

1. Exploratory analysis of the Titanic Data Set

In [3]:

```
import pandas as pd
```

In [4]:

```
df=pd.read_csv('d:/train.csv')
df
```

Out[4]:

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
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [5]:

```
len(df)
```

Out[5]:

891

In [6]:

```
df.head()
```

Out[6]:

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

In [7]:

```
df.count()
```

Out[7]:

```
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 [8]:

```
df['Age'].min(), df['Age'].max()
```

Out[8]:

```
(0.42, 80.0)
```

In [9]:

```
df['Survived'].value_counts()
```

Out[9]:

```
0    549
1    342
Name: Survived, dtype: int64
```

In [10]:

```
df['Survived'].value_counts()*100/len(df)
```

Out[10]:

```
0    61.616162
1    38.383838
Name: Survived, dtype: float64
```

In [11]:

```
df['Sex'].value_counts()
```

Out[11]:

```
male    577
female  314
Name: Sex, dtype: int64
```

In [12]:

```
df['Pclass'].value_counts()
```

Out[12]:

```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [13]:

```
import matplotlib.pyplot as plt
```

```
import matplotlib.pyplot as plt
```

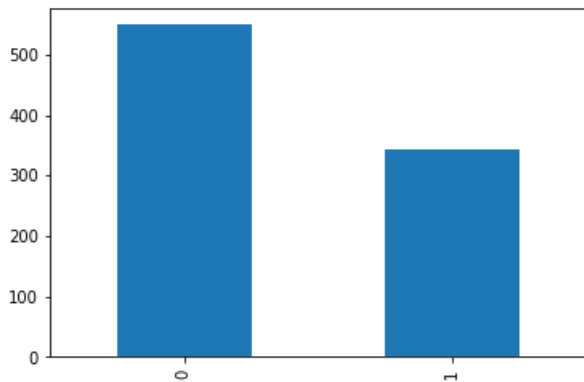
```
%matplotlib inline
```

```
alpha_color = 0.5
```

```
df['Survived'].value_counts().plot(kind= 'bar')
```

Out[13]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31dd20408>

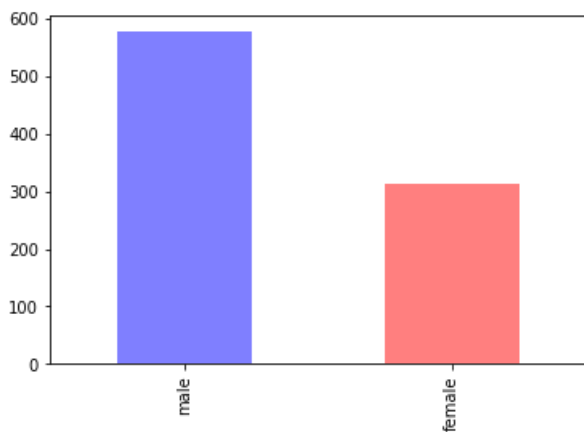


In [14]:

```
df['Sex'].value_counts().plot(kind = 'bar',  
                             color = ['b', 'r'],  
                             alpha = alpha_color)
```

Out[14]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31e15c5c8>

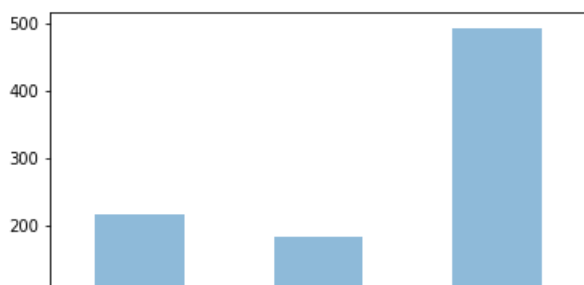


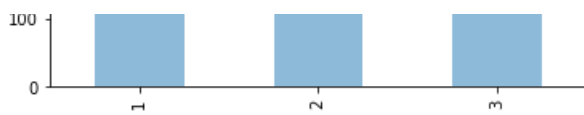
In [15]:

```
df['Pclass'].value_counts().sort_index().plot(kind = 'bar', alpha = alpha_color)
```

Out[15]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31e1fc6c8>



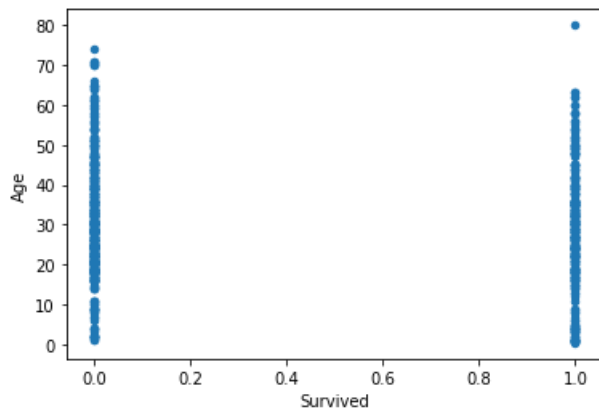


In [16]:

```
df.plot(kind='scatter', x = 'Survived', y = 'Age')
```

Out[16]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31e258d88>

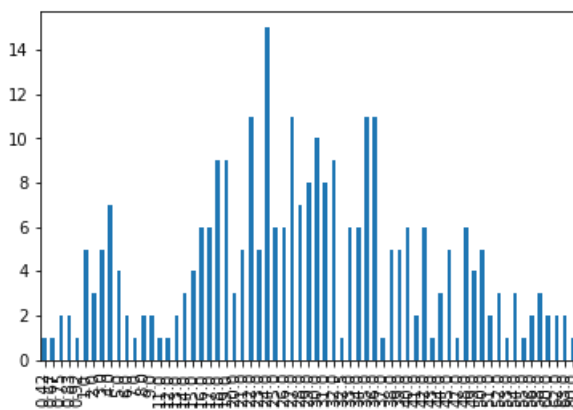


In [17]:

```
df[df['Survived'] == 1]['Age'].value_counts().sort_index().plot(kind = 'bar')
```

Out[17]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31e1edec8>



In [19]:

```
bins = [0, 10, 20, 30, 40, 50, 60, 70, 80]
df['AgeBin'] = pd.cut(df['Age'], bins)
```

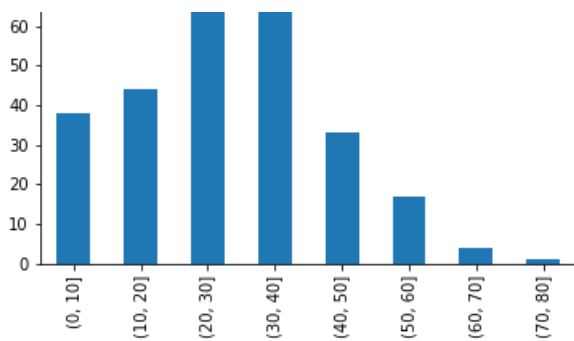
In [21]:

```
df[df['Survived'] == 1]['AgeBin'].value_counts().sort_index().plot(kind = 'bar')
```

Out[21]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31f797348>



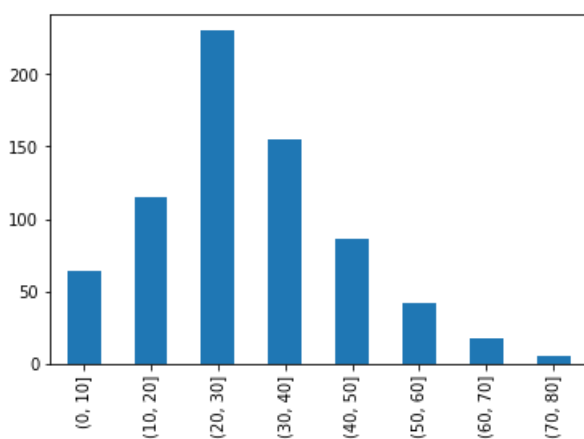


In [22]:

```
df['AgeBin'].value_counts().sort_index().plot(kind = 'bar')
```

Out[22]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31f7f9188>

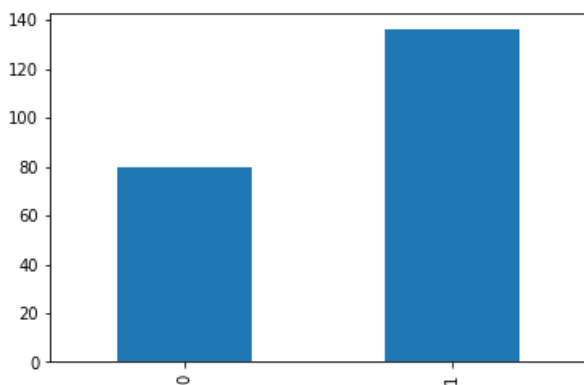


In [28]:

```
df[df['Pclass'] == 1]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[28]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fa25fc8>



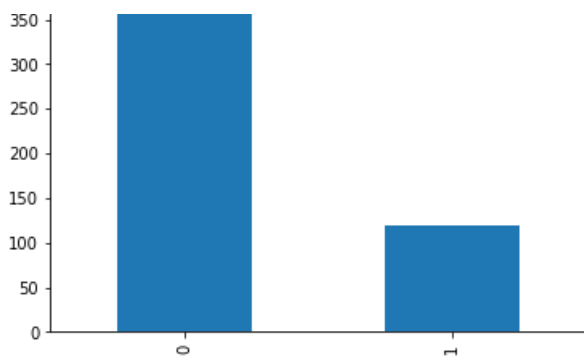
In [29]:

```
df[df['Pclass'] == 3]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[29]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fa847c8>



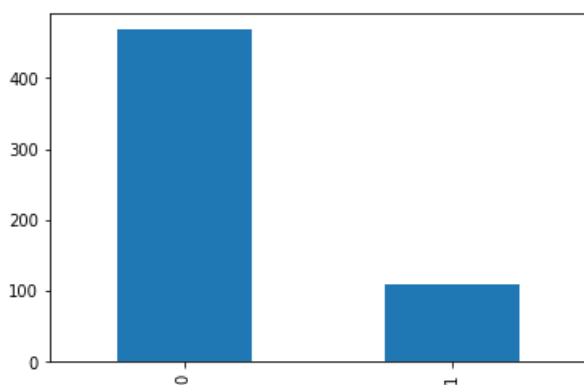


In [31]:

```
df[df['Sex'] == 'male']['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[31]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fae5c88>

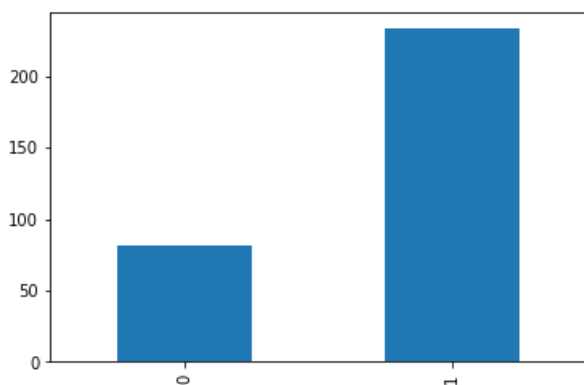


In [32]:

```
df[df['Sex'] == 'female']['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[32]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fb46088>

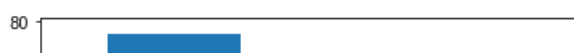


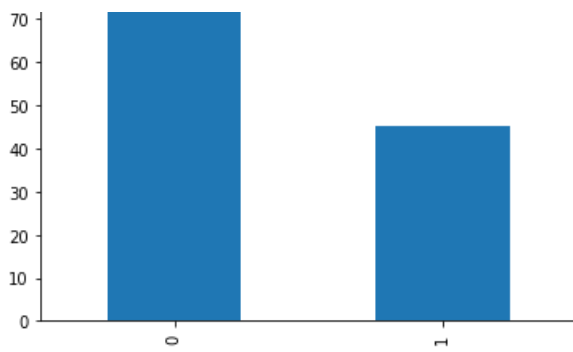
In [34]:

```
df[(df['Sex'] == 'male') & (df['Pclass'] == 1)]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[34]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fc02e48>



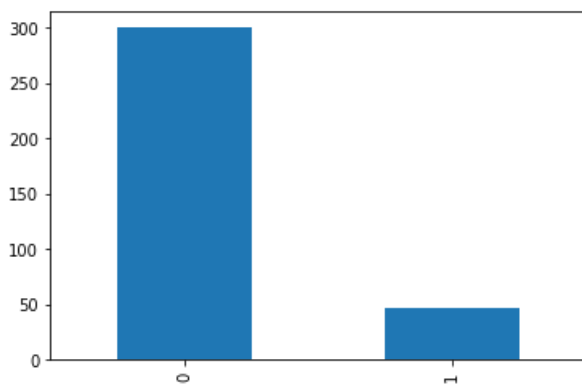


In [36]:

```
df[(df['Sex'] == 'male') & (df['Pclass'] == 3)]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[36]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fc6d848>

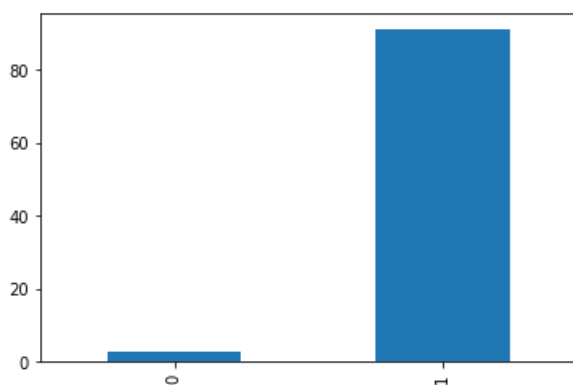


In [37]:

```
df[(df['Sex'] == 'female') & (df['Pclass'] == 1)]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[37]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fd275c8>

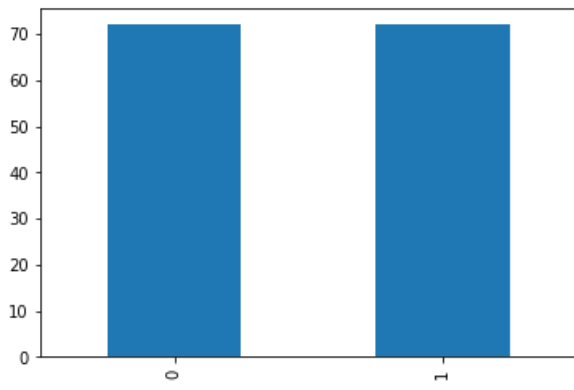


In [38]:

```
df[(df['Sex'] == 'female') & (df['Pclass'] == 3)]['Survived'].value_counts().sort_index().plot(kind = 'bar')
```

Out[38]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f31fd90b08>



In []: