

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

```
train=pd.read_csv('train.csv')
```

```
train.head()
```

	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 (Lilv Mav Peel)	female	35.0	1	0	113803	53.1000	C123	S

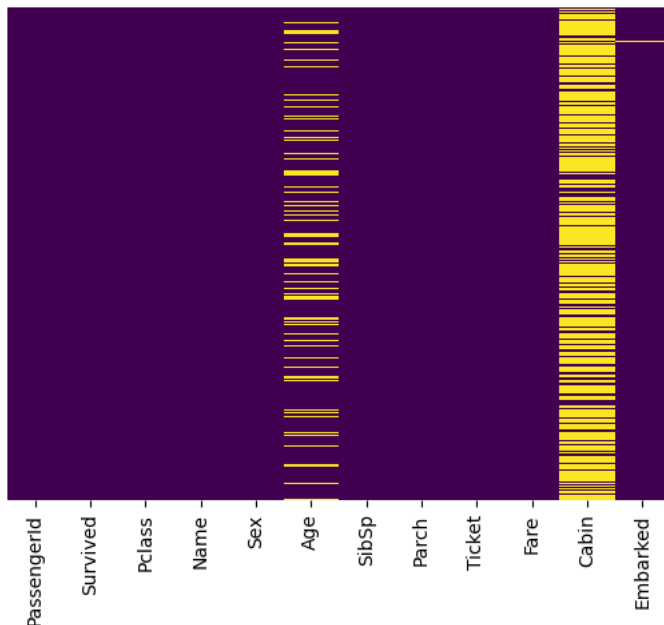
```
train.isnull().#Missing data
```

	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

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

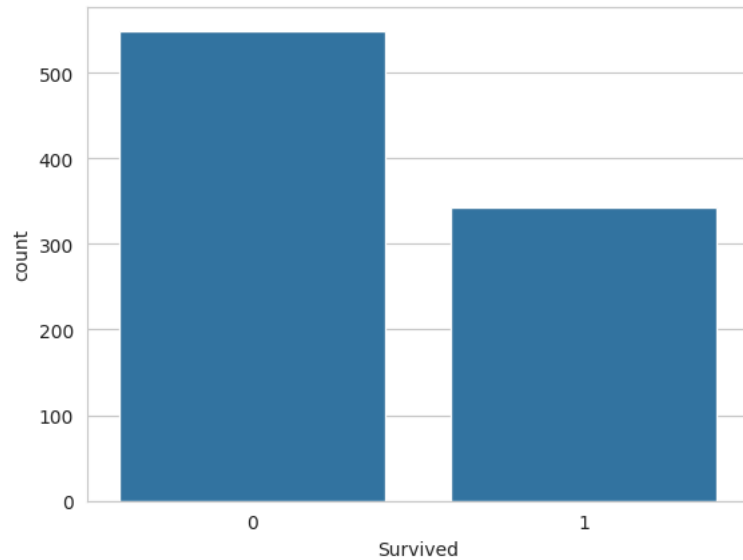
<Axes: >



```
sns.set_style('whitegrid')
sns.countplot(x='Survived',data=train)
```

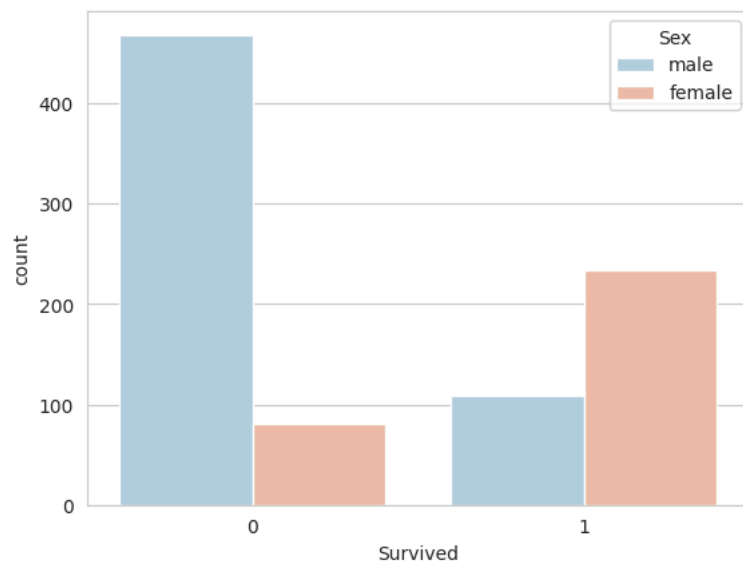


<Axes: xlabel='Survived', ylabel='count'>



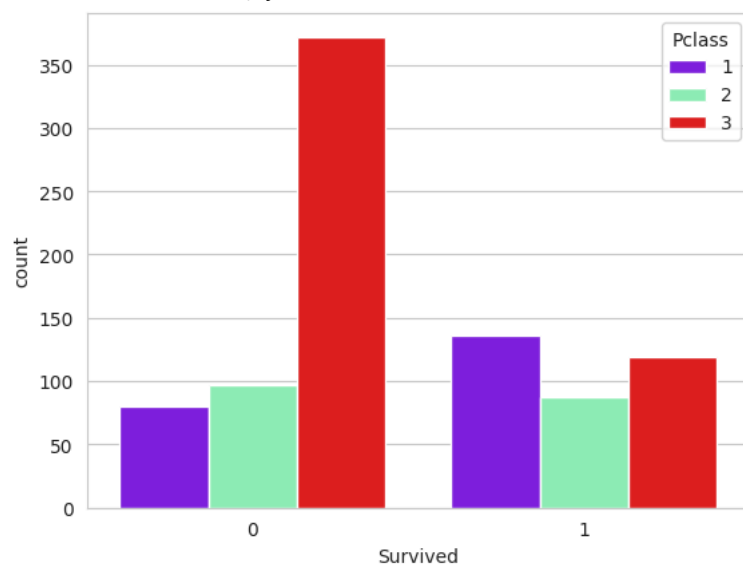
```
sns.set_style('whitegrid')  
sns.countplot(x='Survived',hue='Sex',data=train,palette='RdBu_r')
```

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



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

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



```
sns.distplot(train['Age'].dropna(),kde=False,color='darkred',bins=40)
```

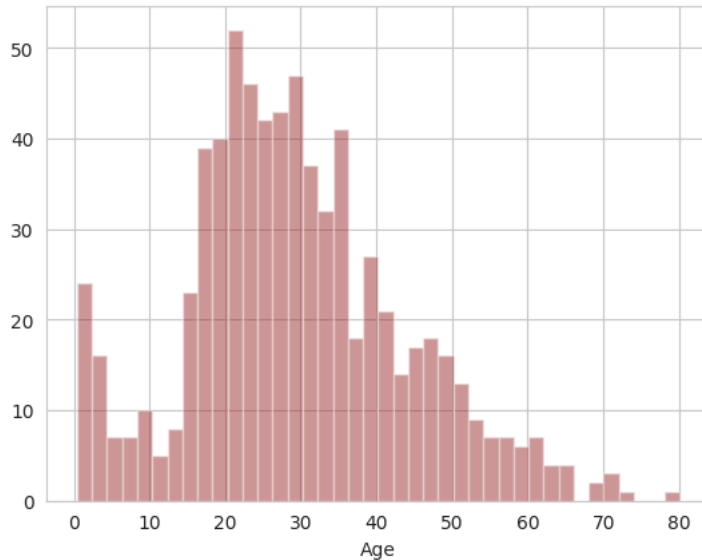
/tmp/ipython-input-2002818437.py:1: UserWarning:

`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 with similar flexibility) or `histplot` (an axes-level function for histograms).

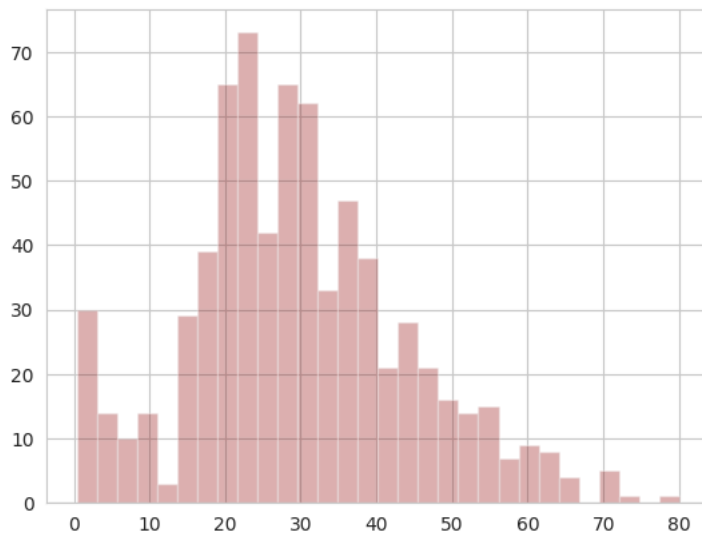
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(train['Age'].dropna(),kde=False,color='darkred',bins=40)
<Axes: xlabel='Age'>
```



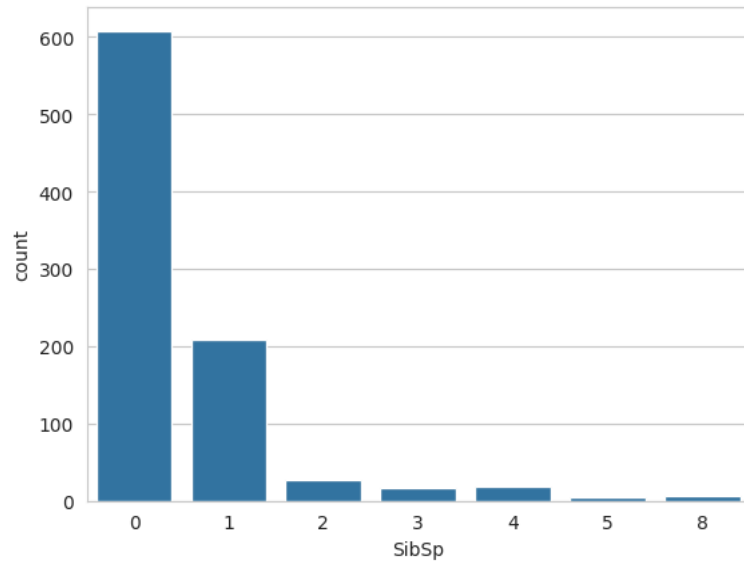
```
train['Age'].hist(bins=30,color='darkred',alpha=0.3)
```

<Axes: >



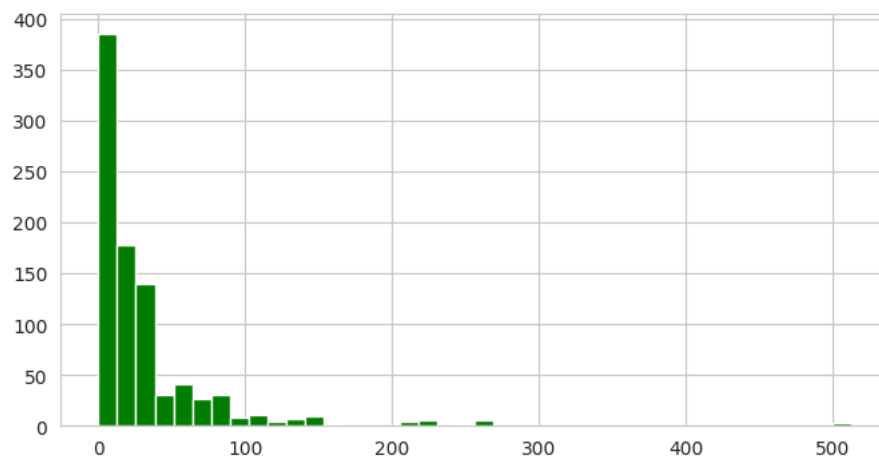
```
sns.countplot(x='SibSp',data=train)
```

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



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

<Axes: >

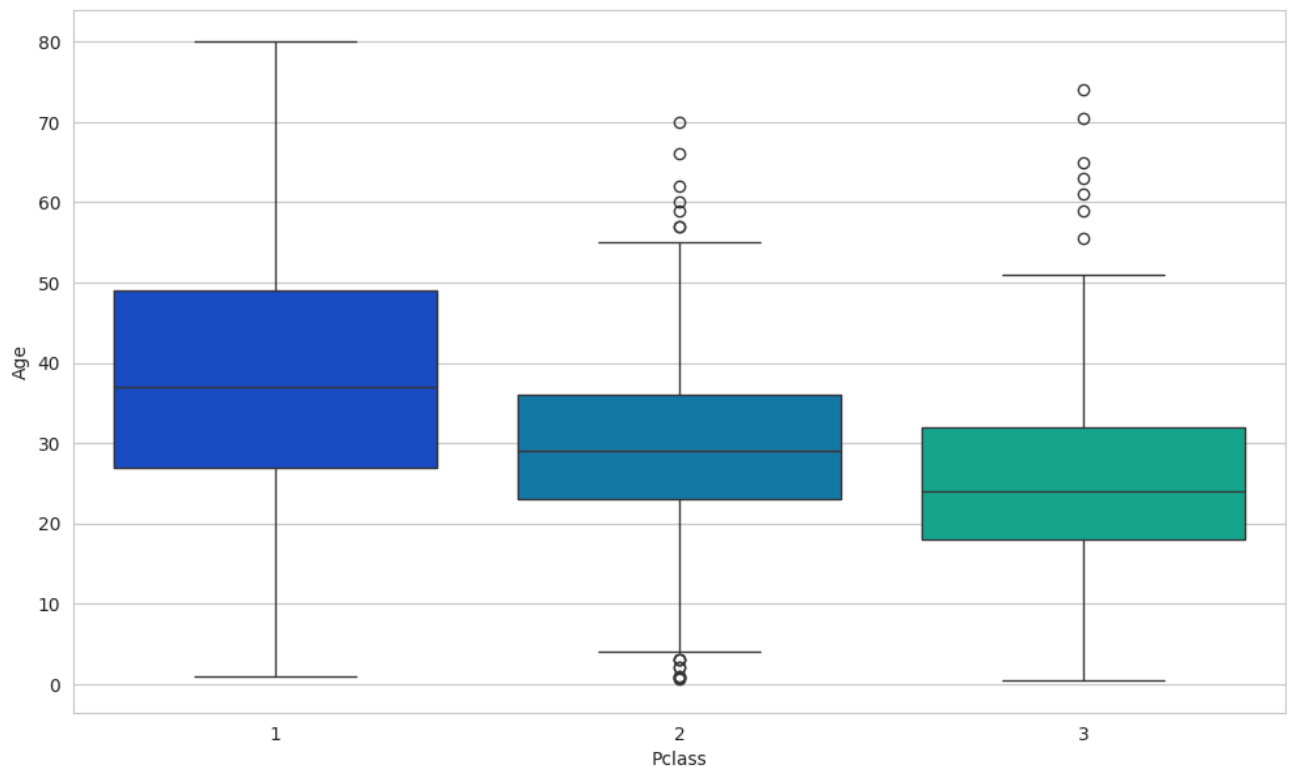


```
#Data Cleaning
plt.figure(figsize=(12,7))
sns.boxplot(x='Pclass',y='Age',data=train,palette='winter')
```

```
/tmp/ipython-input-1432412761.py:3: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and see

```
sns.boxplot(x='Pclass',y='Age',data=train,palette='winter')
<Axes: xlabel='Pclass', ylabel='Age'>
```



```
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
```

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

```
/tmp/ipython-input-3805301510.py:2: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future ver
Age=cols[0]
```

```
/tmp/ipython-input-3805301510.py:3: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future ver
Pclass=cols[1]
```

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

<Axes: >



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



```
train.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs), Mrs.	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S

```
train.dropna(inplace=True)
```

train

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

<Axes: >

