EXPERIMENT 10 OUTPUT:-

READING THE CSV FILE

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds.head(3))
```

[1] ✓ 1.6s Python

```
   # P_ID   T_ID   P_Name  P_Gender   Class  Amount  Time_ID    S_ID S_Name
0   30001  40035     Yash         M   First     100   900001  700001    CHE
1   30002  40036   Vedant         M  Second      75   900002  700002    TMN
2   30003  40037    Manav         M  Second      75   900003  700003    KOL
```

DATATYPES OF PARTICULAR COLUMN

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds[' P_Gender'].dtype)
```

[4] ✓ 0.3s Python

```
object
```

COLUMN NAMES

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
columns_list = list(ds.columns)
print('List of columns is:-', columns_list)
```

[5] ✓ 0.3s Python

```
List of columns is:- ['# P_ID', 'T_ID', ' P_Name', ' P_Gender', 'Class', 'Amount', 'Time_ID', 'S_ID', 'S_Name']
```

DATA TYPES

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds.dtypes)
```

✓ 0.4s Python

```
# P_ID        int64
T_ID          int64
 P_Name      object
 P_Gender    object
Class        object
Amount        int64
Time_ID       int64
S_ID          int64
S_Name       object
dtype: object
```

NO OF DIMENSIONS

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds.ndim)
```

✓ 0.3s Python

```
2
```

MAX()

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print('Max Ticket Cost:-')
print(ds['Amount'].max())
```

✓ 0.3s Python

```
Max Ticket Cost:-
100
```

MINIMUM()

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print('Minimum Ticket Cost:-')
print(ds['Amount'].min())
```

[12] ✓ 0.4s Python

```
Minimum Ticket Cost:-
25
```

MEAN()

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print('Mean Ticket Cost:-')
print(ds['Amount'].mean())
```

[14] ✓ 0.2s Python

```
Mean Ticket Cost:-
66.5
```

MEDIAN()

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print('Median of Ticket Cost:-')
print(ds['Amount'].median())
```

[21] ✓ 0.2s Python

```
Median of Ticket Cost:-
75.0
```

GROUP BY

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds.groupby('Class')['Amount'].mean())
```

[20] ✓ 0.4s

```
Class
First     94.642857
Second    75.000000
Third     25.000000
Name: Amount, dtype: float64
```

FILTERING

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds[ds['Amount']<75])
```

[23] ✓ 0.5s

	# P_ID	T_ID	P_Name	P_Gender	Class	Amount	Time_ID	S_ID	S_Name
3	30004	40035	Etash	M	Third	25	900004	700004	PUN
10	30011	40045	Amogh	M	Third	25	900011	700011	AMD
17	30018	40036	Ramesh	M	Third	25	900018	700018	MYS
18	30019	40038	Suresh	M	Third	25	900019	700019	JSL
21	30022	40041	Rohan	M	Third	25	900022	700022	KGP
27	30028	40042	Rachna	F	Third	25	900028	700028	SOL
28	30029	40041	Aishwarya	F	Third	25	900029	700029	BNG
32	30033	40035	Hardik	M	First	25	900033	700033	JAI
34	30035	40036	Kunal	M	Third	25	900035	700035	AMD
35	30036	40035	Shweta	F	Third	25	900036	700036	HAR
36	30037	40037	Sarthak	M	Third	25	900037	700037	JOD
41	30042	40041	Riya	F	Third	25	900042	700042	MYS
43	30044	40039	Kabir	M	Third	25	900044	700044	DEL
46	30047	40035	Jay	M	Third	25	900047	700047	SOL
49	30050	40039	Anjali	F	Third	25	900050	700050	DEL

SLICING

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
print(ds.loc[10:20,['Amount','Class']])
```

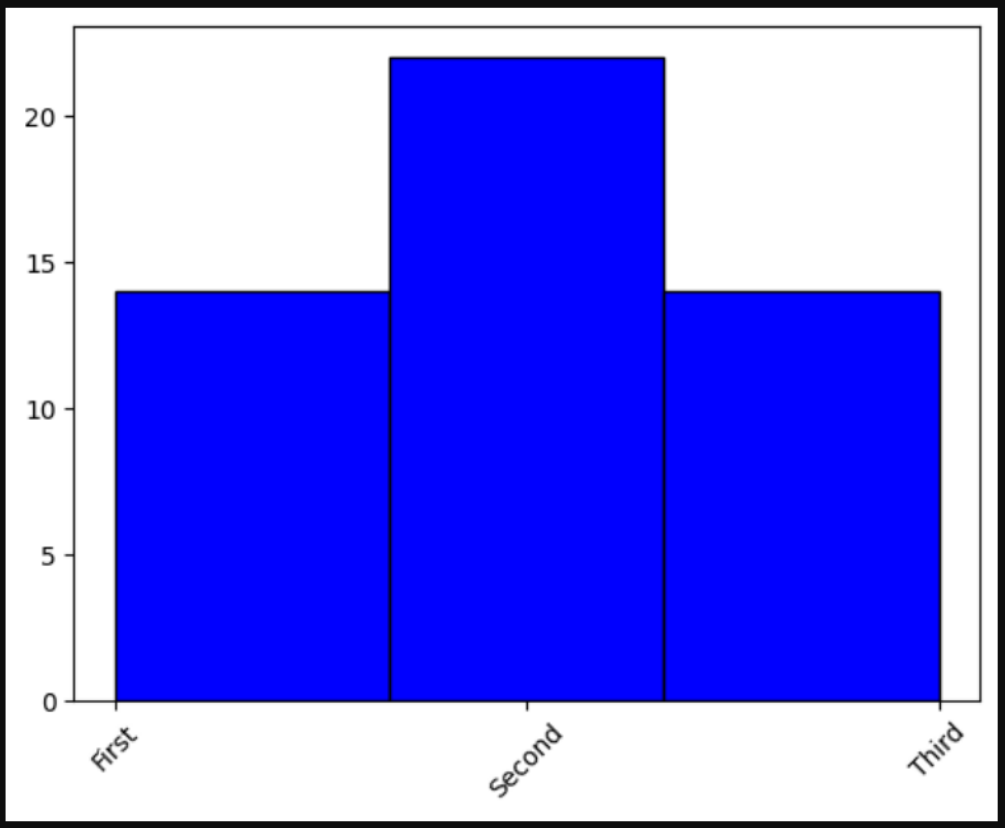
[26] ✓ 0.3s

	Amount	Class
10	25	Third
11	75	Second
12	100	First
13	75	Second
14	100	First
15	75	Second
16	100	First
17	25	Third
18	25	Third
19	75	Second
20	100	First

HISTOGRAM

```
import pandas as pd
import matplotlib.pyplot as plt
ds = pd.read_csv("transport.csv", encoding='latin-1')
fig = plt.figure()
ax = fig.add_subplot(1,1,1)
ax.hist(ds["Class"], bins = 3, facecolor = 'blue', edgecolor = 'black')
plt.xticks(rotation = 45)
plt.show()
```

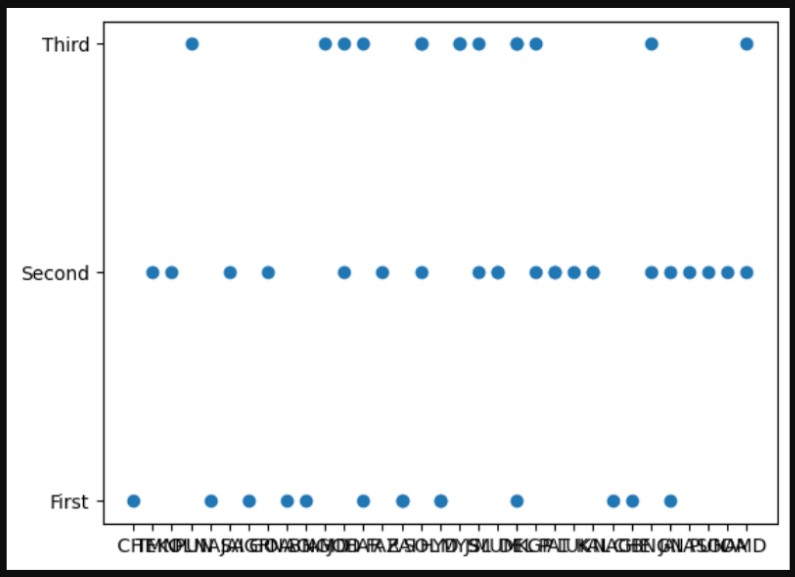
28] ✓ 0.1s



CLUSTERING

```
import pandas as pd
import matplotlib.pyplot as plt
ds = pd.read_csv("transport.csv", encoding='latin-1')
fig = plt.figure()
ax = fig.add_subplot(1,1,1)
ax.scatter(ds['S_Name'], ds['Class'])
plt.show()
```

[53] ✓ 0.1s



NULL COLUMNS WITH 0

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
newds = ds.fillna(0)
print(newds.head())
```

[36] ✓ 0.4s

	# P_ID	T_ID	P_Name	P_Gender	Class	Amount	Time_ID	S_ID	S_Name
0	30001	40035	Yash	M	First	100	900001	700001	CHE
1	30002	40036	Vedant	M	Second	75	900002	700002	TMN
2	30003	40037	Manav	M	Second	75	900003	700003	KOL
3	30004	40035	Etash	M	Third	25	900004	700004	PUN
4	30005	40037	Dalvi	M	First	100	900005	700005	NAS

NOT NULL COLUMNS

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
newds = ds.notnull()
print(newds.head())
```

[37] ✓ 0.2s

	# P_ID	T_ID	P_Name	P_Gender	Class	Amount	Time_ID	S_ID	S_Name
0	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True

DUPLICATE VALUES

```
import pandas as pd
ds = pd.read_csv("transport.csv", encoding='latin-1')
newds = ds.duplicated()
print(newdf)
```

✓ 0.5s

Output exceeds the size limit. Open the full output data in a text editor

#	P_ID	T_ID	P_Name	P_Gender	Class	Amount	Time_ID	S_ID	S_Name
0	30001	40035	Yash	M	First	100	900001	700001	CHE
1	30002	40036	Vedant	M	Second	75	900002	700002	TMN
2	30003	40037	Manav	M	Second	75	900003	700003	KOL
3	30004	40035	Etash	M	Third	25	900004	700004	PUN
4	30005	40037	Dalvi	M	First	100	900005	700005	NAS
5	30006	40038	Krishana	M	Second	75	900006	700006	JAI
6	30007	40039	Ayush	M	First	100	900007	700007	AGR
7	30008	40044	Prajwal	M	Second	75	900008	700008	GOA
8	30009	40035	Yashna	F	First	100	900009	700009	NAG
9	30010	40039	Ananya	F	First	100	900010	700010	BNG
10	30011	40045	Amogh	M	Third	25	900011	700011	AMD
11	30012	40041	Daksha	F	Second	75	900012	700012	JOD
12	30013	40041	Amaan	M	First	100	900013	700013	HAR
13	30014	40042	Bhavesh	M	Second	75	900014	700014	FAZ
14	30015	40045	Komal	F	First	100	900015	700015	RAI
15	30016	40039	Rajesh	M	Second	75	900016	700016	SOL
16	30017	40045	Rakhi	F	First	100	900017	700017	HYD
17	30018	40036	Ramesh	M	Third	25	900018	700018	MYS
18	30019	40038	Suresh	M	Third	25	900019	700019	JSL
19	30020	40037	Nikita	F	Second	75	900020	700020	MUM
20	30021	40044	Sejal	F	First	100	900021	700021	DEL
21	30022	40041	Rohan	M	Third	25	900022	700022	KGP
22	30023	40039	Sailesh	M	Second	75	900023	700023	PAT
23	30024	40035	Archana	F	Second	75	900024	700024	LUK
...									
46	30047	40035	Jay	M	Third	25	900047	700047	SOL
47	30048	40036	Dhruv	M	First	100	900048	700048	RAI
48	30049	40040	Jaya	F	Second	75	900049	700049	JSL
49	30050	40039	Anjali	F	Third	25	900050	700050	DEL