ESSENTIALS OF DATA WAREHOUSING AND DATA MINING

LAB ASSIGNMENT - 12

NAME - Kaparotu Venkata Surya Tharani
USN - 22BTRAD018
BRANCH - AI & DE

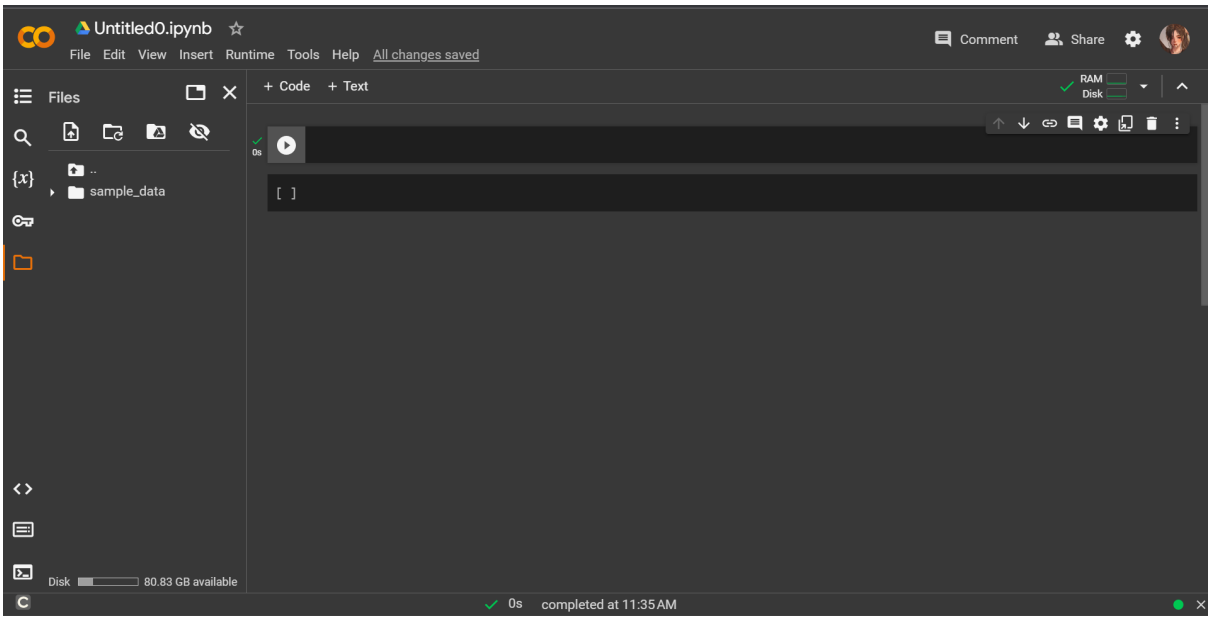
1. Given the irish data. Except for the class attribute, take all attribute value in to a matrix A. Now calculate the corelation coefficiect (without using the direct correlation calculation function from any python library) between each sample (row vector - one irish flower) in matrix A with all the sample in matrix A. And for plotting and validating your calculations correctness, you can use the following code.

Code :-

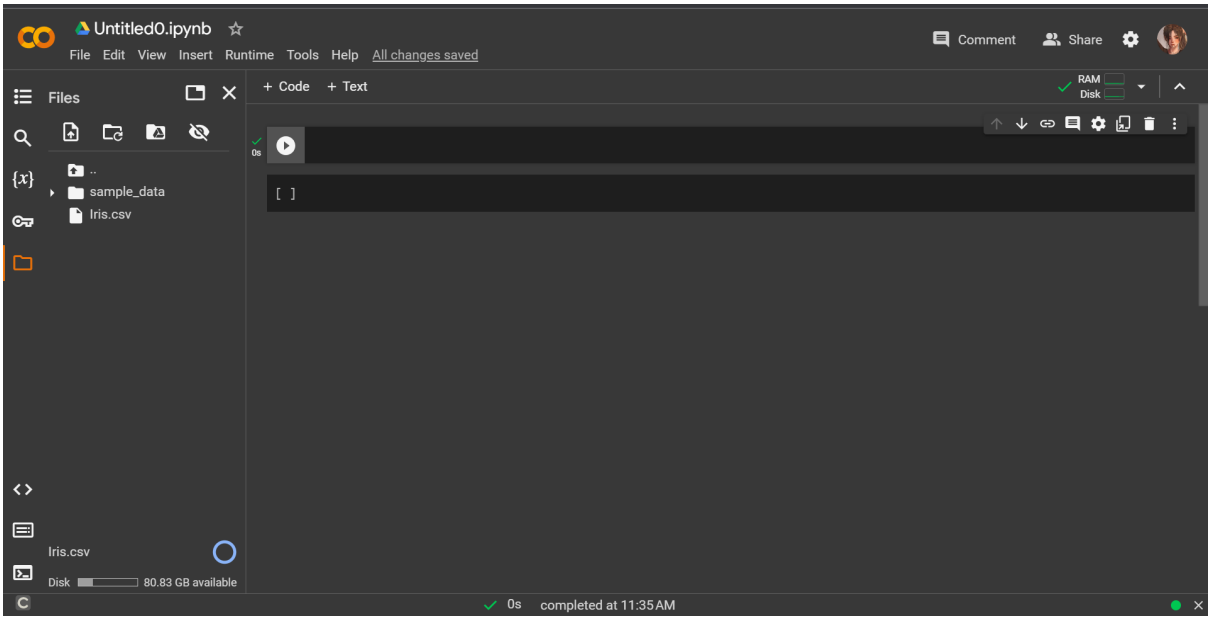
```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

f = pd.read_csv('Iris.csv')

print(f)

print(f.loc[:,:'PetalWidthCm'])

print(f.loc[:,:'PetalWidthCm'].T.corr().loc[:10,:10])

dcorr=f.loc[:,:'PetalWidthCm'].T.corr()

sns.heatmap(dcorr.corr().loc[:10,:10], annot = True, cmap = "coolwarm")
```

SCREENSHOTS :-

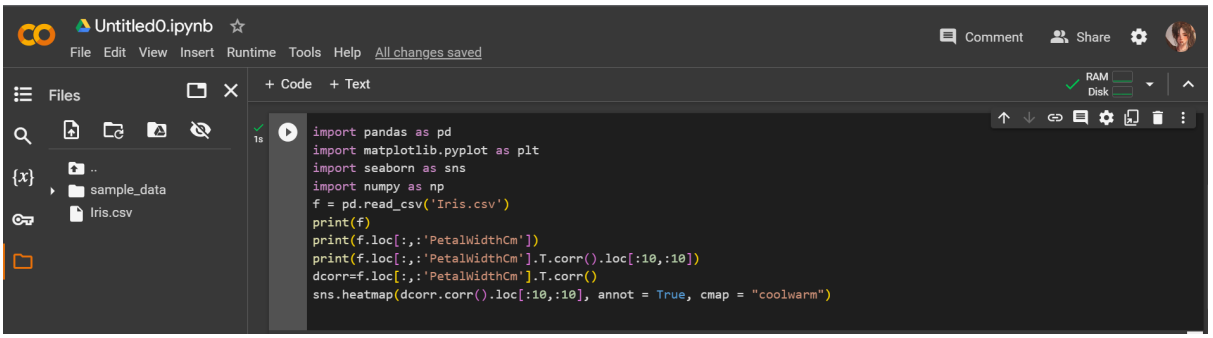
1. Click the folder.



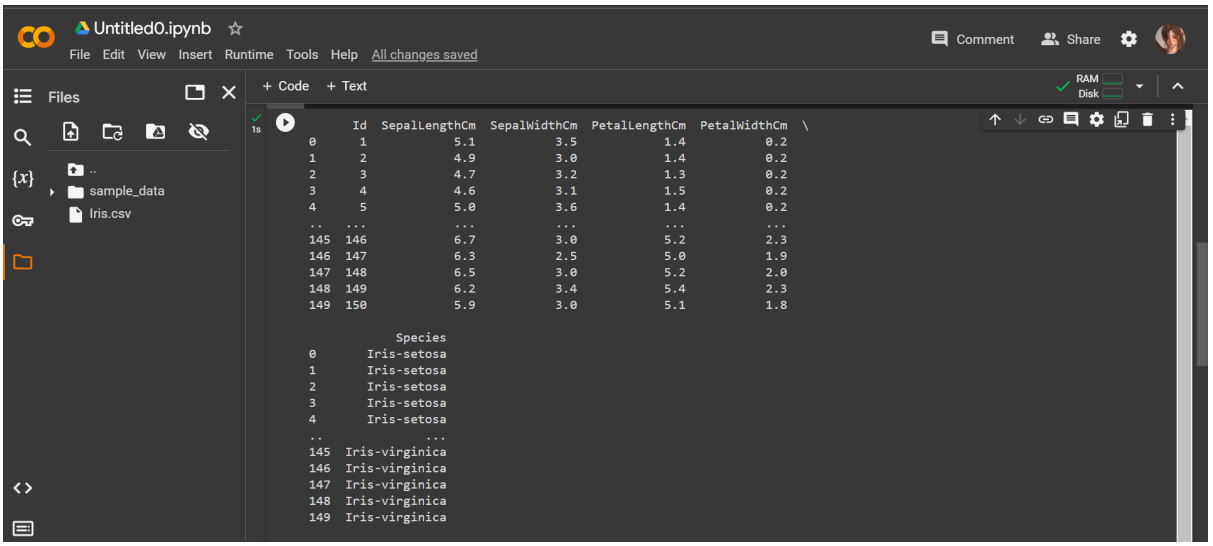
2. Click on upload button and select the iris dataset.

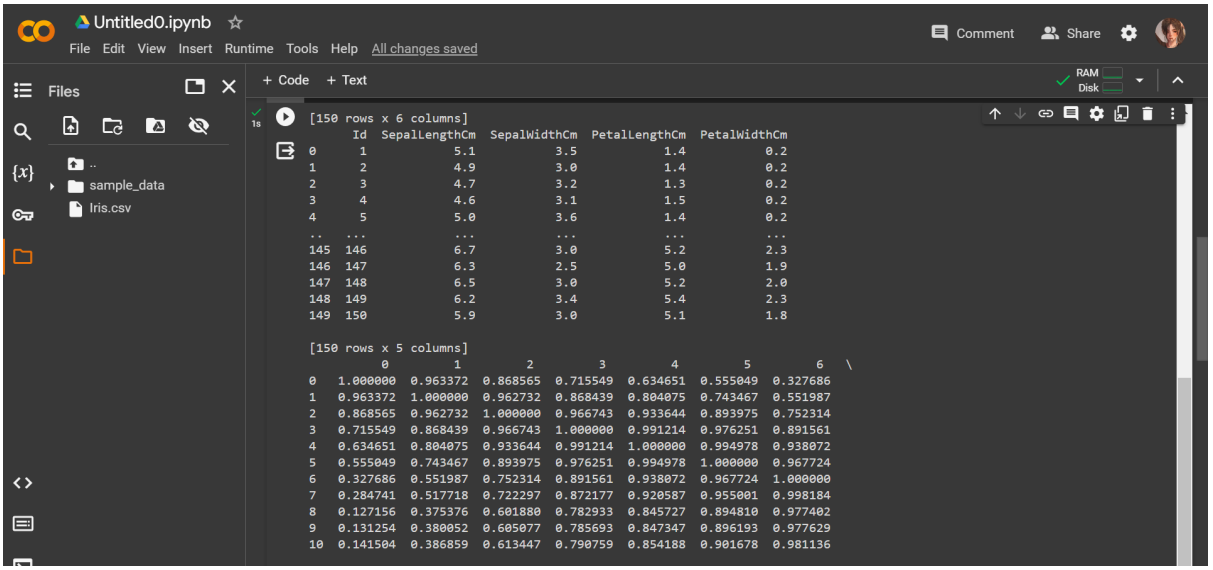


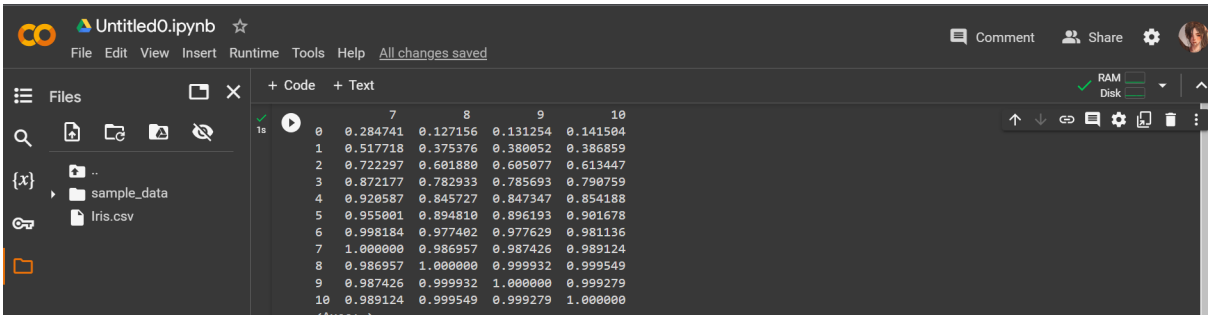
3. Write the python code and execute.

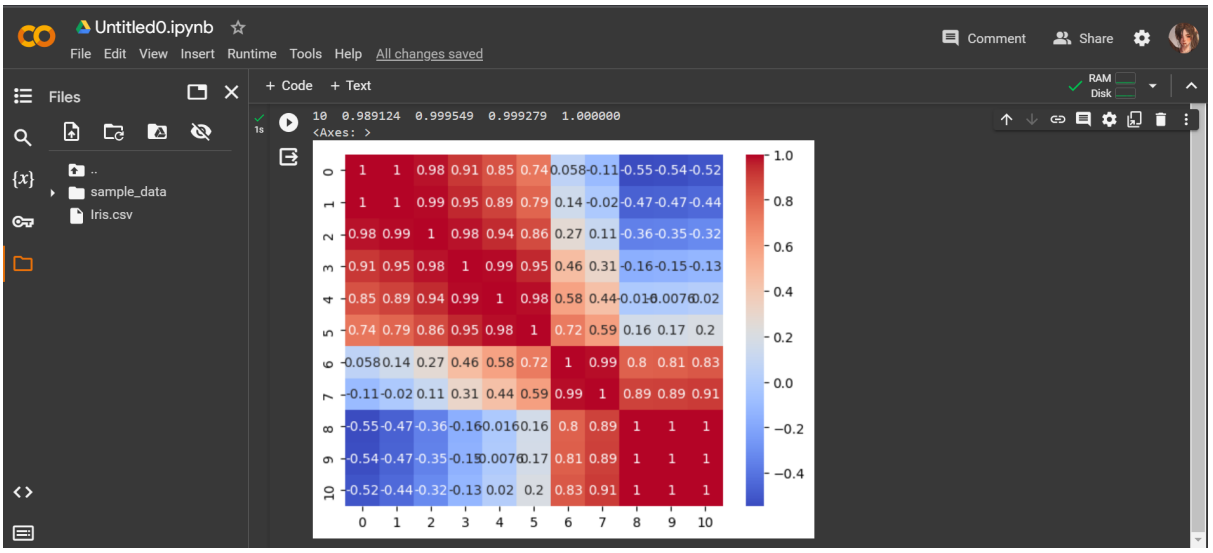


Output :






Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
Iris.csv
7 8 9 10
0 0.284741 0.127156 0.131254 0.141504
1 0.517718 0.375376 0.380052 0.386859
2 0.722297 0.601880 0.605077 0.613447
3 0.872177 0.782933 0.785693 0.790759
4 0.920587 0.845727 0.847347 0.854188
5 0.955001 0.894810 0.896193 0.901678
6 0.998184 0.977402 0.977629 0.981136
7 1.000000 0.986957 0.987426 0.989124
8 0.986957 1.000000 0.999932 0.999549
9 0.987426 0.999932 1.000000 0.999279
10 0.989124 0.999549 0.999279 1.000000


Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
Iris.csv
10 0.989124 0.999549 0.999279 1.000000
<Axes: >
0 1 2 3 4 5 6 7 8 9 10
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4