

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 calculators is allowed.

Q1. Answer the following question

15 Marks

A

5 Marks

Write a python function that accepts a string and calculates the number of upper-case letters and lower-case letter?

B

5 Marks

Write a program which prompts the user to input principal, rate and time and calculate compound interest and visualize the rate of growth of Interest over a line chart?
The formula is: $CI = P(1+R/100)^T - P$

C

5 Marks

Write a program in python, given an integer n, return the number of trailing zeroes in N factorial n!?

Hint -> $7! = 5040$, trailing zeroes =1

Q2. Answer the following Question

15 Marks

Create a 2 dataframes, Company Financial Data (Tab1) and another for Company (Tab2) details.

company	Market value	sales	profits	assets
ICBC	215.6	148.7	42.7	3124.9
China Construction Bank	174.4	121.3	34.2	2449.5
Agricultural Bank of China	141.1	136.4	27	2405.4
JPMorgan Chase	229.7	105.7	17.3	2435.3
Berkshire Hathaway	309.1	178.8	19.5	493.4

Exxon Mobil	422.3	394	32.6	346.8
General Electric	259.6	143.3	14.8	656.6

(Tab -1)

company	sector	industry	continent	country
ICBC	Financials	Major Banks	Asia	China
China Construction Bank	Financials	Regional Banks	Asia	China
Agricultural Bank of China	Financials	Regional Banks	Asia	China
JPMorgan Chase	Financials	Major Banks	North America	United States
Berkshire Hathaway	Financials	Investment Services	North America	United States
Exxon Mobil	Energy	Oil & Gas Operations	North America	United States
General Electric	Industrials	Conglomerates	North America	United States

(Tab -2)

- Store them in df1 and df2. Merge the dataframe based on foreign key? **5 Marks**
- Print the continent with the total number of companies and their total sales? sort them based on highest sales? **5 Marks**
- Print the percentile rank on market value for each company and visualize them on a bar chart with "Country" in legends **5 Marks**

Q3. Attempt question 3A or question 3B.

30 Marks

- Load the Blood_Pressure.csv dataset. The dataset contains information about Blood Pressure abnormality and related factors. Predict the Blood Pressure abnormality with all other factors given in the dataset using logistic regression
 - Read the dataset and split the data in X and Y.
 - Identify independent variables consider it X **(3)**
 - Identify dependent variables consider it Y(Outcome) **(3)**
 - 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)**
 - Perform train-test split with test ratio 20% **(3)**
 - Perform Logistic Regression on train dataset. **(6)**
 - Predict the values on Test dataset. **(6)**
 - Provide the value for Accuracy score. **(3)**

OR

B.

30 Marks

Load the HealthInsurance.csv dataset. The dataset contains information about the insurance price and factors sensitive to it, along with other factors. Predict the Insurance charges 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. Convert the categorical columns (sex, smoke, region) using label encoding method (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)