



Chapter 6 - exercise 4: Titanic Disaster

Thông tin về Titanic Disaster có thể xem tại:

<https://www.kaggle.com/c/titanic/data> (<https://www.kaggle.com/c/titanic/data>)

Cung cấp tập tin train.csv. Hãy thực hiện các yêu cầu sau:

1. Đặt biến titanic chứa dữ liệu đọc được từ tập tin train.csv. In head của titanic. Thiết lập cột index cho titanic là PassengerId. In head của titanic lúc này.
2. Tạo pie chart trình bày tỷ lệ nam/nữ. Cụ thể như sau: Tạo biến males, female là tổng nam và tổng nữ. Tạo biến proportions là list có 2 phần tử là male và female. Vẽ biểu đồ: với dữ liệu là proportions, nhãn là ['Males', 'Female'], màu là ['blue', 'red']. Thiết lập title là Sex Proportion
3. Cho biết có bao nhiêu người còn sống sót
4. Vẽ histogram với vé (Fare). Chi tiết như sau: Sắp xếp lại dữ liệu df theo cột Fare giảm dần. In head. Tạo binsVal = np.arange(0,600,10). Vẽ histogram với dữ liệu là df, bins = binsVal. Bổ sung các thông tin xlabel, ylabel, title

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

%matplotlib inline
```

```
In [2]: # Câu 1:
titanic = pd.read_csv('train.csv', sep = ',')
titanic.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nal
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Nal
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Nal



```
In [3]: titanic.set_index('PassengerId').head()
```

```
Out[3]:
```

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
PassengerId										
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

```
In [4]: # Câu 2: Tạo pie chart trình bày tỷ lệ nam/nữ.
'''
Tạo biến males, female là tổng nam và tổng nữ.
Tạo biến proportions là list có 2 phần tử là male và female
Vẽ biểu đồ: với dữ liệu là proportions, nhãn là ['Males', 'Female'], màu là ['blue', 'red']
Thiết lập title là Sex Proportion
'''

# sum the instances of males and females
males = (titanic['Sex'] == 'male').sum()
females = (titanic['Sex'] == 'female').sum()

# put them into a list called proportions
proportions = [males, females]

# Create a pie chart
plt.pie(
    # using proportions
    proportions,

    # with the labels being officer names
    labels = ['Males', 'Females'],

    # with no shadows
    shadow = False,

    # with colors
    colors = ['blue', 'red'],

    # with one slice exploded out
    explode = (0.15, 0),

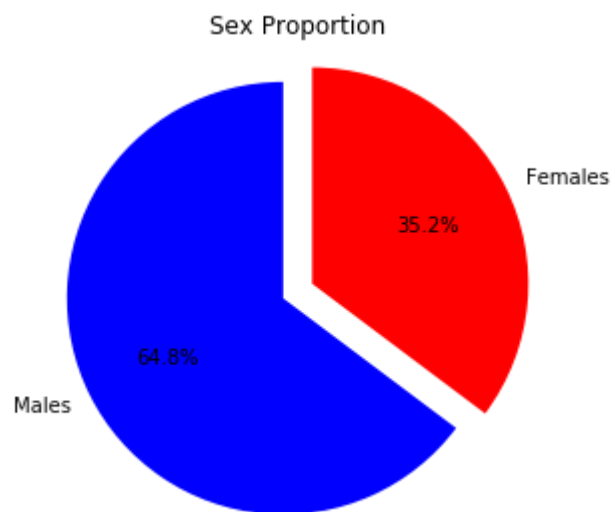
    # with the start angle at 90%
    startangle = 90,

    # with the percent listed as a fraction
    autopct = '%1.1f%%'
)

# View the plot drop above
plt.axis('equal')

# Set Labels
plt.title("Sex Proportion")

# View the plot
plt.tight_layout()
plt.show()
```



```
In [5]: # Câu 3: Cho biết có bao nhiêu người còn sống sót  
titanic.Survived.sum()
```

```
Out[5]: 342
```

```
In [6]: # Câu 4: Vẽ histogram với vé (Fare)
'''
Sắp xếp lại dữ liệu df theo cột Fare giảm dần. In head.
Tạo binsVal = np.arange(0,600,10)
Vẽ histogram với dữ liệu là df, bins = binsVal. Bổ sung các thông tin xlabel, ylabel
'''

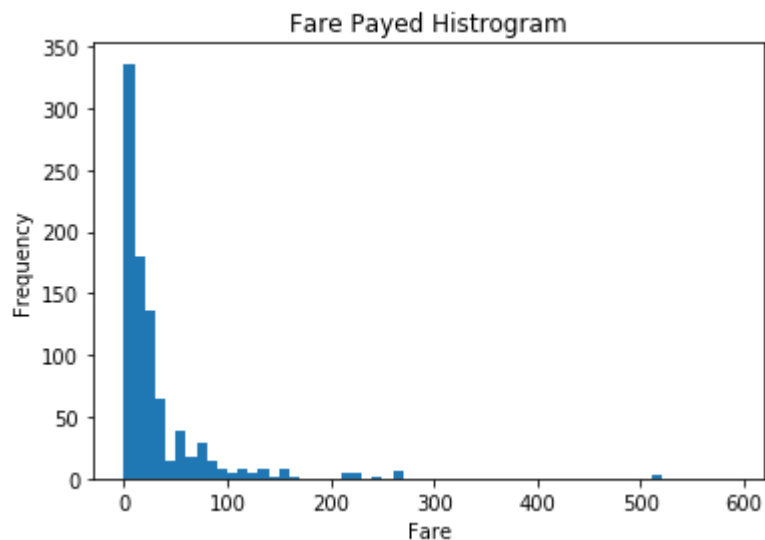
# sort the values from the top to the least value and slice the first 5 items
df = titanic.Fare.sort_values(ascending = False)
df

# create bins interval using numpy
binsVal = np.arange(0,600,10)
binsVal

# create the plot
plt.hist(df, bins = binsVal)

# Set the title and labels
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.title('Fare Payed Histogram')

# show the plot
plt.show()
```



```
In [ ]:
```