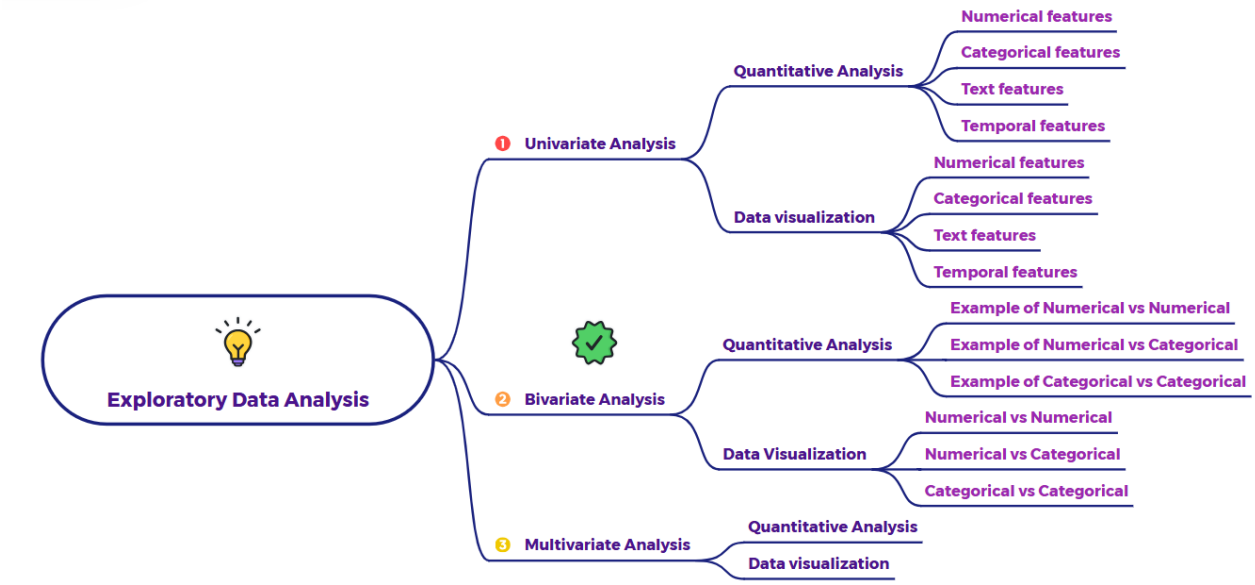
Different types of Bivariate Analysis



1. Numerical vs. Numerical Variables

- **Description:** This type of analysis examines the relationship between two variables that are both quantitative (i.e., they represent measurable quantities). We're often looking for patterns, trends, or correlations.
- **Quantitative Analysis Techniques:**
    - **Correlation:** Measures the strength and direction of a linear relationship. (e.g., Pearson correlation coefficient).
    - **Regression Analysis:** Models the relationship to predict the value of one variable based on the other. (e.g., linear regression).
    - **Covariance:** Measures how two variables change together.
- **Data Visualization Examples:**
    - **Scatter Plot:** Shows the relationship between two numerical variables, with one variable plotted on each axis.
    - **Line Plot:** Shows the trend of one numerical variable against another (often time).
    - **Heatmap (for correlation):** Visualizes the correlation matrix, where color intensity represents the strength of the correlation.

- **Example:**
    - **Scenario:** Analyzing the relationship between a student's study hours and their exam score.
    - **Numerical Variable 1:** Study Hours (e.g., hours per week)
    - **Numerical Variable 2:** Exam Score (e.g., percentage)
    - **Analysis:** We could use a scatter plot to visualize the data and calculate the correlation coefficient to see if there's a linear association between study time and exam performance. Regression analysis could be used to predict exam scores based on study hours.

2. Numerical vs.Categorical Variable

- **Description:** This analysis explores how a numerical variable differs across different categories of a categorical variable. We're looking to see if the central tendency or distribution of the numerical variable changes depending on the group.
- **Quantitative Analysis Techniques:**
    - **Comparing Means:** Using techniques like t-tests (for two categories) or ANOVA (for more than two categories) to see if the average of the numerical variable is significantly different across categories.
    - **Comparing Distributions:** Examining the spread and shape of the numerical variable's distribution for each category.
- **Data Visualization Examples:**
    - **Box Plots:** Show the distribution (median, quartiles, range, outliers) of the numerical variable for each category.
    - **Violin Plots:** Similar to box plots but also show the probability density of the numerical variable for each category.
    - **Histograms (faceted):** Display histograms of the numerical variable separately for each category.

- **Example:**
    - **Scenario:** Analyzing the relationship between the price of a car and its type.
    - **Numerical Variable:** Car Price (e.g., in dollars)
    - **Categorical Variable:** Car Type (e.g., Sedan, SUV, Truck, Sports Car)
    - **Analysis:** We could use box plots to compare the distribution of car prices across different car types. ANOVA could be used to test if the mean price is significantly different for at least one car type.

3. Categorical vs. Categorical Variable

- **Description:** This type of analysis examines the association between two categorical variables. We're looking to see if the occurrence of certain categories in one variable is related to the occurrence of certain categories in the other variable.
- **Quantitative Analysis Techniques:**
    - **Contingency Tables:** Summarize the frequency of observations for each combination of categories.
    - **Chi-Square Test:** Tests whether there's a statistically significant association between the two categorical variables.
    - **Measures of Association:** Calculate the strength of the association (e.g., Cramer's V, Phi coefficient).
- **Data Visualization Examples:**
    - **Stacked Bar Charts:** Show the composition of one categorical variable within each category of the other.
    - **Grouped Bar Charts:** Compare the frequency of categories in one variable across categories of the other.
    - **Heatmaps (for contingency tables):** Visualize the frequencies in a contingency table with color intensity representing the count.

- **Example:**
    - **Scenario:** Analyzing the relationship between gender and political party preference.
    - **Categorical Variable 1:** Gender (e.g., Male, Female)
    - **Categorical Variable 2:** Political Party Preference (e.g., Party A, Party B, Party C)
    - **Analysis:** We could create a contingency table to show how many males and females prefer each political party. A Chi-Square test can determine if gender and party preference are independent or associated.