Group By function

In [1]:
```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
data = pd.read_csv('Titanic.csv')
data.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	Embarked	WikiId	Name_wiki	Age_wiki	Hometown	B
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	S	691.0	Braund, Mr. Owen Harris	22.0	Bridgerule, Devon, England	South
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	C	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New York, New York, US	Ch
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	S	865.0	Heikkinen, Miss Laina	26.0	Jyväskylä, Finland	South
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	S	127.0	Futrelle, Mrs. Lily May (née Peel)	35.0	Scituate, Massachusetts, US	South
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	S	627.0	Allen, Mr. William Henry	35.0	Birmingham, West Midlands, England	South

5 rows × 21 columns

In [3]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 21 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  1309 non-null   int64
 1   Survived     891 non-null    float64
 2   Pclass       1309 non-null   int64
 3   Name         1309 non-null   object
 4   Sex          1309 non-null   object
 5   Age          1046 non-null   float64
 6   SibSp        1309 non-null   int64
 7   Parch        1309 non-null   int64
 8   Ticket       1309 non-null   object
 9   Fare         1308 non-null   float64
 10  Cabin        295 non-null    object
 11  Embarked     1307 non-null   object
 12  WikiId       1304 non-null   float64
 13  Name_wiki    1304 non-null   object
```

Number of passengers in each class

In [4]:
```
data.Pclass.value_counts()
```

Out[4]:
```
3    709
1    323
2    277
Name: Pclass, dtype: int64
```

Single group single column

Count of passengers who survived from each class

In [5]:
```
pclass_survived = data.groupby(['Pclass'])['Survived']
```

In [6]:
```
pclass_survived.sum()
```

Out[6]:
```
Pclass
1    136.0
2     87.0
3    119.0
Name: Survived, dtype: float64
```

Multiple columns

Total amount paid by each class who survived

In [7]:
```
pclass_survived_fare = data.groupby(['Pclass'])['Survived', 'Fare']
```

In [8]:
```
pclass_survived_fare.sum()
```

Out[8]:

Pclass	Survived	Fare
1	136.0	28265.4043
2	87.0	5866.6374
3	119.0	9418.4452

Multiple groups

Count of females and males who survived in each class

In [9]:
```
pclass_sex_survived = data.groupby(['Pclass', 'Sex'])['Survived']
```

In [10]:
```
pclass_sex_survived.sum()
```

Out[10]:
```
Pclass  Sex
1       female    91.0
        male      45.0
2       female    70.0
        male      17.0
3       female    72.0
        male      47.0
Name: Survived, dtype: float64
```

In [11]:
```
pclass_sex_survived = data.groupby(['Pclass', 'Sex'], as_index=False)['Survived']
```

In [12]:
```
pclass_sex_survived.sum()
```

Out[12]:

	Pclass	Sex	Survived
0	1	female	91.0
1	1	male	45.0
2	2	female	70.0
3	2	male	17.0
4	3	female	72.0
5	3	male	47.0

Multiple functions

Mean and sum of pclass & survived

In [13]:
```
pclass_survived_mean_sum =data.groupby(['Pclass'])['Survived'].agg(['mean', 'sum'])
```

In [14]:
```
pclass_survived_mean_sum
```

Out[14]:

Pclass	mean	sum
1	0.629630	136.0
2	0.472826	87.0
3	0.242363	119.0

pd.cut method

Pandas cut() function is used to separate the array elements into different bins .

The cut function is mainly used to perform statistical analysis on scalar data.

Creating the bins for ages

In [15]:
```
data['age_bins'] = pd.cut(data.Age, bins=3, labels=('young', 'middle', 'old'))
```

In [16]:
```
data.head(10)
```

ss	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	WikiId	Name_wiki	Age_wiki	Hometown	Boarded	Destination	Lifeboat	Body	Class	age_bins
3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	691.0	Braund, Mr. Owen Harris	22.0	Bridgerule, Devon, England	Southampton	Qu'Appelle Valley, Saskatchewan, Canada	NaN	NaN	3.0	young
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New York, New York, US	Cherbourg	New York, New York, US	4	NaN	1.0	middle
3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	865.0	Heikkinen, Miss Laina	26.0	Jyväskylä, Finland	Southampton	New York City	14?	NaN	3.0	young
1	Futrelle, Mrs. Jacques Heath	female	35.0	1	0	113803	53.1000	...	127.0	Futrelle, Mrs. Lily May (née	35.0	Scituate, Massachusetts,	Southampton	Scituate, Massachusetts,	D	NaN	1.0	middle

Number of survived from each boarding station

In [17]:
```
boarded_survived=data.groupby(['Boarded'])['Survived']
```

In [18]:
```
boarded_survived.sum()
```

Out[18]:
```
Boarded
Belfast          0.0
Cherbourg       93.0
Queenstown      29.0
Southampton    219.0
Name: Survived, dtype: float64
```

Average fare of each Pclass

In [19]:
```
avg_fare_pclass = data.groupby(['Pclass'])['Fare']
```

In [20]:
```
avg_fare_pclass.mean()
```

Out[20]:
```
Pclass
1    87.508992
2    21.179196
3    13.302889
Name: Fare, dtype: float64
```

Number of age category in each Class

In [21]:
```
age_class = data.groupby(['Class'])['age_bins']
```

In [22]:
```
age_class.value_counts()
```

Out[22]:
```
Class
1.0    middle    172
       young      63
       old        52
2.0    middle    132
       young     108
       old        17
3.0    young     301
       middle    190
       old         8
Name: age_bins, dtype: int64
```

Count of gender from each boarding station

In [23]:
```
sex_barding = data.groupby(['Boarded'])['Sex']
```

In [24]:
```
sex_barding.value_counts()
```

Out[24]:
```
Boarded      Sex
Belfast      male       10
Cherbourg    male      145
             female    114
Queenstown   male       61
             female     58
Southampton  male      625
             female    291
Name: Sex, dtype: int64
```

In [25]:
```
data.head()
```

Out[25]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	WikiId	Name_wiki	Age_wiki	Hometown	Boarded	
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	691.0	Braund, Mr. Owen Harris	22.0	Bridgerule, Devon, England	Southampton	Sa
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New York, New York, US	Cherbourg	Ne
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	865.0	Heikkinen, Miss Laina	26.0	Jyväskylä, Finland	Southampton	N
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	127.0	Futrelle, Mrs. Lily May (née Peel)	35.0	Scituate, Massachusetts, US	Southampton	Ma
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	627.0	Allen, Mr. William Henry	35.0	Birmingham, West Midlands, England	Southampton	N

5 rows × 22 columns

Count of each age category from each boarding station

In [26]:
```
age_boarding = data.groupby(['Boarded'])['age_bins']
```

In [27]:
```
age_boarding.value_counts()
```

Out[27]:
```
Boarded
Belfast      middle      1
             old         1
             young       0
Cherbourg    middle     90
             young      84
             old        23
Queenstown   young      25
             middle     17
             old         4
Southampton  middle    386
             young     363
             old        49
Name: age_bins, dtype: int64
```

Count of gender and age class from each Boarding station

In [28]:
```
age_boarding_sex = data.groupby(['Boarded','Sex'], as_index=False)['age_bins']
```

In [29]:
```
age_boarding_sex.value_counts()
```

Out[29]:

	Boarded	Sex	age_bins	count
0	Belfast	female	middle	0
1	Belfast	female	old	0
2	Belfast	female	young	0
3	Belfast	male	old	1
4	Belfast	male	middle	1
5	Belfast	male	young	0
6	Cherbourg	female	middle	45
7	Cherbourg	female	young	42
8	Cherbourg	female	old	11
9	Cherbourg	male	middle	45
10	Cherbourg	male	young	42
11	Cherbourg	male	old	12
12	Queenstown	female	young	13
13	Queenstown	female	middle	9
14	Queenstown	female	old	0
15	Queenstown	male	young	12
16	Queenstown	male	middle	8
17	Queenstown	male	old	4
18	Southampton	female	young	130
19	Southampton	female	middle	121
20	Southampton	female	old	15
21	Southampton	male	middle	265
22	Southampton	male	young	233
23	Southampton	male	old	34