**1. What are the key tasks involved in getting ready to work with machine learning modeling?**

- Data Collection: Gather relevant and representative data for the problem.

- Data Cleaning: Handle missing values, outliers, and inconsistencies.

- Data Exploration: Analyze data distribution, patterns, and relationships.

- Feature Engineering: Select or create appropriate features for the model.

- Data Preprocessing: Normalize, scale, or transform data as needed.

- Train-Test Split: Divide the dataset into training and testing subsets.

- Model Selection: Choose a suitable algorithm or model.

- Model Training: Train the chosen model on the training data.

- Model Evaluation: Assess the model's performance on the testing data.

- Hyperparameter Tuning: Optimize model parameters for better results.

**2. What are the different forms of data used in machine learning? Give a specific example for each of them.**

- Numerical Data: Data represented by numbers. Example: Temperature readings.

- Categorical Data: Data represented by categories or labels. Example: Types of fruits.

- Ordinal Data: Categorical data with a meaningful order. Example: Education levels (high school, college, etc.).

- Text Data: Natural language text. Example: Customer reviews.

- Image Data: Pixel values representing images. Example: Medical scans.

- Time Series Data: Data collected over time intervals. Example: Stock prices.

**3. Distinguish:**

**1. Numeric vs. categorical attributes:**

- Numeric attributes represent quantities and can be measured. Example: Age.

- Categorical attributes represent categories or labels. Example: Gender.

**2. Feature selection vs. dimensionality reduction:**

- Feature selection chooses relevant features from the dataset.

- Dimensionality reduction reduces the number of features while preserving important information.

**4. Make quick notes on any two of the following:**

1. The histogram:

- A histogram is a graphical representation of the distribution of data.

- It consists of bins (intervals) on the x-axis and frequency/count on the y-axis.

2. Use a scatter plot:

- A scatter plot displays data points as individual dots on a 2D graph.

- It shows the relationship between two numerical variables.

**5. Why is it necessary to investigate data? Is there a discrepancy in how qualitative and quantitative data are explored?**

- Investigating data helps identify patterns, outliers, and relationships.

- Qualitative data (categorical) is explored using frequency tables and charts, while quantitative data (numeric) is analyzed with histograms and statistical measures.

**6. What are the various histogram shapes? What exactly are ‘bins'?**

- Histogram shapes include normal (bell-shaped), skewed (positively or negatively), and multimodal.

- Bins are intervals on the x-axis of a histogram that group data points. The height of each bin represents the frequency/count of data points in that interval.

**7. How do we deal with data outliers?**

- Outliers can be handled by removing them, transforming them, or capping/extending them.

- Techniques like Z-score, IQR (Interquartile Range), and box plots help identify outliers.

**8. What are the various central inclination measures? Why does the mean vary too much from the median in certain data sets?**

- Central inclination measures include mean, median, and mode.

- The mean varies from the median in skewed datasets because outliers disproportionately affect the mean.

**9. Describe how a scatter plot can be used to investigate bivariate relationships. Is it possible to find outliers using a scatter plot?**

- A scatter plot shows how two variables are related by plotting their values as points.

- Outliers can often be identified in a scatter plot as points far from the main cluster.

**10. Describe how cross-tabs can be used to figure out how two variables are related.**

- Cross-tabs (cross-tabulation) are used to analyze the relationship between two categorical variables.

- They present the frequency distribution of the two variables in a table format, revealing patterns and associations.