



Data Journal

Date: July 25	Course/topic: Course 1: Data, data everywhere
Prompt:	What does the word “data” mean to you?
Journal Entry:	When I think of data, I think of records and information. Data is a digital record. It could be a record of a sale, GPS records of where a car has been, or posts of images and pictures on social media. Data in itself isn’t useful, because there’s so much of it. However, enough data usually contains information and patterns. This information is valuable, and data analysts are the ones who extract it.
Other thoughts or questions:	The amount of data that is created each year gets bigger and bigger! Data is important! Nowadays, companies that don’t use their data are at a disadvantage to those that do.

Date: July 28	Course/topic: Course 2: Ask Questions to Make Data-Driven Decisions
Prompt:	How does structured thinking contribute to data analysis?
Journal Entry:	Structured thinking is a methodical approach to problem-solving that entails breaking down big situations into smaller, manageable components. It supports the development of logical problem-solving strategies, the organization of ideas, and the identification of important topics in data analysis. Analysts may make sure they ask the proper questions, take into account all pertinent information, and provide thorough answers by using organized thinking. This methodology is particularly beneficial in handling extensive datasets or intricate business issues, since it facilitates concentration and avoids the omission of crucial elements.
Other thoughts or questions:	I'm interested in learning more about specific techniques or frameworks for structured thinking in data analysis. How do experienced analysts apply this concept in their daily work?



Date: Aug 5	Course/topic: Course 3: Prepare Data for Exploration
Prompt:	How do you distinguish between biased and unbiased data?
Journal Entry:	Making the distinction between skewed and objective data is essential to precise analysis. A fair and representative sample of the population or phenomena under study is represented by unbiased data. Conversely, systematic mistakes in biased data cause outcomes to be skewed in one way. Personal preconceptions, data gathering methodologies, and sample strategies are some of the factors that might induce bias. To spot any biases and take them into consideration in the analysis, data analysts must thoroughly evaluate the sources of their data and the procedures used to acquire it.
Other thoughts or questions:	I wonder how often unintentional bias creeps into data collection processes. What are some best practices for minimizing bias in data collection?

Date: Aug 20	Course/topic: Course 4: Process Data from Dirty to Clean
Prompt:	What does data integrity mean to you?
Journal Entry:	The stability, consistency, and correctness of data across the course of its lifespan are referred to as data integrity. It's essential to guaranteeing that reliable and unmodified data persists. Maintaining data integrity in the context of data analysis entails recognizing and reducing risks—such as input mistakes, data corruption, or illegal access—that might jeopardize the accuracy of the data. Maintaining data integrity is essential for data analysts to provide trustworthy and useful insights.
Other thoughts or questions:	I'm curious about the specific SQL functions we'll learn for cleaning string variables. How often do data analysts encounter dirty data in real-world scenarios?



Date: Oct 6	Course/topic: Course 5: Analyze Data to Answer Questions
Prompt:	Why is organizing data before analysis important?
Journal Entry:	Data organization is important before analysis for a number of reasons. It facilitates the easier identification of patterns and trends, lowers the possibility of analytical mistakes, and ultimately saves time. We can rapidly and meaningfully arrange data by using sorts and filters, which makes it easier to identify gaps, inconsistencies, and outliers. Later on in the process, the use of more sophisticated analytical approaches is made easier by this basic arrangement. Well-structured data serves as the basis for precise and effective analysis.
Other thoughts or questions:	I'm excited to learn about SQL queries for combining data from multiple tables. How complex can these queries get in real-world scenarios?

Date: Oct 15	Course/topic: Course 6: Share Data through the Art of Visualisation
Prompt:	What makes a data visualization effective?
Journal Entry:	A well-designed data visualization effectively conveys complicated information in a way that is both aesthetically pleasing and simple to comprehend. Using the proper colors and labels, selecting the suitable chart or graph style for the data, and making sure the visualization tells a cohesive story are all important components. The objective is to draw attention to data-related patterns, trends, or correlations that might not be readily obvious from the raw data. Visualizations that are effective should be simple enough for viewers to immediately understand the key points without requiring a lot of explanation.
Other thoughts or questions:	I'm excited to learn about Tableau and its capabilities. How does it compare to other visualization tools? What are some common pitfalls to avoid when creating data visualizations?



Date: Oct 17	Course/topic: Course 7: Data Analysis with R Programming
Prompt:	What are your initial thoughts on R as a programming language for data analysis?
Journal Entry:	It appears that R is a strong and adaptable programming language made especially for statistical computation and graphics. It is a vital tool for data analysts due to its capacity to manage enormous datasets, carry out intricate statistical analysis, and provide sophisticated visualizations. R's vector notion is especially intriguing since it makes data manipulation more effective. Additionally, it seems that R's usage of pipes simplifies data manipulation procedures, improving the readability and clarity of code.
Other thoughts or questions:	I'm curious about how R compares to other data analysis tools like Python or SQL in terms of performance and ease of use. How steep is the learning curve for R?
