

```
In [2]: import pandas as pd

In [7]: fname = 'D:\Data Analytics\Stats Project\train.csv'

In [8]: data = pd.read_csv(fname)

In [9]: len(data)
Out[9]: 891

In [10]: data.head()
Out[10]:
```

	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	Cummings, 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

```

In [11]: data.count()
Out[11]: PassengerId    891
Survived      891
Pclass        891
Name          891
Sex           891
Age           714
SibSp         891
Parch         891
Ticket        891
Fare          891
Cabin        204
Embarked      889
dtype: int64

In [12]: data['Age'].min(), data['Age'].max()
Out[12]: (0.42, 80.0)

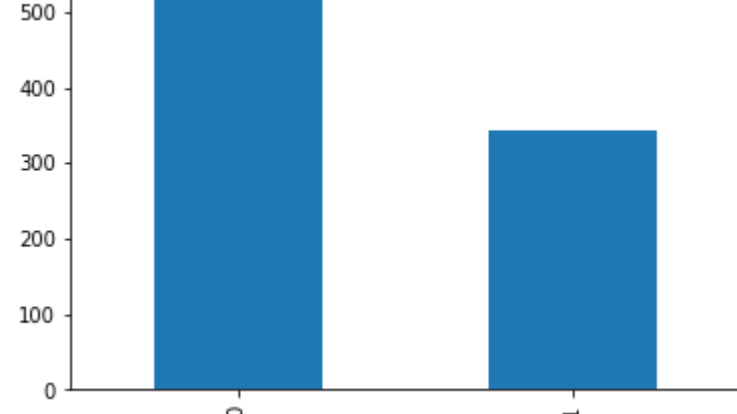
In [13]: data['Survived'].value_counts()
Out[13]: 0    549
         1    342
         Name: Survived, dtype: int64

In [14]: data['Survived'].value_counts() * 100 / len(data)
Out[14]: 0    61.616162
         1    38.383838
         Name: Survived, dtype: float64

In [15]: data['Sex'].value_counts()
Out[15]: male    577
         female  314
         Name: Sex, dtype: int64

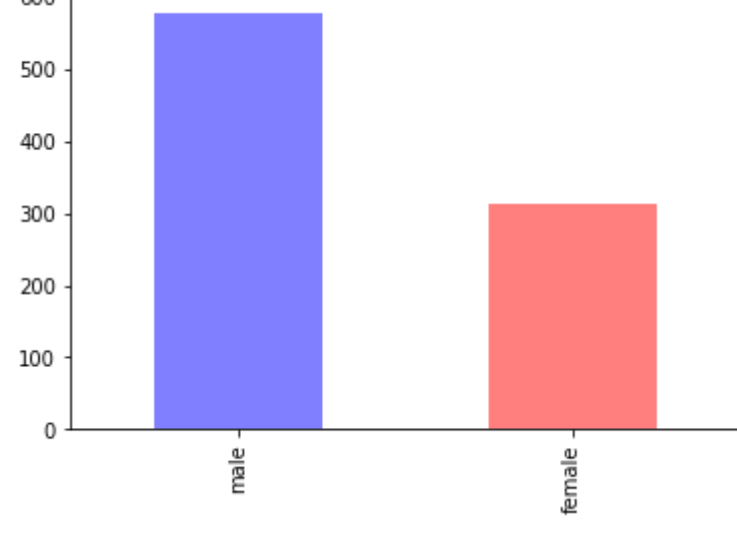
In [16]: data['Pclass'].value_counts()
Out[16]: 3    491
         1    216
         2    184
         Name: Pclass, dtype: int64

In [17]: %matplotlib inline
alpha_color = 0.5
data['Survived'].value_counts().plot(kind='bar')
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0xb4368550>
```



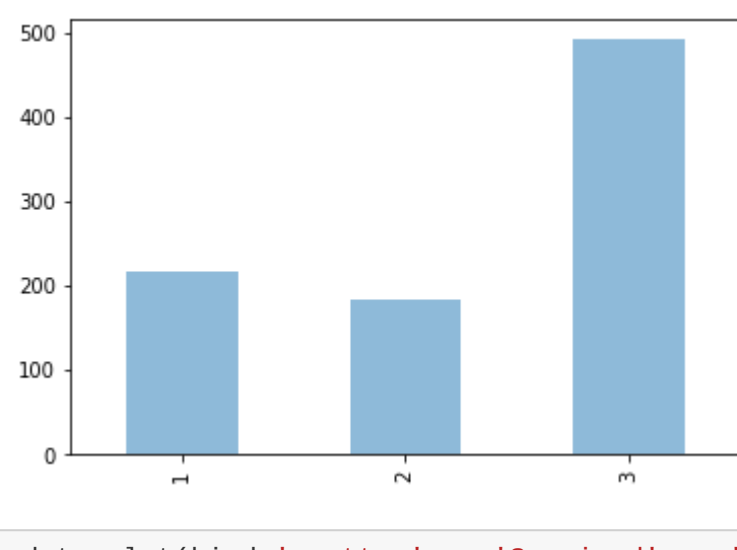
```

In [18]: data['Sex'].value_counts().plot(kind='bar',
                                         color=['b','r'],
                                         alpha=alpha_color)
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0xb22790>
```



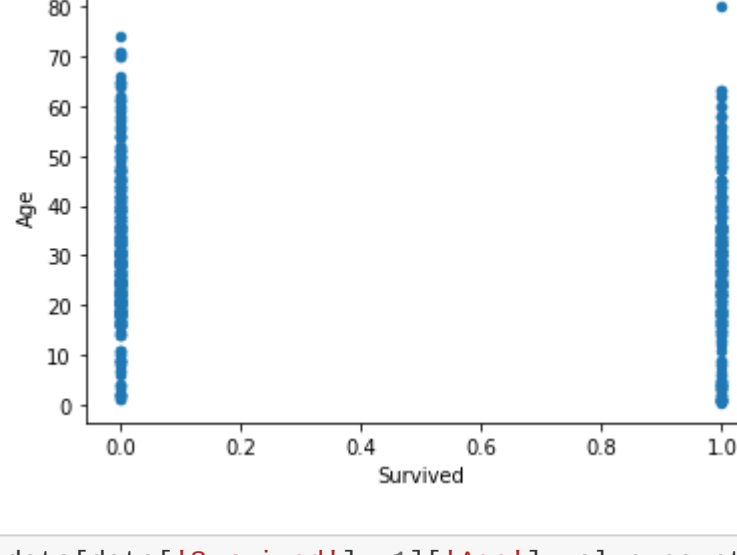
```

In [19]: data['Pclass'].value_counts().sort_index().plot(kind='bar',alpha=alpha_color)
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0xd13b4d0>
```



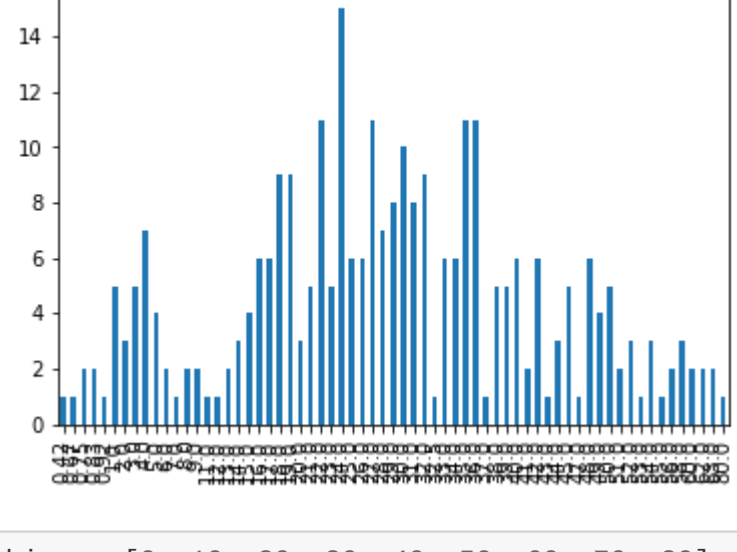
```

In [20]: data.plot(kind='scatter', x='Survived', y='Age')
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0xd3e6290>
```



```

In [21]: data[data['Survived']==1]['Age'].value_counts().sort_index().plot(kind='bar')
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0xd5b1c0>
```



```

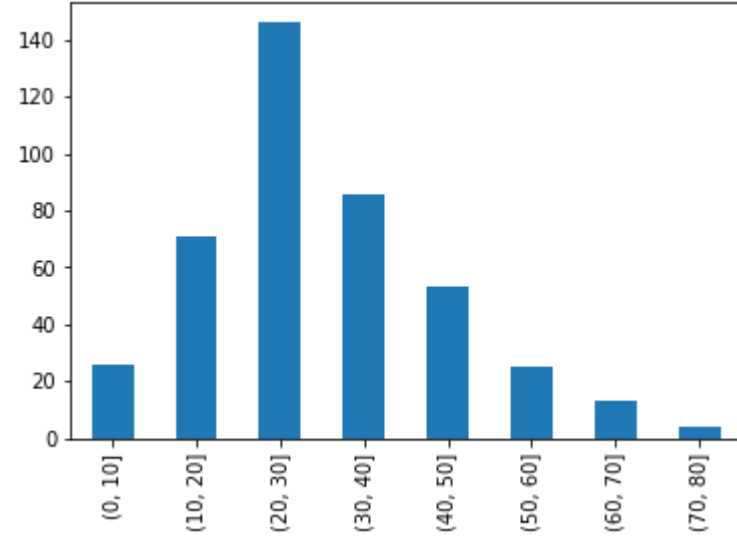
In [22]: bins = [0, 10, 20, 30, 40, 50, 60, 70, 80]
data['AgeBin'] = pd.cut(data['Age'],bins)

In [23]: data[data['Survived']==1]['AgeBin'].value_counts().sort_index().plot(kind='bar')
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0xd1d2dd0>
```



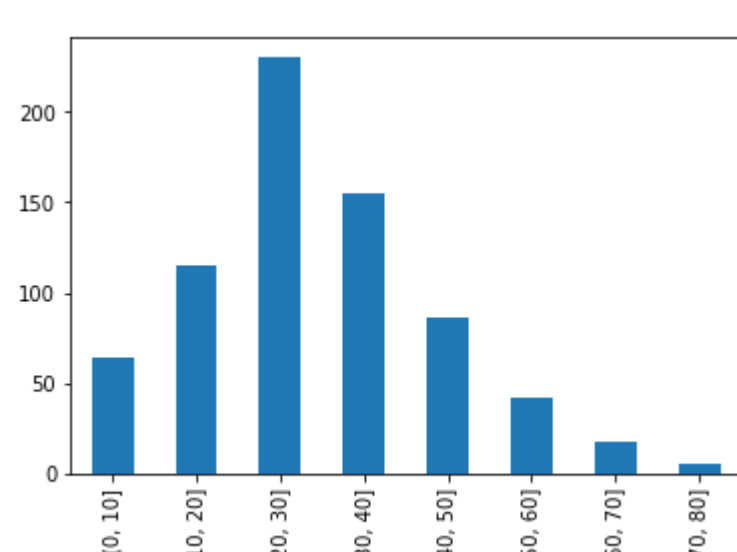
```

In [24]: data[data['Survived']==0]['AgeBin'].value_counts().sort_index().plot(kind='bar')
Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0xd2abfb0>
```



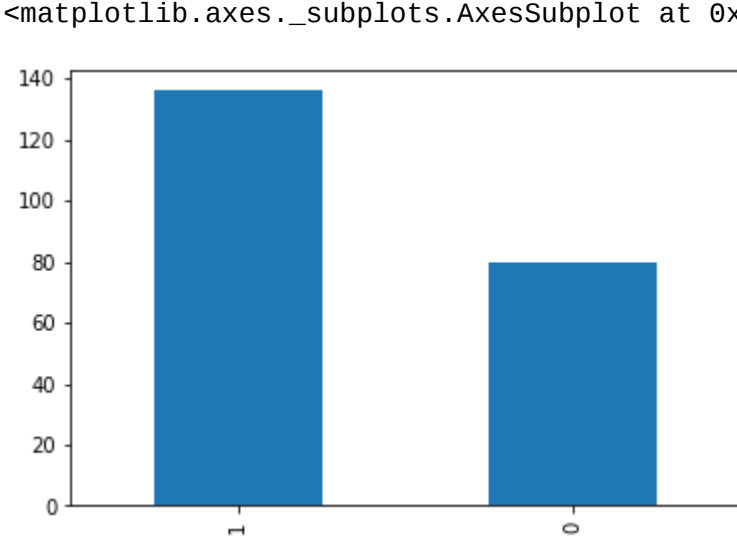
```

In [25]: data['AgeBin'].value_counts().sort_index().plot(kind='bar')
Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0xd477cb0>
```



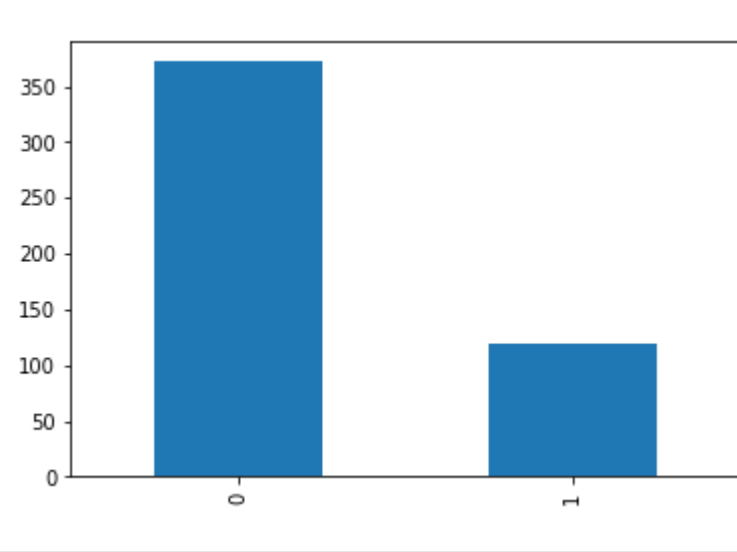
```

In [26]: data[data['Pclass'] == 1]['Survived'].value_counts().plot(kind='bar')
Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0xd4f0910>
```



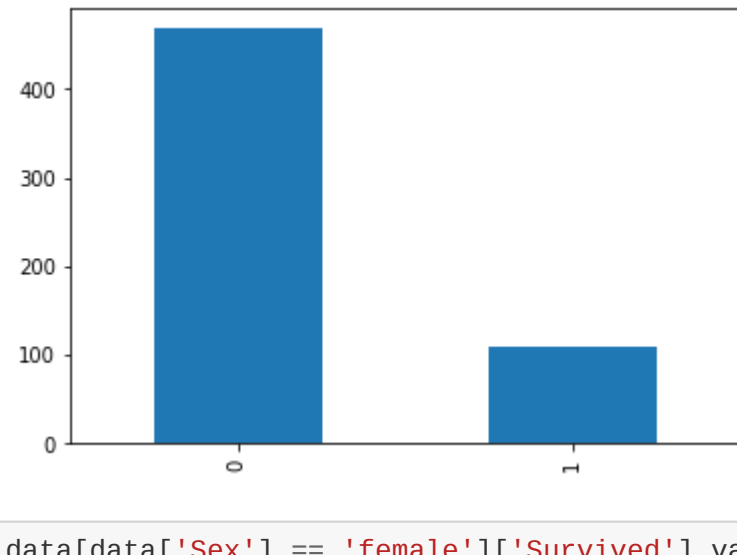
```

In [27]: data[data['Pclass'] == 3]['Survived'].value_counts().plot(kind='bar')
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0xd55c030>
```



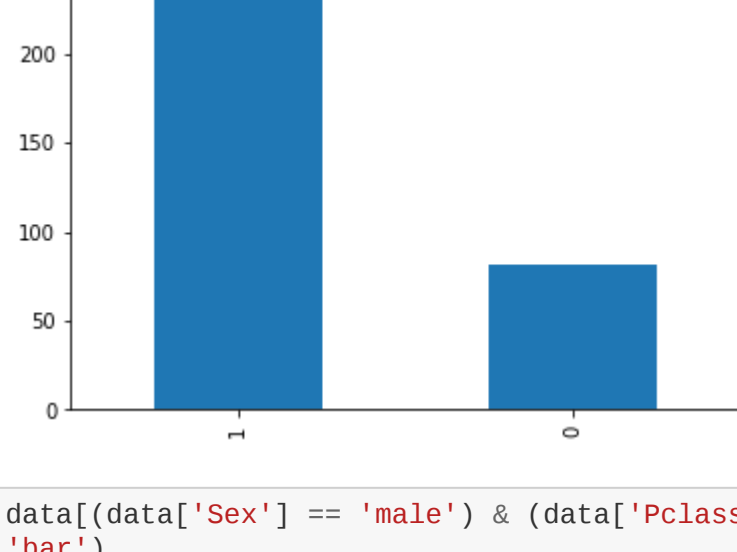
```

In [28]: data[data['Sex'] == 'male']['Survived'].value_counts().plot(kind='bar')
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0xd64e970>
```



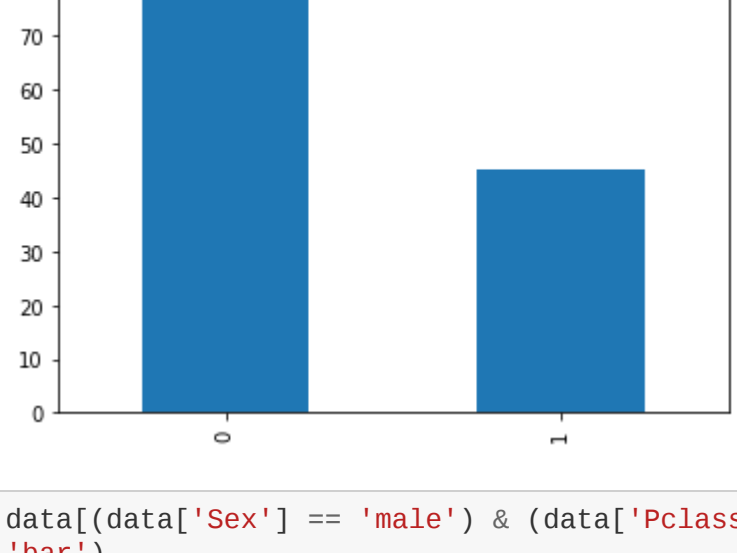
```

In [29]: data[data['Sex'] == 'female']['Survived'].value_counts().plot(kind='bar')
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0xd148e50>
```



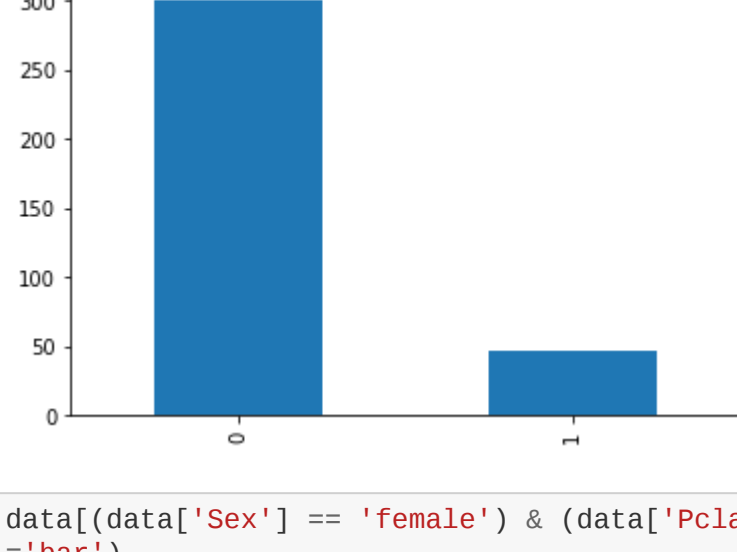
```

In [32]: data[(data['Sex'] == 'male') & (data['Pclass'] == 1)]['Survived'].value_counts().plot(kind='bar')
Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0xcf7f110>
```



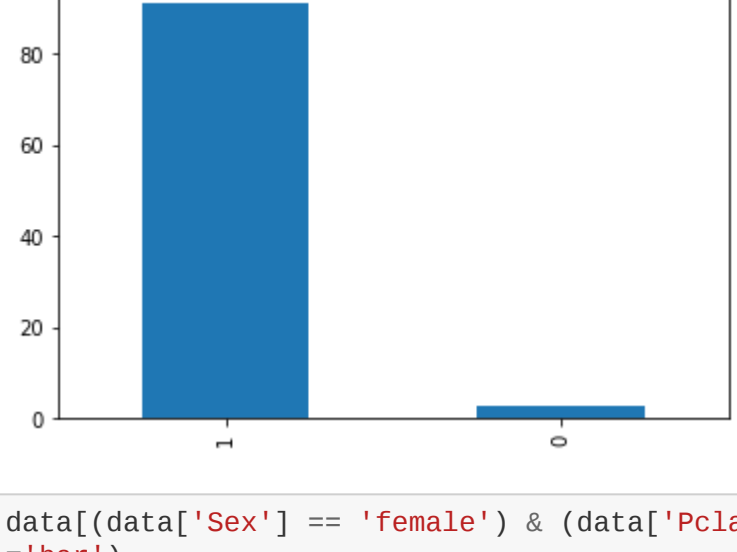
```

In [33]: data[(data['Sex'] == 'male') & (data['Pclass'] == 3)]['Survived'].value_counts().plot(kind='bar')
Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0xd64e990>
```



```

In [34]: data[(data['Sex'] == 'female') & (data['Pclass'] == 1)]['Survived'].value_counts().plot(kind='bar')
Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0xd6a8930>
```



```

In [35]: data[(data['Sex'] == 'female') & (data['Pclass'] == 3)]['Survived'].value_counts().plot(kind='bar')
Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0xd6c4ad0>
```