**Name: CH. Ajith Kumar**

**Reg No:192324257**

**Course Code: CSA0406**

**Course Name: Fundamental Data Science**

11. **Scenario :** You are a data scientist working for a company that sells products online. You have been tasked with creating a simple plot to show the sales of a product over time.

**Question:**

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.

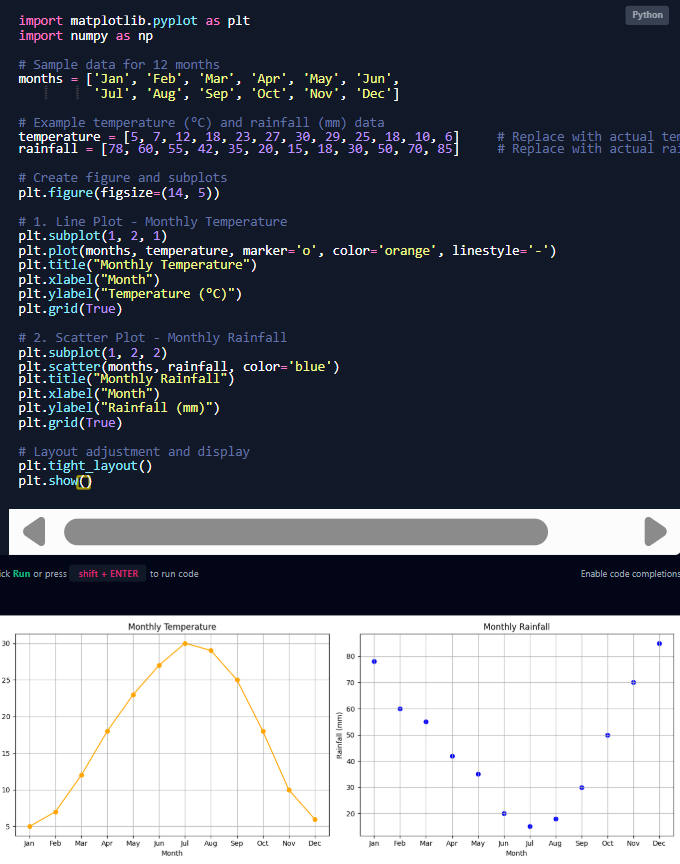


12. **Scenario:** You are working on a data analysis project that involves analyzing the monthly temperature and rainfall data for a city. You have a dataset containing the monthly temperature and rainfall values for each month of a year. Your task is to develop a Python program that generates line plots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

**Question:**

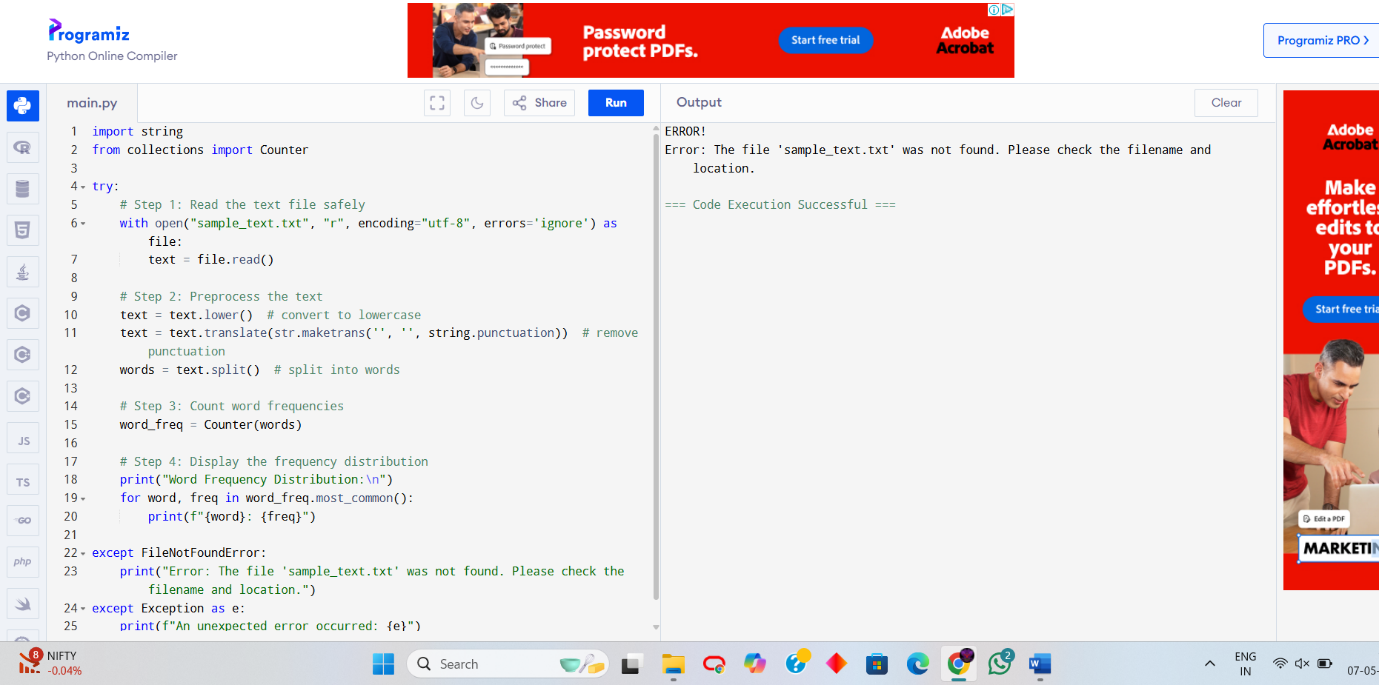
1. Develop a Python program to create a line plot of the monthly temperature data.

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



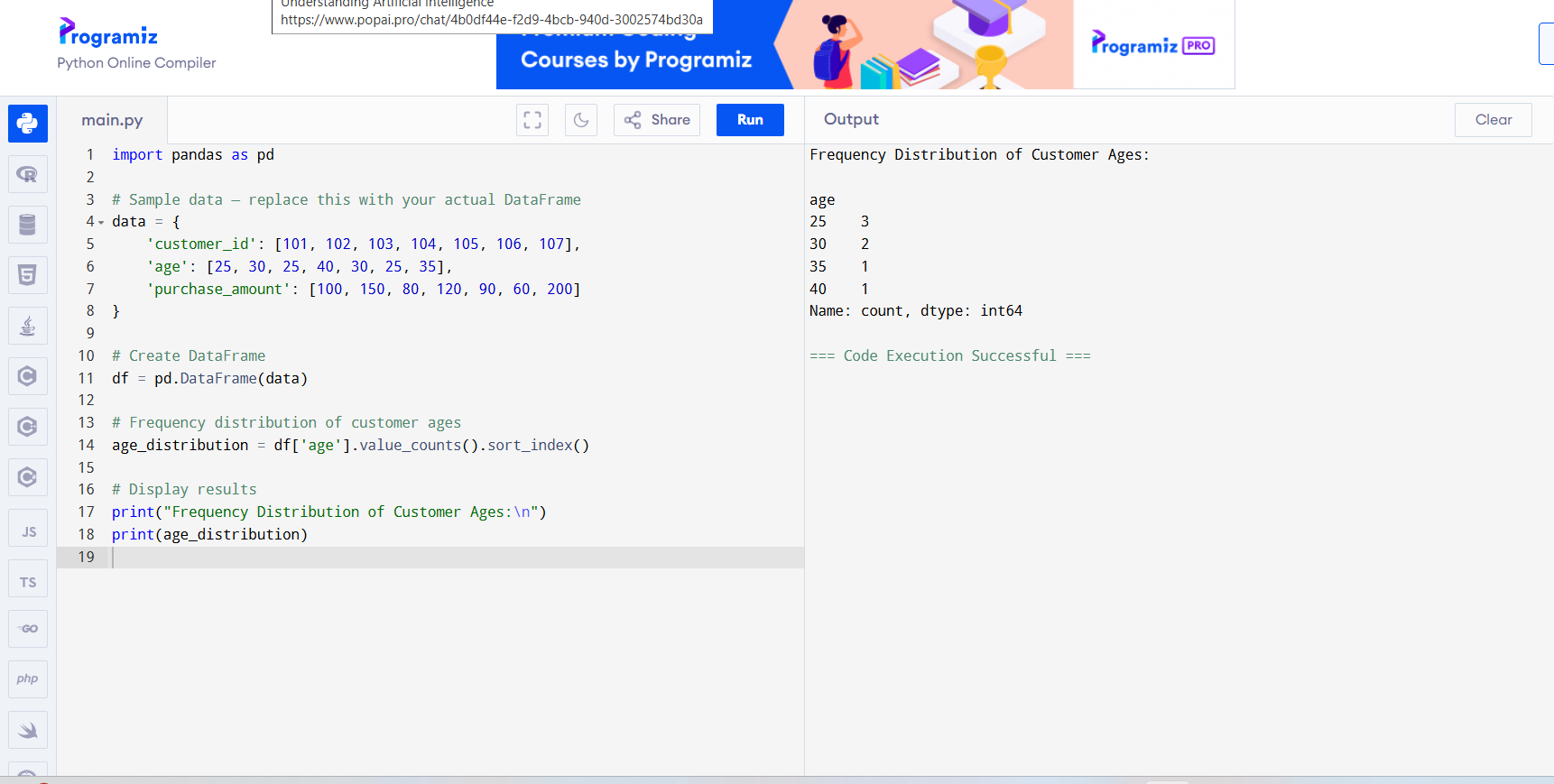
13. **Scenario:** You are working on a text analysis project and need to determine the frequency distribution of words in a given text document. You have a text document named "sample\_text.txt" containing a paragraph of text. Your task is to develop a Python program that reads the text document, processes the text, and generates a frequency distribution of the words.

**Question:** How would you develop a Python program to calculate the frequency distribution of words in a text document?



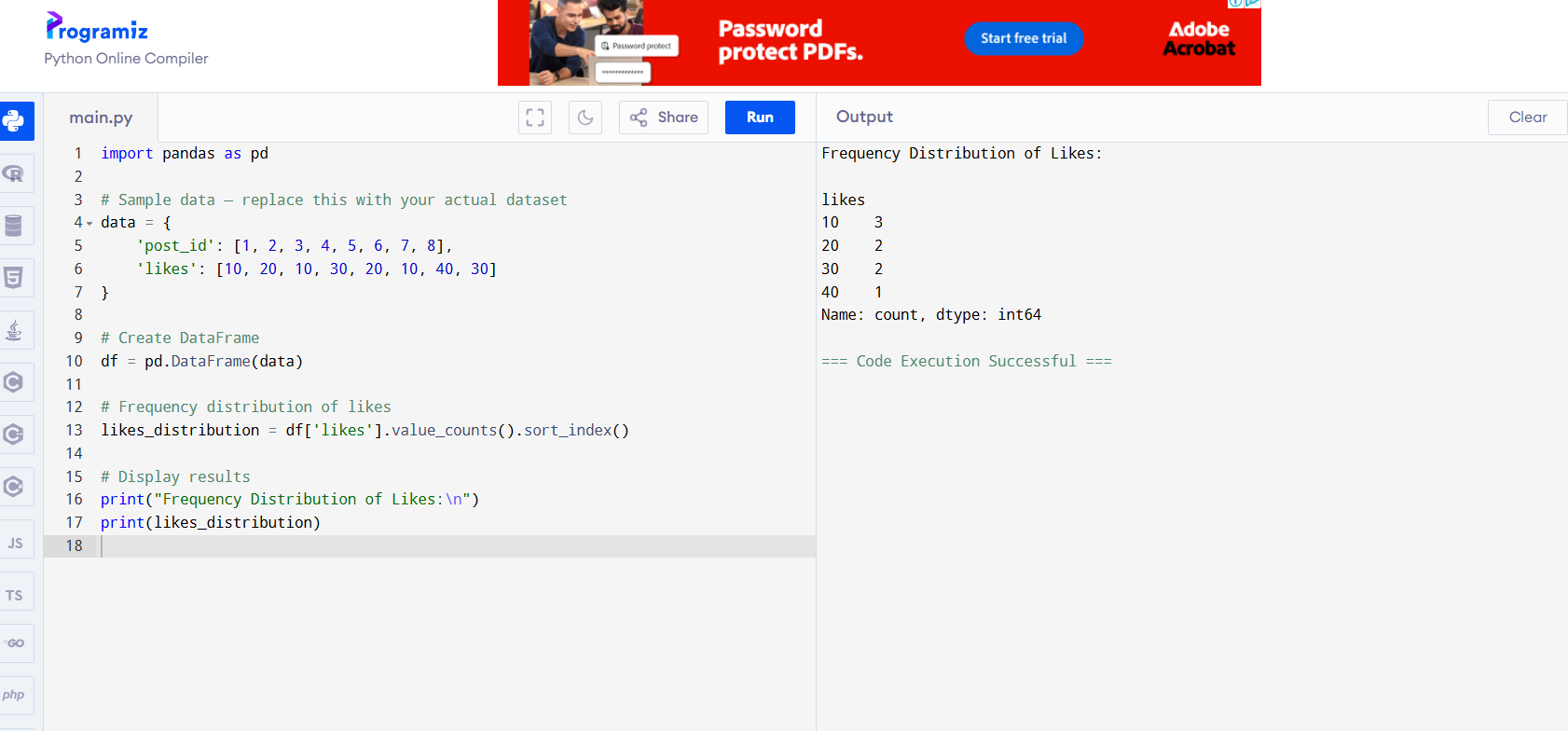
14. **Scenario**: You are a data analyst working for a company that sells products online. You have been tasked with analyzing the sales data for the past month. The data is stored in a Pandas data frame.

**Question:** 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.



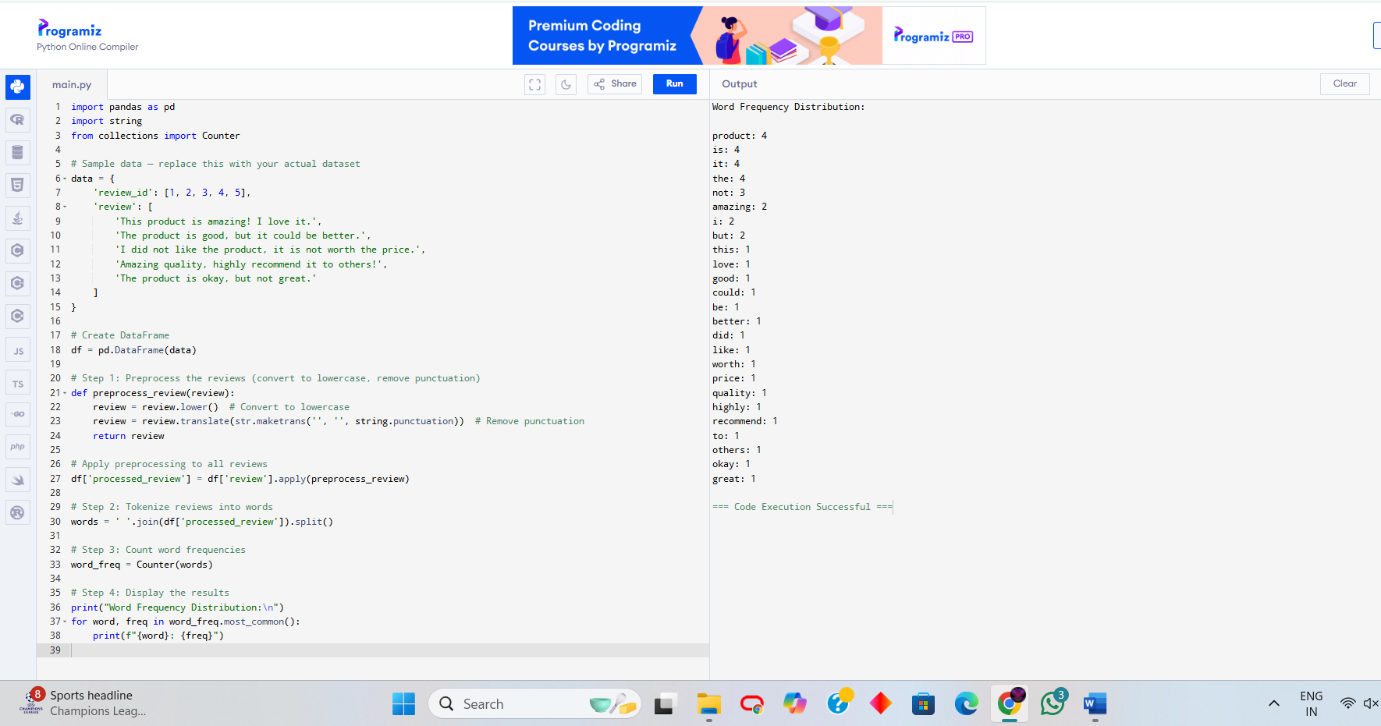
15. **Scenario:** You are a data analyst working for a social media platform. As part of your analysis, you have a dataset containing user interaction data, including the number of likes received by each post. Your task is to develop a Python program that calculates the frequency distribution of likes among the posts.

**Question:** Develop a Python program to calculate the frequency distribution of likes among the posts?



16. **Scenario:** You are working on a project that involves analyzing customer reviews for a product. You have a dataset containing customer reviews, and your task is to develop a Python program that calculates the frequency distribution of words in the reviews.

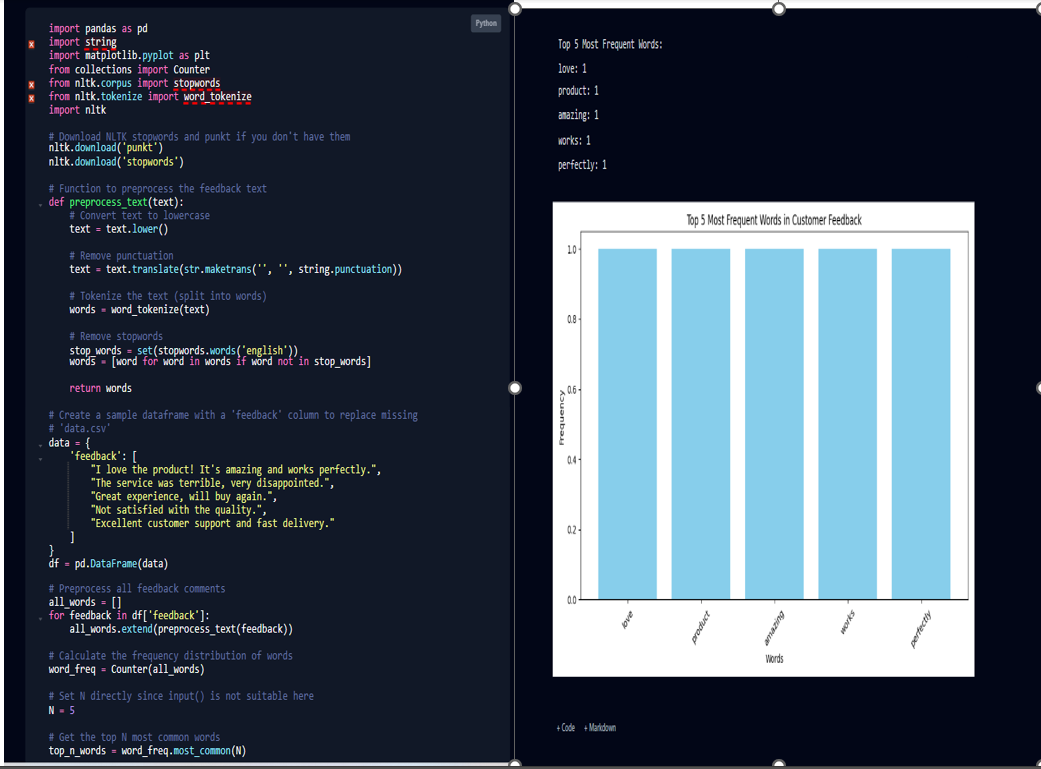
**Question:** Develop a Python program to calculate the frequency distribution of words in the customer reviews dataset?



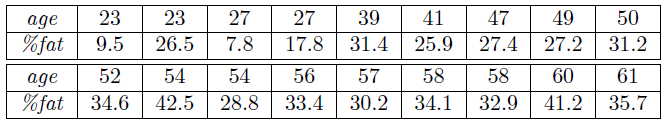
17. **Scenario:** You are a data analyst working for a marketing research company. Your team has collected a large dataset containing customer feedback from various social media platforms. The dataset consists of thousands of text entries, and your task is to develop a Python program to analyze the frequency distribution of words in this dataset. Your program should be able to perform the following tasks:

* Load the dataset from a CSV file (data.csv) containing a single column named "feedback" with each row representing a customer comment.
* Preprocess the text data by removing punctuation, converting all text to lowercase, and eliminating any stop words (common words like "the," "and," "is," etc. that don't carry significant meaning).
* Calculate the frequency distribution of words in the preprocessed dataset.
* Display the top N most frequent words and their corresponding frequencies, where N is provided as user input.
* Plot a bar graph to visualize the top N most frequent words and their frequencies.

**Question**: Create a Python program that fulfills these requirements and gain insights from the customer feedback data.

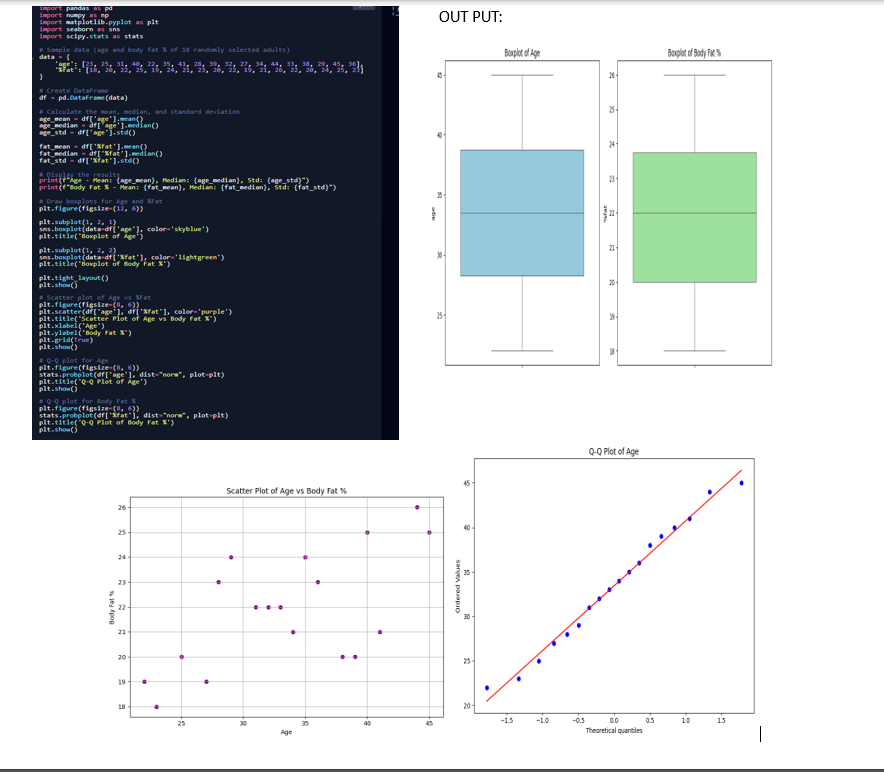


18. Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the following result.

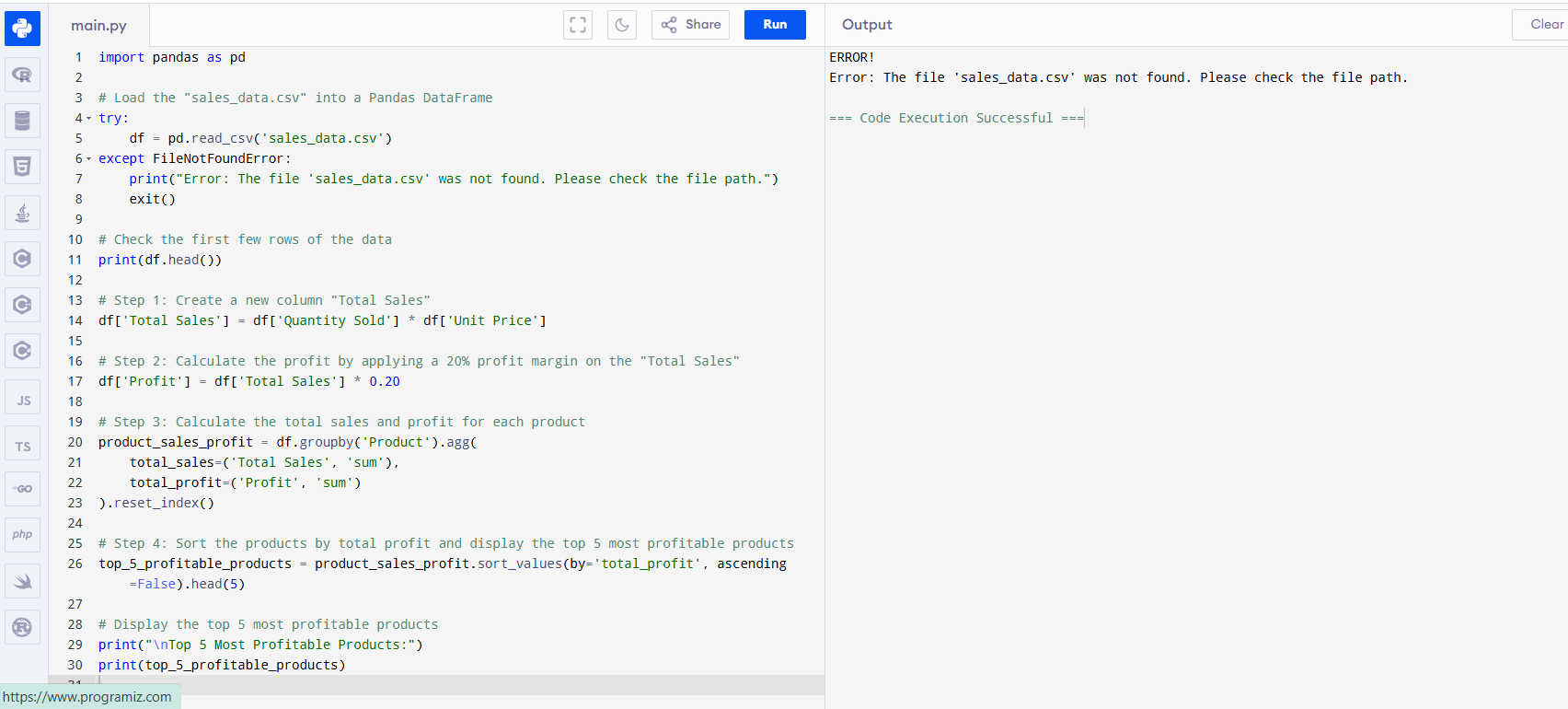


**Question:**

* 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 two variables.



19. Sales and Profit Analysis: a) Load the "sales\_data.csv" file into a Pandas data frame, which contains columns "Date," "Product," "Quantity Sold," and "Unit Price." b) Create a new column named "Total Sales" that calculates the total sales for each transaction (Quantity Sold \* Unit Price). c) Calculate the total sales for each product and the overall profit, considering a 20% profit margin on each product. Display the top 5 most profitable products.



20. Customer Segmentation: a) Load the "customer\_data.csv" file into a Pandas data frame, which contains columns "Customer ID," "Age," "Gender," and "Total Spending." b) Segment customers into three groups based on their total spending: "High Spenders," "Medium Spenders," and "Low Spenders." Assign these segments to a new column in the data frame. c) Calculate the average age of customers in each spending segment.

