In [1]:

```
import pandas as pd
```

In [2]:

```
dataset = pd.read_csv('titanic_train.csv')
```

In [3]:

```
dataset
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [4]:

```
y= dataset['Survived']
```

In [5]:

```
X= dataset[[ 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Cabin', 'Embarked']]
```

In [6]:

```
sex= X['Sex']
```

In [7]:

```
sex
```

Out[7]:

```
0        male
1      female
2      female
3      female
4        male
       ...
886      male
887    female
888    female
889      male
```

```
890      male
Name: Sex, Length: 891, dtype: object
```

In [8]:

```
#One-hot-Encoding: data preprocessing
sex = pd.get_dummies("Sex",drop_first=True)
```

In [9]:

```
pclass = X['Pclass']
```

In [10]:

```
pclass = pd.get_dummies(pclass, drop_first=True)
```

In [11]:

```
sibsp = X['SibSp']
```

In [12]:

```
sibsp = pd.get_dummies(pclass, drop_first=True)
```

In [13]:

```
dataset.isnull()
```

Out[13]:

	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

In [14]:

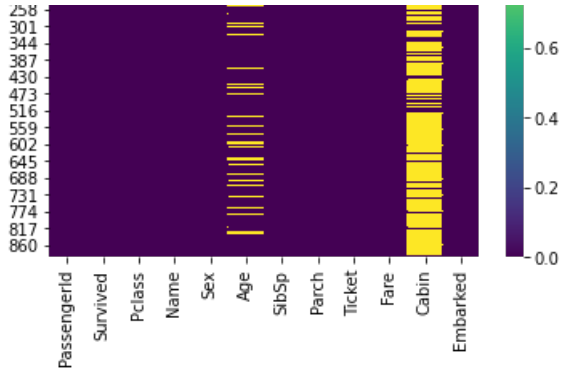
```
import seaborn as sns
```

In [15]:

```
sns.heatmap(dataset.isnull(),cmap='viridis')
```

Out[15]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xc5122f61f0>
```



In [16]:

```
X.drop('Cabin',axis=1,inplace=True)
```

```
C:\Users\Lenovo\anaconda3\lib\site-packages\pandas\core\frame.py:3990: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
```

In [17]:

```
X
```

Out[17]:

	Pclass	Sex	Age	SibSp	Parch	Embarked
0	3	male	22.0	1	0	S
1	1	female	38.0	1	0	C
2	3	female	26.0	0	0	S
3	1	female	35.0	1	0	S
4	3	male	35.0	0	0	S
...	...	...	...	...	...	...
886	2	male	27.0	0	0	S
887	1	female	19.0	0	0	S
888	3	female	NaN	1	2	S
889	1	male	26.0	0	0	C
890	3	male	32.0	0	0	Q

891 rows × 6 columns

In [18]:

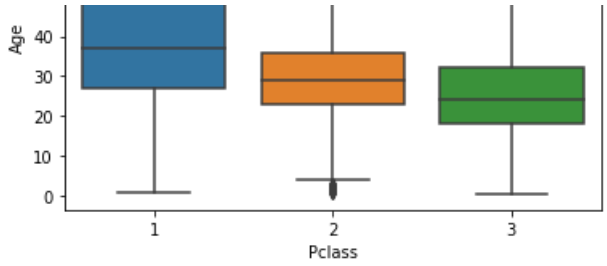
```
age= dataset['Age']
```

In [19]:

```
sns.boxplot(y='Age',x='Pclass',data=dataset)
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xc513296b50>
```



In [20]:

```
def titanic(cols):
    Pclass = cols[1]
    age = cols[0]
    if pd.isnull(age):
        if Pclass == 1:
            return 38
        elif Pclass == 2:
            return 30
        elif Pclass == 3:
            return 30
    else:
        return age
```

In [21]:

```
age= dataset[['Age','Pclass']].apply(titanic , axis=1)
```

In [22]:

```
age
```

Out[22]:

```
0       22.0
1       38.0
2       26.0
3       35.0
4       35.0
        ...
886     27.0
887     19.0
888     30.0
889     26.0
890     32.0
Length: 891, dtype: float64
```

In [23]:

```
X['Age']= age
```

```
<ipython-input-23-20daffa26a16>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  X['Age']= age
```

In [24]:

```
X
```

Out[24]:

	Pclass	Sex	Age	SibSp	Parch	Embarked
0	3	male	22.0	1	0	S

	Pclass	Sex	Age	SibSp	Parch	Embarked
1	1	female	38.0	1	0	C
2	3	female	26.0	0	0	S
3	1	female	35.0	1	0	S
4	3	male	35.0	0	0	S
...	...	...	...	...	...	...
886	2	male	27.0	0	0	S
887	1	female	19.0	0	0	S
888	3	female	30.0	1	2	S
889	1	male	26.0	0	0	C
890	3	male	32.0	0	0	Q

891 rows × 6 columns

In [25]:

```
sns.heatmap(X.isnull(),cmap='viridis')
```

Out[25]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xc50f570a60>
```