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

**MCQs**

**1. What is a key feature of Jupyter Notebooks?**

* a) It supports only Python programming.
* b) It allows for interactive coding, documentation, and visualization.
* c) It is only used for machine learning.
* d) It is a command-line tool.

**Answer**: b

**2. What format does the Iris dataset have?**

* a) JSON
* b) XML
* c) Structured
* d) Unstructured

**Answer**: c

**3. What does the fillna() function do in Pandas?**

* a) Removes missing values.
* b) Fills missing values with the mean or other specified values.
* c) Normalizes the dataset.
* d) Drops duplicate rows.

**Answer**: b

**4. Which library is primarily used for data visualization in this chapter?**

* a) Matplotlib
* b) TensorFlow
* c) PyTorch
* d) Pandas

**Answer**: a

**5. What is the purpose of using StandardScaler in preprocessing?**

* a) To handle missing data.
* b) To normalize data by centering it to have a mean of 0 and a standard deviation of 1.
* c) To categorize data.
* d) To remove outliers.

**Answer**: b

**6. What kind of plots does Seaborn offer for understanding feature relationships?**

* a) Heatmaps
* b) Line plots
* c) Pair plots
* d) Bar charts

**Answer**: c

**7. In data analysis, what is feature normalization?**

* a) Transforming numerical values to a common scale.
* b) Filling missing values with a constant.
* c) Removing outliers from the dataset.
* d) Converting categorical variables into numerical values.

**Answer**: a

**8. What does the function pairplot() in Seaborn visualize?**

* a) Correlation between different datasets.
* b) Pairwise relationships between variables.
* c) Categorical data distributions.
* d) Null value distributions.

**Answer**: b

**9. Why is data cleaning considered crucial in data science?**

* a) To reduce the file size.
* b) To ensure data is accurate and suitable for analysis.
* c) To visualize data faster.
* d) To increase the number of features.

**Answer**: b

**10. Which Python library is primarily used for data manipulation and analysis?**

* a) Numpy
* b) Matplotlib
* c) Pandas
* d) Seaborn

**Answer**: c

**11. What does .head() in Pandas do?**

* a) Shows the first few rows of a dataset.
* b) Displays the column names.
* c) Removes missing values.
* d) Summarizes the dataset.

**Answer**: a

**12. What is the primary advantage of using notebooks in data analysis?**

* a) They are faster than traditional scripts.
* b) They allow for real-time collaboration, visualization, and documentation in one place.
* c) They provide built-in machine learning models.
* d) They are only used for big data analysis.

**Answer**: b

**13. What type of chart is most useful to understand the distribution of a single numerical feature?**

* a) Line chart
* b) Histogram
* c) Pie chart
* d) Heatmap

**Answer**: b

**14. In Seaborn, which function is used to plot pairwise relationships between variables?**

* a) heatmap()
* b) scatter()
* c) pairplot()
* d) lineplot()

**Answer**: c

**15. Why is normalization important in machine learning?**

* a) It helps in handling missing values.
* b) It ensures that features have comparable scales.
* c) It reduces the computational complexity.
* d) It increases the number of features.

**Answer**: b

**16. What is the purpose of using plt.show() in Matplotlib?**

* a) To save the plot to a file.
* b) To display the plot on the screen.
* c) To clear the plot.
* d) To modify the plot.

**Answer**: b

**17. What type of dataset is the Iris dataset?**

* a) Categorical
* b) Structured
* c) Time series
* d) Text

**Answer**: b

**18. Which of the following is used to check for missing values in a Pandas DataFrame?**

* a) isnull()
* b) dropna()
* c) fillna()
* d) mean()

**Answer**: a

**19. In Python, what is the purpose of dropna()?**

* a) Fills missing values with a specified value.
* b) Drops rows with missing values.
* c) Normalizes the dataset.
* d) Adds missing values.

**Answer**: b

**20. What does the StandardScaler() do to the data?**

* a) Replaces missing values with the median.
* b) Converts categorical data to numerical values.
* c) Normalizes data to have zero mean and unit variance.
* d) Removes outliers.

**Answer**: c

**21. Which library is used to load the Iris dataset in this chapter?**

* a) Numpy
* b) Matplotlib
* c) Seaborn
* d) TensorFlow

**Answer**: c

**22. What is the output of iris.fillna(iris.mean())?**

* a) It removes rows with missing values.
* b) It fills missing values with the column mean.
* c) It drops duplicate rows.
* d) It replaces categorical data.

**Answer**: b

**23. What is one reason to use StandardScaler() before building machine learning models?**

* a) It helps reduce the number of features.
* b) It ensures that all features are on the same scale, improving model performance.
* c) It improves visualization of categorical data.
* d) It simplifies the dataset for faster loading.

**Answer**: b

**24. Which visualization technique can show relationships between multiple variables?**

* a) Bar plot
* b) Pair plot
* c) Line plot
* d) Histogram

**Answer**: b

**25. Why is feature engineering important?**

* a) It ensures features are in the right format for modeling.
* b) It removes missing values from the dataset.
* c) It decreases the size of the dataset.
* d) It improves data visualization.

**Answer**: a