## ASSIGNMENT-4

## MODULE-1(INDIVIDUAL TASK)

## CORRELATION EXPLORATION

**Correlation exploration** is the process of analyzing relationships between variables to determine whether and how strongly they are related. It is a key step in **Exploratory Data Analysis (EDA)** used in statistics, data science, and machine learning.

**Positive Correlation**

- When one variable increases, the other also increases.
- Example: Study time ↑ → Exam score ↑

**Negative Correlation**

- When one variable increases, the other decreases.
- Example: Price ↑ → Demand ↓

**No Correlation**

- No relationship between variables.
- Example: Shoe size and intelligence

Common methods:

- **Pearson Correlation** (linear relationship)
- **Spearman Rank Correlation** (non-linear/monotonic relationship)
- **Kendall's Tau**

# Steps in Correlation Exploration

Identify variables
Clean the data
Create scatter plots
Calculate correlation coefficient
Interpret results
Check statistical significance

**Why Correlation Exploration is Important**

- Identifies relationships between features
- Helps in feature selection for ML models
- Detects multicollinearity
- Supports decision-making
- Helps generate hypotheses

**Example in Real Life**

In business analytics:

- Marketing vs Sales revenue
- Website visits vs Product purchases
- Temperature vs Electricity consumption

Correlation exploration is an essential analytical technique used to understand relationships between variables. It plays a critical role in statistics, data science, and machine learning by helping analysts identify patterns and make informed decisions.

If you want, I can also provide:

- Short exam notes
- Numerical example with calculation
- Python code example
- Differences between correlation and regression

Correlation exploration is the process of analyzing and understanding the relationship between two or more variables in a dataset. It helps determine whether changes in one variable are associated with changes in another and measures the strength and direction of that relationship. This is commonly done using statistical measures such as the correlation coefficient, which ranges from –1 to +1, indicating negative, no, or positive correlation. Correlation exploration is an important step in data analysis and machine learning because it helps identify patterns, select relevant features, and detect multicollinearity. However, it is important to remember that correlation does not imply causation, meaning that a relationship between variables does not necessarily mean that one causes the other.

**Correlation exploration** is a statistical and analytical technique used to quantify and interpret the **degree and direction of interdependence** between two or more variables. Beyond simple linear associations, advanced correlation analysis can capture **non-linear relationships**, **partial correlations**, and **conditional dependencies**, which are critical in high-dimensional datasets and predictive modeling.

**Pearson Correlation**

- Measures **linear relationships** between continuous variables.
- Values: $r \in [-1,1]$
- Assumes normality and homoscedasticity.

**Spearman Rank Correlation**

- Measures **monotonic relationships**, not restricted to linearity.
- Uses ranked values instead of raw data.
- Less sensitive to outliers.

**Kendall's Tau**

- Measures ordinal association between variables.
- Focuses on the number of concordant and discordant pairs.
- Useful for small datasets or with many tied ranks.

**Partial Correlation**

- Measures the correlation between two variables **while controlling for one or more additional variables**.
- Useful for multivariate analysis and causal inference.

**Distance Correlation**

- Captures **both linear and non-linear associations**.
- Values: 0 (no dependence) to 1 (perfect dependence).
- Advanced method often used in machine learning feature selection.