In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:
```
titanic = pd.read_csv('D:\\24 - Machine_Learning\\download files\\titanic.csv')
titanic
```

Out[2]:

	PassengerId	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	S
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	C
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	S
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	S
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	C

418 rows × 11 columns

In [ ]:
```
titanic.drop('Cabin', axis=1, inplace=True)
```

In [6]:
```
titanic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Pclass       418 non-null    int64  
 2   Name         418 non-null    object 
 3   Gender       418 non-null    object 
 4   Age          332 non-null    float64
 5   SibSp        418 non-null    int64  
 6   Parch        418 non-null    int64  
 7   Ticket       418 non-null    object 
 8   Fare         417 non-null    float64
 9   Embarked     418 non-null    object 
dtypes: float64(2), int64(4), object(4)
memory usage: 32.8+ KB
```

In [7]:
```
titanic.dropna(inplace=True)
# It will remove any Nan value
```

In [8]:
```
titanic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 331 entries, 0 to 415
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  331 non-null    int64  
 1   Pclass       331 non-null    int64  
 2   Name         331 non-null    object 
 3   Gender       331 non-null    object 
 4   Age          331 non-null    float64
 5   SibSp        331 non-null    int64  
 6   Parch        331 non-null    int64  
 7   Ticket       331 non-null    object 
 8   Fare         331 non-null    float64
 9   Embarked     331 non-null    object 
dtypes: float64(2), int64(4), object(4)
memory usage: 28.4+ KB
```

In [9]:
```
# Convert categorical variable into dummy/indicator variables
pd.get_dummies(titanic['Gender'])
```

Out[9]:

	female	male
0	0	1
1	1	0
2	0	1
3	0	1
4	1	0
...	...	...
409	1	0
411	1	0
412	1	0
414	1	0
415	0	1

331 rows × 2 columns

In [10]:
```
pd.get_dummies(titanic['Gender'], drop_first=True)
```

Out[10]:

	male
0	1
1	0
2	1
3	1
4	0
...	...
409	0
411	0
412	0
414	0
415	1

331 rows × 1 columns

In [11]:
```
gender = pd.get_dummies(titanic['Gender'], drop_first=False)
gender
```

Out[11]:

	female	male
0	0	1
1	1	0
2	0	1
3	0	1
4	1	0
...	...	...
409	1	0
411	1	0
412	1	0
414	1	0
415	0	1

331 rows × 2 columns

In [12]:
```
pd.get_dummies(titanic['Embarked'], drop_first=True)
```

Out[12]:

	Q	S
0	1	0
1	0	1
2	1	0
3	0	1
4	0	1
...	...	...
409	0	1
411	1	0
412	0	1
414	0	0
415	0	1

331 rows × 2 columns

In [13]:
```
embarked = pd.get_dummies(titanic['Embarked'], drop_first=False)
embarked
```

Out[13]:

	C	Q	S
0	0	1	0
1	0	0	1
2	0	1	0
3	0	0	1
4	0	0	1
...	...	...	...
409	0	0	1
411	0	1	0
412	0	0	1
414	1	0	0
415	0	0	1

331 rows × 3 columns

In [14]:
```
titanic.drop(['Gender', 'Embarked', 'Name', 'Ticket'], axis=1, inplace=True)
```

In [15]:
```
titanic.head()
```

Out[15]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare
0	892	3	34.5	0	0	7.8292
1	893	3	47.0	1	0	7.0000
2	894	2	62.0	0	0	9.6875
3	895	3	27.0	0	0	8.6625
4	896	3	22.0	1	1	12.2875

In [16]:
```
titanic = pd.concat([titanic,gender,embarked],axis=1)
```

In [17]:
```
titanic.head()
```

Out[17]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	female	male	C	Q	S
0	892	3	34.5	0	0	7.8292	0	1	0	1	0
1	893	3	47.0	1	0	7.0000	1	0	0	0	1
2	894	2	62.0	0	0	9.6875	0	1	0	1	0
3	895	3	27.0	0	0	8.6625	0	1	0	0	1
4	896	3	22.0	1	1	12.2875	1	0	0	0	1