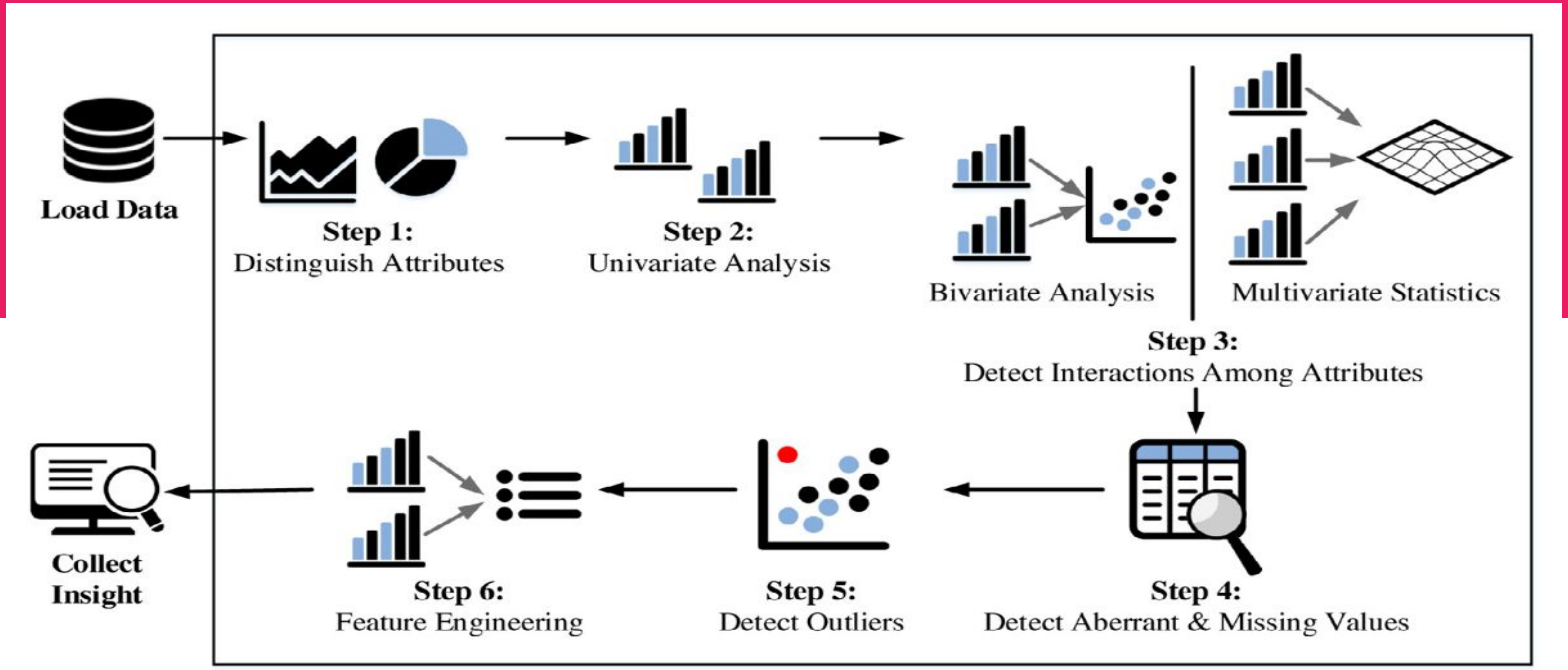
Exploratory Data Analysis



What is EDA?

Analysis of data in terms of various numerical description and graphs

Identifying terms, outliers, trends (making a sense of the story present in the data)

Figuring out the unexpected while validating the expected

Key Aspects of EDA

- Distribution of Data: Examining the distribution of data points to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).
- Graphical Representations: Utilizing charts such as histograms, box plots, scatter plots, and bar charts to visualize relationships within the data and distributions of variables.
- Outlier Detection: Identifying unusual values that deviate from other data points. Outliers can influence statistical analyses and might indicate data entry errors or unique cases.

Key Aspects of EDA <cont>

- Correlation Analysis: Checking the relationships between variables to understand how they might affect each other. This includes computing correlation coefficients and creating correlation matrices.
- Handling Missing Values: Detecting and deciding how to address missing data points, whether by imputation or removal, depending on their impact and the amount of missing data.
- Summary Statistics: Calculating key statistics that provide insight into data trends and nuances.
- Testing Assumptions: Many statistical tests and models assume the data meet certain conditions (like normality or homoscedasticity). EDA helps verify these assumptions.

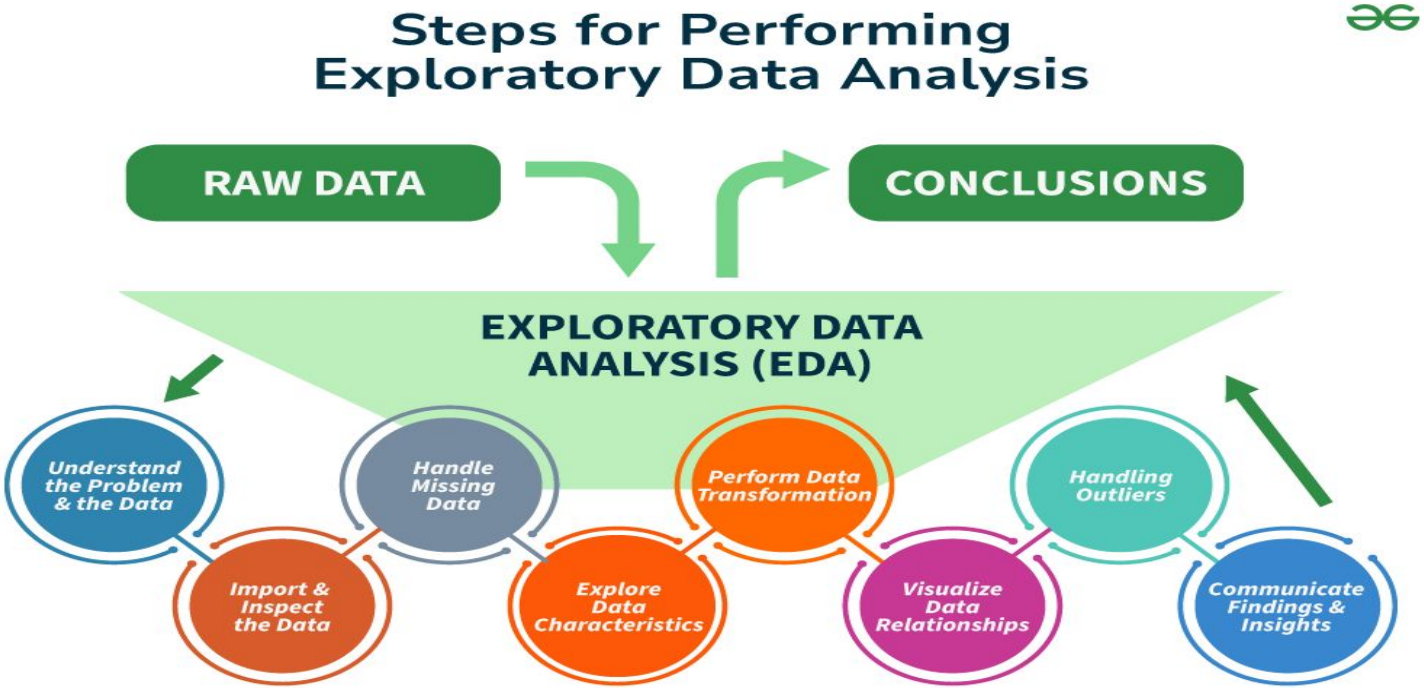
What does EDA do?

- Finds relationships, trends and outliers
- Helps understand data
- Test Assumptions
- Identify important feature
- Building block of model design
- Facilitate data cleaning
- Helps make data driven decisions

Types of EDA

- Univariate → Histogram, Boxplot, Bar charts and Summary statistics
- Bivariate → Scatter plot, Correlation Coefficient, Cross-tabulation, Line graphs and Covariants
- Multivariate → Pair plots and Principal Component Analysis (PCA)

Steps of EDA


Steps for Performing
Exploratory Data Analysis
RAW DATA
CONCLUSIONS
EXPLORATORY DATA
ANALYSIS (EDA)
Understand the Problem & the Data
Import & Inspect the Data
Handle Missing Data
Explore Data Characteristics
Perform Data Transformation
Visualize Data Relationships
Handling Outliers
Communicate Findings & Insights