SAVEETHA SCHOOL OF ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

SAVEETHA UNIVERSITY

Department of Computer Science and Engineering

Course Code: DSA0416 (C-Slot)	**Course Name: Fundamentals of Data Science for Risk Management**	
Branch: AI&DS	**Year: II/III/IV**	**Date: 18.07.2024**

1.you are a data analyst working for an e-commerce company. The company has provided you with a dataset containing information about customer orders. The dataset includes columns such as OrderID, CustomerID, ProductID, Quantity, and TotalPrice. Your task is to use Panda's data frames to analyze and derive insights from the dataset.

2.You are a data scientist working for a medical research institute. The institute is conducting a study to understand the relationship between smoking habits and the incidence of lung cancer among a group of individuals. As part of your analysis, you are tasked with calculating the correlation coefficient between smoking and lung cancer rates and creating a scatter plot to visualize the data.

3.You work as a data analyst for a large e-commerce company that sells a variety of products online. Your company has collected sales data over the past year and wants to analyze and visualize this data to gain insights into sales trends, product performance, and customer behavior. To understand which product categories are most popular, create line, scatter and bar plots that display sales distribution across different product categories. Each plot has to represents a category, and the height of the bar indicates the total sales

4.Scenario: You are working on a data analysis project that involves analyzing the monthly temperature and rainfall data for a city. You have a dataset containing the monthly temperature and rainfall values for each month of a year. Your task is to develop a Python program that generates line plots and scatter plots to visualize the temperature and rainfall Data.

5.Scenario: You are working on a data visualization project and need to create basic plots using Matplotlib. You have a dataset containing the monthly sales data for a company, including the month and corresponding sales values. Your task is to develop a Python

program that generates line plots and bar plots to visualize the sales data.
Question:
1. How would you develop a Python program to create a line plot of the monthly sales data?

2: How would you develop a Python program to create a bar plot of the monthly sales data?