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:

```
x = list(data['Gender'])
y = list(data['Sem1_SPI'])

coef,y= spearmanr(x, y)
print('spearman correlation between gender and Semester marks.:', y)
```

```
spearman correlation between gender and Semester marks.: 0.13791870988533134
```

(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:

```
x = list(data['Address'])
y = list(data['Sem1_SPI'])

coef,y= spearmanr(x, y)
print('spearman correlation between Address and Semester marks.:', y)
```

spearman correlation between Address and Semester marks.: 0.376622656066376

answer:Address and semester correlation is highly corelated compare to gender and semester marks.

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:

```
[137] y=data[['Difficulty','Subject_marks']].corr(method='spearman')
      y
```

	Difficulty	Subject_marks
Difficulty	1.000000	0.934199
Subject_marks	0.934199	1.000000

correlation between Difficulty level and Subject_marks:0.9.34199