Case study: How does a bike-share navigate speedy success?

Scenario

You are a junior data analyst working on the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

**Characters and teams**

* **Cyclistic:** A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheel bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use the bikes to commute to work each day.
* **Lily Moreno:** The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.
* **Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic’s mission and business goals-as well as how you, as a junior data analyst, can help Cyclistic achieve them.
* **Cyclistic executive team:** The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

**About the company**

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program 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 anytime.

Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however the team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

# Ask

* **Three questions:**

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?

* **Business task**

As a data analyst, my primary responsibility lies in exploring and analyzing the behavioral patterns of both casual riders and member riders. This in-depth analysis is crucial for attracting and converting casual riders into members, thereby driving membership growth for the company.

* **Key stakeholders**
  + Lily Moreno: The director of marketing.
  + Cyclistic marketing analytics team.
  + Cyclistic executive team.

# Prepare

* **Where is your data located?**

The data is located at https://divvy-tripdata.s3.amazonaws.com/index.html, specifically in the files corresponding to the year 2023. It is a public dataset provided by Motivate International Inc.

* **How is the data organized?**

The data for the year 2023 is organized monthly-wise in the above-mentioned location, with filenames following a format such as 202301-divvy-tripdata.zip for January 2023, 202302-divvy-tripdata.zip for February 2023, and so on.

* **Are there issues with bias or credibility in this data? Does your data ROCCC?**

Since the data is collected by the company itself as a first party, its credibility is high, and there is a low chance of bias. This dataset is reliable, original, comprehensive, current, and cited, therefore it has ROCCC (Reliable, Original, Comprehensive, Current, Cited, and Clear).

* **How are you addressing licensing, privacy, security, and accessibility?**

The data has been made available by Motivate International Inc. under the data license agreement provided at https://ride.divvybikes.com/data-license-agreement. As it is a public dataset, it is accessible to anyone for analysis purposes. Additionally, the dataset doesn't contain any personal data, ensuring riders' privacy.

* **How did you verify the data’s integrity?**

The dataset for the year 2023 used in this analysis has been checked for consistency, ensuring that column names and data types are consistent throughout.

* **How does it help you answer your question?**

The dataset for 2023 contains important data fields such as started\_at, ended\_at, started\_station, and member\_casual, which are useful for identifying bike riders' behaviors and patterns. Additionally, the latitude and longitude columns for stations aid in accurately visualizing station locations.

* **Are there any problems with the data?**

One potential issue is the lack of detailed location data for bike stations, such as addresses or postcodes. Additionally, there may be limitations in rider data, such as the inability to determine the number of unique riders.

* **A description of all data sources used:**

The data source consists of 12 .csv files for each month in the year 2023.

# Process

* **What tools are you choosing and why?**

I chose to use SQL on the PostgreSQL database management system. SQL is well-suited for handling large datasets efficiently, and PostgreSQL provides powerful features for data manipulation and analysis. It allows me to perform complex queries and transformations directly on the database, which can be more efficient than using spreadsheets or desktop tools like R for large datasets.

* **Have you ensured your data’s integrity?**

Yes, I have ensured the data's integrity by performing various checks and validations. I have examined the columns and their data types after any changes or manipulations to ensure consistency. PostgreSQL's data integrity constraints, such as NOT NULL and foreign key constraints, also help maintain data integrity at the database level.

* **What steps have you taken to ensure that your data is clean?**

To ensure the data is clean, I took the following steps:

* Removed rows with ride\_length < 1 minute.
* Deleted rows with missing end station for classic bikes and docked bikes.
* Performed data type conversions and formatting as needed.
* Verified the consistency and correctness of the data after each transformation.
* **How can you verify that your data is clean and ready to analyze?**

I can verify that the data is clean and ready for analysis by running queries to check for any inconsistencies or anomalies in the dataset. Additionally, I can examine the data visually or generate summary statistics to ensure that it meets the requirements for analysis.

* **Have you documented your cleaning process so you can review and share those results?**

Yes, I have documented the cleaning process, including the steps taken and any transformations applied, along with relevant SQL queries used for cleaning the dataset. This documentation allows for easy review and sharing of the cleaning results with others involved in the analysis process.

* **Load** **data**

A screenshot of a computer code

Description automatically generated

* **Combine all monthly data sets to one data set for the whole year**

A screenshot of a computer

Description automatically generated

* **Create ride length column to calculate the length of each ride**

A white screen with black text

Description automatically generated

* **Calculate the day of week that each ride started**

A computer screen shot of a computer code

Description automatically generated

* **Delete any rides that last less than 1 minute**

A screenshot of a computer code

Description automatically generated

* **Delete rides without ending stations for classic and docked bikes**

A close-up of a white background

Description automatically generated

# Analyze

* **Calculate Average, Maximum, and Minimum Ride Length for each member type.**

A screenshot of a computer

Description automatically generated

In 2023, average trip duration for casual riders is 21 minutes and for member riders is 12 minutes. The maximum trip for casual riders is 8 days 16 minutes and for member riders is 1 days 57 minutes.

* Number of rides by member type and day of week.

A close up of text

Description automatically generated

A graph of a number of people

Description automatically generated with medium confidence

Casual riders tend to use bikes extensively on weekends, while members rely on bike services for their daily commuting needs throughout the entire week.

* **Popular bike seasons**

A graph of a bar chart

Description automatically generated with medium confidence

The data reveals a notable surge in bike usage among both member groups during the summer months of June, July, and August. This trend aligns with the typical increase in bike usage for recreational purposes during the warmer summer season.

* **Popular bike type**

A graph of a bar

Description automatically generated with medium confidence

It is evident that e-bikes are favored by casual riders. However, among member riders, while both types of bikes are utilized similarly, there is a slight preference for classic bikes over e-bikes.

* **Popular Bike Location**

A graph of a bar graph

Description automatically generated

There are 10 popular start stations and Streeter Dr & Grand Ave is the most popular starting location as it is an important street located in the Near North Side area of Chicago. It runs parallel to the Lake Michigan shoreline, creating a beautiful setting and a favorite spot for strolling, enjoying the scenery, and suburban activities. Streeter Drive is also a road adjacent to a number of high-end residential areas and entertainment facilities.

# Act

* Offer Weekend-Friendly Memberships: Introduce flexible membership packages tailored for weekend use, enticing riders with special rates or perks for weekend rentals. Accompany this with weekend-exclusive entertainment events to attract new members.
* Target Casual Riders with Strategic Marketing: Implement targeted marketing campaigns at high-traffic bike stations, showcasing the advantages of membership. Highlight benefits such as convenience, cost savings, and access to premium bikes.
* Summer Promotions for Members: Roll out summer promotions specifically designed for members, including discounted bike rentals and rewards for frequent summer riders. This can incentivize membership sign-ups and boost overall usage during the warmer months.
* Emphasize Membership Benefits: Clearly communicate the advantages of membership, such as reduced rental fees and occasional free usage days. Emphasize how joining as a member not only saves costs but also enhances the overall biking experience.