```
[27]:  import pandas as pd
       import numpy as np
       from sklearn.preprocessing import LabelEncoder, MinMaxScaler
```

```
[48]:  df = pd.read_csv("Titanic-Dataset.csv")
```

```
[50]:  df.head()
```

[50]:

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

```
[52]:  df.isnull()
```

[52]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns

```
[54]:  df.isnull().sum()
```

```
[54]:  PassengerId      0
       Survived         0
       Pclass           0
       Name             0
       Sex              0
       Age            177
       SibSp            0
       Parch            0
       Ticket           0
       Fare             0
       Cabin          687
       Embarked         2
       dtype: int64
```

```
[56]: df.dtypes
```

```
[56]: PassengerId      int64
      Survived         int64
      Pclass           int64
      Name            object
      Sex             object
      Age            float64
      SibSp            int64
      Parch            int64
      Ticket          object
      Fare           float64
      Cabin           object
      Embarked        object
      dtype: object
```

```
[58]: df['Age'].fillna(df['Age'].median(), inplace=True)  # Fill missing with median
      df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)  # Fill missing with mode
      df.drop(columns=['Cabin'], inplace=True) # drop because more nullval
```

```
[64]: df.head()
```

[64]:

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

```
[66]: df.dropna(subset=['Fare'], inplace=True) # droping row of missing value
```

```
[70]: df.drop(columns=['Ticket'], inplace=True) # droping column of missing value
```

```
[72]: df.head()
```

[72]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	S

```
[60]: df.isnull().sum()
```

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