Shiv Nadar University Chennai

Mid Semester Examinations 2023-2024 Even

Question Paper

Name of the Program: B.Tech. Artificial Intelligence and Data Science		Semester: II
Course Code & Name: CS1704 Foundations of Data Science + Lab		
Regulation 2021		
Time: 2 Hours	Answer All Questions	Maximum: 50 Marks

Q.No.		Questions	Marks	CO	KL
1	a	Compare the advantages of NumPy over lists.	3	CO1	KL4
2	a	What is Variance? How is it different from standard deviation?	2	CO2	KL2
3	a	As a data analyst tasked with enhancing the performance of an e-commerce seller based on provided sales data spanning several years, address the following key aspects by finding the right plots for each scenario and write about the plot, syntax, use case, model plot in this scenario: 1. Investigate Sales Trends and Seasonality: — Analyse the sales data over the given period to identify any discernible trends or seasonality monthly. 2. Identify Underperforming Products: — Evaluate the sales performance of each product and recommend whether there are any products with consistently low performance that the seller should consider discontinuing. 3. Optimize Marketing through Age Distribution Analysis: — Utilize age distribution analysis to enhance marketing strategies. Provide insights on how understanding the age demographics of the customer base can contribute to more effective and targeted advertising. 4. Expense Analysis for Seller Understanding: — Help the seller gain a comprehensive understanding of their expenses by breaking down and visualizing the expense categories. Offer insights into areas where cost optimization or reallocation may be beneficial. 5. Explore Correlation Between Ads Clicks and Conversion Rate: — Investigate the correlation between the number of ads clicks and the conversion rate. Provide insights into how these two factors interact, enabling the seller to make informed decisions regarding their advertising strategies.	15	CO5	KL3
4	a	For the given e-commerce data perform the following operations: 1. Perform the Z score normalisation on the sales column to understand the above and below average performance. Also find any abnormality in the sales using outlier detection.	15	CO2	KL4

		2. Perform Min-Max scaling on the price of the products sold, help to make strategic pricing decisions by analysing the correlation between price and sales. 3. Perform IQR based outlier detection on the reviews column to find products with unusually high or low reviews.			

Product	Sales	Price	Reviews
A	100	200	15
B	278	400	25
C	30	600	35
D	40	800	450
E	50	1000	55

5	a	Create a data frame for this data and perform the following operations using pandas. Explain the operation, illustrate with the given data, provide the right pandas syntax. Name&Age&City&Salary&Experience Alice&25&New York&60000&3 Bob&30&San Francisco&80000&5 Charlie&22&Los Angeles&55000& David&35&Chicago&90000&8 Emily&28&Boston&70000&4	15	CO1	KL3
		1. There are missing values in the 'Experience' column. What simple method can you use to fill in these missing values, and why might this be a good choice? 2. Add a new column called 'AgeGroup' that categorises individuals into 'Young', 'Middle-aged', and 'Senior' based on age. Explain your reasoning for creating these categories. 3. How would you filter the Data Frame to show only those individuals who live in 'New York' or 'Boston'? 4. Calculate the average age and average salary of everyone in the Data Frame. What kind of insights can these averages provide? 5. Sort the Data Frame based on 'Salary' in descending order. Why might you want to do this kind of sorting?			

KL – Bloom's Taxonomy Levels
(KL1: Remembering, KL2: Understanding, KL3: Applying, KL4: Analyzing, KL5: Evaluating, KL6: Creating)
CO – Course Outcomes
