Analysis of Cyclistic Bike-Share Usage Patterns to Inform Marketing Strategy

By Author: Tejas Kadam

Introduction

Welcome to the comprehensive analysis of Cyclistic, a leading bike-share company in Chicago known for its robust and inclusive offerings, including traditional bicycles, reclining bikes, hand tricycles, and cargo bikes. This analysis focuses on unraveling the usage patterns of Cyclistic's two primary user groups: casual riders and annual members. By understanding these patterns, Cyclistic aims to refine its marketing strategies to increase the conversion of casual riders into annual members, thereby maximizing profitability and market share.

Cyclistic's success hinges on its ability to cater to a diverse customer base while continuously adapting to their changing needs. The insights derived from this study will provide a foundation for targeted marketing initiatives designed to enhance customer engagement and loyalty. This report is prepared under the guidance of Lily Moreno, Director of Marketing at Cyclistic, and aims to support the company's strategic decisions with data-driven evidence.

Business Task

Objective: The primary objective of this analysis is to identify and understand the differences in usage patterns between annual members and casual riders of Cyclistic's bike-share program. By doing so, the company seeks to develop tailored marketing strategies that effectively convert casual riders into annual members.

Scope of Analysis:

1. **Analytical Focus**: The analysis will focus on the following key aspects:

- Frequency of use: How often do casual riders and annual members use the service?
- Duration and timing of rides: What are the typical durations and peak usage times for each user group?
- Type of bikes used: Are there preferences for specific types of bikes between the two groups?

2. **Data Utilization**: Utilize the last 12 months of trip data provided by Cyclistic, ensuring all analyses adhere to privacy standards and data integrity.

3. **Outcome Expectations**: The ultimate goal is to craft actionable insights that:

- Inform the development of promotional campaigns specifically tailored to convert casual riders into annual members.
- Support strategic decisions by the executive team with compelling data insights and professional data visualizations.

Deliverables:

- A detailed report including a comparison of usage patterns between the two user groups.
- Data-driven recommendations for targeted marketing strategies.
- Visualizations that highlight key differences and support the analytical findings.

By addressing these areas, the marketing team aims to not only increase the number of annual memberships but also enhance the overall customer experience, contributing to Cyclistic's long-term growth and success.

Data Preparation and Analysis

Data Preparation

The dataset provided by the Google Data Analytics Professional Course encompassed 12 months of usage data for Cyclistic bike-share. The initial step involved combining the monthly data files into a single dataset using SQL Server Management System. This allowed for a unified view of the data for subsequent analysis.

After consolidating the datasets, the comprehensive data file was exported to a CSV format for more detailed analysis using Python in Google Colab.

Data Cleaning and Transformation in Python

The CSV file was loaded into a Pandas DataFrame, and several data cleaning steps were performed to prepare the dataset for analysis. These steps included renaming columns for clarity, handling missing values, and converting date and time columns into a suitable format for analysis.

```
[ ] from google.colab import drive
    drive.mount('/content/drive')

    Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
[ ] file_path = '/content/drive/MyDrive/Casestudy_dataset/Cyclist_bike_dataset.csv'
```

```
[ ] import pandas as pd
```

```
[ ] df = pd.read_csv('/content/drive/MyDrive/Casestudy_dataset/Cyclist_bike_dataset.csv', header=None)
```

```
new_headers = ['ride_id', 'rideable_type', 'started_at', 'ended_at', 'start_station_name', 'end_station_name', 'member_casual']

df.columns = new_headers

df.head()
```

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual
0	6842AA605EE9FBB3	electric_bike	2023-03-16 08:20:34.0000000	2023-03-16 08:22:52.0000000	Clark St & Armitage Ave	Larrabee St & Webster Ave	member
1	F984267A75B99A8C	electric_bike	2023-03-04 14:07:06.0000000	2023-03-04 14:15:31.0000000	Public Rack - Kedzie Ave & Argyle St	NaN	member
2	FF7CF57CFE026D02	classic_bike	2023-03-31 12:28:09.0000000	2023-03-31 12:38:47.0000000	Orleans St & Chestnut St (NEXT Apts)	Clark St & Randolph St	member
3	6B61B916032CB6D6	classic_bike	2023-03-22 14:09:08.0000000	2023-03-22 14:24:51.0000000	Desplaines St & Kinzie St	Sheffield Ave & Kingsbury St	member
4	E55E61A5F1260040	electric_bike	2023-03-09 07:15:00.0000000	2023-03-09 07:26:00.0000000	Walsh Park	Sangamon St & Lake St	member

4 E55E61A5F1260040 electric_bike 2023-03-09 07:15:00.0000000 2023-03-09 07:26:00.0000000 Walsh Park Sangam

```
[ ] df.to_csv('your_file_with_new_headers.csv', index=False)
```

```
[ ] df.isnull().sum()

    ride_id                    0
    rideable_type              0
    started_at                 0
    ended_at                   0
    start_station_name    866105
    end_station_name      920524
    member_casual              0
    dtype: int64
```

```
[ ]
    df = df.dropna()

    df['start_station_name'].fillna('Unknown', inplace=True)
    df['end_station_name'].fillna('Unknown', inplace=True)
```

```
df.isnull().sum()

ride_id               0
rideable_type         0
started_at            0
ended_at              0
start_station_name    0
end_station_name      0
member_casual         0
dtype: int64
```

```
[ ] import pandas as pd


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

```
[ ] start_station_name    0
    end_station_name      0
    member_casual         0
    dtype: int64
```

```python
import pandas as pd


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


df['start_date'] = df['started_at'].dt.date
df['end_date'] = df['ended_at'].dt.date


df['start_hour'] = df['started_at'].dt.hour
df['end_hour'] = df['ended_at'].dt.hour


df[['started_at', 'start_date', 'start_hour', 'ended_at', 'end_date', 'end_hour']].head()
```

	started_at	start_date	start_hour	ended_at	end_date	end_hour
0	2023-03-16 08:20:34	2023-03-16	8	2023-03-16 08:22:52	2023-03-16	8
2	2023-03-31 12:28:09	2023-03-31	12	2023-03-31 12:38:47	2023-03-31	12
3	2023-03-22 14:09:08	2023-03-22	14	2023-03-22 14:24:51	2023-03-22	14
4	2023-03-09 07:15:00	2023-03-09	7	2023-03-09 07:26:00	2023-03-09	7
5	2023-03-22 17:47:02	2023-03-22	17	2023-03-22 18:01:29	2023-03-22	18

```
[ ] df.head()
```

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	start_date	end_date	start_hour	end_hour
0	6842AA605EE9FBB3	electric_bike	2023-03-16 08:20:34	2023-03-16 08:22:52	Clark St & Armitage Ave	Larrabee St & Webster Ave	member	2023-03-16	2023-03-16	8	8
			2023-03-31	2023-03-	Orleans St & Chestnut	Clark St & Randolph			2023-03-		

```python
[ ]
df['total_ride'] = 1


df.head()
```

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	start_date	end_date	start_hour	end_hour	tota
0	6842AA605EE9FBB3	electric_bike	2023-03-16 08:20:34	2023-03-16 08:22:52	Clark St & Armitage Ave	Larrabee St & Webster Ave	member	2023-03-16	2023-03-16	8	8	
2	FF7CF57CFE026D02	classic_bike	2023-03-31 12:28:09	2023-03-31 12:38:47	Orleans St & Chestnut St (NEXT Apts)	Clark St & Randolph St	member	2023-03-31	2023-03-31	12	12	
3	6B61B916032CB6D6	classic_bike	2023-03-22 14:09:08	2023-03-22 14:24:51	Desplaines St & Kinzie St	Sheffield Ave & Kingsbury St	member	2023-03-22	2023-03-22	14	14	
4	E55E61A5F1260040	electric_bike	2023-03-09 07:15:00	2023-03-09 07:26:00	Walsh Park	Sangamon St & Lake St	member	2023-03-09	2023-03-09	7	7	
5	123AAD676850F53C	classic_bike	2023-03-22 17:47:02	2023-03-22 18:01:29	Orleans St & Chestnut St (NEXT Apts)	Halsted St & Wrightwood Ave	member	2023-03-22	2023-03-22	17	18	

```python
df['total_duration_minutes'] = (df['ended_at'] - df['started_at']).dt.total_seconds() / 60


df.head()
```

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	start_date	end_date	start_hour	end_hour	tota

```
df['day_of_week'] = df['started_at'].dt.day_name()

df.head()
```

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	start_date	end_date	start_hour	end_hour	total
0	6842AA605EE9FBB3	electric_bike	2023-03-16 08:20:34	2023-03-16 08:22:52	Clark St & Armitage Ave	Larrabee St & Webster Ave	member	2023-03-16	2023-03-16	8	8	
2	FF7CF57CFE026D02	classic_bike	2023-03-31 12:28:09	2023-03-31 12:38:47	Orleans St & Chestnut St (NEXT Apts)	Clark St & Randolph St	member	2023-03-31	2023-03-31	12	12	
3	6B61B916032CB6D6	classic_bike	2023-03-22 14:09:08	2023-03-22 14:24:51	Desplaines St & Kinzie St	Sheffield Ave & Kingsbury St	member	2023-03-22	2023-03-22	14	14	
4	E55E61A5F1260040	electric_bike	2023-03-09 07:15:00	2023-03-09 07:26:00	Walsh Park	Sangamon St & Lake St	member	2023-03-09	2023-03-09	7	7	
5	123AAD676850F53C	classic_bike	2023-03-22 17:47:02	2023-03-22 18:01:29	Orleans St & Chestnut St (NEXT Apts)	Halsted St & Wrightwood Ave	member	2023-03-22	2023-03-22	17	18	

[ ]

```
df['started_at'] = pd.to_datetime(df['started_at'], errors='coerce')

df['month_of_year'] = df['started_at'].dt.month

df.head()
```

[ ]

	ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	start_date	end_date	start_hour	end_hour
0	6842AA605EE9FBB3	electric_bike	2023-03-16 08:20:34	2023-03-16 08:22:52	Clark St & Armitage Ave	Larrabee St & Webster Ave	member	2023-03-16	2023-03-16	8	8
2	FF7CF57CFE026D02	classic_bike	2023-03-31 12:28:09	2023-03-31 12:38:47	Orleans St & Chestnut St (NEXT Apts)	Clark St & Randolph St	member	2023-03-31	2023-03-31	12	12
3	6B61B916032CB6D6	classic_bike	2023-03-22 14:09:08	2023-03-22 14:24:51	Desplaines St & Kinzie St	Sheffield Ave & Kingsbury St	member	2023-03-22	2023-03-22	14	14
4	E55E61A5F1260040	electric_bike	2023-03-09 07:15:00	2023-03-09 07:26:00	Walsh Park	Sangamon St & Lake St	member	2023-03-09	2023-03-09	7	7
5	123AAD676850F53C	classic_bike	2023-03-22 17:47:02	2023-03-22 18:01:29	Orleans St & Chestnut St (NEXT Apts)	Halsted St & Wrightwood Ave	member	2023-03-22	2023-03-22	17	18

[ ]

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

[ ]

```
processed_file_path = '/content/drive/My Drive/clean_data.csv'

df.to_csv(processed_file_path, index=False)
```

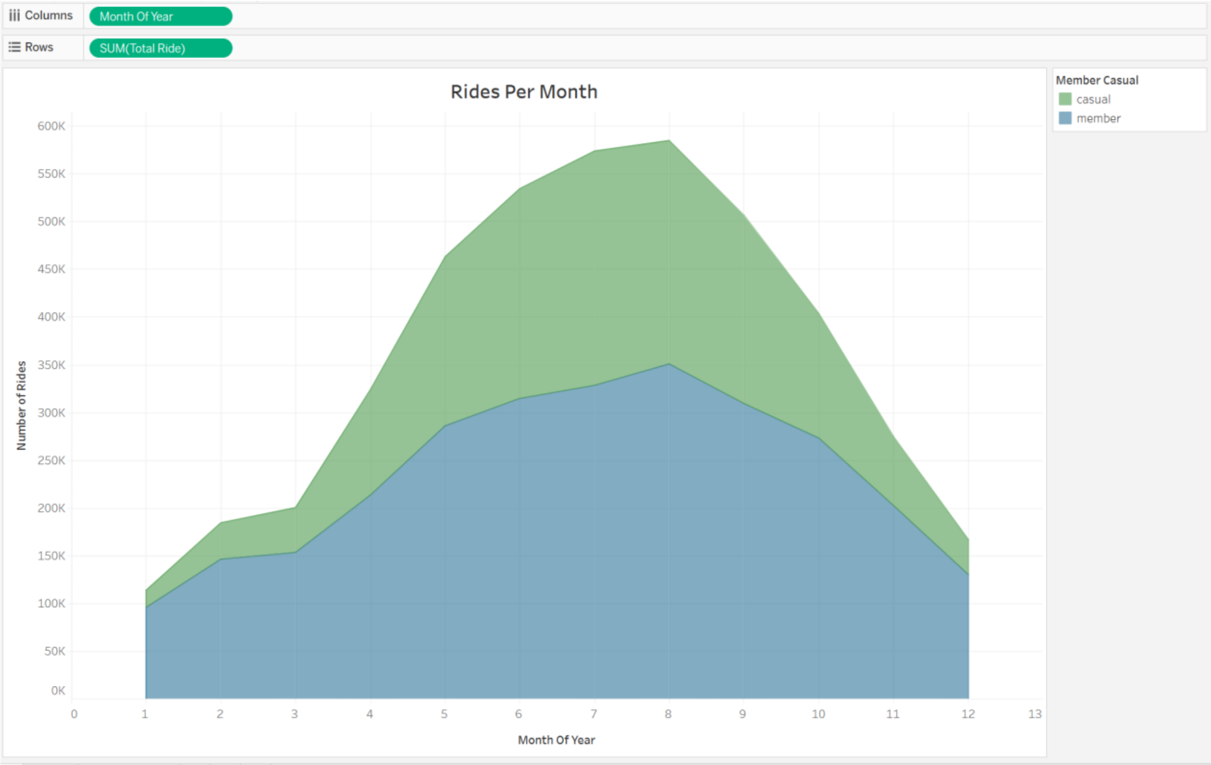
Analysis and Insights

Using the cleaned data, several analyses were performed to understand the differences in usage patterns between casual riders and annual members. Key metrics calculated included the total number of rides, average ride duration, and usage patterns by day of the week and month.

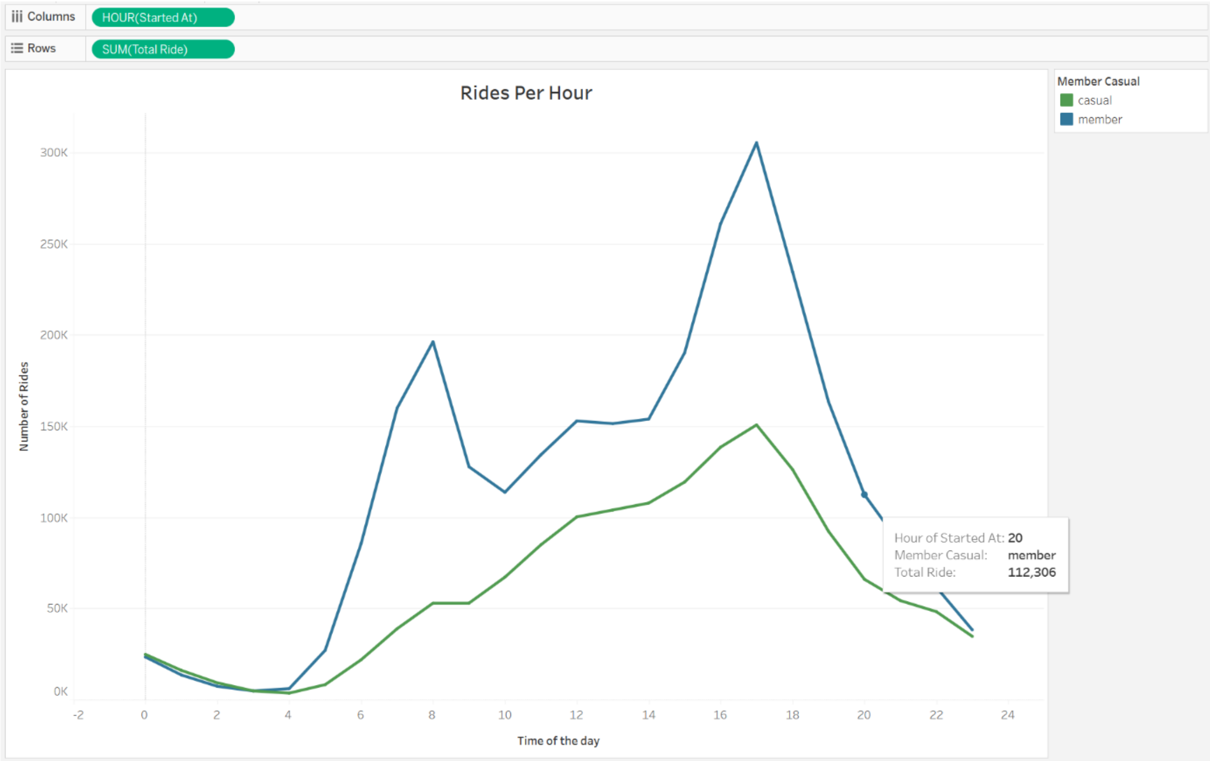
Visualizing Insights with Tableau

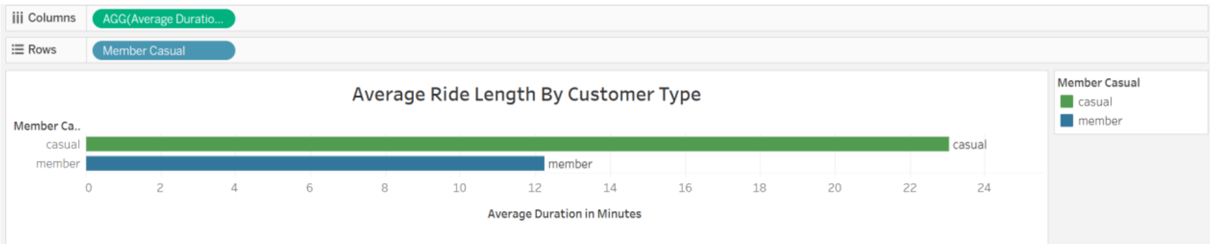
The insights garnered from the analysis were visualized using Tableau, which provided an interactive platform to explore the nuances of the data. Visualizations created included:

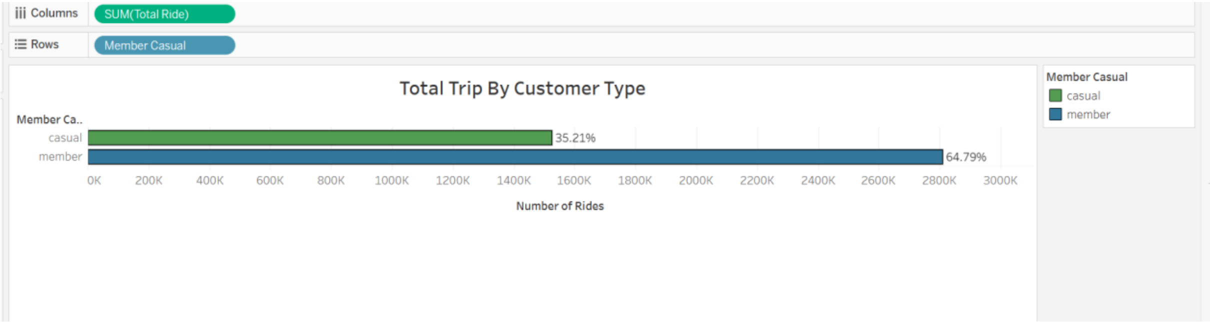
- Distribution of ride durations.
- Comparison of usage patterns between member types.
- Area chart of ride frequency by time and day.




Columns
Member Casual
Day Of Week
Rows
SUM(Total Ride)
Day of Week
casual
member
Number of Rides
450K
400K
350K
300K
250K
200K
150K
100K
50K
0K
Sunday Monday Tuesday Wednes.. Thursday Friday Saturday Sunday Monday Tuesday Wednes.. Thursday Friday Saturday
Day Of Week
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


Columns
HOUR(Started At)
Rows
SUM(Total Ride)
Rides Per Hour
Member Casual
casual
member
Number of Rides
300K
250K
200K
150K
100K
50K
0K
Hour of Started At: 20
Member Casual: member
Total Ride: 112,306
-2 0 2 4 6 8 10 12 14 16 18 20 22 24
Time of the day


Columns
AGG(Average Duratio...
Rows
Member Casual
Average Ride Length By Customer Type
Member Casual
casual
member
Member Ca..
casual
member
0 2 4 6 8 10 12 14 16 18 20 22 24
Average Duration in Minutes


Columns
SUM(Total Ride)
Rows
Member Casual
Total Trip By Customer Type
Member Ca..
casual
member
35.21%
64.79%
0K
200K
400K
600K
800K
1000K
1200K
1400K
1600K
1800K
2000K
2200K
2400K
2600K
2800K
3000K
Number of Rides
Member Casual
casual
member


Columns
Member Casual
Rideable Type
Rows
SUM(Total Ride)
Total Rides by Bike Type
casual
member
Total Ride
1800K
1600K
1400K
1200K
1000K
800K
600K
400K
200K
0K
casual
casual
casual
member
member
classic_bike
docked_bike
electric_bike
classic_bike
electric_bike

Key Findings and Visualizations

Total Trip by Customer Type

The first visualization illustrates the distribution of total trips between casual and member riders. It is evident that members account for a significantly larger proportion of total rides at approximately 65%, while casual riders contribute around 35%. This indicates a robust base of recurring users within the membership structure, highlighting the potential revenue stability from members as opposed to casual riders.

Average Ride Length by Customer Type

The analysis of average ride length reveals that members generally have shorter trip durations compared to casual riders. This suggests that members may be using the service for routine or commute purposes, which are typically shorter but more frequent trips. In contrast, casual riders, who might use the service sporadically, possibly for leisure or exploration, tend to have longer ride times.

Rides Per Hour

Ride frequency by hour shows clear peak times, with member usage peaking during typical rush hours, reinforcing the notion that the service is utilized for commuting. Casual riders, however, show a more spread out pattern, indicating less predictability and a wider range of usage times, likely reflecting a more leisurely use of the service.

Rides by Day of the Week

When examining the distribution of rides throughout the week, both user types show increased activity on weekends, with casual riders showing a more pronounced increase. This further supports the idea that casual riders are more likely to use bikes for leisure activities during weekends.

Rides Per Month

Seasonality plays a significant role in bike-share usage, with a marked increase in rides during the warmer months for both member and casual riders. Members consistently use the service throughout the year, whereas casual use dips more dramatically during colder months.

Total Rides by Bike Type

Finally, the total rides by bike type visualization reveals preferences for bike types among the two rider groups. Members show a strong preference for classic bikes, while casual riders have a more varied usage pattern, including docked and electric bikes, suggesting that they are more open to trying different types of rides.

Recommendations Based on Analysis

1. **Targeted Marketing for Peak Hours**: Since members primarily use the service during rush hours, marketing efforts could be geared towards promoting membership as a reliable and convenient commuting option.
2. **Weekend Promotions**: Given the increased activity of casual riders on weekends, offering weekend promotions or events could incentivize casual riders to convert to members, perhaps through a points system that rewards frequent use.
3. **Seasonal Campaigns**: Addressing the seasonality in usage with targeted campaigns during warmer months could help in retaining casual riders and converting them into members by capitalizing on their higher engagement periods.