

Titanic数据 - 可视化



导入数据

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 import numpy as np
```

```
1 titanic = pd.read_csv('data/Titanic.csv')
2 titanic.head(3)
```

```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>PassengerId</th>
4   <th>Survived</th>
5   <th>Pclass</th>
6   <th>Name</th>
7   <th>Sex</th>
8   <th>Age</th>
9   <th>SibSp</th>
10  <th>Parch</th>
11  <th>Ticket</th>
12  <th>Fare</th>
13  <th>Cabin</th>
14  <th>Embarked</th>
15 </tr>
```

```
1 <tr>
2   <th>0</th>
3   <td>1</td>
4   <td>0</td>
5   <td>3</td>
6   <td>Braund, Mr. Owen Harris</td>
7   <td>male</td>
8   <td>22.0</td>
9   <td>1</td>
10  <td>0</td>
11  <td>A/5 21171</td>
12  <td>7.2500</td>
13  <td>NaN</td>
14  <td>S</td>
15 </tr>
16 <tr>
17   <th>1</th>
18   <td>2</td>
19   <td>1</td>
20   <td>1</td>
21   <td>Cumings, Mrs. John Bradley (Florence Briggs Th...
    </td>
22   <td>female</td>
23   <td>38.0</td>
24   <td>1</td>
25   <td>0</td>
26   <td>PC 17599</td>
27   <td>71.2833</td>
28   <td>C85</td>
29   <td>C</td>
30 </tr>
31 <tr>
32   <th>2</th>
33   <td>3</td>
```

```
34 <td>1</td>
35 <td>3</td>
36 <td>Heikkinen, Miss. Laina</td>
37 <td>female</td>
38 <td>26.0</td>
39 <td>0</td>
40 <td>0</td>
41 <td>STON/O2. 3101282</td>
42 <td>7.9250</td>
43 <td>NaN</td>
44 <td>S</td>
45 </tr>
```

统计年龄大于20的人数

```
1 (titanic['Age'] > 20).sum()
```

```
1 535
```

将PassengerId设置为索引

```
1 titanic.set_index('PassengerId')
2 titanic.head(3)
```

```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>PassengerId</th>
4   <th>Survived</th>
5   <th>Pclass</th>
6   <th>Name</th>
7   <th>Sex</th>
8   <th>Age</th>
9   <th>SibSp</th>
10  <th>Parch</th>
11  <th>Ticket</th>
12  <th>Fare</th>
13  <th>Cabin</th>
14  <th>Embarked</th>
15 </tr>
```

```
1 <tr>
2   <th>0</th>
3   <td>1</td>
4   <td>0</td>
5   <td>3</td>
6   <td>Braund, Mr. Owen Harris</td>
7   <td>male</td>
8   <td>22.0</td>
9   <td>1</td>
10  <td>0</td>
11  <td>A/5 21171</td>
12  <td>7.2500</td>
13  <td>NaN</td>
14  <td>S</td>
15 </tr>
16 <tr>
17   <th>1</th>
18   <td>2</td>
19   <td>1</td>
20   <td>1</td>
21   <td>Cumings, Mrs. John Bradley (Florence Briggs Th...
    </td>
22   <td>female</td>
23   <td>38.0</td>
24   <td>1</td>
25   <td>0</td>
26   <td>PC 17599</td>
27   <td>71.2833</td>
28   <td>C85</td>
29   <td>C</td>
30 </tr>
31 <tr>
32   <th>2</th>
33   <td>3</td>
```

```

34 <td>1</td>
35 <td>3</td>
36 <td>Heikkinen, Miss. Laina</td>
37 <td>female</td>
38 <td>26.0</td>
39 <td>0</td>
40 <td>0</td>
41 <td>STON/O2. 3101282</td>
42 <td>7.9250</td>
43 <td>NaN</td>
44 <td>S</td>
45 </tr>

```

绘制一个展示男女乘客比例的扇形图

```

1 male = (titanic['Sex'] == 'male').sum()
2 female = (titanic['Sex'] == 'female').sum()
3 proportions = [male, female]
4 plt.pie(proportions, labels=['Male', 'Female'],
    shadow=False, colors=['green', 'red'], explode=[0.02, 0],
    startangle=90, autopct='%1.1f%%')
5 plt.axis('equal')
6 plt.title('Sex proportions')
7 plt.tight_layout()
8 plt.show()

```

绘制一个展示船票Fare, 与乘客年龄和性别的散点图

```
1 lm = sns.lmplot(x = 'Age', y = 'Fare', data = titanic, hue =  
  'Sex', fit_reg=False)  
2 lm.set(title = 'Fare x Age')  
3  
4 axes = lm.axes  
5 axes[0, 0].set_ylim(-5, )  
6 axes[0, 0].set_xlim(-5, 85)  
7 plt.show()
```

有多少人生还？

```
1 titanic.Survived.sum()
```

```
1 342
```

绘制一个展示船票价格的直方图

```
1 df = titanic.Fare.sort_values(ascending=False)  
2 binsVal = np.arange(0, 600, 10)  
3 plt.hist(df, bins=binsVal)  
4 plt.xlabel('Fare')  
5 plt.ylabel('Frequency')  
6 plt.title('Fare Payed Histrogram')  
7 plt.show()
```