PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS901**

JULY 2021 END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER I

UE20CS901 - Python for Data Science

Time: 3 Hrs	Answer All Questions	Max Marks: 80

INSTRUCTIONS

- All questions are compulsory.
- Section A should be handwritten in the answer script provided.
- Section B and C are coding questions which have to be answered in the system.

SECTION - A

1	a)	What is the difference between List and Tuple?	2
	b)	What is the difference between index() and find() method with respect to string	2
	c)	What is type casting? Explain with respect to Python	2
	d)	How to change the value associated with the key in dictionary?	2
	e)	What are anonymous functions in Python?	2
2	a)	What is an identity matrix? How to create it using numpy library?	2
	b)	Explain split() method for arrays.	2
	c)	Explain the reshape() method.	2
	d)	What is the difference between sort_values() and sort_index() method?	2
	e)	Which plot is used to visualize distribution of data? Which methods are available for it?	2

SECTION - B

3	a)	A computer hardware shop provides the service of computer assembly. Prepare the bill of assembled computer based on the following information (8 Marks) a) Prepare the pricelist of different spare parts as follows(2marks) HDD 1TB - 5000/- , 2TB - 7500/- , 4 TB - 10000 RAM 8GB - 4000/- 16GB - 6000/- Processor - I5 - 15000/- I7 - 18000/- Display - 24" - 3500/- 26" - 4500/- Keyboard - Normal - 1800/- Wireless -2200/- Other charges - 4000/- b) Print the prices of all items. Ask the user choices for all spare parts(3 marks) c)Prepare the bill adding 12% GST to the total cost.(3 marks) **SAMPLE INPUT AND OUTPUT:** Please find the price list Item - HDD Type : 1TB : 5000 Type : 2TB : 7500 Type : 4TB : 10000 Item - RAM Type : 8GB : 4000 Type : 16GB : 6000 Item - Processor Type : I5 : 15000 Type : I7 : 18000 Item - Display Type : 24 Inches : 3500 Type : 26 Inches : 4500 Item - Keyboard	8

		Type : Normal : 1800 Type : Wireless : 2200 Type : Other : 4000 Enter the HDD choice1TB Enter the RAM choice16GB Enter the processor choice I5 Enter the Display choice 24 Inches Enter the keyboard choiceNormal The Computer Bill ========== Item Price HDD 1TB 5000 RAM 16GB 6000 Processor I5 15000 Display 24 Inches 3500 Keyboard Normal 1800 Other 4000 12%GST 4236.0 Total 39536.0	
	b)	Write a user defined function to solve the knapsack problem. It is a problem in combinatorial optimization: Given a set of items, each with a weight and a profit, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total profit is as large as possible. (12 Marks) Sample input/output: profit = [1,2,3,4] weight = [6,3,8,10] Output : Selected items and their portions are: [0, 1, 0.875, 1] Profit earned: 8.625	12

	c)	Write a program to generate n terms of series as below 0, 1, 3, 6, 10, 15, 21, 28, 36, 45, 55..... Write a user defined function to give sum of terms in the range of k to l where k,l<=n Accept n,k,l from the user (10 Marks)	10
		SECTION - C	
4	a)	The dataset is a collection of website ratings and ratings for different products customers purchased userid - Unique id given to each user who visits the website Ucredit - The credit given by user website's content Ureview - Review given by user for the website's usability Web_review - Review given for website's security consecutive_usage - How frequently user visits the website Exp_review - Review for overall experience assigned _metric - Metric used to calculate rating assigned_rating - Rating given using the metric Product_1- Product_id of first purchase rating_1 - Rating given to the first product product_2 - product_id of the second purchase rating_2 - Rating given to the second product product_3 - product_id of the third purchase rating_3 - Rating given to the third product Read the dataset and understand the following questions 1. Read the "Meta-data" of the dataset (5 Marks) **Step1:** Read the first 10 rows of the dataset	15

		Step2: Check the data type of each column **Step3:** Display the number of numerical and non-numerical columns **Step4:** Display the statistical summary for numerical columns 2. Calculate the average rating given by the user (5 Marks). **Step1:** Consider Ucredit, Ureview, Web_review columns and calculate the average **Step2:** Round the average to 1 decimal place **Step3:** Add a new column avg_rating and fill the values, calculated in above step 3. Count the number of distinct products that are purchased and rated by user (3 Marks) **(The product can appear in first, second or third purchase) 4. Plot a boxplot for the 'web review' and 'Exp_review' (2 Marks)	
	b)	1.Which metric is used for the most of ratings (5 Marks) 2. Create a dataframe of users with consecutive usage more than 4. What is the average user review for this group (5 Marks) 3. Display the heatmap representing the correlation between Ucredit Ureview Web_review consecutive_usage Exp_review (2 Marks) 4. Plot a count-plot for different assigned ratings (2 Marks) 5. Plot a distribution graph for average rating. Use 30 bins (1 Marks)	15