

# Case study (Pandas)

Using Titanic dataset from Seaborn library

```
In [ ]: #importing libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]: #getting the dataset and storing it in a variable

titanic_ship_dataset = sns.load_dataset('titanic')
titanic_ship_dataset.head(3)
```

```
Out[ ]:
```

	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

```
In [ ]: #saving dataframe into csv

titanic_ship_dataset.to_csv('titanic_dataset.csv')
```

```
In [ ]: #basic statistics and summary

titanic_ship_dataset.describe()
```

```
Out[ ]:
```

	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 [ ]: titanic_ship_dataset.head(3)
```

```
Out[ ]:
```

	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

```
In [ ]: #dropping columns from the dataset
```

```
titanic_ship_dataset.drop(['deck', 'embark_town', 'alone'],axis=1)
```

```
Out[ ]:
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True
...	...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False
889	1	1	male	26.0	0	0	30.0000	C	First	man	True
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True

891 rows × 12 columns

```
In [ ]: titanic_ship_dataset.mean()
```

C:\Users\Muhammad Mustafa\AppData\Local\Temp\ipykernel\_13684\1242517853.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric\_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
titanic_ship_dataset.mean()
```

```
Out[ ]:
```

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

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

```
Out[ ]:
```

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

```
In [ ]: titanic_ship_dataset.value_counts('survived')
```

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

```
In [ ]: # as we can see that the number of females are very high in survival rate lets see if
titanic_ship_dataset[titanic_ship_dataset['age'] < 18].mean()
```

C:\Users\Muhammad Mustafa\AppData\Local\Temp\ipykernel\_13684\1336266491.py:2: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric\_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
titanic_ship_dataset[titanic_ship_dataset['age'] < 18].mean()
```

```
Out[ ]: 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
```

```
In [ ]: #now we can see how many childrens were survived

titanic_ship_dataset[titanic_ship_dataset['age'] < 18].groupby(['sex', 'class']).mean()
```

```
Out[ ]:
```

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

```
In [ ]: print(titanic_ship_dataset['age'].mode())
```

```
0    24.0
Name: age, dtype: float64
```

```
In [ ]: #importing the Pakistan hunger and food insecurity
```

```
PK_HFI = pd.read_csv('hunger and food insecurity Pakistan dataset.csv')
```

```
In [ ]: PK_HFI.head()
```

```
Out[ ]:
```

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Item	Year Code	Year	
0	FS	Suite of Food Security Indicators	165	Pakistan	6132	Value	210011	Number of people undernourished (million) (3-y...	20002002	2000-2002	r
1	FS	Suite of Food Security Indicators	165	Pakistan	6132	Value	210011	Number of people undernourished (million) (3-y...	20012003	2001-2003	r
2	FS	Suite of Food Security Indicators	165	Pakistan	6132	Value	210011	Number of people undernourished (million) (3-y...	20022004	2002-2004	r
3	FS	Suite of Food Security Indicators	165	Pakistan	6132	Value	210011	Number of people undernourished (million) (3-y...	20032005	2003-2005	r
4	FS	Suite of Food Security Indicators	165	Pakistan	6132	Value	210011	Number of people undernourished (million) (3-y...	20042006	2004-2006	r

```
In [ ]: PK_HFI.describe()
```

```
Out[ ]:
```

	Area Code	Element Code	Item Code	Year Code	Value	Note
count	19.0	19.0	19.0	1.900000e+01	19.000000	0.0
mean	165.0	6132.0	210011.0	2.009201e+07	28.152632	NaN
std	0.0	0.0	0.0	5.627877e+04	2.177248	NaN
min	165.0	6132.0	210011.0	2.000200e+07	25.200000	NaN
25%	165.0	6132.0	210011.0	2.004701e+07	26.650000	NaN
50%	165.0	6132.0	210011.0	2.009201e+07	28.000000	NaN
75%	165.0	6132.0	210011.0	2.013702e+07	29.450000	NaN
max	165.0	6132.0	210011.0	2.018202e+07	32.500000	NaN

