

*Italics* → Notes

Normal → Self-reflection

## *Case Study: How Does a Bike-Share Navigate Speedy Success?*

### *Introduction*

- *Company: Cyclistic, a bike-share company in Chicago*
- *By the end of this lesson, you will have a portfolio-ready case study*

### *Scenario*

- *Me: junior data analyst working in the marketing analyst team*
- *“The company’s future success depends on maximizing the number of annual memberships.” - Director of Marketing*
- *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*
    - ***Cyclistic executives must approve your recommendations, so they **must be backed up with compelling data insights and professional data visualizations*****

### *Characters and teams*

- *Cyclistic*
  - *Bike-share program*
  - *Features more than 5800 bicycles and 600 docking stations*
  - *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-wheeled bike*
  - *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 them to commute to work each day*
- *Lily Moreno*
  - *Director of Marketing and my manager*
  - *Responsible for the development of campaigns and initiatives to promote the bike-share program (email, social media, and other channels)*
- *Cyclistic marketing analytics team*
  - *Team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy*
  - *I joined this team 6 months ago*
- *Cyclistic executive team*
  - *Notoriously detail-oriented*
  - *Will decide whether to approve the recommended program*

### *About the Company*

- *Launched a successful bike-share offering in 2016*
- *5824 bicycles that are geotracked and locked into a network of 692 stations across Chicago*
- *Bikes can be unlocked from one station and returned to any other station in the system anytime*
- *Currently, 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-rides, full-day passes, and annual memberships.*
    - *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*
    - *Moreno believes there is a very good chance 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.*
- *Goal: Design marketing strategies aimed at converting casual users into annual members*
  - *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.*

### **ASK**

Three questions will guide the marketing strategies to be designed aimed at converting casual users into annual members:

- 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?

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

I will produce a report with the following deliverables:

1. A clear statement of the business task
2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendations based on your analysis

The problem I am trying to solve is helping the marketing analyst team better understand how annual members and casual riders differ. My insights will enable Cyclistic to design marketing strategies aimed at converting casual users into annual members.

## PREPARE

### Previous 12 months of Cyclistic trip data

- Datasets are appropriate and will enable you to answer the business questions
- Made available by Motivate International Inc. under this [license](#).

My data is located on this [website](#), licensed by Motivate International Inc. under this [license](#). The data is organized in .csv format, 13 columns each. Number of rows varies for each file.

Reliable? → MED (missing data in some columns/rows)

Original? → HIGH (comes from Divvy's trip data database)

Comprehensive? → HIGH

Current? → HIGH

Cited? → HIGH

There are some problems with the data

- ☐ Certain entries in the data frames have missing data (blank). Will first combine, then remove.
- ☐ Additional columns - like date, day, month, year, weekday, time, and ride\_length need to be created to aggregate data.
- ☐ Some trips are less than 1 minute or more than 24 hours.
- ☐ Some rides' trip duration came out as negative and a couple of bikes have been pulled for quality control. So these need to be removed.
- ☐ The purpose of a docked\_bike is unknown. Further research is needed to determine whether to keep the entries or remove them.

I will use the 12 months ranging from March 2021 to February 2022 of Cyclistic's trip data for this analysis.

## PROCESS

I'm going to use R to clean, analyze, and visualize data.

To ensure that the data is clean:

- ☐ See if the columns for the 12 months are identical.
- ☐ Combine all 12 months into one, big data frame on R.
- ☐ Clean the data frame by removing rows with duplicate/missing/incomplete data

In addition to cleaning the data, I will further aggregate the data by adding the following columns to the data frame:

- ☐ Add a column called "ride\_length"
- ☐ Add a column called "weekday"
- ☐ Add a column called "date"
- ☐ Add a column called "month"
- ☐ Add a column called "day"
- ☐ Add a column called "year"
- ☐ Add a column called "time"

- Details of the data cleaning in R can be found on github titled, "Cyclistic-Capstone.docx"

## ANALYZE

With all 12 datasets merged into one huge data frame and properly formatted through cleaning and aggregation in R, we can now analyze the data.

- What trends or relationships did you find in the data?
  - Trend: Members predominantly use the bikes on weekdays
  - Trend: Casuals predominantly use the bikes on weekends
  - Trend: Bike usage is high in the summer months and low in the winter months.
    - Weather conditions? Events?
    - November 2021 - February 2022
      - Omicron/Delta Variant Surge
  - Shorter ride lengths by members on weekdays (may be connected to using the bike service to travel to work)
  - Longer ride lengths by casuals across the week (something to tap into to increase casuals to members)
  - Classic bikes seem to be the go-to bike for casuals and members (something to tap into to increase casuals to members)
  - For members, ride demands increase between 4 AM to 9 AM, and again from 12 PM - 5:30 PM
    - These times correspond to the typical 9-to-5 working hours (may also include the lunch hour block as well for those who bike to travel somewhere to eat)
  - For casuals, ride demands have a larger interval, starting at 4 PM, reaching its peak at 5:30.

- Demand for bikes for both members and casuals decrease around 5:30 PM  
(something to tap into to increase casuals to members)

**SHARE + ACT**

[Powerpoint](#)