# Case Study 1

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


### Introduction

Welcome to the Cyclistic bike-share analysis case study! In this case study, you will perform many real-world tasks of a junior data analyst. You will work for a fictional company, Cyclistic, and meet different characters and team members. In order to answer the key business questions, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path.

By the end of this lesson, you will have a portfolio-ready case study. Download the packet and reference the details of this case study anytime. Then, when you begin your job hunt, your case study will be a tangible way to demonstrate your knowledge and skills to potential employers.

## Scenario

You are 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, 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-wheeled 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 them 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, 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. 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 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.

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.

## Ask

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?

You 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

Use the following Case Study Roadmap as a guide. Note: Completing this case study within a week is a good goal.

| Case Study Roadmap - Ask |
| --- |
| Guiding questions<br>● What is the problem you are trying to solve?<br>+ The problem I'm trying to solve is to understand how annual members and occasional riders use Cyclistic bikes differently.<br>● How can your insights drive business decisions?<br>+ My information can help stakeholders better understand their customers, which can help them make decisions that could impact the business. |
| Key tasks<br>1. Identify the business task<br>+ The company wants to launch a marketing campaign aimed at occasional customers and wants to know its customers better (occasional customers and annual customers), the objective being to change the occasional to annual subscriptions, annual subscriptions can increase the profit of the company .<br>2. Consider key stakeholders<br>+ The stakeholders are Lily Moreno: the director of marketing and my manager, the rest of the marketing analytics team. |
| Deliverable<br>☐ A clear statement of the business task<br>+ Identify the key differences between casual and member riders and launch a marketing campaign to turn casual riders into member riders. |

Question:

+ What factors might motivate occasional users to purchase an annual subscription?

## Prepare

You will use Cyclistic's historical trip data to analyze and identify trends. Download the previous 12 months of Cyclistic trip data here. (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this license.) This is public data that you can use to explore how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit you from using riders' personally identifiable information. This means that you won't be able 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. Now, prepare your data for analysis using the following Case Study Roadmap as a guide:

| Case Study Roadmap - Prepare |
| --- |
| Guiding questions<br>• Where is your data located?<br>+ The data can be found on Kaggle.<br>• How is the data organized?<br>+ The data is organized in different folders, each folder containing the data of a specific month saved in csv file.<br>• Are there issues with bias or credibility in this data? Does your data ROCCC?<br>+ This data does not present problems of bias or credibility, as it comes directly from the company's customers. |

This data is also ROCCC too.

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

  + The data set is licensed to the company and does not contain any personally identifiable information about the cycliste.

- How did you verify the data's integrity?

  + To verify the integrity of the data, I made sure that they were accurate, complete, and of high quality.

  I also verified that they did not have too many errors and that they were well protected against cybersecurity threats.

- How does it help you answer your question?

  we can obtain essential information about the cyclist.

- Are there any problems with the data?

Key tasks

1. Download data and store it appropriately.

   Done.

2. Identify how it's organized.

   Done.

3. Sort and filter the data.

   Done

4. Determine the credibility of the data.

   Done

Deliverable

☐ A description of all data sources used

## Process

Then, process your data for analysis using the following Case Study Roadmap as a guide:

| Case Study Roadmap - Process |
| --- |
| Guiding questions<br>● What tools are you choosing and why?<br>The tools I will choose are Excels, and R.<br>● Have you ensured your data's integrity?<br>Yes<br>● What steps have you taken to ensure that your data is clean?<br>● How can you verify that your data is clean and ready to analyze?<br>● Have you documented your cleaning process so you can review and share those results? |
| Key tasks<br>1. Check the data for errors.<br>The data contains errors such as empty cells and duplicate data.<br>2. Choose your tools.<br>I will use R language for this analysis<br>3. Transform the data so you can work with it effectively.<br>4. Document the cleaning process. |

| |
|---|
| In the first time I check for blank sell and for inconsistent value.<br>I sorted et filtered to understand the dataset<br>I created a new column call ride_length, it is a difference between the started_at and ended_at.<br>I delate blank cell.<br>From the started column retrieve the date, the month, the year, the week day , and the hours<br>Create a new data frame without records that have ride length <= zero minute OR > 24 h |
| Deliverable<br>☐ Documentation of any cleaning or manipulation of data |

Follow these steps:

1. Download the previous 12 months of Cyclistic trip data.
2. Unzip the files.
3. Create a folder on your desktop or Drive to house the files. Use appropriate file-naming conventions.
4. Create subfolders for the .CSV file and the .XLS or Sheets file so that you have a copy of the original data. Move the downloaded files to the appropriate subfolder.
5. Follow these instructions for either Excel (a) or Google Sheets (b):
    a. Launch Excel, open each file, and choose to Save As an Excel Workbook file. Put it in the subfolder you created for .XLS files.
    b. Open each .CSV file in Google Sheets and save it to the appropriate subfolder.
6. Open your spreadsheet and create a column called “ride_length.” Calculate the length of each ride by subtracting the column “started_at” from the column “ended_at” (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55.

7. Create a column called “day_of_week,” and calculate the day of the week that each ride started using the “WEEKDAY” command (for example, =WEEKDAY(C2,1)) in each file. Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday.
8. Proceed to the analyze step.

If you like, continue working with the data to better familiarize yourself and perhaps even identify new approaches to answering the business questions.

## Analyze

Now that your data is stored appropriately and has been prepared for analysis, start putting it to work. Use the following Case Study Roadmap as a guide:

| Case Study Roadmap - Analyze |
| --- |
| Guiding questions<br>● How should you organize your data to perform analysis on it?<br>● Has your data been properly formatted?<br>● What surprises did you discover in the data?<br>● What trends or relationships did you find in the data?<br>● How will these insights help answer your business questions? |

Key tasks

1. Aggregate your data so it's useful and accessible.

   Done

2. Organize and format your data.

   Done

3. Perform calculations.

   Done

4. Identify trends and relationships.

   Done

Deliverable

☐ A summary of your analysis

I aggregate the data, organize, perform ;some calculation and identify trends .

I do some calculation and I store it in a new data frame .

## Share

You can find the final R-markdown at the following link.