Assignment 1 - NPTEL Python for Data Science

Name: Nehan Shetty

SRN: PES2UG22EC086

MCQs (1 Mark Each)

Q1. Which library in Python is mainly used for numerical computing?
A) Pandas
B) NumPy
C) Matplotlib
D) Seaborn
Answer: B) NumPy

Q2. In Pandas, which function is used to read a CSV file?
A) pd.read_csv()
B) pd.read_excel()
C) pd.read_data()
D) pd.read_table()
Answer: A) pd.read_csv()

Descriptive Questions (5–7 Marks Each)

Q3. What is **Pandas DataFrame**? How is it created, and why is it useful in data science? Demonstrate with an example.

Answer:

- A **DataFrame** is a **two-dimensional, tabular data structure** in Pandas with labeled rows and columns.
- It is similar to an Excel sheet or SQL table.
- It allows storing and manipulating heterogeneous data (integers, floats, strings, etc.) in columns.
- Widely used in **data cleaning, manipulation, and analysis**.

Creating a DataFrame:

import pandas as pd

```
data = {
   "Name": ["Alice", "Bob", "Charlie"],
   "Age": [25, 30, 35],
   "Salary": [50000, 60000, 75000]
}


df = pd.DataFrame(data)
print(df)
```

Output:

```
      Name  Age  Salary
0    Alice   25   50000
1      Bob   30   60000
2  Charlie   35   75000
```

Uses in Data Science:

- Easy to filter, group, and aggregate data.
- Built-in functions for handling missing data, merging datasets, and statistical analysis.
- Works seamlessly with other libraries like NumPy, Matplotlib, and Scikit-learn.

Q4. What is the role of NumPy in data science?

Answer:

- NumPy is a Python library for **numerical computing**.
- Provides efficient array operations (faster than lists).
- Supports mathematical functions (mean, std, dot product).
- Forms the base for **Pandas, Matplotlib, Scikit-learn**.
- Widely used in **data analysis, linear algebra, and scientific computing**.