Exercise1:

Visualization Strategy Questions

1. How would you visualize the relationship between two categorical variables (e.g., gender and product preference)?
    - Suggest two visualizations and explain why they are appropriate.
2. You have a dataset with a time series of daily sales over the past year. How would you visualize seasonal trends and anomalies?
    - Provide two visualization methods that can effectively highlight these trends.
3. How would you visualize the distribution of a numerical variable (e.g., income) across different regions or categories?
    - Suggest two different types of visualizations and explain their strengths.
4. You are given a dataset with multiple numerical features, and you want to understand the relationships between all of them. Which visualization would you use?
    - Explain how a pair plot or correlation heatmap could help in this situation.
5. How would you visualize changes in customer behavior over time in a dataset with categorical data (e.g., product purchased) and time stamps?
    - Suggest two visualization techniques and explain why they would work.

Pattern Recognition and Interpretation Questions

1. Given a heatmap showing correlations between different features in a dataset, how would you identify which features are most strongly related?
    - What would you look for in terms of color patterns or numerical values?
2. You have a scatter plot with multiple groups, color-coded by category (e.g., age group, income group). How would you assess whether the groups are well separated or overlapping?
    - What patterns or features would you focus on in the scatter plot?
3. Given a bar chart comparing sales across multiple categories (e.g., region or product type), what patterns would suggest which category has the highest performance or growth?
    - How would you identify trends in the data from the chart?
4. In a line chart showing monthly website traffic for two different regions, what patterns would indicate seasonal spikes or anomalies?
    - Describe how the line chart could reveal these patterns.
5. Given a scatter plot of two variables with a trend line, how would you interpret the strength and direction of the relationship between the variables?
    - What would indicate a strong, weak, or no correlation?

Feature and Data Transformation Questions

1. You are given a dataset with skewed data for a numerical feature (e.g., income). How would you visualize the impact of transforming the data (e.g., log transformation) on its distribution?
    - Suggest two visualizations to compare the original and transformed data.
2. How would you visualize the impact of a feature selection technique (e.g., removing correlated variables) on a machine learning model's performance?
    - Suggest a way to visualize feature importance before and after feature selection.
3. Given a dataset with outliers, how would you visualize their impact on the overall distribution?
    - Suggest a visualization that could help identify outliers and their effect on the data.
4. How would you visualize a large dataset with numerous features to understand which features are most influential in predicting a target variable?
    - Suggest a feature importance plot or a dimensionality reduction technique, and explain how each could provide insights

Advanced/Reflective Questions

1. How would you visualize the effectiveness of a marketing campaign in increasing sales across different age groups or income levels?

- Suggest visualizations that could reveal trends and group differences over time.

2. Given a confusion matrix, how would you visualize the performance of a classification model?
   - What chart or plot would provide an intuitive understanding of the model's accuracy, precision, and recall?
3. You want to visualize the impact of different hyperparameters (e.g., learning rate, batch size) on model performance. What type of plot would you use to compare these metrics?
   - Suggest a visualization that can effectively show the relationships between hyperparameters and model performance.
4. In a dataset with multiple variables, what types of visualizations would allow you to identify potential multicollinearity?
   - How can a pair plot or correlation matrix help you identify multicollinearity?
5. Given a decision tree model's output, how would you visualize the decision-making process of the model?
   - Suggest a visualization that can explain the splits made by the decision tree.