1. Analyzing Student Performance:

Scenario: You are analyzing factors that affect student performance in a standardized test. You collect data on study hours, attendance rates, and socioeconomic background.

Question: How would you set up your linear regression model? What considerations would you make regarding the interpretation of coefficients in this context?

2. Time Series Data

Scenario: You have monthly data on electricity consumption over several years and want to predict future consumption based on trends and seasonal patterns.

Question: Can linear regression be effectively used in this scenario? If so, how would you incorporate time as a variable in your model?

3. Evaluating Marketing Strategies

Scenario: A marketing team has created a decision tree to evaluate different advertising strategies for a new product launch.

Question: How would you assess the effectiveness of the current decision tree structure? What specific metrics or outcomes would you analyze to determine if any adjustments are necessary?

4. Financial Investment Decisions

Scenario: An investment firm uses a decision tree to analyze potential investment opportunities based on market trends and economic indicators.

Question: What process would you follow to update the decision tree as market conditions change? How would you visualize these changes for clarity among team members?

5. Disease Diagnosis

Scenario: A healthcare provider uses patient data (symptoms, age, medical history) to diagnose diseases using K-NN.

Question: What considerations should you take into account when choosing the value of $$ K $$? How would you ensure that the model is robust against overfitting?

6. Image Classification\*\*

Scenario: You are developing an application that classifies images of animals (e.g., cats vs. dogs) using K-NN.

Question: Given the high dimensionality of image data, what techniques could you use to optimize the performance of K-NN? How would you measure the accuracy of your model?

7. Credit Risk Assessment

Scenario: A bank uses logistic regression to determine the probability of a loan applicant defaulting on their loan based on their credit score, income level, and employment status.

Question: How would you interpret the coefficients of your logistic regression model in this context? What implications do these coefficients have for risk assessment?

8. Email Classification

Scenario: You are tasked with building a spam filter using SVM to classify emails as either "spam" or "not spam."

Question: How would you decide between using a linear SVM and a non-linear SVM for this classification problem? What features would you consider important for your model?

9. Real Estate Price Prediction

Scenario: A real estate company wants to predict house prices based on features like location, size, and number of bedrooms.

Question: If your model's R-squared value is 0.85, what does this indicate about the model's performance? Are there any limitations to using R-squared as the sole metric?

10. Gene Expression Analysis

Scenario: In a biological study, researchers have collected gene expression data from thousands of genes across multiple samples.

Question: How would PCA help in identifying key patterns or variations in gene expression? What challenges might arise when interpreting the principal components?

11.Weather Prediction

Scenario: You create a simple linear regression model to predict daily temperatures based on historical weather data.

Question: If your predictions are consistently inaccurate because they do not capture seasonal variations, what does this indicate about bias? How might you improve your model's accuracy?

12. Predicting Credit Card Fraud

Scenario: A financial institution is using logistic regression to identify fraudulent transactions based on transaction amount, location, and user behavior.

Question: In setting up your data pipeline, how would you ensure that the model is trained on balanced classes? What techniques could you employ to address class imbalance?