UNIT-1

What does EDA stand for in the context of data analysis?

a) Early Data Assessment

b) Exploratory Data Analysis

c) Efficient Data Analytics

d) Extended Data Accumulation

Answer: b) Exploratory Data Analysis

What is the primary goal of Exploratory Data Analysis (EDA)?

a) To make predictions

b) To summarize data

c) To perform hypothesis testing

d) To clean data

Answer: b) To summarize data

Public data is typically:

a) Accessible to anyone

b) Restricted to authorized personnel

c) Encrypted and secure

d) Stored offline

Answer: a) Accessible to anyone

Private data is usually:

a) Easily downloadable from the internet

b) Openly shared with the public

c) Restricted to certain individuals or organizations

d) Available without any access control

Answer: c) Restricted to certain individuals or organizations

What is web scraping?

a) A method for securing data

b) A technique for extracting information from websites

c) A type of data visualization

d) A form of data encryption

Answer: b) A technique for extracting information from websites

Which of the following is not a popular programming language for web scraping?

a) Python

b) JavaScript

c) Ruby

d) HTML

Answer: d) HTML

Which data type is used to represent whole numbers in Python?

a) Float

b) Integer

c) String

d) Boolean

Answer: b) Integer

What data type is suitable for storing text data?

a) Integer

b) Float

c) String

d) Boolean

Answer: c) String

What does it mean to merge two data frames in pandas?

a) Add new columns to one data frame

b) Combine rows from two data frames

c) Remove duplicate rows

d) Split a data frame into two

Answer: b) Combine rows from two data frames

What pandas function is used to drop columns from a data frame?

a) delete()

b) remove_column()

c) drop()

d) exclude()

Answer: c) drop()

What is imputation in data analysis?

a) A method for creating missing values

b) A technique for replacing missing values with estimated or calculated values

c) A method for removing all missing values from the dataset

d) A type of visualization technique

Answer: b) A technique for replacing missing values with estimated or calculated values

Which pandas function is used to remove rows with missing values in a DataFrame?

a) remove_null()

b) dropna()

c) fillna()

d) missing_data()

Answer: b) dropna()

Outliers are data points that:

a) Are always incorrect and should be removed

b) Fall within the normal range of data

c) Deviate significantly from the rest of the data

d) Are missing values

Answer: c) Deviate significantly from the rest of the data

What is a common technique for handling outliers?

a) Deleting all rows with outliers

b) Replacing outliers with the median value

c) Ignoring outliers and not taking any action

d) Duplicating outliers to emphasize their impact

Answer: b) Replacing outliers with the median value

Standardization in data preprocessing involves:

a) Converting data to a different data type

b) Scaling data to have a mean of 0 and a standard deviation of 1

c) Removing all missing values

d) Replacing outliers with the maximum value in the dataset

Answer: b) Scaling data to have a mean of 0 and a standard deviation of 1

Which library in Python is commonly used for standardizing data?

a) Pandas

b) Numpy

c) Scikit-Learn

d) Matplotlib

Answer: c) Scikit-Learn

Invalid values in a dataset are typically:

a) Values that do not conform to the data type or expected range

b) Missing values

c) Outliers

d) Duplicate values

Answer: a) Values that do not conform to the data type or expected range

What is a common approach to handling invalid values?

a) Ignoring them and leaving them as is

b) Replacing them with zeros

c) Removing the entire row with invalid values

d) Adding them to the mean of the dataset

Answer: c) Removing the entire row with invalid values

Filtering data in a DataFrame involves:

a) Removing all rows and columns

b) Selecting a subset of rows or columns based on specified conditions

c) Rearranging the rows and columns randomly

d) Sorting the data in ascending order

Answer: b) Selecting a subset of rows or columns based on specified conditions

Which pandas function is commonly used for filtering data?

a) filter()

b) select()

c) query()

d) subset()

Answer: c) query()

UNIT-2

What is the primary goal of univariate analysis?

a) Explore relationships between variables

b) Analyze multiple variables together

c) Examine a single variable in isolation

d) Predict future outcomes

Answer: c) Examine a single variable in isolation

Which of the following is NOT a common statistical measure used in univariate analysis?

a) Mean

b) Median

c) Correlation coefficient

d) Standard Deviation

Answer: c) Correlation coefficient

In categorical unordered univariate analysis, what type of data is typically examined?

a) Continuous data

b) Ordinal data

c) Nominal data

d) Time-series data

Answer: c) Nominal data

What is a common visualization technique for categorical unordered data?

a) Histogram

b) Box plot

c) Bar chart

d) Scatter plot

Answer: c) Bar chart

Ordinal data is often analyzed using which of the following?

a) Pie chart

b) Box plot

c) Line chart

d) Heatmap

Answer: b) Box plot

What does ordinal data represent?

a) Categories with no specific order

b) Categories with a natural order or ranking

c) Numeric values

d) Binary data

Answer: b) Categories with a natural order or ranking

What does the median represent in a set of numerical data?

a) The most frequently occurring value

b) The middle value when data is sorted

c) The average value

d) The highest value

Answer: b) The middle value when data is sorted

Which statistic provides a measure of the spread or variability of numerical data?

a) Mean

b) Median

c) Variance

d) Mode

Answer: c) Variance

What type of analysis explores the relationship between two numerical variables?

a) Categorical analysis

b) Numeric-categorical analysis

c) Numeric-numeric analysis

d) Multivariate analysis

Answer: c) Numeric-numeric analysis

In a scatter plot, what does each point represent?

a) A category

b) A pair of numerical values

c) A bar chart

d) A single value

Answer: b) A pair of numerical values

What does correlation measure?

a) A cause-and-effect relationship

b) The strength and direction of a linear relationship between two variables

c) The variability within a single variable

d) The mean of a dataset

Answer: b) The strength and direction of a linear relationship between two variables

What is the key difference between correlation and causation?

a) Correlation implies causation

b) Causation implies correlation

c) Correlation suggests a relationship, but causation implies a cause-and-effect relationship

d) There is no difference between the two terms

Answer: c) Correlation suggests a relationship, but causation implies a cause-and-effect relationship

In numerical-categorical analysis, what type of analysis is performed?

a) Examining the relationship between two numerical variables

b) Analyzing a single numerical variable

c) Analyzing a single categorical variable

d) Exploring the relationship between a numerical and a categorical variable

Answer: d) Exploring the relationship between a numerical and a categorical variable

What visualization technique is commonly used for numerical-categorical analysis?

a) Box plot

b) Scatter plot

c) Bar chart

d) Histogram

Answer: a) Box plot

What type of analysis explores relationships between two categorical variables?

a) Categorical unordered analysis

b) Categorical ordered analysis

c) Multivariate analysis

d) Categorical-categorical analysis

Answer: d) Categorical-categorical analysis

What is a common visualization technique for categorical-categorical analysis?

a) Scatter plot

b) Heatmap

c) Line chart

d) Histogram

Answer: b) Heatmap

Multivariate analysis involves the examination of:

a) A single variable

b) Multiple variables together

c) Only categorical data

d) Data collected over time

Answer: b) Multiple variables together

What is the goal of multivariate analysis?

a) To examine the relationship between two variables

b) To explore the distribution of a single variable

c) To understand complex interactions among multiple variables

d) To perform univariate analysis

Answer: c) To understand complex interactions among multiple variables

What is a common approach to handling missing values in a dataset?

a) Ignore them and proceed with analysis

b) Remove all rows with missing values

c) Replace them with arbitrary values

d) Impute or fill them with appropriate values

Answer: d) Impute or fill them with appropriate values

Data visualization is primarily used for:

a) Hiding data from viewers

b) Communicating information and patterns in data

c) Conducting hypothesis testing

d) Storing data in a database

Answer: b) Communicating information and patterns in data

Unit-3

What is a permutation?

a) A way to arrange objects in a specific order

b) A way to select objects without regard to order

c) A type of probability distribution

d) A random variable

Answer: a) A way to arrange objects in a specific order

What is a combination?

a) A way to arrange objects in a specific order

b) A way to select objects without regard to order

c) A type of probability distribution

d) A random variable

Answer: b) A way to select objects without regard to order

What does probability measure?

a) Certainty

b) Possibility

c) Likelihood

d) Determinism

Answer: c) Likelihood

In probability, what is the complement of an event?

a) An unrelated event

b) The event itself

c) The event not happening

d) An independent event

Answer: c) The event not happening

What is a mutually exclusive event?

a) Two events that always occur together

b) Two events that can never occur together

c) Two events that are unrelated

d) Two events that are independent

Answer: b) Two events that can never occur together

What is a dependent event?

a) An event that is not influenced by other events

b) An event that is influenced by other events

c) An event that has a high probability of occurring

d) An event that is certain to happen

Answer: b) An event that is influenced by other events

What does the Addition Rule of Probability state?

a) The probability of both events A and B occurring is the sum of their individual probabilities.

b) The probability of both events A and B occurring is the product of their individual probabilities.

c) The probability of either event A or event B occurring is the sum of their individual probabilities.

d) The probability of neither event A nor event B occurring is the product of their individual probabilities.

Answer: c) The probability of either event A or event B occurring is the sum of their individual probabilities.

What are mutually exclusive events?

a) Events that have no intersection

b) Events that always occur together

c) Events that are dependent

d) Events that are not related to each other

Answer: a) Events that have no intersection

What does the Multiplication Rule of Probability state?

a) The probability of both events A and B occurring is the sum of their individual probabilities.

b) The probability of both events A and B occurring is the product of their individual probabilities.

c) The probability of either event A or event B occurring is the sum of their individual probabilities.

d) The probability of neither event A nor event B occurring is the product of their individual probabilities.

Answer: b) The probability of both events A and B occurring is the product of their individual probabilities.

What are independent events?

a) Events that have no intersection

b) Events that always occur together

c) Events that are dependent on each other

d) Events that are not related to each other

Answer: d) Events that are not related to each other

What is the probability of an impossible event?

a) 1

b) 0

c) 0.5

d) -1

Answer: b) 0

What is the probability of a certain event?

a) 1

b) 0

c) 0.5

d) -1

Answer: a) 1

What is a random variable?

a) A variable that is always constant

b) A variable that can take on random values

c) A variable that is determined by the experimenter

d) A variable that cannot be measured

Answer: b) A variable that can take on random values

What is the difference between a discrete random variable and a continuous random variable?

a) A discrete random variable can take on any value within a range, while a continuous random variable can only take on specific values.

b) A discrete random variable can only take on specific values, while a continuous random variable can take on any value within a range.

c) They are the same; there is no difference between them.

d) Discrete random variables are used in experiments, while continuous random variables are used in simulations.

Answer: b) A discrete random variable can only take on specific values, while a continuous random variable can take on any value within a range.

What is a probability distribution?

a) A list of random numbers

b) A mathematical function that describes the likelihood of each possible outcome of a random variable

c) A set of rules for conducting experiments

d) A measure of central tendency

Answer: b) A mathematical function that describes the likelihood of each possible outcome of a random variable

In a probability distribution, what must the sum of all probabilities equal?

a) 0

b) 1

c) 2

d) It can be any value

Answer: b) 1

What does the expected value (mean) of a random variable represent?

a) The most frequently occurring value

b) The median value

c) The long-term average value over many trials

d) The highest value

Answer: c) The long-term average value over many trials

How is the expected value calculated for a discrete random variable?

a) By summing the product of each possible value and its corresponding probability

b) By taking the median value

c) By finding the most frequent value

d) By taking the square root of the variance

Answer: a) By summing the product of each possible value and its corresponding probability

What is a binomial distribution?

a) A probability distribution for continuous random variables

b) A probability distribution for discrete random variables with a fixed number of trials

c) A probability distribution for discrete random variables with infinite trials

d) A type of normal distribution

Answer: b) A probability distribution for discrete random variables with a fixed number of trials

In a binomial distribution, what are the two parameters that define the distribution?

a) Mean and median

b) Variance and standard deviation

c) Number of trials (n) and probability of success (p)

d) Skewness and kurtosis

Answer: c) Number of trials (n) and probability of success (p)