**Practical-01**

**AIM: Prepare synthetic data set for student data, consisting of Enrollment number, name, gender, semester wise, subject wise marks, difficulty level of the subject, SPI (Semester Index), address with geographical location .**

**a.**

**(i) Write a program to find correlation between gender and Semester marks.**

**Program:**

import pandas as pd

data=pd.read\_csv('/content/sample\_data/DAV\_1\_Data1.csv')

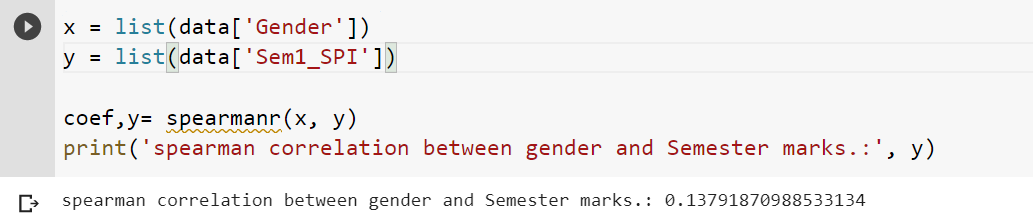
x = list(data['Gender'])

y = list(data['Sem1\_SPI'])

coef,y= spearmanr(x, y)

print('spearman correlation between gender and Semester marks: ', y)

**output:**



**(ii) Write a program to find correlation between geographical location and semester marks. Analyze which two are highly correlated.**

**Program:**

import pandas as pd

data=pd.read\_csv('/content/sample\_data/DAV\_1\_Data1.csv')

x = list(data['Address'])

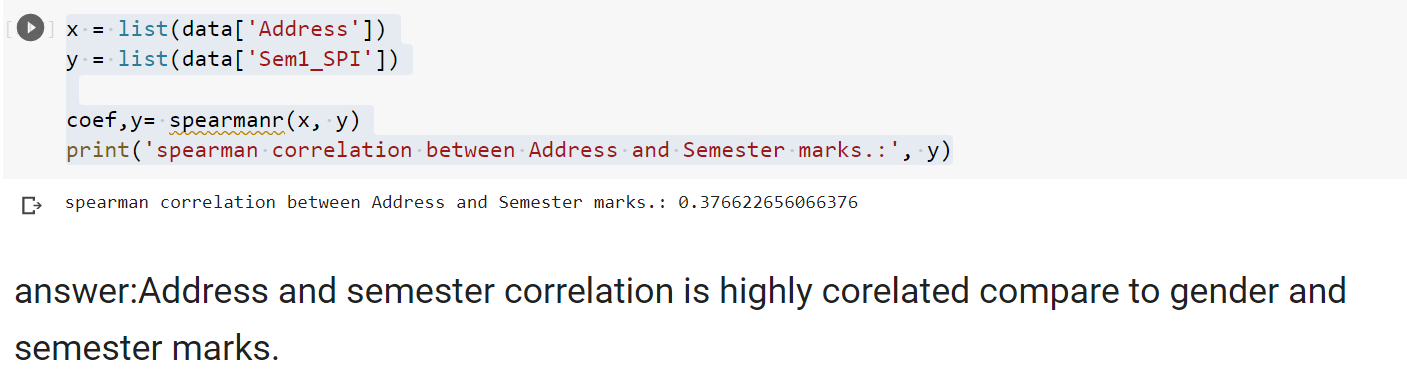
y = list(data['Sem1\_SPI'])

coef, y= spearmanr(x, y)

print('spearman correlation between Address and Semester

marks:', y)

**output:**



**b. Write a program to calculate correlation between difficulty level and subject marks. The higher the difficulty level the marks should be less. The two should be negatively correlated. Analyze the correlation.**

**Program:**

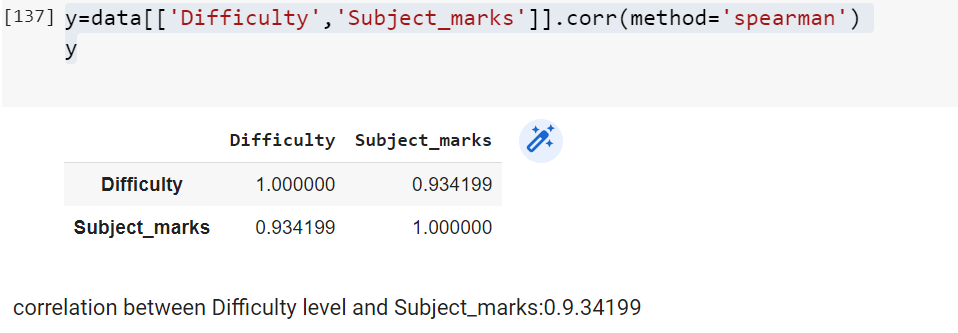
import pandas as pd

df=pd.read\_csv('/content/sample\_data/DAV2.csv')

y=data[['Difficulty','Subject\_marks']].corr(method='spearman')

y

**output:**

****