

3 Main Components:

1. **Data Analysis Approach:** (main)
 - a. Prepare and analyze data insights using the 6 steps of the data analysis process. (Ask, Prepare, Process, Analyze, Share, and Act)
 - b. Create clear, meaningful, and accessible visualizations.
 2. **Marketing Research Approach:** (before and after)
 - a. Clearly define (3) “business management question, marketing research problem, and research problems.”
 - b. Offer recommendations for possible future initiatives based on insights. (further actions)
 3. **Technical Approach:** (throughout)
 - a. Track everything; Measure everything.
 - i. Tools & resources.
 - ii. Problems & issues.
 - iii. Lessons Learned & Improved Steps.
 - iv. Changes & alterations.
 1. Reasons why?
 - v. Chronological progress.
-

Data Analysis Approach:

1) Ask:

Guiding questions:

- What is the problem you are trying to solve?
 - “How do annual members and casual riders use Cyclistic bikes differently?”
 - Preferences
 - Usage patterns
- How can your insights drive business decisions?
 - The main goal is to “Design a marketing strategy to convert casual riders into annual members”.
 - With these insights, our marketing team can design data-driven strategies to target casual riders by highlighting the benefits of annual memberships based on preferences and usage patterns of casual riders.

Key task:

- Identify the business task:
 - Access Cyclistic’s historical bike trip data.
 - Clean and normalize data if necessary.
 - Analyze data against the research problem & discover any visible trends.
 - Visualize the data in an accurate, unbiased, appropriate, and accessible manner.
 - Document all findings.
- Consider key stakeholders:
 - Lily Moreno (Director of Marketing): Oversees the development of marketing campaigns.
 - Cyclistic Marketing Analytics Team: Analyzes the data and prepares recommendations.
 - Cyclistic Executive Team: Decides on the approval and execution of the marketing strategy.
 - Casual Riders and Annual Members: The user groups whose behaviors are being studied to inform the marketing strategy.

☒ **Deliverable:**

- A clear statement of the business task:
 - “Analyze how annual members and casual riders use Cyclistic bikes differently to uncover patterns that will guide the development of targeted marketing strategies aimed at converting casual riders into annual members.”

2) Prepare:

Guiding questions:

- Where is your data located?
 - Amazon AWS
 - <https://divvy-tripdata.s3.amazonaws.com/index.html>
 - Made available by Motivate International Inc. under a non-exclusive, royalty-free, limited, perpetual license to access, reproduce, analyze, copy, modify, distribute, and use the data for any lawful purpose.
 - <https://divvybikes.com/data-license-agreement>
- How is the data organized?
 - .csv files of individual month data within .zip files
 - 1 workbook with 12 spreadsheets for each of the 12 months in the past year.
 - Each spreadsheet starts with 13 columns with the headers, “ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, member_casual”

ride_id: (categorical)	(a unique string of random number & uppercase characters of length 16; “ex: EF7EA123D46B98DA”)
rideable_type: (categorical)	(classic_bike, electric_bike, electric_scooter)
started_at: (datetime)	(mm/dd/yy hh:mm:ss AM, mm/dd/yy hh:mm:ss PM)
ended_at: (datetime)	(mm/dd/yy hh:mm:ss AM, mm/dd/yy hh:mm:ss PM)
start_station_name: (categorical)	(an address)
start_station_id: (categorical)	(number of varying length; a string of 2 uppercase characters followed by numbers)
end_station_name: (categorical)	(an address)
end_station_id: (categorical)	(number of varying length; a string of 2 uppercase characters followed by numbers)
start_lat: (numeric)	(latitude as a decimal number)
start_lng: (numeric)	(longitude as a decimal number)

end_lat: (numeric)	(latitude as a decimal number)
end_lng: (numeric)	(longitude as a decimal number)
member_casual: (categorical)	(casual, member)

- Are there issues with bias or credibility in this data? (ROCCC)

-

Reliable:	collected by Bikeshare LLC and under the ownership of the City of Chicago: released to the public.
Original:	It is unknown whether the data was transformed or modified before public release; however, this is likely for the need to protect the privacy of the individual from whom the data was gathered.
Comprehensive:	containing details on trips, usage times, and customer types (without personally identifiable information)
Current:	covers roughly the last 12 months of data from last year.
Cited:	https://divvy-tripdata.s3.amazonaws.com/index.html ; https://divvybikes.com/data-license-agreement

- How are you addressing licensing, privacy, security, and accessibility?
 - That data is used under the license indicated:
 - <https://divvybikes.com/data-license-agreement>
 - “The City of Chicago owns all right, title, and interest in the Data.”
 - Following data privacy regulations, personally identifiable information (PII) is not included in the final report and analysis to ensure privacy and security.
- How did you verify the data’s integrity?
 - There is consistent formatting between the spreadsheets.
 - There are some missing values in columns, “start_station_name, start_station_id, end_station_name, end_station_id, end_lat, end_lng, ”
 - There are some impossible “started_at” and “ended_at” times.
 - The column fields are related to our business task.
 - We have an adequate amount of data from which to derive insight.
- How does it help you answer your question?

- The trip data provides insights into the behavior of annual members and casual riders, such as:
 - Trip frequency by day and month
 - Preferred bike types
 - Trip durations

- Are there any problems with the data?
 - Missing/Incomplete records
 - Stored as incorrect data type.
 - Note: data are stored in the dataset month the bike was returned not when it was taken.

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:
 - "The data used for this analysis consists of 12 months of Cyclistic's historical trip data, publicly provided by Motivate International Inc. The datasets are organized by month and include details about each bike trip, such as trip duration, start/end times, station locations, bike types, and user types (casual riders or annual members). All personally identifiable information has been excluded to maintain privacy, and the data is licensed for public use."

3) Process:

Guiding questions:

- What tools are you choosing and why?
 - I'm using Excel to evaluate its performance under heavy loads, enhance my proficiency in optimization techniques to reduce processing time, and improve my familiarity with handling time formats. I will also be using Excel for visualization to assess its capabilities and eventually compare it with more standard tools like Power BI and Tableau.
- Have you ensured your data's integrity?
 - Yes, all column field in each spreadsheet has a standardized format.
 - I checked for consistency, completeness, and accuracy in all fields.
 - Especially, removing column data that does not address the business task.
- What steps have you taken to ensure that your data is clean?
 - Handling missing values: the column fields in which null values where null values existed bordered on potential data privacy concerns and did not address the business task.
 - **Outliers**: took note of the records with a "BikeType" of "electric_scooter" that existed in 09-2024 data, and removed records with impossible started-ended time stamps.
 - Standardizing data formats: Ensure that all date and time fields are in the correct format.
- How can you verify that your data is clean and ready to analyze?
 - Run summary statistics on the dataset to check for any irregularities or inconsistencies in the cleaned data.
 - Perform sample checks of random rows to ensure that the data transformations were applied correctly.
- Have you documented your cleaning process so you can review and share those results?
 - Yes, located in "Cyclistic Capstone Project (Log)"
 - **Link:**

Key Tasks:

- Check the data for errors.
- Choose tools.
- Transform the data so you can work with it effectively.
- Document the cleaning process.

☒ **Deliverable:**

- Documentation of any cleaning or manipulation of data:
 - In, "Cyclistic Capstone Project (Log)"
 - **Link:**

4) Analyze:

Guiding Questions:

- How should you organize your data to perform analysis on it?
 - Excluding privacy-sensitive data, by variables of interest (both independent and dependent). MembershipType, RideType, Start and End Times, and any other information that can be derived from the previous columns.
- What surprises did you discover in the data?
 - Casual riders had a longer average use duration than Cyclistic members.
- What trends or relationships did you find in the data?
 - Target months - June, July, August, September (summer into the beginning of fall)
 - Casual members have a longer average ride duration than Cyclistic members, although showing greater dispersity within the group.
 - During the target months, electric bikes are favored over classic bikes.
 - Casual riders tend to peak at 5 pm.
 - Cyclistic members tend to peak at 8 am and 5 pm.
 - Average ride duration tends to be greater during weekends (Sa, S)
 - Most riders ride for around 5 minutes.
- How will these insights help answer your business questions?
 - They compare the usage of the two memberships of interest.

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:
 - Document in "Cyclistic Capstone Project (Log)"
 - [Link:](#)

5) Share:

Guiding Questions:

- Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
- What story does your data tell?
- How do your findings relate to your original question?
- Who is your audience? What is the best way to communicate with them?
- Can data visualization help you share your findings?
- Is your presentation accessible to your audience?

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

6) Act:

Guiding Questions:

- What is your final conclusion based on your analysis?
- How could your team and business apply your insights?
- What next steps would you or your stakeholders take based on your findings?
- Is there additional data you could use to expand on your findings?

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

Marketing Research Approach:

1) Business management question:

- “How do I increase annual memberships, specifically, convert casual riders to annual members?”

2) Marketing research problem:

- “How do annual members and casual riders use Cyclistic bikes differently?”

3) Research problems:

- “Which riders (annual/casual) are more likely to ride for leisure?”
- “Which riders (annual/casual) are more likely to commute to work using rental bikes?”
- “What types of bikes do casual riders use?”
- “What types of bikes do annual members use?”
- “What is the average ride time of casual riders?”
- “What is the average ride time of annual members?”