

**PUSASQF405**  
**APPLICATIONS OF IT - BASICS OF PYTHON**

**Time: 2 hours**

**Total Marks: 60**

**Note:**

1. The candidate has the option to either question 3A or question 3B. Rest all questions are mandatory.
2. Numbers to the right indicate full marks.
3. The candidates will be provided with the formula sheet and graphs (if required) for the examination.
4. Use of approved scientific calculator is allowed.

**Q1. Attempt All questions.**

A. Write a python program with function to find the area of a square given its side **5 Marks.**

B. Write a program in python to calculate the interest earned by a customer. from the bank after depositing a principal amount for 5 years. Take input. from the customer for principal amount and the interest provided by the the bank was at 6.5%.  
(Formula:  $\text{principal\_amount} \times \text{intrest\_rate} \times \text{years}$ ) **5 Marks**

C. Create a dataframe 2 dataframes, one for students and another for academic details. Store them in df1 and df2. Split the data based on condition. Result and merge the data based on name and on students whose result was P in data post-split. **5 Marks**

```
student_details:
data = {
    "Name": ["Alice", "Bob", "Charlie", "David"],
    "Age": [30, 40, 25, 50],
    "Gender": ["F", "M", "M", "M"],
    "City": ["New York", "Los Angeles", "Chicago", "Houston"]
}

academic_details:
data = {
    "Name": ["Alice", "Charlie", "David", "Bob", "Alice", "David"],
    "Marks": [75, 80, 55, 78, 40, 67],
    "Result": ["P", "P", "P", "P", "F", "P"]
}
```

**Q2. Attempt All questions.**

**A. Write a python code for the below dictionary:**

```
i. data = {'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Emily'],  
          'Age': [25, 30, 35, 40, 45],  
          'Country': ['USA', 'Canada', 'USA', 'Canada', 'USA']}
```

ii. Store the data in a dataframe df

iii. Filter the dataframe where country is USA

**5 Marks**

**B. Write the below python code to create a dataframe**

```
df = pd.DataFrame({'A': [1, 2, 3, np.nan, 5],  
                  'B': [6, np.nan, 8, 9, 10],  
                  'C': [11, 12, 13, 14, np.nan],  
                  'D': [15, np.nan, 17, 18, 19],  
                  'E': [20, 21, np.nan, 23, 24]})
```

i. Check if there are any NA values in the dataframe

ii. If there are NA values in the dataframe, replace the NA values with its mode.

**C. A company Z is well known for manufacturing balcony blinds. Create a python function such that the price of the balcony blind is Rs. 350 per square feet and do the following: 5 Marks**

i. Ask user to input the area of the blind. **(2.5)**

ii. Create a function such that the input is passed to calculate the final bill amount which includes 5% of manufacturing cost of the total bill amount, 9% CGST and 9% SGST of the total bill amount and a discount of 15%. Provide the final bill charges. **(2.5)**

```
manufacturing_amount = 0.05 * bill_amount
```

```
cgst = 0.09 * bill_amount
```

```
sgst = 0.09 * bill_amount
```

```
total_bill = bill_amount + manufacturing_amt + cgst + sgst - discount
```

**Q3. Attempt question 3A or question 3B.**

**A. Read insurance.csv dataset in python. Perform Linear Regression on the dataset 30 Marks**

- a. Read the dataset. (2)
- b. Convert categorical columns sex, smoker, region to numerical. (3)
- c. Store the whole data frame except for charges column in X and charges column in Y (7)
- d. Split the dataset into train and test with test size 20% (7)
- e. Perform Linear Regression (7)
- f. Provide the value for RMSE. (4)

**OR**

**B. Load the diabetes.csv dataset. Build a model to predict the chances of a patient having diabetes or not. Predict using Logistic Regression. 30 Marks**

- a. Read the dataset. (5)
- b. Split the dataset in X and Y where X will have all columns except for Outcome, Y will have the column Outcome (5)
- c. Perform train test split with test size 20% (5)
- d. Run the logistic regression model. (5)
- e. Predict the values for test dataset. (5)
- f. Find the accuracy of the model. (5)