## **Google Data Analytics Capstone: Complete a Case Study**

**Cyclistic Case Study – Data Analysis Process**

**Introduction**

Welcome to the Cyclistic bike-share analysis case study! In this case study, I (assume that) work for a fictional company, Cyclistic, along with some key team members. In order to answer the business questions, follow the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act. Along with the Case Study.

**Scenario**

I (assume that) am 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, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, my 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 rider’s option 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 my (assume that) 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. I (assume that) joined this team six month ago and have been busy learning about Cyclistic’s mission and business goals—as well as how I (assume that), 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 al-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 me (assume that) the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

I will produce a report with the following deliverables:

1. A clear statement of the business task

2. A description of al 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.

|  |  |
| --- | --- |
| **Date:** 2025-07-12 | **Topic:** Ask |
| **Prompt:** | * How do annual members and casual riders use Cyclistic bikes differently? * Create a report with business task, data sources, cleaning, analysis, visualizations, and recommendations. |
| **Journal Entry:** | * The business task is to understand the behavioral differences between casual riders and annual members. * This insight will support marketing strategies to convert casual riders to members. * Key stakeholders include Lily Moreno (Marketing Director) and the executive team. |
| **Other thoughts or questions:** | * What specific marketing goals does Lily prioritize most? * Are there any current hypotheses about user behavior? |

|  |  |
| --- | --- |
| **Date:**2025-07-12 | **Topic:** Prepare Phase |
| **Prompt:** | Use Cyclistic’s historical trip data to analyze and identify trends. Describe all data sources used and assess their credibility, structure, and relevance to the business task.. |
| **Journal Entry:** | Today, I downloaded and reviewed Cyclistic’s historical trip data, which is publicly available via [Divvy 2020 Q1](https://docs.google.com/spreadsheets/d/179QVLO_yu5BJEKFVZShsKag74ZaUYIF6FevLYzs3hRc/template/preview#gid=640449855). The data consists of 1 month (Jan-2020) of trip history files in .xlsx format. It contains information such as ride ID, start and end time, start and end station, user type (casual vs. member), bike type, and ride duration.  The data is organized in a tabular format, and each row represents a unique bike trip. The most critical column for my analysis is member\_casual, which distinguishes between members and casual riders.  To ensure the data’s credibility, I verified that:   * The source is official and publicly licensed. * The dataset matches across months and has consistent formatting. * Personally identifiable information is not included, maintaining user privacy.   This dataset helps answer the key question: **How do annual members and casual riders use Cyclistic bikes differently?** It allows me to compare ride duration, frequency, days of the week, time of day, and bike type between casual and member riders.  I noticed minor issues, such as missing values and inconsistent station names, which I plan to clean in the next phase. The data **ROCCC**:   * **R**eliable (comes from a trusted public source), * **O**riginal (direct from the provider), * **C**omprehensive (includes all relevant fields), * **C**urrent (covers a full month), * **C**ited (licensed by Motivate International Inc.). |
| **Other thoughts or questions:** | * I wonder if there are seasonal patterns in how each rider type uses the bikes. * Could weather data enhance the analysis? * How should I handle missing or null station names in trip records? |

|  |  |
| --- | --- |
| **Date:** 2025-07-12 | **Topic:** Process Phase |
| **Prompt:** | Process your data to prepare it for analysis. Ensure data integrity, clean and transform it as needed, and document the cleaning steps for reproducibility. |
| **Journal Entry:** | Today, I began processing the Cyclistic trip data from the last 1 month to prepare it for analysis. I used **Google Sheets** for data processing because it allows for cloud storage, collaboration, and basic formula-based data transformation — and it works well within the memory limits. Tools Chosen:  * **Google Sheets** (for cleaning and transformations) * **Google Drive** (for organizing and storing data)  Key Steps Taken:  1. **Downloaded the** .xlsx files. 2. Created folders on my Google Drive:    * One for **original .xlsx files**    * One for **converted Google Sheets** 3. **Uploaded the .xlsx into Google Sheets** and saved a copy in the appropriate subfolder. 4. For the dataset:    * Added a column ride\_length using the formula =ended\_at - started\_at (in HH:MM:SS format using Format > Number > Time)    * Added a column day\_of\_week using the formula =WEEKDAY(started\_at,1) (where 1 = Sunday, 7 = Saturday)    * Verified time zone consistency across datasets.    * Checked for:      + Missing values (e.g. station names or user type)      + Negative ride lengths or zero duration      + Duplicated or incomplete rows  Cleaning & Transformation Summary:  * Removed rows where ride\_length was negative or zero. * Standardized column names. * Filtered out rows with missing station data. * Ensured that only valid ride types and user types (member/casual) remained * Verified formatting consistency in rideable\_type and member\_casual columns.  Data Integrity:  * The file was successfully processed with the same structure and format * Confirmed that no PII (personally identifiable information) is present * Retained only relevant fields: start time, end time, station info, ride type, user type, etc. |
| **Other thoughts or questions:** |  Should I combine all 12 months into one master dataset before analysis?   Would it be helpful to add columns like month or hour\_of\_day for deeper pattern recognition? |

|  |  |
| --- | --- |
| **Date:** 2025-07-13 | **Topic:** Analyze Phase |
| **Prompt:** | Aggregate and analyze the data to identify patterns and insights. Perform calculations to support answering the key business question: How do annual members and casual riders use Cyclistic bikes differently? |
| **Journal Entry:** | Tool Chosen:  * **RStudio Destop** (for analysed data)  How I Organized and Analyzed the Data:  * Ensured uniform formatting across columns (ride\_id, rideable\_type, started\_at, ended\_at, ride\_length, member\_casual, etc.). * **Grouped data** by member\_casual and:   + Calculated average ride length   + Counted rides per day of the week   + Summarized usage by time of day  Key Trends & Relationships:  |  |  |  | | --- | --- | --- | | **Metric** | **Casual Riders** | **Annual Members** | | **Average Ride Length** | ~96 minutes | ~13 minutes | | **Peak Days** | Weekends | Weekdays | | **Time of Day (Peak Use)** | Afternoons (11–7 PM) | Morning & Evening Commutes |  Surprises Discovered:  * Casual riders often take significantly longer trips, indicating leisure-focused behaviour. * Casual riders show sharp seasonal patterns, especially peaking in weekends * Members consistently ride during the workweek, supporting the idea that they use bikes for commuting.  How This Helps Answer the Business Question: These insights help clarify usage patterns:   * Casual riders prefer weekends and longer rides → possibly tourists or leisure users. * Members prefer shorter, regular weekday rides → commuters or local users.   Understanding these behavioural differences supports crafting **targeted marketing messages** (e.g., highlighting cost savings and convenience for commuters to casual users). |
| **Other thoughts or questions:** | * Could weather data deepen seasonal trend analysis? * Should I break down behaviour by neighbourhood or station clusters next? |

|  |  |
| --- | --- |
| **Date:** 2025-07-13 | **Topic:** Share Phase |
| **Prompt:** | Create sophisticated and accessible visualizations to communicate how annual members and casual riders use Cyclistic bikes differently. Share findings in a way that engages the executive team and informs decision-making. |
| **Journal Entry:** | Today I focused on creating data visualizations in R to share how annual members and casual riders use Cyclistic bikes differently. I chose ggplot2 because it allows detailed customization, which helps produce polished and professional charts.  Steps I followed:   1. Sketched initial ideas on paper — I wanted to highlight ride duration differences, ride counts by day of week, and bike type usage between rider types. 2. Opened RStudio and loaded the ggplot2 and dplyr packages for plotting and data manipulation. 3. Created several charts:    * **Bar chart** comparing ride length by member type to show distribution and median ride times.    * **Bar chart** showing total rides per day of week by rider type, using colour to differentiate.    * **Line chart** to display proportions of Ride Usage by Hour of Day. 4. Added meaningful titles, axis labels, and legends with contrasting colours for clarity. 5. Saved plots as high-resolution images for presentation slides. |
| **Other thoughts or questions:** | * Would interactive dashboards (e.g., using Shiny) add value for future presentations? * How can I further simplify charts for a non-technical executive audience? * Explore additional colour-blind-friendly palettes to improve accessibility. |

|  |  |
| --- | --- |
| **Date:** 2025-07-14 | **Topic:** Act Phase |
| **Prompt:** | Summarize your final conclusions from the analysis, state how the business can apply the insights, and prepare your top three recommendations for the marketing team. Prepare your case study portfolio and practice your presentation. |
| **Journal Entry:** | **Final Conclusion:** The analysis clearly shows that annual members use Cyclistic bikes differently from casual riders. Members take longer rides, ride more frequently on weekdays (suggesting commute use), and tend to use traditional bike types more often than casual riders. Casual riders, in contrast, mostly ride shorter distances and prefer weekends, indicating leisure use.  **Business Application:** These insights can help Cyclistic design targeted marketing strategies. By emphasizing benefits that appeal to casual riders — such as cost savings for frequent riders, convenience for daily commuters, and access to diverse bike types — the company can effectively encourage casual riders to convert to annual memberships.  **Next Steps:**   * Develop marketing campaigns highlighting value and convenience tailored to casual riders’ patterns. * Explore digital media channels that casual riders frequently use, focusing on weekends and leisure-related messaging. * Monitor conversion rates post-campaign to measure effectiveness and adjust as needed.   **Additional Data to Expand Findings:** Incorporating demographic data or survey feedback from riders could deepen understanding of motivations and barriers to membership.  **Top 3 Recommendations:**   1. **Target weekend casual riders with promotions that emphasize cost-effectiveness and convenience of annual membership for more frequent, longer rides.** 2. **Highlight the benefits of diverse bike options (e.g., assistive bikes) to appeal to a broader customer base and differentiate Cyclistic from competitors.** 3. **Use digital marketing campaigns focused on social media and email channels popular among casual riders, with clear calls to action for membership sign-up.** |
| **Other thoughts or questions:** |  How to best measure the success of marketing strategies post-launch?   What tools or dashboards can help the marketing team track membership conversion in real-time?   Explore potential partnerships with local businesses or workplaces to promote memberships. |

|  |  |
| --- | --- |
| **Date:** 2025-07-14 | **Topic:** Wrap-up |
| **Prompt:** | Reflect on completing the Cyclistic case study, summarize your experience, and consider your next steps for building your data analytics portfolio. |
| **Journal Entry:** | I’ve successfully completed the Cyclistic bike-share case study, covering all stages from asking the business questions to acting on data-driven recommendations. This experience deepened my skills in data cleaning, analysis, visualization, and communicating insights effectively for business decisions.  Following the structured roadmap helped me stay organized and ensure my work addressed key stakeholder needs. I’m now more confident presenting data-driven stories to leadership and making actionable recommendations.  Going forward, I plan to either take on another case study or design an independent project using the Ask-Prepare-Process-Analyze-Share-Act framework to continue strengthening my portfolio. This will enhance my readiness for job interviews and real-world analytics challenges.  Best of luck to all analysts pursuing data-driven success! |
| **Other thoughts or questions:** | * What additional datasets would be good to explore next? * How can I practice storytelling with data more effectively? * Should I learn advanced visualization tools like Tableau or Power BI next? |