

In [62]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

Filename = '/users/ambilurama/Downloads/Nano/data analysis/survival - Sheet1.csv'

survival_df = pd.read_csv(Filename ,index_col='PassengerId' )
# using Passengerid as the index
```

In [63]:

```
survival_df.head()
```

Out[63]:

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

In [64]:

```
survival_df.groupby('Survived').groups
```

Out[64]:

```
{0: [1,
      5,
      6,
      7,
      8,
      13,
      14,
      15,
      17,
      19,
      21,
      25,
      27,
      28,
      30,
      31,
      34,
      35,
```

In [65]:

```
survival_df.groupby('Survived').size()
```

Out[65]:

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

In [66]:

```
survival_df.groupby('Sex').size()
```

Out[66]:

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

In []:

In [67]:

```
survival_df.groupby(['Survived','Sex']).size()
```

Out[67]:

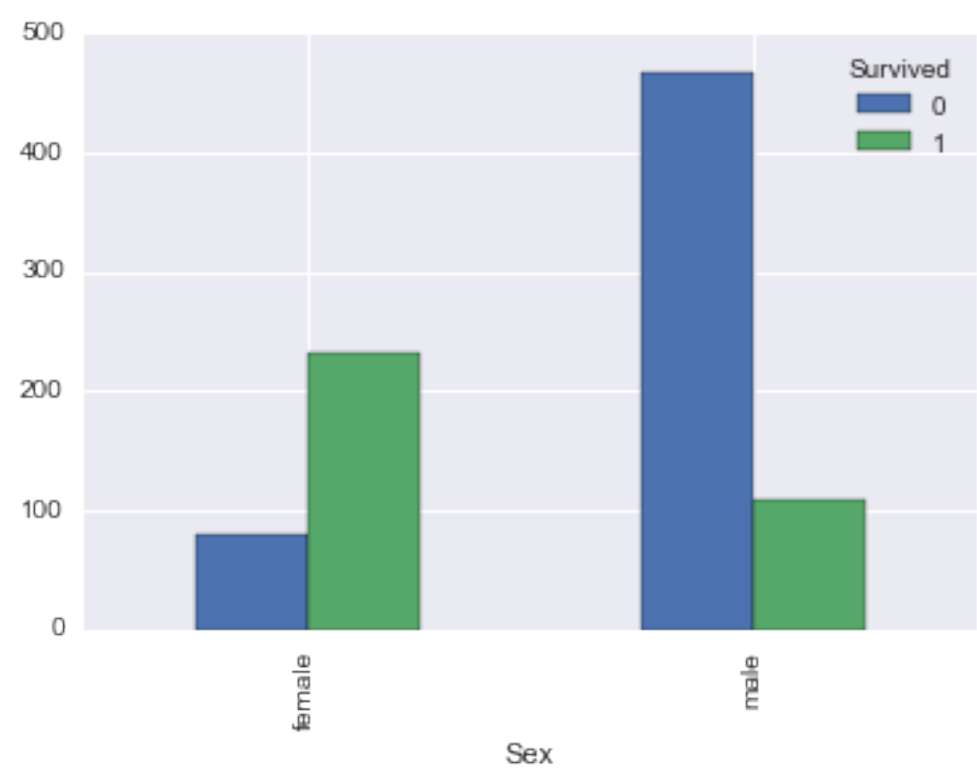
Sex	female	male
Survived		
0	81	468
1	233	109

In [68]:

```
survival_df.groupby(['Sex','Survived']).size().unstack().plot.bar()
```

Out[68]:

<matplotlib.axes._subplots.AxesSubplot at 0x11d522290>



In [69]:

```
survival_df.groupby(['Survived', 'Name']).groups
```

Out[69]:

```
{(0, 'Abbing, Mr. Anthony'): [846],
 (0, 'Abbott, Mr. Rossmore Edward'): [747],
 (0, 'Abelson, Mr. Samuel'): [309],
 (0, 'Adahl, Mr. Mauritz Nils Martin'): [366],
 (0, 'Adams, Mr. John'): [402],
 (0, 'Ahlin, Mrs. Johan (Johanna Persdotter Larsson)'): [41],
 (0, 'Alexander, Mr. William'): [811],
 (0, 'Alhomaki, Mr. Ilmari Rudolf'): [841],
 (0, 'Ali, Mr. Ahmed'): [211],
 (0, 'Ali, Mr. William'): [785],
 (0, 'Allen, Mr. William Henry'): [5],
 (0, 'Allison, Miss. Helen Loraine'): [298],
 (0, 'Allison, Mrs. Hudson J C (Bessie Waldo Daniels)'): [499],
 (0, 'Allum, Mr. Owen George'): [835],
 (0, 'Andersson, Master. Sigvard Harald Elias'): [851],
 (0, 'Andersson, Miss. Ebba Iris Alfrida'): [814],
 (0, 'Andersson, Miss. Ellis Anna Maria'): [120],
 (0, 'Andersson, Miss. Ingeborg Constanzia'): [542],
```

In [70]:

```
survival_df.groupby(['Survived', 'Age']).groups
```

Out[70]:

```
{(0, 7.0): [51, 279],
 (0, 20.0): [13, 92, 114, 132, 379, 405, 442, 641, 683, 726, 841, 877]
 ,
 (0, 23.5): [297],
 (0, nan): [6],
 (0, nan): [681],
 (0, 10.0): [420, 820],
 (0, nan): [215],
 (0, 17.0): [115, 164, 434, 501, 533, 722, 845],
 (0, 20.5): [228],
 (0, 28.5): [58, 736],
 (0, 61.0): [171, 327, 626],
 (0, nan): [503],
 (0, nan): [186],
 (0, 32.0): [71, 207, 383, 520, 637, 658, 666, 770, 891],
 (0, nan): [816],
 (0, nan): [869],
 (0, 70.0): [673, 746],
```

```
survival_df.loc[18]
```

In [72]:

```
survival_df.loc[156]
```

In [73]:

```
survival_df.groupby('Survived').get_group(0)
```

Out[73]:

[illegible]

6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4
7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51
8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21
13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0
14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8
17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18
21	0	2	Fynney, Mr. Joseph J	male	35.0	0	0	239865	26
25	0	3	Palsson, Miss. Torborg Danira	female	8.0	3	1	349909	21
27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2
28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	26
30	0	3	Todoroff, Mr. Lalio	male	NaN	0	0	349216	7.8
31	0	1	Uruchurtu, Don. Manuel	male	40.0	0	0	PC 17601	27

			E						
34	0	2	Wheadon, Mr. Edward H	male	66.0	0	0	C.A. 24579	10
35	0	1	Meyer, Mr. Edgar Joseph	male	28.0	1	0	PC 17604	82
36	0	1	Holverson, Mr. Alexander Oskar	male	42.0	1	0	113789	52
38	0	3	Cann, Mr. Ernest Charles	male	21.0	0	0	A./5. 2152	8.0
39	0	3	Vander Planke, Miss. Augusta Maria	female	18.0	2	0	345764	18
41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0	1	0	7546	9.4
42	0	2	Turpin, Mrs. William John Robert (Dorothy Ann ...	female	27.0	1	0	11668	21
43	0	3	Kraeff, Mr. Theodor	male	NaN	0	0	349253	7.8
46	0	3	Rogers, Mr. William John	male	NaN	0	0	S.C./A.4. 23567	8.0
47	0	3	Lennon, Mr. Denis	male	NaN	1	0	370371	15
49	0	3	Samaan, Mr. Youssef	male	NaN	2	0	2662	21
50	0	3	Arnold-Franchi, Mrs. Josef (Josefine Franchi)	female	18.0	1	0	349237	17
			Panula,						

51	0	3	Master. Juha Niilo	male	7.0	4	1	3101295	39
52	0	3	Nosworthy, Mr. Richard Cater	male	21.0	0	0	A/4. 39886	7.8
...
845	0	3	Culumovic, Mr. Jeso	male	17.0	0	0	315090	8.6
846	0	3	Abbing, Mr. Anthony	male	42.0	0	0	C.A. 5547	7.5
847	0	3	Sage, Mr. Douglas Bullen	male	NaN	8	2	CA. 2343	69
848	0	3	Markoff, Mr. Marin	male	35.0	0	0	349213	7.8
849	0	2	Harper, Rev. John	male	28.0	0	1	248727	33
851	0	3	Andersson, Master. Sigvard Harald Elias	male	4.0	4	2	347082	31
852	0	3	Svensson, Mr. Johan	male	74.0	0	0	347060	7.7
853	0	3	Boulos, Miss. Nourelain	female	9.0	1	1	2678	15
855	0	2	Carter, Mrs. Ernest Courtenay (Lilian Hughes)	female	44.0	1	0	244252	26
860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2
861	0	3	Hansen, Mr. Claus Peter	male	41.0	2	0	350026	14
862	0	2	Giles, Mr. Frederick Edward	male	21.0	1	0	28134	11
864	0	3	Sage, Miss. Dorothy Edith	female	NaN	8	2	CA. 2343	69

			"Dolly"						
865	0	2	Gill, Mr. John William	male	24.0	0	0	233866	13
868	0	1	Roebbling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50
869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5
871	0	3	Balkic, Mr. Cerin	male	26.0	0	0	349248	7.8
873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0
874	0	3	Vander Cruyssen, Mr. Victor	male	47.0	0	0	345765	9.0
877	0	3	Gustafsson, Mr. Alfred Ossian	male	20.0	0	0	7534	9.8
878	0	3	Petroff, Mr. Nedelio	male	19.0	0	0	349212	7.8
879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8
882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8
883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10
884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10
885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0
886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29
			Montvila,						

887	0	2	Rev. Juozas	male	27.0	0	0	211536	13
889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23
891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

549 rows × 11 columns

In [74]:

```
unsurvived_df=survival_df.groupby( 'Survived' ).get_group(0)
```

In [75]:

```
survival_df.groupby( 'Survived' ).get_group(1)
```

Out[75]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
PassengerId											
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.00	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.00	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily	female	35.00	1	0	113803	53.1000	C123	S

In [76]:

```
survival_df.groupby( 'Survived' ).get_group(1)[ 'Name' ]
```

Out[76]:

```

PassengerId
2      Cumings, Mrs. John Bradley (Florence Briggs Th...
3                        Heikkinen, Miss. Laina
4      Futrelle, Mrs. Jacques Heath (Lily May Peel)
```

Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
Nasser, Mrs. Nicholas (Adele Achem)
Sandstrom, Miss. Marguerite Rut
Bonnell, Miss. Elizabeth
Hewlett, Mrs. (Mary D Kingcome)
Williams, Mr. Charles Eugene
Masselmani, Mrs. Fatima
Beesley, Mr. Lawrence
McGowan, Miss. Anna "Annie"
Sloper, Mr. William Thompson
Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...
O'Dwyer, Miss. Ellen "Nellie"
Spencer, Mrs. William Augustus (Marie Eugenie)
Glynn, Miss. Mary Agatha
Mamee, Mr. Hanna
Nicola-Yarred, Miss. Jamila
Laroche, Miss. Simonne Marie Anne Andree
Devaney, Miss. Margaret Delia
O'Driscoll, Miss. Bridget
Harper, Mrs. Henry Sleeper (Myna Haxtun)
Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkin...
Woolner, Mr. Hugh
Rugg, Miss. Emily
West, Miss. Constance Mirium
Icard, Miss. Amelie
Moubarek, Master. Gerios
Nye, Mrs. (Elizabeth Ramell)
...
Chambers, Mrs. Norman Campbell (Bertha Griggs)
Hays, Mrs. Charles Melville (Clara Jennings Gr...
Lulic, Mr. Nikola
Moor, Mrs. (Beila)
Mallet, Master. Andre
McCormack, Mr. Thomas Joseph
Stone, Mrs. George Nelson (Martha Evelyn)
Yasbeck, Mrs. Antoni (Selini Alexander)
Richards, Master. George Sibley
Compton, Miss. Sara Rebecca
Chip, Mr. Chang
Marechal, Mr. Pierre
Serepeca, Miss. Augusta
Goldenberg, Mrs. Samuel L (Edwiga Grabowska)
Lines, Miss. Mary Conover
Aks, Mrs. Sam (Leah Rosen)
Wick, Mrs. George Dennick (Mary Hitchcock)
Daly, Mr. Peter Denis
Baclini, Mrs. Solomon (Latifa Qurban)
Swift, Mrs. Frederick Joel (Margaret Welles Ba...
Bystrom, Mrs. (Karolina)
Duran y More, Miss. Asuncion
Johnson, Master. Harold Theodor
Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
Abelson, Mrs. Samuel (Hannah Wizosky)

```
876             Najib, Miss. Adele Kiamie "Jane"
880         Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)
881         Shelley, Mrs. William (Imanita Parrish Hall)
888             Graham, Miss. Margaret Edith
890             Behr, Mr. Karl Howell
```

```
Name: Name, dtype: object
```

```
In [ ]:
```

```
In [77]:
```

```
survival_df.groupby('Parch').size()
```

```
Out[77]:
```

```
Parch
0      678
1      118
2       80
3        5
4         4
5         5
6         1
dtype: int64
```

```
In [ ]:
```

In [78]:

```
survival_df.groupby('Survived').get_group(1)
```

Out[78]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
PassengerId											
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.00	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.00	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily	female	35.00	1	0	113803	53.1000	C123	S

In [79]:

```
survival_df.groupby('Parch').get_group(2)['Name']
```

Out[79]:

PassengerId

9 Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)

28 Fortune, Mr. Charles Alexander

44 Laroche, Miss. Simonne Marie Anne Andree

59 West, Miss. Constance Mirium

60 Goodwin, Master. William Frederick

64 Skoog, Master. Harald

69 Andersson, Miss. Erna Alexandra

72 Goodwin, Miss. Lillian Amy

79 Caldwell, Master. Alden Gates

89 Fortune, Miss. Mabel Helen

94 Dean, Mr. Bertram Frank

120 Andersson, Miss. Ellis Anna Maria

137 Newsom, Miss. Helen Monypeny

141 Boulos, Mrs. Joseph (Sultana)

148 Ford, Miss. Robina Maggie "Ruby"

149 Navratil, Mr. Michel ("Louis M Hoffman")

154 van Billiard, Mr. Austin Blyler

160 Sage, Master. Thomas Henry

166 Goldsmith, Master. Frank John William "Frankie"

181 Sage, Miss. Constance Gladys

183 Asplund, Master. Clarence Gustaf Hugo

185 Kink-Heilmann, Miss. Luise Gretchen

202 Sage, Mr. Frederick
 234 Asplund, Miss. Lillian Gertrud
 238 Collyer, Miss. Marjorie "Lottie"
 248 Hamalainen, Mrs. William (Anna)
 255 Rosblom, Mrs. Viktor (Helena Wilhelmina)
 256 Touma, Mrs. Darwis (Hanne Youssef Razi)
 262 Asplund, Master. Edvin Rojj Felix
 298 Allison, Miss. Helen Loraine
 ...
 534 Peter, Mrs. Catherine (Catherine Rizk)
 536 Hart, Miss. Eva Miriam
 540 Frolicher, Miss. Hedwig Margaritha
 541 Crosby, Miss. Harriet R
 542 Andersson, Miss. Ingeborg Constanzia
 543 Andersson, Miss. Sigrid Elisabeth
 551 Thayer, Mr. John Borland Jr
 586 Taussig, Miss. Ruth
 594 Bourke, Miss. Mary
 609 Laroche, Mrs. Joseph (Juliette Marie Louise La...
 616 Herman, Miss. Alice
 635 Skoog, Miss. Mabel
 643 Skoog, Miss. Margit Elizabeth
 660 Newell, Mr. Arthur Webster
 684 Goodwin, Mr. Charles Edward
 686 Laroche, Mr. Joseph Philippe Lemercier
 743 Ryerson, Miss. Susan Parker "Suzette"
 755 Herman, Mrs. Samuel (Jane Laver)
 764 Carter, Mrs. William Ernest (Lucile Polk)
 784 Johnston, Mr. Andrew G
 789 Dean, Master. Bertram Vere
 793 Sage, Miss. Stella Anna
 803 Carter, Master. William Thornton II
 814 Andersson, Miss. Ebba Iris Alfrida
 820 Skoog, Master. Karl Thorsten
 828 Mallet, Master. Andre
 847 Sage, Mr. Douglas Bullen
 851 Andersson, Master. Sigvard Harald Elias
 864 Sage, Miss. Dorothy Edith "Dolly"
 889 Johnston, Miss. Catherine Helen "Carrie"

Name: Name, dtype: object

In [80]:

```
survival_df.groupby('Parch').size()
```

Out[80]:

```
Parch
0      678
1      118
2       80
3        5
4         4
5         5
6         1
dtype: int64
```

In [81]:

```
Alive_df=pd.DataFrame.from_dict(survival_df.groupby('Survived').get_group(1))
print Alive_df
```

PassengerId	Survived	Pclass \
2	1	1
3	1	3
4	1	1
9	1	3
10	1	2
11	1	3
12	1	1
16	1	2
18	1	2
20	1	3
22	1	2
23	1	3
24	1	1
26	1	3
29	1	3
32	1	1
33	1	3
37	1	3

In [82]:

```
people_Alive_df=pd.DataFrame.from_dict(survival_df.groupby('Survived').get_group(1))
```

In [83]:

```
people_Alive_df
```

Out[83]:

	Name

PassengerId	
2	Cumings, Mrs. John Bradley (Florence Briggs Th...
3	Heikkinen, Miss. Laina
4	Futrelle, Mrs. Jacques Heath (Lily May Peel)
9	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
10	Nasser, Mrs. Nicholas (Adele Achem)
11	Sandstrom, Miss. Marguerite Rut
12	Bonnell, Miss. Elizabeth
16	Hewlett, Mrs. (Mary D Kingcome)
18	Williams, Mr. Charles Eugene
20	Masselmani, Mrs. Fatima
22	Beesley, Mr. Lawrence
23	McGowan, Miss. Anna "Annie"
24	Sloper, Mr. William Thompson
26	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...
29	O'Dwyer, Miss. Ellen "Nellie"
32	Spencer, Mrs. William Augustus (Marie Eugenie)
33	Glynn, Miss. Mary Agatha
37	Mamee, Mr. Hanna
40	Nicola-Yarred, Miss. Jamila
44	Laroche, Miss. Simonne Marie Anne Andree
45	Devaney, Miss. Margaret Delia
48	O'Driscoll, Miss. Bridget
53	Harper, Mrs. Henry Sleeper (Myna Haxtun)
54	Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkin...
56	Woolner, Mr. Hugh
57	Rugg, Miss. Emily
59	West, Miss. Constance Mirium
62	Icard, Miss. Amelie
66	Moubarek, Master. Gerios
67	Nye, Mrs. (Elizabeth Ramell)

...	...
810	Chambers, Mrs. Norman Campbell (Bertha Griggs)
821	Hays, Mrs. Charles Melville (Clara Jennings Gr...
822	Lulic, Mr. Nikola
824	Moor, Mrs. (Beila)
828	Mallet, Master. Andre
829	McCormack, Mr. Thomas Joseph
830	Stone, Mrs. George Nelson (Martha Evelyn)
831	Yasbeck, Mrs. Antoni (Selini Alexander)
832	Richards, Master. George Sibley
836	Compton, Miss. Sara Rebecca
839	Chip, Mr. Chang
840	Marechal, Mr. Pierre
843	Serepeca, Miss. Augusta
850	Goldenberg, Mrs. Samuel L (Edwiga Grabowska)
854	Lines, Miss. Mary Conover
856	Aks, Mrs. Sam (Leah Rosen)
857	Wick, Mrs. George Dennick (Mary Hitchcock)
858	Daly, Mr. Peter Denis
859	Baclini, Mrs. Solomon (Latifa Qurban)
863	Swift, Mrs. Frederick Joel (Margaret Welles Ba...
866	Bystrom, Mrs. (Karolina)
867	Duran y More, Miss. Asuncion
870	Johnson, Master. Harold Theodor
872	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
875	Abelson, Mrs. Samuel (Hannah Wzosky)
876	Najib, Miss. Adele Kiamie "Jane"
880	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)
881	Shelley, Mrs. William (Imanita Parrish Hall)
888	Graham, Miss. Margaret Edith
890	Behr, Mr. Karl Howell

342 rows × 1 columns

In []:

In [84]:

```
def split_name(data):  
  
    full_name=data.split(" ")  
  
    last_name=full_name[0]  
    salutation=full_name[1]  
    first_name_1=full_name[2]  
    if len(full_name)==3:  
        first_name_2= ' '  
    else:  
        first_name_2=full_name[3]  
  
    return (last_name,salutation,first_name_1,first_name_2)
```

In []:

In [85]:

```
print split_name("Lines Miss Mary Conover")  
  
('Lines', 'Miss', 'Mary', 'Conover')
```

In []:

In [86]:

```
def clear_name(data):  
    return (data['Name'].str.translate(None, "(,)."))
```

In [87]:

```
clear_name(people_Alive_df)
```

Out[87]:

```
PassengerId  
2      Cumings Mrs John Bradley Florence Briggs Thayer  
3      Heikkinen Miss Laina
```

4 Futrelle Mrs Jacques Heath Lily May Peel
9 Johnson Mrs Oscar W Elisabeth Vilhelmina Berg
10 Nasser Mrs Nicholas Adele Achem
11 Sandstrom Miss Marguerite Rut
12 Bonnell Miss Elizabeth
16 Hewlett Mrs Mary D Kingcome
18 Williams Mr Charles Eugene
20 Masselmani Mrs Fatima
22 Beesley Mr Lawrence
23 McGowan Miss Anna "Annie"
24 Sloper Mr William Thompson
26 Asplund Mrs Carl Oscar Selma Augusta Emilia Jo...
29 O'Dwyer Miss Ellen "Nellie"
32 Spencer Mrs William Augustus Marie Eugenie
33 Glynn Miss Mary Agatha
37 Mamee Mr Hanna
40 Nicola-Yarred Miss Jamila
44 Laroche Miss Simonne Marie Anne Andree
45 Devaney Miss Margaret Delia
48 O'Driscoll Miss Bridget
53 Harper Mrs Henry Sleeper Myna Haxtun
54 Faunthorpe Mrs Lizzie Elizabeth Anne Wilkinson
56 Woolner Mr Hugh
57 Rugg Miss Emily
59 West Miss Constance Mirium
62 Icard Miss Amelie
66 Moubarek Master Gerios
67 Nye Mrs Elizabeth Ramell
...
810 Chambers Mrs Norman Campbell Bertha Griggs
821 Hays Mrs Charles Melville Clara Jennings Gregg
822 Lulic Mr Nikola
824 Moor Mrs Beila
828 Mallet Master Andre
829 McCormack Mr Thomas Joseph
830 Stone Mrs George Nelson Martha Evelyn
831 Yasbeck Mrs Antoni Selini Alexander
832 Richards Master George Sibley
836 Compton Miss Sara Rebecca
839 Chip Mr Chang
840 Marechal Mr Pierre
843 Serepeca Miss Augusta
850 Goldenberg Mrs Samuel L Edwiga Grabowska
854 Lines Miss Mary Conover
856 Aks Mrs Sam Leah Rosen
857 Wick Mrs George Dennick Mary Hitchcock
858 Daly Mr Peter Denis
859 Baclini Mrs Solomon Latifa Qurban
863 Swift Mrs Frederick Joel Margaret Welles Barron
866 Bystrom Mrs Karolina
867 Duran y More Miss Asuncion
870 Johnson Master Harold Theodor
872 Beckwith Mrs Richard Leonard Sallie Monypeny

```
875                Abelson Mrs Samuel Hannah Wizosky
876                Najib Miss Adele Kiamie "Jane"
880                Potter Mrs Thomas Jr Lily Alexenia Wilson
881                Shelley Mrs William Imanita Parrish Hall
888                Graham Miss Margaret Edith
890                Behr Mr Karl Howell
```

```
Name: Name, dtype: object
```

```
In [ ]:
```

```
In [88]:
```

```
def split_names(names):
    return names.apply(split_name)
```

```
In [89]:
```

```
split_names(clear_name(people_Alive_df))
```

```
Out[89]:
```

```
PassengerId
2          (Cumings, Mrs, John, Bradley)
3          (Heikkinen, Miss, Laina,  )
4          (Futrelle, Mrs, Jacques, Heath)
9          (Johnson, Mrs, Oscar, W)
10         (Nasser, Mrs, Nicholas, Adele)
11         (Sandstrom, Miss, Marguerite, Rut)
12         (Bonnell, Miss, Elizabeth,  )
16         (Hewlett, Mrs, Mary, D)
18         (Williams, Mr, Charles, Eugene)
20         (Masselmani, Mrs, Fatima,  )
22         (Beesley, Mr, Lawrence,  )
23         (McGowan, Miss, Anna, "Annie")
24         (Sloper, Mr, William, Thompson)
26         (Asplund, Mrs, Carl, Oscar)
29         (O'Dwyer, Miss, Ellen, "Nellie")
32         (Spencer, Mrs, William, Augustus)
33         (Glynn, Miss, Mary, Agatha)
37         (Mamee, Mr, Hanna,  )
40         (Nicola-Yarred, Miss, Jamila,  )
44         (Laroche, Miss, Simonne, Marie)
45         (Devaney, Miss, Margaret, Delia)
48         (O'Driscoll, Miss, Bridget,  )
53         (Harper, Mrs, Henry, Sleeper)
54         (Faunthorpe, Mrs, Lizzie, Elizabeth)
56         (Woolner, Mr, Hugh,  )
57         (Rugg, Miss, Emily,  )
59         (West, Miss, Constance, Mirium)
62         (Icard, Miss, Amelie,  )
```

66 (Moubarek, Master, Gerios,)
67 (Nye, Mrs, Elizabeth, Ramell)
...
810 (Chambers, Mrs, Norman, Campbell)
821 (Hays, Mrs, Charles, Melville)
822 (Lulic, Mr, Nikola,)
824 (Moor, Mrs, Beila,)
828 (Mallet, Master, Andre,)
829 (McCormack, Mr, Thomas, Joseph)
830 (Stone, Mrs, George, Nelson)
831 (Yasbeck, Mrs, Antoni, Selini)
832 (Richards, Master, George, Sibley)
836 (Compton, Miss, Sara, Rebecca)
839 (Chip, Mr, Chang,)
840 (Marechal, Mr, Pierre,)
843 (Serepeca, Miss, Augusta,)
850 (Goldenberg, Mrs, Samuel, L)
854 (Lines, Miss, Mary, Conover)
856 (Aks, Mrs, Sam, Leah)
857 (Wick, Mrs, George, Dennick)
858 (Daly, Mr, Peter, Denis)
859 (Baclini, Mrs, Solomon, Latifa)
863 (Swift, Mrs, Frederick, Joel)
866 (Bystrom, Mrs, Karolina,)
867 (Duran, y, More, Miss)
870 (Johnson, Master, Harold, Theodor)
872 (Beckwith, Mrs, Richard, Leonard)
875 (Abelson, Mrs, Samuel, Hannah)
876 (Najib, Miss, Adele, Kiamie)
880 (Potter, Mrs, Thomas, Jr)
881 (Shelley, Mrs, William, Imanita)
888 (Graham, Miss, Margaret, Edith)
890 (Behr, Mr, Karl, Howell)
Name: Name, dtype: object

In [90]:

```
survival_df.groupby(['Survived', 'Parch']).size()
```

Out[90]:

Survived	Parch	
0	0	445
	1	53
	2	40
	3	2
	4	4
	5	4
	6	1
1	0	233
	1	65
	2	40
	3	3
	5	1

dtype: int64

In [91]:

```
def age_categorize(age):  
  
    if age>=90:  
        return '90+'  
    elif age>=80:  
        return '80-90'  
    elif age>=70:  
        return '70 -80'  
    elif age>=60:  
        return '60-70'  
    elif age>=50:  
        return '50-60'  
    elif age>=40:  
        return '40-50'  
    elif age>=30:  
        return '30-40'  
    elif age>=20:  
        return '20-30'  
    elif age>=10:  
        return '10-20'  
    elif age<10:  
        return '0-10'
```

In [92]:

```
age_categorize(38)
```

Out[92]:

'30-40'

In [93]:

```
def comp_age_categorize(data):  
    return data.apply(age_categorize)
```

In [94]:

In []:

In [95]:

```
Age_categorize_df.groupby('Age').size()
```

Out[95]:

```
Age  
0-10      62  
10-20    102  
20-30    220  
30-40    167  
40-50     89  
50-60     48  
60-70     19  
70 -80      6  
80-90      1  
dtype: int64
```

In [96]:

```
Age_categorize_df.groupby('Age').get_group('0-10')
```

Out[96]:

	Age
PassengerId	
8	0-10
11	0-10
17	0-10
25	0-10

23	0-10
44	0-10
51	0-10
59	0-10
64	0-10
79	0-10
120	0-10
148	0-10
165	0-10
166	0-10
172	0-10
173	0-10
183	0-10
184	0-10
185	0-10
194	0-10
206	0-10
234	0-10
238	0-10
262	0-10
279	0-10
298	0-10
306	0-10
341	0-10
349	0-10
375	0-10
382	0-10
...	...
446	0-10
449	0-10
470	0-10
480	0-10

480	0-10
481	0-10
490	0-10
531	0-10
536	0-10
542	0-10
550	0-10
619	0-10
635	0-10
643	0-10
645	0-10
692	0-10
721	0-10
751	0-10
752	0-10
756	0-10
778	0-10
788	0-10
789	0-10
804	0-10
814	0-10
825	0-10
828	0-10
832	0-10
851	0-10
853	0-10
870	0-10

62 rows × 1 columns

In [97]:

```
not_alive_df=survival_df.groupby('Survived').get_group(0)
print not_alive_df
```

PassengerId	Survived	Pclass	\
1	0	3	
5	0	3	
6	0	3	
7	0	1	
8	0	3	
13	0	3	
14	0	3	
15	0	3	
17	0	3	
19	0	3	
21	0	2	
25	0	3	
27	0	3	
28	0	1	
30	0	3	
31	0	1	
34	0	2	
35	0	1	

In [98]:

```
print Alive_df
```

PassengerId	Survived	Pclass	\
2	1	1	
3	1	3	
4	1	1	
9	1	3	
10	1	2	
11	1	3	
12	1	1	
16	1	2	
18	1	2	
20	1	3	
22	1	2	
23	1	3	
24	1	1	
26	1	3	
29	1	3	
32	1	1	
33	1	3	
37	1	2	

In [99]:

```
Age_alive=pd.DataFrame.from_dict(comp_age_categorize(Alive_df['Age']))
Age_not_alive=pd.DataFrame.from_dict(comp_age_categorize(not_alive_df['Age']))
```

In [100]:

```
print Age_alive.groupby('Age').size()
print Age_not_alive.groupby('Age').size()
```

```
Age
0-10      38
10-20     41
20-30     77
30-40     73
40-50     34
50-60     20
60-70      6
80-90      1
dtype: int64
Age
0-10      24
10-20     61
20-30    143
30-40     94
40-50     55
50-60     28
60-70     13
70 -80      6
dtype: int64
```

In [101]:

```
x=pd.DataFrame.from_dict({'Survived': Age_alive.groupby('Age').size()})
y=pd.DataFrame.from_dict({'Lost': Age_not_alive.groupby('Age').size()})
```

In []:

In [102]:

```
z=pd.concat([x,y],axis=1)
print z
z.sum()
```

	Survived	Lost
0-10	38.0	24.0
10-20	41.0	61.0
20-30	77.0	143.0
30-40	73.0	94.0
40-50	34.0	55.0
50-60	20.0	28.0
60-70	6.0	13.0
70 -80	NaN	6.0
80-90	1.0	NaN

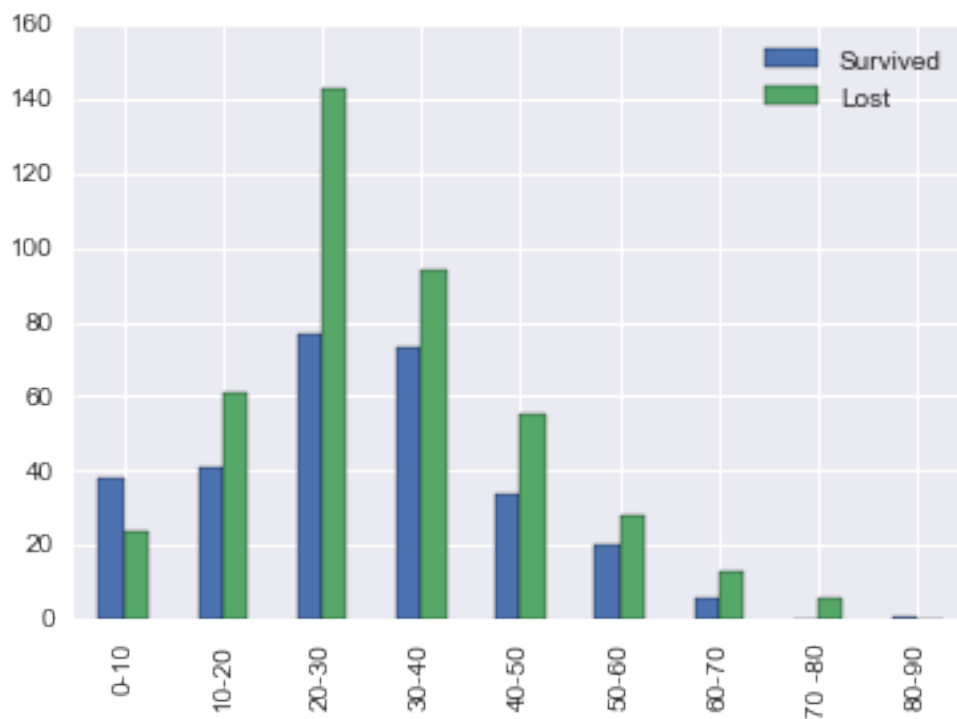
Out[102]:

```
Survived    290.0
Lost        424.0
dtype: float64
```

In [103]:

```
%matplotlib inline
```

```
z.plot.bar()
plt.show()
```



In [104]:

```
Alive_df.groupby(['SibSp', 'Parch']).get_group((0,0)).sum()['Survived']
```

Out[104]:

163.0

In [105]:

```
len(not_alive_df.groupby(['SibSp', 'Parch']).get_group((0,0)))
```

Out[105]:

374

In []:

In [106]:

```
survival_df.groupby('Age').get_group(5)
```

Out[106]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
PassengerId										
59	1	2	West, Miss. Constance Mirium	female	5.0	1	2	C.A. 34651	27.7500	N
234	1	3	Asplund, Miss. Lillian Gertrud	female	5.0	4	2	347077	31.3875	N
449	1	3	Baclini, Miss. Marie Catherine	female	5.0	2	1	2666	19.2583	N
778	1	3	Emanuel, Miss. Virginia Ethel	female	5.0	0	0	364516	12.4750	N

In [107]:

```
survival_df.groupby(['Embarked', 'Survived']).size()
```

Out[107]:

Embarked	Survived	
C	0	75
	1	93
Q	0	47
	1	30
S	0	427
	1	217

dtype: int64

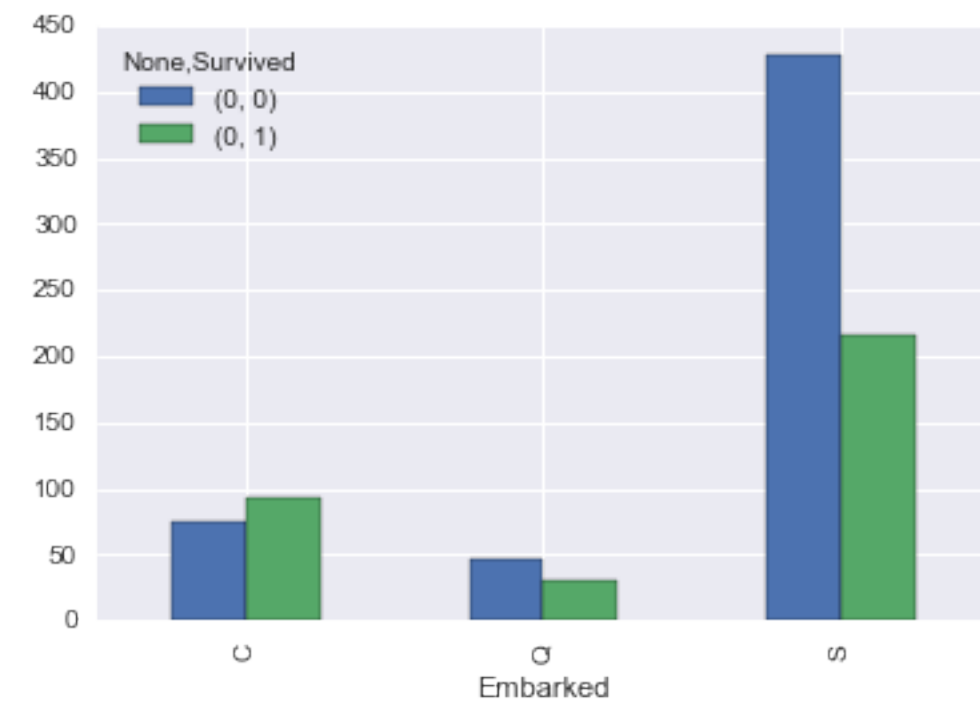
In [108]:

```
Embarked=pd.DataFrame.from_dict(survival_df.groupby(['Embarked', 'Survived']).size()  
print Embarked  
Embarked.plot.bar()
```

	0	1
Survived	0	1
Embarked		
C	75	93
Q	47	30
S	427	217

Out[108]:

<matplotlib.axes._subplots.AxesSubplot at 0x11d534910>



In [126]:

```
def use_survived_lost(survived):  
  
    if survived==0:  
        return 'LostLife'  
    elif survived==1:  
        return 'Survived'
```

In [133]:

```
use_survived_lost(1)
```

Out[133]:

```
'Survived'
```

In [131]:

```
passenger_class= pd.DataFrame.from_dict({'Alive':survival_df.groupby(['Pclass','Survived'])  
print passenger_class
```

	Alive	
Survived	0	1
Pclass		
1	80	136
2	97	87
3	372	119

In [132]:

```
def DF_survived_lost(data):  
    return data.apply(use_survived_lost)  
print DF_survived_lost(survival_df['Survived'])
```

PassengerId	
1	LostLife
2	Survived
3	Survived
4	Survived
5	LostLife
6	LostLife
7	LostLife
8	LostLife
9	Survived
10	Survived
11	Survived
12	Survived
13	LostLife
14	LostLife
15	LostLife
16	Survived


```
17      LostLife
18      Survived
19      LostLife
20      Survived
21      LostLife
22      Survived
23      Survived
24      Survived
25      LostLife
26      Survived
27      LostLife
28      LostLife
29      Survived
30      LostLife

...
862     LostLife
863     Survived
864     LostLife
865     LostLife
866     Survived
867     Survived
868     LostLife
869     LostLife
870     Survived
871     LostLife
872     Survived
873     LostLife
874     LostLife
875     Survived
876     Survived
877     LostLife
878     LostLife
879     LostLife
880     Survived
881     Survived
882     LostLife
883     LostLife
884     LostLife
885     LostLife
886     LostLife
887     LostLife
888     Survived
889     LostLife
890     Survived
891     LostLife
Name: Survived, dtype: object
```

In []:

In [115]:

```
passclass = pd.Series(survival_df.groupby(['Pclass', 'Survived']).size()).unstack()  
  
print passclass
```

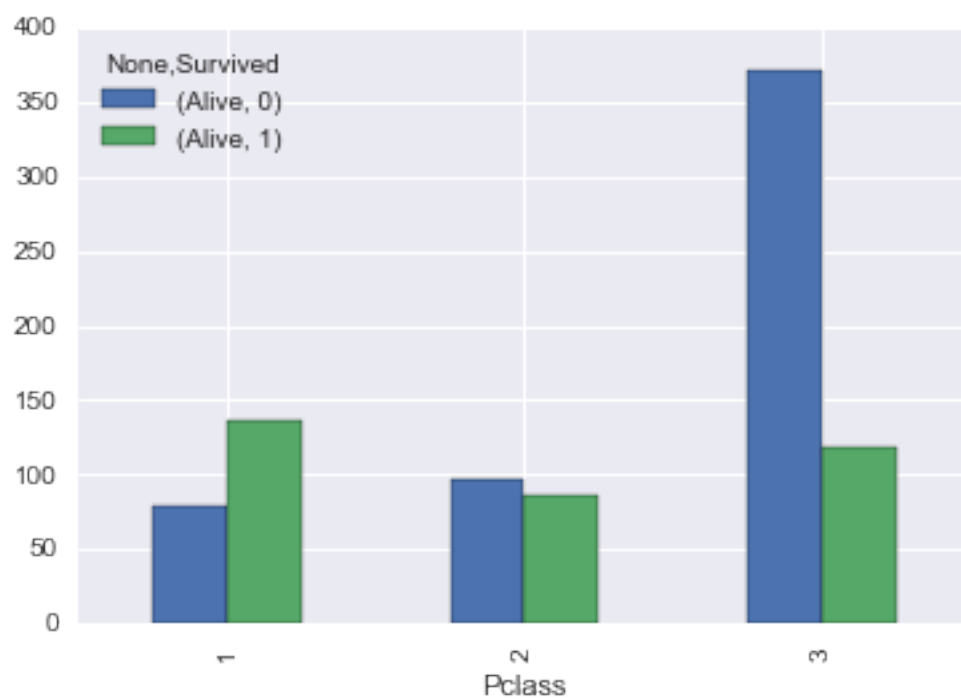
```
Survived    0    1  
Pclass  
1           80   136  
2           97    87  
3          372   119
```

In [111]:

```
passenger_class.plot.bar()
```

Out[111]:

<matplotlib.axes._subplots.AxesSubplot at 0x11d89bc90>



In []:

```
def split_name(data):

    full_name=data.split(" ")

    last_name=full_name[0]
    salutation=full_name[1]
    first_name_1=full_name[2]
    if len(full_name)==3:
        first_name_2= ' '
    else:
        first_name_2=full_name[3]

    return salutation

print split_name("Lines Miss Mary Conover")
```

In []:

```
def split_names(names):
    return names.apply(split_name)
```

In [2]:

```
salutations = pd.DataFrame.from_dict({'Salutations':split_names(Raw_survival_df['Name'])})
```

```
-----
-----
NameError                                Traceback (most recent call
last)
<ipython-input-2-3e8ab3f5049d> in <module>()
----> 1 salutations = Pd.DataFrame.from_dict({'Salutations':split_name
s(Raw_survival_df['Name'])})
```

NameError: name 'Pd' is not defined

In []:

```
print salutations.tail()
vv=salutations.groupby('Salutations').size()

print vv
```

In []:

```
import heapq

def get_others(data):
    others=0
    if (data< 7):
        others +=data

    return others
```

In []:

```
def get_top6(data):
    return data.apply(get_others).sum()

s1= pd.Series(heapq.nlargest(5,vv),index=[ 'Mr' , 'Miss' , 'Mrs' , 'Master' , 'Dr' ],name='Salutations')

s2=pd.Series([get_top6(vv)],index=[ 'others' ],name='Salutations')
```

In []:

```
Top6 =s1.append(s2)
print Top6
```

In []:

```
Top6.plot.pie(subplots=True)
```