PPSAS306
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 calculators is allowed.**

Q1 Attempt the following: **15 Marks**

A. **5 Marks**

WAP to check if value 140 exists in the following dictionary.
my_dict = {'a': 100, 'b': 120, 'c': 140, 'd': 160}

B. **5 Marks**

For the given list

States_list = [‘MAH’,’GUJ’,’DEL’,’PUN’]
Update ‘states_list’ by adding ‘UP’, ‘MP’ and ‘AP’

Insert UP on index 4 and ad AP at index 3

C. **5 Marks**

For the given data frame consisting id, name, maths_marks, chem_marks, eng_marks, french_marks perform group by function based on:

i. Name with respect to french_marks (2)
ii. Name with respect to eng_marks (1)
iii. Name with respect to maths_marks (1)
iv. Name with respect to chem_marks (1)

```
data = {
 'id': [1, 2, 3, 4, 5],
 'name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],
 'maths_marks': [90, 85, 78, 92, 88],
 'chem_marks': [80, 75, 88, 72, 90],
 'eng_marks': [85, 78, 90, 82, 88],
 'french_marks': [88, 92, 78, 85, 82]
}
```

Q2. Attempt the following: **15 Marks.**

A. **5 Marks**

Write a python program to print fibonacci series by taking an input from the user and print the series till the number n.

Example:
Enter a number (n) to generate Fibonacci series up to n: 10
Fibonacci series up to 10 is:
[0, 1, 1, 2, 3, 5, 8]

B. **5 Marks**

For the below provided data frame write a python program to create a bar plot to visualize count of different categories.

```
data = {

    'Category' : ['Fruit', 'Vegetable', 'Dairy', 'Grains', 'Meat'],
    'Count' : [30, 25, 40, 20, 35]
}
```

i. Store the data in a dataframe named "df" (1)

ii. Create a bar plot (4)

C. **5 Marks**

Personal data of 5 individuals having various attributes is mentioned below. Perform feature engineering on numeric data to derive new features that can provide insights to individuals' characteristics.

```
data = {

'Age' : [25, 30, 22, 28, 35],
'Height_cm' : [165, 172, 155, 168, 180],
'Weight_kg' : [ 60, 70, 50, 65, 75],
'Income_USD' : [50000, 60000, 45000, 55000, 70000]

}
```

i. Perform min-max scalar (1)

ii. Calculate BMI from weight and height as mentioned in below code (1)

iii. Create a function get_bmi_category where: (3)

a. bmi < 18.5 'Underweight'
b. bmi >= 18.5 and bmi < 25 'Normal'
c. bmi > = 25 and bmi < 30 'Overweight'
else 'Obese'

Q3 Attempt question Q3A or Q3B **30 Marks.**

A.

You are given the "diabetes" dataset from scikit-learn's built-in datasets. This dataset contains features representing various aspects of sugar levels and lifestyles and their respective values. Your task is to build a linear regression model to predict the value based on the given features. Perform the following steps:

a.	Load the dataset	(2)
b.	Check missing values and handle them appropriately	(2)
c.	Encode categorical variables using one hot encoding if any	(2)
d.	Split the data into train and test	(2)
e.	Train a linear regression model on the training data.	(6)
f.	Build a linear regression model to predict the values	(6)
g.	Make predictions on test data using trained linear regression model	(4)
h.	Evaluate performance of linear regression model using appropriate metrics	(4)
i.	Create a regression plot to visualize relationship between predicted housing prices and actual house price	(2)

OR

B. **30 Marks**

You have been provided with the famous Iris dataset, containing measurements of various iris flowers along with their species classification. Your task is to perform a logistic regression analysis on this dataset to classify iris flowers into their respective species based on their features.

a. Load the iris dataset and display the first few rows (2)

b. Check for missing values in the dataset. If there are any missing values, handle them appropriately. State whether there are any missing values or not. (2)

c. Encode categorical variables using one-hot encoding. Are there any categorical variables in this dataset that need encoding? If so, perform the encoding; otherwise, explain why it's not necessary. (2)

d. Split the data into training and testing sets using an 80-20 split ratio. How many samples are in the training set and the testing set, respectively (2)

e. Perform exploratory data analysis (EDA) by creating visualizations to gain insights into the relationships between features and the target variable. Include at least two different types of visualizations. Interpret any notable findings from the visualizations. (6)

f. Build a logistic regression model using the training data. Train the model and display the model's coefficients for each feature. (6)

g. Make predictions on the test data using the trained logistic regression model. Calculate the accuracy of the model on the test set. Provide an interpretation of the accuracy. (4)

h. Create a confusion matrix and a classification report for the model's predictions on the test set. Explain the key metrics from the classification report. (6)