

Exploratory Data Analysis

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [2]: t = pd.read_csv(r'J:\cf\titanic.csv')
t
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8

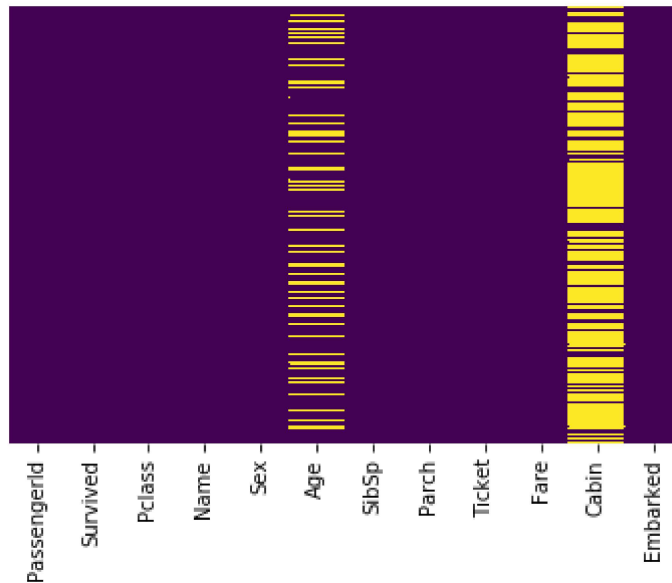
```
In [3]: t.isnull()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	
1	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	True	
3	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	True	
5	False	False	False	False	False	True	False	False	False	False	True	
6	False	False	False	False	False	False	False	False	False	False	False	
7	False	False	False	False	False	False	False	False	False	False	True	
8	False	False	False	False	False	False	False	False	False	False	True	
9	False	False	False	False	False	False	False	False	False	False	True	
10	False	False	False	False	False	False	False	False	False	False	False	

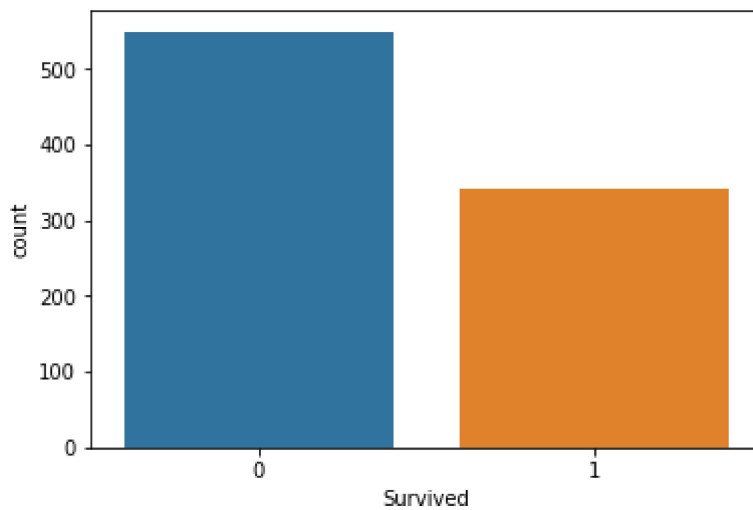
```
In [5]: sns.heatmap(t.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x295db34aac8>
```



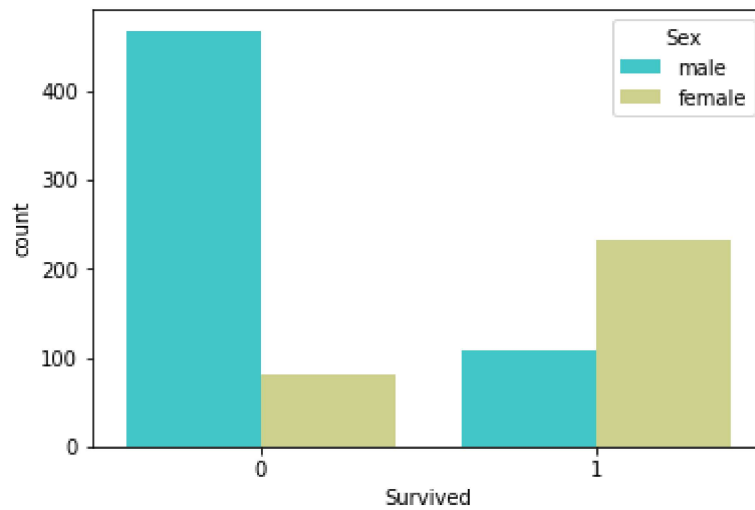
```
In [6]: sns.countplot(x='Survived',data=t)
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x295db39a7b8>
```



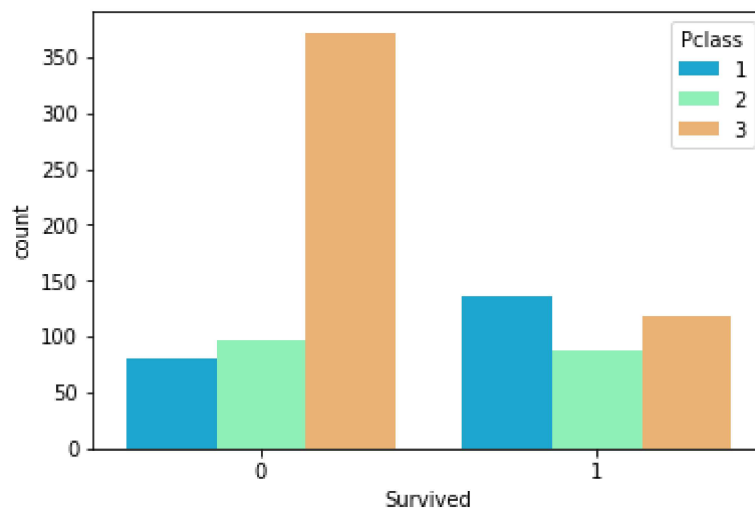
```
In [7]: sns.countplot(x='Survived',hue='Sex',data=t,palette='rainbow')
```

```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x295db3f8ba8>
```



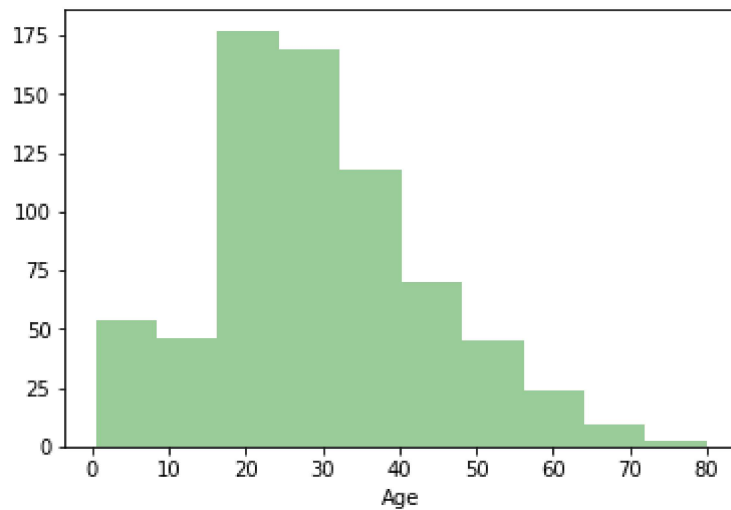
```
In [8]: sns.countplot(x='Survived',hue='Pclass',data=t,palette='rainbow')
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x295db45b7f0>
```



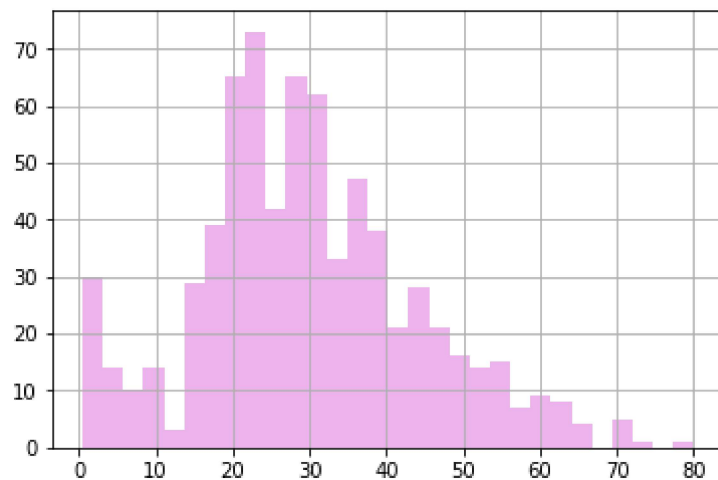
```
In [10]: sns.distplot(t['Age'].dropna(),kde=False,color='g',bins=10)
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x295db54be48>
```



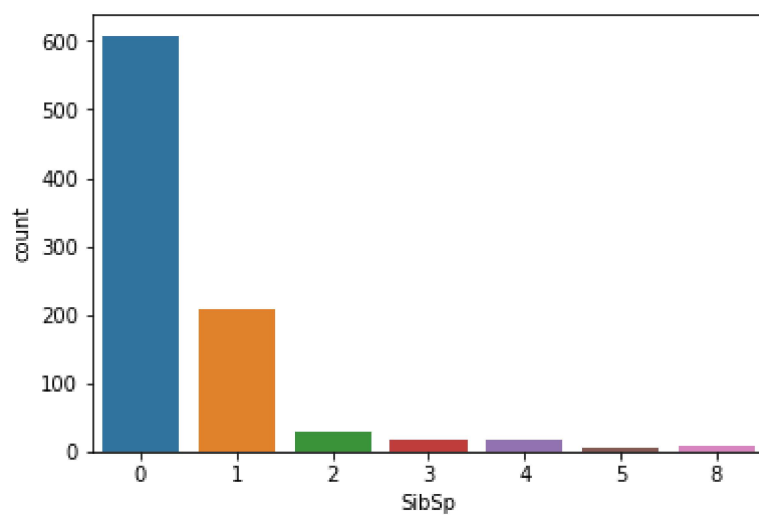
```
In [12]: t['Age'].hist(bins=30,color='m',alpha=0.3)
```

```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x295db6848d0>
```



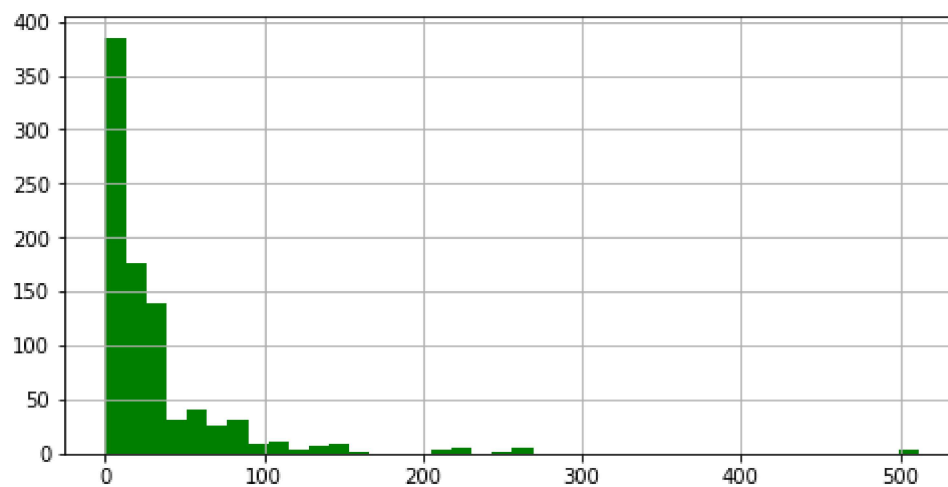
```
In [13]: sns.countplot(x='SibSp',data=t)
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x295db705d68>
```



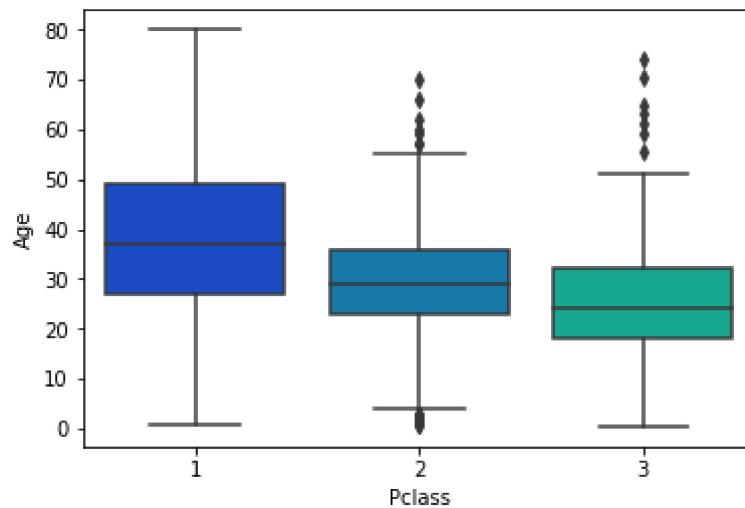
```
In [15]: t['Fare'].hist(color='g',bins=40,figsize=(8,4))
```

```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x295db80eba8>
```



```
In [16]: sns.boxplot(x='Pclass',y='Age',data=t,palette='winter')
```

```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x295db8d46a0>
```

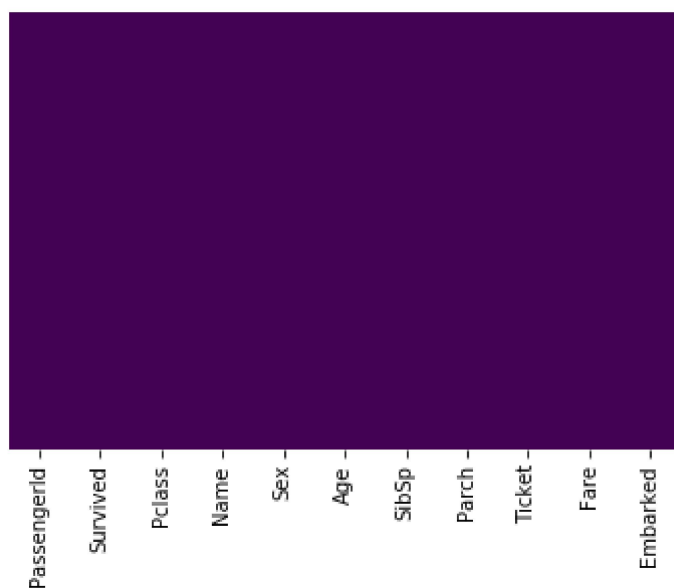


```
In [18]: def cha(cols):  
    Age=cols[0]  
    Pclass=cols[1]  
    if pd.isnull(Age):  
        if Pclass==1:  
            return 37  
        elif Pclass==2:  
            return 29  
        else:  
            return 24  
    else:  
        return Age
```

```
In [19]: t['Age'] = t[['Age','Pclass']].apply(cha,axis=1)
```

```
In [23]: sns.heatmap(t.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x295d9fb1c50>
```



```
In [21]: t.drop('Cabin',axis=1,inplace=True)
```

In [22]:

t

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8

In []: