ASSIGNMENT : 1

Assignment 1: Python for Data Science

Multiple Choice Questions (1 Mark each)

1.What will be the output after the following statements are executed?

```
X=300
Y= 17
X%=Y
print(X)
```

a) 11

b)17.6

c)300

d)17

Accepted Answers:

A)11

---

2. Which of the following python sequence data type is immutable?

a) list

b)dictionary

c)tuple

d)array

Accepted Answers:

b)tuple

Descriptive Questions (5 or 7 Marks each)

1. The video discusses a "Data Science Workflow" with several key steps. List and briefly explain at least four of these steps. For each step, provide a practical example of a task a data scientist might perform.

Answer:

The Data Science Workflow consists of the following key steps:

- **1. Reading Data:** This is the initial step of acquiring data from various sources. The data can be in different formats like CSV, Excel, or from a database.
    - **Example:** A data scientist might import a customer dataset from a .csv file into their Python environment to begin analysis.
- **2. Data Processing and Cleaning:** This step involves handling inconsistencies in the data. This includes dealing with missing values, correcting inaccuracies, and structuring the data in a usable format.
    - **Example:** If a customer's age is missing in some rows of a dataset, a data scientist might fill in these missing values with the average age of all customers.
- **3. Data Summarization:** This step focuses on understanding the basic characteristics of the data by calculating summary statistics.
    - **Example:** Calculating the mean, median, and mode of the 'price' column in a dataset of house sales to understand the central tendency of house prices.
- **4. Data Visualization:** In this step, data is represented graphically to identify patterns, trends, and outliers that might not be obvious from looking at raw data.
    - **Example:** Creating a bar chart to compare the sales figures of different products or a scatter plot to see the relationship between advertising spending and sales.

---

2. The video explains several reasons why Python is a preferred programming language for Data Science. Discuss at least three of these reasons in detail. For each reason, explain how it benefits a data scientist in their day-to-day work.

Answer:

Python is a top choice for data science professionals due to several key advantages:

- **1. Extensive Libraries and Frameworks:** Python has a vast collection of libraries specifically designed for data science. A library is a collection of pre-written code that data scientists can use to perform common tasks without having to write the code from scratch.

    - **Benefit:** This is a massive time-saver. For example, instead of writing complex code to create a graph, a data scientist can use a library like **Matplotlib** to generate visualizations in just a few lines. This allows the data scientist to focus more on analysis and less on basic coding.

- **2. Simplicity and Ease of Use:** Python's syntax is known for being clean, readable, and similar to plain English. This makes it one of the easiest programming languages to learn and use.
    - **Benefit:** This simplicity allows for rapid prototyping. A data scientist can quickly write a script to explore a dataset or test a hypothesis. The readable code also makes it easier for teams to collaborate on projects.
- **3. Strong Community Support:** Python has a massive and active global community. This community contributes to creating new libraries and providing support for fellow users.
    - **Benefit:** When a data scientist runs into a problem, they can easily find