Understanding Casual and Annual Rider Behavior at Cyclistic Bike-Share
Looking at how people use things to help get more members

1. Overview

In this analysis, we aim to explore the behavioral distinctions between two key Cyclistic Bike-Share user groups: casual riders and annual members. By diving into Cyclistic's historical bike trip data, our goal is to uncover patterns, preferences, and usage behaviors that differentiate these two segments. Utilizing data science methodologies such as data cleaning, exploratory analysis, and visualization, we intend to extract meaningful insights about their interaction with the service.

The driving force behind this research is Cyclistic's strategic priority: increasing the number of annual memberships. Since annual members contribute more significantly to Cyclistic's profitability, converting casual riders into loyal annual members represents a crucial opportunity for growth. Lily Moreno, our Marketing Director, believes that an impactful marketing campaign tailored to casual riders could be the catalyst for this transformation.

This study is designed to support the development of such a campaign by offering data-backed insights that align with the preferences and behaviors of casual riders. By addressing the needs of this user segment, we aim to craft strategies that inspire them to commit to annual memberships. We believe that a deeper understanding of their habits will empower Cyclistic to deliver more targeted and persuasive marketing efforts.

We are eager to share our findings with the Cyclistic team and hope our work will serve as a foundation for shaping future marketing initiatives. Ultimately, this analysis is a step toward advancing Cyclistic's mission of creating a bike-sharing service that is inclusive, accessible, and financially sustainable.

- **Company Background**

  Cyclistic is a bike-sharing company in Chicago, offering a fleet of bicycles for public use across the city. The program was created by Cyclistic has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system a time. The company is working towards a large bike-sharing community, targeting regular users by relying on the flexibility of the fares issued.

- **Data Context**

  We looked at a whole year of bike rides from Cyclistic to see how people use the bikes. We have important information like when each ride started and ended, which stations people started and finished at, what type of bike they rode, and whether they are someone who rides bikes just for fun or someone who pays for membership all year long.

- **Problem Statement**

  We want to learn how people use Cyclistic bikes differently. Some people ride the bikes a lot and are members all year, while others only ride them sometimes. By understanding how they each use the bikes, we can come up with a new plan to encourage the people who ride sometimes to become year-round members!

- **Business Task**

  We want to look at how people who ride bikes just for fun and people who have a membership to ride bikes all the time use Cyclistic bikes in different ways. By understanding these differences, we can come up with a new plan to encourage the fun riders to become members so they can ride bikes more often.

- **Key Stakeholders**
  + Lily Moreno: The director of marketing who is responsible for developing campaigns and initiatives to promote the bike-sharing program.
  + Cyclistic marketing analytics team: This team is responsible for collecting, analyzing, and reporting data that guides Cyclistic's marketing strategy.
  + Cyclistic's executive team: This detail-oriented executive team will decide whether to approve the recommended marketing program.

- **Key Questions (3)**
  + How do annual members and casual riders use Cyclistic bikes differently?
  + Why would casual riders buy Cyclistic annual memberships?
  + How can Cyclistic use digital media to influence casual riders to become members?

3. Case Study Roadmap - 'Prepare'

- **Data Location and Organization:**
  Motivate International Inc. has supplied the data under this [license](#). I will utilize this publicly available data to investigate how various consumer categories use Cyclistic bikes. It is made up of monthly files that span 12 consecutive months. The information may be in an organized format, such as Excel or CSV. There are thirteen columns about bike trips in each file in the dataset. The ride ID, rideable type, start and end station IDs and locations, coordinates, and membership type are among the comprehensive details about each ride that are provided in these columns.

- **Data Credibility and Bias:**
  We must know that the data may contain biases and whether it is good by determining ROCCC (Reliable, Original, Comprehensive, Current, Cited). Since the data is provided by the bike sharing service operator and has a certificate and citation, it is reliable. This data source is collected from 2022, it has been 2 years since I did this capstone project in 2024, the data has been updated to Oct 8th 2024. This shows that the data will be effective in the analysis process and is really good to start a perfect project.

- **Data Integrity:**
  Exploratory data analysis (EDA) will be used to find any discrepancies, missing values, outliers, or duplicate entries in order to confirm the integrity of the data.
- **Data Problems:** Missing data, errors format, or bias data, v.v
- **Dataset:** [divvy-tripdata](#)
- **Key tasks:**
  + Download and store dataset: I download the data from the provided link and store it in a local and Google Drive location.
  + Explore the dataset to understand its structure
  + Process and clean the data based on variables that are significant to my analysis. Remove redundant and erroneous data (missing data, data with incorrect format, etc.)
  + Determine the credibility of the data

4. Case Study Roadmap - 'Process'
- Data Cleaning in Python Language

  + Dropping Irrelevant Columns:
    The columns start_station_name, start_station_id, end_station_name, and end_station_id contain a large amount of missing data (NaN). while calculating the proportion of non-missing and missing data in the four columns above (per the graphic). The

aforementioned columns include over 800,000 thousand missing values. I discovered that start_station_name and end_station_name had no bearing on the data we examine. I thus choose to eliminate these two columns.

+ Checking for Duplicates Data
The duplicate values were examined using the built-in "drop_duplicates" method. This helped ensure data integrity and eliminate any duplicated entries.

+ Missing Data:
Realizing that there was a serious lack of data, I took action to handle the missing data including numeric data (filling in the average value) and string data (filling in the common value). However, the two columns 'start_station_name' and 'end_station_name' in particular needed to be handled, I replaced the rare labels with 'Other' values and filled in the remaining missing values as 'Unknown'.

+ Adding the Ride Length Column
A new column named "ride_length" was added to calculate the duration of each ride. The value in the ride_length column was obtained by subtracting the started_at timestamp from the ended_at timestamp.

+ Setting the Time Format
The ride_length column was formatted as "HH:MM:SS". This ensured that the ride duration was presented in a standardized time format (date/time). Create new columns by splitting data from the started_at column into related information like hour, month, year, ride_length_category, and distance_km.

+ Setting the Day of the Week
Create a new column called day_of_week, and calculate the day of the week that each ride started. Format it as General or as a number with no decimals, such as 1 = Sunday and 7 =Saturday.

- Data Transformation in SQL

SQL Server was used for data processing in order to convert data. From October 2023 to September 2024, the data from each month was imported and combined into a single database named "TripData-20232024."The following are the steps in this process:
+ Importing Data: SQL Server was utilized to import the monthly files that included the riding data.

- + Merging Data: The combined data for the full year was put into a new table called "Trip-Data-20232024."The UNION ALL statement was used to combine the data from the tables into the "TripData-20232024" table.
- + Manipulate Data: A new table called "Analyze-TripData-20232024" was created once the data was combined.

The process phase is followed by the link: here

5. Case Study Roadmap - 'Analyze'

5.1. Overview

Cyclistic Bike Share is a bike-sharing service with two main user groups: members and casual users. Power BI data presents information on the number of trips, types of bikes used, trip times, and trends related to user behavior. This report focuses on analyzing data from November 2023 to September 2024 with the following key aspects:

- ➢ Number of trips over time.
- ➢ Bike usage behavior by day of the week.
- ➢ Trip duration classification.
- ➢ Bike types used.
- ➢ The ratio of casual users to members.
- ➢ Most popular starting points.

5.2. Main operations

- Data Organization and Formatting:
  In this phase, it's essential to organize the data in a manner that makes it accessible and convenient for performing various analyses. This involves appropriately formatting data, ensuring consistency across all variables, and creating a unified view of the data.

- Descriptive Analysis:
  We'll conduct descriptive analysis, which helps understand the central tendencies and distribution of the data. This provides an overview of the patterns and trends within the data and can reveal surprising insights.

- Identifying Trends and Relationships:
  In this step, we'll identify key trends, patterns, and relationships between different variables in our data. This may involve looking at correlations between variables, analyzing patterns over time, or identifying factors that influence a particular outcome.

- Relevance to the Business Question:
  This phase of analysis is critical to answering our main business question: how do casual riders and annual members use Cyclistic bikes

differently? The patterns, trends, and relationships we uncover will contribute to our understanding of user behavior.

- Key Tasks:
1. Aggregate the data: Aggregating the data in different ways can reveal new insights and make the data easier to work with. 2. Organize and format the data: Proper organization and formatting of the data is essential for effective analysis. 3. Perform calculations: Calculations can help us understand the data better and can form the basis for our insights. 4. Identify trends and relationships: Identifying key trends and relationships in the data is a critical part of the analysis process.

5.3. Detailed analysis

- Total number of trips per month
The monthly total trips chart shows a clear fluctuation in the number of trips: The month with the highest number of trips: Sep 2024 with 0.63 million trips. The month with the lowest number of trips: Jan 2024 with 0.13 million trips.
In the spring and summer (from May to September), the number of trips increases sharply, reflecting the trend of users increasing outdoor activities during favorable weather.

- Number of trips by day of the week
The highest number of trips is on Thursday and Saturday, with 0.73 million trips per day. The lowest number of trips is on Monday, with 0.64 million trips.
This suggests that mid-week trips may be related to the need to travel to work or school (the "member" user group). On weekends, the demand for bicycle use increases, especially from the "casual" group for recreational purposes.

- Trip time classification
Trip duration is divided into 7 groups, of which:
    The < 10-minute group accounts for the highest proportion with 2,414,450 trips (50.3% of total trips).
    The 10-20 minute group takes second place with 1,371,239 trips.
    Trips longer than 60 minutes account for only 140,121 trips, a very small proportion.

    This analysis shows that the majority of users use the service for short trips, suitable for their travel needs within the city.

- Type of vehicle used
Number of trips by vehicle type:

Classic bike: 2.8 million trips.
Electric bike: 1.9 million trips.
Electric scooter: 0.1 million trips.
Classic bicycles are still the most used vehicle, accounting for a higher proportion than both electric bikes and scooters.

- Ratio of casual and member users
User Percentage: Casual: 50.4% and Member: 49.6%.
The two user groups have almost equal proportions, indicating that both groups contribute significantly to the total number of trips.

- Most Popular Departures
The departures with the most trips: Streeter Dr & Grand Ave: 64,000 trips. and DuSable Lake Shore Dr & Monroe St: 43,000 trips.
These are high-traffic locations, likely located near tourist areas or downtown.

5.4. Conclusion
Based on the analysis data, we draw the following conclusions:

- Seasonality has a significant impact on the number of trips: Summer months have a significantly higher number of trips than winter.
- Casual and member groups have different behaviors:
Casual groups are more active on weekends.
Member groups tend to use the service on weekdays, which is suitable for commuting to work.
- Bicycles are usually the most popular means of transport, but electric bicycles are gradually becoming popular.
- Short trips account for the largest proportion, indicating that the service mainly meets the need for quick travel within the city.

6. Case Study Roadmap - 'Share'
6.1. Overview
- Communication of Insights:
In this phase, our key task is to communicate our insights to the intended audience. This involves creating clear and impactful visualizations that can illustrate the key trends and patterns we have discovered in our analysis.

- Data Visualization:
Data visualizations are a powerful way to communicate our findings. They enable us to present complex data in an accessible and engaging manner. Our aim is to create visualizations that are clear, compelling, and accurately represent our data.

- **Presentation:**
  The presentation of our findings should be tailored to our audience. We need to consider what information is most relevant to them, how to present it in a way that is easy to understand, and how to highlight the key insights that we want them to take away from our presentation.

- **Accessibility:**
  Accessibility is a key consideration in the sharing phase. We need to ensure that our visualizations and presentation are accessible to all members of our audience. This means considering factors such as color contrast, font size, and clarity of language.

- **Key Tasks:**
  1. Determine the best way to share your findings: Think about your audience and the best way to communicate your insights to them. 2. Create effective data visualizations: Use appropriate visualization techniques to present your data in a clear and compelling way. 3. Present your findings: Craft a narrative that links your findings to the business question and highlights the key insights. 4. Ensure your work is accessible: Make sure that your visualizations and presentation are accessible to all members of your audience.

6.2. Final Summary and Key Insights for Cyclistic's Conversion Strategy

Cyclistic Bike Share is a notable bike-sharing service with two main target groups: casual and member. Through data analysis from November 2023 to September 2024, we have discovered the service usage trends, behavioral characteristics of each user group, and seasonal demand differences. This report not only provides a comprehensive view of how the service is used, but also provides a foundation for proposing appropriate business strategies.

- **Distinct Usage Patterns**
  Different usage behavior between casual and member groups:
  + **Casual group:**
    Use the service mainly on weekends (Thursday to Saturday), with short trips (<10 minutes) accounting for the highest proportion.
    The main purpose of this group is often entertainment or sightseeing, which is shown by their high proportion at stations near tourist areas.
  + **Member group:**
    Use the service more on weekdays, especially during morning and afternoon rush hours. These can be trips for daily travel, such as going to work or school. They tend to use shorter and more frequent trips than the casual group.
  + **Preferred type of bike:**

Classic bikes are still the top choice (2.8 million trips), but electric bikes (1.9 million trips) are gradually becoming a competitive option, especially for longer trips.

- Seasonal Trends
    + **Peak season:** From May to September (spring and summer), the number of trips increases sharply, especially in July 2024 (0.63 million trips). This shows that favorable weather plays an important role in promoting demand for the service.
    + **Low season:** In winter (January 2024), the number of trips decreases significantly, reaching only 0.13 million trips.
    + **Insight:** The service is seasonal in nature, requiring different strategies to optimize operations in each season. This also shows the popularity of the service for entertainment purposes during the weekend.

- **Preferred Stations and Geographical Preferences**
    + The casual group tends to use services at stations near tourist attractions, parks, or lakeside areas.
    + The member group often focuses on stations near residential areas or office buildings, serving daily travel needs.

7. Case Study Roadmap - 'Act'

- Action on Insights
  Based on the analysis of Cyclistic Bike Share data, we identified key insights into user behavior, seasonal trends in service usage, and differences between casual and member groups. To leverage this information, specific actions were proposed, focusing on:

    ★ Optimizing services during low season.
    ★ Improving the experience and increasing the loyalty of each user group.
    ★ Investing in potential areas and suitable vehicles.

- Application of Insights
  Data from November 2023 to September 2024 show the following trends:

    + Seasonality: Trips peak in summer (July 2024) and decline sharply in winter (January 2024).
    + Trip duration: Most trips last less than 10 minutes, suitable for short-distance travel in urban areas.

+ Vehicle type: Classic bikes are the most popular, but electric bikes are increasingly popular.
+ Different usage behavior:
  Casual groups focus on weekends and tourist areas.
  Member groups use the service mainly on weekdays and during peak hours.

- Recommendations
  1. Promote and enhance service experience during low season (winter)
     + Action:
       ➔ Implement a discount or special offer for both user groups during winter.
       ➔ Develop a promotional campaign targeting the casual group, focusing on outdoor activities or winter travel.
       ➔ Improve vehicle maintenance and introduce safety initiatives in winter weather conditions.
     + Objective: Increase the number of trips during low season, and reduce revenue differences between seasons.

  2. Develop a personalized strategy for each user group
     + Casual group:
       ➔ Increase service at tourist attractions, parks, or recreational areas on weekends.
       ➔ Create flexible rental packages (by the hour, by the day) to suit recreational purposes.
     + Member group:
       ➔ Offer preferential membership packages, discounts when renewing.
       ➔ Upgrade service quality at stations near office areas or residential areas, especially during peak hours.

  3. Invest in e-bikes and appropriate infrastructure
     + Action:
       ➔ Expand the number of e-bikes in high-demand areas, especially long routes or complex terrain.
       ➔ Improve station infrastructure, ensuring adequate space for e-bike parking and charging.
     + Goal: Meet growing user demand and enhance service experience.

  4. Optimize stations in downtown and tourist areas
     + Action:
       ➔ Upgrade popular stations such as Streeter Dr & Grand Ave and DuSable Lake Shore Dr & Monroe St to accommodate high traffic volumes.
       ➔ Invest more in areas near attractions or lakefront routes to attract casual groups.
     + Goal: Increase service capacity and maximize the potential of high-traffic areas.

8.1. Promote and Enhance the Experience in the Low Season
- Feasibility: High. Promotional and advertising campaigns can be implemented quickly and at low cost. However, increasing the number of trips in winter may be limited by weather factors.
- Risk: Demand will be difficult to increase sharply if the climate conditions are too harsh.
-
8.2. Develop a personalized strategy by user group
- Feasibility: Medium. Analyzing user behavior and building a suitable service package requires time and resources. However, if implemented effectively, this will increase customer loyalty.
- Risk: Inconsistency in strategy can cause confusion or not achieve the expected results.
-
8.3. Invest in electric bicycles and appropriate infrastructure
- Feasibility: High. The expansion of electric bicycles is an inevitable trend, especially when the demand for use is high. However, the initial investment costs for vehicles and infrastructure (such as charging stations) will be quite high.
- Risks: Without optimal planning, slow or ineffective capital recovery may occur.

8.4. Optimize stations in central and tourist areas
- Feasibility: High. This is an easy-to-implement method and brings immediate results, as these stations already have high traffic volumes.
- Risks: Additional investment in operating and maintenance costs is required to ensure that the stations operate well during peak seasons.