
- **Usage by Weekdays**:
    - **Casual riders** ride longer on **weekends (Saturday & Sunday)**, with ride durations peaking on **Saturdays (~29 min)** and **Sundays (~28 min)**.
    - **Annual members** maintain **consistent shorter ride times (~12–14 min)** across all days, reinforcing the idea that they primarily use bikes for commuting rather than leisure.



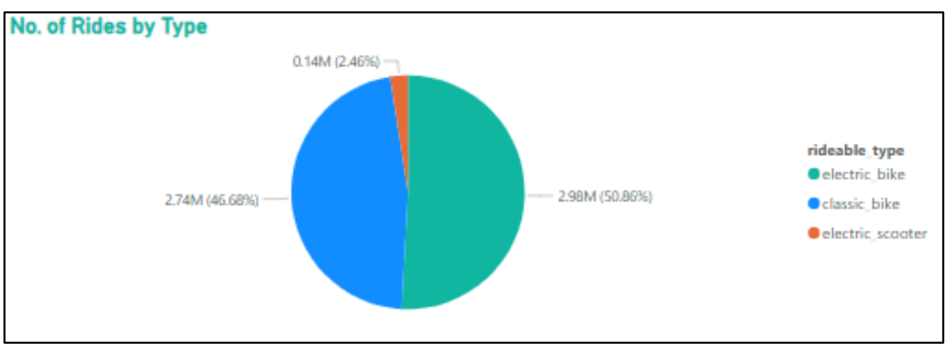
- **Total Rides by Weekday**:
    - The highest number of rides occurs on **Saturdays (~0.94M rides)**, followed by **Wednesdays (~0.89M rides)**.
    - Monday has the lowest number of trips (**~0.78M rides**), indicating reduced weekday commuting activity.



- **Usage by Rideable Type:** Classic bike rides average 23 minutes, much longer than electric bikes (12.4 minutes) or scooters (10.4 minutes), which are similar. This suggests different usage: classic bikes for longer trips, electric for shorter commutes.



- **Count by Types:** Classic and electric bikes dominate rides (50% and 46.7%), while scooters are barely used (2.5%).



These patterns suggest that casual riders **prefer weekends for longer leisure trips**, while annual members have **a consistent, commuter-driven usage pattern**.

Implications from the Data

The data tells a clear story about how different user segments engage with the Cyclistic bike-sharing service:

- Casual riders are leisure-oriented users, preferring weekend rides and longer trip durations.
- Annual members are regular commuters, taking short and frequent trips throughout the week.
- Electric bikes (12 min avg. ride time) and electric scooters (10 min avg.) are used for quick mobility, while classic bikes (23 min avg.) are favored for longer rides.
- Wednesdays are the most frequent day for rides, suggesting mid-week commuting peaks.

Phase-6. Act–Recommendations & Conclusion

Top 3 Recommendations for Cyclistic:

1. **Introduce a "Weekend Membership Plan"**

- Casual riders use bikes primarily on weekends for longer trips. Offering weekend-specific discounts or a limited-time membership could increase conversions.

2. **Enhance Bike Availability Based on Demand**

- Ensure classic bikes are readily available at high-demand stations during weekends.
- Deploy more electric bikes in commuter-heavy areas to attract daily users.

3. **Targeted Digital Marketing to Convert Casual Riders**

- Send app notifications & email offers to casual riders after a ride, promoting the benefits of membership**.**
- Offer **"first-month free" or seasonal discounts** to encourage sign-ups.

🔹 **Conclusion**: Casual riders represent a significant conversion opportunity, and targeted marketing efforts should focus on encouraging them to purchase memberships through incentives that match their riding habits.
