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 work for a fictional company, Cyclistic, along with some key team members. In order to answer the business questions, follow the steps ofi 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.

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-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 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, and annual memberships. Customers who purchase single-ride or full-day passes are refierred 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 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 reasonable goal.

Case Study Roadmap - Ask
Guiding questions ● What is the problem you are trying to solve? How do annual members and casual riders use Cyclistic bikes differently? ● How can your insights drive business decisions? Must be backed up with compelling data insights and professional data visualizations.
Key tasks ● Identify the business task ● Consider key stakeholders
Deliverable ● A clear statement of the business task Understand what the difference between annual members and casual riders. To Lily Moreno and Cyclistic executive team

Prepare

Use Cyclistic's historical trip data to analyze and identify trends. If you will NOT be using R, download the [previous 12 months of Cyclistic trip data.](#) If you are planning on using Posit's RStudio, use the [Divvy 2019 Q1](#) and [Divvy 2020 Q1](#) datasets. Choosing other data might lead to errors because the data exceeds the memory available in the free plan. (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 • Where is your data located? [previous 12 months of Cyclistic trip data.](#) • How is the data organized? In [Divvy 2019 Q1](#) and [Divvy 2020 Q1](#) datasets. • Are there issues with bias or credibility in this data? [Does your data ROCCC?](#) There are no issues and its Reliable, Original, Comprehensive, Current, and Cited. • How are you addressing licensing, privacy, security, and accessibility? Use the license that I agree with their limitation, ensure privacy of customers and company, ensure the data is secure by using ROCCC, and accessibility depends on the file I want to share and how I want others do on it. • How did you verify the data's integrity? By disallowing replication, transfer, and manipulation and use common format. • How does it help you answer your question? By using ROCCC and addressing licensing, privacy, security, and accessibility • Are there any problems with the data? There are some fields with blanks in gender and birthyear in 2019 and ride_id without any other information in 2020
Key tasks • Download data and store it appropriately. • Identify how it's organized. • Sort and filter the data. • Determine the credibility of the data.
Deliverable • A description of all data sources used Using this source [previous 12 months of Cyclistic trip data.](#) to get [Divvy 2019 Q1](#) and [Divvy 2020 Q1](#) datasets. With ensure of data's integrity

Process

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

Case Study Roadmap - Process
Guiding questions ● What tools are you choosing and why? Using different Excel tools to ensure the data have no error like TIRM, TEXT, and Filter ● Have you ensured your data's integrity? Yes, following Principles of data integrity: Completeness, Validity, Accuracy, and Consistency ● What steps have you taken to ensure that your data is clean? Remove duplicate, Outdated, Incomplete, Incorrect/inaccurate, Inconsistent data, and check for spelling errors, document errors, misfielded values, missing values, business objectives, source of the error, analyzing the system prior to data cleaning, backing up data prior to data cleaning, and accounting for data cleaning in deadlines/process for all data ● How can you verify that your data is clean and ready to analyze? By removing all errors from the data, the data becomes ready for analysis ● Have you documented your cleaning process so you can review and share those results? Yes
Key tasks ● Check the data for errors. ● Choose your tools. ● Transform the data so you can work with it effectively. ● Document the cleaning process.
Deliverable ● Documentation of any cleaning or manipulation of data 1. There are blanks highlighted in dark red in both files 2. Remove any additional space in the data in both files 3. Delete row with no value except two column that not have ID in Divvy_2020_Q1 and 4. Change between started_at and ended_at for number of rows because it reversed 5. Transform the data into new Excel sheet

Follow these steps:

1. Download the previous 12 months of trip data.
   Note: If you are planning on using Posit's RStudio, use the Divvy 2019 Q1 and Divvy 2020 Q1 datasets. Choosing other data might lead to errors because the data exceeds the memory available in the free plan.
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 • How should you organize your data to perform analysis on it? By cleaning, formatting, and structuring data perfectly • Has your data been properly formatted? Yes • What surprises did you discover in the data? There are many relations between the data from two spreadsheets • What trends or relationships did you find in the data? Average length by day of week, Average length by user type, Average length by time of day, and Most start stations and end stations used • How will these insights help answer your business questions? By finding the relationships and trends this help answer business questions
Key tasks • Aggregate your data so it's useful and accessible. • Organize and format your data. • Perform calculations. • Identify trends and relationships.
Deliverable • A summary of your analysis I find there are days of week average length, and most start and end stations used different from casual and member, and both most time use in afternoon, other time, then morning, and number of members grater than casuals

Follow these steps for using spreadsheets

Open your spreadsheet application, then complete the following steps:
1. Where relevant, make columns consistent and combine them into a single worksheet.
2. Clean and transform your data to prepare for analysis.
3. Conduct descriptive analysis.
4. Run a few calculations in one file to get a better sense of the data layout. Options:
   - Calculate the mean of `ride_length`
   - Calculate the max `ride_length`
   - Calculate the mode of `day_of_week`
5. Create a pivot table to quickly calculate and visualize the data. Options:
   - Calculate the average `ride_length` for members and casual riders. Try rows = `member_casual`; Values = Average of `ride_length`.
   - Calculate the average `ride_length` for users by `day_of_week`. Try columns = `day_of_week`; Rows = `member_casual`; Values = Average of `ride_length`.
   - Calculate the number of rides for users by `day_of_week` by adding Count of `trip_id` to Values.

6. Open another file and perform the same descriptive analysis steps. Explore different seasons to make some initial observations.
7. Once you have spent some time working with the individual spreadsheets, merge them into a full-year view. Do this with the tool you have chosen to use to perform your final analysis, either a spreadsheet, a database and SQL, or R Studio.
8. Export a summary file for fiurther analysis.

Follow these steps for using SQL

Open your SQL tool of choice, then complete the following steps:
1. Import your data.
2. Explore your data, perhaps looking at the total number of rows, distinct values, maximum, minimum, or mean values.
3. Where relevant, use JOIN statements to combine your relevant data into one table.
4. Create summary statistics.
5. Investigate interesting trends and save that information to a table.

Follow these steps for using R

Open your preferred version of R, click this link, and select "Use template." Then, copy and paste the text from the template into an R script.
1. Import your data from Divvy 2019 Q1 and Divvy 2020 Q1.
2. Make columns consistent and merge them into a single data frame.
3. Clean up and add data to prepare for analysis.
4. Conduct descriptive analysis.
5. Export a summary file for further analysis.

Share

Now that you have performed your analysis and gained some insights into your data, create visualizations to share your findings. Moreno has reminded you that they should be sophisticated and polished in order to effectively communicate to the executive team. Use the following Case Study Roadmap as a guide:

Case Study Roadmap – Share

Guiding questions
- Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently? By the trips, day, and stations used
- What story does your data tell?
  About the difference between members and casual different by trips, day, and stations
- How do your findings relate to your original question?
  Is related to it and answer the question for this analysis
- Who is your audience? What is the best way to communicate with them?
  The bikers and can communicate with them by advertising at bike stations and if members by number entered when subscribed
- Can data visualization help you share your findings?
  Yes, by simplifying the words and numbers in shapes, so it will be easy to understand
- Is your presentation accessible to your audience?
  Depending on the company policy and privecy

Key tasks
- Determine the best way to share your findings.
- Create effective data visualizations.
- Present your findings.
- Ensure your work is accessible.

Follow these steps:
1. Take out a piece of paper and a pen and sketch some ideas for how you will visualize the data.
2. Once you choose a visual form, open your tool of choice to create your visualization. Use a presentation software, such as PowerPoint or Google Slides, your spreadsheet program, Tableau, or R.
3. Create your data visualization, remembering that contrast should be used to draw your audience's attention to the most important insights. Use artistic principles including size, color, and shape.
4. Ensure clear meaning through the proper use of common elements, such as headlines, subtitles, and labels.
5. Refine your data visualization by applying deep attention to detail.

Act

Now that you have finished creating your visualizations, act on your findings. Prepare the deliverables Morena asked you to create, including the three top recommendations based on your analysis. Use the following Case Study Roadmap as a guide:

Case Study Roadmap – Act
Guiding questions • What is your final conclusion based on your analysis? The days and stations used by both members and casuals is different, so the company need to advertise their services to extend and convert casuals to members • How could your team and business apply your insights? Depending on the analysis graph and final recommendation or findings • What next steps would you or your stakeholders take based on your findings? Advertise, Promotions, and Extends • Is there additional data you could use to expand on your findings? There is ride length can expand findings to average use of bikes
Key tasks • Create your portfolio. • Add your case study. • Practice presenting your case study to a friend or family member.
Deliverable • Your top three recommendations based on your analysis Advertise, Promotions, and Extends at stations and in days that have highest number of casuals

Follow these steps:
1. If you do not have one already, create an online portfolio.
2. Consider how you want to feature your case study in your portfolio.
3. Upload or link your case study findings to your portfolio.
4. Write a brief paragraph describing the case study, your process, and your discoveries.
5. Add the paragraph to introduce your case study in your portfolio.

Wrap-up

Congratulations on finishing the Cyclistic bike-share case study! If you like, complete another case study to continue growing your portfolio. Or, use the steps from the **Ask**, **Prepare**, **Process**, **Analyze**, **Share**, and **Act** Case Study Roadmap to create a new project that's all your own. Best of luck on your job search!