In [ ]:
```
NAME : SHINDE SHUBHAM DNYANDEV,     ROLL NO. : EN23107121,    BATCH : C
```

In [1]:
```
import pandas as pd
```

In [7]:
```
df = pd.read_csv("/home/admin1/Titanic.csv")
df
```

Out[7]:

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

891 rows × 10 columns

In [23]:
```
df.head()
```

Out[23]:

	sex	age	sibsp	parch	fare	embarked	class	who	alone	survived
0	male	22.0	1	0	7.2500	S	Third	man	False	0
1	female	38.0	1	0	71.2833	C	First	woman	False	1
2	female	26.0	0	0	7.9250	S	Third	woman	True	1
3	female	35.0	1	0	53.1000	S	First	woman	False	1
4	male	35.0	0	0	8.0500	S	Third	man	True	0

In [25]:
```
df.isnull()
```

Out[25]:

	sex	age	sibsp	parch	fare	embarked	class	who	alone	survived
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False
887	False	False	False	False	False	False	False	False	False	False
888	False	True	False	False	False	False	False	False	False	False
889	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False

891 rows × 10 columns

In [35]:
```
df.isnull().sum()
```

Out[35]:
```
sex           0
age         177
sibsp         0
parch         0
fare          0
embarked      2
class         0
who           0
alone         0
survived      0
dtype: int64
```

In [37]:
```
df.describe()
```

Out[37]:

	age	sibsp	parch	fare	survived
count	714.000000	891.000000	891.000000	891.000000	891.000000
mean	29.699118	0.523008	0.381594	32.204208	0.383838
std	14.526497	1.102743	0.806057	49.693429	0.486592
min	0.420000	0.000000	0.000000	0.000000	0.000000
25%	20.125000	0.000000	0.000000	7.910400	0.000000
50%	28.000000	0.000000	0.000000	14.454200	0.000000
75%	38.000000	1.000000	0.000000	31.000000	1.000000
max	80.000000	8.000000	6.000000	512.329200	1.000000

In [43]:
```
df.dtypes
```

Out[43]:
```
sex          object
age         float64
sibsp         int64
parch         int64
fare        float64
embarked     object
class        object
who          object
alone          bool
survived      int64
dtype: object
```

In [45]:
```
df.shape
```

Out[45]:
```
(891, 10)
```

In [47]:
```
df.size
```

Out[47]:
```
8910
```

In [51]:
```
df.ndim
```

Out[51]:
```
2
```

In [57]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 10 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   sex       891 non-null    object
 1   age       714 non-null    float64
 2   sibsp     891 non-null    int64
 3   parch     891 non-null    int64
 4   fare      891 non-null    float64
 5   embarked  889 non-null    object
 6   class     891 non-null    object
 7   who       891 non-null    object
 8   alone     891 non-null    bool
 9   survived  891 non-null    int64
dtypes: bool(1), float64(2), int64(3), object(4)
memory usage: 63.6+ KB
```

In [101…
```
df['sex'] = df['sex'].astype('category')
```

In [103…
```
df.dtypes
```

Out[103…
```
sex         category
age          float64
sibsp          int64
parch          int64
fare           int64
embarked    category
class         object
who           object
alone           bool
survived       int64
dtype: object
```

In [105…
```
df['embarked'] = df['embarked'].astype('category')
```

In [107…
```
df.dtypes
```

Out[107…
```
sex          category
age           float64
sibsp           int64
parch           int64
fare            int64
embarked     category
class          object
who            object
alone            bool
survived        int64
dtype: object
```

In [109…
```
count = df.select_dtypes('int64')
count
```

Out[109…

	sibsp	parch	fare	survived
0	1	0	7	0
1	1	0	71	1
2	0	0	7	1
3	1	0	53	1
4	0	0	8	0
...	...	...	...	...
886	0	0	13	0
887	0	0	30	1
888	1	2	23	0
889	0	0	30	1
890	0	0	7	0

891 rows × 4 columns

In [111…
```
df
```

Out[111…

	sex	age	sibsp	parch	fare	embarked	class	who	alone	survived
0	male	22.0	1	0	7	S	Third	man	False	0
1	female	38.0	1	0	71	C	First	woman	False	1
2	female	26.0	0	0	7	S	Third	woman	True	1
3	female	35.0	1	0	53	S	First	woman	False	1
4	male	35.0	0	0	8	S	Third	man	True	0
...	...	...	...	...	...	...	...	...	...	...
886	male	27.0	0	0	13	S	Second	man	True	0
887	female	19.0	0	0	30	S	First	woman	True	1
888	female	NaN	1	2	23	S	Third	woman	False	0
889	male	26.0	0	0	30	C	First	man	True	1
890	male	32.0	0	0	7	Q	Third	man	True	0

891 rows × 10 columns

In [113…
```
df['sex'].replace(['male','female'],['female','male'])
```

Out[113…
```
0      female
1        male
2        male
3        male
4      female
        ...
886    female
887      male
888      male
889    female
890    female
Name: sex, Length: 891, dtype: category
Categories (2, object): ['female', 'male']
```

In [115…
```
df.age.unique()
```

Out[115…
```
array([22.  , 38.  , 26.  , 35.  ,   nan, 54.  ,  2.  , 27.  , 14.  ,
        4.  , 58.  , 20.  , 39.  , 55.  , 31.  , 34.  , 15.  , 28.  ,
        8.  , 19.  , 40.  , 66.  , 42.  , 21.  , 18.  ,  3.  ,  7.  ,
       49.  , 29.  , 65.  , 28.5 ,  5.  , 11.  , 45.  , 17.  , 32.  ,
       16.  , 25.  ,  0.83, 30.  , 33.  , 23.  , 24.  , 46.  , 59.  ,
       71.  , 37.  , 47.  , 14.5 , 70.5 , 32.5 , 12.  ,  9.  , 36.5 ,
       51.  , 55.5 , 40.5 , 44.  ,  1.  , 61.  , 56.  , 50.  , 36.  ,
       45.5 , 20.5 , 62.  , 41.  , 52.  , 63.  , 23.5 ,  0.92, 43.  ,
       60.  , 10.  , 64.  , 13.  , 48.  ,  0.75, 53.  , 57.  , 80.  ,
       70.  , 24.5 ,  6.  ,  0.67, 30.5 ,  0.42, 34.5 , 74.  ])
```

In [117…
```
df.age.max()
```

Out[117…
```
80.0
```

In [119…
```
df.age.min()
```

Out[119…
```
0.42
```

In [135…
```
from sklearn.preprocessing import LabelEncoder
```

In [137…
```
LE = LabelEncoder()
LE
```

Out[137…
▾ LabelEncoder

LabelEncoder()

In [141…
```
df.embarked = LE.fit_transform(df.embarked)
```

In [143…
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 10 columns):
 #   Column    Non-Null Count  Dtype   
---  ------    --------------  -----   
 0   sex       891 non-null    category
 1   age       714 non-null    float64 
 2   sibsp     891 non-null    int64   
 3   parch     891 non-null    int64   
 4   fare      891 non-null    int64   
 5   embarked  891 non-null    int64   
 6   class     891 non-null    object  
 7   who       891 non-null    object  
 8   alone     891 non-null    bool    
 9   survived  891 non-null    int64   
dtypes: bool(1), category(1), float64(1), int64(5), object(2)
memory usage: 57.7+ KB
```

In [ ]: