We'll check data from "Titanic"(Kashti) data set

Import Libraries

```
In [2]:  import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
```

```
In [10]: kashti = sns.load_dataset("titanic")
         kashti.head()
```

Out[10]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN

Saving DataFrame into CSV File

```
In [4]:  kashti.to_csv("kashti.csv")
```

Saving DataFrame into Excel File

Library for excel\ *pip install openpyxl*

```
In [5]:  kashti.to_excel("kashti.xlsx")
```

Basic Satistics

In [6]: `kashti.describe()`

Out[6]:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [12]: `kashti.head()`

Out[12]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN

Droping few column and make a new Data Set

In [14]: `new_kashti = kashti.drop(['deck','embark_town','alone'],axis=1)`

In [18]: `new_kashti.head()`

Out[18]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	no

Mean of Dataset

In [19]: `kashti.mean()`

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_22004\3332994036.py:1: FutureWarning: Drop
ping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprec
ated; in a future version this will raise TypeError.  Select only valid columns befor
e calling the reduction.
  kashti.mean()
```

Out[19]:
```
survived        0.383838
pclass          2.308642
age            29.699118
sibsp           0.523008
parch           0.381594
fare           32.204208
adult_male      0.602694
alone           0.602694
dtype: float64
```

Mean using Groupby

In [21]: `kashti.groupby(['sex','class']).mean()`

Out[21]:

sex	class	survived	pclass	age	sibsp	parch	fare	adult_male	alone
female	First	0.968085	1.0	34.611765	0.553191	0.457447	106.125798	0.000000	0.361702
	Second	0.921053	2.0	28.722973	0.486842	0.605263	21.970121	0.000000	0.421053
	Third	0.500000	3.0	21.750000	0.895833	0.798611	16.118810	0.000000	0.416667
male	First	0.368852	1.0	41.281386	0.311475	0.278689	67.226127	0.975410	0.614754
	Second	0.157407	2.0	30.740707	0.342593	0.222222	19.741782	0.916667	0.666667
	Third	0.135447	3.0	26.507589	0.498559	0.224784	12.661633	0.919308	0.760807

Value Counts of column variables

In [22]: `kashti.value_counts(['survived'])`

Out[22]:
```
survived
0          549
1          342
dtype: int64
```

Children ansd Women are first

That's why female survival rate is more than male.

In [24]: `kashti[kashti['age']<18].mean()`

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_22004\3332001198.py:1: FutureWarning: Drop
ping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprec
ated; in a future version this will raise TypeError.  Select only valid columns befor
e calling the reduction.
  kashti[kashti['age']<18].mean()
```

Out[24]:   survived   0.539823
            pclass   2.584071
            age   9.041327
            sibsp   1.460177
            parch   1.053097
            fare   31.220798
            adult_male   0.159292
            alone   0.203540
            dtype: float64

Children Survival Rate

In [25]:
```python
kashti[kashti['age']<18].groupby(['sex','class']).mean()
```

Out[25]:

sex	class	survived	pclass	age	sibsp	parch	fare	adult_male	alone
female	First	0.875000	1.0	14.125000	0.500000	0.875000	104.083337	0.000000	0.125000
	Second	1.000000	2.0	8.333333	0.583333	1.083333	26.241667	0.000000	0.166667
	Third	0.542857	3.0	8.428571	1.571429	1.057143	18.727977	0.000000	0.228571
male	First	1.000000	1.0	8.230000	0.500000	2.000000	116.072900	0.250000	0.000000
	Second	0.818182	2.0	4.757273	0.727273	1.000000	25.659473	0.181818	0.181818
	Third	0.232558	3.0	9.963256	2.069767	1.000000	22.752523	0.348837	0.232558