

PRACTICAL NO. 01

PROGRAM :

a.

```
import pandas as pd
mtcars = pd.read_csv('mtcars.csv')
# Display summary statistics
print("\n1] Summary Statistics for mtcars dataset:\n")
print(mtcars.describe())
# Display structure information
print("\n2] Structure Information for mtcars dataset:\n")
print(mtcars.info())
# Use the quantile() method to get the quartile values for a specific
column
print("\n3] Quartile Information for mtcars dataset:\n")
print(mtcars['mpg'].quantile([0.25, 0.5, 0.75]))
```

b.

```
import pandas as pd
iris = pd.read_csv('iris.csv')
# Use subset() function to select only rows where Sepal.Width > 3
setosa_subset = iris[iris['sepal_width'] >= 3.9]
print("\n1] Subset of Iris dataset with only rows where Sepal.Width>3.9 :\n")
print(setosa_subset)
# Use aggregate() function to calculate mean sepal length for each species
aggregate_result = iris.groupby('species').mean()
print("\n2] Aggregate result - Mean sepal length and sepal width for each
species:\n")
print(aggregate_result)
```

OUTPUT :

a.

1] Summary Statistics for mtcars dataset:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.0000
mean	20.090625	6.187500	230.721875	146.687500	3.596563	3.217250	17.848750	0.437500	0.406250	3.687500	2.8125
std	6.026948	1.785922	123.938694	68.562868	0.534679	0.978457	1.786943	0.504016	0.498991	0.737804	1.6152
min	10.400000	4.000000	71.100000	52.000000	2.760000	1.513000	14.500000	0.000000	0.000000	3.000000	1.0000
25%	15.425000	4.000000	120.825000	96.500000	3.080000	2.581250	16.892500	0.000000	0.000000	3.000000	2.0000
50%	19.200000	6.000000	196.300000	123.000000	3.695000	3.325000	17.710000	0.000000	0.000000	4.000000	2.0000
75%	22.800000	8.000000	326.000000	180.000000	3.920000	3.610000	18.900000	1.000000	1.000000	4.000000	4.0000
max	33.900000	8.000000	472.000000	335.000000	4.930000	5.424000	22.900000	1.000000	1.000000	5.000000	8.0000

2] Structure Information for mtcars dataset:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 32 entries, 0 to 31

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	model	32 non-null	object
1	mpg	32 non-null	float64
2	cyl	32 non-null	int64
3	disp	32 non-null	float64
4	hp	32 non-null	int64
5	drat	32 non-null	float64
6	wt	32 non-null	float64
7	qsec	32 non-null	float64
8	vs	32 non-null	int64
9	am	32 non-null	int64
10	gear	32 non-null	int64
11	carb	32 non-null	int64

dtypes: float64(5), int64(6), object(1)

memory usage: 3.1+ KB

None

3] Quartile Information for mtcars dataset:

0.25 15.425

0.50 19.200

0.75 22.800

Name: mpg, dtype: float64

b.

1] Subset of Iris dataset with only rows where Sepal.Width > 3.8 :

	sepal_length	sepal_width	petal_length	petal_width	species
5	5.4	3.9	1.7	0.4	setosa
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
32	5.2	4.1	1.5	0.1	setosa
33	5.5	4.2	1.4	0.2	setosa

2] Aggregate result - Mean sepal length and sepal width for each species:

	sepal_length	sepal_width	petal_length	petal_width
species				
setosa	5.006	3.418	1.464	0.244
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026