RAMAIAH Institute of Technology

Department of Artificial Intelligence and Data Science

QUESTION BANK FOR VI Sem (Term: May-August 2024)

Data Analytics Laboratory (ADL65)

I.A. Marks: 50 **Exam Hours: 03**
Credits: 0:0:1 **Exam Marks: 50**

Sl. No	Question	CO's Mapping	PO and PSO Mapping
1	You are working as a cashier at a grocery store. Your task is to create a program that simulates the checkout process for a customer's shopping cart. The program should calculate the total cost of the items, including tax, and provide a detailed receipt. i. Define a list of products, each represented as a dictionary with keys: "name", "price", and "quantity". ii. Allow the cashier to input the products in the customer's shopping cart, including the name, price, and quantity of each item. iii. Calculate the subtotal (price * quantity) for each item and display a detailed receipt withproduct names, quantities, prices, and subtotals. iv. Calculate the total cost of the items in the cart before tax. v. Apply a tax rate (e.g., 8%) to the total cost to calculate the tax amount. vi. Calculate the final total cost by adding the tax amount to the total cost before tax.	**CO1**	**PO1,3,5 and PSO 1,2**
2	You have been tasked with creating a program that calculates and assigns grades for students enrolled in multiple courses. The program will take input for the marks obtained by 10 students in 5 different courses, compute the total and average marks for each student, and assign corresponding grades based on their average performance. Declare constants for the number of students (num_students) and the number of courses (num_courses). Initialize an empty list to store student information.For each student: i. Input the student's name. ii. Input marks for each of the 5 courses. iii. Calculate the total marks and average marks. iv. Determine the grade based on the average marks using a grading scale. v. Display the student information, including their name, individual course marks, total marks,average marks, and the assigned grade.	**CO1**	**PO1,3,5 a 1,2**
3	You are developing a library management system that needs a fine calculation feature. Write a program that takes the number of days a book is overdue and calculates the fine amount based on the library's policy. The policy states that for the first 7 days, there is no fine. After 7 days, a fixed fine per day is charged. Additionally, there's a cap on the fine amount after 30 days. Input the number of days the book is overdue. i. Use conditional statements to calculate the fine amount based on the library's policy. ii. Display the fine amount along with a message indicating whether the fine is within the cap or exceeded it. Modify this code for 100 students and 5 courses.	**CO1**	**PO1,3,5 and PSO 1,2**
4	You are developing an inventory management system for a small store. The system needs to handle inventory items and their quantities. Write a program that uses arrays to store inventory items and their quantities, and includes functions to add new items, update quantities, and display the inventory. i. Define an array to store inventory items. ii. Define an array to store corresponding quantities. iii. Implement functions to: - Add a new item along with its quantity. - Update the quantity of an existing item. - Display the inventory items and quantities. iv. Use the functions to manage the inventory and handle user interactions.	**CO1**	**PO1,3,5 and PSO 1,2**

<table>
<tr><td>5</td><td>You are working as an educational analyst and need to analyze the performance of students in a school. You have data on student names, their scores in different subjects, and attendance. Write a program that uses data frames to manage and analyze student data, including calculating average scores, identifying students with low attendance, and generating a report.
Create a data frame to store student information with columns: "Name", "Math_Score","Science_Score", "History_Score", "Attendance".
Implement functions to:
i. Calculate the average scores for each student.
ii. Identify students with attendance below a certain threshold.
iii. Generate a report with student names, average scores, and attendance status.
Use the functions to analyze student performance and generate the report</td><td>CO2</td><td>PO1,3,5 and PSO 1,2</td></tr>
<tr><td>6</td><td>You are a data analyst at a retail company that sells products online. The company is interested in predicting sales for the upcoming months to better manage inventory and plan marketing strategies. As part of your role, you need to develop a program that utilizes time series analysis to forecast sales based on a historical sales dataset.
Write an R program to forecast sales for the next three months using time series analysis techniques. The program should perform the following steps:
i. Load the required libraries, including the forecast package.
ii. Create a data frame with two columns: Month and Sales. The Month column should contain a sequence of dates from January 2023 to June 2023 (inclusive), and the Sales column should contain the corresponding sales amounts (12000, 15000, 18000, 16000, 20000, 22000).
iii. Convert the sales data into a time series object with a monthly frequency.
iv. Fit an ARIMA (AutoRegressive Integrated Moving Average) model to the sales time seriesusing the auto.arima() function.
v. Forecast sales for the next three months using the fitted ARIMA model and the forecast()function.
vi. Display the forecasted sales results, including point forecasts and prediction intervals.</td><td>CO2</td><td>PO1,3,5 and PSO 1,2</td></tr>
<tr><td>7</td><td>You are a data analyst working for an e-commerce company that specializes in selling a variety of products online. The company aims to analyze customer purchase data comprehensively to gain insights into customer behaviour and spending patterns.
Your goal is to develop a R program that performs an in-depth analysis of customer purchase data. You will calculate various statistical measures and generate visualizations to understand the distribution of purchase amounts among customers.
Note: Load the necessary libraries, including the dplyr and ggplot2 packages.
Given the example customer purchase data provided below, create a data frame named purchase_data with two columns: CustomerID and PurchaseAmount.
Calculate and display the following statistical measures:
i. Mean (average) purchase amount
ii. Median purchase amount
iii. Standard deviation of purchase amounts
iv. 1st quartile (25th percentile) of purchase amounts
v. 3rd quartile (75th percentile) of purchase amounts
Create a histogram to visualize the distribution of purchase amounts using the ggplot2 package.Display the histogram with appropriate labels and titles.
Example Customer Purchase Data:
<table><tr><th>Customer ID</th><th>Purchase Amount</th></tr><tr><td>101</td><td>150</td></tr><tr><td>102</td><td>200</td></tr><tr><td>103</td><td>120</td></tr><tr><td>104</td><td>300</td></tr><tr><td>105</td><td>80</td></tr></table></td><td>CO2</td><td>PO1,3,5 and PSO 1,2</td></tr>
<tr><td>8</td><td>Write an R program that generates two matrices, matrix_A and matrix_B, and conducts operations including element-wise addition, scalar multiplication, matrix transpose, and multiplication.</td><td>CO1</td><td>PO1,3,5 and PSO 1,2</td></tr>
</table>

9	You are a data analyst tasked with analyzing and visualizing a dataset. The dataset contains information about student performance in a course. You need to create a program that generates various types of plots to help understand and present the data effectively. Write a program that performs data analysis and generates visualizations for a given dataset. The program should: a. Load the necessary libraries (ggplot2). b. Prepare example data with columns for student names, scores, and attendance percentages. c. Perform the following tasks: i. Create a scatter plot to visualize the relationship between scores and attendancepercentages. ii. Generate a bar plot to show the distribution of scores among different students. iii. Create a line plot to display the trend of scores over time (assuming different students'scores were collected at different time intervals). iv. Generate a histogram to visualize the distribution of scores. d. Customize the appearance of each plot, such as color, labels, and titles. e. Arrange the plots in a way that they are easy to compare and understand. f. Provide appropriate titles for each plot and the axes.	**CO3**	**PO1,3,5 and PSO 1,2**
10	In this exercise, you will delve into data manipulation using the dplyr package in R. You will perform operations that empower you to filter, select, mutate, group, summarize, arrange, and join data frames. The dplyr package provides a streamlined approach to enhance your ability to manipulate and transform data efficiently, making it a vital tool for data analysis tasks. i. Filter and Select: Apply the filter() function to extract rows that satisfy certain conditions from a given data frame. Additionally, employ the select() function to choose specific columns from the data frame. ii. Mutate: Utilize the mutate() function to create new variables or modify existing ones within the data frame, thus enriching your dataset with calculated values. iii. Group and Summarize: Harness the power of the group_by() function to group data based on specific variables. Then,employ the summarize() function to compute summary statistics within each group. iv. Arrange: Leverage the arrange() function to sort the data frame based on the values of selectedvariables, helping you gain insights from organized data. v. Join: Explore the join() function, which allows you to merge multiple data frames based oncommon variables, thereby combining information for a comprehensive analysis.	**CO3**	**PO1,3,5 and PSO 1,2**
11	You are a data analyst at an e-commerce company that sells a variety of products online. The company has provided you with a dataset containing information about customer purchases. Your task is to perform a comprehensive data analysis to gain insights into customer behavior and spending patterns. *Dataset Description:* The dataset customer_purchases.csv contains the following columns: Customer ID: Unique identifier for each customer. Purchase Amount: The amount spent by the customer on a purchase. **Problem Tasks:** You are required to perform the following tasks using R: Task 1: Load the Dataset Load the necessary libraries, including readr and dplyr. Read the dataset customer_purchases.csv into a data frame named purchase_data. Task 2: Data Summary Calculate and display the total number of records in the dataset. Calculate and display the total number of unique customers in the dataset.	**CO3**	**PO1,3,5 and PSO 1,2**

	Task 3: Calculate Statistical Measures Calculate and display the mean (average) purchase amount.Calculate and display the median purchase amount. Calculate and display the standard deviation of purchase amounts. Task 4: Customer Segmentation Create a new column named Segment in the purchase_data data frame based on thefollowing criteria: "Low Spender" if the purchase amount is less than the median. "High Spender" if the purchase amount is greater than or equal to the median. Task 5: Visualize Data Create a histogram to visualize the distribution of purchase amounts using the ggplot2package. Customize the plot with appropriate labels and titles.		
12	You are a data analyst tasked with performing an Exploratory Data Analysis (EDA) on the Indian Premier League (IPL) dataset. The IPL dataset contains information about various IPL matches, including match dates, teams, venues, outcomes, and performance metrics. Your objective is to gain insights into the dataset by conducting an in-depth analysis using R programming. Dataset Description: The dataset named "ipl_data.csv" includes the following columns: Match_ID: Unique identifier for each match.Date: Date of the match. Team1: Name of the irst team participating in the match. Team2: Name of the second team participating in the match.Venue: Stadium where the match was played. Winner: Name of the winning team. Total.Runs: Total runs scored in the match. Total.Wickets: Total wickets taken in the match. Other relevant columns (if any). **Problem Tasks:** Your task is to perform the following Exploratory Data Analysis (EDA) tasks using R:Task 1: Data Overview and Structure Display the structure of the dataset using str() function.\ Output summary statistics of numerical columns using summary (). Task 2: Basic Data Insights Calculate and display the total number of matches in the dataset. Determine the number of unique teams that have participated in IPL matches.List the unique teams from both Team1 and Team2. Task 3: Team Performance Analysis Calculate the number of matches won by each team and display the results.Compute the average total runs scored in the matches. Calculate the average total wickets taken in the matches.\ Task 4: Venue Insights Identify and display the most frequently used venue for matches. Task 5: Visualization Create a bar plot to visualize the number of matches won by each team.	**CO3**	**PO1,3,5 and PSO 1,2**

Marks Distribution:

Write-Up (8)	Execution (35)	Viva/Demo	Change of Program	Total
08	35	07	-10 Marks	50 Marks