**EXPLORATORY DATA ANALYSIS (EDA)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Aspect** | **univariate** | **bivariate** | **multi-variate** |
| Definition | Examines a single variable at a time. It focuses on understanding the distribution and characteristics of one variable. | Analyses the relationship between two variables simultaneously. It explores how one variable may be influenced by another. | Studies the interactions and relationships between three or more variables together, often using advanced techniques. |
| Primary Purpose | Understand the individual characteristics and behaviour of a variable. | Explore relationships, associations, or dependencies between two variables. | Investigate complex relationships and interactions between multiple variables. |
| Examples : | # Histogram for a numeric variable  plt.hist(df['Age'], bins=20 , color = 'black')  plt.xlabel('Age')  plt.ylabel('Count')  plt.title('Histogram')  plt.show() | # Scatter plot for two numeric variables  plt.scatter(df['Age'], df['Fare'], color='black')  plt.xlabel('Age')  plt.ylabel('Fare')  plt.title('Scatter Plot')  plt.show()  # Box plot for numeric vs. categorical variable  sns.boxplot(df['Age'], color='black')  plt.xlabel('Sex - male / female')  plt.ylabel('Age')  plt.title('Box Plot')  plt.xticks(rotation=45)  plt.show()  # Correlation matrix  correlation\_matrix = df.corr()  sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')  plt.title('Correlation Heatmap')  plt.show() | # Pairplot for multiple numeric variables  sns.pairplot(df[['Pclass', 'Age', 'Fare']])  plt.title('Pairplot')  plt.show()  # Grouped bar plot for multiple categorical variables  sns.barplot(x='Sex', y='Age', hue='Pclass', data=df, color='black')  plt.xlabel('Sex')  plt.ylabel('Age')  plt.title('Grouped Bar Plot')  plt.xticks(rotation=45)  plt.show() |