ADS – Assignment 2 ("Understanding Data" Assignment 1 Quiz Questions")

Chaitanya Inamdar

1. What is the primary focus of data visualization?
    - a) To format the data
    - b) To extract insights from the dataset
    - c) To clean the data
    - d) To remove outliers

Answer: b

2. Which of the following methods normalizes the data in the notebook?
    - a) fillna()
    - b) StandardScaler()
    - c) isnull()
    - d) pairplot()

Answer: b

3. What is the purpose of the fillna() function?
    - a) It fills missing data with specific values, such as mean or median.
    - b) It removes duplicates in the dataset.
    - c) It drops rows with missing data.
    - d) It normalizes the data.

Answer: a

4. Which of the following features is not present in the Iris dataset?
    - a) Sepal length
    - b) Petal width
    - c) Sepal height
    - d) Petal length

Answer: c

5. How does normalization affect machine learning algorithms?
    - a) It removes outliers from the data.
    - b) It ensures that features are on the same scale, which can improve performance.
    - c) It fills missing data with the mean value.
    - d) It categorizes continuous data.

Answer: b

6. What is the use of sns.pairplot() in Seaborn?
    - a) To plot relationships between categorical variables
    - b) To plot pairwise relationships between numerical variables
    - c) To create histograms
    - d) To create line plots

Answer: b

7. Why do we need to handle missing data in a dataset?
    - a) Missing data can distort the analysis and introduce bias.
    - b) Missing data increases the number of features.
    - c) Missing data speeds up computation.
    - d) Missing data makes the dataset smaller.

Answer: a

8. Which function in Pandas is used to fill missing values in a dataset?
    - a) dropna()
    - b) fillna()
    - c) replace()
    - d) isnull()

Answer: b

9. What does a pair plot typically display?
    - a) Individual data points
    - b) Relationships between pairs of numerical variables
    - c) Missing data patterns
    - d) Categorical feature distributions

Answer: b

10. Why is scaling data important before applying machine learning models?
    - a) It removes missing data.
    - b) It ensures features are on the same scale, which helps improve model accuracy.
    - c) It reduces the number of features.
    - d) It speeds up the loading of data.

Answer: b

11. What is the purpose of plt.title() in Matplotlib?
    - a) To add a label to the x-axis
    - b) To add a label to the y-axis
    - c) To display the title of the plot
    - d) To save the plot to a file

Answer: c

12. In the Iris dataset, which feature is most predictive of species classification based on the pair plot?
    - a) Sepal width
    - b) Petal length
    - c) Sepal length
    - d) Petal width

Answer: b

13. Which function allows you to normalize the dataset in Python?
    - a) dropna()
    - b) StandardScaler()
    - c) fillna()
    - d) head()

Answer: b

14. What does the sns.pairplot() function help visualize?
    - a) Distribution of a single feature
    - b) Pairwise relationships between all variables
    - c) The mean of all features
    - d) Categorical variables only

Answer: b

15. Why are notebooks like Jupyter and Google Colab commonly used in data science?
    - a) They support only Python programming.
    - b) They allow for the integration of code, visualization, and documentation in one place.
    - c) They can only be used for small datasets.
    - d) They are faster than traditional IDEs.

Answer: b

16. What does the Pandas function .head() return?
    - a) The last 5 rows of the dataset
    - b) The first 5 rows of the dataset
    - c) The column names
    - d) The number of rows in the dataset

Answer: b

17. In data cleaning, what is the purpose of imputation?
    - a) To remove outliers
    - b) To fill missing values with substitute values
    - c) To normalize features
    - d) To reduce the size of the dataset

Answer: b

18. Why is data visualization important in data analysis?
    - a) It helps in scaling the data.
    - b) It allows analysts to identify trends, patterns, and outliers that are not easily visible in raw data.
    - c) It speeds up data loading.
    - d) It reduces missing data.

Answer: b

19. What does StandardScaler() do?
    - a) It normalizes data by centering it with a mean of 0 and standard deviation of 1.
    - b) It fills missing values.
    - c) It drops duplicate rows.
    - d) It adds more features to the dataset.

Answer: a

20. What is the primary function of Seaborn in this chapter?
    - a) Data scaling
    - b) Data visualization
    - c) Data imputation
    - d) Data cleaning

Answer: b

21. How do you handle missing data in the notebook?
    - a) Using the dropna() function
    - b) Using the fillna() function to fill missing values with the mean
    - c) Ignoring the missing values
    - d) Manually adding the missing values

Answer: b

22. Why is it necessary to scale features in the Iris dataset before building models?
    - a) To visualize the data
    - b) To ensure that all features have comparable scales, which improves model performance
    - c) To remove outliers
    - d) To reduce the number of features

Answer: b

23. What does the plt.hist() function display?
    - a) A histogram of the data
    - b) A scatter plot of the data
    - c) A line plot of the data
    - d) A bar chart of the data

Answer: a

24. Why are Jupyter Notebooks preferred in data analysis and machine learning workflows?
    - a) They allow only for coding.
    - b) They integrate coding, documentation, and real-time visualization into one environment.
    - c) They are faster than any other IDE.
    - d) They are used only for big data processing.

Answer: b

25. In a pair plot, what do the diagonal plots represent?
    - a) Pairwise correlations
    - b) The distribution of each variable
    - c) Missing data patterns
    - d) The median of each feature

Answer: b