



October 2024: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND ARTIFICIAL INTELLIGENCE_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 and signed at the end of the same. Section B and C are coding questions which have to be answered in the system.

SECTION A – 20 MARKS

1	a)	How do you generate a random float number between 0 and 1 in Python?	2
	b)	What is the difference between pop() and remove() methods in a Python list?	2
	c)	Explain the purpose of the filter() function in Python. Provide an example where it is used to filter even numbers from a list.	2
	d)	What is the purpose of the enumerate() function in Python? Provide an example of its usage in a loop.	2
	e)	In NumPy, what is the difference between arange() and linspace()?	2

2	a)	How do you select a subset of rows based on a condition in a pandas DataFrame? Provide an example.	2
	b)	Explain the concept of groupby in pandas with its syntax. Provide an example of how to use the groupby() function to group data and perform an aggregate operation (like sum or mean) on a DataFrame.	2
	c)	How can you remove missing data from a pandas DataFrame?	2
	d)	What is the difference between apply() and map() in pandas?	2
	e)	How can you reset the index of a pandas DataFrame? Explain with an example.	2

SECTION B – 40 MARKS

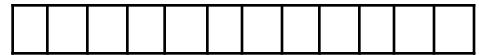
3	a)	Write a Python program that takes a list of integers as input and returns a new list containing only the prime numbers from the original list. If no prime numbers are found, return an empty list.	8
	b)	Write a Python program to manage a basic inventory system for a small store. Your program should: <ol style="list-style-type: none"> Allow the user to add a new item with its name, quantity, and price. Implement a function to update the quantity of an existing item. 	8



	c.	Implement a function to retrieve the total value of all items in the inventory (sum of quantity * price). d. Implement a function to display all items in the inventory.	
c)		Create a Python function that takes a dictionary of employee names with their corresponding (years of experience, current salaries, and performance ratings) (scale of 1-5). The function should adjust the salaries based on the following rules: a. If an employee has more than 10 years of experience and their salary is below 70,000 dollars, increase their salary by 15% plus an additional amount equal to 2% of their current salary multiplied by their performance rating. b. If they have between 5 and 10 years of experience and their salary is below 50,000, increase their salary by 10% plus an additional amount equal to 1.5% of their current salary multiplied by their performance rating. c. If they have less than 5 years of experience and their salary is below 40,000, increase their salary by 5% plus an additional amount equal to 1% of their current salary multiplied by their performance rating. d. Employees not fitting these criteria receive no adjustment. e. The function should return a new dictionary with the adjusted salaries.	8
d)		Write a Python program that takes a list of sentences and returns a dictionary where the keys are the unique words (case insensitive) that are at least 4 characters long, and the values are the counts of those words. Additionally, create a histogram that shows the distribution of word lengths (the length of each unique word) in the input sentences. Ignore punctuation and consider only alphabetic characters.	8
e)		Write a function longest_consecutive_sequence(nums) that takes a list of integers and returns the length of the longest consecutive elements sequence. The sequence should be considered as consecutive if they follow each other in integer value.	8

SECTION C – 40 MARKS

4	a)	Using the car_info dataset answer the following: (i) Load the dataset and find the number of cars with a price greater than \$25,000. Calculate the average mileage for these cars. Write your inference. (4 Marks) (ii) Create two separate DataFrames for cars with fuel type "Petrol" and cars with fuel type "Diesel." Display the average price for each fuel type. (5 Marks) (iii) Find the average price and maximum mileage for each car make. (3 Marks) (iv) Create a new column that categorizes cars based on their mileage as 'Low', 'Moderate', or 'High'. (4 Marks) - Mileage >0 and < 30000 --> Low - Mileage >30000 and <100000 --> Moderate - Mileage > 100000 --> High (v) Identify the most expensive car for each make. (4 Marks)	20
4	b)	Using the car_info dataset (without loading the data again) answer the following (i) Based on the relationship between Price and Age of the car for different fuel types comment on the pricing trends across different mileage ranges. (4 marks)	20



	<p>(ii) How can we calculate the average price for each car make while filtering out listings with missing values in the Price or Mileage columns. (4 Marks)</p> <p>(iii) Show the distribution of Fuel_Type across Make, revealing the most popular fuel types for different car brands. (4 Marks)</p> <p>(iv) How do car prices vary with age, and are there differences in price trends between manual and automatic transmission cars? Comment on both of these scenarios with relevant plots.(4 marks)</p> <p>(v) Which car make are the most listed in the dataset? Support your answer using relevant plots (4 marks)</p>	
--	--	--