

Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306
POST GRADUATE DIPLOMA IN MANAGEMENT (2023-25)
END TERM Examination (Set ~ B)
TERM-IV (Oct'24)

Subject Name: **Python for Business Analytics (PBA)**

Time: **02:00 Hrs.**

Sub. Code: **PGIT31**

Max Marks: **40**

INSTRUCTIONS

1. All questions are to be solved using Jupyter notebook on individual Computers/LAPTOPS.
 - o The first Cell (of Jupyter Notebook): Enter Subject Name, Date, Course & Session (PGDM 2023-25), SEMESTER, Student's Name, ROLL NUMBER on top eg. # Subject Name, # Date and so on
 - o Use ONLY one notebook file and write # Question number before starting each question
 - o Save the file using your full name, RollNumber and Section for example (Name_PGDMnnnnn_Sec) with '.ipynb' extension and submit the soft copies using a PD to the invigilator.
 - o Sharing of PD for submitting final ANSWER/SOLUTION is not allowed and would lead to cancellation of the answers submitted.
2. During examination, no student is allowed to use mobile phones/Smart watch/Internet in any conditions.
3. NO Data sheets (.csv file/Excel) will be provided as a soft copy on the Desktops/Laptops
4. There are 2 pages in this question paper.

COs	Statement of Course Outcomes	Bloom's Taxonomy
CO-1	Demonstrate proficient use of Python programming.	L-3, L4
CO-2	Skillfully operate upon arrays using NumPy in Python.	L-3, L4
CO-3	Illustrate use of Pandas to create and analyze datasets to solve business problems.	L-3, L-4
CO-4	Develop impressive data visualizations using matplotlib and seaborn libraries	L-5

All questions are compulsory and carry equal marks

[8 Marks x 5=40 Marks]

SECTION - A

Q1. Attempt all parts of the question. Each subpart is of **1 Marks each.**

(CO1, CO2)

A) Write the Python queries for the given data 'x':

`x = np.array([[0,1,2,3], [4,5,6,7]])`

- a. Display the x's array attributes (shape, dimensions, data type and size)
- b. Change the shape to (2, 4)
- c. Print column-wise sum and row-wise sum
- d. Write code to extract number 7 from the array.

- B) a. Create a NumPy array of 20 elements with name **np_array**.
- b. Reshape it to 5 row and 4 columns dimension with name **np_array2**
- c. Convert it to a Pandas DataFrame with column headings **col1, col2, col3, col4** with name **df**.
- d. Transpose DataFrame **df**

SECTION - B

Q2. Given the dataset, create a .csv, to: [2 marks each] (CO1,CO2)

- i. a) Check for duplicates,
b) Remove duplicates in the dataset
- ii. Fill the missing values in the 'Age' and 'Salary' columns, with the **mean** value of the respective columns. Write this data as **Q2data.csv**.
- iii. Encode the categorical variables 'Country' and 'Purchased' columns into numerical data.
- iv. Check for potential outliers in the 'Age' and 'Salary' columns. (hint: remove any rows with values beyond 1.5 times the IQR from the quartiles).

	Country	Age	Salary	Purchased
0	France	44.0	72000.0	No
1	Spain	27.0	48000.0	Yes
2	Germany	30.0	54000.0	No
3	Spain	38.0	61000.0	No
4	Germany	40.0	NaN	Yes
5	France	35.0	58000.0	Yes
6	Spain	NaN	52000.0	No
7	France	48.0	79000.0	Yes
8	Germany	50.0	83000.0	No
9	France	37.0	67000.0	Yes

Q3. Create a csv file with following data and use it to answer the following question: (CO3)

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	NaN	148	406.0
5	60	2020/12/06	102	127	NaN
6	60	'2020/12/07'	110	136	374.0
7	45	'2020/12/07'	110	136	374.0

- a. Find the empty data, wrong format, wrong data and duplicate data using Pandas
- b. Remove the duplicate data and display.
- c. Save it in text file

Q4. Given the two DataFrames (**Product_info** & **sales_trnx**), solve the queries given below: (CO3, CO4)

Product_info

Product_ID	Name	Uprice	Category	Sales_Target
101	Pencil HB	70	Pencil	500
102	Eraser	21	Eraser	500
103	Sharpener	22	Sharpener	500
104	Pencil H2	80	Pencil	500
105	Pencil E2	120	Pencil	500
106	Pencil HH	120	Pencil	500

Sales_trnx

Product_ID	Sales_Qty	Bill_Date
101	88	10-05-2023
102	75	12-05-2023
105	93	15-05-2023
106	81	17-05-2023

- a) Merge the **Product_info** and **sales_trnx** DataFrames on **Product_ID**. [4]
- b) Create a plot that shows the actual **sales vs. sales target**. [4]

SECTION - C

Q5. While using the dataset of **Q2data.csv** (Q2. ii), solve the queries given below: (CO3, CO4)

- a) Write code to create pivot-table/crosstab having columns Country and Average Salary, as **df_cross**. [4]
- b) Create a distribution plot with KDE on the above data **df_cross**. [2]
- c) Perform descriptive analysis on the above dataset **df_cross**. [2]

-- O --

Kindly fill the total marks allocated to each CO's in the table below:

COs	Marks Allocated

CO1	08 Marks
CO2	08 Marks
CO3	16 Marks
CO4	8 Marks

(Please ensure the conformity of the CO wise marks allocation as per your TLEP.)

Blooms Taxonomy Levels given below for your ready reference:

- L1= Remembering
- L2= Understanding
- L3= Apply
- L4= Analyze
- L5= Evaluate
- L6= Create