Data-Driven Decisions

PART 1: Data for Supervised Learning

[Step 1] Choose a scenario Choose one scenario from the provided list that you find interesting or relevant.	• Predict which customers are likely to cancel their subscriptions or stop using a service. • Classify emails as spam or not spam. • Predict the selling price of houses based on various features.
[Step 2] Identify the data you would use to train a model Identify at least three types of data you would need to collect or have access to develop a machine-learning model for your chosen scenario. Provide a brief rationale for your choices.	Understanding past customer behavior, including average subscription duration and reasons for cancellation (e.g., high cost or product overstock), is crucial for improving retention strategies. Additionally, analyzing how long customers typically use similar products/services from other brands provides valuable benchmarks for setting realistic expectations and tailoring marketing efforts.
[Step 3] Discuss Data Challenges Describe any potential hurdles or constraints in gathering the data or leveraging it for your model	It may not always be possible to determine why customers discontinue a subscription or to gather feedback on why they stop purchasing. This limitation can result in insufficient or skewed data, making it challenging to produce an unbiased analysis.

PART 2: Data for Unsupervised Learning

[Step 1] Choose a scenario Choose one scenario from the provided list that you find interesting or relevant.	• Group customers into segments based on purchasing behavior. • Identify fraudulent transactions in financial data. • Recommend articles, movies, or products to users based on their interests.
[Step 2] Identify the data you would use to train a model	Since we lack existing data or examples to guide the generative AI in formulating responses on a specific topic, it would need to actively engage with users by asking

Identify at least three types of data you would need to collect or have access to develop a machine learning model for your chosen scenario. Provide a brief rationale for your choices.	questions and probing for preferences across various media or products. As users provide answers, the AI can refine its questions, making them more specific and improving the quality of recommendations. Additionally, we could leverage patterns from our existing customers' shopping habits, compare them to other brands' recommendations, and use this data to enhance the AI's suggestion accuracy.
[Step 3] Discuss Data Challenges Describe any potential hurdles or constraints in gathering the data or leveraging it for your model	Without sufficient data to guide its outputs, the AI may struggle to meet user expectations. It could lack the necessary knowledge to fulfill the request effectively, as it may not have enough examples, classifications, or language models to reference. This limitation can hinder its ability to provide meaningful support.

PART 3: Data for Generative AI

[Step 1] Choose a scenario Choose one scenario from the provided list that you find interesting or relevant.	• Generate images of customers wearing products without physical try-on. • Create advanced chatbots capable of handling a wide range of customer service inquiries with natural, human-like responses. • Produce creative text and images for advertising campaigns.
[Step 2] Identify the data you would use to train a model Identify at least three types of data you would need to collect or have access to develop a machine learning model for your chosen scenario. Provide a brief rationale for your choices.	Using generative AI to build advanced chatbots enables the AI to learn how to respond effectively to various types of user input, even when questions are phrased uniquely or in unfamiliar ways. Unlike traditional AI, generative AI can create responses independently, without predefined content, allowing it to provide accurate answers without requiring human intervention. It remains active continuously, learning from past interactions with both customers and staff to expand its knowledge base and improve response accuracy over time.
[Step 3] Discuss Data Challenges Describe any potential hurdles or constraints in gathering the data or leveraging it for your model	If the AI has limited data or an inadequate approach for generating accurate responses, it can lead to customer frustration, incorrect purchases, dissatisfaction, product returns, and negative reviews. If the AI isn't trained to continuously learn about the brand's offerings, FAQs, best practices, and appropriate customer communication, it risks

	damaging the customer experience. This can ultimately result in lost revenue and a decline in brand reputation.