CS3362 DATA SCIENCE LABORATORY

L T P C 0 0 4 2

**COURSE OBJECTIVES:**

* To understand the python libraries for data science
* To understand the basic Statistical and Probability measures for data science.
* To learn descriptive analytics on the benchmark data sets.
* To apply correlation and regression analytics on standard data sets.
* To present and interpret data using visualization packages in Python.

**LIST OF EXPERIMENTS:**

1. Download, install and explore the features of NumPy, SciPy, Jupyter, Statsmodels and Pandas packages.

2. Working with Numpy arrays

3. Working with Pandas data frames

4. Reading data from text files, Excel and the web and exploring various commands for doing descriptive analytics on the Iris data set.

5. Use the diabetes data set from UCI and Pima Indians Diabetes data set for performing the following:

a. Univariate analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation,Skewness and Kurtosis.

b. Bivariate analysis: Linear and logistic regression modeling

c. Multiple Regression analysis

d. Also compare the results of the above analysis for the two data sets.

6. Apply and explore various plotting functions on UCI data sets.

a. Normal curves b. Density and contour plots

c. Correlation and scatter plots d. Histograms

e. Three dimensional plotting

7. Visualizing Geographic Data with Basemap

**LIST OF TOOLS :**

Python, Numpy, Scipy, Matplotlib, Pandas, statmodels, seaborn, plotly, bokeh

**Note:**

Example data sets like: UCI, Iris, Pima Indians Diabetes etc.

|  |  |  |
| --- | --- | --- |
| **Exp.No:1** | **Download, install and explore the features of NumPy, SciPy, Jupyter, Statsmodels and Pandas packages.** | **Date:** |

**Aim:**

To Download, install and explore the features of NumPy, SciPy, Jupyter, Statsmodels and Pandas packages.

**Tools Required:**

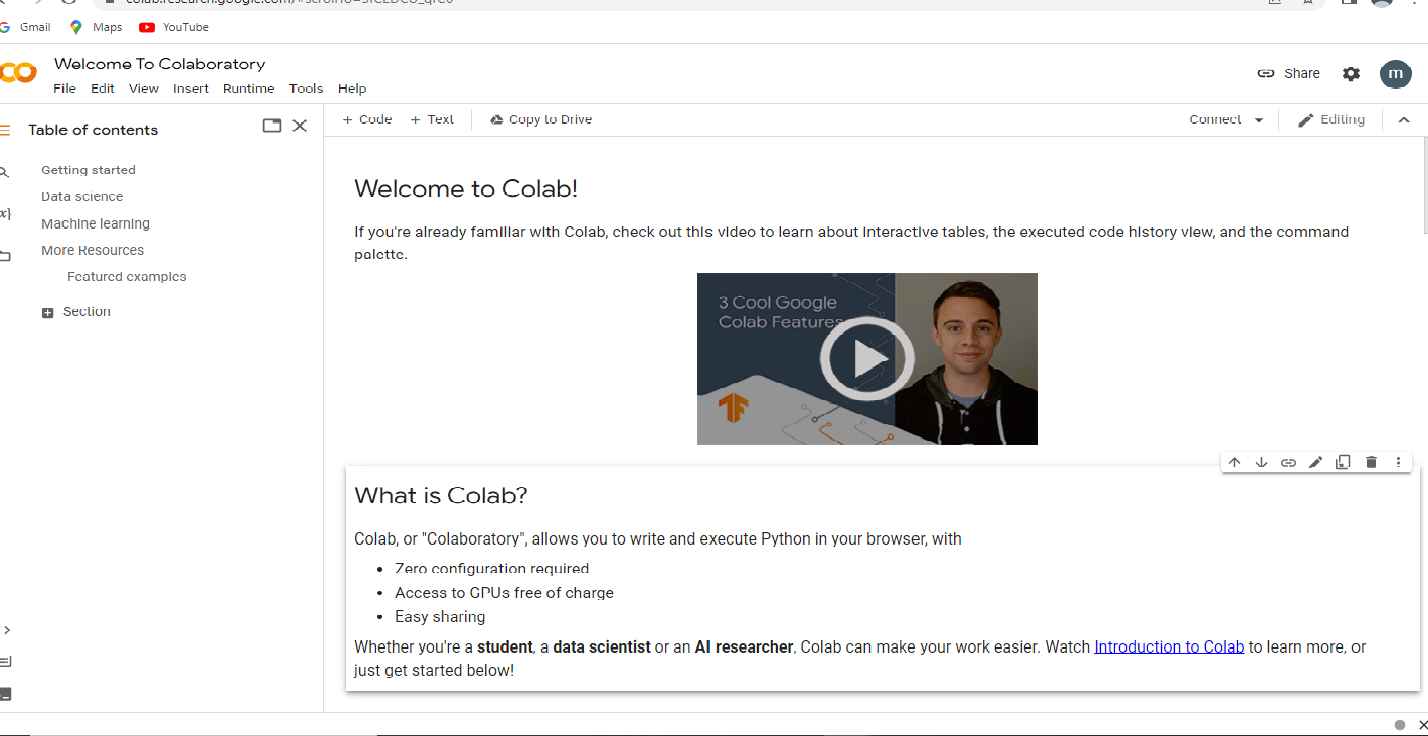
Google Collab (<https://colab.research.google.com/> ) , Browser (Any).

**Pakages:**

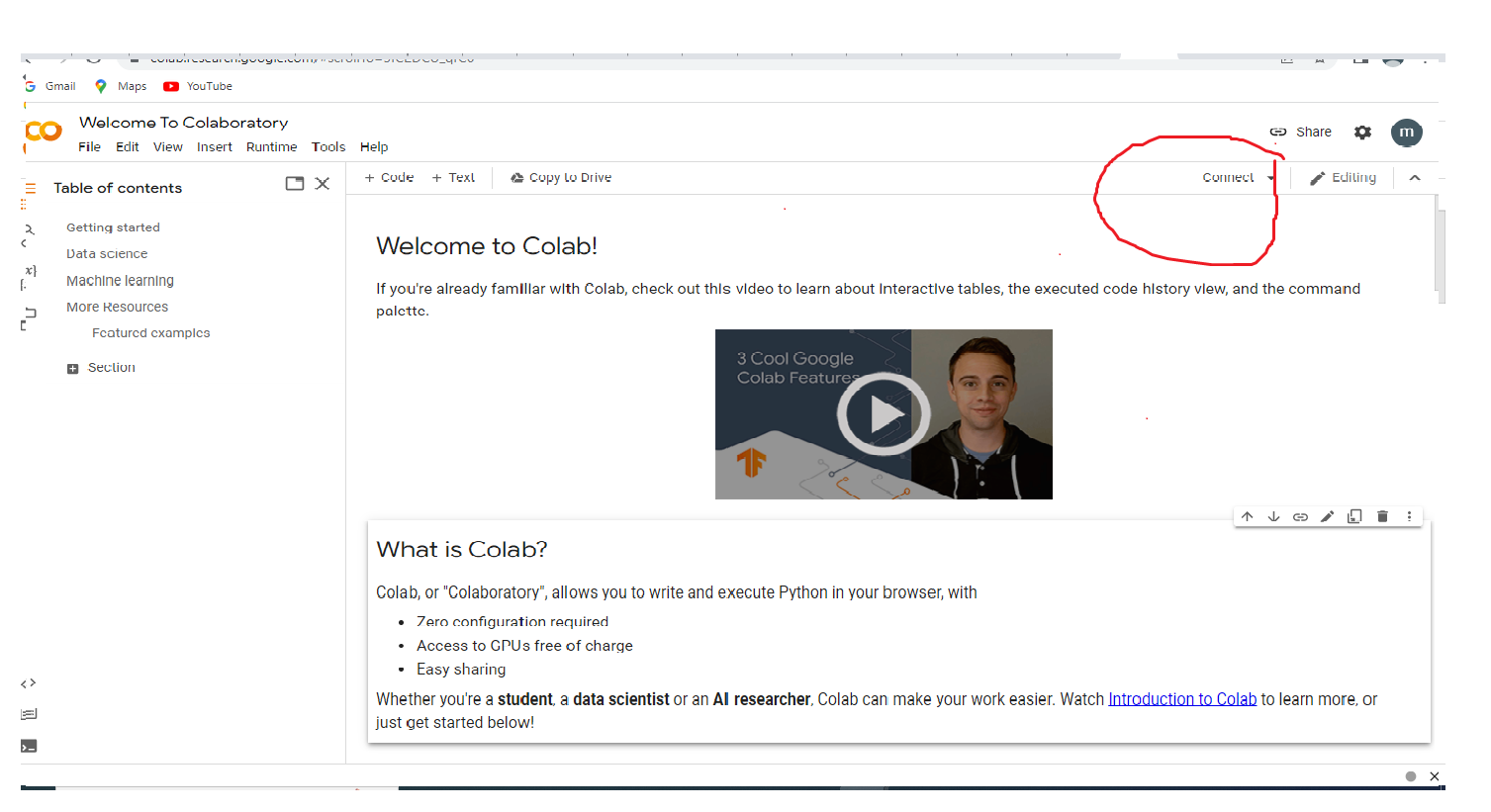
* NumPy
* SciPy
* Jupyter
* Statsmodels
* Pandas

**Steps:**

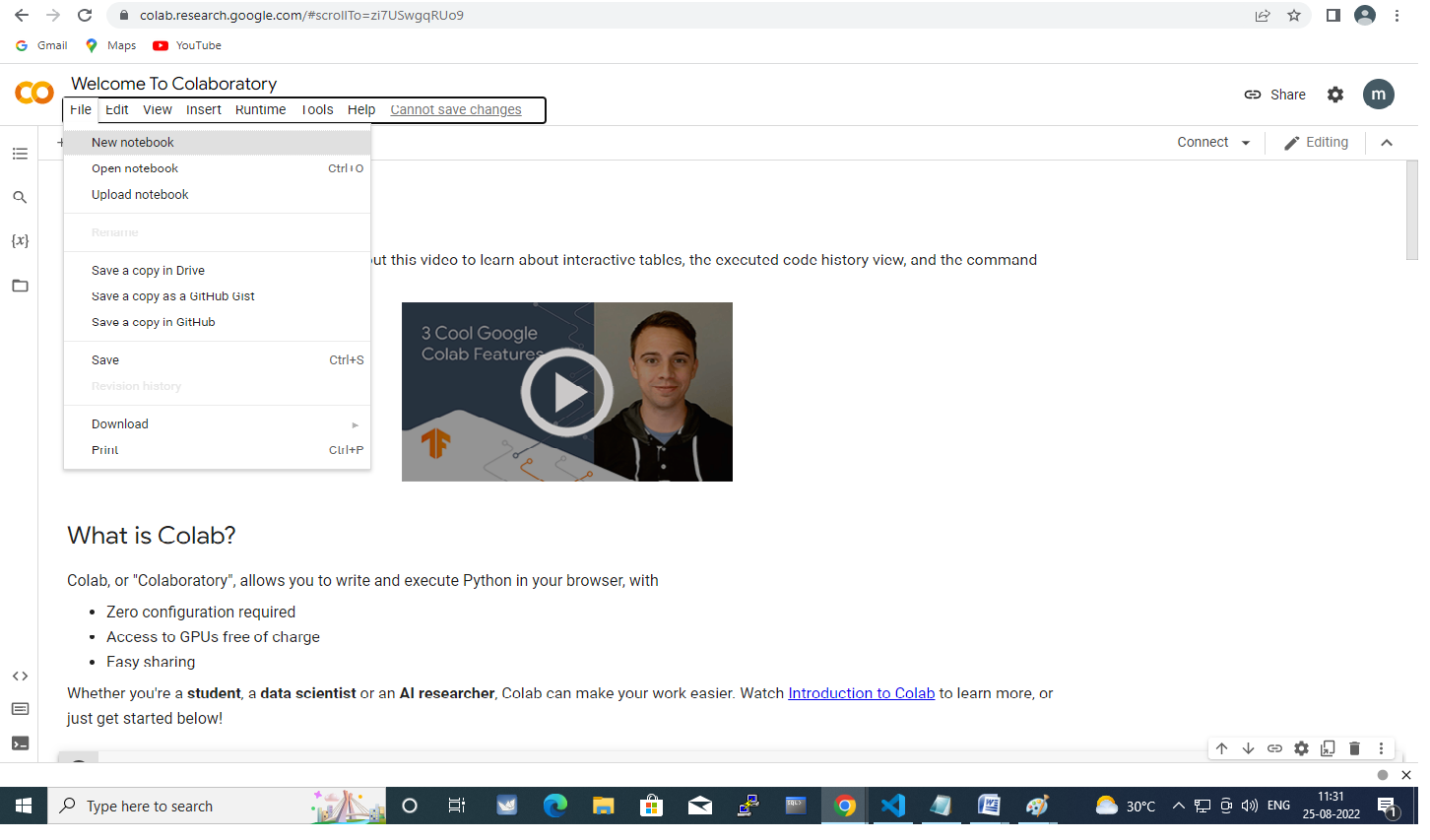
1.Go to Link <https://colab.research.google.com/> In browser.



2.Click on The Connect Button .



3. Go to Files in menu Bar and click File ->new Notebook.



4.Type the codes given below in the code section.

**Install:**

Code:

!pip install numpy

Code:

!pip install SciPy

Code:

!pip install Jupyter

Code:

!pip install Statsmodels

Code:

!pip install Pandas

5. Now Explore Their features in internet.

Numpy - <https://numpy.org/doc/stable/user/whatisnumpy.html>

Scipy - <https://scipy.org/>

Jupyter - <https://jupyter.org/hub>

Stastmodel - <https://www.statsmodels.org/stable/index.html>

Pandas - <https://pandas.pydata.org/>

|  |  |  |
| --- | --- | --- |
| **Exp.No:2** | **Working with Numpy arrays** | **Date:** |

**Aim:**

**To Work with numpy Arrays.**

CODE:

EXAMPLE 1

Using 1-dimentional array

Description:An array that has 0-D arrays as its elements is called uni-dimensional or 1-D array.

These are the most common and basic arrays.

Program:

import numpy as np

arr = np.array([1, 2, 3, 4, 5]) #creating np array

print(arr) #prints elements in array

print(type(arr)) #prints type of array

OUTPUT:

[1 2 3 4 5]

<class 'numpy.ndarray'>

EXAMPLE 2

Using 2-dimentional array

Description:An array that has 1-D arrays as its elements is called a 2-D array.

These are often used to represent matrix or 2nd order tensors.

Program:

import numpy as np

arr = np.array([[1, 2, 3], [4, 5, 6]]) #creating np array

print(arr) #prints elements in array

print(type(arr)) #prints type of array

OUTPUT:

[

[1 2 3] [4 5 6]

]

EXAMPLE 3

Using 3-dimentional array

Description:An array that has 2-D arrays (matrices) as its elements is called 3-D array.

These are often used to represent a 3rd order tensor.

Program:

import numpy as np

arr = np.array([[[1, 2, 3], [4, 5, 6]], [[1, 2, 3], [4, 5, 6]]])

print(arr)

OUTPUT:

[

[[1 2 3] [4 5 6]]

[[1 2 3] [4 5 6]]

]