

Bike Share Case Study Detailed Work

Company – Cyclistic is a bike-share company based in Chicago which offers services as a day to day pay & use basis & also as a monthly subscription plan.

Aim –

- company's future success depends on maximizing the number of annual memberships
- 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

Characters and teams –

1. Cyclistic:
 - Offers A bike-share program that features more than 5,800 bicycles and 692 docking stations
 - Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes
 - Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day
2. Lily Moreno: The director of marketing and your manager
3. Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy
4. 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 geo-tracked and locked into a network of 692 stations across Chicago.
- Customers who purchase single-ride or 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
- 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 very good chance to convert casual riders into members.
- **Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst 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.

Project Objective

Three questions will guide the future marketing program:

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?

Moreno has assigned you the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

Guiding questions-----

- What is the problem you are trying to solve?
-- Getting insights on, how do annual members and casual riders use Cyclistic bikes differently?
- How can your insights drive business decisions?
-- My insights can be used by marketing analyst team who 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.

Key tasks-----

1. Identify the business task
-- Design marketing strategies aimed at converting casual riders into annual members
2. Consider key stakeholders
-- Lily Moreno: The director of marketing and your manager

-- Cyclistic marketing analytics team

Data Preparation

- will be using Cyclistic's historical trip data to analyze and identify trends.
- going to use latest data between January to December 2021.
- The data has been made available by Motivate International Inc. so it is reliable and can be trusted for the analysis.

Guiding questions-----

- Where is your data located?

- Data is located on a server (<https://divvy-tripdata.s3.amazonaws.com/index.html>)

- How is the data organized?

- Data is not organized properly the naming is also confusing so I had to pay extra attention to find the right data.

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

- Data is reliable as it is collected by a credible organization, also data is original as it is a public dataset from a credible organization. Data is almost complete as it covers most of the months also it is comprehensive as it explains itself what is it about.

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

-- Data-privacy issues prohibit you from using riders' personally identifiable information. This means that we are not supposed to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

- How did you verify the data's integrity?

-- Data is almost accurate, complete and reliable for our analysis but it does require some cleaning.

- How does it help you answer your question?

-- It contains the information from which we can find the distance traveled and time interval between each trip so we can find the patterns between them. Also, we can find how do different type of riders use different bikes in different hour, in different week day, in different months & different seasons of the year.

- Are there any problems with the data?

-- yes, it needs some cleaning and organization.

Key tasks-----

I have downloaded the required data and then organized it accordingly then by sorting and filtering I got some idea about the data.

Deliverable—

12 csv files for 12 months data in a separate folder.

Data Processing

Performing required cleaning and preparatory steps on data to get it ready for analysis.

Guiding Questions--

- What tools are you choosing and why?

--Firstly, I used Excel to take a first look at the data and get a sense of it as the data for each month is small it is easier in excel. Then I moved to SQL and python for further analysis and computations. Then I used tableau for data visualization.

- Have you ensured your data's integrity?

-- Yes, I have to check for accuracy, completeness and consistency. As the data is come from a credible organization it is trustworthy.

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

-- First, I checked for data types of each column, then the data range & data format. I also checked for duplicate data, incomplete data, inconsistent data then I cleaned rows with blank data if it is in necessary fields like ride_id, start time & so on. Also, I cleaned the text columns with trim function.

- How can you verify that your data is clean and ready to analyze?

-- I checked for duplicates, removed extra spaces in text columns, removed blanks from each column also in python with the help of data visualization I cleaned the outliers & data accuracy with respect to the information it is representing.

Key tasks---

Checked the data for errors, used required tools at each step, completed required transformation of the data and documented the cleaning process.

Documentation of cleaning and manipulation process---

1. Loaded the data in excel to get the idea about the rows & columns and checked for formatting and proper names
2. Checked for duplicated
3. Cleaned with trim function on text columns

4. Checked for blank cell in each column and removed the complete row
5. Created columns "Trip_duration" & "Day_of_week" as 1 = Sunday and 7 = Saturday.
6. Created column season according to Chicago Months & Seasons--- (Winter Months- December January February, Spring Months- March April May, Summer Months- June July August, Autumn Months- September October November.
7. Loaded the 12 files into one in python

Analysis

Steps--

1. Loaded all the 12 files in python and checked for formatting and names of columns & their sizes to combine them in one data frame
2. Combined the data in one data frame in python
3. Sorted the data according to start date
4. Used info() & head() methods to get an idea about the data
5. Calculated the Distance between each trip in km & Trip Duration in time format
6. Sorted the value according to Distance
7. There was one outlier which was skewing the result so eliminated the row from analysis data for accurate analysis
8. Checked for the rows which contains zero values in coordinate columns and cleaned it as they were giving wrong information about the Distance values
9. Imported the cleaned data to perform analysis in Tableau
10. Saved the required columns from cleaned data set in a separate data frame to perform analysis and save time in process
11. Used "describe () "method to understand the initial statistics of data
12. After that created initial plots and graphs to better understand the distribution
13. Aggregated the data as required and did some analysis on that to find out trends and relationships of different parameters with respect to casual and member riders.