Import The Necessary Libraries

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

Import Dataset

```
dataset=pd.read_csv("Titanic-Dataset.csv")
dataset
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	

Checking For Null Values

```
dataset.isnull().any()
```

```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

```
dataset["Age"].fillna(dataset["Age"].mean(),inplace=True)
dataset
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	NaN

```
dataset["Embarked"].fillna(dataset["Embarked"].mode()[0],inplace=True)
dataset
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0500	NaN

```
dataset["Cabin"].fillna(dataset["Cabin"].mode()[1],inplace=True)
dataset
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	C23 C25 C27
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9250	C23 C25 C27
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0500	C23 C25 C27
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13.0000	C23 C25 C27

```
dataset.isnull().sum()
```

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

Data Visualization

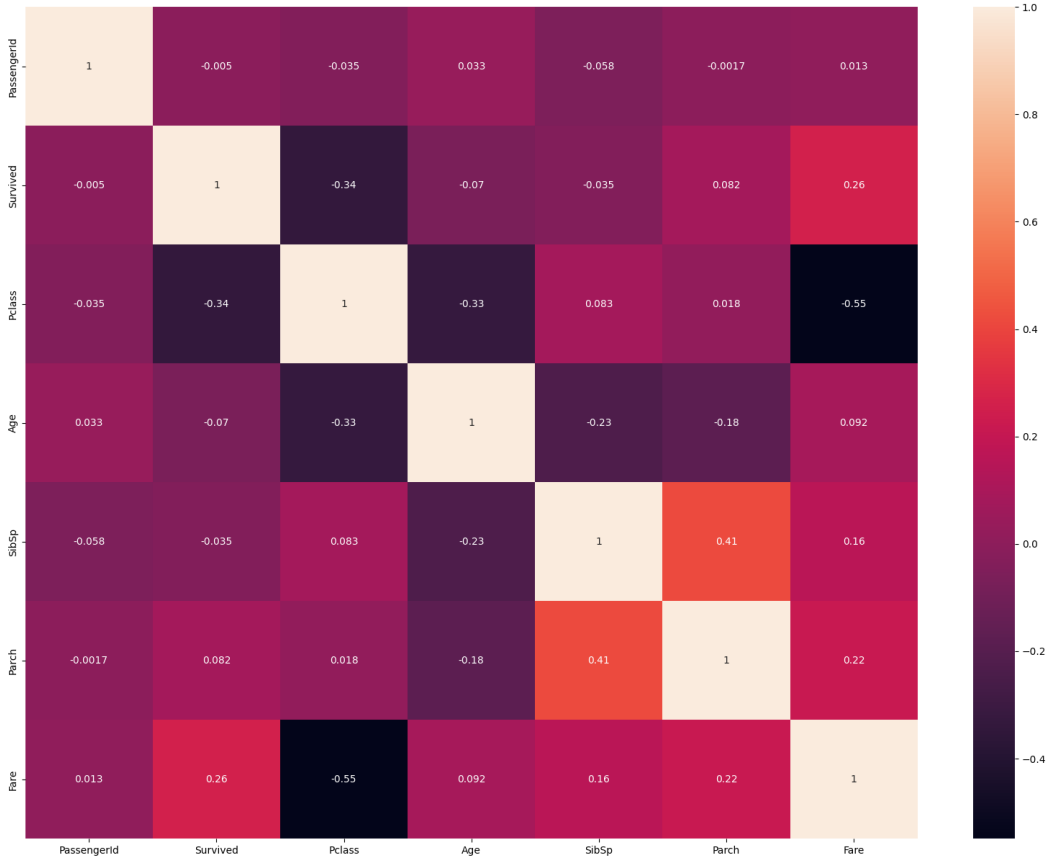
```
corr=dataset.corr()
corr
```

```
<ipython-input-8-f22ca9e9dc13>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is
  corr=dataset.corr()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.033207	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.069809	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.331339	0.083081	0.018443	-0.549500
Age	0.033207	-0.069809	-0.331339	1.000000	-0.232625	-0.179191	0.091566
SibSp	-0.057527	-0.035322	0.083081	-0.232625	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.179191	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.091566	0.159651	0.216225	1.000000

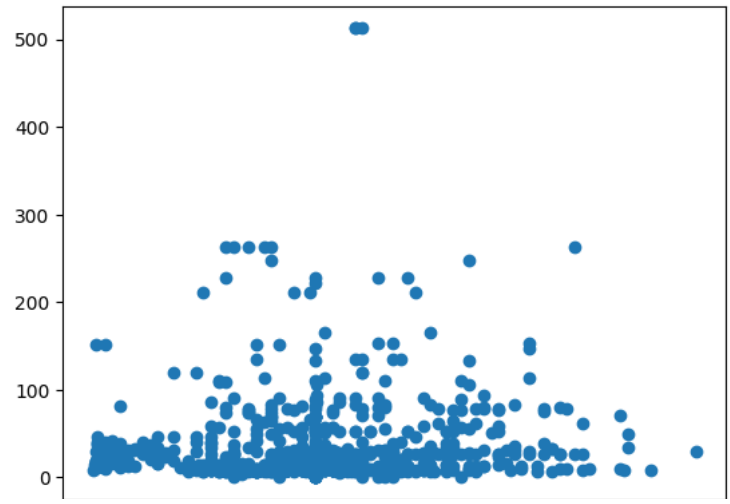
```
plt.subplots(figsize=(20,15))
sns.heatmap(corr,annot=True)
```

```
<Axes: >
```




```
plt.scatter(dataset["Age"],dataset["Fare"])
```

```
<matplotlib.collections.PathCollection at 0x7d41e6a164d0>
```



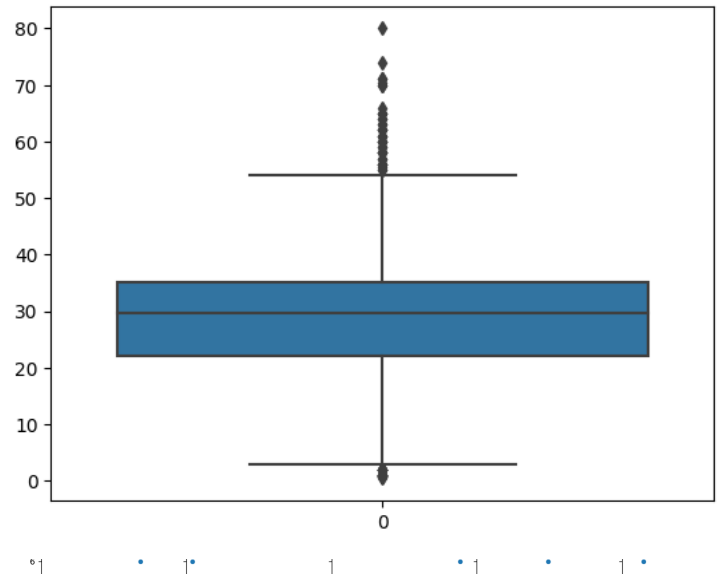

```
sns.pairplot(dataset)
```

```
<seaborn.axisgrid.PairGrid at 0x7d41e67746a0>
```

Outlier Detection

```
sns.boxplot(dataset.Age)
```

```
<Axes: >
```




```
q1=dataset.Age.quantile(0.565)
q1
```

```
29.69911764705882
```

```
q2=dataset.Age.quantile(0.68)
q2
```

```
32.0
```

```
q3=dataset.Age.quantile(1)
q3
```

```
80.0
```

```
IQR=q3-q1
```

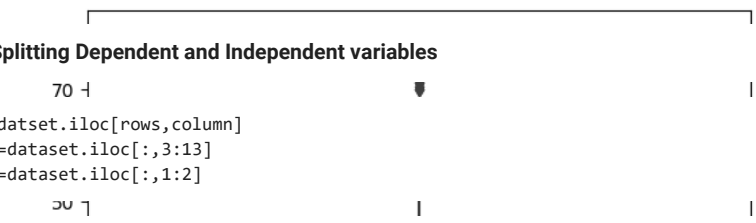
```
upper_limit=q3+1.5*IQR
upper_limit
```

```
155.45132352941175
```

```
dataset['Age']=np.where(dataset['Age']>upper_limit,30,dataset['Age'])
```

```
sns.boxplot(dataset.Age)
```

```
<Axes: >
```



Splitting Dependent and Independent variables

```
#datset.iloc[rows,column]
x=dataset.iloc[:,3:13]
y=dataset.iloc[:,1:2]
```

```
y.head()
```

	Survived
0	0
1	1
2	1
3	1
4	0

```
x.head()
```

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	C23 C25 C27	S
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	C23 C25 C27	S
3	Futrelle, Mrs. Jacques Heath								

Perform Encoding

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x["Sex"]=le.fit_transform(x["Sex"])
x["Sex"]
```

```
0      1
1      0
2      0
3      0
4      1
      ..
886    1
887    0
888    0
889    1
890    1
Name: Sex, Length: 891, dtype: int64
```

```
x.head()
```

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	C23 C25 C27	S
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	C85	C
2	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	C23 C25 C27	S
3	Futrelle, Mrs. Jacques Heath (Lily								

```
x.Embarked.value_counts()
```

```
S    646
C    168
Q     77
Name: Embarked, dtype: int64
```

```
Embarked=pd.get_dummies(x["Embarked"],drop_first=True)
Embarked
```

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1
...	...	...
886	0	1
887	0	1
888	0	1
889	0	0
890	1	0

891 rows × 2 columns

```
x=pd.concat([x,Embarked],axis=1)
```

```
x.head()
```

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Q	S
0	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	C23 C25 C27	S	0	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	C85	C	0	0
2	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	C23 C25 C27	S	0	1
3	Futrelle, Mrs. Jacques Heath										

```
x.drop(["Embarked"],axis=1,inplace=True)
x.head(6)
```

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Q	S
0	Braund, Mr. Owen Harris	1	22.000000	1	0	A/5 21171	7.2500	C23 C25 C27	0	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.000000	1	0	PC 17599	71.2833	C85	0	0
2	Heikkinen, Miss. Laina	0	26.000000	0	0	STON/O2. 3101282	7.9250	C23 C25 C27	0	1
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.000000	1	0	113803	53.1000	C123	0	1

Splitting Data into Train and Test

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.5,random_state=0)
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
((445, 10), (446, 10), (445, 1), (446, 1))
```

Feature Scaling

```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
y_train=sc.fit_transform(y_train)
y_test=sc.fit_transform(y_test)
y_train
```

```
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453]])
```

```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_train=sc.fit_transform(y_train)
x_test=sc.fit_transform(y_test)
x_train
```

```
array([[-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
```

```
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [-0.82032453],
       [ 1.21902975],
```