CASE STUDY: HOW DOES A BIKE-SHARE NAVIGATE SPEEDY SUCCESS?

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 tracked 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 at 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 targeting all-new customers, Moreno believes there is a 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.

Characters and teams

- Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations.
- 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 and 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.

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.

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

Ask

Guiding questions What is the problem you are trying to solve?	**Answer to Guiding Questions** Categorizing things to understand how casual riders and annual members use Cyclistic bikes differently.
How can your insights drive business decisions?	The insight will help the marketing team in designing a new marketing strategy to convert casual riders into annual members.
Key tasks Identify the business task	**Answer to Key task** The business task is to: Understand how casual riders and annual members use cyclistic bikes differently and use this insight to Design a new marketing strategy that will convert casual riders into annual members
Consider key stakeholders	The key stakeholder is The Cyclistic marketing analytics team, Lily Moreno and the Cyclistic executive team.
Deliverable A clear statement of the business task	**Answer to Deliverable** Analysing how casual riders and annual members use cyclistic bikes differently, using this insight to Design a new marketing strategy that is aimed at converting casual riders into annual members. Thereby Maximizing the number of annual memberships.

Prepare

Guiding questions Where is your data located?	**Answer to Guiding Questions** The data is collected by observation and is a third-party data located online.
How is the data organized?	The data is organized in a structured format in a spreadsheet, they are in a folder, organized in a logical hierarchy, and renamed with the existing file name including dates using an underscore. ☐ Created a folder on your desktop or Drive to house the files. Used appropriate file-naming conventions. ☐ Created subfolders for the file cleaned, .CSV file and the .XLS or Sheets file so that I have a copy of the original data. ☐ Move the downloaded files to the appropriate subfolder.
Are there issues with bias or credibility in this data? Does your data ROCCC?	The data is credible and there are no bias issues with it, it is also reliable. However, the data may not be 100% original, but it is comprehensive, current and cited.
How are you addressing licensing, privacy, security, and accessibility?	Privacy is address by Preserving the data subject's privacy or sensitive data, information and activity any time a data transaction occurs, encrypting it with a key to guarantee the security of the data.
How did you verify the data's integrity?	By checking it authenticity, completeness, consistency, accuracy, timeline, transparency and Compliance of the data.
How does it help you answer your question?	The data helps to answer my question by providing the necessary information needed for analyses such as the start date/time, end date/time, id and membership status of each cyclistic bike rider.
Are there any problems with the data	There is no problem with the data.
Key tasks Download data and store it appropriately.	**Answer to Key task** ☐ Created a folder on my desktop/ Drive to house the files. Used appropriate file-naming conventions like underscore, date. ☐ Created subfolders for the file cleaned, .CSV file and the .XLS or

	Sheets file so that I have a copy of the original data. - Move the downloaded files to the appropriate subfolder.
Identify how it's organized.	The data file is named with the alignment of the company name including dates using an underscore (cyclistic_bike_202201). I created a folder and subfolders in a logical hierarchy so related files are stored together. The files are automatically backed up.
Sort and filter the data.	Yes, I filtered and sorted the data
Determine the credibility of the data.	The data is credible, since there are no bias issues with the data, however, the data is reliable, and may not be 100% original, but it is comprehensive, current and cited.
Deliverable A description of all data sources used	**Answer to Deliverable** The data is collected by observation and is a third-party data located online, in a CSV file format, the data is reliable but not original, comprehensive, current and not cited.

Process

Guiding questions What tools are you choosing and why?	**Answer to Guiding Questions** For practice's sake, I will be using google spreadsheets and excel
Have you ensured your data's integrity?	Yes
What steps have you taken to ensure that your data is clean?	- Sorted the sheet using columns. - Filtered each datasheet to check for blank spaces, and whitespaces, remove duplicates, check spellings - Checked length in columns and removed cells not equal to the average length in a column. - Changed the data format for some columns. - Renamed some columns to be more descriptive, hid and deleted some columns.
How can you verify that your data is clean and ready to analyze?	By ensuring the data is well cleaned.

Have you documented your cleaning process so you can review and share those results?	Yes
Key tasks Check the data for errors.	**Answer to Key task** Done
Choose your tools.	I choose google spreadsheets and excel
Transform the data so you can work with it effectively.	The data has been properly transformed.
Document the cleaning process	The cleaning process have been properly documented.
Deliverable Documentation of any cleaning or manipulation of data	☐ Sorted the sheet using first column. ☐ Filtered each datasheet to check for blank spaces, and whitespaces, remove duplicates, check spellings ☐ Checked column ride_id length and removed ride_id that is not equal to the average length. ☐ Renamed columns to be more descriptive: ride_id to trip_id, rideable_type to bikeid and casual_member to usertype, and hid start_date_time, start_station_name, end_date_time, end_station_name. ☐ Added 2 new columns (ride_length and day_of_week). ☐ Changed the data format for column day_of_week to number, column ride_length to time and changed the data format for columns start_date_time and end_date_time to date and time. ☐ Column ride_length is for calculating the length of each ride by subtracting column end_date_time from column start_date_time. Column day_of_week is for calculating the day of the week that each ride started using the "WEEKDAY" command and formatted as general or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday.

Analyze

Guiding questions How should you organize your data to perform analysis on it?	By Sorting and filtering data

Has your data been properly formatted?	The rows and columns are properly formatted
What surprises did you discover in the data?	Members have more ride length and users than Casual riders.
What trends or relationships did you find in the data?	- The minimum ride length for all users was in Jan 2023 at 00:13:00 so was the minimum for average ride length for member was in Jan 23 at 00:10:22 - The maximum for average ride length for member is in May 22 at 00:21:59 but Day 1 in May 22 as maximum casual members length by day of week at 00:25:43 - The maximum for average ride length for casual users is in Mar 22 at 00:32:38 and Day 1 in MAR-22 as the maximum casual users ride length by day of week at 00:38:48.
How will these insights help answer your business questions?	By analyzing the most busy day of the week for all users, the most and least rides for users, the least and most month and total number of trip id.
Key tasks Aggregate your data so it's useful and accessible.	Aggregated the data by calculating the mean and max for ride_length and mode of day of week.
Organize and format your data.	Done
Perform calculations.	Done
Identify trends and relationships	Done
Deliverable A summary of your analysis	I conducted a few descriptive analyses on each month for the last 13 months 01/2022-01/2023 to calculate: - Mean and Maximum of ride length. - Mode of day of week. - Average ride length for members and casual riders. - Average ride length for users by day of week. - Counted the number of rides for users by day of week and trip id. - Then I merged them into a summary file to get: The minimum ride length for all users was in Jan 2023 at 00:13:00 - The maximum ride length for all users was in Sep 2022 at 23:05:00 - The mode of day of week is day 5 and 7.

	- The average total for average ride length for casual users is 00:22:22 - The average total for average ride length for members is 00:12:49. - The minimum for average ride length for casual users is in Aug 22 at 00:14:09 - The minimum for average ride length for member is in Jan 23 at 00:10:22 - The maximum for average ride length for casual users is in Mar 22 at 00:32:38 - The maximum for average ride length for member is in May 22 at 00:21:59 - Day 4 in SEP-22 has the minimum casual users ride length by day of week at 00:11:12 - Day 1 in MAR-22 maximum casual users ride length by day of week at 00:38:48 - Total ride length for casual users ride length by day of week at 09:10:59 - Day 5 in Jan 23 has the minimum members ride length by day of week at 00:09:50 - Day 1 in May 22 maximum casual members length by day of week at 00:25:43 - Total ride length for members ride length by day of week at 19:29:16 - Total average count casual users 5392.21978 - Day 3 in Oct-22 has the minimum casual users count trip id by day of week at 396 - Day 7 in Apr-22 has the minimum casual users count trip id by day of week at 36324 - Total count casual sum at 490692 - Total average count members 13095.63736 - Day 1 in Aug-22 has the minimum members count trip id by day of week at 317 - Day 4 in Nov-22 has the minimum members count trip id by day of week at 47437

	Total count casual sum at 1191703

Share

Guiding questions Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?	Yes, Casual users takes more average ride length compare to members. Casual users also have higher minimum and maximum average ride length compare members with lower minimum and maximum average ride length.
What story does your data tell?	That Casual user has the total highest average rides length for all users and members with the lowest. That Member has the total highest total rides length for all users and casual with the lowest. There more low rides length in Jan 2023 for all users but members have the lowest ride length in Day 5 of Jan 2023. There is more high ride length in Sep 2022 for all users but Casual users have the lowest ride in Day 4 of Sep 2022. The most day of the week is Day 5 and 7. In March 2022 Casual users has the maximum average ride and in Day 1 of March 2022 they recorded highest ride. In May 2022 Members has the maximum average ride and in Day 1 of May 2022 they recorded highest ride. In Aug 2022 Casual has the minimum ride length but Day 1 in Aug 2022 the count trip Id was at its lowest for members Lowest trip id for casual was in Day 3 Oct 2022 however Day 7 April 2022 they had the highest trip Id. Day 4 Nov 2022 has the highest trip Id for member and total member has the highest count trip ID.
How do your findings relate to your original question?	
What trends or relationships did you find in the data?	The minimum ride length for all users was in Jan 2023 at 00:13:00 so was the minimum for average ride length for member was in Jan 23 at 00:10:22

	☐ The maximum for average ride length for member is in May 22 at 00:21:59 but Day 1 in May 22 as maximum casual members length by day of week at 00:25:43 ☐ The maximum for average ride length for casual users is in Mar 22 at 00:32:38 and Day 1 in MAR-22 as the maximum casual users ride length by day of week at 00:38:48.
Who is your audience? What is the best way to communicate with them?	☐ Cyclistic ☐ Lily Moreno ☐ Cyclistic marketing analytics team ☐ Cyclistic executive team
Can data visualization help you share your findings?	Yes
Is your presentation accessible to your audience?	Yes
Key tasks Determine the best way to share your findings.	
Create effective data visualizations.	
Present your findings.	
Ensure your work is accessible.	
Deliverable Supporting visualizations and key findings.	