

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

```
In [2]: data = pd.read_csv('train.csv')
data.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

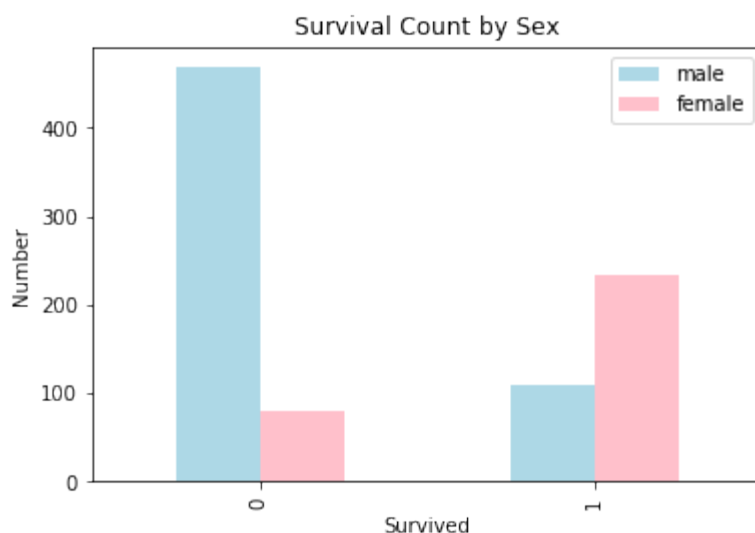
```
In [3]: data.describe()
```

Out[3]:

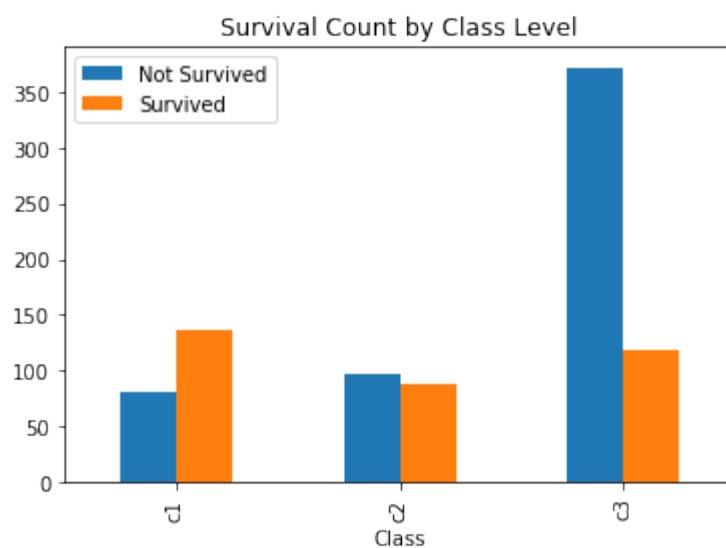
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
In [4]: female = data[data['Sex'] == 'female'].Survived.value_counts()
male = data[data['Sex'] == 'male'].Survived.value_counts()
df = pd.DataFrame({'male': male, 'female': female})
df.plot.bar(color=['lightblue', 'pink'])
plt.xlabel('Survived')
plt.ylabel('Number')
plt.title('Survival Count by Sex')
```

Out[4]: Text(0.5,1,'Survival Count by Sex')



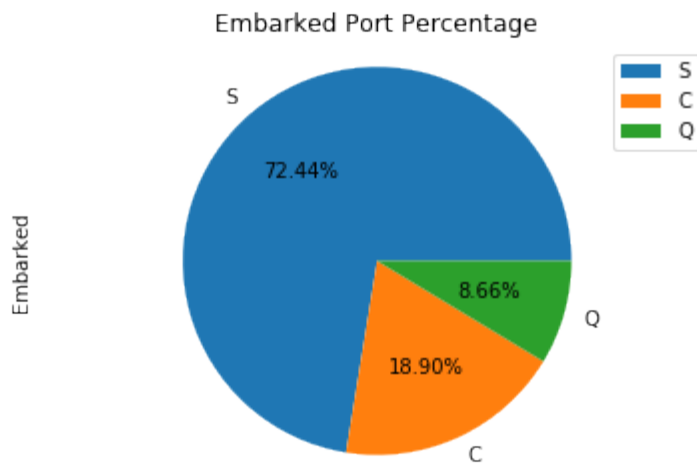
```
In [18]: class1 = data[data['Pclass'] == 1].Survived.value_counts()
class2 = data[data['Pclass'] == 2].Survived.value_counts()
class3 = data[data['Pclass'] == 3].Survived.value_counts()
df = pd.DataFrame({'c1': class1, 'c2': class2, 'c3': class3})
df.index = ['Not Survived', 'Survived']
df.T.plot.bar()
plt.title('Survival Count by Class Level')
plt.xlabel('Class')
plt.ylabel = 'Survival Count'
```



```
In [42]: embark = data['Embarked'].value_counts()
embark.plot.pie(autopct='%2f%%')
```

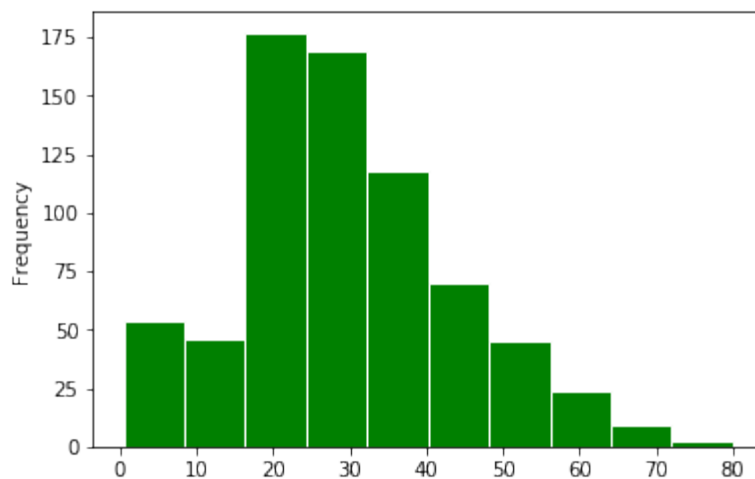
```
plt.title('Embarked Port Percentage')
plt.legend()
plt.axis('equal')
```

```
Out[42]: (-1.115244536548557,
          1.100725939222069,
          -1.1207980301925962,
          1.1184376185399005)
```



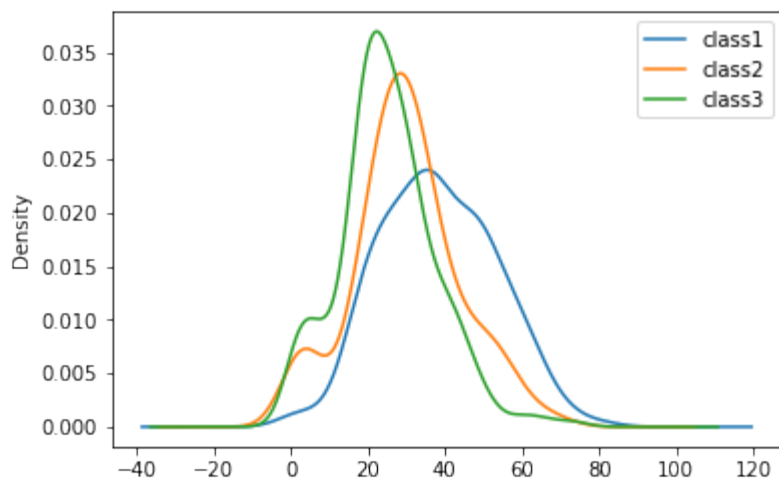
```
In [54]: data['Age'].plot(kind='hist', facecolor='green', edgecolor='white')
```

```
Out[54]: <matplotlib.axes._subplots.AxesSubplot at 0x2786d4bdb08>
```



```
In [78]: class1 = data[data['Pclass'] == 1].Age
          class2 = data[data['Pclass'] == 2].Age
          class3 = data[data['Pclass'] == 3].Age
          df = pd.DataFrame({'class1': class1, 'class2': class2, 'class3': class3})
          df.plot.kde()
```

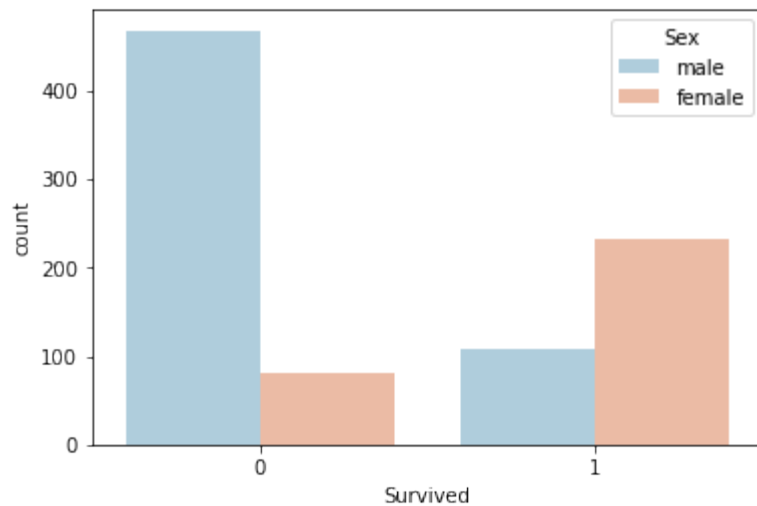
```
Out[78]: <matplotlib.axes._subplots.AxesSubplot at 0x2786fbd23c8>
```



```
In [68]: import seaborn as sns
```

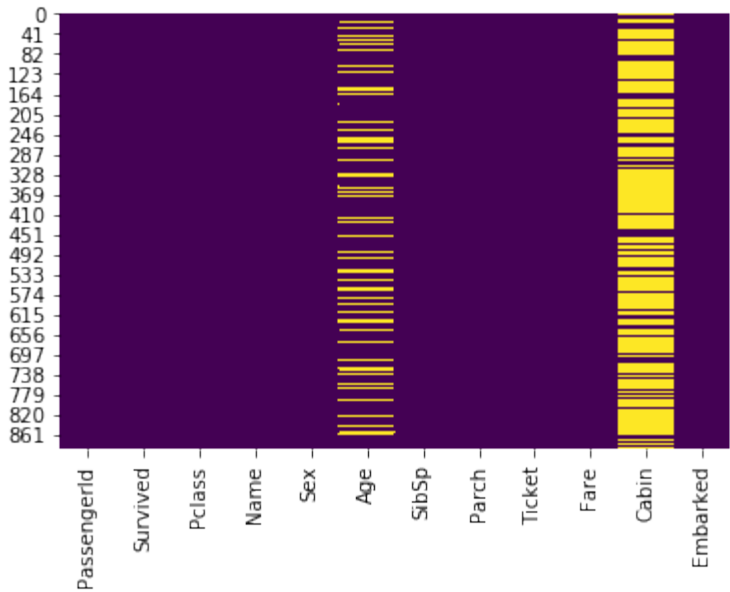
```
In [71]: sns.countplot(x='Survived', data=data, hue='Sex', palette='RdBu_r')
```

```
Out[71]: <matplotlib.axes._subplots.AxesSubplot at 0x2786e77a8c8>
```



```
In [74]: sns.heatmap(data.isna(), cbar=False, cmap='viridis')
```

```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x2786e94f5c8>
```



```
In [79]: sns.boxplot(x='Pclass', y='Age', data=data)
```

Out[79]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2786fbc2408>

