PUSASQF405
APPLICATIONS OF IT - BASICS OF PYTHON (SET 2)

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 All questions.

A) **5 Marks**

Write a program to find the number of occurrences of elements from list1 to list2?
Hint – No. of common elements between the two list
list1=[1,2,3,4,5]
list2= [2,8,9,4,5]

B) **5 Marks**

Write a program which prompts the user to input principal, the rate is 6% annually, calculate the final amount (after 10 years). Plot the amount over 10 years on a bar chart
The formula is: CI = P(1+R/100)^T

C) **5 Marks**

Write a program in python to find the missing number in the array?
input_list = [10,7,4,1,2,5,8,9,6]

Note – Don't use a constant list or numbers

Q2 Answer the following **15 Marks**

Create a 2 dataframes, Purchase order details and another for Purchase store details.

Name	Purchase Date	Purchaser Amt	MRP
Rahul	31-01-2023	148.7	150
Pinky	30-01-2023	121.3	130
John	29-01-2023	136.4	159
Shubham	31-01-2023	105.7	110
Aman	27-01-2023	178.8	178.8
Parul	29-01-2023	394	400
Abhay	31-01-2023	143.3	150

Name	sector	Location	Region
Rahul	FMCG	Delhi	North India
Pinky	GARMENTS	Delhi	North India
John	FMCG	Delhi	North India
Shubham	FMCG	Mumbai	West India
Aman	GARMENTS	Mumbai	West India
Parul	MEDICINE	Mumbai	West India
Abhay	MEDICINE	Gujrat	West India

a) Store them in df1 and df2. Merge the dataframe based on name **5 Marks**

b) Print the date on which maximum purchases were made, along with it print the region wise split of the total purchase amount **5 Marks**

Sample output -

Date	PurchaseAmt	North India	West India
XXX	xxx	xxx	xxx

c) Create a line chart using seaborn to represent the percentage of discount offered on each day? Discount % -> (Mrp-Purchase Amt)/Mrp*100 **5 Marks**

.

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

A.

Load the Loan.csv dataset. The dataset contains information about the loan eligibility and other eligibility factors. Predict the loan 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(Loan_Status) **(3)**

b. Perform feature engineering, Fill all missing values in categorical columns with 0 or ignore the line items and fill the numerical columns with its mean value. **(6)**

c. Perform train-test split with test ratio 25% **(3)**

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

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

f. Provide the value for Accuracy score. **(3)**

OR

B. **30 Marks**

Load the House_Rent_Dataset.csv dataset. The dataset contains information.
About house rent and the factors that affect it. Predict the rent charges
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(Rent) **(3)**

b. Perform feature engineering using label encoding on at least 3 columns of your choice **(6)**

c. Create a train-test split with 20% test ratio. **(3)**

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

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

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