Data Journal

Date: 9 July	**Course/topic: Google Data Analytics Capstone - Ask Phase**
Prompt:	What is the business task?
Journal Entry:	The business task is to find out how casual riders and annual members use Cyclistic bikes differently. The goal is to provide useful insights that can help the marketing team convert more casual riders into annual members. This will involve analyzing usage patterns such as when, how often, and for how long both rider types use the bikes.
Other thoughts or questions:	I'm curious to see what specific behaviors separate casual users from loyal members and how that data can guide Cyclistic's marketing strategy.

Date: 9 July	**Course/topic: Google Data Analytics Capstone - Prepare Phase**
Prompt:	What data did you use and where did you get it?
Journal Entry:	I used the public Cyclistic trip data from the last 12 months, available on the official Divvy Bikes website. I downloaded the monthly .csv files, extracted them from zip folders, and opened them in Excel to review. During this, I noticed that the July 2024 dataset had a large number of missing values, especially in station name and ID columns. The August 2024 dataset was much cleaner and more reliable for analysis. All data is made available under a public license by Motivate International Inc. and does not include any personally identifiable information, which helps ensure data privacy. I stored the files securely and verified that each file had the expected structure and key columns needed for analysis.
Other thoughts or questions:	Still deciding whether to use the July file after cleaning or drop it completely.

Date: 9 July	**Course/topic: Google Data Analytics Capstone - Process Phase**
Prompt:	What cleaning steps did you take and why?
Journal Entry:	After loading the Cyclistic trip data in Excel, I began cleaning the dataset. First, I removed all rows that contained blank or null values in key columns like station names and ride types. I also deleted irrelevant columns such as ride_id, start_lat, start_lng, end_lat, and end_lng, as they were not required for my analysis. Additionally, I verified that all data types were in the correct format—especially date and time fields. I ensured that all monthly files were structurally consistent and removed duplicate entries wherever found. These steps helped prepare a clean and reliable dataset for analysis.
Other thoughts or questions:	The July 2024 data had too many blanks. I'm unsure whether to keep it after cleaning or drop it altogether.

Date: 9 July	**Course/topic: Google Data Analytics Capstone - Analyze Phase**
Prompt:	What did you find from the data?
Journal Entry:	From the cleaned dataset, I discovered some interesting trends. Member riders tend to use bikes more on weekdays—especially Mondays, Thursdays, and Fridays—while casual riders are more active on weekends. I also found that the average ride duration between both groups is not significantly different, which was surprising. Additionally, ride counts for both rider types are quite close on Sundays. These insights can help the marketing team focus on when and how to target casual users to increase membership.
Other thoughts or questions:	I might explore other dimensions like bike type usage or ride start hours to deepen the analysis.

Date: 9 July	**Course/topic:** Google Data Analytics Capstone – Share Phase
Prompt:	How will you share your findings?
Journal Entry:	I created a clear and simple report summarizing my analysis of Cyclistic data, including key charts on ride patterns, user types, and ride durations. I formatted visuals in Excel for easy understanding by stakeholders. My aim is to deliver actionable insights to the marketing team through reports or presentations that help improve membership conversion strategies.
Other thoughts or questions:	Should I prepare a slide deck as well? What's the best way to ensure non-technical stakeholders understand the data?

Date: 9 July	**Course/topic:** Google Data Analytics Capstone – Act Phase
Prompt:	What actions should be taken based on your analysis?
Journal Entry:	Based on my analysis, I recommend Cyclistic focus marketing efforts on casual riders active on weekends by offering targeted promotions. Additionally, improving bike availability during peak times can enhance user experience. Further analysis using SQL to identify high-demand hours and locations is suggested. These actions aim to increase annual memberships and optimize operations.
Other thoughts or questions:	How can the impact of these recommendations be effectively measured? Should Cyclistic implement a pilot program before full rollout?