

# Exploratory Data Analysis (EDA): Identifying Patterns, Anomalies, and Insights in Raw Data

## Objective

The objective of this analysis is to explore raw data to uncover hidden patterns, detect anomalies, and generate insights without making prior assumptions. EDA serves as a crucial step for understanding the structure and characteristics of datasets before applying advanced modeling techniques.

## Materials and Methods

### Materials:

- Raw datasets (unstructured or structured)
- Visualization tools (e.g., Matplotlib, Seaborn, Excel)
- Statistical computing platforms (e.g., Python, R, SPSS)

### Methods:

1. Data Import: Load raw data into analysis tools.
2. Data Cleaning: Handle missing values, duplicates, and inconsistencies.
3. Summary Statistics: Calculate descriptive metrics (mean, median, standard deviation).
4. Visualization: Create plots such as histograms, scatter plots, and boxplots to examine distributions and relationships.
5. Pattern and Anomaly Detection: Identify trends, clusters, or outliers in the data.
6. Insights Extraction: Formulate initial hypotheses and questions for deeper analysis.

## Results

- EDA revealed significant outliers in transaction amounts for Q2.
- Scatter plot analysis showed a nonlinear relationship between age and purchase frequency.
- Histogram visualizations indicated data skewness that required transformation before modeling.

## Conclusion

Exploratory Data Analysis is a vital step in the data science workflow, enabling analysts to understand dataset structure and quality, detect anomalies, and generate meaningful questions. Insights from EDA guide subsequent stages of predictive modeling and hypothesis testing.