Each row's first student will write the SET A question paper, the second student will write SET B, and so on.

Write your code in Jupiter Notebook and upload it separately on VTOP as follows:

First question → Upload as DA4

Second question → Upload as DA5

In the provided answer sheet, include:

Aim

Procedure

Methodology

Analysis

Inference/Conclusions
Get the signature from the TRA and submit the answer sheet to the TRA.

Programme Name & Branch	: M.Sc. & Data Science
Slot	: L29+L30
Course Name & code	: Regression Analysis and Predictive Models Lab & PMDS504P
Faculty Name	: Dr. Jisha Francis
Due Date & Max. Marks	: 5 April, 2025 & 40 Marks

Instructions

(a) Answer all questions provided below.

(b) Ensure your code is well-commented and easy to understand.

(c) **Each question carries 20 marks, distributed as follows:**

- **10 marks for correct Python code.**
- **5 marks for correct objective, methodology, and data analysis.**
- **5 marks for conclusions and inferences.**

Dataset: The dataset `shampoo-sales.csv` contains monthly shampoo sales data for a period of three years. This section is divided into two parts for ease of execution and assessment.

Question 1 (20 Marks) - Data Understanding & Visualization

You are provided with a dataset `shampoo-sales.csv` containing monthly shampoo sales over a three-year period.

(a) **Data Preprocessing (6 Marks)**

- Load the dataset using pandas.
- Rename the columns to appropriate names.
- Check and handle missing values.

(b) **Exploratory Data Analysis (14 Marks)**

- Display summary statistics (count, mean, std, min, max, etc.).
- Plot a line chart to visualize monthly shampoo sales over time.
- Include labeled axes and a grid for clarity.

Question 2 (20 Marks) - Stationarity & Autocorrelation

(a) **Stationarity Check using ADF Test (8 Marks)**

- Perform the Augmented Dickey-Fuller (ADF) test.
- Print the ADF statistic, p-value, and critical values.
- Add a comment interpreting whether the series is stationary based on the p-value.

(b) **Autocorrelation Analysis (7 Marks)**

- Plot the Autocorrelation Function (ACF) up to 20 lags.
- Comment on the autocorrelation behavior of the time series.

****

Programme Name & Branch	: M.Sc. & Data Science
Slot	: L29+L30
Course Name & code	: Regression Analysis and Predictive Models Lab & PMDS504P
Faculty Name	: Dr. Jisha Francis
Due Date & Max. Marks	: 5 April, 2025 & 10 Marks

Instructions

(a) Answer all questions provided below.

(b) Ensure your code is well-commented and easy to understand.

(c) **Each question carries 20 marks, distributed as follows:**

- **10 marks for correct Python code.**
- **5 marks for correct objective, methodology, and data analysis.**
- **5 marks for conclusions and inferences.**

Dataset: You are provided with a monthly sunspot activity dataset `Sunspot.csv`. This dataset contains monthly sunspot observations over several decades. Perform the following tasks using Python:

Question 1 (20 Marks) Data Understanding & Visualization

(a) **Data Preprocessing (5 Marks)**

- Load the dataset.
- Rename columns appropriately (if necessary).
- Convert the `Month` column to datetime and set as index.
- Handle any missing values.

(b) **Exploratory Data Analysis (10 Marks)**

- Display summary statistics (count, mean, std, min, max, etc.).
- Plot a line chart showing sunspot activity over time.

Question 2 (20 Marks) -Stationarity Check & Autocorrelation Analysis

(a) **Stationarity Check (10 Marks)**

- Apply the Augmented Dickey-Fuller (ADF) test.
- Print the ADF statistic, p-value, and critical values.
- Add a comment interpreting whether the series is stationary based on the p-value.

(b) **Autocorrelation Analysis (5 Marks)**

- Plot the Autocorrelation Function (ACF) for up to 40 lags.
- Add a comment interpreting the autocorrelation behavior.

****