LAB No. 1

Introduction to VS Code and Google Colab for Data Analysis Using Python

LAB Description

In this lab, students will learn how to work with Python programming environments using Visual Studio Code (VS Code) and Google Colab. VS Code is a lightweight and powerful source-code editor that runs on a local system and supports Python development through extensions. It allows users to write, run, and debug Python programs efficiently on their own computer.

Google Colab is a cloud-based Python notebook environment provided by Google that runs in a web browser. It does not require any local installation and provides free access to computing resources. Colab is especially useful for data analysis and visualization, as it supports interactive notebooks, built-in libraries, and easy file uploads.

Using either VS Code or Google Colab, students will create a dataset in Python, upload or load the data, and perform basic data analysis operations. The lab focuses on understanding how datasets are handled, how simple statistics are calculated, and how graphical representations help in interpreting data.

Lab Objective

The objectives of this lab are:

- To understand the basic functioning of VS Code and Google Colab
- To learn how to create and upload a dataset using Python
- To perform basic statistical analysis (such as mean, median, and count)
- To visualize data using simple graphs (line charts, bar charts, or histograms)
- To develop foundational skills in data analysis and visualization using Python

How to use python in VS code?

1. Create environment named 'venv'

python -m venv

2. Activate environment

venv\Scripts\activate

3. Select Environment in VS CodeAfter creating the environment:

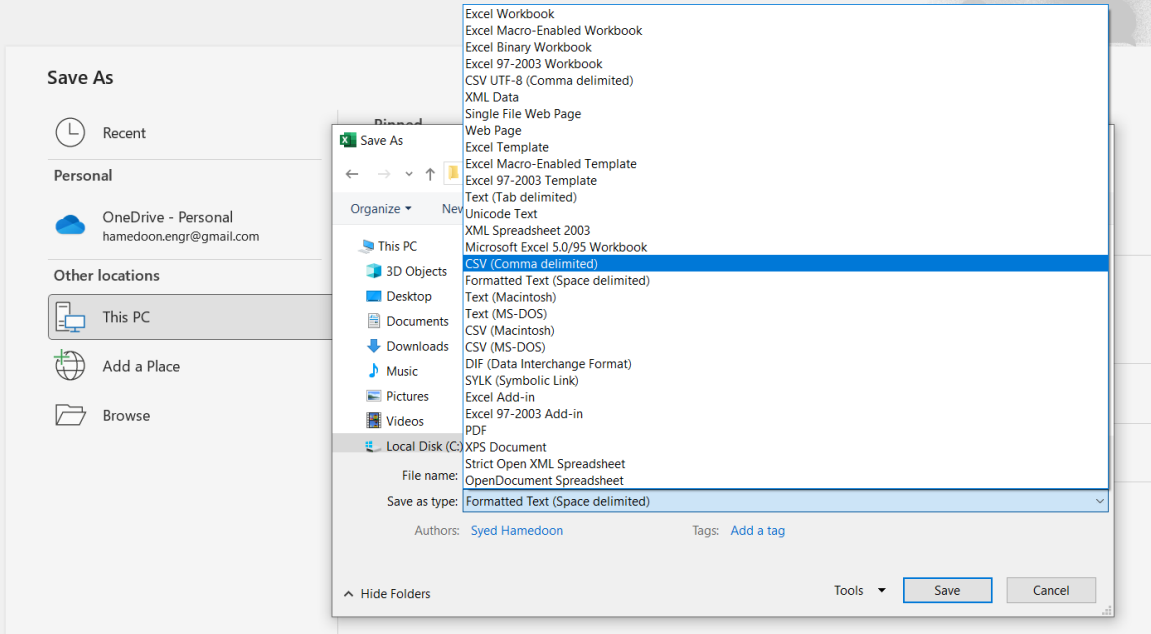
1) Open your project folder in VS Code.
2) Press Ctrl + Shift + P → search for Python:
3) Select Interpreter.
4) Choose your newly created environment (venv or myenv).

4. Install Machine Learning and Deep Learning Libraries

- pip install –upgrade pip
- pip install pandas
- pip install matplotlib
- pip install seaborn
- pip install scikit-learn
- pip install scipy
- pip install numpy
- pip install xgboost
- pip install lightgbm
- pip install catboost
- pip install tensorflow
- pip install keras
- pip install torch

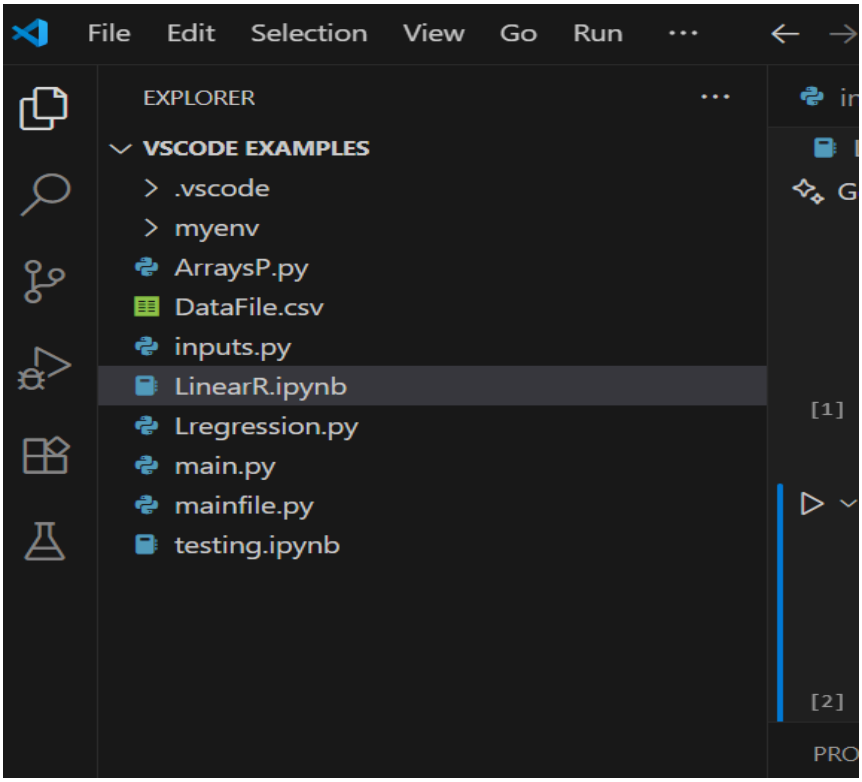
Task 1:

Save Excel file as .csv (Comma delimited)

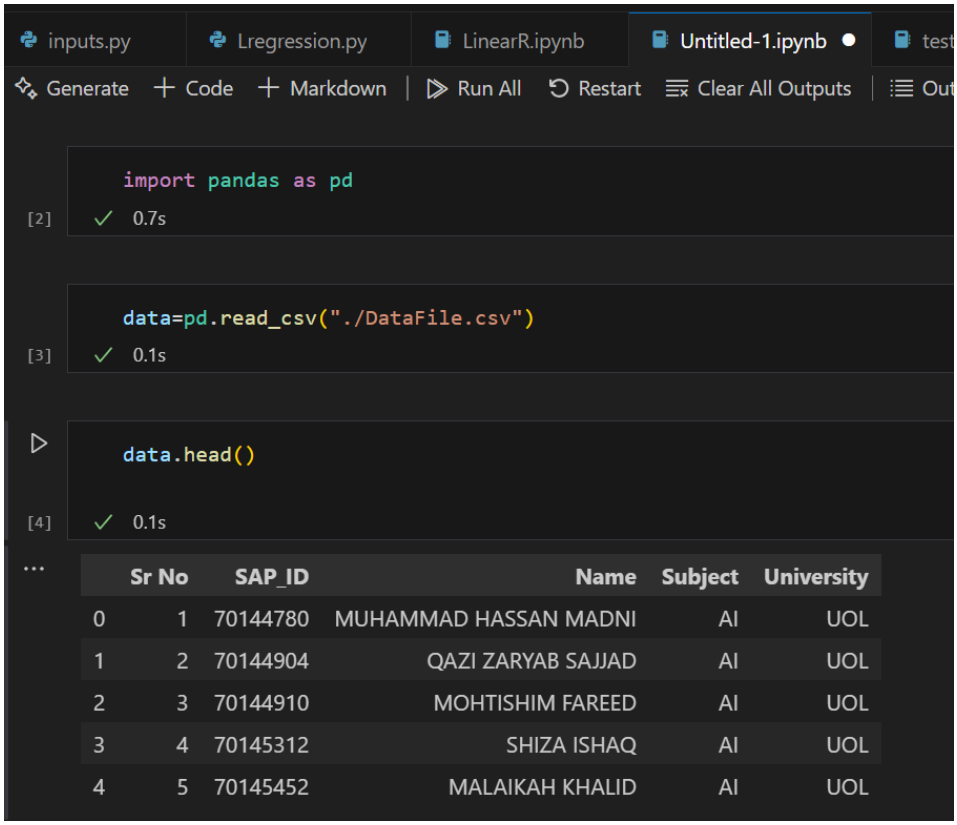


Task 2:

Save .csv file in directory and folder where we created a python environment



Here we can see our file successfully

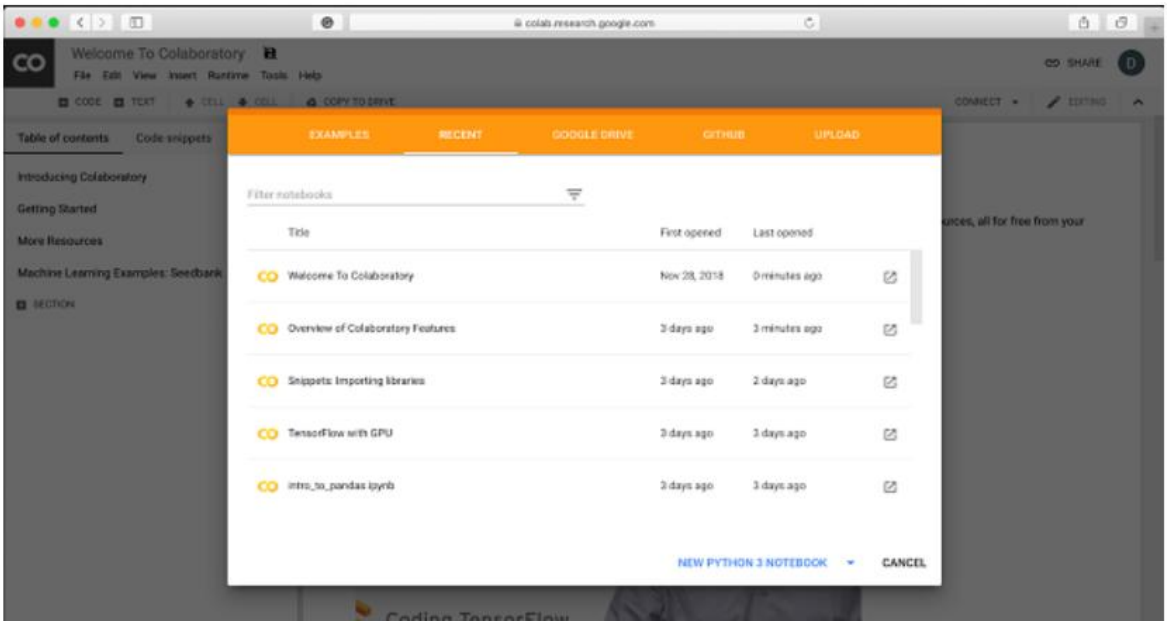


	Sr No	SAP_ID	Name	Subject	University
0	1	70144780	MUHAMMAD HASSAN MADNI	AI	UOL
1	2	70144904	QAZI ZARYAB SAJJAD	AI	UOL
2	3	70144910	MOHTISHIM FAREED	AI	UOL
3	4	70145312	SHIZA ISHAQ	AI	UOL
4	5	70145452	MALAIKAH KHALID	AI	UOL

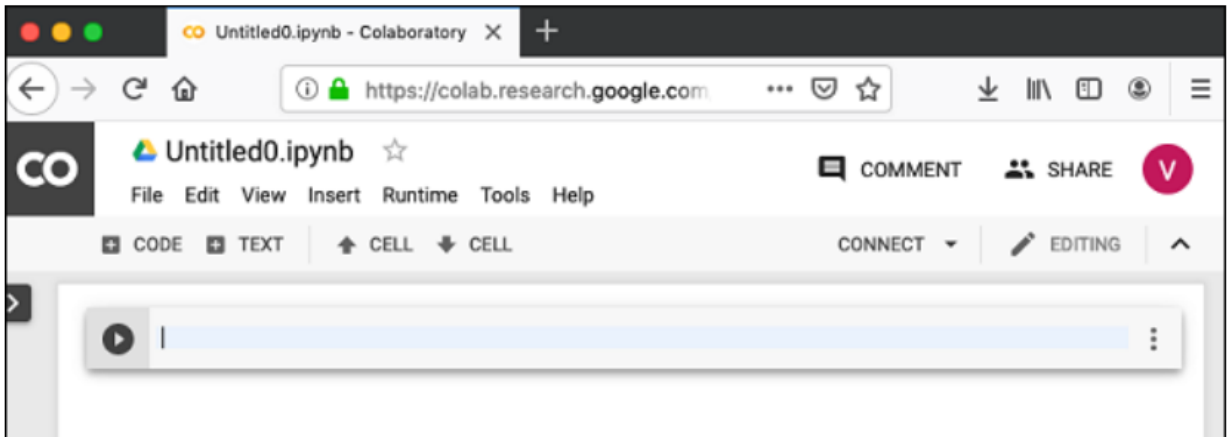
Introduction of Google Colab

- Colab is a free notebook environment that runs entirely in the cloud. It lets you and your team members edit documents, the way you work with Google Docs.
- Colab supports many popular machine learning libraries which can be easily loaded in your notebook.
- Another attractive feature that Google offers to the developers is the use of GPU. Colab supports GPU and it is totally free. The reasons for making it free for public could be to make its software a standard in the academics for teaching machine learning and data science.
- Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook

- **Step 1** – Open the following URL in your browser – https://colab.research.google.com Your browser would display the following screen (assuming that you are logged into your Google Drive) –

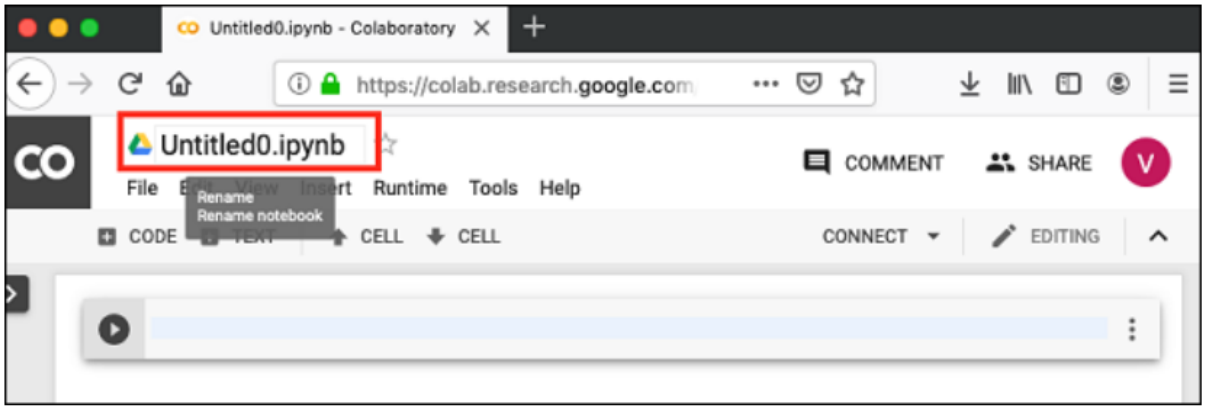


- **Step 2** – Click on the **NEW PYTHON 3 NOTEBOOK** link at the bottom of the screen. A new notebook would open up as shown in the screen below.

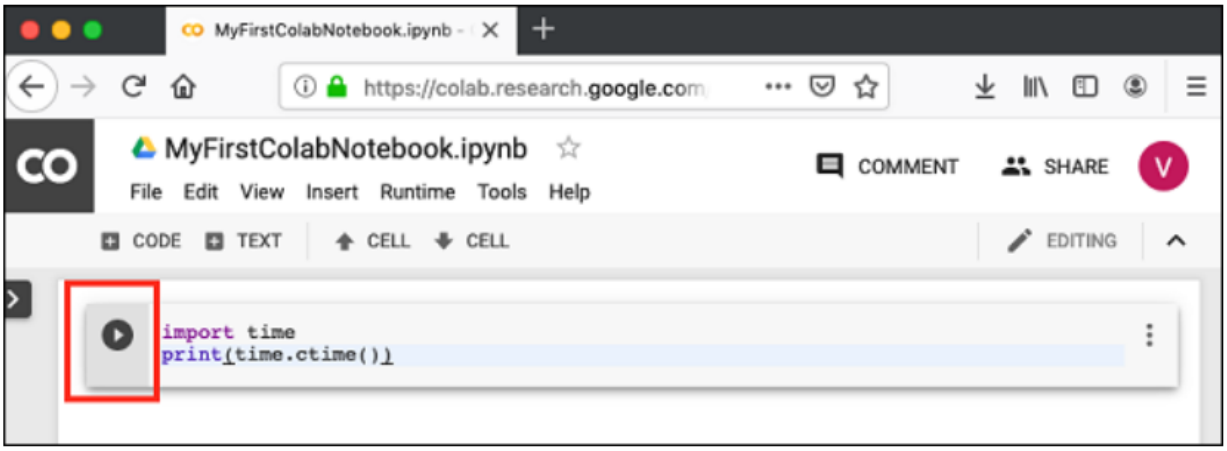


- **Setting Notebook Name**

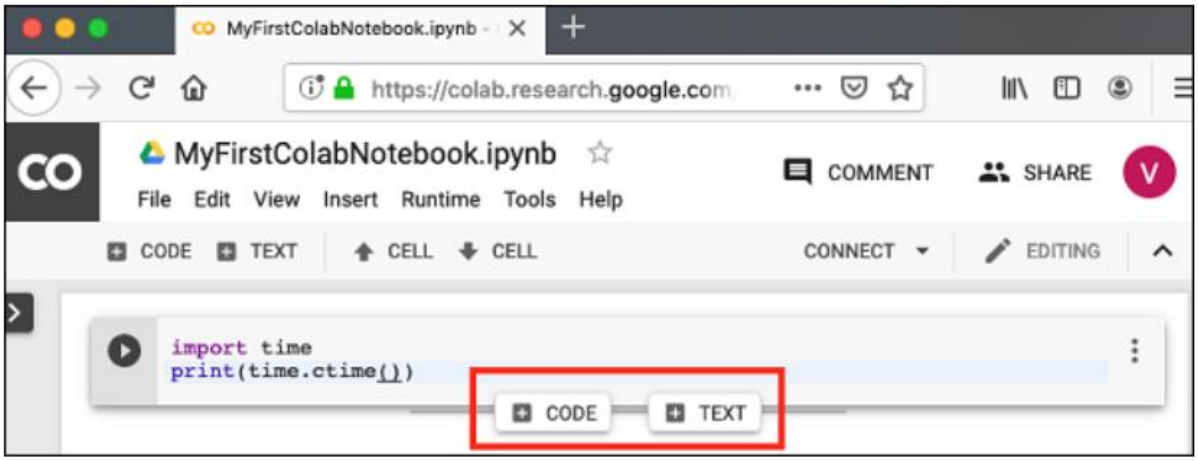
By default, the notebook uses the naming convention UntitledXX.ipynb. To rename the notebook, click on this name and type in the desired name in the edit box as shown here −



- Entering Code

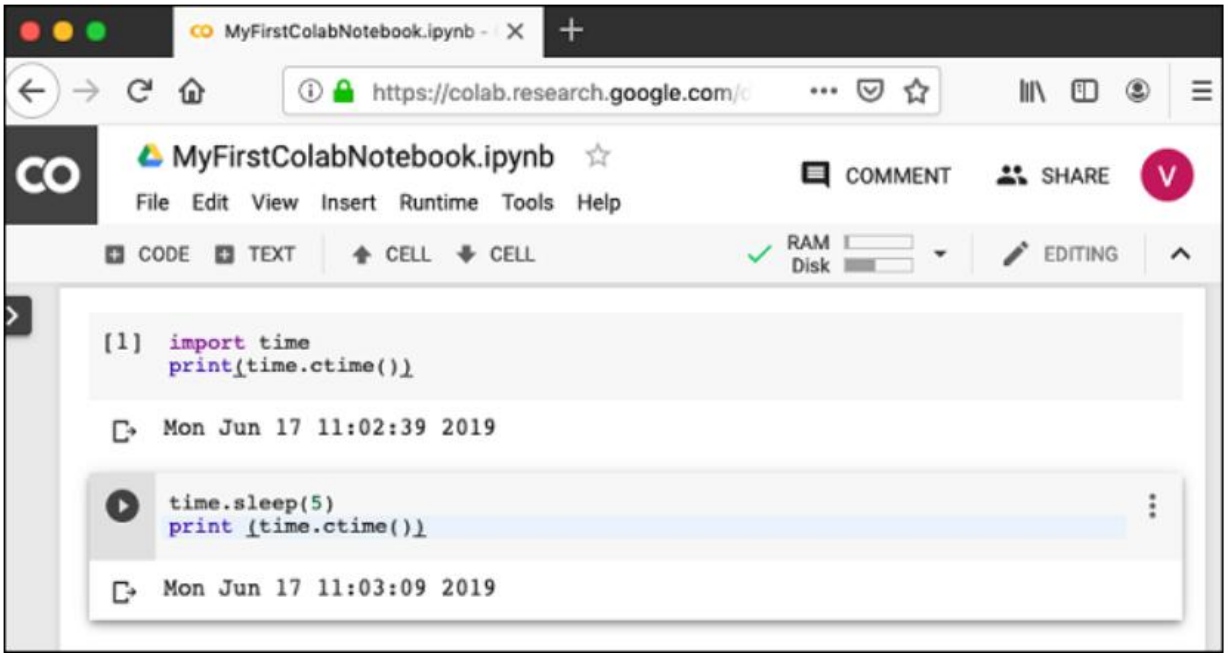


- Adding Code Cells
- To add more code to your notebook, select the following **menu** options −

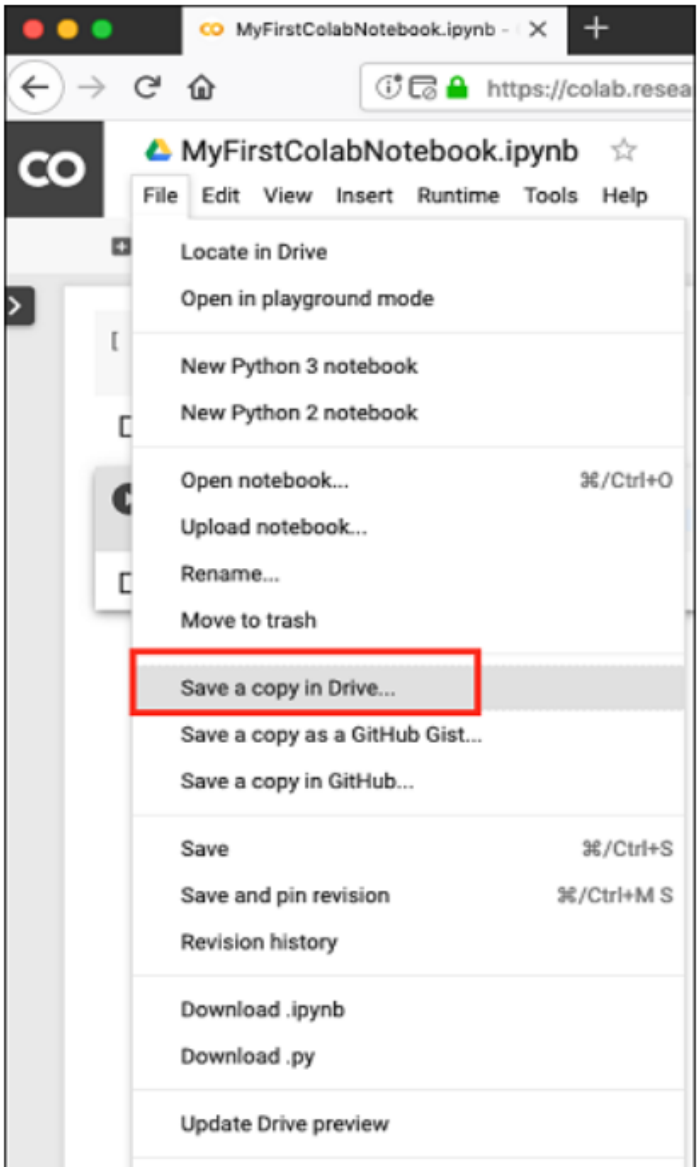


- **Run All**

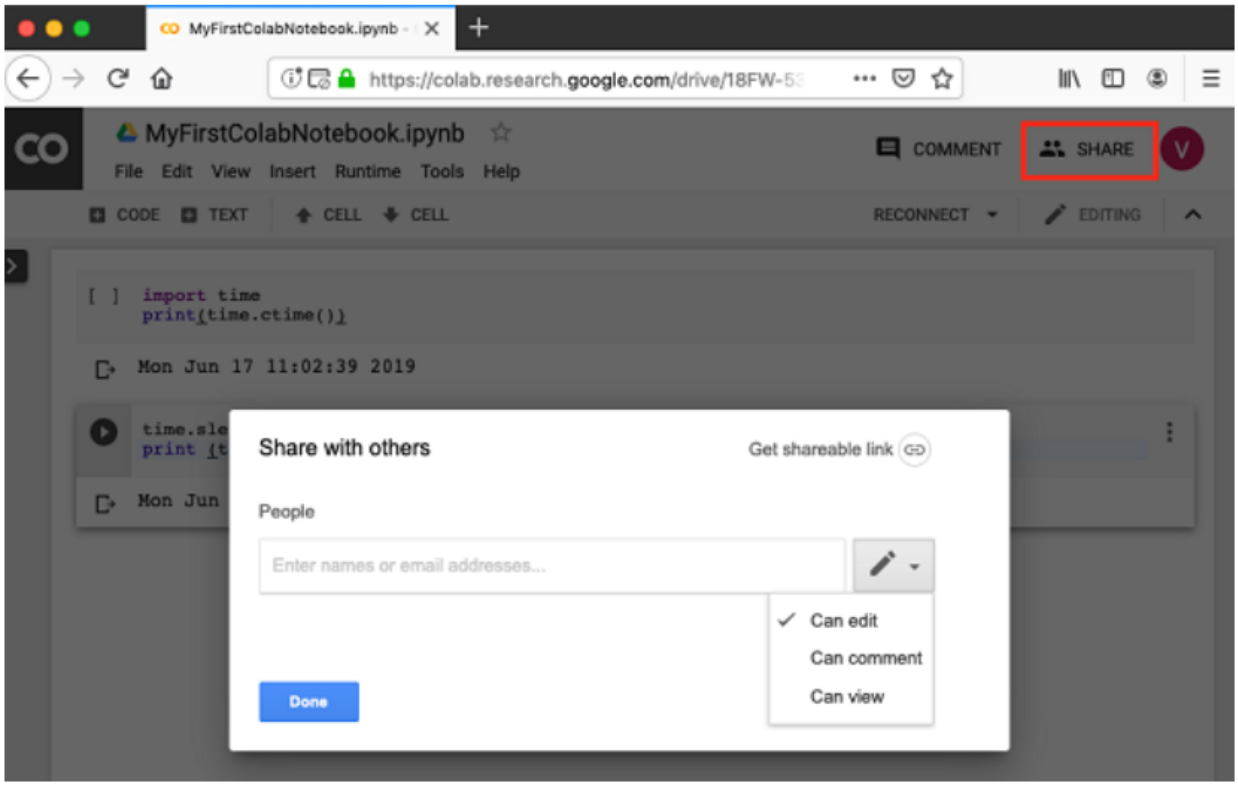
  To run the entire code in your notebook without an interruption, execute the following menu options −



Saving to Google Drive



Google Colab - Sharing Notebook



LAB Assignment No. 1

Lab Assignment – Dataset Creation & Analysis Objective

To learn how to create and upload a dataset in Python, perform basic statistical analysis, and visualize data using graphs.

Tasks Q1: Create a Dataset Manually

- Create a dataset of at least **10 students** with the following columns:
    - o Student_ID, o Name, o Age, o

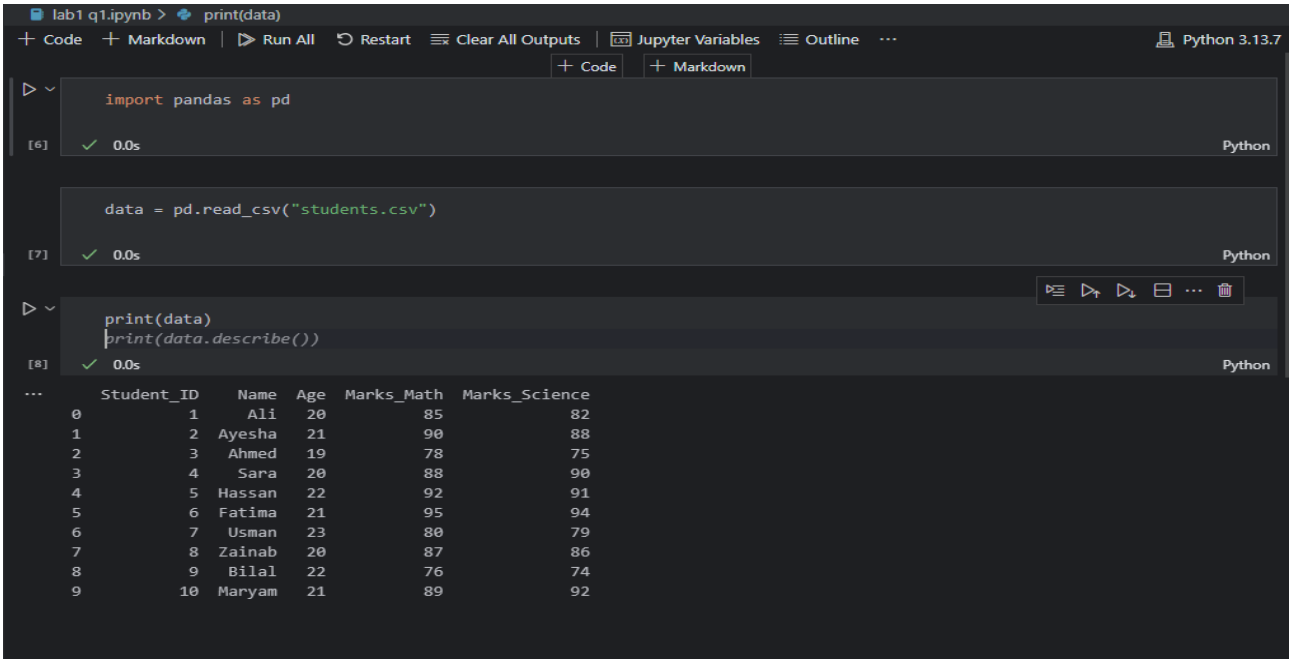
Marks_Math, o

Marks_Science.

- Store the dataset in a **CSV file** named students.csv.

Q2: Upload Dataset in Python

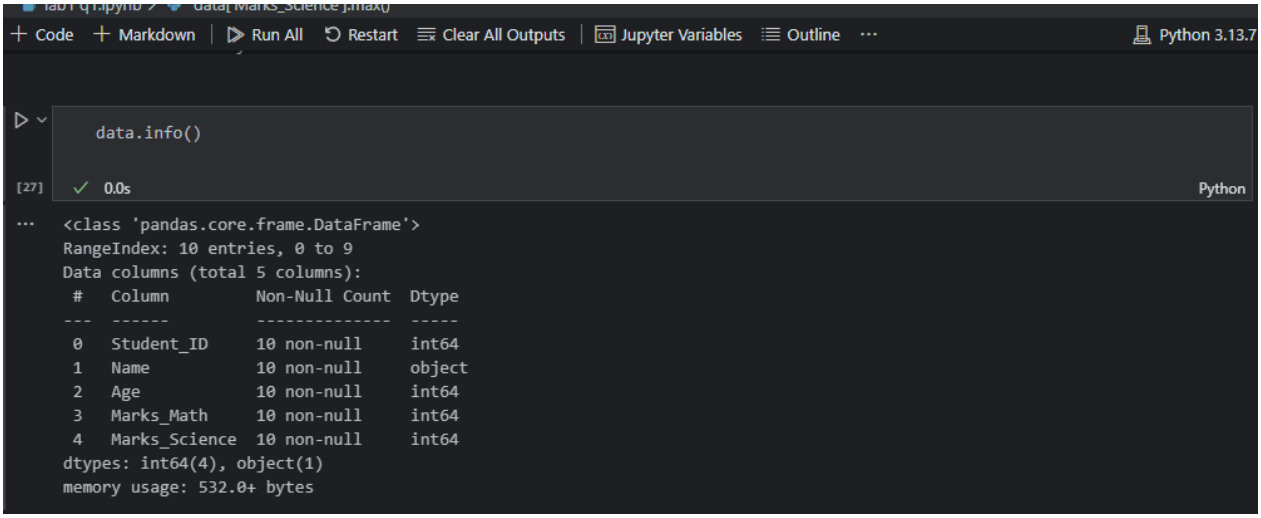
- Use **Pandas** to load the dataset.



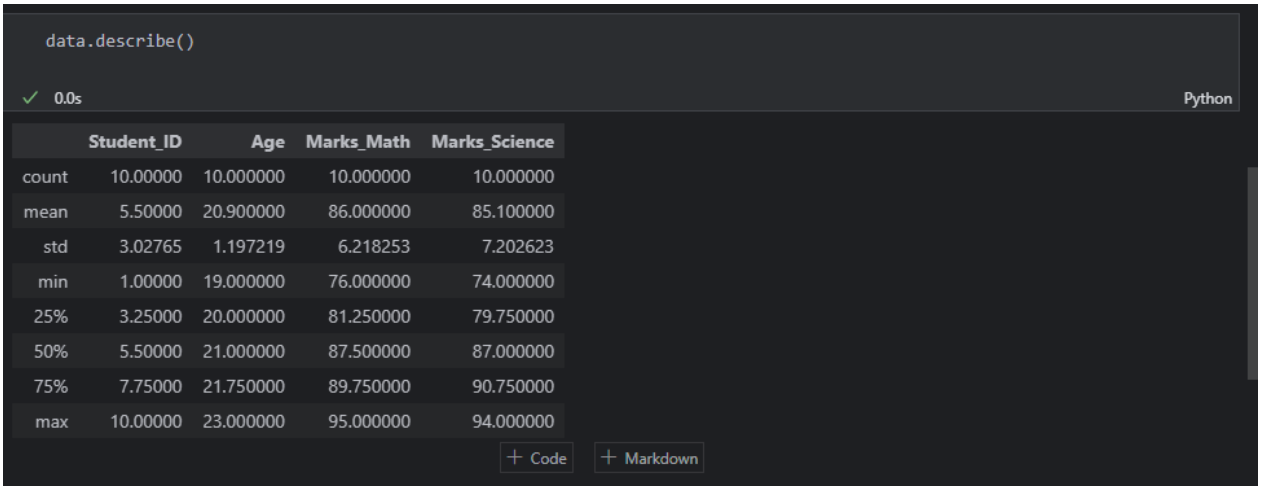
Q3: Observe Dataset Information

Run the following commands and explain the output:

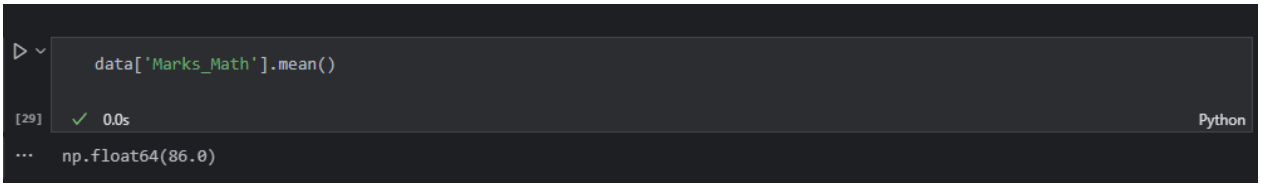
1. data.info() → Dataset structure



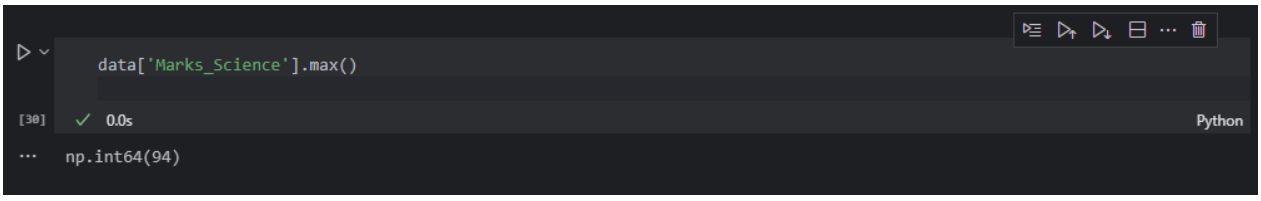
2. data.describe() → Summary statistics (mean, std, min, max, etc.)



3. data['Marks_Math'].mean() → Mean of Math marks

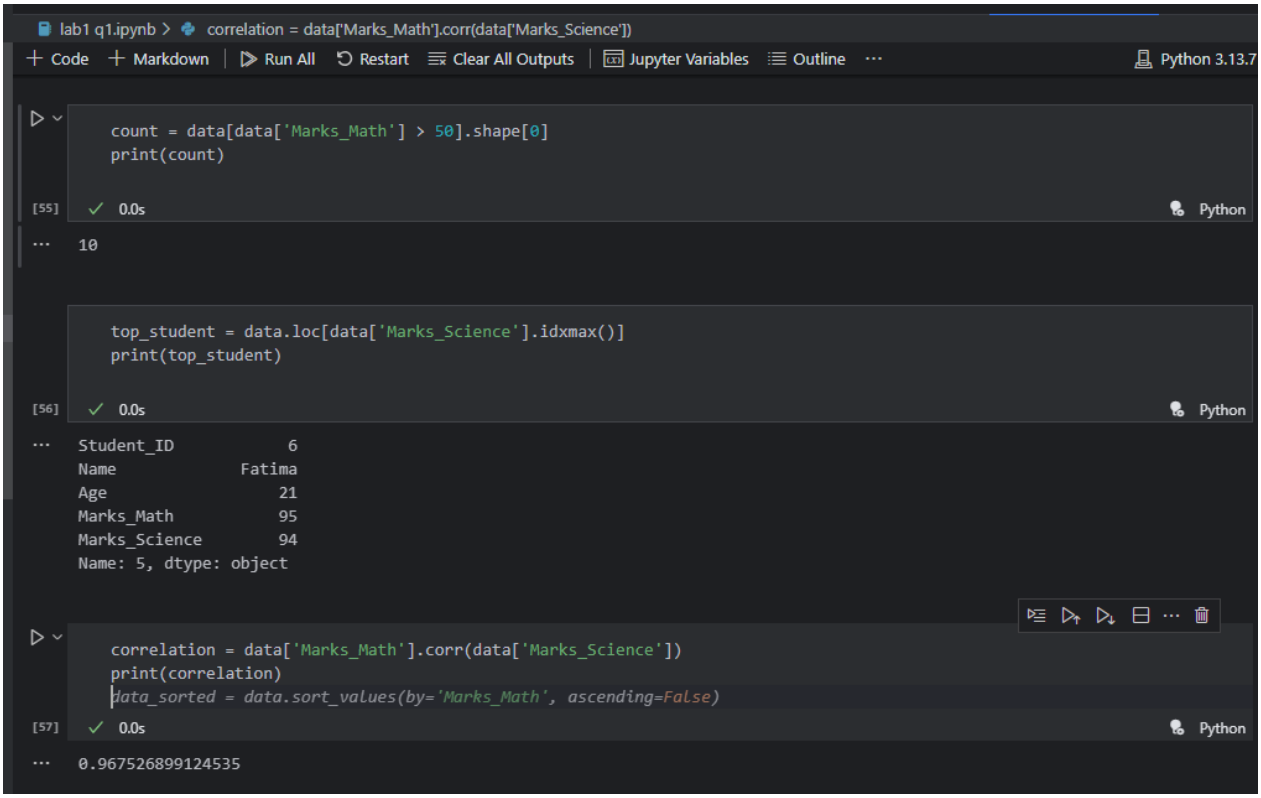


4. data['Marks_Science'].max() → Maximum Science marks



Q4: Perform Some Data Analysis

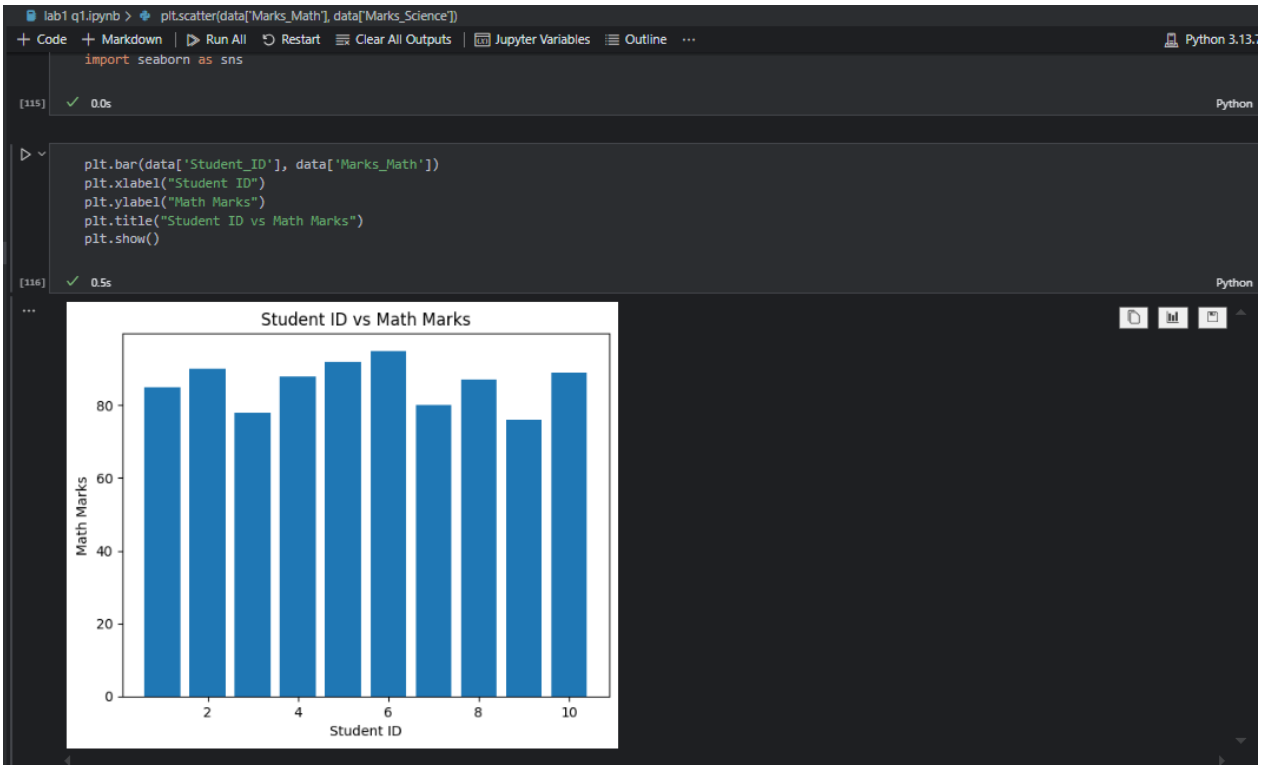
- Find how many students have Marks_Math > 50.
- Find the student with the **highest Science marks**.
- Calculate the **correlation** between Marks_Math and Marks_Science.



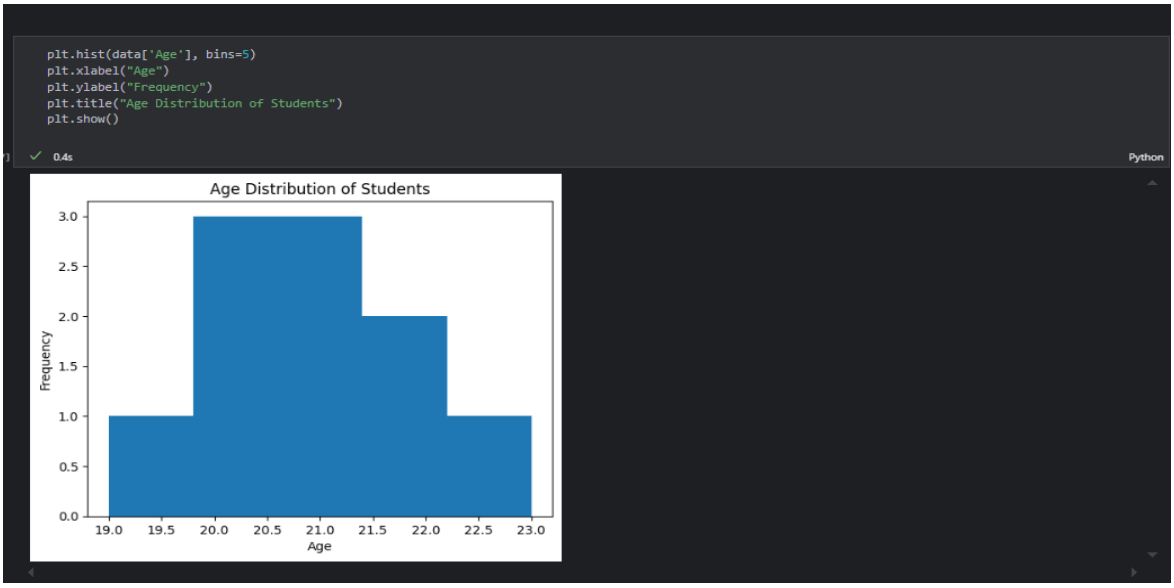
Q5: Data Visualization

Use **Matplotlib/Seaborn** to create graphs:

1. A bar chart of Student_ID vs Marks_Math.



A histogram of Age.



2. A scatter plot of Marks_Math vs Marks_Science.

```
plt.scatter(data['Marks_Math'], data['Marks_Science'])
plt.xlabel("Math Marks")
plt.ylabel("Science Marks")
plt.title("Math vs Science Marks")
plt.show()
```

0.3s Python