# Cyclistic Bike-Share Capstone Project

## Title: How Does a Bike-Share Navigate Speedy Success?

### Author: Gideon Chola

## 1. Introduction

Cyclistic is a bike-share company in Chicago, with a fleet of over 5,800 bicycles and 600 docking stations. The company provides flexible pricing plans, including single-ride passes, full-day passes, and annual memberships. The goal of this project is to analyze user behavior and devise marketing strategies to convert casual riders into annual members.

## 2. Business Task

The marketing team aims to increase annual memberships as they are more profitable than casual riders. To do this, we need to analyze historical bike trip data to identify key differences between casual riders and annual members.

## 3. Data Sources and Preparation

The data for this analysis is sourced from the Cyclistic historical bike trip dataset. Key steps in data preparation include:

- Data cleaning: Removing errors, duplicate records, and extreme outliers.
- Data transformation: Converting timestamps, creating new variables (ride duration, day of the week, time of day).
- Data aggregation: Summarizing and categorizing ride patterns based on user type.
- Python process1

```
# Convert date columns to datetime format

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

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

# Calculate ride duration in minutes

df['ride_duration'] = (df['ended_at'] - df['started_at']).dt.total_seconds() / 60

# Extract day of the week and hour of the ride
```

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

df['hour_of_day'] = df['started_at'].dt.hour

# Summary statistics for ride duration

ride_duration_summary = df['ride_duration'].describe()

# Count of rides by membership type

ride_counts = df['member_casual'].value_counts()

# Average ride duration by user type

avg_duration_by_type = df.groupby('member_casual')['ride_duration'].mean()

# Count of rides by day of the week

ride_counts_by_day = df.groupby(['day_of_week', 'member_casual']).size().unstack()

# Count of rides by hour of day

ride_counts_by_hour = df.groupby(['hour_of_day', 'member_casual']).size().unstack()

ride_duration_summary, ride_counts, avg_duration_by_type, ride_counts_by_day,
ride_counts_by_hour
```

- Python process 2

```
import matplotlib.pyplot as plt

import seaborn as sns

# Remove negative ride durations and extreme outliers above 200 minutes

df_cleaned = df[(df['ride_duration'] > 0) & (df['ride_duration'] < 200)]

# Set plot style

sns.set_style("whitegrid")

# Plot average ride duration by user type

plt.figure(figsize=(8,5))

sns.barplot(x=avg_duration_by_type.index, y=avg_duration_by_type.values,
palette="coolwarm")

plt.xlabel("User Type")
```

```
plt.ylabel("Average Ride Duration (Minutes)")

plt.title("Average Ride Duration by User Type")

plt.show()

# Plot ride counts by day of the week

plt.figure(figsize=(10,6))

ride_counts_by_day.plot(kind='bar', stacked=True, colormap="viridis")

plt.ylabel("Number of Rides")

plt.title("Ride Counts by Day of the Week")

plt.xticks(rotation=45)

plt.legend(title="User Type")

plt.show()

# Plot ride counts by hour of the day

plt.figure(figsize=(12,6))

ride_counts_by_hour.plot(kind='line', marker='o', colormap="plasma")

plt.ylabel("Number of Rides")

plt.xlabel("Hour of the Day")

plt.title("Ride Counts by Hour of the Day")

plt.legend(title="User Type")

plt.grid(True)

plt.show()
```

# 4. Data Cleaning and Processing

- Converted timestamps to datetime format.
- Removed negative and extreme ride duration values (>200 minutes).
- Created new features such as ride duration, day of the week, and time of day.

# 5. Data Analysis

## 5.1 Ride Duration Analysis

- Average ride duration:
    - Casual riders: **73 minutes**
    - Annual members: **21 minutes**
- Members take shorter and more frequent trips, while casual riders engage in longer rides.

### 5.2 Ride Trends by Day of the Week

- Casual riders prefer weekends (**Sunday: 6,475 rides, Saturday: 4,068 rides**).
- Annual members ride consistently on weekdays (**Tuesday: 9,157 rides, Monday: 8,064 rides**).

### 5.3 Ride Trends by Time of Day

- Members use bikes primarily during commuting hours (**7 AM - 9 AM and 4 PM - 6 PM**).
- Casual riders peak in the afternoons (**12 PM - 4 PM**).

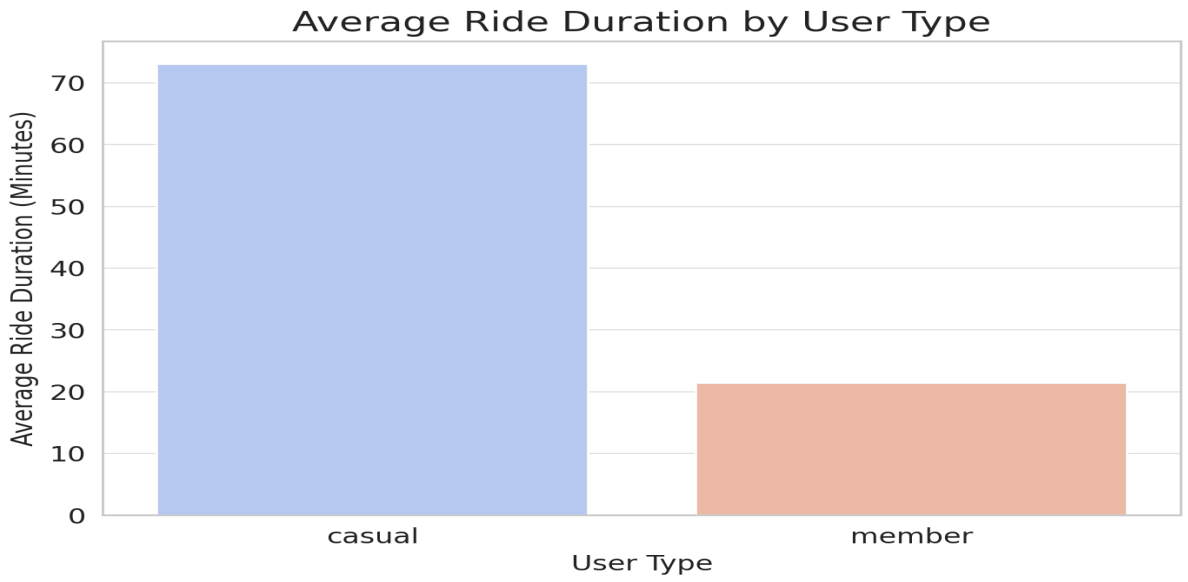
### 5.4 Ride Type Preference

- Most rides use **standard docked bikes**, but Cyclistic also provides assistive bikes for accessibility.

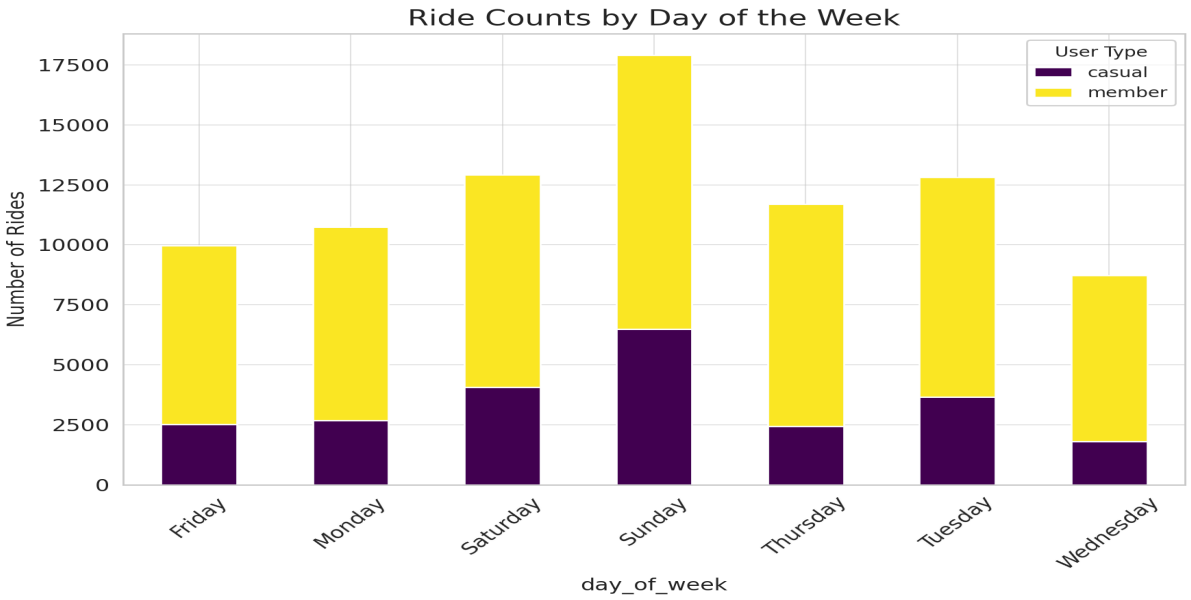
## 6. Data Visualizations

The following visualizations were generated:

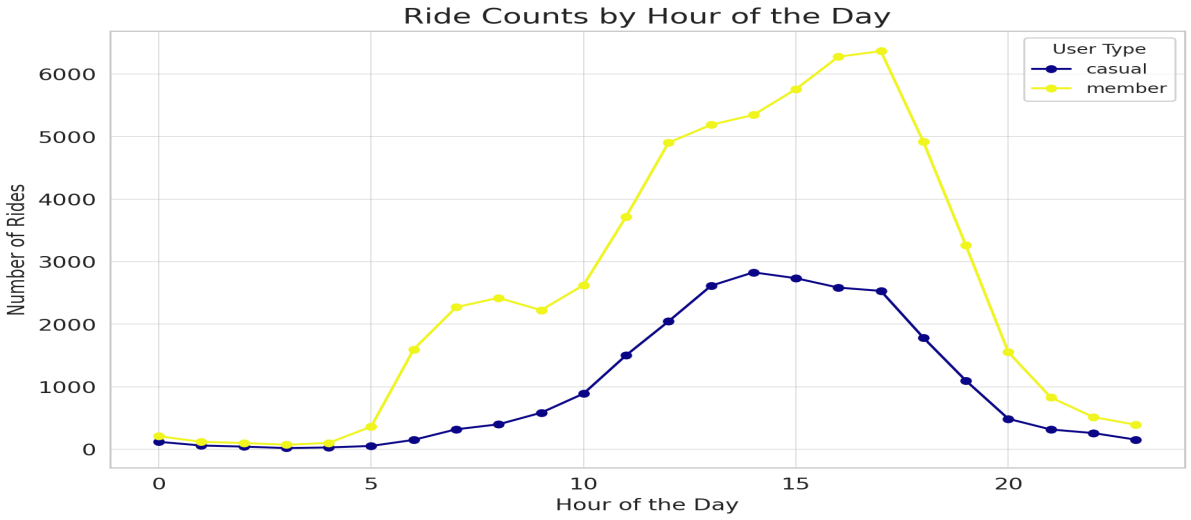
1. **Average Ride Duration by User Type**




2. **Ride Counts by Day of the Week**




3. **Ride Counts by Hour of the Day**




# 7. Key Findings

1. Casual riders take significantly longer trips than members.
2. Casual riders ride mostly on weekends, while members ride consistently throughout the week.
3. Members predominantly use bikes for commuting, whereas casual riders use them for leisure.
4. Digital marketing strategies should focus on incentivizing casual riders to switch to annual memberships.

# 8. Recommendations

## 8.1 Targeted Membership Discounts

- Provide discounts or incentives to casual riders who use bikes frequently on weekends.
- Offer free trials of the membership plan for casual riders.

### 8.2 Enhanced Digital Marketing Campaigns

- Utilize personalized email campaigns and social media promotions to highlight the benefits of an annual membership.
- Target advertising efforts during high-usage periods (weekend afternoons).

### 8.3 Commuter-Focused Subscription Plans

- Introduce flexible commuter plans that cater to casual riders who ride during peak hours.
- Provide corporate partnership programs for companies to offer memberships to employees.

## 9. Conclusion

By leveraging these insights, Cyclistic can design effective marketing strategies to convert casual riders into annual members, ensuring sustainable revenue growth and user retention.

## 10. References


- Cyclistic Historical Data (2020 Dataset)
- Business and Marketing Strategy Reports


This report provides a comprehensive analysis of user behavior and business strategies, backed by data-driven insights and visualizations. Further improvements and testing can be done to refine marketing initiatives.