importing the packages

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
from sklearn.model_selection import train_test_split
```

importing the dataset

In [2]:
```
df=pd.read_csv(r"C:\Users\MY HOME\Desktop\datascience\passenger.csv")
df
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.
...	...	...	...	...	...	...	...	...	...	
413	1305	0	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.
416	1308	0	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.
417	1309	0	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.

418 rows × 12 columns

Data cleaning

In [3]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [4]:
```
1 df.describe()
```

Out[4]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	332.000000	418.000000	418.000000	417.000000
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.481622	0.841838	14.181209	0.896760	0.981429	55.907576
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	0.000000	1.000000	21.000000	0.000000	0.000000	7.895800
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	0.000000	14.454200
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	0.000000	31.500000
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	512.329200

Filling the null values

In [5]:
```
df.fillna(method="ffill",inplace=True)
df
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.
...	...	...	...	...	...	...	...	...	...	
413	1305	0	3	Spector, Mr. Woolf	male	28.0	0	0	A.5. 3236	8.
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.
416	1308	0	3	Ware, Mr. Frederick	male	38.5	0	0	359309	8.
417	1309	0	3	Peter, Master. Michael J	male	38.5	1	1	2668	22.

418 rows × 12 columns

In [6]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          418 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         418 non-null    float64
 10  Cabin        406 non-null    object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [7]:
```
1 df.fillna(method="ffill",inplace=True)
```

In [8]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          418 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         418 non-null    float64
 10  Cabin        406 non-null    object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [9]:
```
1 df.drop('Cabin',axis=1,inplace=True)
```

In [10]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          418 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         418 non-null    float64
 10  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 36.1+ KB
```

In [11]:
```
1 df.isna().any()
```

Out[11]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Embarked       False
dtype: bool
```

In [12]:
```
1 df.shape
```

Out[12]:
```
(418, 11)
```

In [13]:
```
1 df.isnull().sum()
```

Out[13]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [14]:
```
convert={"Sex":{"female":0,"male":1}}
df=df.replace(convert)
df
```

Out[14]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	892	0	3	Kelly, Mr. James	1	34.5	0	0	330911	7.82
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	0	47.0	1	0	363272	7.00
2	894	0	2	Myles, Mr. Thomas Francis	1	62.0	0	0	240276	9.68
3	895	0	3	Wirz, Mr. Albert	1	27.0	0	0	315154	8.66
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	0	22.0	1	1	3101298	12.28
...	...	...	...	...	...	...	...	...	...	
413	1305	0	3	Spector, Mr. Woolf	1	28.0	0	0	A.5. 3236	8.05
414	1306	1	1	Oliva y Ocana, Dona. Fermina	0	39.0	0	0	PC 17758	108.90
415	1307	0	3	Saether, Mr. Simon Sivertsen	1	38.5	0	0	SOTON/O.Q. 3101262	7.25
416	1308	0	3	Ware, Mr. Frederick	1	38.5	0	0	359309	8.05
417	1309	0	3	Peter, Master. Michael J	1	38.5	1	1	2668	22.35

418 rows × 11 columns

In [15]:
```
df["Embarked"].value_counts()
```

Out[15]:
```
Embarked
S    270
C    102
Q     46
Name: count, dtype: int64
```

In [16]:
```
convert={"Embarked":{"S":0,"C":1,"Q":2}}
df=df.replace(convert)
df
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	892	0	3	Kelly, Mr. James	1	34.5	0	0	330911	7.82
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	0	47.0	1	0	363272	7.00
2	894	0	2	Myles, Mr. Thomas Francis	1	62.0	0	0	240276	9.68
3	895	0	3	Wirz, Mr. Albert	1	27.0	0	0	315154	8.66
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	0	22.0	1	1	3101298	12.28
...	...	...	...	...	...	...	...	...	...	
413	1305	0	3	Spector, Mr. Woolf	1	28.0	0	0	A.5. 3236	8.05
414	1306	1	1	Oliva y Ocana, Dona. Fermina	0	39.0	0	0	PC 17758	108.90
415	1307	0	3	Saether, Mr. Simon Sivertsen	1	38.5	0	0	SOTON/O.Q. 3101262	7.25
416	1308	0	3	Ware, Mr. Frederick	1	38.5	0	0	359309	8.05
417	1309	0	3	Peter, Master. Michael J	1	38.5	1	1	2668	22.35

418 rows × 11 columns

Train a model

In [17]:
```
x=["PassengerId","Pclass","Sex","Embarked","Parch"]
y=["0","1"]
all_inputs=df[x]
all_classes=df["Survived"]
x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,t
```

In [18]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [19]:
```
clf=DecisionTreeClassifier(random_state=3)
clf.fit(x_train,y_train)
score=clf.score(x_test,y_test)
print(score)
```

```
1.0
```

In [20]:
```
from sklearn.linear_model import Lasso
```

In [21]:
```
lasso=Lasso(alpha=0)
lasso.fit(x_train,y_train)
clf=lasso.score(x_test,y_test)
print(clf)
```

```
0.9999999999740633

C:\Users\MY HOME\AppData\Local\Temp\ipykernel_9788\3365621483.py:2: UserWa
rning: With alpha=0, this algorithm does not converge well. You are advise
d to use the LinearRegression estimator
  lasso.fit(x_train,y_train)
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\linear_model\_coordinate_descent.py:631: UserWarning: Coordinate
descent with no regularization may lead to unexpected results and is disco
uraged.
  model = cd_fast.enet_coordinate_descent(
```

In [ ]:
```

```