



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

	<p>Type : Normal : 1800</p> <p>Type : Wireless : 2200</p> <p>Type : Other : 4000</p> <p>Enter the HDD choice1TB</p> <p>Enter the RAM choice16GB</p> <p>Enter the processor choice I5</p> <p>Enter the Display choice 24 Inches</p> <p>Enter the keyboard choiceNormal</p> <p>The Computer Bill</p> <p>=====</p> <table><tr><td>Item</td><td>Price</td></tr><tr><td>HDD 1TB</td><td>5000</td></tr><tr><td>RAM 16GB</td><td>6000</td></tr><tr><td>Processor I5</td><td>15000</td></tr><tr><td>Display 24 Inches</td><td>3500</td></tr><tr><td>Keyboard Normal</td><td>1800</td></tr><tr><td>Other</td><td>4000</td></tr><tr><td>12%GST</td><td>4236.0</td></tr><tr><td>Total</td><td>39536.0</td></tr></table>	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	
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)	<p>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)</p> <p>Sample input/output: profit = [1,2,3,4] weight = [6,3,8,10]</p> <p>Output : Selected items and their portions are: [0, 1, 0.875, 1]</p> <p>Profit earned: 8.625</p>	12																		

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

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