In [37]:
```
import pandas as pd
df = pd.read_csv('train.csv')
df
```

Out[37]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	

891 rows × 12 columns

In [11]:
```
df.columns
```

Out[11]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [12]:
```
df.shape
```

Out[12]:
```
(891, 12)
```

In [15]:
```
df['Pclass'].count()
```

Out[15]:
```
891
```

In [17]:
```
df['Name'].count()
```

Out[17]:
```
891
```

In [18]:
```
df['Pclass'].unique()
```

Out[18]:
```
array([3, 1, 2], dtype=int64)
```

In [20]:
```
df['Sex'].unique()
```

Out[20]:
```
array(['male', 'female'], dtype=object)
```

In [21]:
```
df['Sex'].count()
```

Out[21]:
```
891
```

In [23]:
```
df['Sex'].value_counts()
```

Out[23]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [24]:
```
df['Survived'].value_counts()
```

Out[24]:
```
0    549
1    342
Name: Survived, dtype: int64
```

In [30]:
```
df['Sex'].min()
```

Out[30]:
```
'female'
```

In [32]:
```
df['Survived'].max()
```

Out[32]:
```
1
```

In [31]:
```
df['Sex'].max()
```

Out[31]:
```
'male'
```

In [47]:
```
df['Survived']
```

Out[47]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

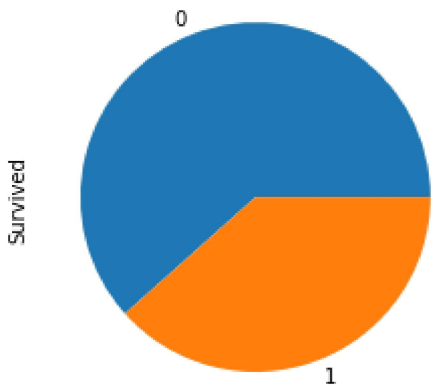
In [97]:
```
a = df['Survived'].value_counts()
a
```

Out[97]:
```
0    549
1    342
Name: Survived, dtype: int64
```

In [49]:
```
%matplotlib inline
```

In [55]:
```
a.plot(kind = 'pie')
```

Out[55]:
```
<AxesSubplot:ylabel='Survived'>
```

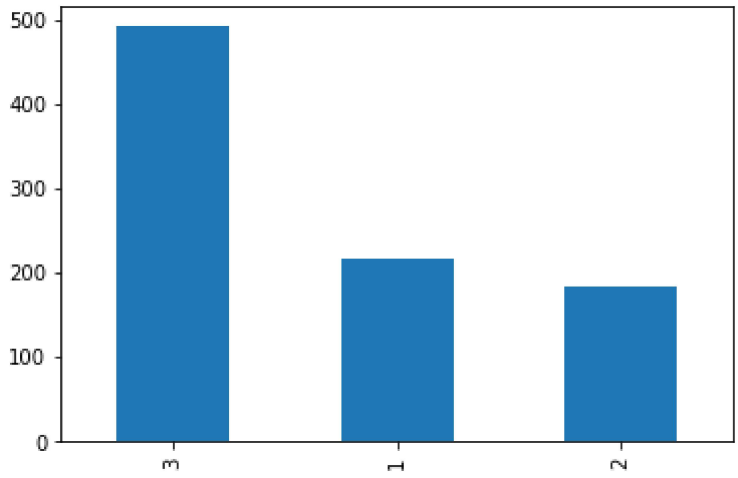


In [59]:
```
b = df.mean()
b
```

Out[59]:
```
PassengerId    446.000000
Survived         0.383838
Pclass           2.308642
Age             29.699118
SibSp            0.523008
Parch            0.381594
Fare            32.204208
dtype: float64
```

In [72]:
```
df['Pclass'].value_counts().plot(kind = 'bar')
```

Out[72]:
```
<AxesSubplot:>
```



In [90]:
```
df
```

Out[90]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	female	35.0	1	0	113803	53.1000	S
4	5	0	3	male	35.0	0	0	373450	8.0500	S
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	male	27.0	0	0	211536	13.0000	S
887	888	1	1	female	19.0	0	0	112053	30.0000	S
888	889	0	3	female	NaN	1	2	W./C. 6607	23.4500	S
889	890	1	1	male	26.0	0	0	111369	30.0000	C
890	891	0	3	male	32.0	0	0	370376	7.7500	Q

891 rows × 10 columns

In [104]:
```
grp = df.groupby('Sex')
```

In [105]:
```
grp.count()
```

Out[105]:

Sex	PassengerId	Survived	Pclass	Age	SibSp	Parch	Ticket	Fare	Embarked
female	314	314	314	261	314	314	314	314	312
male	577	577	577	453	577	577	577	577	577

In [106]:
```
df.count()
```

Out[106]:
```
PassengerId    891
Survived       891
Pclass         891
Sex            891
Age            714
SibSp          891
Parch          891
Ticket         891
Fare           891
Embarked       889
dtype: int64
```

In [107]:
```
grp
```

Out[107]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000024288730A00>
```

In [108]:
```
grp2 = df.groupby(['Survived','Sex'])
```

In [109]:
```
grp2.count()
```

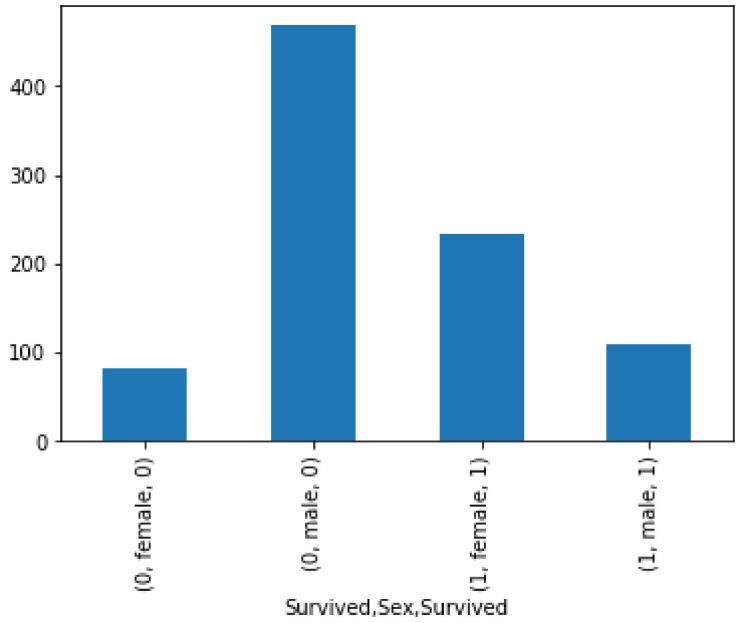
Out[109]:

		PassengerId	Pclass	Age	SibSp	Parch	Ticket	Fare	Embarked
Survived	**Sex**								
0	**female**	81	81	64	81	81	81	81	81
	male	468	468	360	468	468	468	468	468
1	**female**	233	233	197	233	233	233	233	231
	male	109	109	93	109	109	109	109	109

In [124]:
```
k=grp2['Survived'].value_counts()
```

In [119]:
```
k.plot(kind = 'bar')
```

Out[119]:
```
<AxesSubplot:xlabel='Survived,Sex,Survived'>
```



In [120]:
```
grp3 = df.groupby(['Survived','Pclass'])
```

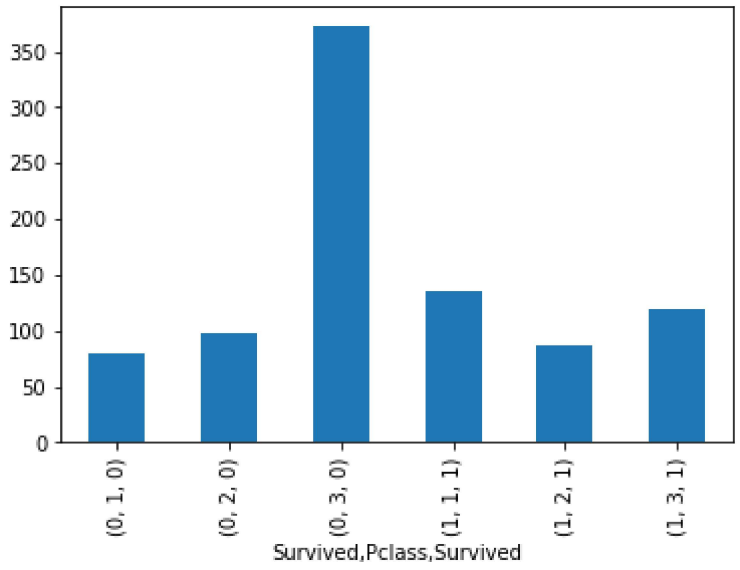
In [121]:
```
grp3.count()
```

Out[121]:

Survived	Pclass	PassengerId	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	80	80	64	80	80	80	80	80
	2	97	97	90	97	97	97	97	97
	3	372	372	270	372	372	372	372	372
1	1	136	136	122	136	136	136	136	134
	2	87	87	83	87	87	87	87	87
	3	119	119	85	119	119	119	119	119

In [129]:
```
h=grp3['Survived'].value_counts()
h.plot(kind = 'bar')
```

Out[129]: <AxesSubplot:xlabel='Survived,Pclass,Survived'>



In [125]:
```
grp4 = df.groupby(['Survived','Sex','Pclass'])
```

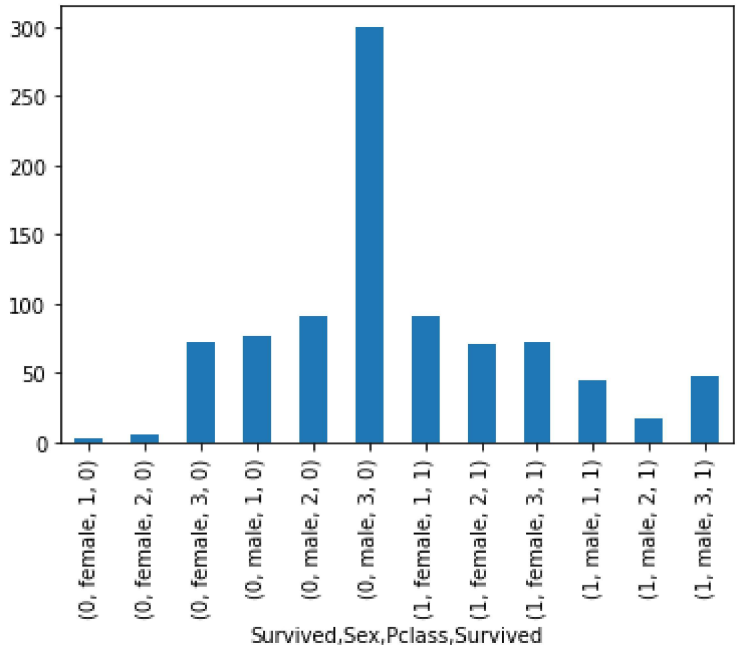
In [126]:
```
grp4.count()
```

Out[126]:

Survived	Sex	Pclass	PassengerId	Age	SibSp	Parch	Ticket	Fare	Embarked
0	female	1	3	3	3	3	3	3	3
		2	6	6	6	6	6	6	6
		3	72	55	72	72	72	72	72
	male	1	77	61	77	77	77	77	77
		2	91	84	91	91	91	91	91
		3	300	215	300	300	300	300	300
1	female	1	91	82	91	91	91	91	89
		2	70	68	70	70	70	70	70
		3	72	47	72	72	72	72	72
	male	1	45	40	45	45	45	45	45
		2	17	15	17	17	17	17	17
		3	47	38	47	47	47	47	47

In [128]:
```
i=grp4['Survived'].value_counts()
i.plot(kind = 'bar')
```

Out[128]: <AxesSubplot:xlabel='Survived,Sex,Pclass,Survived'>



In [131]:
```
df['SibSp'].unique()
```

Out[131]:
```
array([1, 0, 3, 4, 2, 5, 8], dtype=int64)
```

In [144]:
```
grp5 = df.groupby(['Pclass','SibSp'])
```

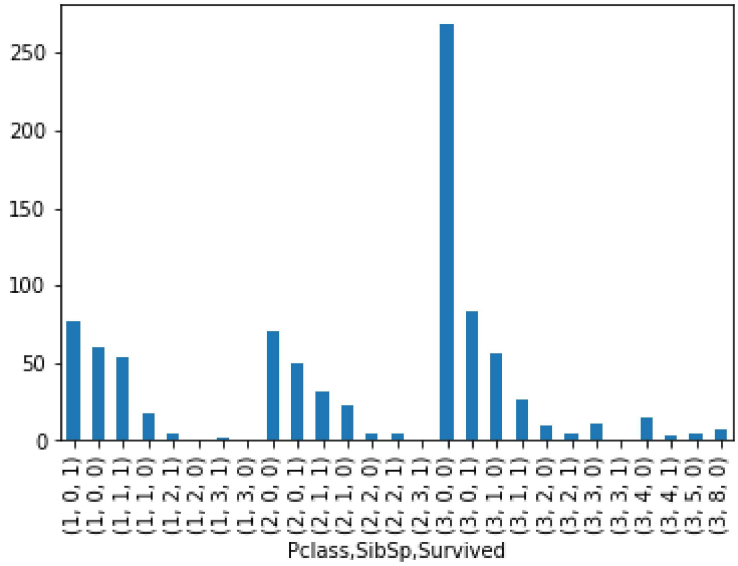
In [135]:
```
grp5.count()
```

Out[135]:

		PassengerId	Survived	Sex	Age	Parch	Ticket	Fare	Embarked
Pclass	**SibSp**								
1	**0**	137	137	137	113	137	137	137	135
	1	71	71	71	65	71	71	71	71
	2	5	5	5	5	5	5	5	5
	3	3	3	3	3	3	3	3	3
2	**0**	120	120	120	109	120	120	120	120
	1	55	55	55	55	55	55	55	55
	2	8	8	8	8	8	8	8	8
	3	1	1	1	1	1	1	1	1
3	**0**	351	351	351	249	351	351	351	351
	1	83	83	83	63	83	83	83	83
	2	15	15	15	12	15	15	15	15
	3	12	12	12	8	12	12	12	12
	4	18	18	18	18	18	18	18	18
	5	5	5	5	5	5	5	5	5
	8	7	7	7	0	7	7	7	7

In [140]:
```
l=grp5['Survived'].value_counts()
l.plot(kind = 'bar')
```

Out[140]: <AxesSubplot:xlabel='Pclass,SibSp,Survived'>



In [141]:
```
df['Parch'].unique()
```

Out[141]: array([0, 1, 2, 5, 3, 4, 6], dtype=int64)

In [143]:
```
grp6 = df.groupby(['Pclass','Parch'])
```

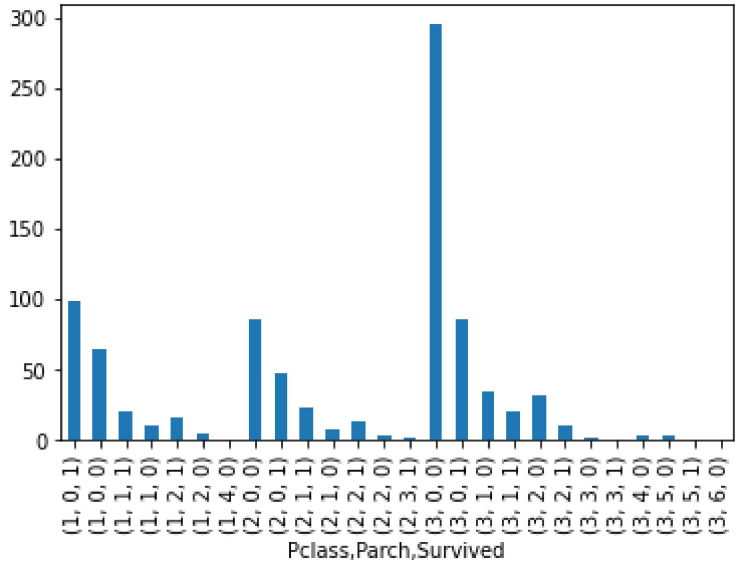
In [145]:
```
grp6.count()
```

Out[145]:

Pclass	Parch	PassengerId	Survived	Sex	Age	SibSp	Ticket	Fare	Embarked
1	**0**	163	163	163	134	163	163	163	161
	1	31	31	31	30	31	31	31	31
	2	21	21	21	21	21	21	21	21
	4	1	1	1	1	1	1	1	1
2	**0**	134	134	134	123	134	134	134	134
	1	32	32	32	32	32	32	32	32
	2	16	16	16	16	16	16	16	16
	3	2	2	2	2	2	2	2	2
3	**0**	381	381	381	264	381	381	381	381
	1	55	55	55	48	55	55	55	55
	2	43	43	43	31	43	43	43	43
	3	3	3	3	3	3	3	3	3
	4	3	3	3	3	3	3	3	3
	5	5	5	5	5	5	5	5	5
	6	1	1	1	1	1	1	1	1

In [146]:
```
m=grp6['Survived'].value_counts()
m.plot(kind = 'bar')
```

Out[146]: <AxesSubplot:xlabel='Pclass,Parch,Survived'>



In [149]:
```
df['Embarked'].unique()
```

Out[149]: array(['S', 'C', 'Q', nan], dtype=object)

In [ ]:
```

```