

# EDA-Multivariate Analysis

---

## Multivariate Analysis (MVA) for Exploratory Data Analysis (EDA)

Multivariate analysis is a set of statistical techniques used to analyze data with multiple variables. In the context of exploratory data analysis, MVA helps identify patterns, relationships, and correlations among multiple features or variables in a dataset.

### Key Concepts:

1. **Dimensionality:** The number of variables in your dataset.
2. **Independence:** Whether each variable is independent of others (i.e., no multicollinearity).
3. **Correlation:** Measure of how closely related two or more variables are.

## Multivariate Analysis Techniques for EDA:

### 1. Principal Component Analysis (PCA)

PCA reduces dimensionality by identifying new axes that capture the most variance in the data. It helps identify patterns and relationships between variables.

- Example:
  - Suppose you're analyzing customer satisfaction ratings based on three features: price, quality, and service.
  - PCA can help identify which combination of these features best explains overall satisfaction.

### 2. Factor Analysis (FA)

FA is similar to PCA but focuses on identifying underlying factors that explain the correlations between variables.

- Example:
  - Suppose you're analyzing survey data with multiple questions related to customer preferences (e.g., product design, packaging, price).
  - FA can help identify which underlying factors (e.g., sustainability concerns, convenience priorities) drive these preferences.

### 3. Cluster Analysis

Cluster analysis groups similar observations into clusters based on their feature values.

- Example:
  - Suppose you're analyzing data from customers with multiple purchases.
  - Cluster analysis can help identify customer segments based on purchase history, location, and other characteristics.

#### **4. Canonical Correlation Analysis (CCA)**

CCA measures the correlation between two sets of variables (e.g., predictor and response variables).

- Example:
  - Suppose you're analyzing data from an experiment with multiple treatments (predictors) affecting a single outcome variable.
  - CCA can help identify which predictors are most strongly related to the outcome variable.

#### **5. Multidimensional Scaling (MDS)**

MDS visualizes the similarity or dissimilarity between observations as points in a lower-dimensional space.

- Example:
  - Suppose you're analyzing customer reviews with multiple features (e.g., sentiment, topic).
  - MDS can help visualize how similar or different these reviews are based on their feature values.

These multivariate analysis techniques for EDA can help you identify patterns, relationships, and correlations among your data variables. By applying these methods, you can gain a deeper understanding of your data and inform business decisions accordingly.