Section – A (20 questions)

1. **What is the goal of Exploratory Data Analysis (EDA)?**

a) Cleaning raw data b) Understanding data patterns

c) Normalizing data values d) Applying machine learning algorithms

**2. Which library is commonly used for creating box plots in Python?**

a) Seaborn b) Matplotlib c) Plotly d) NumPy

**3. What library specializes in creating interactive visualizations in Python?**

a) Matplotlib b) Seaborn c) Plotly d) Pandas

**4. What is the primary purpose of data visualization best practices?**

a) Making visualizations more complex b) Enhancing aesthetics

c) Improving communication of data insights d) Reducing visualization variety

**5. What are the basic components of supervised learning?**

a) Features and testing b) Labels and training c) Features and labels d) Training and testing

**6. What is the primary purpose of the scikit-learn library in Python?**

a) Web development b) Machine learning algorithms c) Data visualization d) Text processing

**7. What distinguishes Artificial Intelligence from traditional software?**

a) AI is self-aware b) AI learns and adapts

c) AI is always rule-based d) AI cannot perform complex tasks

**8. In Machine Learning, what does the term "label" refer to?**

a) Input data b) Output prediction c) Model parameters d) Training set

9. **Which Python library is commonly used for scientific computing and mathematical operations in AI and ML?**

a) TensorFlow b) PyTorch c) NumPy d) Keras

**10. In Pandas, what is the primary purpose of the DataFrame?**

a) Storing one-dimensional labeled data b) Performing matrix operations

c) Organizing data in tabular form d) Creating visualizations

**11. How can you write a DataFrame to a CSV file in Pandas?**

a) df.save\_csv() b) df.write\_csv() c) df.to\_csv() d) df.export\_csv()

**12. What does the method `df.describe()` provide in Pandas?**

a) Basic statistics of numeric columns b) Total number of rows and columns

c) A summary of non-numeric columns d) Data types of each column

**13. What is the purpose of the standard deviation in descriptive statistics?**

a) Measure of central tendency b) Measure of data spread

c) Median value d) Measure of skewness

**14. What does the term "skewness" refer to in descriptive statistics?**

a) Data distribution asymmetry b) Mean value c) Data variability d) Standard deviation

**15. Which Python library is commonly used for creating interactive statistical graphics?**

a) Plotly b) Seaborn c) Matplotlib d) Pandas

**16. In Matplotlib, what is the purpose of the `plt.scatter()` function?**

a) Create a bar plot b) Create a scatter plot c) Generate a line plot d) Display a histogram

**17. Which library is known for its high-level interface for drawing attractive statistical graphics?**

a) Matplotlib b) Seaborn c) Plotly d) Pandas

**18. What is the advantage of using Plotly for data visualization?**

a) Simplicity b) High interactivity c) Lightweigh d) Limited chart types

**19. What is the key difference between supervised and unsupervised learning?**

a) The need for labeled data in supervised learning

b) The involvement of neural networks in unsupervised learning.

c) The use of regression models in supervised learning**.**

d) The absence of training data in unsupervised learning

**20. What is the purpose of the training set in machine learning?**

a) To test the model b) To evaluate the accuracy c) To train the model d) To validate the results

Section – B (15 questions)

**21. Given the following dataset: `[10, 15, 20, 25, 30]`, what is the mean?**

a) 18 b) 20 c) 22.5 d) 25

**22. Calculate the median for the dataset: `[5, 8, 12, 15, 20]`.**

a) 8 b) 12 c) 15 d) 20

**23. If a dataset has a variance of 16, what is the standard deviation?**

a) 2 b) 4 c) 8 d) 16

**24. Consider a DataFrame with the following column values: `[10, 20, 30, 40, 50]`. What is the result of doubling each value?**

a) `[10, 20, 30, 40, 50]` b) `[20, 40, 60, 80, 100]` c) `[5, 10, 15, 20, 25]` d) `[1, 2, 3, 4, 5]`

**25. If you have a DataFrame with three columns 'A', 'B', and 'C', and you want to find the sum of each column, what Pandas function would you use?**

a) `df.total()` b) `df.sum()` c) `df.add()` d) `df.aggregate()`

**26. In a classification problem, if a model correctly predicts 80 out of 100 instances, what is its accuracy?**

a) 0.8 b) 0.2 c) 0.6 d) 0.4

**27. For a regression model, if the mean squared error (MSE) is 25, what is the root mean squared error (RMSE)?**

a) 5 b) 25 c) 125 d) 625

**28. Calculate the variance of the dataset: `[12, 15, 18, 22, 25]`.**

a) 9.5 b) 12.5 c) 15.5 d) 18.5

**29. Given a dataset with the values `[8, 10, 12, 15, 20]`, calculate the interquartile range (IQR).**

a) 3 b) 5 c) 7 d) 10

**30. Consider a DataFrame with two columns 'Sales' and 'Expenses'. Calculate the profit by subtracting Expenses from Sales. What is the profit for the row with Sales=150 and Expenses=80?**

a) 50 b) 70 c) 100 d) 120

**31. If you have a DataFrame with a column 'Population' representing the population of cities, and you want to calculate the percentage of the population for each city relative to the total population, what Pandas function would you use?**

a) `df.percent()` b) `df.divide()` c) `df.apply()` d) `df.divide(df['Population'].sum())`

**32. In a binary classification problem, a model predicts 120 instances as positive, but only 80 of them are actually positive. Calculate precision.**

a) 0.33 b) 0.50 c) 0.66 d) 0.75

**33. For a regression model, if the coefficient of determination (R-squared) is 0.85, what percentage of the variance in the dependent variable is explained by the model?**

a) 15% b) 35% c) 65% d) 85%

**34. Which scikit-learn module is commonly used for model evaluation and parameter tuning?**

a) `sklearn.evaluate` b) `sklearn.tune` c) `sklearn.metrics` d) `sklearn.model`

**35. For a binary classification problem, if the true positive rate (sensitivity) is 0.80 and the false positive rate is 0.10, what is the specificity?**

a) 0.90 b) 0.80 c) 0.70 d) 0.60