

Cyclistic Bike-Share Case Study

Case Background

I am a junior data analyst working in 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, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve my recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Business Background

Business Model:

- Bikes are geotracked and can be unlocked from one station and returned to any other station in the system at any time.
- Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.
- The flexible pricing plans include single-ride passes, full-day passes, and annual memberships.

Product Information:

- A fleet of 5,824 bicycles
- 692 stations across Chicago
- Different bikes offered include traditional bicycles, reclining bikes, hand tricycles, and cargo bikes
- More than 50% of riders opt for traditional bikes
- 8% of riders use the assistive options
- Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work

Ask Phase

Guiding Questions:

1. What is the problem I am trying to solve?
 - a. The problem I am trying to solve is how casual riders and annual members use Cyclistic bikes differently.
2. How can my insights drive business decisions?
 - a. My insights can inform stakeholders how casual riders and annual members use Cyclistic bikes differently to design better marketing strategies to convert casual riders into annual members.

Key Tasks:

1. Identify the business task.
 - a. This key task is written out in the deliverable section for this phase.
2. Consider key stakeholders.
 - a. Lily Moreno: The director of marketing and my manager. She is responsible for the development of campaigns and initiatives to promote the bike share program.
 - b. Cyclistic executive team: Detail-oriented executive team that will decide whether to approve the recommended marketing program.

Deliverable:

1. A clear statement of the business task.
 - a. Analyze how annual members and casual riders use Cyclistic bikes differently to recommend marketing strategies for increasing the number of annual members.

Prepare Phase

Guiding Questions:

1. Where is my data located?
 - a. Cyclistic's historical trip data is located and downloaded from the link provided [here](#).
2. How is the data organized?
 - a. This guiding question's answer is written out in the Key Tasks section for this phase.
3. Are there issues with bias or credibility in this data?
 - a. This guiding question's answer is written out in the Key Tasks section for this phase.
4. How am I addressing licensing, privacy, security, and accessibility?
 - a. The data has been made available by Motivative International Inc. under this [license](#).
The privacy of the data is confirmed by anonymizing any personally identifiable information. The data is located in a cloud repository (AWS), so it is secured through their services. The data is accessible to the public with minimal constraints (no authorization needed).
5. How did I verify the data's integrity?
 - a. The general accuracy and consistency of the data are not compromised because it was collected through a trustworthy source (Divvy powered by Lyft).
6. How does it help me answer my question?
 - a. Analyzing the historical data about casual riders and annual members will help me notice patterns that could help with predicting future trends. In addition, it could indicate how casual riders and members use Cyclistic differently.

Key Tasks:

1. Download data and store it appropriately.
 - a. I have downloaded the previous 12 months of Cyclistic trip data and stored it in a dedicated folder. I have also created a copy of all this data in case I run into any issues and need to restart the data manipulation and cleaning.
2. Identify how it's organized.
 - a. The data is organized monthly. Each month has a CSV file and all the headers for each CSV file are the same.
3. Determine the credibility of the data.

- a. The data is not completely reliable because there were minimal restraints to ensure its validity. A small portion of the data is incomplete (missing values), and has some logical errors such as having a starting time after the ending time.

Deliverable:

- 1. A description of all data sources used.
 - a. The past 12 months of historical trip data from December 2020 to November 2021.

Process Phase

Guiding Questions:

1. What tools am I choosing to use and why?
 - a. Excel to explore the data quickly and understand the structure of the worksheets.
 - b. SQL to perform data cleaning, data manipulation, and data analysis.
 - c. Tableau to create data visualizations for stakeholders to quickly understand my findings.
2. Have I ensured my data's integrity?
 - a. The data's integrity is not as good as it could be because some of the data is not accurate, complete, and consistent. I fixed these issues by transforming the data to restore its integrity.
3. What steps have I taken to ensure that my data is clean?
 - a. I have fixed outdated data, incomplete data, incorrect/inaccurate data, and inconsistent data, and checked to see if there is duplicate data.
4. How can I verify that my data is clean and ready to analyze?
 - a. I can verify that the data is free of duplicates, the date falls within the appropriate range, correct incomplete and inaccurate data, and correct inconsistencies within the data.
5. Have I documented my cleaning process so I can review and share these results?
 - a. Yes, I have created a separate document for the data cleaning process and that can be accessed [here](#).

Key Tasks:

1. Check the data for errors.
 - a. I have checked for errors in the data, and this can be seen in the documentation of the cleaning process.
2. Choose your tools.
 - a. Excel, SQL, and Tableau.
3. Transform the data so I can work with it effectively.
 - a. All the monthly files for Cyclistic are formatted the same so I combined them into one file to manipulate and cleanse the data more efficiently.
4. Document the cleaning process.
 - a. This key task is written out in the deliverable section for this phase.

Deliverable:

1. Documentation of any cleaning or manipulation of data.
 - a. The documentation for cleaning and manipulating the data can be found [here](#).

Analyze Phase

Guiding Questions:

1. How should I organize my data to perform analysis on it?
 - a. I organized the data by converting the twelve separate monthly files into one large file. This allowed me to clean, manipulate, and analyze the data more efficiently.
2. Has my data been properly formatted?
 - a. Yes, the data has been properly formatted after data cleaning and performing data manipulations.
3. What surprises did I discover in the data?
 - a. The average ride duration and the number of rides did not have a positive relationship. The casual riders had longer ride durations, but the members used the bikes more often.
4. What trends or relationships did I find in the data?
 - a. Members would use the bikes the most at 8 am and 5 pm every weekday. Also, the only time casuals would use the bikes more than members was during the weekends.
5. How will these insights help answer my business question?
 - a. These insights can help inform stakeholders how casual riders and annual members use Cyclic bikes differently to design better marketing strategies to convert casual riders into annual members.

Key Tasks:

1. Aggregate my data so it's useful and accessible.
 - a. The data has been aggregated into one file for better accessibility and easier use.
2. Organize and format my data.
 - a. The data has been properly formatted and organized during the Process Phase. This step needed to be done before any analysis could be performed on the data because the data must be reliable to produce reliable results.
3. Perform calculations.
 - a. Some of the calculations performed include the length of the ride, the most popular day of the week for riding, the average ride length, the maximum ride length, and the most popular bike for each day of the week.
4. Identify trends and relationships.
 - a. This key task is written out in the deliverable section for this phase.

Deliverable:

1. A summary of my analysis.
 - a. Casuals dominate in bike rides during the summer season. Members dominate in bike rides in the other three seasons. Casual riders only have more rides than members during the weekend in spring and summer. But casual riders have a longer average ride duration year-round compared to members. Casuals ride most often during the weekend. Whereas members' rides are distributed more evenly throughout the week. Members' number of rides peak year-round during weekday rush hour. On weekends casuals ride more than members in the afternoon except for in winter. Casuals ride longer than members year-round. The casual's number of rides is greater than members' only during the summer season. Casuals have longer ride durations than members for every rideable type. Casuals dominate docked bike usage compared to classic and

electric bikes. Members have similar ride durations for every rideable type. Members are more spread out throughout the city. Casuals ride more by the coast and are not riding much in the city during winter.

Share Phase

Guiding Questions:

1. Was I able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
 - a. I was able to identify certain trends, but I need more data to make a solid conclusion.
2. How do my findings relate to my original question?
 - a. My findings indicate that casuals and members use Cyclistic bikes differently in the past 12 months.
3. Who is my audience? What is the best way to communicate with them?
 - a. My audience includes Lily Moreno: The director of marketing and my manager, and the Cyclistic executive team.
4. Can data visualization help me share my findings?
 - a. Yes. Data visualization is the best way to share my findings.
5. Is my presentation accessible to my audience?
 - a. Yes. The data visualization is easy to understand and colorblind-friendly.

Key Tasks:

1. Determine the best way to share my findings.
 - a. The best way to share my findings is through Tableau.
2. Present my findings.
 - a. My findings are presented in a storyboard format which is easier to understand and highlights the important parts of each section.
3. Ensure my work is accessible.
 - a. My work is accessible to anyone that has the link and is easy to follow for anyone including color blind people.

Deliverable:

1. Supporting visualizations and key findings.
 - a. The Tableau visualizations can be found at this link [here](#).

Act Phase

Guiding Questions:

1. What is my conclusion based on my analysis?
 - a. The 12 months of historical data does indicate that there are some trends between the two types of users, however, more data would be needed for a concrete recommendation.
2. How could my team and business apply my insights?
 - a. They could use the current data to come up with a solution, but more data would be needed for a better recommendation.
3. What next steps would I or my stakeholders take based on my findings?
 - a. Gather more data about Cyclistic bikes if there is enough time or the ability to push the project timeline. Otherwise, use the current recommendations.
4. Is there additional data I could use to expand on my findings?
 - a. Yes. Additional data would be useful to expand on my findings and create a better understanding of how the users differ.

Deliverable:

1. My top three recommendations based on my analysis.
 - a. Partner up with companies in the area. The members' bike rides peak during rush hour on weekdays which indicates that they are mainly using it for their commute to work. Cyclistic could partner up with companies in the area to incentivize their employees to ride to work by giving them a discounted price. Also, the company could create a contract with Cyclistic in which they would pay a large sum of money to allow all their employees to use Cyclistic bikes for free.
 - b. The only time casual riders use Cyclistic more than members is during the summer season and on weekends. Cyclistic could create a summer membership deal that would only last for a few months. This would be more appealing to users that do not want to commit to an annual membership since they are mainly riding for a few months of the year. Also, there could be a weekend-only membership option since casual riders are mainly using the bikes from Friday to Sunday.
 - c. Add additional benefits for members. There should be options and services only available to members. This would incentivize casual users to commit to a membership if there are services that they can use only when they become members.