```
import pandas as pd
from sklearn.preprocessing import LabelEncoder as le
from sklearn.preprocessing import OneHotEncoder as ohe
import copy
```

```
data = pd.read_csv('titanic.csv')
```

```
data.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

Next steps: Generate code with data | View recommended plots | New interactive sheet

```
data.shape
```

```
(418, 12)
```

```
data.size
```

```
5016
```

```
data.dtypes
```

	0
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

```
data.isnull().sum()
```

	0
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	86
SibSp	0
Parch	0
Ticket	0
Fare	1
Cabin	327
Embarked	0

dtype: int64

```
data.describe()
```

	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

ENCODING

LABEL ENCODING

```
dflabel=copy.deepcopy(data)
```

```
labelencoder_x=le()
dflabel['Sex']=labelencoder_x.fit_transform(dflabel['Sex'])
```

```
dflabel.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	1	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	0	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	1	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	1	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	0	22.0	1	1	3101298	12.2875	NaN	S

Next steps: Generate code with `dflabel` View recommended plots New interactive sheet

```
labelencoder_x=le()
dflabel['Embarked']=labelencoder_x.fit_transform(dflabel['Embarked'])
```

```
dflabel.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	1	34.5	0	0	330911	7.8292	NaN	1
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	0	47.0	1	0	363272	7.0000	NaN	2
2	894	0	2	Myles, Mr. Thomas Francis	1	62.0	0	0	240276	9.6875	NaN	1
3	895	0	3	Wirz, Mr. Albert	1	27.0	0	0	315154	8.6625	NaN	2
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	0	22.0	1	1	3101298	12.2875	NaN	2

Next steps: Generate code with dflabel | View recommended plots | New interactive sheet

```
dflabel.drop(['Name', 'Cabin','Ticket','Fare'], axis=1,inplace=True)
```

```
dflabel.head(30)
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
0	892	0	3	1	34.5	0	0	1
1	893	1	3	0	47.0	1	0	2
2	894	0	2	1	62.0	0	0	1
3	895	0	3	1	27.0	0	0	2
4	896	1	3	0	22.0	1	1	2
5	897	0	3	1	14.0	0	0	2
6	898	1	3	0	30.0	0	0	1
7	899	0	2	1	26.0	1	1	2
8	900	1	3	0	18.0	0	0	0
9	901	0	3	1	21.0	2	0	2
10	902	0	3	1	NaN	0	0	2
11	903	0	1	1	46.0	0	0	2
12	904	1	1	0	23.0	1	0	2
13	905	0	2	1	63.0	1	0	2
14	906	1	1	0	47.0	1	0	2
15	907	1	2	0	24.0	1	0	0
16	908	0	2	1	35.0	0	0	1
17	909	0	3	1	21.0	0	0	0
18	910	1	3	0	27.0	1	0	2
19	911	1	3	0	45.0	0	0	0
20	912	0	1	1	55.0	1	0	0
21	913	0	3	1	9.0	0	1	2
22	914	1	1	0	NaN	0	0	2
23	915	0	1	1	21.0	0	1	0
24	916	1	1	0	48.0	1	3	0
25	917	0	3	1	50.0	1	0	2
26	918	1	1	0	22.0	0	1	0
27	919	0	3	1	22.5	0	0	0
28	920	0	1	1	41.0	0	0	2
29	921	0	3	1	NaN	2	0	0

Next steps: Generate code with dflabel | View recommended plots | New interactive sheet

```
dflabel.dtypes
```

	0
PassengerId	int64
Survived	int64
Pclass	int64
Sex	int64
Age	float64
SibSp	int64
Parch	int64
Embarked	int64

dtype: object

ONE HOT ENCODING

```
dfhotencoder=copy.deepcopy(data)
```

```
result=ohe().fit_transform(dfhotencoder['Sex'].values.reshape(-1,1)).toarray()
dfhotencoder[['Female','Male']]=pd.DataFrame(result,index=dfhotencoder.index)
```

```
dfhotencoder.drop('Sex',axis=1,inplace=True)
```

```
result=ohe().fit_transform(dfhotencoder['Embarked'].values.reshape(-1,1)).toarray()
dfhotencoder[['Q','S','C']]=pd.DataFrame(result,index=dfhotencoder.index)
```

```
dfhotencoder.drop('Embarked',axis=1,inplace=True)
```

```
dfhotencoder.drop(['Name', 'Cabin','Ticket','Fare'], axis=1,inplace=True)
```

```
dfhotencoder.head(10)
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Female	Male	Q	S	C
0	892	0	3	34.5	0	0	0.0	1.0	0.0	1.0	0.0
1	893	1	3	47.0	1	0	1.0	0.0	0.0	0.0	1.0
2	894	0	2	62.0	0	0	0.0	1.0	0.0	1.0	0.0
3	895	0	3	27.0	0	0	0.0	1.0	0.0	0.0	1.0
4	896	1	3	22.0	1	1	1.0	0.0	0.0	0.0	1.0
5	897	0	3	14.0	0	0	0.0	1.0	0.0	0.0	1.0
6	898	1	3	30.0	0	0	1.0	0.0	0.0	1.0	0.0
7	899	0	2	26.0	1	1	0.0	1.0	0.0	0.0	1.0
8	900	1	3	18.0	0	0	1.0	0.0	1.0	0.0	0.0
9	901	0	3	21.0	2	0	0.0	1.0	0.0	0.0	1.0

Next steps: Generate code with dfhotencoder | View recommended plots | New interactive sheet