

Exploratory Data Analysis (EDA)

- **Introduction:**

The objective of this Exploratory Data Analysis (EDA) task is to extract meaningful insights from the dataset using statistical and visual methods. The process involves understanding the dataset structure, identifying data types, missing values, and outliers, performing univariate, bivariate, and correlation analysis, and summarizing key findings.

- **Dataset Overview:**

Dataset: Excel-Formulas-and-Functions-Complete-Sheet.csv Shape: 892 rows × 12 columns.

- **Univariate Analysis:**

Numerical Variables: - Plotted histograms with KDE to visualize distributions. - Observed skewness in some numeric features. Categorical Variables: - Plotted bar charts for top categories. - Some columns have dominant categories with high frequency.

- **Bivariate Analysis:**

Correlation Heatmap: - Showed positive and negative correlations. Category vs Numeric: - Boxplots showed variation across categories.

- **Outlier Detection:**

Boxplots identified outliers in several numeric columns. Some features showed significant deviations beyond the expected range.

- **Missing Value Treatment:**

Numerical columns: Filled with median. Categorical columns: Filled with mode. This approach preserves data distribution without introducing bias.

- **Key Insights:**

The dataset contains both categorical and numerical features. - Some numeric features are skewed. - Strong correlations between certain variables. - Few categories dominate in categorical variables. - Outliers detected in multiple numeric columns.

- **Conclusion:**

The EDA process provided a comprehensive understanding of the dataset's structure and patterns. These insights will guide future steps such as feature engineering, model selection, and predictive analysis.

- **References:**

Python Pandas Documentation - Seaborn Documentation - Matplotlib Documentation.