In [1]:

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

In [27]:

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

In [28]:

```
data.head()
```

Out[28]:

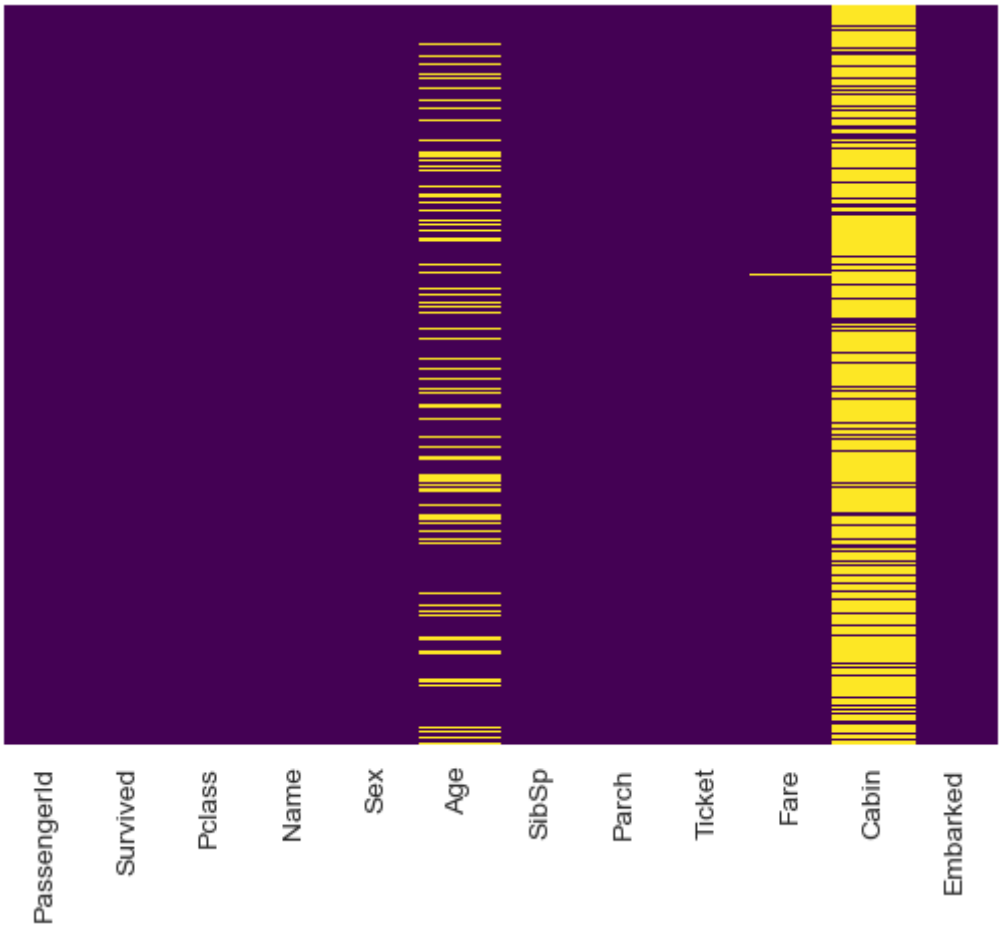
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154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

In [29]:

```
1 sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[29]:

<Axes: >

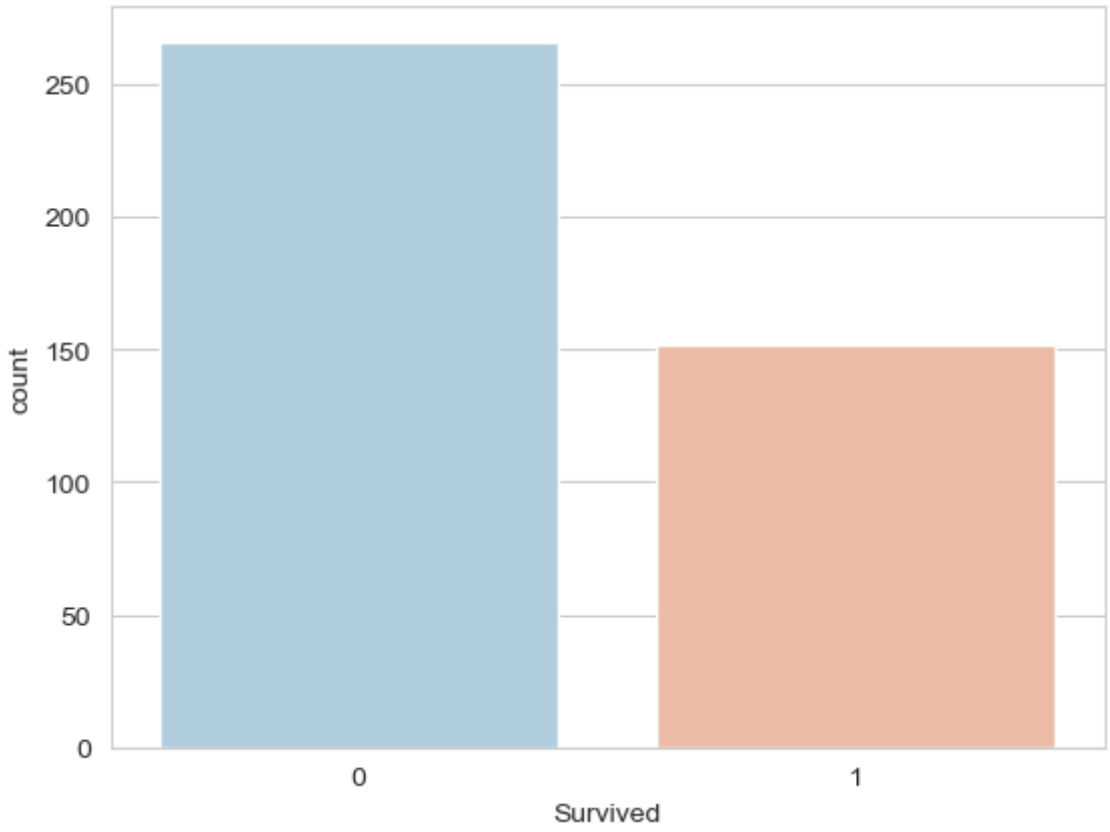


In [30]:

```
1 sns.set_style('whitegrid')
2 sns.countplot(x='Survived',data=data,palette='RdBu_r')
```

Out[30]:

```
<Axes: xlabel='Survived', ylabel='count'>
```

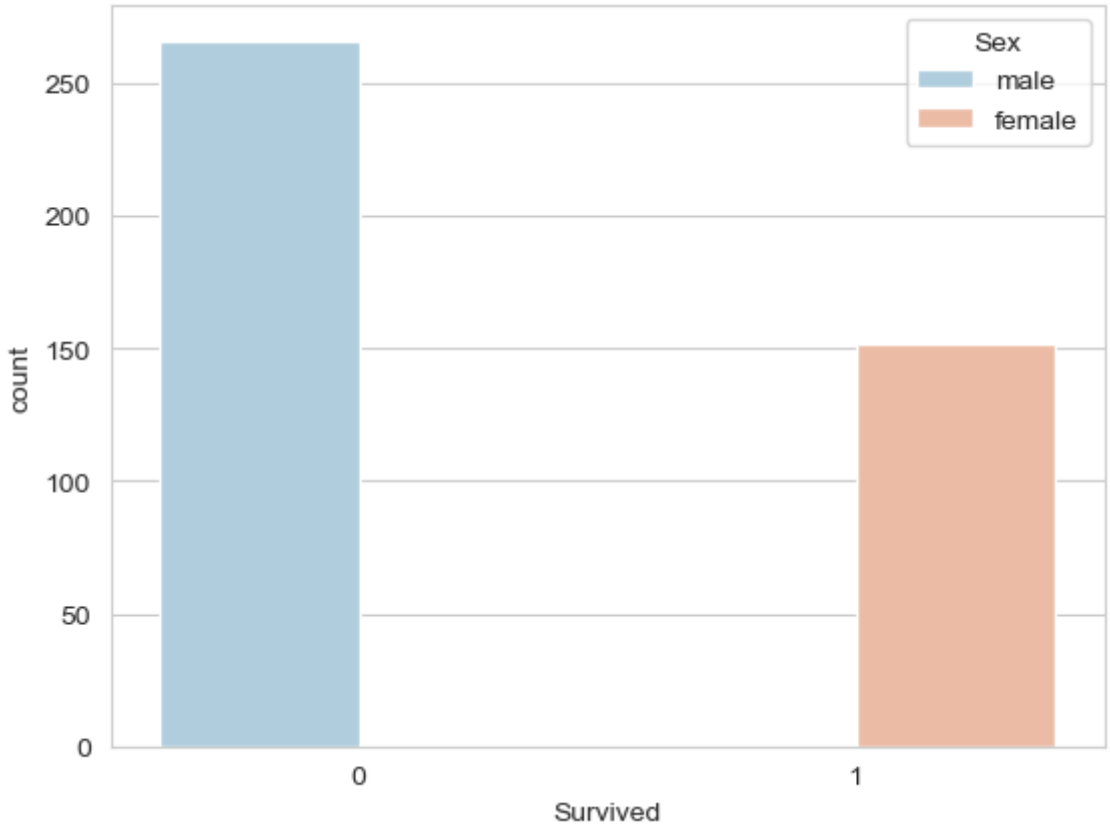


In [31]:

```
1 sns.countplot(x='Survived',hue='Sex',data=data,palette='RdBu_r')
```

Out[31]:

```
<Axes: xlabel='Survived', ylabel='count'>
```

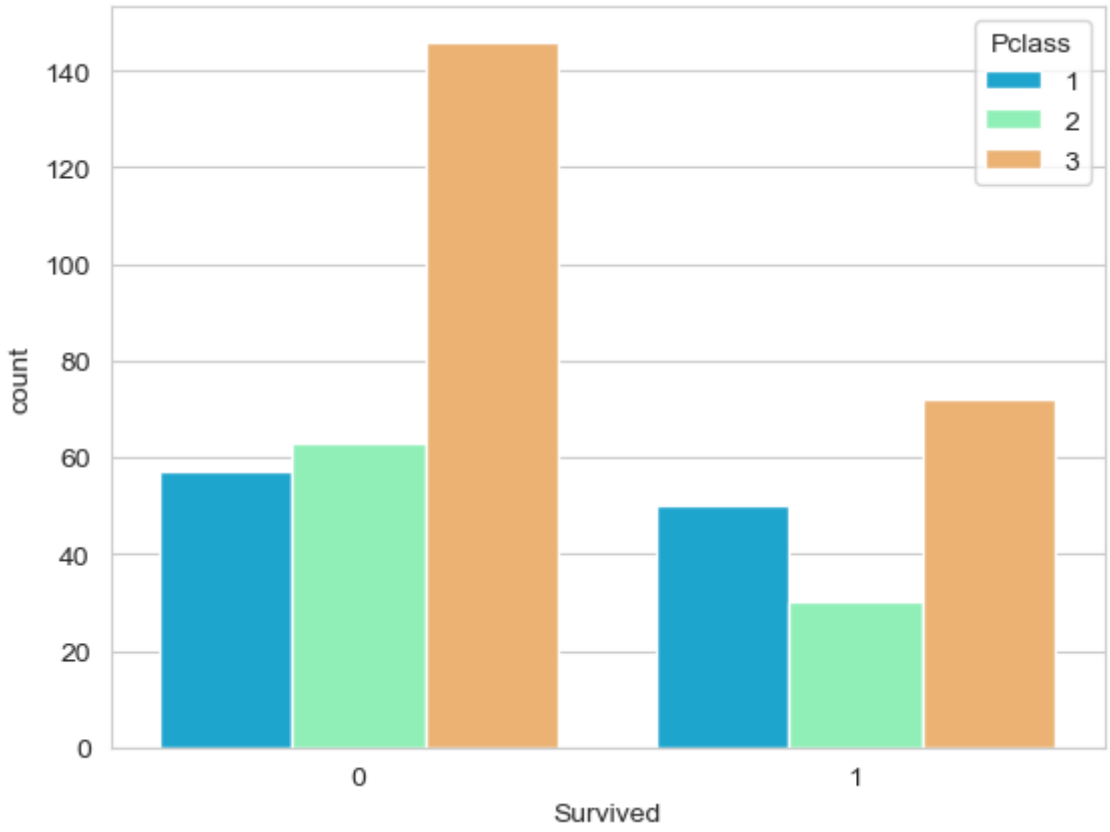


In [32]:

```
1 sns.set_style('whitegrid')
2 sns.countplot(x='Survived',hue='Pclass',data=data,palette='rainbow')
```

Out[32]:

```
<Axes: xlabel='Survived', ylabel='count'>
```



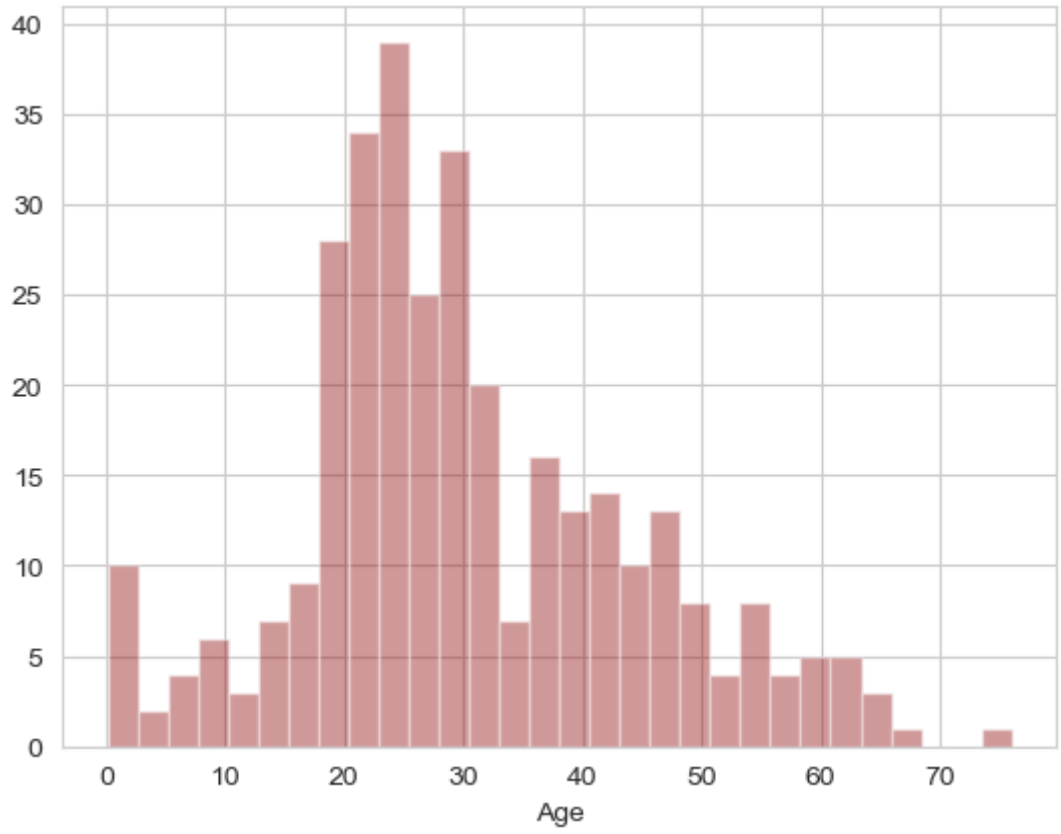
In [33]:

```
1 sns.distplot(data['Age'].dropna(),kde=False,color='darkred',bins=30)
```

```
C:\Users\suraj\AppData\Local\Temp\ipykernel_10448\4202982456.py:1: UserWa
rning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.
0.

Please adapt your code to use either `displot` (a figure-level function w
ith
similar flexibility) or `histplot` (an axes-level function for histogram
s).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (http
s://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot(data['Age'].dropna(),kde=False,color='darkred',bins=30)
```

Out[33]:

```
<Axes: xlabel='Age'>
```

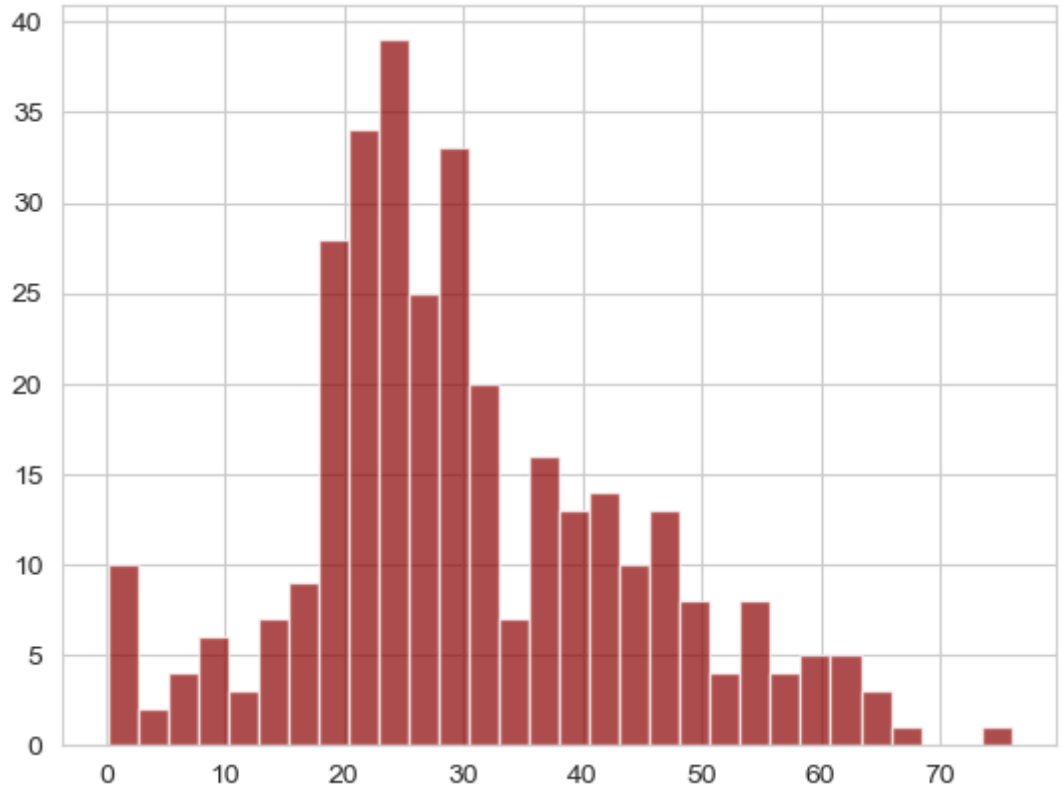


In [34]:

```
1 data['Age'].hist(bins=30,color='darkred',alpha=0.7)
```

Out[34]:

<Axes: >

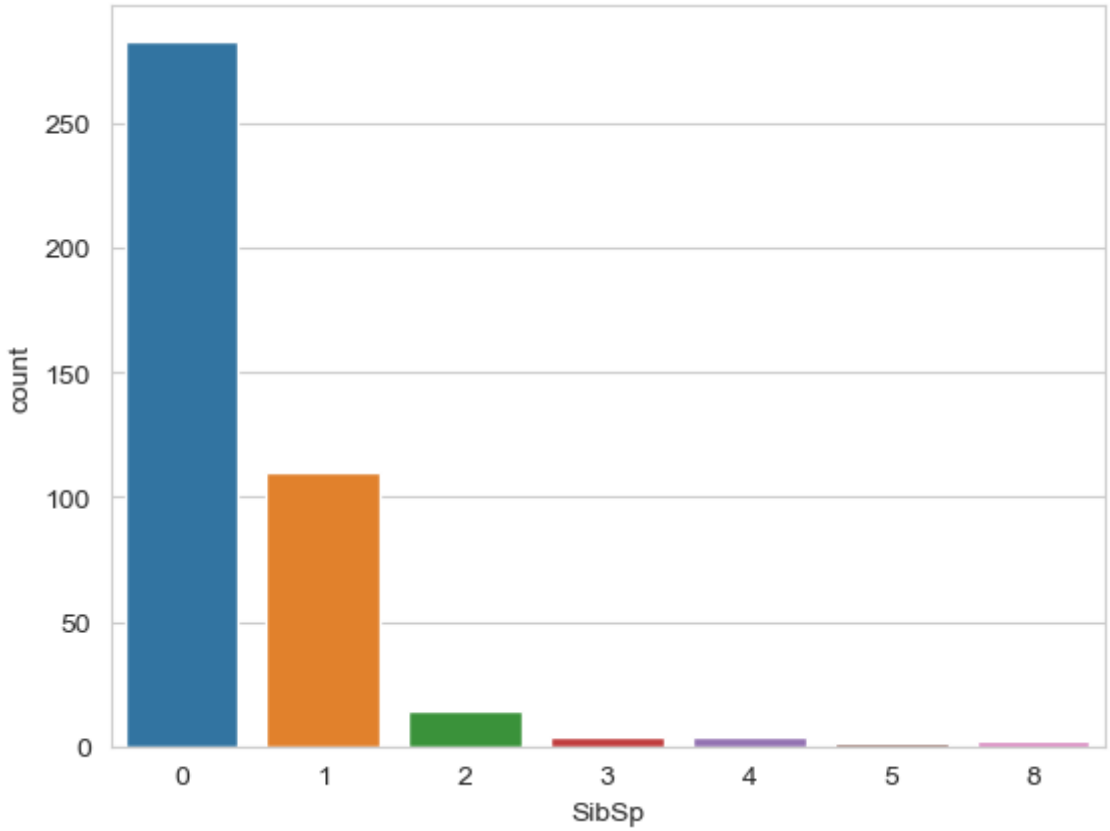


In [35]:

```
1 sns.countplot(x='SibSp',data=data)
```

Out[35]:

```
<Axes: xlabel='SibSp', ylabel='count'>
```

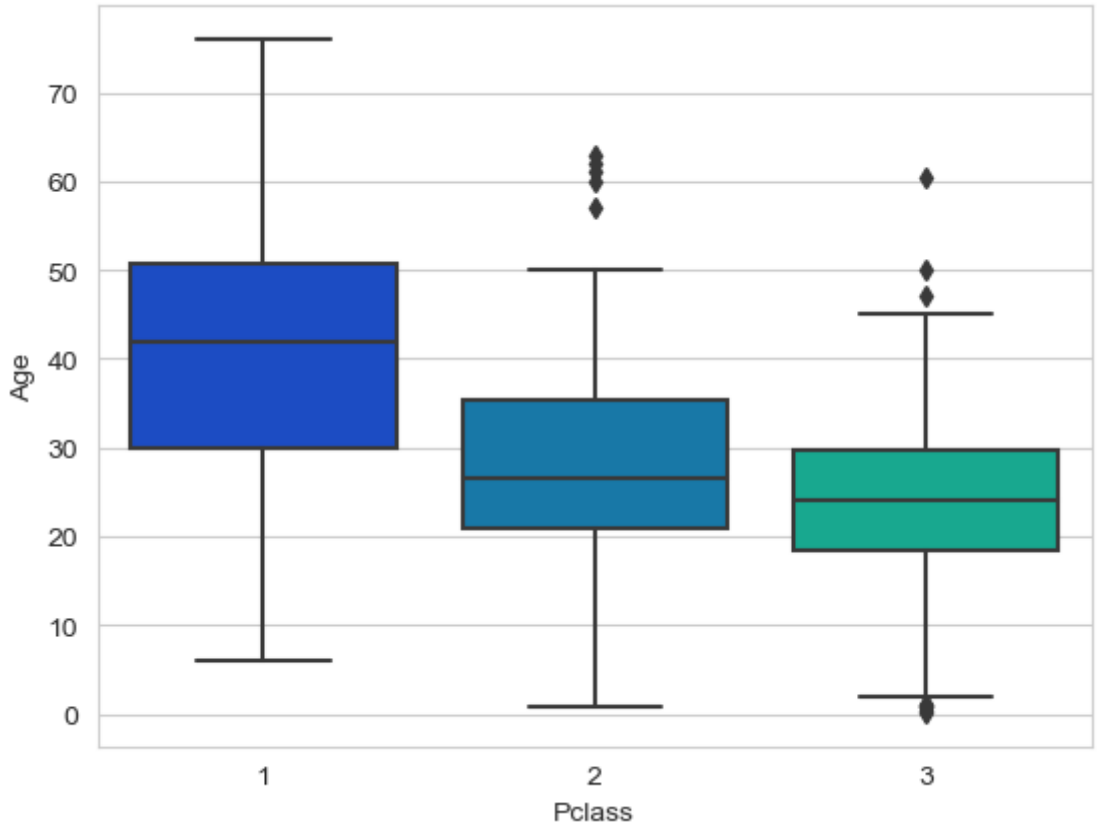


In [36]:

```
1 sns.boxplot(x='Pclass',y='Age',data=data,palette='winter')
```

Out[36]:

```
<Axes: xlabel='Pclass', ylabel='Age'>
```

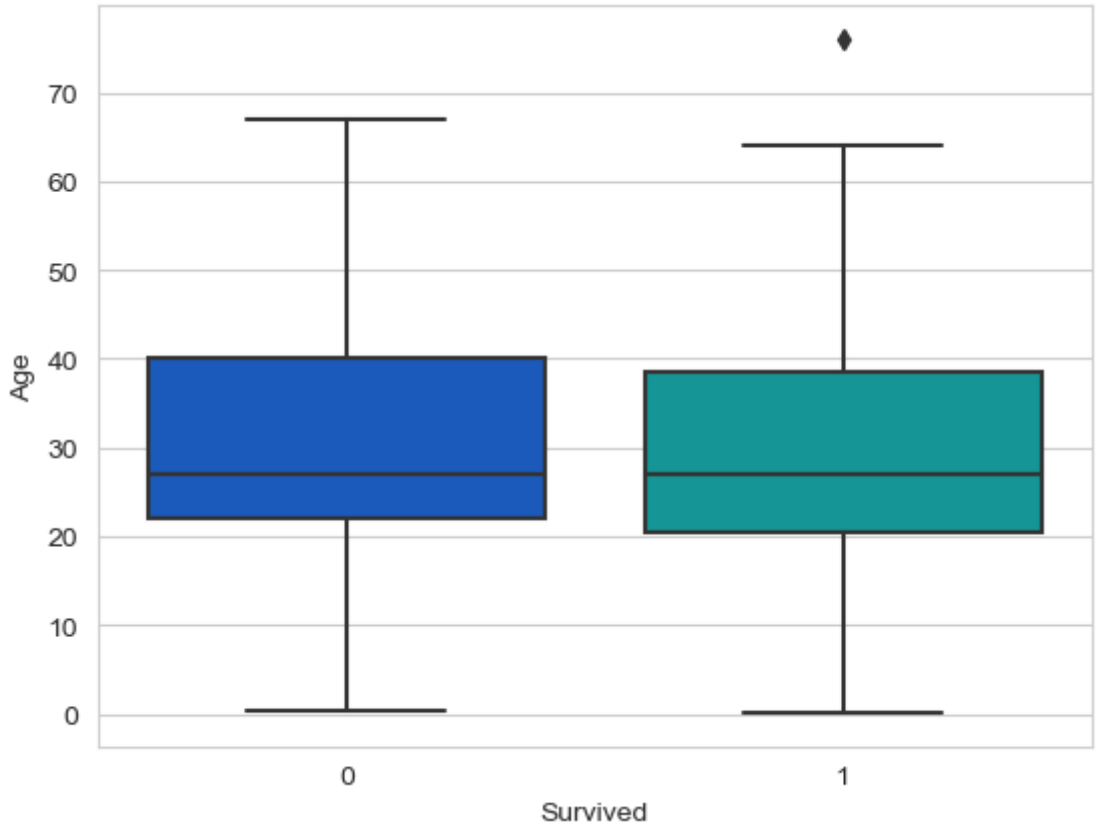


In [37]:

```
1 sns.boxplot(x='Survived',y='Age',data=data,palette='winter')
```

Out[37]:

```
<Axes: xlabel='Survived', ylabel='Age'>
```



In [38]:

```
1  def impute_age(cols):
2      Age = cols[0]
3      Pclass = cols[1]
4
5      if pd.isnull(Age):
6
7          if Pclass == 1:
8              return 37
9
10         elif Pclass == 2:
11             return 29
12
13         else:
14             return 24
15
16     else:
17         return Age
```

In [39]:

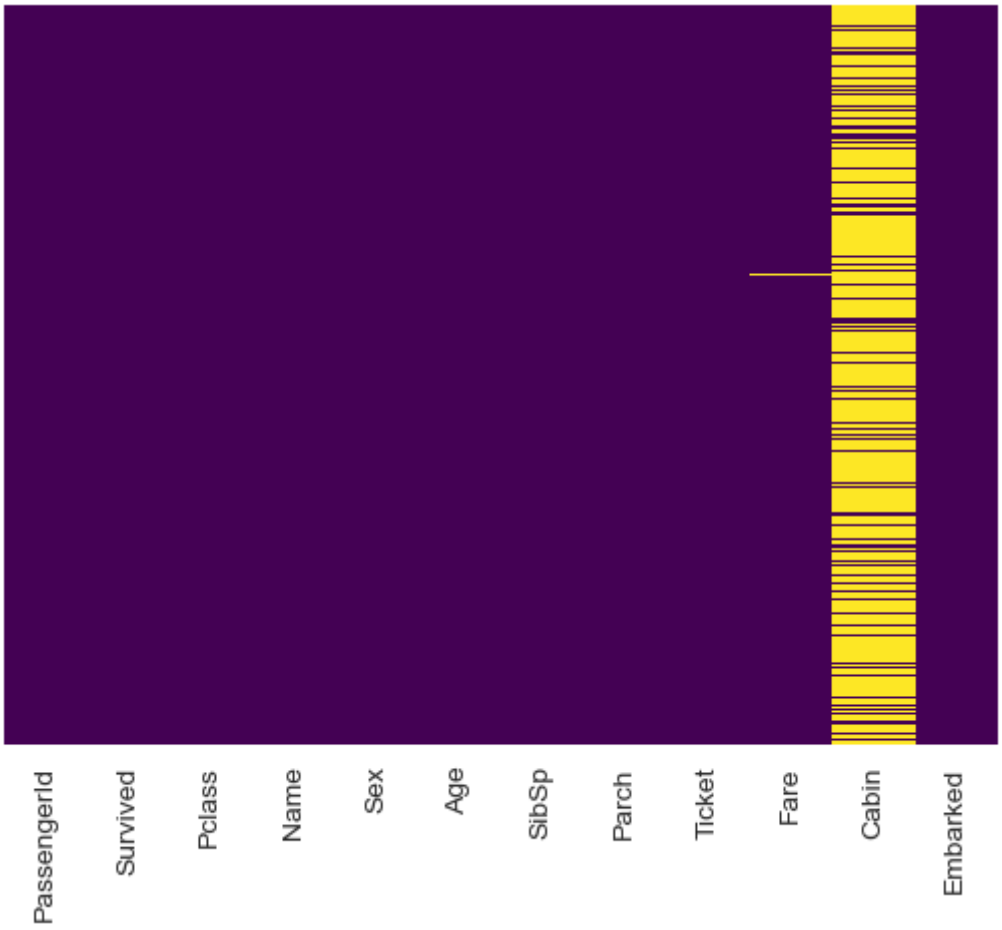
```
1 data['Age'] = data[['Age','Pclass']].apply(impute_age,axis=1)
```

In [40]:

```
1 sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[40]:

<Axes: >



In [41]:

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

In [42]:

```
1 data.head()
```

Out[42]:

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

In [43]:

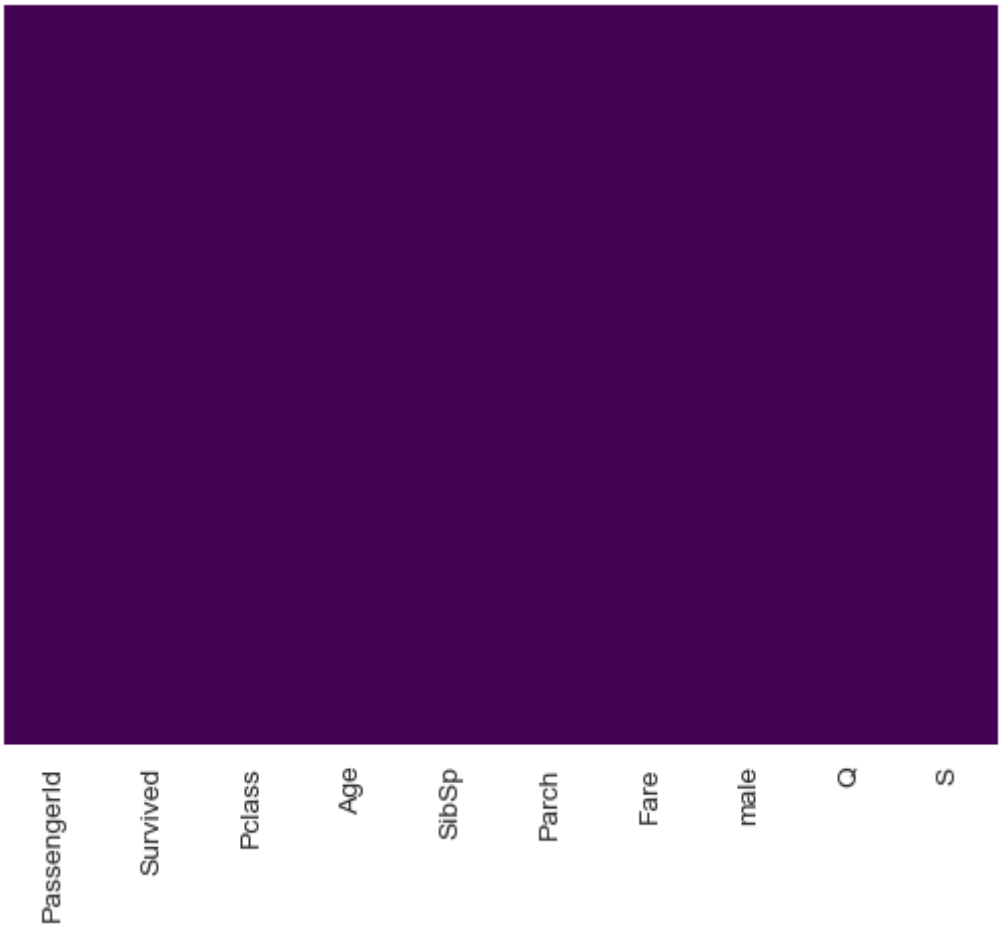
```
1 data.dropna(inplace=True)
```

In [61]:

```
1 sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[61]:

<Axes: >



In [44]:

```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 417 entries, 0 to 417
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  417 non-null    int64  
 1   Survived     417 non-null    int64  
 2   Pclass       417 non-null    int64  
 3   Name         417 non-null    object 
 4   Sex          417 non-null    object 
 5   Age          417 non-null    float64
 6   SibSp        417 non-null    int64  
 7   Parch        417 non-null    int64  
 8   Ticket       417 non-null    object 
 9   Fare         417 non-null    float64
 10  Embarked     417 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 39.1+ KB
```

In [45]:

```
sex = pd.get_dummies(data['Sex'],drop_first=True)
embark = pd.get_dummies(data['Embarked'],drop_first=True)
```

In [46]:

```
data.drop(['Sex','Embarked','Name','Ticket'],axis=1,inplace=True)
```

In [47]:

```
data = pd.concat([data,sex,embark],axis=1)
```

In [48]:

```
data.head()
```

Out[48]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	892	0	3	34.5	0	0	7.8292	1	1	0
1	893	1	3	47.0	1	0	7.0000	0	0	1
2	894	0	2	62.0	0	0	9.6875	1	1	0
3	895	0	3	27.0	0	0	8.6625	1	0	1
4	896	1	3	22.0	1	1	12.2875	0	0	1

In [49]:

```
from sklearn.model_selection import train_test_split
```

In [51]:

```
X = train.drop('Survived',axis=1)
y = train['Survived']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_sta
```

In [52]:

```
from sklearn.linear_model import LinearRegression
```

In [53]:

```
lm = LinearRegression()
```

In [54]:

```
1 lm.fit(X_train,y_train)
```

Out[54]:

```
LinearRegression
LinearRegression()
```

In [55]:

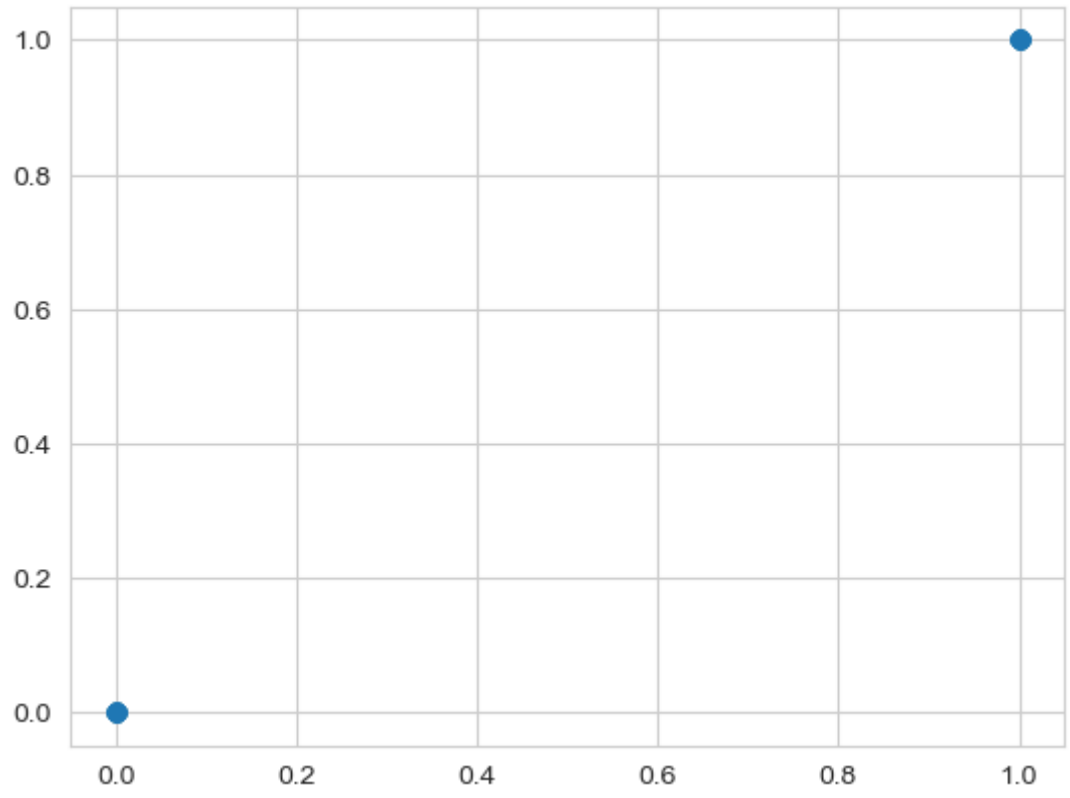
```
1 predictions = lm.predict(X_test)
```

In [56]:

```
1 plt.scatter(y_test,predictions)
```

Out[56]:

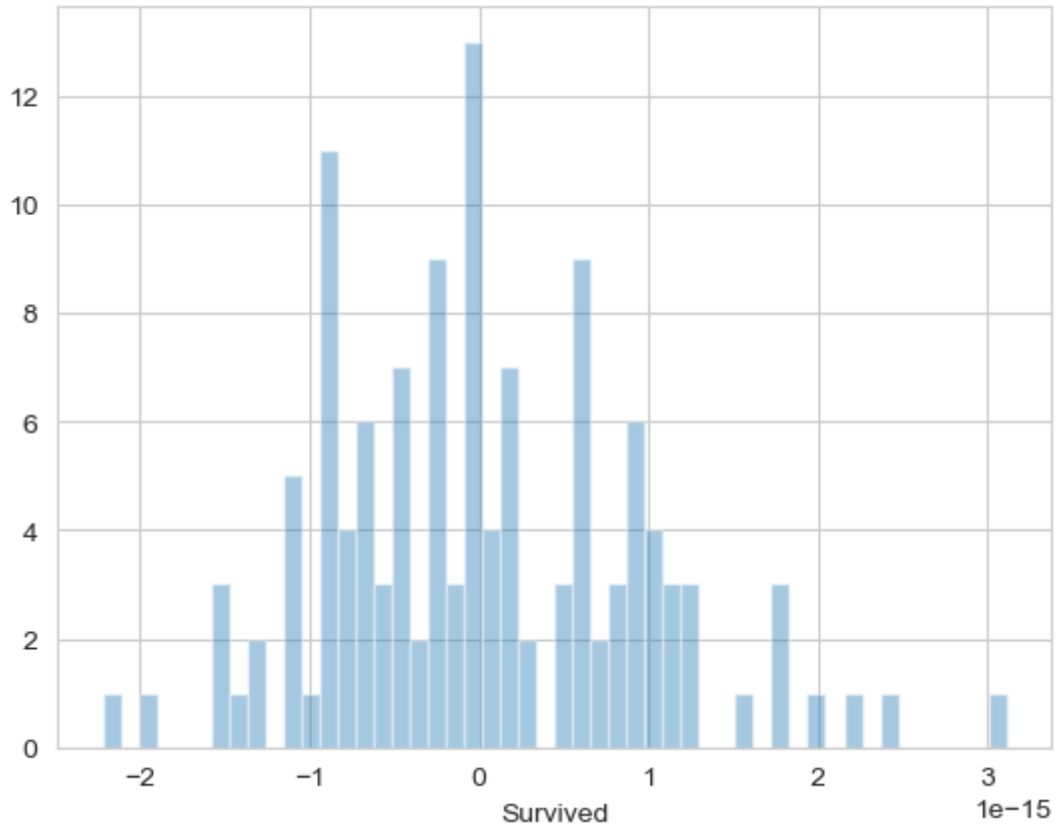
```
<matplotlib.collections.PathCollection at 0x229b2c18610>
```



In [58]:

```
1 sns.distplot((y_test-predictions),bins=50,kde=False);
```

```
C:\Users\suraj\AppData\Local\Temp\ipykernel_10448\3145007558.py:1: UserWa
rning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.
0.

Please adapt your code to use either `displot` (a figure-level function w
ith
similar flexibility) or `histplot` (an axes-level function for histogram
s).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (http
s://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot((y_test-predictions),bins=50,kde=False);
```



In [59]:

```
1 from sklearn import metrics
```

In [60]:

```
print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))

```

```
MAE: 7.128336721601204e-16
MSE: 8.392407869408556e-31
RMSE: 9.16100860681211e-16
```

In [ ]:

```

```