**Skill Stork International School**

**Worksheet Class: CBSE XI Subject : AI**

#### ****Section A: Data Cleaning****

1. **Data Types Identification**
   * Q1: Given a dataset, identify and list the data types (e.g., integer, float, string) of each column in Python.
2. **Handling Missing Data**
   * Q2: Write code to check for missing values in each column and calculate the percentage of missing values.
   * Q3: Write a code snippet to fill missing numerical data in a dataset with the mean of that column.
   * Q4: Demonstrate how to remove rows with missing data in Python.
3. **Duplicate Data**
   * Q5: Write a function to check for duplicate entries in a dataset and count the number of duplicates.
   * Q6: Write code to drop duplicate rows while keeping the first instance of each duplicate.
4. **Data Standardization**
   * Q7: Write code to convert a column containing text values (e.g., "Yes", "No") to numerical values (1, 0).
   * Q8: Given a column with categorical data (e.g., color names), write code to convert these categories into one-hot encoded columns.
5. **Outlier Detection and Handling**
   * Q9: Write code to detect outliers in a column using the IQR (Interquartile Range) method.
   * Q10: Write a Python function to replace outliers in a dataset with the median of that column.
6. **Data Scaling and Normalization**
   * Q11: Write a code snippet to standardize a dataset using z-score normalization.
   * Q12: Write code to normalize a column of data to be between 0 and 1.

#### ****Section B: Data Analysis****

1. **Descriptive Statistics**
   * Q13: Write code to calculate and display the mean, median, mode, and standard deviation of a given numerical column in a dataset.
2. **Data Visualization**
   * Q14: Write code to plot a histogram for a numerical column in a dataset.
   * Q15: Write a code snippet to create a boxplot for a specified numerical column to show the distribution and detect outliers.
   * Q16: Use code to create a scatter plot to examine the relationship between two numerical columns.
3. **Correlation and Covariance**
   * Q17: Write code to calculate the correlation between two columns and interpret the results.
   * Q18: Demonstrate how to calculate covariance for two columns and interpret the results.
4. **Aggregating Data**
   * Q19: Write code to group a dataset by a categorical column and calculate the average value for each group.
   * Q20: Write a Python function to calculate the total, minimum, and maximum values of a numerical column, grouped by another column (e.g., city or category).

*Link for Dataset:* *https://github.com/ATinyLearner/AI-CBSE/blob/main/studentPerformance.csv*

**General Instruction for students:**

* For codes , write the code on code editor
* Take screenshot of the code snippet
* Paste it on a solution docx file in your computer
* Send your solution docx to avikghosh@skillstork.org
* Submission Date & Time : 3rd November 2024 , 7.00pm