8

April 15, 2024

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

[3]:
```
pwd
```

[3]:
```
'/home/student'
```

[4]:
```
train = pd.read_csv('titanic_train.csv')
```

[5]:
```
train.head()
```

[5]:

	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

[6]:
```
train.isnull()
```

[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	\
0	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
5	False	False	False	False	False	True	False	False	False	
6	False	False	False	False	False	False	False	False	False	
7	False	False	False	False	False	False	False	False	False	
8	False	False	False	False	False	False	False	False	False	
9	False	False	False	False	False	False	False	False	False	
10	False	False	False	False	False	False	False	False	False	
11	False	False	False	False	False	False	False	False	False	
12	False	False	False	False	False	False	False	False	False	
13	False	False	False	False	False	False	False	False	False	
14	False	False	False	False	False	False	False	False	False	
15	False	False	False	False	False	False	False	False	False	
16	False	False	False	False	False	False	False	False	False	
17	False	False	False	False	False	True	False	False	False	
18	False	False	False	False	False	False	False	False	False	
19	False	False	False	False	False	True	False	False	False	
20	False	False	False	False	False	False	False	False	False	
21	False	False	False	False	False	False	False	False	False	
22	False	False	False	False	False	False	False	False	False	
23	False	False	False	False	False	False	False	False	False	
24	False	False	False	False	False	False	False	False	False	
25	False	False	False	False	False	False	False	False	False	
26	False	False	False	False	False	True	False	False	False	
27	False	False	False	False	False	False	False	False	False	
28	False	False	False	False	False	True	False	False	False	
29	False	False	False	False	False	True	False	False	False	
..	…	…	…	…	…	…	…	…		
861	False	False	False	False	False	False	False	False	False	
862	False	False	False	False	False	False	False	False	False	
863	False	False	False	False	False	True	False	False	False	
864	False	False	False	False	False	False	False	False	False	
865	False	False	False	False	False	False	False	False	False	
866	False	False	False	False	False	False	False	False	False	
867	False	False	False	False	False	False	False	False	False	
868	False	False	False	False	False	True	False	False	False	
869	False	False	False	False	False	False	False	False	False	
870	False	False	False	False	False	False	False	False	False	
871	False	False	False	False	False	False	False	False	False	
872	False	False	False	False	False	False	False	False	False	
873	False	False	False	False	False	False	False	False	False	
874	False	False	False	False	False	False	False	False	False	
875	False	False	False	False	False	False	False	False	False	

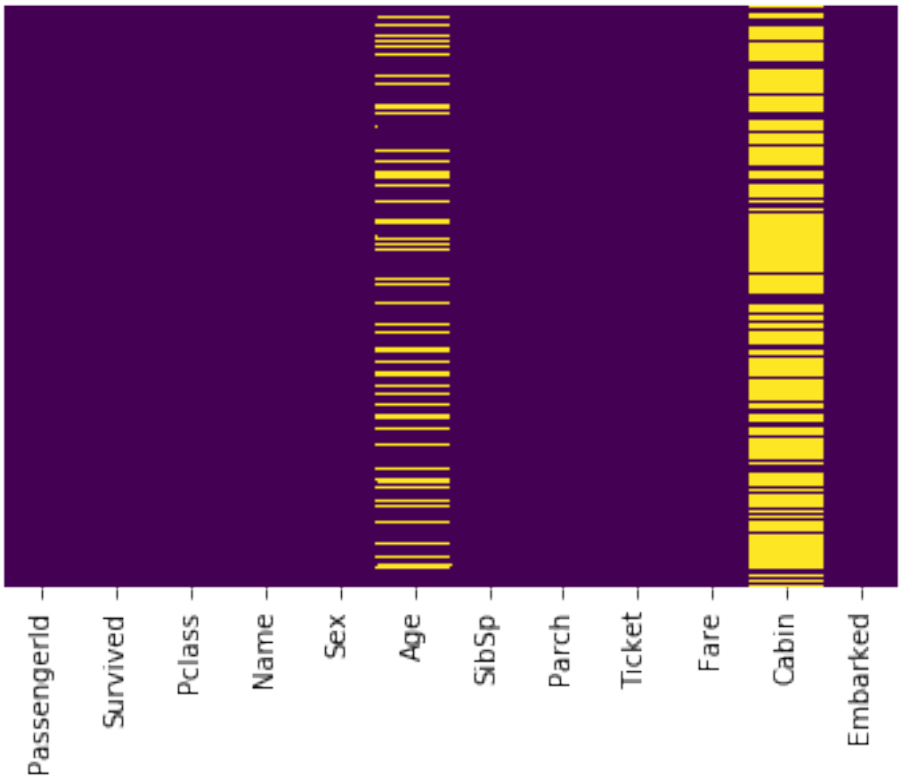
876	False	False	False	False	False	False	False	False	False
877	False	False	False	False	False	False	False	False	False
878	False	False	False	False	False	True	False	False	False
879	False	False	False	False	False	False	False	False	False
880	False	False	False	False	False	False	False	False	False
881	False	False	False	False	False	False	False	False	False
882	False	False	False	False	False	False	False	False	False
883	False	False	False	False	False	False	False	False	False
884	False	False	False	False	False	False	False	False	False
885	False	False	False	False	False	False	False	False	False
886	False	False	False	False	False	False	False	False	False
887	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False
889	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False

	Fare	Cabin	Embarked
0	False	True	False
1	False	False	False
2	False	True	False
3	False	False	False
4	False	True	False
5	False	True	False
6	False	False	False
7	False	True	False
8	False	True	False
9	False	True	False
10	False	False	False
11	False	False	False
12	False	True	False
13	False	True	False
14	False	True	False
15	False	True	False
16	False	True	False
17	False	True	False
18	False	True	False
19	False	True	False
20	False	True	False
21	False	False	False
22	False	True	False
23	False	False	False
24	False	True	False
25	False	True	False
26	False	True	False
27	False	False	False
28	False	True	False
29	False	True	False

```
..      …     …        …
861  False   True     False
862  False  False     False
863  False   True     False
864  False   True     False
865  False   True     False
866  False   True     False
867  False  False     False
868  False   True     False
869  False   True     False
870  False   True     False
871  False  False     False
872  False  False     False
873  False   True     False
874  False   True     False
875  False   True     False
876  False   True     False
877  False   True     False
878  False   True     False
879  False  False     False
880  False   True     False
881  False   True     False
882  False   True     False
883  False   True     False
884  False   True     False
885  False   True     False
886  False   True     False
887  False  False     False
888  False   True     False
889  False  False     False
890  False   True     False

[891 rows x 12 columns]
```

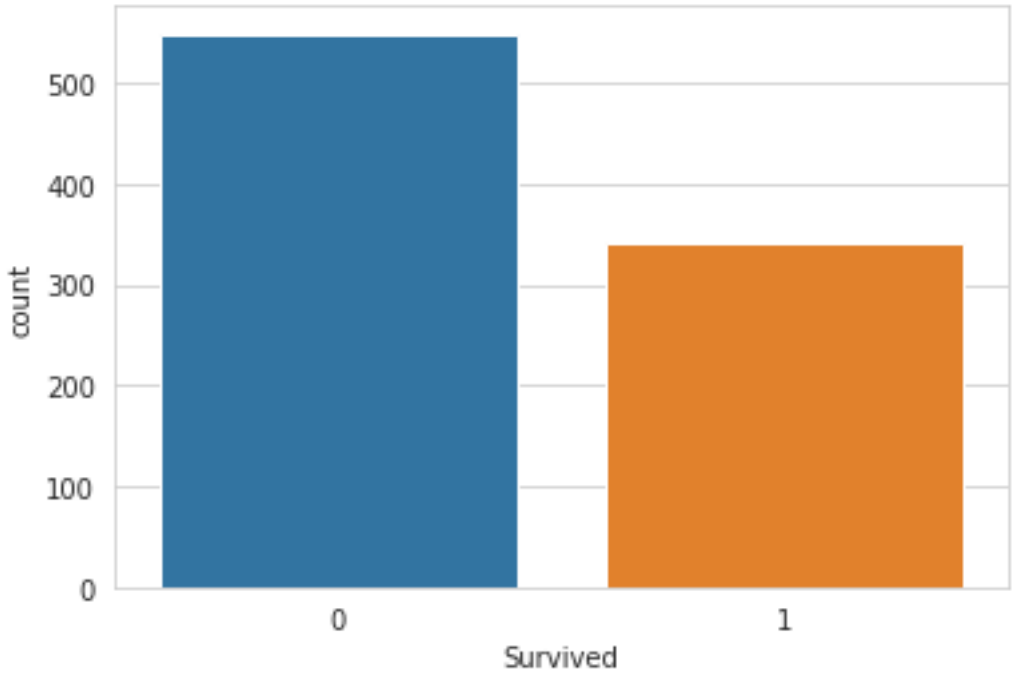
[7]:
```
sns.heatmap(train.isnull(),yticklabels = False, cbar = False, cmap = 'viridis')
```

[7]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa2f864a6d8>
```



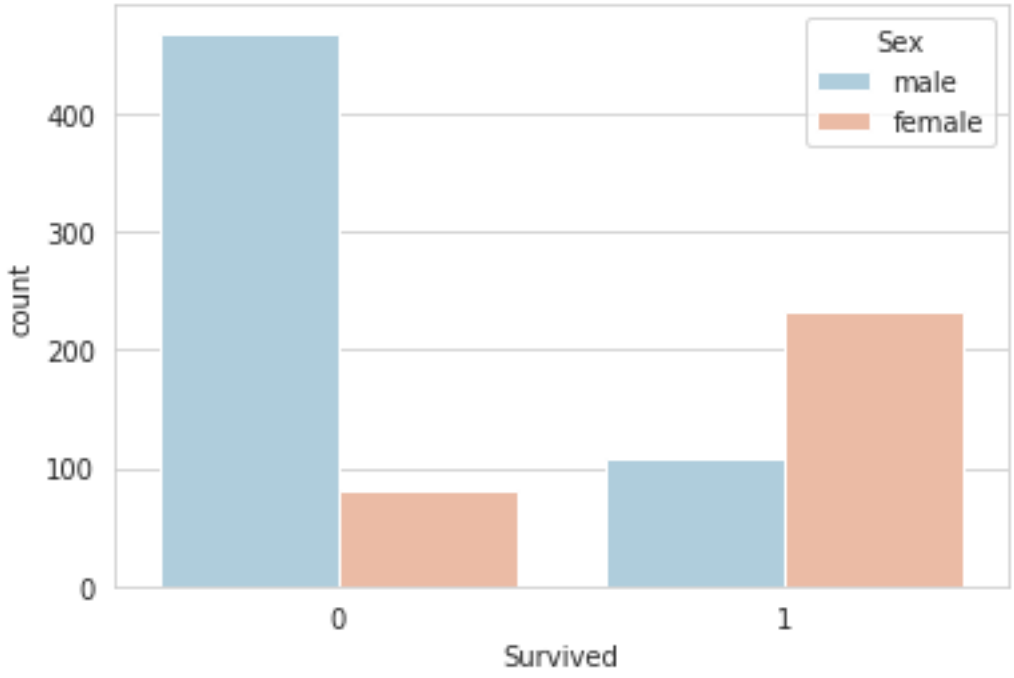
[8]:
```
sns.set_style('whitegrid')
sns.countplot(x = 'Survived', data = train)
```

[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f634ab00>



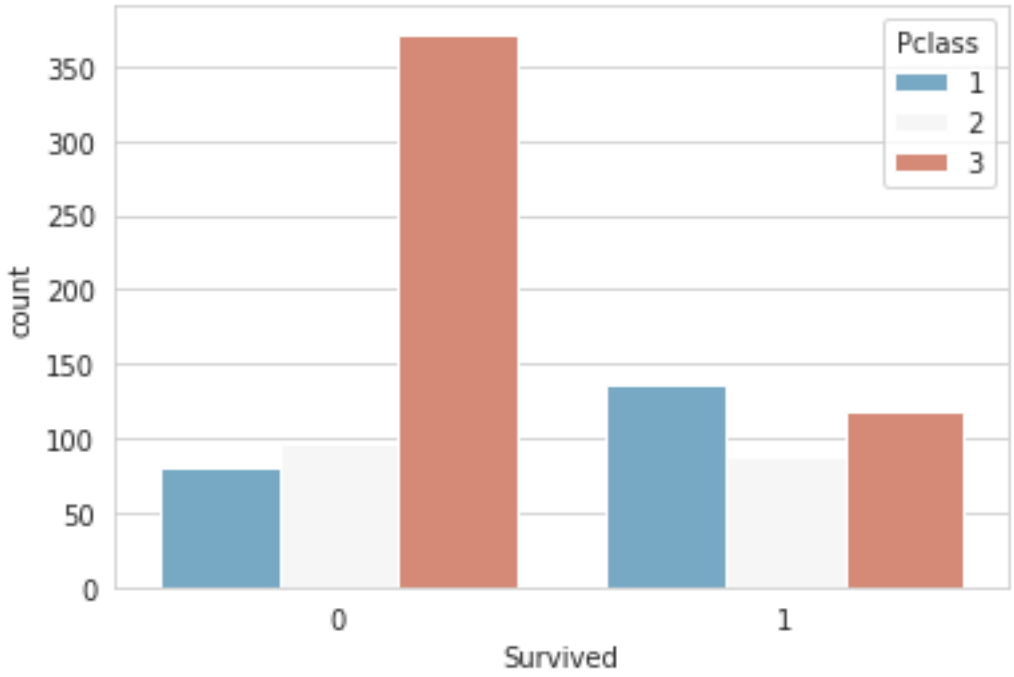
[9]:
```
sns.set_style('whitegrid')
sns.countplot(x = 'Survived', hue = 'Sex', data = train, palette = "RdBu_r" )
```

[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f62b7a58>



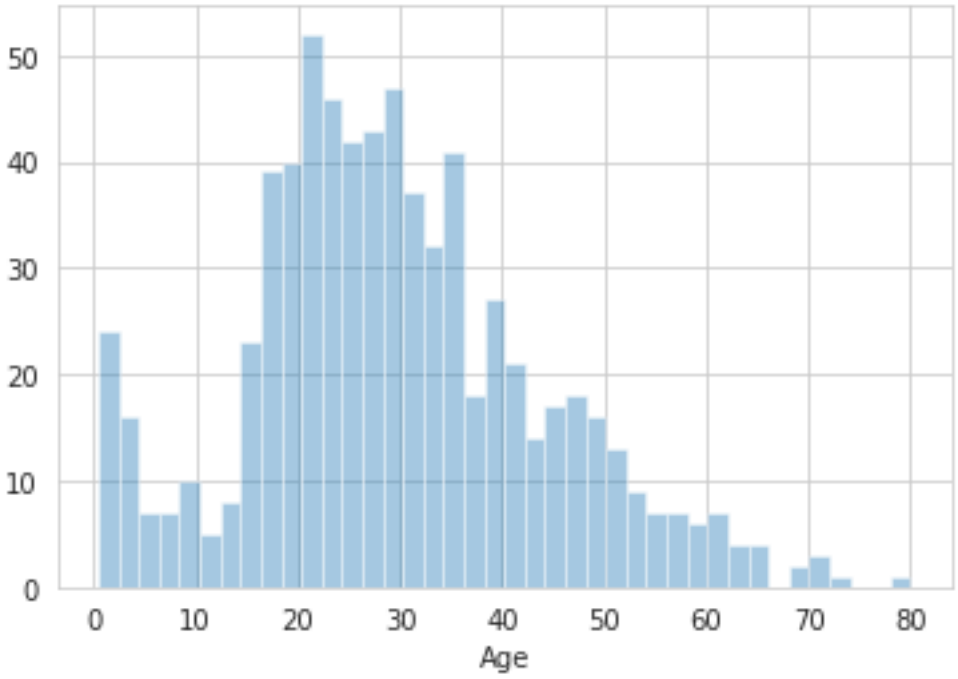
[10]:
```
sns.set_style('whitegrid')
sns.countplot(x = 'Survived', hue = 'Pclass', data = train, palette = "RdBu_r" )
```

[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f629c278>



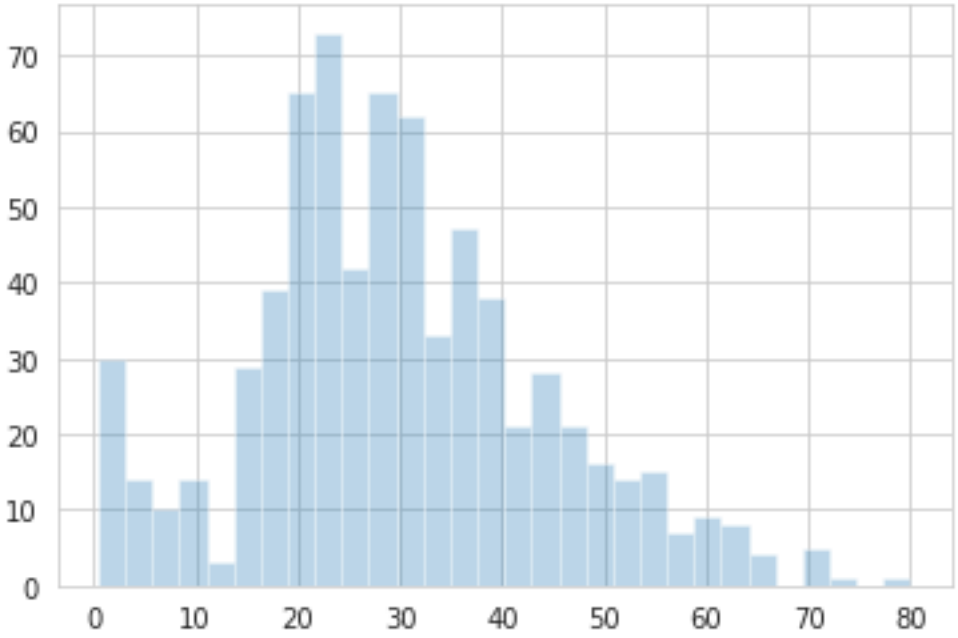
[11]:
```
sns.distplot(train["Age"].dropna(),kde = False, bins = 40)
```

[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f620f358>



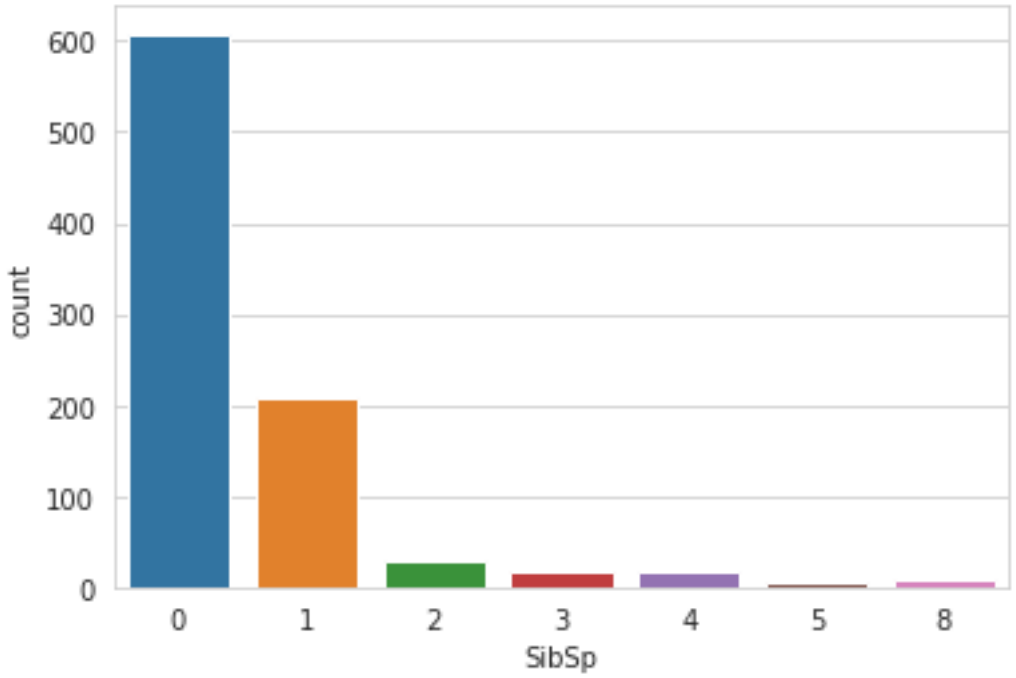
[12]:
```
train['Age'].hist( bins= 30, alpha = 0.3)
```

[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f60f6588>



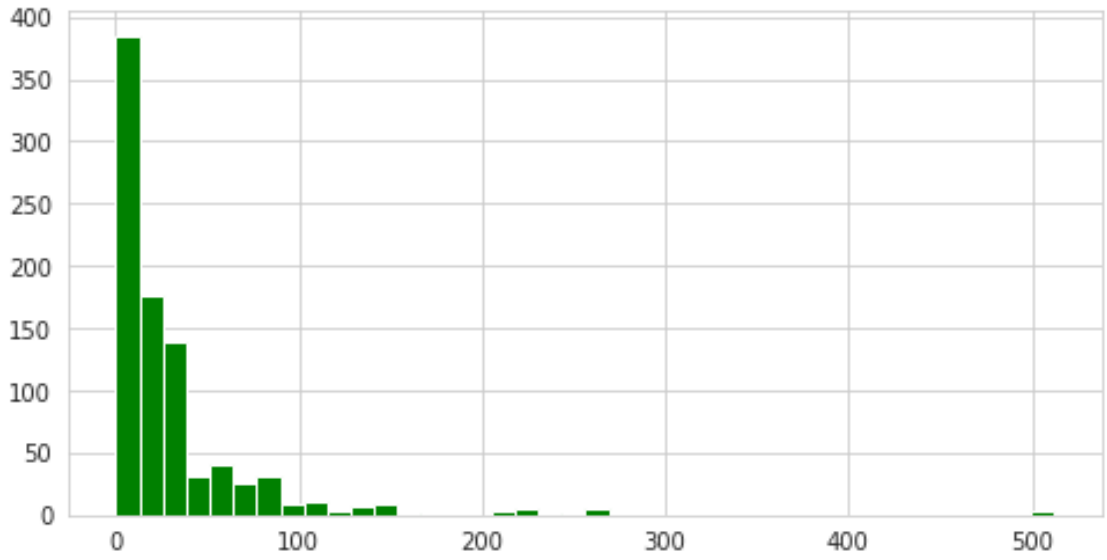
[13]:
```
sns.countplot(x ='SibSp' , data = train)
```

[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f607b358>



[14]:
```
train['Fare'].hist(color = 'green', bins = 40, figsize = (8, 4))
```

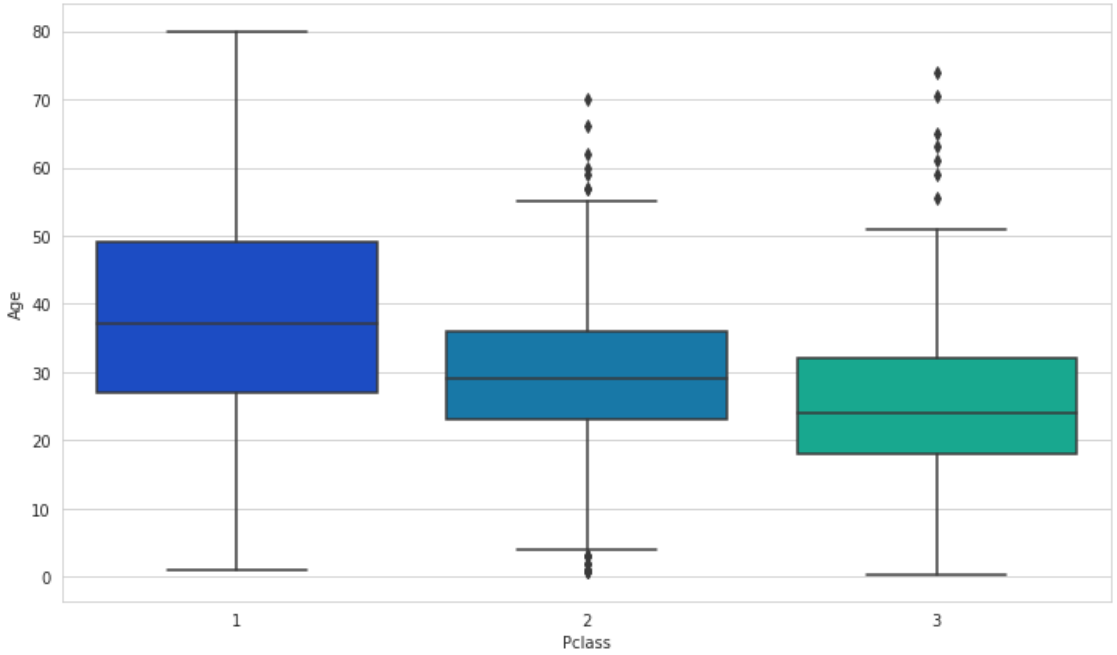
[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f6004080>



[15]:
```
#Data Cleaning
```

[16]:
```
plt.figure(figsize = (12, 7))
sns.boxplot(x = 'Pclass', y = 'Age', data = train, palette = 'winter')
```

[16]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa2f5f79828>
```



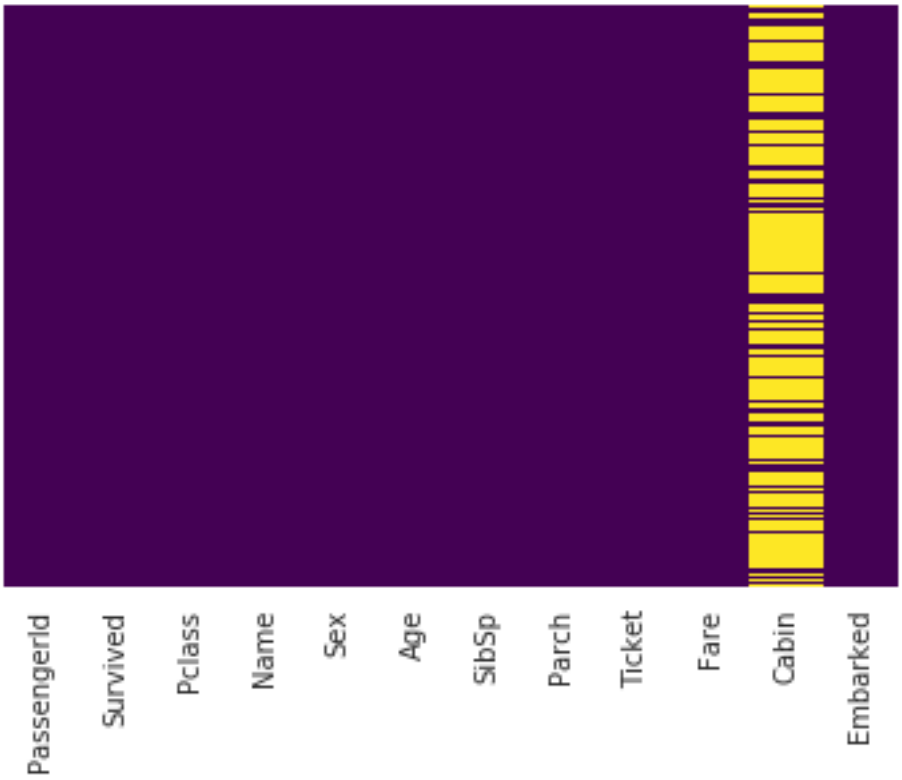
[17]:
```
def impute_age(cols):
    Age = cols[0]
    Pclass = cols[1]

    if pd.isnull(Age):

        if Pclass == 1:
            return 37

        elif Pclass == 2:
            return 29

        else:
            return 24

    else:
```

```
        return Age
```

[18]:
```
train['Age'] = train[['Age', 'Pclass']].apply(impute_age, axis = 1)
```

[19]:
```
sns.heatmap(train.isnull(), yticklabels = False, cbar = False, cmap = 'viridis')
```

[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2f5cc9ac8>



[20]:
```
train.drop('Cabin', axis = 1, inplace = True)
```

[21]:
```
train.head()
```

[21]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	\
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	1	
2	3	1	3					
3	4	1	1					
4	5	0	3					

```
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[22]:
```
train.dropna(inplace = True)
```

[23]:
```
train.head()
```

[23]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[24]:
```
pd.get_dummies(train ["Embarked"], drop_first = True).head()
```

[24]:
```
   Q  S
0  0  1
1  0  0
2  0  1
3  0  1
4  0  1
```

[25]:
```
sex = pd.get_dummies(train ["Sex"], drop_first = True)
embark = pd.get_dummies(train['Embarked'], drop_first = True)
```

[26]:
```
train.head()
```

[26]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[27]:
```
train.head()
```

[27]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[28]:
```
train.drop(['Sex', 'Embarked', 'Name', 'Ticket'], axis = 1, inplace = True)
```

[29]:
```
train.head()
```

[29]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
0	1	0	3	22.0	1	0	7.2500
1	2	1	1	38.0	1	0	71.2833
2	3	1	3	26.0	0	0	7.9250
3	4	1	1	35.0	1	0	53.1000
4	5	0	3	35.0	0	0	8.0500

[30]:
```
train = pd.concat([train, sex, embark], axis = 1)
```

[31]:
```
train.head()
```

[31]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1	0	7.2500	1	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0
2	3	1	3	26.0	0	0	7.9250	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	1
4	5	0	3	35.0	0	0	8.0500	1	0	1

[32]:
```
train.drop('Survived', axis = 1).head()
```

[32]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	3	22.0	1	0	7.2500	1	0	1
1	2	1	38.0	1	0	71.2833	0	0	0
2	3	3	26.0	0	0	7.9250	0	0	1
3	4	1	35.0	1	0	53.1000	0	0	1
4	5	3	35.0	0	0	8.0500	1	0	1

[33]:
```
train['Survived'].head()
```

[33]:
```
0    0
1    1
2    1
3    1
4    0
Name: Survived, dtype: int64
```

[34]:
```
from sklearn.model_selection import train_test_split
```

[35]:
```
X_train, X_test, y_train, y_test = train_test_split(train.drop('Survived', axis␣
↪= 1),
                                                    train['Survived'],␣
↪test_size = 0.3, random_state = 1)
```

[36]:
```
from sklearn.linear_model import LogisticRegression
```

[37]:
```
logmodel = LogisticRegression()
logmodel.fit(X_train, y_train)
```

```
/home/student/anaconda3/lib/python3.7/site-
packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver
will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

[37]:
```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='warn', n_jobs=None, penalty='l2',
                   random_state=None, solver='warn', tol=0.0001, verbose=0,
                   warm_start=False)
```

[38]:
```
predictions = logmodel.predict(X_test)
```

[39]:
```
from sklearn.metrics import confusion_matrix
```

[40]:
```
accuracy=confusion_matrix(y_test,predictions)
```

[41]:
```
accuracy
```

[41]:
```
array([[144,  22],
       [ 22,  79]])
```

[42]:
```
from sklearn.metrics import accuracy_score
```

[43]:
```
accuracy=accuracy_score(y_test,predictions)
accuracy
```

[43]:
```
0.8352059925093633
```

[44]:
```
predictions
```

[44]:
```
array([1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0,
       0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0,
       0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1,
       0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0,
       1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0,
       0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0,
       1, 0, 0])
```

[45]:
```
from sklearn.metrics import classification_report
```

[46]:
```
print(classification_report(y_test,predictions))
```

```
              precision    recall  f1-score   support

           0       0.87      0.87      0.87       166
           1       0.78      0.78      0.78       101

    accuracy                           0.84       267
   macro avg       0.82      0.82      0.82       267
weighted avg       0.84      0.84      0.84       267
```