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

XXXX: 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: 100

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)	How will you demonstrate the following "Python is case sensitive programming language"?	2
	b)	Provide 2 examples of implicit type casting	2
	c)	What is a complex data type in Python?	2
	d)	What is the default parameter in user-defined function?	2
	e)	Explain arithmetic operators which can be used with strings.	2
2	a)	Define significant features of Numpy library.	2
	b)	Explain the difference between pivot table and cross table?	2
	c)	Explain the random package of Numpy?	2
	d)	How to delete rows and columns from a Dataframe?	2
	e)	How to rank the rows of Dataframe?	2

SECTION B (40 Marks)			
3	a)	For the given list perform the following operations list_name = ['Ann', 'Pat', 'David','Tisha','Sumantha'] 1. Create a list with values equal to the length of each string 2. Create a list with names of length less than 5 3. Create a dictionary with strings as key and length of the string as values 4. Sort the list with respect to the length of the string keeping the largest string at the first position 5. Find the most frequently occurring character throughout the list	15

	b)	A farmer wants to sell his yields to the market. He has a truck with a capacity of 500kg to carry. The farmer yields tomatoes, potatoes, garlic, and brinjals. Following are the quantity, selling price of 10kg unit of each vegetable. Tomato (Yield 150kg) (SP 500/-) Potatoes (Yield 200 kg) (SP 420/-) Garlic (Yield 250 kg) (SP 700/-) Brinjals (Yield 100 kg) (SP 600/-) For every vegetable there is a must sell quantity (As the vegetable may get stale if not sold in a specific period) Tomato - 50 kg Potato - 70 kg Garlic - 70 kg Brinjal - 40 kg In addition to the above products what other products the farmer should carry to earn maximum money. Provide the quantity of each vegetable that the farmer should carry in the truck	20

	c)	Write a user-defined function to solve the given task. From the given vehicle number plates determine are these numbers even or odd. Output the dictionary with key as number and value as (even/odd) as appropriate(5 Marks) number_plates = ['MH 12 XJ - 2234', 'UP 04 LG - 2455', 'GJ 34 RV - 2442', 'KL 07 AP-2433']	5
SECTION C (40 Marks)			
4	a)	The Dataset Dataset Content This dataset contains the stressed condition of students during pandemic. The data is collected for students across 12 different countries in Europe. Citation: Attribute Information: Attributes for datasets: sex - student's sex (binary: 'F' - female or 'M' - male) age - student's age (numeric: from 15 to 22) famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3) Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart) Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 -higher education) Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 -5th to 9th grade, 3 - secondary education or 4 - higher education) Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'athome' or 'other') Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'athome' or 'other') guardian - student's guardian (nominal: 'mother', 'father' or 'other') studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	10

		schoolsup - extra educational support (binary: yes or no) famsup - family educational support (binary: yes or no) paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) activities - extra-curricular activities (binary: yes or no) internet - Internet access at home (binary: yes or no) romantic - with a romantic relationship (binary: yes or no) famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent) freetime - free time after school (numeric: from 1 - very low to 5 - very high) goout - going out with friends (numeric: from 1 - very low to 5 - very high) health - current health status (numeric: from 1 - very bad to 5 - very good) absences - number of school absences (numeric: from 0 to 93) These grades are related with the course subject for the evaluation conducted during pandemic: G1 - first period grade (numeric: from 0 to 20) G2 - second period grade (numeric: from 0 to 20) G3 - final grade (numeric: from 0 to 20,) Stress - 0 or 1. 0 indicates not stressed, 1 indicates stressed	

		Perform the following tasks after reading the dataset 1. Calculate the % of stressed students. 2. What is the average age of female students 3. Add a column as avg_grade which is equivalent to the average of 3 grades. 4. Find the correlation between studytime and average grades 5. Check the distribution of average grade	
	b)	Find the role of the family in managing stress for teenager students by answering the following questions 1. Create a cross table with values of family relation quality against stress status.(2 Mark) 2. Create a pivot table to include stressed student's count with respect to information of the size of family, family's support for education (3 Marks) 3. What is minimum and maximum average grade scored by students whose parents have their education level as higher education (3 Marks) 4. How much average free time does a student living in a family with greater than 3 members. How many students with free time less than average free time in such families are stressed (5 Marks) 5. Create a new column to store the status of the mother's occupation. Keep only 2 status working_mother - yes/no. Plot the working mother's impact on the count of the stress of students. (5 Marks)	18

	c)	Solve the following 1. Plot the boxplot of absences against the gender of students (2 Mark) 2. Plot the piechart for the different occupations of the father of the student (3 Marks) 3. How many students perform progressively? That their scores improved in three evaluations.(3 Marks) 4. How many students in romantic relationships perform progressively better? Display their gender wise distribution (4 Marks)¶	12