**Mock Test I**

**MCQs**

1. Which of the following libraries in Python is commonly used for numerical computing and working with arrays?

A) Pandas

B) Seaborn

C) NumPy

D) Matplotlib

*NumPy provides powerful n-dimensional arrays (ndarrays) that are more efficient and convenient than Python's built-in lists for numerical operations.*

*NumPy is implemented in C and optimized for performance, making it much faster for numerical computations compared to standard Python operations.*

*It includes a vast array of mathematical functions for operations on arrays, including linear algebra, statistics, and Fourier transforms.*

*NumPy supports broadcasting, allowing operations on arrays of different shapes, which simplifies coding and enhances performance.*

*NumPy serves as the foundation for many other libraries in the scientific computing ecosystem, such as SciPy, Pandas, and Matplotlib, making it essential for data science and analysis.*

*NumPy arrays require less memory than lists and are more efficient for large datasets, making them ideal for handling large-scale numerical data.These features make NumPy the go-to library for numerical computing in Python.*

1. Which of the following methods is used to remove missing data from a Pandas DataFrame?

A) dropna()

B) fillna()

C) replace()

D) isnull()

*The dropna() method is specifically designed to remove missing values (NaNs) from a Pandas DataFrame or Series. It allows for flexibility in how missing data is handled, such as specifying whether to drop rows or columns and setting thresholds for dropping. In data cleaning, it is a common step to remove incomplete data entries to ensure the integrity of analyses.*

*Contrast with Other Options:*

*fillna(): This method is used to fill missing values with a specified value, not to remove them.*

*replace(): This method is used to replace specified values in a DataFrame, not specifically for handling NaNs.*

*isnull(): This method checks for missing values and returns a boolean DataFrame, but it does not remove them.*

1. In a Seaborn pairplot(), what does the diagonal plot show by default?

A) Scatter plot

B) KDE (Kernel Density Estimate)

C) Box plot

D) Histogram

*In Seaborn's pairplot(), the diagonal plots represent the distribution of each variable. By default, this is displayed as a histogram. Histograms effectively visualize the frequency distribution of a single variable, making them suitable for understanding the underlying distribution of the data. The histograms on the diagonal allow for quick visual comparisons of the distributions of different variables, which is useful in exploratory data analysis. Although the default is a histogram, users can customize the diagonal plots to display other types of plots, such as KDE, by using the diag\_kind parameter. This reinforces that the default choice is indeed a histogram.*

1. What is the output of the following code?

X = **12**

**if** (X>**10** & X<**15**):

print('YES')

**else**:

print('No')

* YES
* NO
* Error
* No Output

***Bitwise AND Operator:*** *The & operator is a bitwise AND operator, not a logical AND. In conditional statements, you should use and for logical comparisons.*

***Indentation Error:*** *The print statements are not indented correctly under the if and else blocks. In Python, proper indentation is crucial for defining blocks of code.*

1. Complete the piece of code to print a maximum of three variables, a,b and c.

a=**10**

b=**16**

c=**20**

**if**(\*\*\*missing part1\*\*\*\*):

print("a")

**elif**(\*\*\*\*\*missing part **2**\*\*\*\*\*):

print("b")

**else**:

print("c")

* Missing Part 1: a > b,c

Missing Part 2: b > a,c

* Missing Part 1: a > b and a > c

Missing Part 2: b > a and b > c

* Missing Part 1: c < a and c < b

Missing Part 2: a < b

* Missing Part 1: (b < a, a > c)

Missing Part 2: (b > a, c < b)

*To determine which variable is the maximum, the conditions need to check if one variable is greater than the other two.*

*Part 1: a > b and a > c checks if a is greater than both b and c.*

*Part 2: b > a and b > c checks if b is greater than both a and c.*

*The structure of the if and elif statements logically allows for the correct identification of the maximum value. If neither a nor b is the maximum, it defaults to c, ensuring all cases are covered.*

*The other options do not correctly assess the maximum value:*

***First Option:*** *Incorrect logical operators for finding the maximum.*

***Third Option:*** *Incorrectly checks if c is less than both a and b, which does not identify the maximum.*

***Fourth Option:*** *Uses tuples instead of proper logical comparisons.*

**Subjective Questions**

1. Define the term 'Data Wrangling in Data Analytics.

Data wrangling in data analytics refers to the process of cleaning, transforming, and organizing raw data into a usable format. It involves handling missing values, correcting inaccuracies, and ensuring consistency across datasets. This process is crucial for preparing data for analysis, as it enhances data quality and reliability. Techniques such as filtering, merging, and reshaping data are commonly employed during wrangling. Ultimately, effective data wrangling enables analysts to derive meaningful insights and make informed decisions based on accurate data.

1. How do you treat outliers in a dataset?

To treat outliers in a dataset, you can first identify them using describe function and the boxplots or the IQR (interquartile range). Once identified, you have several options: you can remove the outliers if they are errors or irrelevant, or you can transform them by applying techniques such as scaling, mean or median of the non-outlier data.

1. How would you handle missing data in a dataset?

We can handle missing data by various approaches. Some of them are, If the missing data is relatively small and random, removing the rows may be a good option. If entire columns have a significant amount of missing data, you may choose to drop those columns. For categorical variables, you can replace missing values with a default category like 'Unknown' or 'None'. For numerical data, replacing missing values with the mean, median, or mode is common.

1. What are the most commonly used Python libraries for data analysis and their functions?

Pandas: Data manipulation and analysis, especially for structured data (e.g., tables, CSV files).

Numpy: Numerical computation and efficient manipulation of large arrays and matrices.

Matplotlib: Data visualization through static, animated, and interactive plots.

Seaborn: Statistical data visualization, built on top of Matplotlib.

1. How do you manipulate and aggregate data using pandas in Python?

Manipulating and aggregating data using Pandas in Python involves several operations that allow you to clean, reshape, and summarize data effectively.

Selecting columns, creating new columns , using loc and iloc functions, summarizing using groupby and pivot tables.

**Coding Questions**

1. Write a code to check if the string that the user enters starts with a vowel or not. Print capital YES or NO. For example, if input = 'analytics' then, your output should print 'YES'.

**Sample Input:**

Alpha, analytics

**Sample Output:**

YES

**Sample Input:**

Time, Sample

**Sample Output:**

NO

Code Answer:

def check\_vowel(input\_string):

# Convert the string to lowercase to make the comparison case-insensitive

vowels = 'aeiouAEIOU'

# Check if the first character is a vowel

if input\_string[0] in vowels:

print("YES")

else:

print("NO")

1. Given a single positive integer n, create a NumPy array of size (n x n) with alternating ones and zeros to form a checkerboard pattern.

**Sample Input:**

6

**Sample Output:**

[[1 0 1 0 1 0]

[0 1 0 1 0 1]

[1 0 1 0 1 0]

[0 1 0 1 0 1]

[1 0 1 0 1 0]

[0 1 0 1 0 1]]

import numpy as np

def create\_checkerboard(n):

# Create a 2D numpy array of zeros with shape (n, n)

checkerboard = np.zeros((n, n), dtype=int)

# Set the alternating pattern starting with 1 at (0, 0)

checkerboard[::2, ::2] = 1 # Set 1s on even rows at even columns

checkerboard[1::2, 1::2] = 1 # Set 1s on odd rows at odd columns

return checkerboard

# Sample input

n = 6

print(create\_checkerboard(n))