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

**Answer: Key tasks in getting ready for machine learning modeling:**

* Data collection
* Data preprocessing (cleaning, handling missing values, etc.)
* Feature selection and engineering
* Splitting data into training, validation, and test sets
* Choosing a suitable model
* Training and evaluating the model

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

**Answer: Different forms of data in machine learning with examples:**

* Numerical Data (e.g., Age of a person)
* Categorical Data (e.g., Gender: Male/Female)
* Ordinal Data (e.g., Customer satisfaction levels: Low, Medium, High)
* Time-series Data (e.g., Stock prices over time)
* Text Data (e.g., Product reviews)
* Image Data (e.g., X-ray images for medical diagnosis)

**3. Distinguish:**

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

* Numeric attributes represent continuous or discrete numbers (e.g., height, weight).
* Categorical attributes represent distinct groups or categories (e.g., colors, countries).

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

* Feature selection chooses the most relevant features.
* Dimensionality reduction transforms features into a lower-dimensional space (e.g., PCA).

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

1. The histogram

* **Histogram**: A graphical representation of data distribution using bars.

2. Use a scatter plot

* **Scatter Plot**: A plot to visualize relationships between two numerical variables.

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

**Answer:**

Necessary to identify errors, patterns, and trends.

Qualitative Data: Explored using thematic analysis.

Quantitative Data: Explored using statistical summaries and visualizations

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

**Answer:**

**Shapes:** Normal, Skewed, Bimodal, Multimodal, Uniform.

**Bins:** Intervals that group data into frequency ranges.

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

**Answer:**

Removing extreme values

Transforming data

Using robust statistical methods (e.g., median instead of mean

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

**Answer:**

**Measures:** Mean, Median, Mode

The mean is sensitive to outliers, while the median remains stable.

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

**Answer:**

Shows correlation between two variables.

Outliers appear as points distant from the trend.

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

**Answer:**

Summarizes data in a matrix format.

Useful in categorical data analysis (e.g., gender vs. product preference).