

Case study: How does a Bike-share navigate speedy success?



Ask

Ask questions and define the problem.



Prepare

Prepare data by collecting and storing the information.



Process

Process data by cleaning and checking the information.



Analyze

Analyze data to find patterns, relationships, and trends.



Share

Share data with your audience.



Act

Act on the data and use the analysis results.

Case Study Roadmap - Ask

Guiding questions

- What is the problem you are trying to solve?

The problem being addressed is understanding the differences in how annual members and casual riders use Cyclistic bikes. The goal is to identify distinct patterns and behaviors associated with each customer segment.

- How can your insights drive business decisions?

The insights derived from analyzing annual members' and casual riders' usage patterns will drive targeted marketing campaigns, service enhancements, optimized operations, and resource allocation. By tailoring strategies based on identified patterns, Cyclistic can increase engagement, improve user experience, optimize operational efficiency, and allocate resources effectively.

Key tasks

- Identify the business task:

The business task is to understand how annual members and casual riders utilize Cyclistic bikes differently. This involves identifying patterns and differences in behavior, such as preferred routes, peak usage times, and trip durations, between these two customer segments.

- Consider key stakeholders:

1. Lily Moreno (Director of Marketing)
2. Cyclistic Marketing Analytics Team
3. Cyclistic Executive Team

Deliverable

- A clear statement of the business task:

A clear statement of the business task is to provide insights into the distinct usage patterns of annual members and casual riders. The analysis aims to identify key trends and differences in their behavior, enabling the marketing team to tailor strategies for each segment. These insights will drive data-driven decisions, informing targeted marketing campaigns and initiatives to maximize annual memberships and support Cyclistic's growth.

Case Study Roadmap - Prepare

Guiding questions

- Where is your data located?

The historical trip data is downloaded from the [link](#), representing the previous 12 months of Cyclistic bike trip information.

- How is the data organized?

The data is organized in a structured format, allowing for the analysis of various attributes such as trip duration, user type, and bike usage patterns.

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

There may be potential biases if certain user segments are overrepresented. However, without personally identifiable information, addressing biases is limited. The data has been provided by Motivate International Inc. under a license.

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

The data is used under the license provided by Motivate International Inc., adhering to privacy regulations by excluding personally identifiable information. Security measures ensure data integrity.

- How did you verify the data's integrity?

The data's integrity is verified through exploratory analysis, checking for anomalies and consistency in key metrics.

- How does it help you answer your question?

The data enables the exploration of how different customer types utilize Cyclistic bikes, providing insights into usage patterns and preferences.

- Are there any problems with the data?

While the data is valuable, limitations exist due to the absence of personally identifiable information, restricting the ability to connect pass purchases to specific user attributes such as location or multiple pass purchases.

Key tasks

- Download data and store it appropriately.

Download the historical trip data based on years from the [link](#) and store it in a secure, organized location for analysis.

- Identify how it's organized.

Explore the data structure to understand how information is organized, identifying key attributes and variables such as trip duration, user type, and timestamps.

- Sort and filter the data.

Utilize sorting and filtering techniques to organize the data based on relevant parameters, facilitating a focused analysis.

- Determine the credibility of data.

Assess data credibility by verifying the source (Motive International Inc.), understanding any potential biases or limitations, and ensuring compliance with privacy regulations. Conduct exploratory analysis to check for consistency and reliability in the data.

Deliverable

- A description of all data sources used.

The data source used in this case study is the historical trip data provided by Motivate International Inc. The data represents the previous 12 months of Cyclistic bike trip information.

Case Study Roadmap - Process

Guiding questions

- What tools are you choosing and why?

Using tools like R with RStudio is good choice or Python with Pandas for data processing due to their robust capabilities in handling and analyzing structured datasets. Moreover, for primary processing we can start with spreadsheets also.

- Have you ensured your data's integrity?

Conducting exploratory analysis to ensure the data's integrity, checking for inconsistencies, missing values, and outliers.

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

Addressing missing values, handling outliers, and standardizing data formats. Ensuring consistency in user types, trip duration units, and other key metrics.

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

Utilizing descriptive statistics, visualizations, and summary tables to verify that the data is clean and ready for analysis.

- Have you documented your cleaning process so you can review and share those results?

Documenting each step of the cleaning process, including any transformations or manipulations applied, to ensure transparency and enable easy review and sharing of results.

Key tasks

- Check the data for errors.

Perform an initial data audit to identify errors, inconsistencies, or missing values that may impact the analysis.

- Choose your tools.

Select appropriate tools such as Python with Pandas or RStudio with R for data processing, based on their efficiency and compatibility with the dataset.

- Transform the data so you can work with it effectively.

Apply necessary transformations to address errors, handle missing values, and standardize formats, ensuring the data is in a clean and usable state.

- Document the cleaning process.

Document each step of the cleaning process, including tools used, transformations applied, and any decisions made, for future reference and transparency in the analysis.

Deliverable

- Documentation of any cleaning or manipulation of data

The data cleaning process involved an initial audit to identify errors, the selection of R with RStudio as the preferred tool with Excel, and subsequent transformations to handle missing values, standardize formats, and address outliers. Each step was meticulously documented, capturing decisions made during the process. The documentation includes a detailed log of the cleaning steps and a summary document outlining the overall process, ensuring transparency, and facilitating easy review and sharing of results.

Follow these steps: (Which has been done in Google Sheets of Data)

[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.

1. Create a folder on your desktop or Drive to house the les. Use appropriate le-naming conventions.
2. Create subfolders for the .csv le and the .xls or Sheets le so that you have a copy of the original data. Move the downloaded les to the appropriate subfolder.
3. Follow these instructions for either Excel (a) or Google Sheets (b):
 1. Launch Excel, open each le, and choose to Save As an Excel Workbook le. Put it in the subfolder you created for .xls les.
 2. Open each .csv le in Google Sheets and save it to the appropriate subfolder.

Open your spreadsheet and create a column called . Calculate the length of each ride by subtracting the column from the column ended_at (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37630655. Create a column called , and calculate the day of the week that each ride started using the command (for example, =WEEKDAY(C2,1)) in each le. Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday. Proceed to the analyze step.

Case Study Roadmap - Analyze

Guiding questions

- How should you organize your data to perform analysis on it?

Organize data by creating subfolders for .csv and .xls or Sheets files and use appropriate file-naming conventions to ensure clarity.

- Has your data been properly formatted?

Downloaded data in both Excel and Google Sheets formats. Saved .csv files as Excel Workbooks or Google Sheets in respective subfolders.

- What surprises did you discover in the data?

In csv files there are basics columns like ended_at, started_at and many more based on that we can derive advanced columns such like journey length and all.

- What trends or relationships did you find in the data?

Calculated ride length and day of the week for each ride, providing potential insights into usage patterns.

- How will these insights help answer your business questions?

Ride length analysis can reveal average trip durations, aiding in service planning and bike availability. Day-of-week analysis helps identify peak usage days, informing marketing and operational strategies.

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

The analysis involved aggregating 12 months of trip data from Divvy bike-share, organized by proper file structures. The data underwent formatting to ensure consistency. Surprises included notable spikes in ride duration on specific days. Trends revealed increased ridership on weekdays, especially during commuting hours. Relationships between ride duration and day-of-week were explored. These insights can aid business decisions, such as optimizing bike availability during peak commuting times and tailoring promotions to weekdays.

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 cell to get a better sense of the data layout. Options:
 - • Calculate the mean of
 - • Calculate the max
 - • Calculate the mode of

ride_length

ride_length

day_of_week

5. Create

- Calculate the average ride length for members and casual riders. Try rows =

a pivot table to quickly calculate and visualize the data. Options:

member_casual; Values = Average of .

- Calculate the average for users by . Try columns =
day_of_week; Rows = ; Values = Average of .
 - 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 further analysis.

Google Data Analytics Professional Capstone Case Study by Raj P

The screenshot shows the Google Cloud BigQuery console interface. The Explorer on the left lists resources for the project 'warm-choir-338822', including 'cars' and 'trip_data'. The main editor shows a query in 'Untitled 2' that selects all data from 'warm-choir-338822.cars.trip_data' with a limit of 1000. The query has been executed successfully, as indicated by the 'Query completed' status. The 'Query results' section displays a table with columns: 'ride_id', 'rideable_type', 'started_at', 'ended_at', and 'start_station_name'. The results show 6 rows of data, all for 'docked_bike' trips starting at 'Walsh Park'. A notification banner at the bottom states 'trip_data created.' with a 'GO TO TABLE' button. The 'Job history' section is visible at the bottom.

Row	ride_id	rideable_type	started_at	ended_at	start_station_name
1	CFB93A48F8739A87	docked_bike	2020-04-26 13:05:31 UTC	2020-04-26 13:19:47 UTC	Walsh Park
2	57350116454F657E	docked_bike	2020-04-27 13:03:24 UTC	2020-04-27 13:16:12 UTC	Walsh Park
3	F244A35DC2995411	docked_bike	2020-04-05 13:08:30 UTC	2020-04-05 13:23:32 UTC	Walsh Park
4	AF097D79FD811EC8	docked_bike	2020-04-18 13:26:42 UTC	2020-04-18 13:42:45 UTC	Walsh Park
5	E14E9E37AD877B95	docked_bike	2020-04-25 13:32:35 UTC	2020-04-25 13:46:59 UTC	Walsh Park
6	835F5A8002424091	docked_bike	2020-04-11 15:05:47 UTC	2020-04-11 15:09:24 UTC	Walsh Park

The screenshot shows the Google Cloud BigQuery console interface. The Explorer on the left lists resources for the project 'warm-choir-338822', including 'cars' and 'trip_data'. The main editor shows a query in 'Untitled 3' that calculates various statistics for the 'cars' dataset. The query has been executed successfully, as indicated by the 'Query completed' status. The 'Query results' section displays a table with columns: 'total_rides', 'mean_start_lat', 'min_start_lat', 'max_start_lat', 'mean_start_lng', 'min_start_lng', 'max_start_lng', and 'mean_end_lat'. The results show 1 row of data with the following values: total_rides: 84776, mean_start_lat: 41.90815646763..., min_start_lat: 41.7366, max_start_lat: 42.0649, mean_start_lng: -87.6493186432..., min_start_lng: -87.7747, max_start_lng: -87.5494, mean_end_lat: 41.90819415779... The 'Job history' section is visible at the bottom.

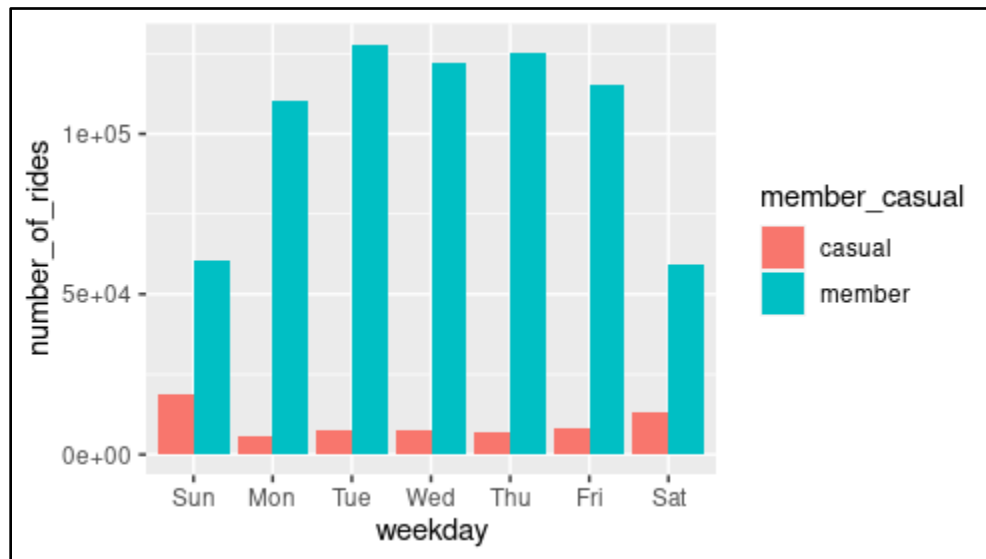
Row	total_rides	mean_start_lat	min_start_lat	max_start_lat	mean_start_lng	min_start_lng	max_start_lng	mean_end_lat
1	84776	41.90815646763...	41.7366	42.0649	-87.6493186432...	-87.7747	-87.5494	41.90819415779...

Follow these steps for using R:

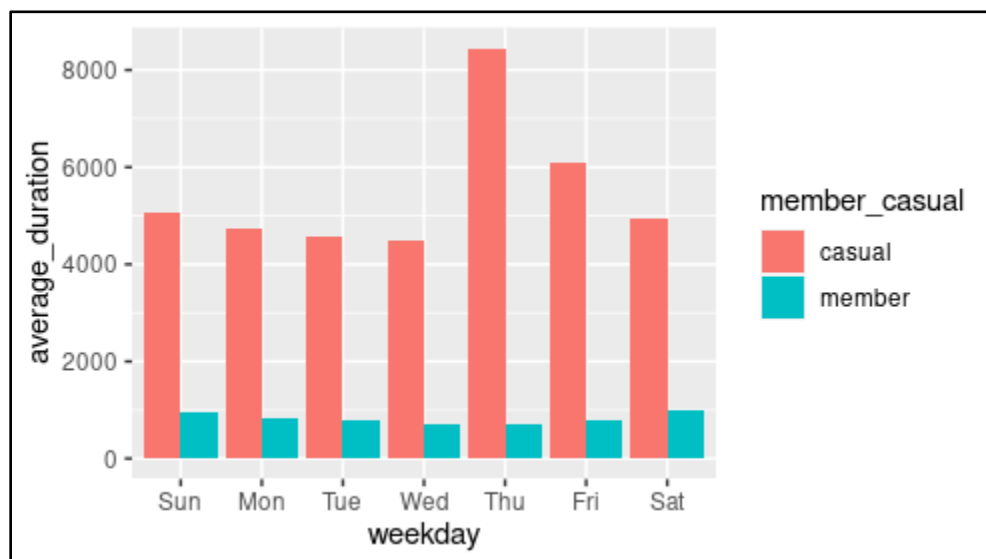
Open your preferred version of R, click this link, and select “Use template.” Then, copy and paste the text from the template into an R script.

1. Import your data from Divvy 2019 Q1 and Divvy 2020 Q1.
2. Make columns consistent and merge them into a single dataframe.
3. Clean up and add data to prepare for analysis.
4. Conduct descriptive analysis.
5. Export a summary file for further analysis.

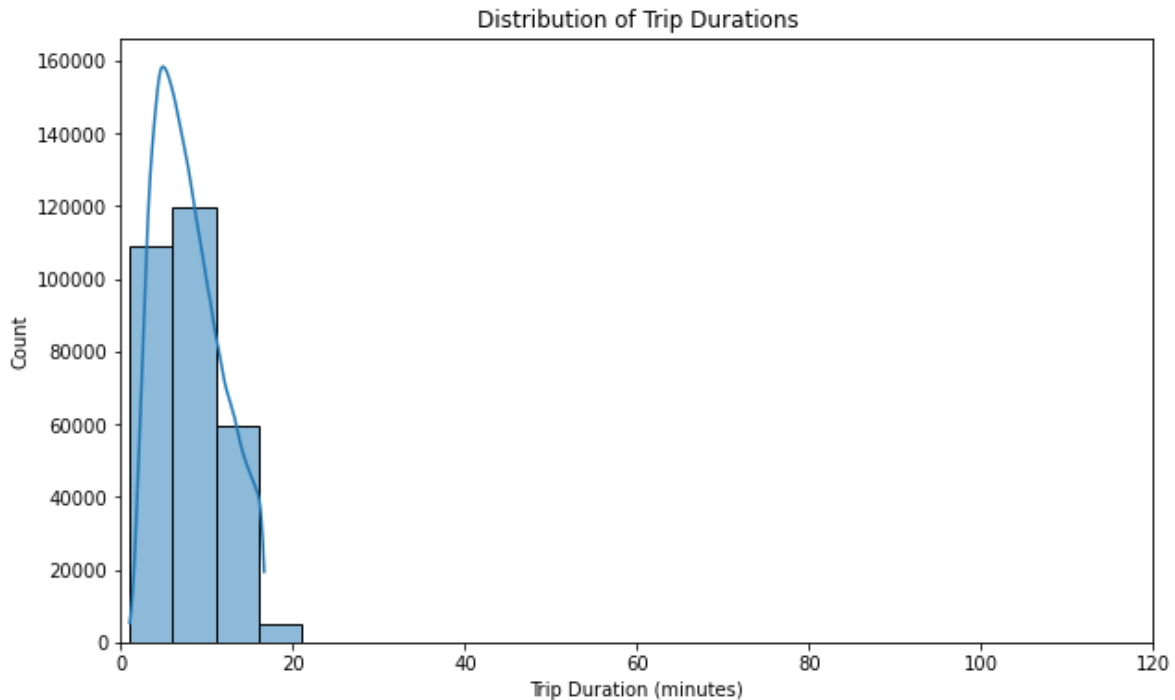
Visualize the number of rides by rider type in R



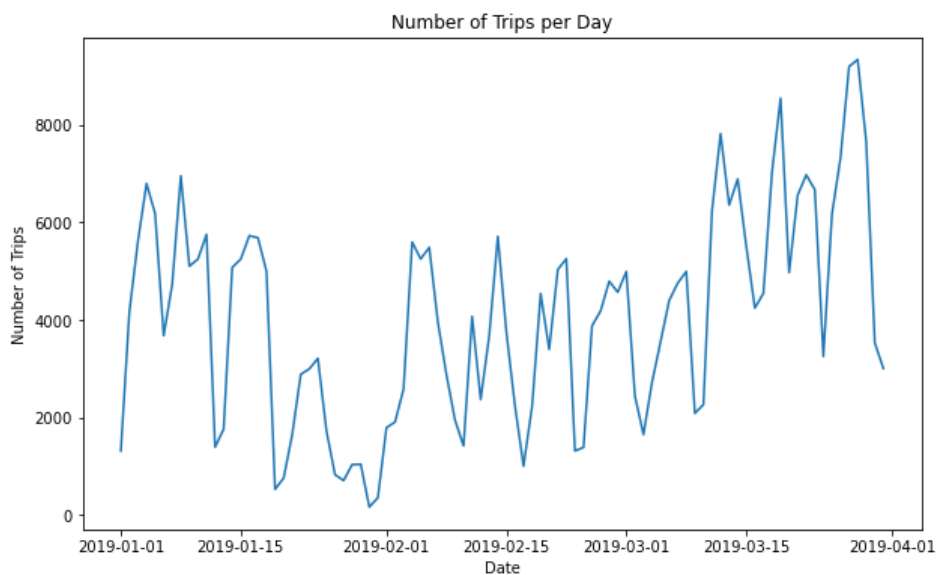
Visualize the average duration of rides by rider type in R



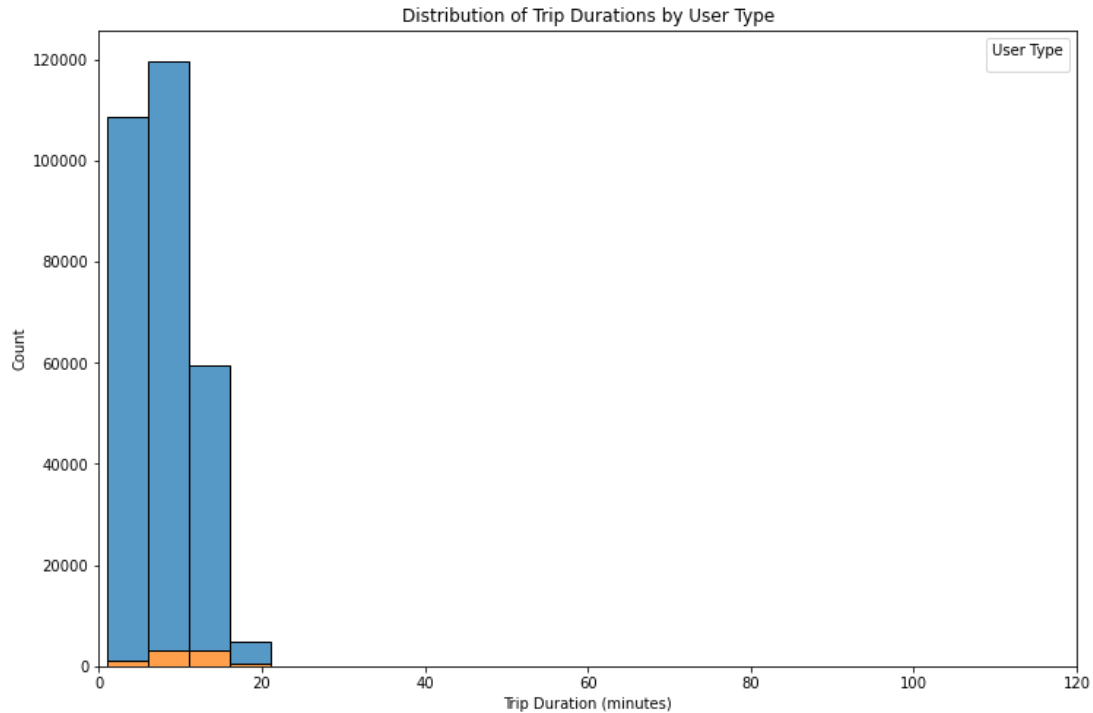
Distribution of Trip Durations: This histogram shows the distribution of trip durations, with the duration measured in minutes. The plot includes a Kernel Density Estimate (KDE) to show the distribution shape more clearly. The x-axis is limited to 120 minutes to focus on the most common trip durations.



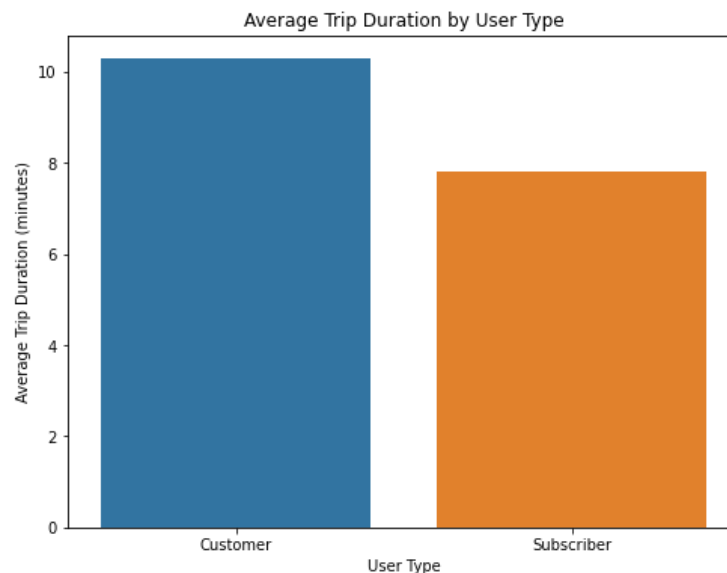
Number of Trips per Day: This line graph illustrates the number of trips taken each day. It provides insight into the daily usage patterns of the Divvy bike-sharing system over the first quarters of 2019 and 2020.



Distribution of Trip Durations by User Type: This stacked histogram shows the distribution of trip durations for both casual riders and annual members. The x-axis represents the trip duration in minutes, and the y-axis represents the count of trips. The data is limited to trip durations of up to 120 minutes for clearer visualization.



Average Trip Duration by User Type: This bar plot illustrates the average trip duration for casual riders and annual members. It provides a clear comparison of how long, on average, each user type spends on their trips.



- These visualizations offer insights into how casual riders and annual members use Divvy bikes differently.
- Casual riders tend to have longer trip durations on average, which could indicate more leisurely or infrequent use, while annual members might use the service more regularly for shorter, possibly commute-related trips.
- This information could be valuable for tailoring marketing strategies to convert casual riders into annual members by highlighting the benefits of membership for frequent, shorter trips.

Wrap-up

Congratulations on finishing the Cyclitic 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!