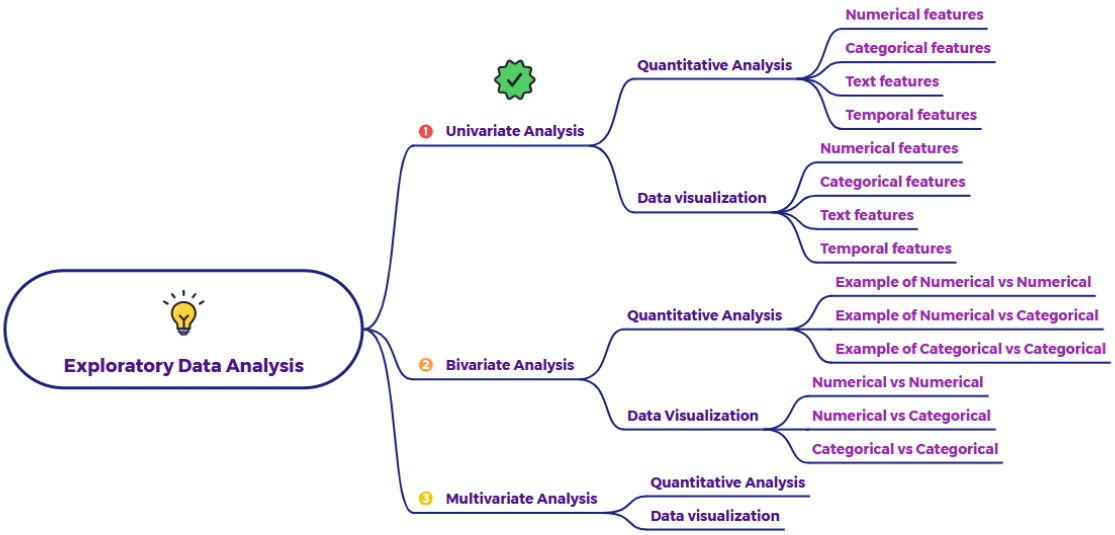
What is Univariate Analysis?



Univariate Analysis

Univariate analysis is the simplest form of data analysis, where we examine a single variable or feature in isolation. It helps us understand the distribution, characteristics, and patterns of that variable without considering any relationships with other variables.

Here's how we approach it with different types of variables:

1. Numerical Features

- **Quantitative Analysis**:
    - We use descriptive statistics to summarize the distribution's properties.
    - Examples:
        - Mean: Average value (e.g., average customer age).
        - Median: Middle value when data is ordered (e.g., median income).
        - Standard Deviation: Measures the amount of variation or dispersion of a set of values (e.g., standard deviation of test scores).

        - Percentiles: Values below which a certain percentage of the data falls (e.g., 25th, 50th, 75th percentiles of house prices).

- **Data Visualization**:
    - We use charts to visualize the distribution.
    - Examples:
        - Histogram: Shows the frequency distribution of data, grouping values into ranges (e.g., histogram of customer ages).
        - Box Plot: Displays the distribution and highlights key statistics like median, quartiles, and outliers (e.g., box plot of income).
        - Density Plot: Shows the probability density function of the data, providing a smooth representation of the distribution (e.g., density plot of exam scores).

2. Categorical Features

- **Quantitative Analysis**:
    - We analyze the frequency and proportion of each category.
    - Examples:
        - Frequency Count: Number of occurrences of each category (e.g., count of males and females in a dataset).
        - Proportions or Percentages: Share of each category relative to the total (e.g., percentage of customers who prefer each product).
- **Data Visualization**:
    - We use charts to visualize the distribution of categories.
    - Examples:
        - Bar Chart: Shows the frequency or proportion of each category (e.g., bar chart of preferred colors).

        - Pie Chart: Shows the relative proportion of each category as a slice of a circle (e.g., pie chart of market share for different brands).

3. Text Features

- **Quantitative Analysis**:
    - We analyze text characteristics and patterns.
    - Examples:
        - Word Count: Number of words in a text (e.g., average words in product descriptions).
        - Frequency of Specific Words: How often certain words appear (e.g., frequency of keywords in customer reviews).
        - Character Count: Number of characters in a text.
        - Average word length
- **Data Visualization**:
    - We use visualizations to represent text data characteristics.
    - Examples:
        - Word Cloud: Visual representation of word frequencies, where larger words indicate higher frequency (e.g., word cloud of customer feedback).
        - Bar chart of top N words: Show the most frequent words.

4. Temporal Features

- **Quantitative Analysis**:
    - We analyze time-based components and trends.
    - Examples:
        - Distribution of Dates: Range, most frequent year, month, day of the week (e.g., distribution of transaction dates).
        - Time between events: Average time between customer logins.

- **Data Visualization**:
  - We use charts to visualize time-based patterns.
  - Examples:
    - Histogram of events per month: Shows how events are distributed over months.
    - Line chart of website traffic over time: Visualizes trends and seasonality.
    - Bar chart of transactions by day of the week