

151050035 匡卉

我有很多不懂的地方，所以部分内容是参考钱超峰同学的作业

In [1]:

```
import pandas as pd
```

In [4]:

```
train_url = "http://s3.amazonaws.com/assets.datacamp.com/course/Kaggle/train.csv"
train = pd.read_csv(train_url)
test_url = "http://s3.amazonaws.com/assets.datacamp.com/course/Kaggle/test.csv"
test = pd.read_csv(test_url)
train.head()
print('完成')
```

完成

In [5]:

```
test.head()
```

Out[5]:

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

In [6]:

```
train.shape
```

Out[6]:

(891, 12)

In [7]:

```
train.describe()
```

Out[7]:

	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	32.000000
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.000000
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	7.250000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.500000
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	512.000000

In [9]:

```
train[:3]
```

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	53.00
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	51.00

In [10]:

```
train["Survived"][:5]
```

Out[10]:

```
0    0
1    1
2    1
3    1
4    0
```

Name: Survived, dtype: int64

In [12]:

```
train.sort_values(by = ['Age'])[:5]
```

Out[12]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
803	804	1	3	Thomas, Master. Assad Alexander	male	0.42	0	1	2625
755	756	1	2	Hamalainen, Master. Viljo	male	0.67	1	1	250649
644	645	1	3	Baclini, Miss. Eugenie	female	0.75	2	1	2666
469	470	1	3	Baclini, Miss. Helene Barbara	female	0.75	2	1	2666
78	79	1	2	Caldwell, Master. Alden Gates	male	0.83	0	2	248738

In [13]:

```
train.sort_values(by = ['Age'], ascending = False)[:5]
```

Out[13]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
630	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	27042	3
851	852	0	3	Svensson, Mr. Johan	male	74.0	0	0	347060	7
493	494	0	1	Artagaveytia, Mr. Ramon	male	71.0	0	0	PC 17609	4
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	PC 17754	3
116	117	0	3	Connors, Mr. Patrick	male	70.5	0	0	370369	7

In [14]:

```
train.groupby(['SibSp', 'Survived']).sum()
```

Out[14]:

		PassengerId	Pclass	Age	Parch	Fare
SibSp	Survived					
0	0	184300	1004	9672.50	43	6611.6071
	1	92565	426	5115.75	70	9009.1459
1	0	40630	232	2739.00	71	3256.1710
	1	51273	198	2767.42	66	5970.6294
2	0	5221	39	364.00	5	511.2000
	1	6327	27	201.50	13	937.9041
3	0	3901	34	57.00	17	539.6918
	1	1244	7	110.00	4	562.8500
4	0	6304	45	102.00	21	502.7000
	1	565	9	25.00	6	70.7000
5	0	1684	15	51.00	10	234.5000
8	0	3372	21	NaN	14	486.8500