Investigatoion On Titanic Dataset

```
As a part of data analysis my intreast to survey on the titanic passenger information of 891 peopele from 2224 passenger
s.The given data contains PassengerId,Name,Pcalss (it is a socio-economic status of passenger belongs to 1/2/3 classe
s.),sex,age,sibsp & parch(these two says number of pepole parent/childeren/spouse staying in abroad of the passenger),ti
cket number as ticket,fare,cabin,embarked(Port of Embarkation(C=Cherbourg;Q=Queenstown;S=Southampton)and it also contain
 whether the passenger survived or not (0=no,1=yes). Here we can notice that most of the people were died in titanic, No
w here our survey is based two main questions?:
1)How many people are Surviveded from different Pclasses?
2)For each  person how many of there relations are in abroad?
```

In the below cell we are calling the libraries, to investigate the data. Most likely we are going to use the pandas for these project.

In [127]:
```
import plotly
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import plotly.tools as tls
import csv
import plotly.plotly as py
import plotly.graph_objs as go
```

To investigate our data I'm reading the files from folder where they are located using the read command in pandas and the files are saved in csv format.

In [120]:
```
train_df=pd.read_csv("C:/Users/midid/Desktop/dataanlysis/project_titanic/train.csv") #survey is done on these file
test_df=pd.read_csv("C:/Users/midid/Desktop/dataanlysis/project_titanic/test_2.1.csv") #having look at these file
gender_submission_df=pd.read_csv("C:/Users/midid/Desktop/dataanlysis/project_titanic/gender_submission_2.2.csv")
```

The below cells describes the data we are going to survey. First takig look at Embarkation and the number of people survived, it explains about the people who are entered form three diffrent sations and there socio-economic classes.

In [57]:
```
train_df #survey data
```

Out[57]:

	PassengerId	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
Pclass											
3	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
1	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
3	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
3	6	0	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
1	7	0	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
3	8	0	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
3	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
2	10	1	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
3	11	1	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
1	12	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
3	13	0	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
3	14	0	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S
3	15	0	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S
2	16	1	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000	NaN	S
3	17	0	Rice, Master. Eugene	male	2.0	4	1	382652	29.1250	NaN	Q
2	18	1	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S
3	19	0	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0000	NaN	S
3	20	1	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250	NaN	C
2	21	0	Fynney, Mr. Joseph J	male	35.0	0	0	239865	26.0000	NaN	S
2	22	1	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.0000	D56	S
3	23	1	McGowan, Miss. Anna "Annie"	female	15.0	0	0	330923	8.0292	NaN	Q
1	24	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5000	A6	S
3	25	0	Palsson, Miss. Torborg Danira	female	8.0	3	1	349909	21.0750	NaN	S
3	26	1	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38.0	1	5	347077	31.3875	NaN	S
3	27	0	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250	NaN	C
1	28	0	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.0000	C23 C25 C27	S
3	29	1	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792	NaN	Q
3	30	0	Todoroff, Mr. Lalio	male	NaN	0	0	349216	7.8958	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...
2	862	0	Giles, Mr. Frederick Edward	male	21.0	1	0	28134	11.5000	NaN	S
1	863	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48.0	0	0	17466	25.9292	D17	S
3	864	0	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500	NaN	S
2	865	0	Gill, Mr. John William	male	24.0	0	0	233866	13.0000	NaN	S
2	866	1	Bystrom, Mrs. (Karolina)	female	42.0	0	0	236852	13.0000	NaN	S

	PassengerId	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
Pclass											
2	867	1	Duran y More, Miss. Asuncion	female	27.0	1	0	SC/PARIS 2149	13.8583	NaN	C
1	868	0	Roebling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50.4958	A24	S
3	869	0	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000	NaN	S
3	870	1	Johnson, Master. Harold Theodor	male	4.0	1	1	347742	11.1333	NaN	S
3	871	0	Balkic, Mr. Cerin	male	26.0	0	0	349248	7.8958	NaN	S
1	872	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
1	873	0	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S
3	874	0	Vander Cruyssen, Mr. Victor	male	47.0	0	0	345765	9.0000	NaN	S
2	875	1	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	NaN	C
3	876	1	Najib, Miss. Adele Kiamie "Jane"	female	15.0	0	0	2667	7.2250	NaN	C
3	877	0	Gustafsson, Mr. Alfred Ossian	male	20.0	0	0	7534	9.8458	NaN	S
3	878	0	Petroff, Mr. Nedelio	male	19.0	0	0	349212	7.8958	NaN	S
3	879	0	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	NaN	S
1	880	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C
2	881	1	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
3	882	0	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	NaN	S
3	883	0	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S
2	884	0	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
3	885	0	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
3	886	0	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
2	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
1	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
3	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
1	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
3	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 11 columns

In [36]:
```
train_df.describe()
```

Out[36]:

	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 [79]:
```
train_df['Sex']
```

Out[79]:
```
0       male
1     female
2     female
3     female
4       male
5       male
6       male
7       male
8     female
9     female
10    female
11    female
12      male
13      male
14    female
15    female
16      male
17      male
18    female
19    female
20      male
21      male
22    female
23      male
24    female
25    female
26      male
27      male
28    female
29      male
       ...
861     male
862   female
863   female
864     male
865   female
866   female
867     male
868     male
869     male
870     male
871   female
872     male
873     male
874   female
875   female
876     male
877     male
878     male
879   female
880   female
881     male
882   female
883     male
884     male
885   female
886     male
887   female
888   female
889     male
890     male
Name: Sex, dtype: object
```

From the above data we can observe that in test_df the column Survived is seperately noted in other file named as gender_submission_df so, now we are merging the data for two files.

In [5]:
```
test_gender = test_df.merge(gender_submission_df, on='PassengerId', how='outer')
```

In [7]:
```
test_gender.head()
```

Out[7]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Survived
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q	0
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S	1
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q	0
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S	0
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S	1

In the above cell shows some of the cells(small data frame) which we merged with the missing column in the outer layer.

we are looking only only 891 people as per the given in project details(data in train_df).Before doing these we are going to group the passengers according to there stages and adding them. And also have a look about the fare of the stations.

In [ ]:
```
train_group_df=train_df.groupby('Embarked')    #grouping the people accoring to there port of entries
```

In [174]:
```
print train_group_df.sum()['Survived']
```

```
Embarked
C     93
Q     30
S    217
Name: Survived, dtype: int64
```

In [175]:
```
print train_group_df.mean()['Fare'] #average of ticket fare for three different port of embarkations.
```

```
Embarked
C    59.954144
Q    13.276030
S    27.079812
Name: Fare, dtype: float64
```

At these movement we can notice that at the port of 's' the cost of a ticket is very less and the survived passengers are high. By comparing other two stations at 'c' the number of people entered higher then station 'Q' and also there fare of a station 'c' ticket is more then the station 'Q'.So, we can note that most of the people prefer to travell form port of embarktion at 'S. And most of them survived those who entered in station 's'.

Lets have look by grouping the passengers by there pclass, as shown in the below cases it explains us most of them belongs to the 1st calss and secondly from the 3rd calss so, very less passengers are from the 2 calss are survived. As shown in the histogram between port of embarkation and Pclass there mean we acn see that dark color lines most of them survived in the 1st class then the other classes.

In [15]:
```
train_group_by_pclass=train_df.groupby('Pclass') #grouping the people according to Pclass.
```

In [16]:
```
print train_group_by_pclass.sum()['Survived']  #printing how many are survived in each class.
```

```
Pclass
1    136
2     87
3    119
Name: Survived, dtype: int64
```

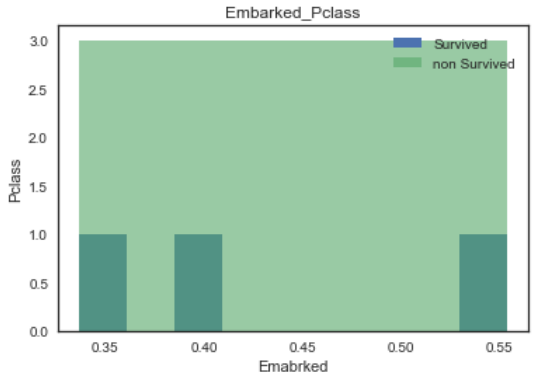
In [ ]:
```

```

In [96]:
```
pclass_embarked=train_df.groupby('Embarked')['Pclass','Survived'].mean()   ##plotting a histogram between Pclass an
d
                                                                                #   survived by taking there average
embarked_plot=pclass_embarked['Survived']                                  # and grouped them according to there emabarkat
ion


plt.hist(embarked_plot, bins=9,label="Survived")
plt.hist(embarked_plot, bins = 1,label="non Survived", alpha=.6)
plt.legend(loc='upper right')
plt.xlabel("Emabrked")
plt.ylabel("Pclass")
sns.set_style('white')
plt.title("Embarked_Pclass")
```

Out[96]: <matplotlib.text.Text at 0xbfb7ac8>



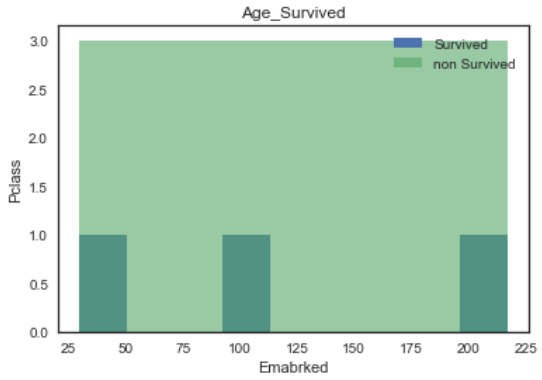
In [95]:
```
pclass_embarked=train_df.groupby('Embarked')['Age','Survived'].sum()     #plotting a histogram between age and
                                                                            #   survived by taking there sum
                                                                          # and grouped them according to there emabarkat
ion

embarked_plot=pclass_embarked['Survived']



plt.hist(embarked_plot, bins=9,label="Survived")
plt.hist(embarked_plot, bins = 1,label="non Survived", alpha=.6)
plt.legend(loc='upper right')
plt.xlabel("Emabrked")
plt.ylabel("Pclass")
sns.set_style('white')
plt.title("Age_Survived")
```

Out[95]: <matplotlib.text.Text at 0xc1a6c18>



Now let us see how many dependers of these passengers are having the dependers in the abroad like parents, childrean, siblings and spouse by investigatting them in using multiple angles. in the below we can see a short look that 340 parents/children and 466 siblings/spouse people in abroad have realation with 891 passengers,in total 797 people are in abroad and we can also say that 6 parents/children , 8 siblings/spouse are belongs to one or two persons from the described data in the above cell.

In [101]:
```
print train_df.sum()['Parch']   #sum of parents/children in abroad of the passengers
```

```
340.0
```

In [100]:
```
print train_df.sum()['SibSp']      #sum of sibling/'spouse in abroad of the passengers
```

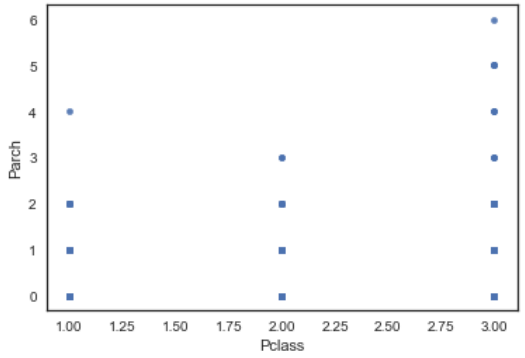
```
466.0
```

In [82]:
```
print (total_in_abroad(train_df['SibSp'],train_df['Parch'])).sum()   #sum of people who are in abroad of the passengers
```

```
797.0
```

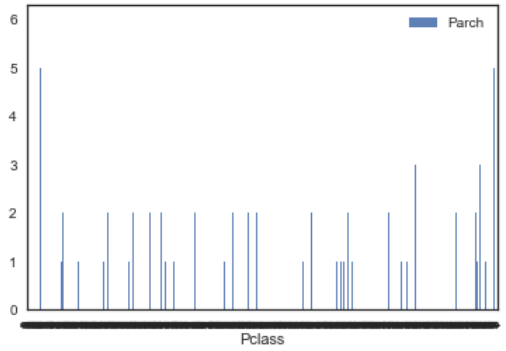
In [110]:
```
train_df.plot.scatter('Pclass', 'Parch',alpha=0.9) #plot between realted people in abroad
```

Out[110]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xc447b70>
```



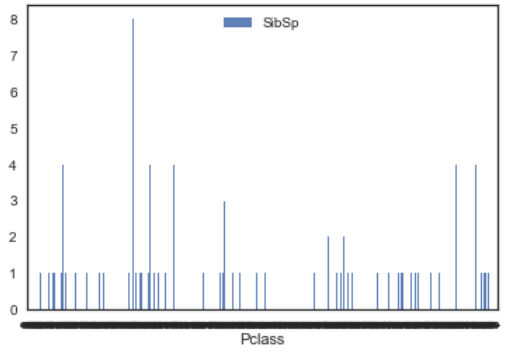
In [111]:
```
train_df.plot.bar('Pclass', 'Parch',alpha=0.9) #bar graph between pclass and parch
```

Out[111]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xcd0c198>
```



In [165]:
```
train_df.plot.bar('Pclass', 'SibSp',alpha=0.9) # plot between pclass and sibsp
```

Out[165]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x258078d0>
```



In [84]:
```
def total_in_abroad(SibSp,Parch):    #caluclating reltaed persons who are in abord accoring to there embarkation.
    return (SibSp.shift(7)+Parch.shift(8)).groupby(train_df['Embarked'])
print (total_in_abroad(train_df['SibSp'],train_df['Parch'])).sum()
```

```
Embarked
C    147.0
Q     62.0
S    588.0
dtype: float64
```

In [86]:
```
def total_in_abroad(SibSp,Parch):     #caluclating reltaed persons who are in abord accoring to there classes.
    return (SibSp.shift(7)+Parch.shift(8)).groupby(train_df['Pclass'])
print (total_in_abroad(train_df['SibSp'],train_df['Parch'])).sum()
```

```
Pclass
1    223.0
2    177.0
3    397.0
dtype: float64
```

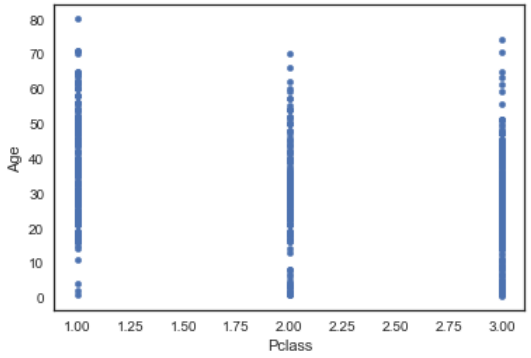
In [116]:
```
train_df.groupby('Pclass')['Fare'].mean() # cost of ticket according to the class.
```

Out[116]:
```
Pclass
1    84.154687
2    20.662183
3    13.675550
Name: Fare, dtype: float64
```

Most of the related abroad people passeger belongs to 3rd class and prefer to enter from the 's' enterence and seconderly belongs to 1st class and the people of age above 10 to below 70 years pefer to 1st calss there fare is more then the 3rd class people thosewhose age is high. In second class from age 10 to 60 year people is are plotted in sctter graph those average fare is less compare to the 1st class. and coming to the 3rd class people most of the below 10 year and below 50 year people are travelled in it there average fare is very less compare two other classes because most of them are children.

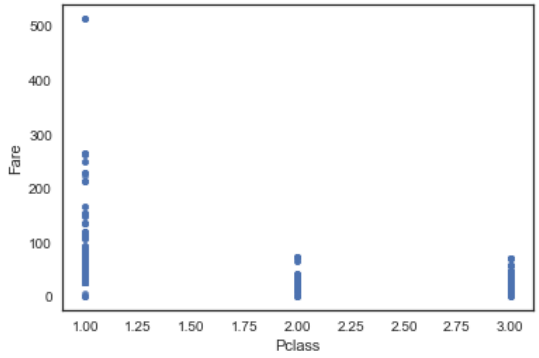
In [130]:
```
train_df.plot.scatter('Pclass', 'Age')  #plot between pclass and age.
```

Out[130]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x161d3358>
```



In [157]:
```
train_df.plot.scatter('Pclass', 'Fare',alpha=0.9) #cost of pclasses.
```

Out[157]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x16b930b8>
```



To survey the data we done in different ways by taking there a bar and sctter graphs between age,pclass,fare and the relations in abroad. we can say that most of them are travelled in 1st class in betweeen age above 10 to below 65 years are survived and in in the 3rd class most of them are chlidren (we can say by using down table) and below 50 age people travlled are some of them alive. Coming to the 2nd class people very less passengers age in between 10 to 60 years old are travelled they hardly survived also there fare is also very low compare to the other two classes.

In [115]:
```
train_df[train_df['Age'].isnull()]
```

Out[115]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250	NaN	C
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250	NaN	C
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792	NaN	Q
29	30	0	3	Todoroff, Mr. Lalio	male	NaN	0	0	349216	7.8958	NaN	S
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	NaN	1	0	PC 17569	146.5208	B78	C
32	33	1	3	Glynn, Miss. Mary Agatha	female	NaN	0	0	335677	7.7500	NaN	Q
36	37	1	3	Mamee, Mr. Hanna	male	NaN	0	0	2677	7.2292	NaN	C
42	43	0	3	Kraeff, Mr. Theodor	male	NaN	0	0	349253	7.8958	NaN	C
45	46	0	3	Rogers, Mr. William John	male	NaN	0	0	S.C./A.4. 23567	8.0500	NaN	S
46	47	0	3	Lennon, Mr. Denis	male	NaN	1	0	370371	15.5000	NaN	Q
47	48	1	3	O'Driscoll, Miss. Bridget	female	NaN	0	0	14311	7.7500	NaN	Q
48	49	0	3	Samaan, Mr. Youssef	male	NaN	2	0	2662	21.6792	NaN	C
55	56	1	1	Woolner, Mr. Hugh	male	NaN	0	0	19947	35.5000	C52	S
64	65	0	1	Stewart, Mr. Albert A	male	NaN	0	0	PC 17605	27.7208	NaN	C
65	66	1	3	Moubarek, Master. Gerios	male	NaN	1	1	2661	15.2458	NaN	C
76	77	0	3	Staneff, Mr. Ivan	male	NaN	0	0	349208	7.8958	NaN	S
77	78	0	3	Moutal, Mr. Rahamin Haim	male	NaN	0	0	374746	8.0500	NaN	S
82	83	1	3	McDermott, Miss. Brigdet Delia	female	NaN	0	0	330932	7.7875	NaN	Q
87	88	0	3	Slocovski, Mr. Selman Francis	male	NaN	0	0	SOTON/OQ 392086	8.0500	NaN	S
95	96	0	3	Shorney, Mr. Charles Joseph	male	NaN	0	0	374910	8.0500	NaN	S
101	102	0	3	Petroff, Mr. Pastcho ("Pentcho")	male	NaN	0	0	349215	7.8958	NaN	S
107	108	1	3	Moss, Mr. Albert Johan	male	NaN	0	0	312991	7.7750	NaN	S
109	110	1	3	Moran, Miss. Bertha	female	NaN	1	0	371110	24.1500	NaN	Q
121	122	0	3	Moore, Mr. Leonard Charles	male	NaN	0	0	A4. 54510	8.0500	NaN	S
126	127	0	3	McMahon, Mr. Martin	male	NaN	0	0	370372	7.7500	NaN	Q
128	129	1	3	Peter, Miss. Anna	female	NaN	1	1	2668	22.3583	F E69	C
140	141	0	3	Boulos, Mrs. Joseph (Sultana)	female	NaN	0	2	2678	15.2458	NaN	C
154	155	0	3	Olsen, Mr. Ole Martin	male	NaN	0	0	Fa 265302	7.3125	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
718	719	0	3	McEvoy, Mr. Michael	male	NaN	0	0	36568	15.5000	NaN	Q
727	728	1	3	Mannion, Miss. Margareth	female	NaN	0	0	36866	7.7375	NaN	Q
732	733	0	2	Knight, Mr. Robert J	male	NaN	0	0	239855	0.0000	NaN	S
738	739	0	3	Ivanoff, Mr. Kanio	male	NaN	0	0	349201	7.8958	NaN	S
739	740	0	3	Nankoff, Mr. Minko	male	NaN	0	0	349218	7.8958	NaN	S
740	741	1	1	Hawksford, Mr. Walter James	male	NaN	0	0	16988	30.0000	D45	S
760	761	0	3	Garfirth, Mr. John	male	NaN	0	0	358585	14.5000	NaN	S

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
766	767	0	1	Brewe, Dr. Arthur Jackson	male	NaN	0	0	112379	39.6000	NaN	C
768	769	0	3	Moran, Mr. Daniel J	male	NaN	1	0	371110	24.1500	NaN	Q
773	774	0	3	Elias, Mr. Dibo	male	NaN	0	0	2674	7.2250	NaN	C
776	777	0	3	Tobin, Mr. Roger	male	NaN	0	0	383121	7.7500	F38	Q
778	779	0	3	Kilgannon, Mr. Thomas J	male	NaN	0	0	36865	7.7375	NaN	Q
783	784	0	3	Johnston, Mr. Andrew G	male	NaN	1	2	W./C. 6607	23.4500	NaN	S
790	791	0	3	Keane, Mr. Andrew "Andy"	male	NaN	0	0	12460	7.7500	NaN	Q
792	793	0	3	Sage, Miss. Stella Anna	female	NaN	8	2	CA. 2343	69.5500	NaN	S
793	794	0	1	Hoyt, Mr. William Fisher	male	NaN	0	0	PC 17600	30.6958	NaN	C
815	816	0	1	Fry, Mr. Richard	male	NaN	0	0	112058	0.0000	B102	S
825	826	0	3	Flynn, Mr. John	male	NaN	0	0	368323	6.9500	NaN	Q
826	827	0	3	Lam, Mr. Len	male	NaN	0	0	1601	56.4958	NaN	S
828	829	1	3	McCormack, Mr. Thomas Joseph	male	NaN	0	0	367228	7.7500	NaN	Q
832	833	0	3	Saad, Mr. Amin	male	NaN	0	0	2671	7.2292	NaN	C
837	838	0	3	Sirota, Mr. Maurice	male	NaN	0	0	392092	8.0500	NaN	S
839	840	1	1	Marechal, Mr. Pierre	male	NaN	0	0	11774	29.7000	C47	C
846	847	0	3	Sage, Mr. Douglas Bullen	male	NaN	8	2	CA. 2343	69.5500	NaN	S
849	850	1	1	Goldenberg, Mrs. Samuel L (Edwiga Grabowska)	female	NaN	1	0	17453	89.1042	C92	C
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292	NaN	C
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500	NaN	S
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000	NaN	S
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	NaN	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

177 rows × 12 columns

In [107]:
```
train_df.groupby('Embarked')['Pclass','Survived'].sum()
```

Out[107]:

	Pclass	Survived
Embarked		
C	317	93
Q	224	30
S	1514	217

In [71]:
```
print train_df.groupby('Age').sum()['Survived']  #it shows how many passengers belong to certain age
```

```
Age
0.42      1
0.67      1
0.75      2
0.83      2
0.92      1
1.00      5
2.00      3
3.00      5
4.00      7
5.00      4
6.00      2
7.00      1
8.00      2
9.00      2
10.00     0
11.00     1
12.00     1
13.00     2
14.00     3
14.50     0
15.00     4
16.00     6
17.00     6
18.00     9
19.00     9
20.00     3
20.50     0
21.00     5
22.00    11
23.00     5
         ..
44.00     3
45.00     5
45.50     0
46.00     0
47.00     1
48.00     6
49.00     4
50.00     5
51.00     2
52.00     3
53.00     1
54.00     3
55.00     1
55.50     0
56.00     2
57.00     0
58.00     3
59.00     0
60.00     2
61.00     0
62.00     2
63.00     2
64.00     0
65.00     0
66.00     0
70.00     0
70.50     0
71.00     0
74.00     0
80.00     1
Name: Survived, dtype: int64
```

Now making a realation between two variables and by standardizing them to x-axis and y-axis seeing there relation between them.

In [121]:
```
def correlation(x,y):
     return ((x-x.mean())*(y-y.mean())).mean()/x.std(ddof=0)/y.std(ddof=0)


age = train_df['Age']
Survived=train_df['Survived']
Embarked=train_df['Embarked']
Fare=train_df['Fare']
Pclass=train_df['Pclass']
```

In [38]:
```
print correlation(age,Survived)
```

```
-0.0779826784139
```

In [39]:
```
print correlation(Survived,Pclass)
```

```
-0.338481035961
```

In [126]:
```
print correlation(Pclass,Fare)
```

```
-0.549499619944
```

From the above correlation it can be concluded that all the realations are below the standard deviation so, passengers survived are very hardly. In our survey given there are three port of Embarkation(entrise) they are c,q,s here,most of the people preffred to come form the port s the average fare form 's' is moderate then the other two and 217 are survived those who came from thesse port,From 'c' port 93 people were survived and the average fare is more then other two. And 30 people survived those who entered from port 'q' wnd there average fare is very less compare to the other two cases.
Most of the children and below 50 years old people are in 3rd class, there realations with abrod people is almost 397 and 119 people were survived.Seconderly, in the 1st class 10-65 old people travlled 136 are survived and approxmatley 223 relatives are in abroad. Atlast coming to the 2nd class 10-60 age passengers are travelling in that 136 are alive and have 223 abroad people having relationship with these passengers.

From the above survey we can predict that most people of age 10-65 years age are in 1st class has there port of embarkation is 'S',there ticket fare is modarate,those are survived and most of them havig there parents/children or sibling/spouse are in abroad.