**DSA0402 – FUNDAMENTAL OF DATA SCIENCE**

**OUTPUTS**

1. to calculate the average score for each subject and determine the subject with the highest average score? Assume 4x4 matrix that stores marks of each student in given How would you find the top 5 products that have been sold the most in the past month?order.

A screenshot of a computer program

Description automatically generated

2. find the average price of all the pr How would you find the top 5 products that have been sold the most in the past month?oducts sold in the past month? Assume 3x3 matrix with each row representing the sales for a different product

A screenshot of a computer

Description automatically generated

3. find the average sale price of houses with more than four bedr How would you find the top 5 products that have been sold the most in the past month?ooms in the neighborhood

A screenshot of a computer program

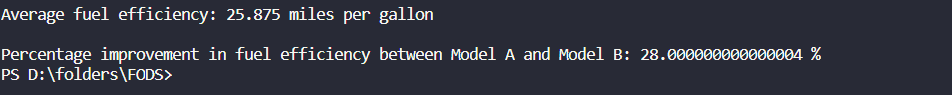
Description automatically generated

4. calculate the total sales for the year and determine the perc How would you find the top 5 products that have been sold the most in the past month?entage increase in sales from the first quarter to the fourth quarter

A close up of a screen

Description automatically generated

5. How would you use NumPy arrays and How would you find the top 5 products that have been sold the most in the past month? arithmetic operations to calculate the average fuel efficiency and determine the percentage improvement in fuel efficiency between two car models



6. Use arithmetic operations to calculate How would you find the top 5 products that have been sold the most in the past month? the total cost of a customer's purchase, including discounts and taxes, given the item prices, quantities, discount rate, and tax rate?

A screenshot of a computer

Description automatically generated

7. Using Pandas DataFrame operations, how would you find the following information

from the order\_data DataFrame:

1. The total number of orders made by each customer.

2. The average order quantity for each product.

3. The earliest and latest order dates in the dataset

A screenshot of a computer

Description automatically generated

8. How would you find the top 5 products that have been sold the most in the past month?

A screen shot of a computer

Description automatically generated

9. Using Pandas DataFrame operations, how would you find the following information

from the property\_data DataFrame:

1. The average listing price of properties in each location.

2. The number of properties with more than four bedrooms.

3. The property with the largest area.

A screenshot of a computer

Description automatically generated

10.

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?

A graph and chart of a graph

Description automatically generated with medium confidence

11.

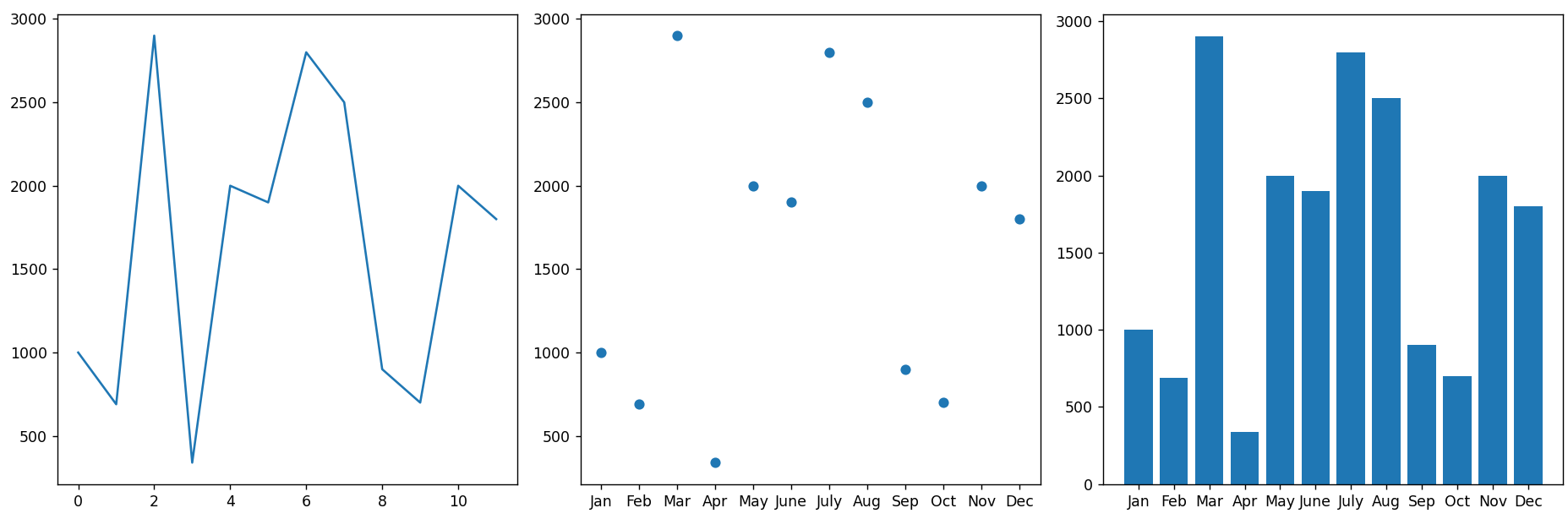
1. Write code to create a simple line plot in Python using Matplotlib to predict sales happened in a

month?

2. Write code to create a scatter plot in Python using Matplotlib to predict sales happened in a

month?

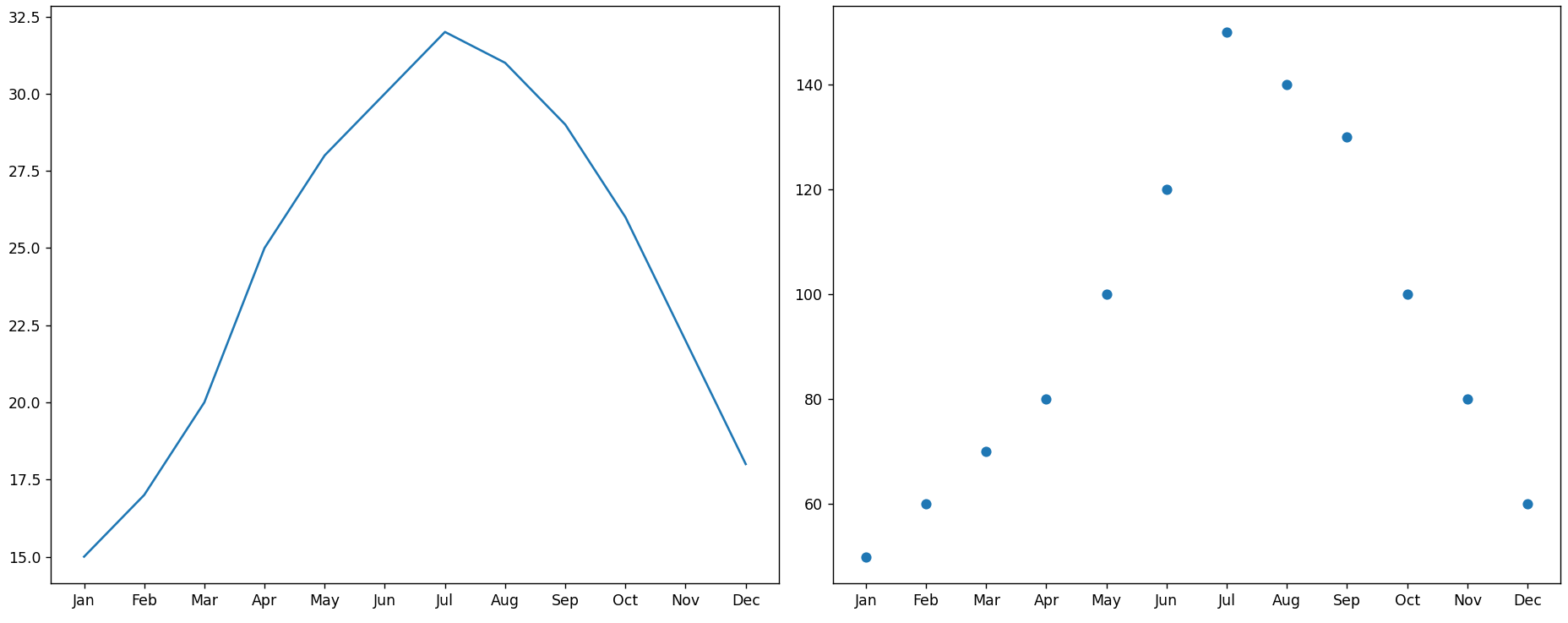
3. Develop a Python program to create a bar plot of the monthly sales data



12.

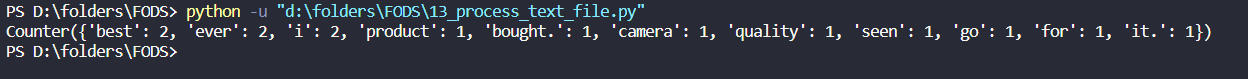
1. Develop a Python program to create a line plot of the monthly temperature data.

2: Develop a Python program to create a scatter plot of the monthly rainfall data



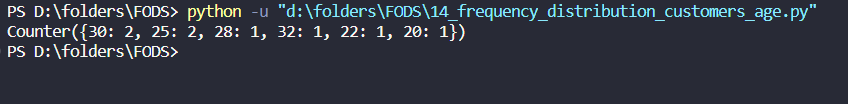
13. How would you develop a Python program to calculate the frequency distribution of

words in a text document?



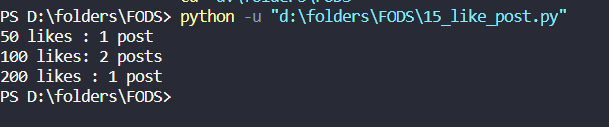
14. Develop a code in python to find the frequency distribution of the ages of the customers

who have made a purchase in the past month



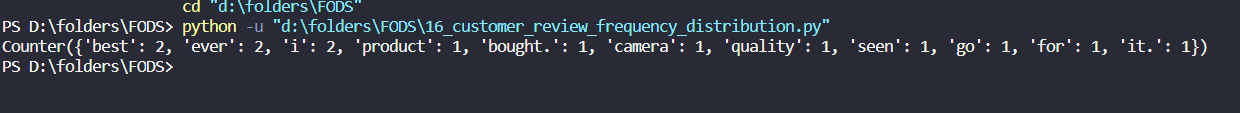
15. Develop a Python program to calculate the frequency distribution of likes among the

posts



16. Develop a Python program to calculate the frequency distribution of words in the

customer reviews dataset?



17. Create a Python program that fulfills these requirements and helps your team gain

insights from the customer feedback data.

A screen shot of a computer

Description automatically generated

A graph of blue rectangular objects

Description automatically generated with medium confidence

18.

 Calculate the mean, median and standard deviation of age and %fat using Pandas.  Draw the boxplots for age and %fat.  Draw a scatter plot and a q-q plot based on these

A screenshot of a computer

Description automatically generated

A diagram of a graph

Description automatically generated with medium confidence

A graph with blue dots

Description automatically generated

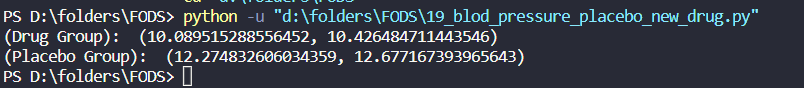
A red line with blue dots

Description automatically generated

19. What is the 95% confidence interval for the mean reduction in blood pressure for patients who

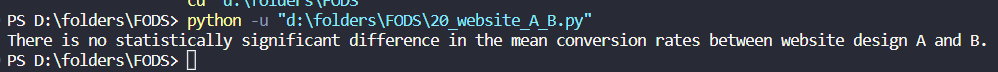
received the new drug? Also, what is the 95% confidence interval for the mean reduction in blood

pressure for patients who received the placebo?

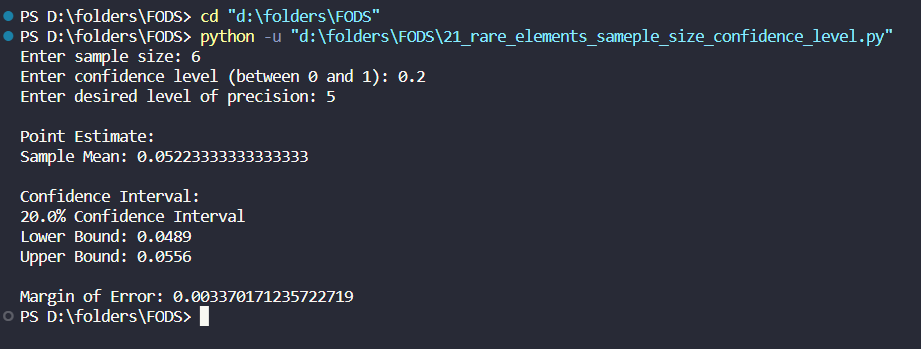


20. "Based on the data collected from the A/B test, is there a statistically significant difference in the

mean conversion rates between website design A and website design B?"



21. write a Python program that allows the user to input the sample size, confidence level, and desired level of precision.



22. You will use the pandas library to calculate confidence intervals to estimate the true population

mean rating. You have been provided with a CSV file named "customer\_reviews.csv," which contains customer ratings for products in the chosen category

A screen shot of a computer

Description automatically generated

23. to analyze the data using hypothesis testing and calculate the p-value to determine if the new treatment has a statistically significant effect compared to the placebo. You will use the matplotlib library to visualize the data and the p-value.

A screen shot of a computer

Description automatically generated

24. K-Nearest Neighbors (KNN) Classifier

A computer screen shot of a computer code

Description automatically generated

25. Decision Tree for Iris Flower Classification

A screen shot of a computer

Description automatically generated

26. Linear Regression for Housing Price Prediction

A computer screen shot of a computer code

Description automatically generated

27. Logistic Regression for Customer Churn Prediction

A computer screen with white text

Description automatically generated

28. K-Means Clustering for Customer Segmentation

A screen shot of a computer program

Description automatically generated

29. Evaluation Metrics for Model Performance

A close up of a screen

Description automatically generated

30. Classification and Regression Trees (CART) for Car Price Prediction

A computer screen shot of a computer program

Description automatically generated

31. Classification and Regression Trees (CART) algorithm from scikit-learn to predict the price of the new car based on the input features

A screenshot of a computer screen

Description automatically generated

32. Your task is to use Python and clustering algorithms to segment the customers into different groups based on their behavior and characteristics. The marketing team will use these segments to tailor their marketing campaigns and promotions effectively.

A screen shot of a computer

Description automatically generated

A graph with a red line and blue dots

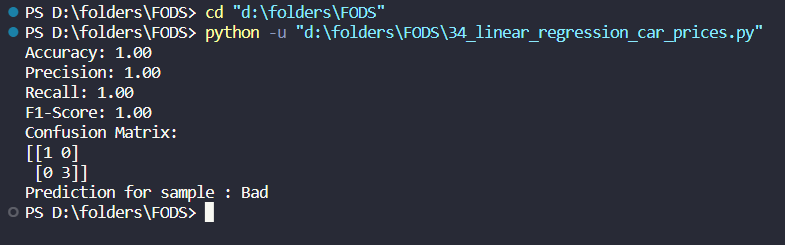
Description automatically generated

33. Your task is write a Python program to perform bivariate analysis and build a linear regression model to predict house prices based on a selected feature (e.g., house size) from the dataset. Additionally, you need to evaluate the model's performance to ensure its accuracy and reliability

A computer screen with white text

Description automatically generated

34. Your task is write a Python program that perform linear regression modeling to predict car prices based on a selected set of features from the dataset. Additionally, you need to evaluate the model's performance and provide insights to the marketing team to understand the most influential factors affecting car prices



35. Your task is to build a classification model using the KNN algorithm to predict the treatment outcome ("Good" or "Bad") for new patients based on their features. Evaluate the model's performance using accuracy, precision, recall, and F1-score.Make predictions on the test set and display the results.

A diagram of a customer segmentation

Description automatically generated

36. Your task is to build a clustering model using the K-Means algorithm to group customers into distinct segments based on their spending patterns

A computer screen shot of a computer code

Description automatically generated

37. Identify any potential correlation between study time and exam scores and explore various plotting functions to visualize this relationship effectively.

A graph with blue dots

Description automatically generated

A computer screen with blue text

Description automatically generated

A graph with blue dots

Description automatically generated

38. Write a python program will take in a dataset containing daily temperature readings for

each city over a year and perform the following tasks:

1. Calculate the mean temperature for each city.

2. Calculate the standard deviation of temperature for each city.

3. Determine the city with the highest temperature range (difference between the highest and

lowest temperatures).

4. Find the city with the most consistent temperature (the lowest standard deviation).

A screen shot of a computer

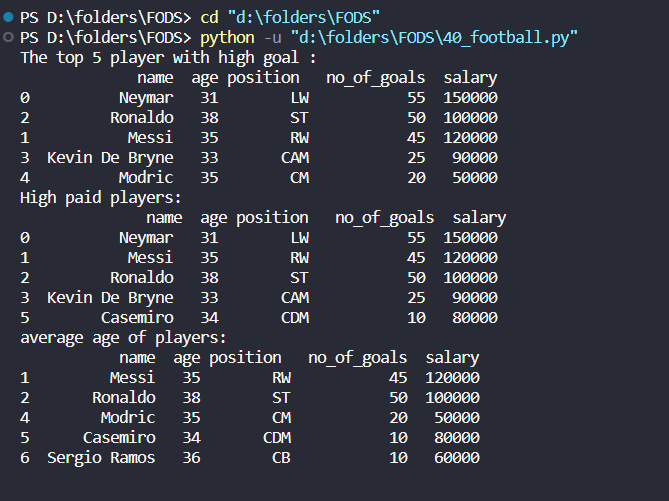
Description automatically generated

39. Build a clustering model using the K-Means algorithm to group customers based on their spending and purchase behavior and visualize the clusters using scatter plots or other appropriate visualizations to gain insights into customer distribution and distinguish different segments.

A graph with colored dots and numbers

Description automatically generated

40. Develop a Python program to read the data from the CSV file into a pandas data frame, to find the top 5 players with the highest number of goals scored and the top 5 players with the highest salaries. Also calculate the average age of players and display the names of players who are above the average age and visualize the distribution of players based on their positions using a bar chart



A graph of blue bars

Description automatically generated with medium confidence