Python for Data Science - Assignment

1 MARKERS:

Q1. Which of the following functions is used to display the first few rows of a DataFrame in Pandas?
a) df.head()
b) df.tail()
c) df.info()
d) df.describe()

Answer: a) df.head()

Q2. Which Python library is mainly used for numerical computations and array operations?
a) Pandas
b) NumPy
c) Matplotlib
d) Scikit-learn

Answer: b) NumPy

5 MARKERS:

Q1. Explain the difference between NumPy arrays and Python lists. Give two advantages of using NumPy arrays in data science.

Answer:

Python List: A list can store different data types (integers, strings, floats). It is flexible but slower for numerical operations.

NumPy Array: A NumPy array stores only one data type (all integers or all floats). It is more efficient and faster for mathematical operations.

Advantages of NumPy arrays:
1. Speed: Operations on NumPy arrays are faster compared to Python lists.
2. Mathematical Functions: NumPy provides many built-in functions (mean, sum, standard deviation) for data analysis.

Q2. What is data visualization in Python? Explain with an example using Matplotlib to plot a simple line graph of student marks versus subjects.

Answer:

Data Visualization: It is the process of representing data in graphical form (like charts, plots, graphs) to understand patterns and insights.

In Python, Matplotlib is a popular library for data visualization.

Example:

```
import matplotlib.pyplot as plt

subjects = ['Math', 'Science', 'English', 'History']
marks = [85, 90, 78, 88]

plt.plot(subjects, marks, marker='o')
plt.title("Student Marks")
plt.xlabel("Subjects")
plt.ylabel("Marks")
plt.show()
```