Cyclistic: 2019 Q1 - Q4

Analysis of Cyclistic Bike-Share Usage Patterns

Shivam Amrutkar

July 22, 2024

Executive Summary

This report provides a comprehensive analysis of bike-sharing data, focusing on the usage patterns of casual riders and annual members. The analysis aims to uncover key differences in trip duration, age distribution, and weekly usage patterns between these two user types.

The findings reveal that casual riders tend to have significantly longer average trip durations, with an average of approximately 48 minutes, compared to annual members, who average around 14 minutes per trip. Additionally, the age distribution indicates that annual members are generally older, with an average age of 36 years, while casual riders average 32 years. This demographic insight suggests that the bike-sharing program attracts a younger audience among casual users.

Furthermore, the analysis of usage patterns by day of the week highlights distinct trends, with variations in the proportion of rides taken by casual riders versus annual members. These insights are crucial for understanding user behavior and optimizing service offerings.

Based on the findings, several recommendations are proposed, including targeted marketing strategies aimed at younger demographics to increase membership sign-ups, adjustments to bike availability based on peak usage days, and initiatives to encourage casual riders to transition to annual memberships.

Key Findings

Usage Patterns:

- Casual riders tend to have longer average ride durations compared to annual members.
- Annual members exhibit more frequent usage, especially on weekdays, indicating a potential preference for commuting purposes.
- Peak usage times for casual riders are during weekends and holidays, while annual members show consistent usage throughout the week.

Parameter	Casual	Member
Frequent rides - Day	Saturday (Weekends & Holidays)	Tuesday (Mostly Weekdays)
Frequent rides - Month	August	August
Average ride duration	48 minutes	14 minutes
Average Age	32 years	36 years

Introduction

This case study focuses on Chicago's bike-sharing program, which has been operational since 2016. The program offers two primary user types: casual riders, who purchase single-ride or day passes, and annual members, who subscribe to the service for long-term use. By analyzing the usage patterns of these two distinct user groups, we aim to gain valuable insights that can inform strategic decisions, improve service offerings, and enhance the overall efficiency of the bike-sharing system.

The findings and recommendations presented in this case study will be valuable not only for Chicago's bike-sharing program but also for similar initiatives in other urban areas. By deepening our understanding of user behavior and preferences, we can contribute to the broader goal of creating more sustainable, efficient, and user-friendly urban transportation systems.

Objectives

- To compare and contrast the usage patterns of casual riders and annual members across various dimensions, including trip duration, frequency, and timing.
- To analyze demographic factors, such as age distribution, that may influence bike-sharing preferences and behaviors.
- To identify trends in usage patterns throughout the week, potentially uncovering opportunities for targeted service improvements.
- To provide data-driven recommendations for optimizing the bike-sharing program, including strategies for converting casual riders to annual members and enhancing overall user experience.

This report leverages datasets collected from divvy-tripdata.s3.amazonaws.com/index.html for 2019 Q1 to Q4. Through rigorous analysis and visualization of this data, we seek to uncover meaningful patterns and trends that can drive strategic decision-making and operational improvements.

Data Collection & Preparation

Source: http://divvy-tripdata.s3.amazonaws.com/index.html

Datasets used for analysis:

1. 2019 Q1: https://divvy-tripdata.s3.amazonaws.com/Divvy_Trips_2019_Q1.zip
2. 2019 Q2: https://divvy-tripdata.s3.amazonaws.com/Divvy_Trips_2019_Q2.zip
3. 2019 Q3: https://divvy-tripdata.s3.amazonaws.com/Divvy_Trips_2019_Q3.zip
4. 2019 Q4: https://divvy-tripdata.s3.amazonaws.com/Divvy_Trips_2019_Q4.zip

Cleaning

- Removing Duplicate Entries: I identified and removed ***114 duplicate ride records*** to prevent data skew.
- Handling Missing Values: ***559,208 records*** with missing gender or age were excluded from the analysis.
- Data Type Conversion: Date and time fields were converted to the appropriate datetime format for easier manipulation.

Feature Engineering

- I calculated ride duration *(trip_duration)* in seconds from the start time and end time values.
- User age *(age)* was calculated based on the birth year and the year of the ride (taken as 2020).
- The day of the week *(day)* was extracted from the start date for temporal analysis.

Final Dataset

After the cleaning and preparation process, our final dataset consisted of ***32,58,682 valid ride records***. This cleaned dataset forms the basis of our subsequent analysis, ensuring that our findings are based on accurate and reliable data.

Analysis Methodology

The analysis of the bike-sharing dataset was conducted using a systematic approach to ensure that the findings are robust, reliable, and actionable.

1. Exploratory Data Analysis (EDA)

- Descriptive Statistics: Calculating summary statistics (mean, median, mode, standard deviation) for key variables such as trip duration, user age, and ride frequency.
- Data Visualization: Creating visual representations of the data using histograms, box plots, and scatter plots to identify trends, patterns, and potential outliers.

2. User Segmentation

- The dataset was segmented into two primary user types: casual riders and annual members. This segmentation allowed for a comparative analysis. Key metrics analyzed included: Average trip duration, Frequency of rides per user, Age distribution

3. Temporal Analysis

- Aggregating Data: Summarizing the number of rides for each user type by day of the week.

4. Data Visualization

- To effectively communicate the findings, various data visualization techniques were employed, including:
    - Bar Plot
    - Histogram
    - Box Plot

5. Interpretation of Results

- The final step involved interpreting the results in the context of the bike-sharing program's objectives. Insights were drawn from the analysis to inform recommendations for improving user engagement, optimizing service offerings, and enhancing overall user experience.

Findings

The analysis of the bike-sharing dataset revealed several significant insights regarding the usage patterns of casual riders and annual members. The key findings are summarized below:

1. Trip Duration

Average Trip Duration: Casual riders exhibited a significantly longer average trip duration of approximately **48 minutes**, compared to annual members, who averaged around **14 minutes**.
This suggests that casual riders may be using the service for leisure or recreational purposes, while annual members tend to use it for shorter, more utilitarian trips.

user_type	count	mean	std	min	25%	50%	75%	max
Casual	344180.0	2869.256720	46106.346662	61.0	795.0	1385.0	2425.0	7420632.0
Member	2914616.0	858.484237	15200.803571	61.0	361.0	587.0	966.0	9056633.0

Note: Data in seconds

2. User Demographics

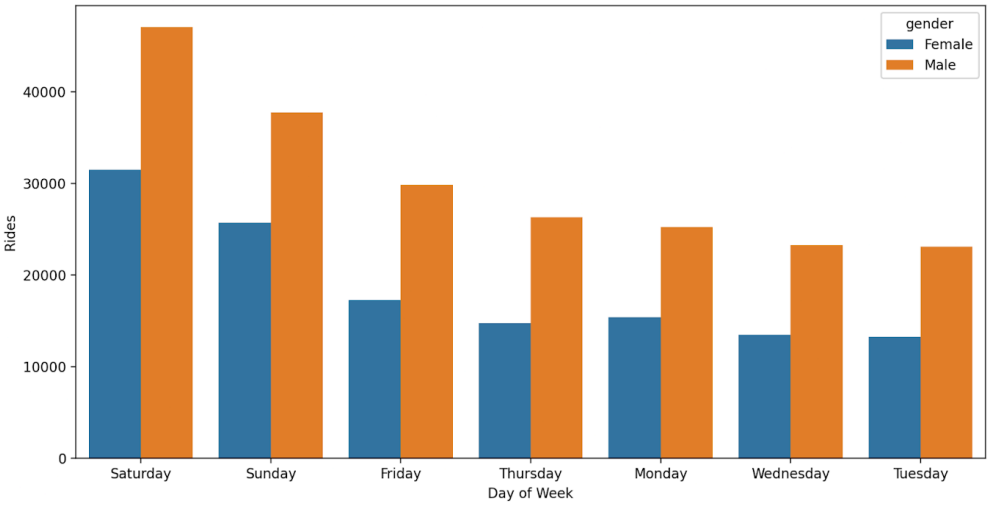
Age Distribution: The average age of casual riders was found to be **32 years**, while annual members had an average age of **36 years**.
This indicates that the bike-sharing program attracts a slightly younger audience among casual users.

user_type	count	mean	std	min	25%	50%	75%	max
Casual	344180.0	31.948786	9.614827	17.0	25.0	29.0	35.0	261.0
Member	2914616.0	36.385330	10.902478	6.0	28.0	33.0	42.0	121.0

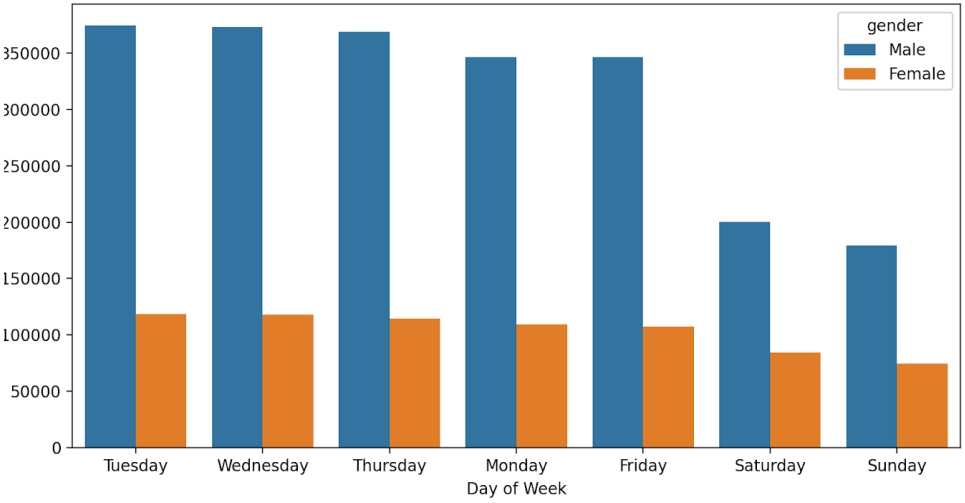
3. Usage Patterns by Day of the Week

Peak Usage Days: ***Casual riders showed a preference for using the service on weekends***, with a notable spike in ridership on Saturdays and Sundays. In contrast, annual members demonstrated more consistent usage throughout the week, with higher ridership on weekdays.

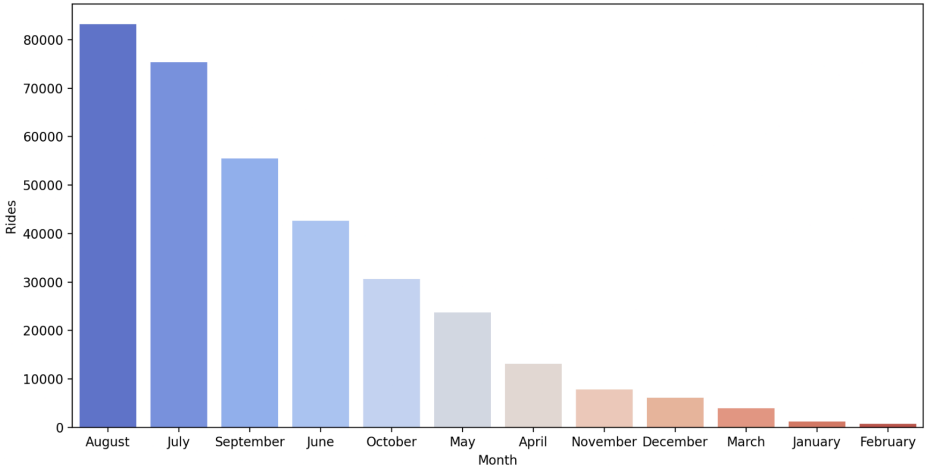
Hourly Trends: The analysis of hourly usage patterns indicated that ***casual riders tended to ride during the late morning and early afternoon***, while annual members showed peak usage during commuting hours (morning and evening).



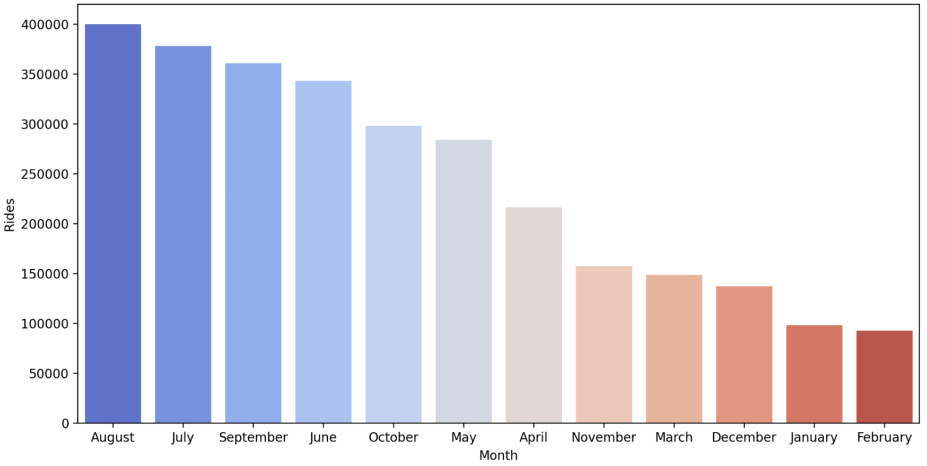
Casual



Members



Casual



Members

Recommendations

Based on the insights derived from the analysis of the bike-sharing dataset, the following recommendations are proposed to enhance user engagement, and optimize service offerings:

1. Weekend/ Holiday Packages

Most of the casual users took the ride on weekends and holidays. ***Providing an annual package that has plans dedicated to weekends and holidays*** can convert potential casual users into members.

2. Personalized Offers

Casual riders having higher average riding times indicate they are potential customers to be targeted by ***offering limited-time discounts on annual memberships or loyalty programs that reward frequent usage.***

3. Increase Bike Availability

Implement strategies to ensure a ***higher availability of bikes during peak usage times, particularly on weekends and holidays***, to accommodate the demand from casual riders.