

Experiment - 19

Aim: Import any CSV file to Pandas DataFrame and perform various statistical operations.

Program:

#Creating a Sample DataFrame

```
import pandas as pd
df = {
    "Name":["Ram", "Sam", "Scott", "Ann", "John"],
    "Mathematics" : [80, 90, 85, 70, 95],
    "Science" : [85, 95, 80, 90, 75],
    "English" : [90, 85, 80, 70, 95]
}
```

```
index_labels=['r1', 'r2', 'r3', 'r4', 'r5']
df = pd.DataFrame(df , index=index_labels)
```

```
print(df)
```

#Perform statistical operations

```
1) sum()
# Add new column Sum
df['Sum'] = df['Mathematics'] + df['Science'] + df['English']
print(df)
```

```
# Select 1 to 3 columns to sum
df['Sum']=df.iloc[:,1:3].sum(axis=1)
print(df)
```

```
# Sum columns Fee and Discount for row from r2 to r3
df['Sum'] = df.loc['r2':'r4',['Mathematics','Science']].sum(axis = 1)
print(df)
```

```
print(df['Sum'].mean())
```

```
print(df['Sum'].median())
```

```
print(df['Sum'].mode())
```

```
print(df['Sum'].count())
```

```
print(df['Sum'].std())
```

```
print(df['Sum'].max())
```

```
print(df['Sum'].min())
```

#Descriptive Statistics

```
print(df.describe())
```

```
# summary statistics of character column
print(df.describe(include=['object']))

# summary statistics of all the column
print(df.describe(include='all'))
```

Output:

	Name	Mathematics	Science	English
r1	Ram	80	85	90
r2	Sam	90	95	85
r3	Scott	85	80	80
r4	Ann	70	90	70
r5	John	95	75	95

1)sum()

	Name	Mathematics	Science	English	Sum
r1	Ram	80	85	90	255
r2	Sam	90	95	85	270
r3	Scott	85	80	80	245
r4	Ann	70	90	70	230
r5	John	95	75	95	265

	Name	Mathematics	Science	English	Sum
r1	Ram	80	85	90	165
r2	Sam	90	95	85	185
r3	Scott	85	80	80	165
r4	Ann	70	90	70	160
r5	John	95	75	95	170

	Name	Mathematics	Science	English	Sum
r1	Ram	80	85	90	NaN
r2	Sam	90	95	85	185.0
r3	Scott	85	80	80	165.0
r4	Ann	70	90	70	160.0
r5	John	95	75	95	NaN

2)mean ()

170.0

3)median ()

165.0

4)mode ()

```
0    160.0
1    165.0
2    185.0
dtype: float64
```

5)count ()

3

6) std()

13.228756555322953

7) max ()

185.0

8) min ()

160.0

9) describe()

	Mathematics	Science	English	Sum
count	5.0	5.0	5.0	3.0
mean	84.0	85.0	84.0	170.0
std	9.6	7.9	9.6	13.2
min	70.0	75.0	70.0	160.0
25%	80.0	80.0	80.0	162.500000
50%	85.0	85.0	85.0	165.0
75%	90.0	90.0	90.0	175.0
max	95.0	95.0	95.0	185.0

	Name
count	5
unique	5
top	Ram
freq	1