PPSASMAJ304
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. Answer the following **15 Marks**

A. **5 Marks**

Write a program that accepts a sentence and calculate the number of letters and digits

B. **5 Marks**

Write a program which prompts the user to input principal, the rate is 7.5% annually, calculate the final amount (after 15 years). Plot the interest amount over 9 years on a bar chart. The formula is: CI = P(1+R/100)^T

C. **5 Marks**

Given a 2D array of user IDs, find the number of occurrences of each number, output the number and occurrence in dictionary.

Input: [[4, -3, 15], [15, 22, 4], [-3, 4, 22]]

Output: {"4": 3, "-3": 2, "15": 2, "22": 2}

Q2. Answer the following **15 Marks**

A. **5 Marks**

Read the bike travel csv file and store them in df1 and convert start time and end time in date time format?

B. **5 Marks**

Find the bike which is used for the most amount of time. Plot the duration and cycle id on bar chart and sort them from increasing to decreasing duration?

C. **5 Marks**

Find the last time each bike was in use. Output both the bike number and the date-timestamp of the bike's last use

Q3. Attempt question 3A or question 3B. **30 Marks**

A. Load the breast cancer dataset. The dataset contains information about the breast cancer related factors. Predict the breast cancer status with all other factors given in the dataset using logistic regression

a. Read the dataset and split the data in X and Y.

i. Identify independent variables consider it X **(3)**

ii. Identify dependent variables consider it Y **(3)**

b. Perform feature engineering, fill all missing values and Perform label encoding or One hot encoding whenever required. **(6)**

c. Perform train-test split with test ratio 30% **(3)**

d. Perform Logistic Regression on train dataset. **(6)**

e. Predict the values on the Test dataset. **(6)**

f. Provide the value for Accuracy score and other classification matrix **(3)**

OR

B. **30 Marks**

Load the Wine juice dataset. The dataset contains information.

About the wine juice quality and the factors that affect it. Predict the wine juice quality using Linear Regression.

a. Read the dataset and split the data in X and Y.

i. Identify independent variables consider it X **(3)**

ii. Identify dependent variables consider it Y **(3)**

b**.** Perform feature engineering, fill all missing values and Perform label encoding or One hot encoding whenever required. **(6)**

c. Create a train-test split with a 10% test ratio. **(3)**

d. Run the Linear Regression model. **(6)**

e. Predict values for test dataset. **(6)**

f. Provide the RMSE & R square value **(3)**