

In [1]:

```
import numpy as np
import pandas as pd
B=pd.read_csv("C:/Users/TERIAK-JB/Desktop/GOMYCODE/titanic.csv",encoding ="iso-8859-1")
B.head(10)
```

Out[1]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
5	1.0	1.0	Anderson, Mr. Harry	male	48.0000	0.0	0.0	19952	26.5500	E12	S	3	NaN	New York, NY
6	1.0	1.0	Andrews, Miss. Kornelia Theodosia	female	63.0000	1.0	0.0	13502	77.9583	D7	S	10	NaN	Hudson, NY
7	1.0	0.0	Andrews, Mr. Thomas Jr	male	39.0000	0.0	0.0	112050	0.0000	A36	S	NaN	NaN	Belfast, NI
8	1.0	1.0	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0000	2.0	0.0	11769	51.4792	C101	S	D	NaN	Bayside, Queens, NY
9	1.0	0.0	Artagaveytia, Mr. Ramon	male	71.0000	0.0	0.0	PC 17609	49.5042	NaN	C	NaN	22.0	Montevideo, Uruguay

In [2]:

```
B.tail(13)
```

Out[2]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
1297	3.0	0.0	Wiseman, Mr. Phillippe	male	NaN	0.0	0.0	A/4. 34244	7.2500	NaN	S	NaN	NaN	NaN
1298	3.0	0.0	Wittevrongel, Mr. Camille	male	36.0	0.0	0.0	345771	9.5000	NaN	S	NaN	NaN	NaN
1299	3.0	0.0	Yasbeck, Mr. Antoni	male	27.0	1.0	0.0	2659	14.4542	NaN	C	C	NaN	NaN
1300	3.0	1.0	Yasbeck, Mrs. Antoni (Selini Alexander)	female	15.0	1.0	0.0	2659	14.4542	NaN	C	NaN	NaN	NaN
1301	3.0	0.0	Youseff, Mr. Gerious	male	45.5	0.0	0.0	2628	7.2250	NaN	C	NaN	312.0	NaN
1302	3.0	0.0	Yousif, Mr. Wazli	male	NaN	0.0	0.0	2647	7.2250	NaN	C	NaN	NaN	NaN
1303	3.0	0.0	Yousseff, Mr. Gerious	male	NaN	0.0	0.0	2627	14.4583	NaN	C	NaN	NaN	NaN
1304	3.0	0.0	Zabour, Miss. Hileni	female	14.5	1.0	0.0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3.0	0.0	Zabour, Miss. Thamine	female	NaN	1.0	0.0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5	0.0	0.0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0	0.0	0.0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0	0.0	0.0	315082	7.8750	NaN	S	NaN	NaN	NaN
1309	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [3]:

```
B.dropna(axis=0,how='all',inplace=True)
```

In [4]:

In [4]:

B

Out[4]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3.0	0.0	Zabour, Miss. Hileni	female	14.5000	1.0	0.0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3.0	0.0	Zabour, Miss. Thamine	female	NaN	1.0	0.0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5000	0.0	0.0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0000	0.0	0.0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0000	0.0	0.0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [5]:

B.isnull().sum()

Out[5]:

pclass 0  
survived 0  
name 0  
sex 0  
age 263  
sibsp 0  
parch 0  
ticket 0  
fare 1  
cabin 1014  
embarked 2  
boat 823  
body 1188  
home.dest 564  
dtype: int64

In [6]:

B["age"].fillna(B["age"].mean(),inplace=True)  
B["age"]=B["age"].apply(np.int)  
B

Out[6]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29	0.0	0.0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0	1.0	2.0	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON

	pclass	survived			name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	
3	1.0	0.0	Allison, Mr. Hudson		Joshua Creighton	male	30	1.0	2.0	113781	151.5500	C22 C26		S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1.0	0.0	Allison, Mrs. Hudson		J C (Bessie Waldo Daniels)	female	25	1.0	2.0	113781	151.5500	C22 C26		S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...			...	...	...	...	...	...	...	...		...	...	...	...
1304	3.0	0.0			Zabour, Miss. Hileni	female	14	1.0	0.0	2665	14.4542	NaN		C	NaN	328.0	NaN
1305	3.0	0.0			Zabour, Miss. Thamine	female	29	1.0	0.0	2665	14.4542	NaN		C	NaN	NaN	NaN
1306	3.0	0.0			Zakarian, Mr. Mapriededer	male	26	0.0	0.0	2656	7.2250	NaN		C	NaN	304.0	NaN
1307	3.0	0.0			Zakarian, Mr. Ortin	male	27	0.0	0.0	2670	7.2250	NaN		C	NaN	NaN	NaN
1308	3.0	0.0			Zimmerman, Mr. Leo	male	29	0.0	0.0	315082	7.8750	NaN		S	NaN	NaN	NaN

1309 rows × 14 columns

In [7]:

```
B["survived"]=B["survived"].apply(np.int)
B
```

Out[7]:

	pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	
0	1.0	1		Allen, Miss. Elisabeth Walton	female	29	0.0	0.0	24160	211.3375	B5		S	2	NaN	St Louis, MO
1	1.0	1		Allison, Master. Hudson Trevor	male	0	1.0	2.0	113781	151.5500	C22 C26		S	11	NaN	Montreal, PQ / Chesterville, ON
2	1.0	0		Allison, Miss. Helen Loraine	female	2	1.0	2.0	113781	151.5500	C22 C26		S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1.0	0		Allison, Mr. Hudson Joshua Creighton	male	30	1.0	2.0	113781	151.5500	C22 C26		S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1.0	0		Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1.0	2.0	113781	151.5500	C22 C26		S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...		...	...	...	...	...	...	...	...		...	...	...	...
1304	3.0	0		Zabour, Miss. Hileni	female	14	1.0	0.0	2665	14.4542	NaN		C	NaN	328.0	NaN
1305	3.0	0		Zabour, Miss. Thamine	female	29	1.0	0.0	2665	14.4542	NaN		C	NaN	NaN	NaN
1306	3.0	0		Zakarian, Mr. Mapriededer	male	26	0.0	0.0	2656	7.2250	NaN		C	NaN	304.0	NaN
1307	3.0	0		Zakarian, Mr. Ortin	male	27	0.0	0.0	2670	7.2250	NaN		C	NaN	NaN	NaN
1308	3.0	0		Zimmerman, Mr. Leo	male	29	0.0	0.0	315082	7.8750	NaN		S	NaN	NaN	NaN

1309 rows × 14 columns

In [8]:

```
B["pclass"]=B["pclass"].apply(np.int)
B
```

Out[8]:

	pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	
0	1	1		Allen, Miss. Elisabeth Walton	female	29	0.0	0.0	24160	211.3375	B5		S	2	NaN	St Louis, MO
1	1	1		Allison, Master. Hudson Trevor	male	0	1.0	2.0	113781	151.5500	C22 C26		S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0		Allison, Miss. Helen Loraine	female	2	1.0	2.0	113781	151.5500	C22 C26		S	NaN	NaN	Montreal, PQ / Chesterville, ON

pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1.0	2.0	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1.0	0.0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1.0	0.0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0.0	0.0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0.0	0.0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0.0	0.0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [9]:

```

B["sibsp"]=B["sibsp"].apply(np.int)
B

```

Out[9]:

pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0.0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2.0	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2.0	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2.0	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0.0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0.0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0.0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0.0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0.0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [10]:

```

B["parch"]=B["parch"].apply(np.int)
B

```

Out[10]:

pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [11]:

```
B
```

Out[11]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [12]:

```
B.isnull().sum()
```

Out[12]:

```

pclass      0
survived    0
name        0
sex         0
age         0
sibsp       0
parch       0
ticket      0
fare        1
cabin      1014
embarked    2
boat        823
body        1188
home.dest   564
dtype: int64

```

In [13]:

B.info()

<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1309 entries, 0 to 1308  
Data columns (total 14 columns):  
# Column Non-Null Count Dtype  
--- ---  
0 pclass 1309 non-null int64  
1 survived 1309 non-null int64  
2 name 1309 non-null object  
3 sex 1309 non-null object  
4 age 1309 non-null int64  
5 sibsp 1309 non-null int64  
6 parch 1309 non-null int64  
7 ticket 1309 non-null object  
8 fare 1308 non-null float64  
9 cabin 295 non-null object  
10 embarked 1307 non-null object  
11 boat 486 non-null object  
12 body 121 non-null float64  
13 home.dest 745 non-null object  
dtypes: float64(2), int64(5), object(7)  
memory usage: 153.4+ KB

In [14]:

B["fare"].fillna(B["fare"].mode(),inplace=True)  
B

Out[14]:

	pclass	survived		name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1		Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1		Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0		Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0		Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0		Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...		...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0		Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0		Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0		Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0		Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0		Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [15]:

B.isnull().sum()

Out[15]:

pclass 0  
survived 0  
name 0  
sex 0  
age 0  
sibsp 0  
parch 0

parch 0  
ticket 1  
fare 1  
cabin 1014  
embarked 2  
boat 823  
body 1188  
home.dest 564  
dtype: int64

In [16]:

```
B.info()
```

<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1309 entries, 0 to 1308  
Data columns (total 14 columns):  
# Column Non-Null Count Dtype  
---  
0 pclass 1309 non-null int64  
1 survived 1309 non-null int64  
2 name 1309 non-null object  
3 sex 1309 non-null object  
4 age 1309 non-null int64  
5 sibsp 1309 non-null int64  
6 parch 1309 non-null int64  
7 ticket 1309 non-null object  
8 fare 1308 non-null float64  
9 cabin 295 non-null object  
10 embarked 1307 non-null object  
11 boat 486 non-null object  
12 body 121 non-null float64  
13 home.dest 745 non-null object  
dtypes: float64(2), int64(5), object(7)  
memory usage: 153.4+ KB

In [17]:

```
B
```

Out[17]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [18]:

```
B['embarked'].value_counts()
```

Out[18]:

```
S 914
C 270
Q 123
Name: embarked, dtype: int64
```

In [19]:

```
B["embarked"].fillna('S',inplace=True)
B
```

Out[19]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [20]:

```
B["fare"].fillna(B["fare"].median(),inplace=True)
B
```

Out[20]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	NaN
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	NaN	NaN
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	NaN



1307	pclass	survived	Zakarian, Mr. Leo	male	29	0	0	315082	7.25	NaN	S	NaN	NaN	NaN
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	NaN

1309 rows × 14 columns

In [21]:

```
B.isnull().sum()
```

Out[21]:

```
pclass      0
survived    0
name        0
sex         0
age         0
sibsp       0
parch       0
ticket      0
fare        0
cabin      1014
embarked    0
boat       823
body       1188
home.dest   564
dtype: int64
```

In [22]:

```
B["home.dest"].value_counts()
```

Out[22]:

```
New York, NY      64
London           14
Montreal, PQ     10
Cornwall / Akron, OH  9
Paris, France     9
..
Barre, Co Washington, VT  1
Mexico City, Mexico      1
New Britain, CT          1
Chelsea, London          1
Victoria, BC             1
Name: home.dest, Length: 369, dtype: int64
```

In [23]:

```
B["home.dest"].fillna("New York NK",inplace=True)
B
```

Out[23]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	NaN	C	NaN	328.0	New York NK

1305	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
			Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	NaN	C	NaN	304.0	New York NK
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	NaN	C	NaN	304.0	New York NK
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	NaN	C	NaN	NaN	New York NK
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	NaN	S	NaN	NaN	New York NK

1309 rows × 14 columns

In [24]:

```
B.drop("cabin",1,inplace=True)
```

In [25]:

```
B.drop("boat",1,inplace=True)
```

In [26]:

```
B.drop("body",1,inplace=True)
```

In [27]:

```
B
```

Out[27]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	embarked	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	S	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON
...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	C	New York NK
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	C	New York NK
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	C	New York NK
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	C	New York NK
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	S	New York NK

1309 rows × 11 columns

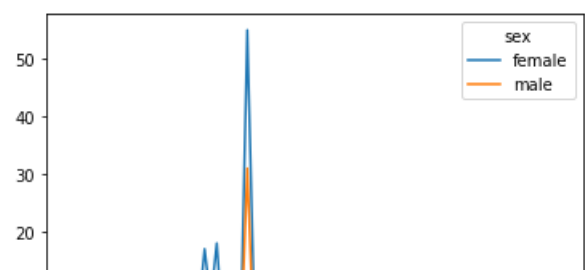
### relation entre age et sex

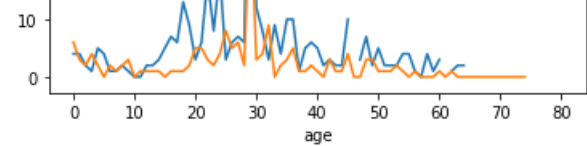
In [28]:

```
B.pivot_table('survived', index='age',columns='sex',aggfunc="sum").plot()
```

Out[28]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x4572248>





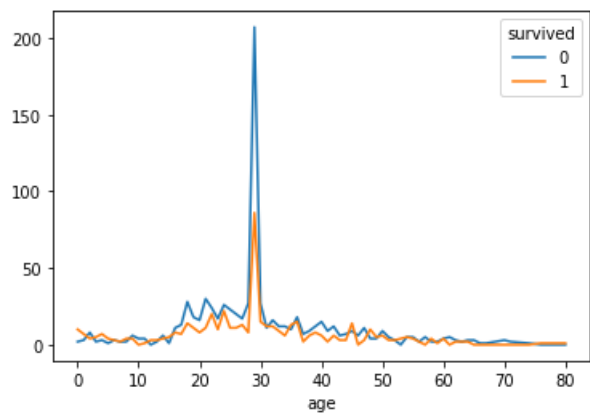
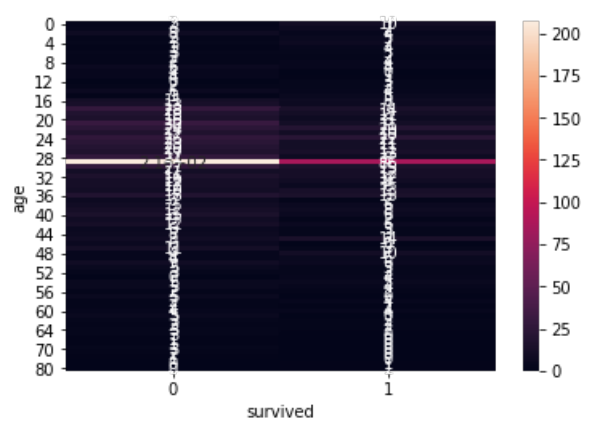
```
In [40]:
B.corr()['survived'].sort_values()
```

```
Out[40]:
pclass    -0.312469
age        -0.047021
sibsp      -0.027825
parch       0.082660
fare       0.244414
survived    1.000000
Name: survived, dtype: float64
```

```
In [:
B
```

```
In [51]:
import seaborn as sns
import pandas as pd
age_survived=pd.crosstab(B['age'],B['survived'])
sns.heatmap (age_survived, annot = True)
age_survived.plot()
```

```
Out[51]:
<matplotlib.axes._subplots.AxesSubplot at 0xc032388>
```



```
In [:
B.info()
```

```
In [:
B.corr().sum()
```

```
B.isnull().sum()
```

In [103]:

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
def plot_correlation_map( df ):

    corr = df.corr()

    s , ax = plt.subplots( figsize =( 12 , 10 ) )

    cmap = sns.diverging_palette( 220 , 10 , as_cmap = True )

    s = sns.heatmap(

        corr,

        cmap = cmap,

        square=True,

        cbar_kws={ 'shrink' : .9 },

        ax=ax,

        annot = True,

        annot_kws = { 'fontsize' : 12 }

    )

plot_correlation_map( B )
```



In []:

```
B[['survived','pclass']].groupby(['pclass'],as_index=True).mean().plot
```

In []:

```
# création de colonne TITRE
```

In [77]:

In [77]:

```
C=[]
df=""
for x in B["name"]:
    df=x.split(",")
    df=df[1].split(".")
    df1=df[0].strip()
    C.append(df1)

B["Title"]=C
B
```

Out[77]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	embarked	home.dest	Family	Title
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	S	St Louis, MO	0	Miss
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Master
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Miss
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Mr
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Mrs
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	C	New York NK	1	Miss
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	C	New York NK	1	Miss
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	C	New York NK	0	Mr
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	C	New York NK	0	Mr
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	S	New York NK	0	Mr

1309 rows x 13 columns

In []:

```
B.isnull().sum()
```

In [82]:

```
Title_Dictionary = {

    "Capt":    "Officer",

    "Col":      "Officer",

    "Major":    "Officer",

    "Dr":       "Officer",

    "Rev":      "Officer",

    "Jonkheer": "Royalty",

    "Don":      "Royalty",

    "Sir":      "Royalty",

    "Lady":     "Royalty",

    "the Countess": "Royalty",

    "Dona":     "Royalty",

    "Mme":      "Miss",

    "Mlle":     "Miss",

    "Miss":     "Miss",

    "Ms":       "Mrs",

    "Mr":       "Mrs",

    "Mrs":      "Mrs",
```

"Master": "Master"

}

In [83]:

```
B["Title"].replace(Title_Dictionary,inplace=True)
B
```

Out[83]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	embarked	home.dest	Family	Title
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	S	St Louis, MO	0	Miss
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Master
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Miss
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Mrs
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3	Mrs
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	3	0	Zabour, Miss. Hileni	female	14	1	0	2665	14.4542	C	New York NK	1	Miss
1305	3	0	Zabour, Miss. Thamine	female	29	1	0	2665	14.4542	C	New York NK	1	Miss
1306	3	0	Zakarian, Mr. Mapriededer	male	26	0	0	2656	7.2250	C	New York NK	0	Mrs
1307	3	0	Zakarian, Mr. Ortin	male	27	0	0	2670	7.2250	C	New York NK	0	Mrs
1308	3	0	Zimmerman, Mr. Leo	male	29	0	0	315082	7.8750	S	New York NK	0	Mrs

1309 rows × 13 columns

In [84]:

```
B["Title"].value_counts()
```

Out[84]:

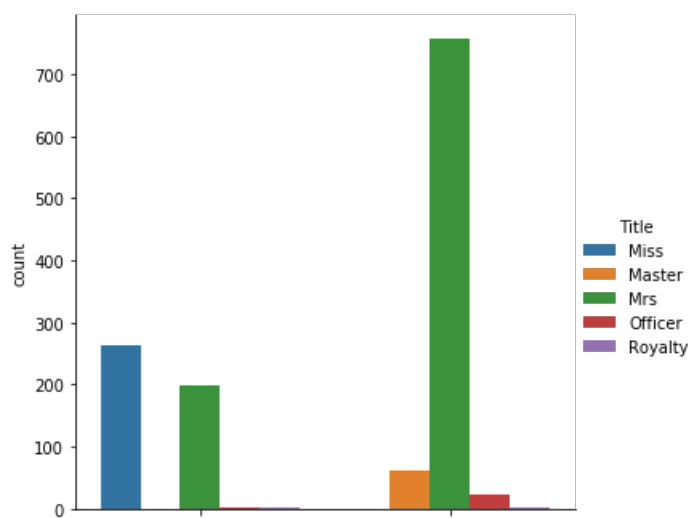
Mrs 956  
Miss 263  
Master 61  
Officer 23  
Royalty 6  
Name: Title, dtype: int64

In [100]:

```
import seaborn as sns
sns.factorplot('sex',data=B,kind="count", hue="Title")
```

Out[100]:

<seaborn.axisgrid.FacetGrid at 0xccd5c48>



female                      male  
sex

In [73]:

B.head(10)

Out[73]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	embarked	home.dest	Family
0	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	S	St Louis, MO	0
1	1	1	Allison, Master. Hudson Trevor	male	0	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3
2	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	S	Montreal, PQ / Chesterville, ON	3
5	1	1	Anderson, Mr. Harry	male	48	0	0	19952	26.5500	S	New York, NY	0
6	1	1	Andrews, Miss. Kornelia Theodosia	female	63	1	0	13502	77.9583	S	Hudson, NY	1
7	1	0	Andrews, Mr. Thomas Jr	male	39	0	0	112050	0.0000	S	Belfast, NI	0
8	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0	11769	51.4792	S	Bayside, Queens, NY	2
9	1	0	Artagaveytia, Mr. Ramon	male	71	0	0	PC 17609	49.5042	C	Montevideo, Uruguay	0

In [97]:

Family\_Title=B[["Family","Title"]].groupby(["Family"],as\_index=True)  
Family\_Title

Out[97]:

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000000CAD4308>

In [52]:

B.name.unique()

Out[52]:

array(['Allen, Miss. Elisabeth Walton', 'Allison, Master. Hudson Trevor',  
      'Allison, Miss. Helen Loraine', ..., 'Zakarian, Mr. Mapriededer',  
      'Zakarian, Mr. Ortin', 'Zimmerman, Mr. Leo'], dtype=object)

In [71]:

Family=pd.DataFrame()  
Family=B["Family"]=B["parch"]+B["sibsp"]  
Family

Out[71]:

0    0  
1    3  
2    3  
3    3  
4    3  
..  
1304 1  
1305 1  
1306 0  
1307 0  
1308 0  
Length: 1309, dtype: int64