Experiment 8 : Exploratory Data Analysis

```
# importing dependencies
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import numpy as np


df = pd.read_csv('/content/Chile.csv')
df
```

	Unnamed: 0	region	population	sex	age	education	income	statusquo	vote
0	1	N	175000	M	65.0	P	35000.0	1.00820	Y
1	2	N	175000	M	29.0	PS	7500.0	-1.29617	N
2	3	N	175000	F	38.0	P	15000.0	1.23072	Y
3	4	N	175000	F	49.0	P	35000.0	-1.03163	N
4	5	N	175000	F	23.0	S	35000.0	-1.10496	N
...	...	...	...	...	...	...	...	...	...
2695	2696	M	15000	M	42.0	P	15000.0	-1.26247	N
2696	2697	M	15000	F	28.0	P	15000.0	1.32950	Y
2697	2698	M	15000	F	44.0	P	75000.0	1.42045	Y
2698	2699	M	15000	M	21.0	S	75000.0	0.18315	NaN
2699	2700	M	15000	M	20.0	PS	35000.0	1.38179	Y

2700 rows × 9 columns

Using describe function for Descriptive statistics

```
df.describe()
```

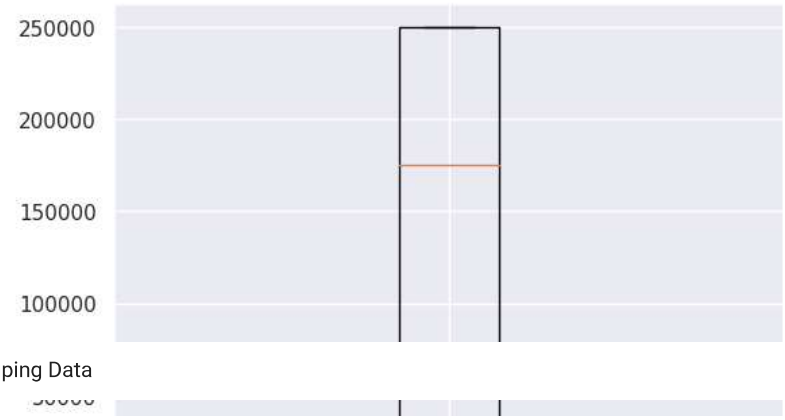
	Unnamed: 0	population	age	income	statusquo
count	2700.000000	2700.000000	2699.000000	2602.000000	2.683000e+03
mean	1350.500000	152222.222222	38.548722	33875.864719	-1.118151e-08
std	779.567188	102198.039602	14.756415	39502.867120	1.000186e+00
min	1.000000	3750.000000	18.000000	2500.000000	-1.803010e+00
25%	675.750000	25000.000000	26.000000	7500.000000	-1.002235e+00
50%	1350.500000	175000.000000	36.000000	15000.000000	-4.558000e-02
75%	2025.250000	250000.000000	49.000000	35000.000000	9.685750e-01
max	2700.000000	250000.000000	70.000000	200000.000000	2.048590e+00

```
df['sex'].value_counts()
```

```
F    1379
M    1321
Name: sex, dtype: int64
```

Box Plot –> outliers and median represantion

```
Y = list(df.population)
plt.boxplot(Y)
plt.show()
```



Grouping Data

```
df.groupby(['education', 'sex']).mean()
```

```
<ipython-input-63-781302169cd9>:1: FutureWarning: The default value of numeric_only in DataFrameGroupBy
  df.groupby(['education', 'sex']).mean()
```

education	sex	Unnamed: 0	population	age	income	statusquo
P	F	1302.410214	131810.131796	43.701812	16699.152542	0.204629
	M	1256.828000	121180.000000	46.004008	18962.655602	0.107066
PS	F	1505.552764	187167.085427	32.321608	70403.645833	-0.145067
	M	1361.562738	179215.779468	35.136882	66749.011858	-0.210100
S	F	1366.724382	169076.855124	34.408127	35324.074074	-0.011237
	M	1407.891697	158944.043321	33.944043	36417.910448	-0.157421

ANOVA –> Analysis of Variance

```
from scipy.stats import f_oneway

group1 = [5,6,3,5,4]
group2 = [33,45,22,44,34]
group3 = [14,12,35,23,12]

# perform ANOVA
f_statistic, p_value = f_oneway(group1, group2, group3)
print(f_statistic)
print(p_value)
```

```
19.15985130111526
0.00018393275764353885
```