PART A – THEORY ANSWERS

Explain Mean, Median, Mode (Income context)

Mean (Average):

In loan analysis, mean income helps bank understand overall earning capacity of applicants.

Median:
Median is the middle income when all incomes are arranged in ascending order.
It is useful when income distribution has outliers.

Mode:
Mode is the most frequently occurring income value in dataset.

Standard Deviation and Variance (Loan Amount)

Variance:
Variance measures how far loan amounts spread from the mean.

Standard Deviation:

If standard deviation is high → Loan amounts vary a lot.
If low → Loan amounts are similar.

What is a Random Variable?

A random variable is a numerical outcome of a random experiment.

Example from dataset:

- Credit_Score of a randomly selected customer.
- Loan_Amount can also be considered a continuous random variable.

Types:

- Discrete (e.g., Loan_Term)
- Continuous (e.g., Income)

Define Conditional Probability (Loan Default Context)

Conditional probability is the probability of an event occurring given that another event has already occurred.

Example:

$$P(\text{Default} \mid \text{Credit Score} < 600)$$

Bayes' Theorem (Banking Application)

Bayes' theorem updates probability based on new evidence.

In banking:

- A = Customer defaults
- B = Customer has low credit score

Bank can update default probability when new financial data is received.

Empirical vs Theoretical Probability

Empirical Probability:
Based on observed data.

$$P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total observations}}$$

Example:
If 800 out of 5000 customers defaulted,

$$P(\text{Default}) = \frac{800}{5000}$$

Theoretical Probability:
Based on mathematical assumption or model.

Empirical = Data based
Theoretical = Formula based

Poisson Distribution (Business Example)

Poisson distribution models number of events in fixed interval of time.

Business Example:
Number of loan applications received per day.

Eigenvalues and Eigenvectors (Data Analysis)

Eigenvectors represent direction of maximum variance.
Eigenvalues represent magnitude of that variance.

They are used in:

- Principal Component Analysis (PCA)
- Dimensionality reduction
- Feature extraction in financial datasets

In loan analysis, PCA can reduce multiple financial variables into fewer components.