

▼ Chapter 6 - Exercise 3: Titanic Disaster

Vào ngày 15 tháng 4 năm 1912, trong chuyến hành trình đầu tiên của mình, tàu Titanic đã chìm sau khi va chạm với một tảng băng trôi, đã có 1502 mãi mãi ra đi trong tổng số 2224 hành khách và phi hành đoàn.

Thông tin về Titanic Disaster có thể xem tại: <https://www.kaggle.com/c/titanic/data>

▼ Dựa trên tập tin *train.csv* , hãy thực hiện các yêu cầu sau:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

%matplotlib inline

# Câu 1:
# a) Đọc dữ liệu từ tập tin train.csv và lưu vào biến titanic.
titanic = pd.read_csv(r'data\train.csv', sep = ',')
# Hiển thị 5 dòng dữ liệu đầu của titanic
titanic.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs TL	female	38.0	1	0	PC 17599	71.

```
# b) Thiết lập cột index cho titanic là PassengerId. Hiển thị lại 5 dòng dữ liệu đầu của t
titanic.set_index('PassengerId').head()
```

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
----------	--------	------	-----	-----	-------	-------	--------	------

Câu 2: Tạo pie chart thể hiện tỷ lệ hành khách nam/nữ trên tàu.

'''

Gợi ý:

- 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

'''

Tạo biến males, female là tổng nam và tổng nữ

males = (titanic['Sex'] == 'male').sum()

females = (titanic['Sex'] == 'female').sum()

Tạo biến proportions là list có 2 phần tử là male và female

proportions = [males, females]

Vẽ biểu đồ

plt.pie(proportions, labels = ['Males', 'Females'], shadow = False, colors = ['blue', 'red'])

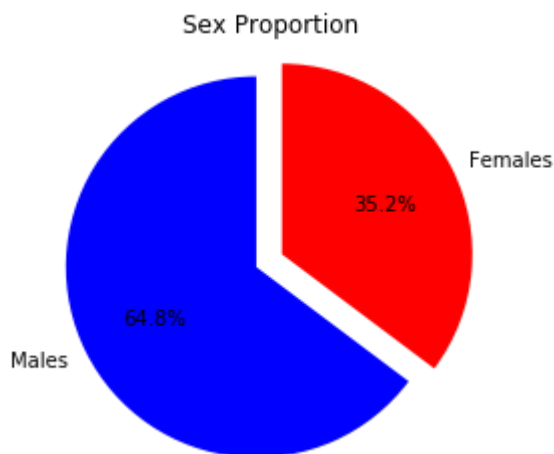
plt.axis('equal')

Thiết lập title là Sex Proportion

plt.title("Sex Proportion")

Show the plot

plt.show()



Câu 3: Cho biết có bao nhiêu người còn sống sót

print('Số người người còn sống sót là: ', titanic.Survived.sum())

Số người người còn sống sót là: 342

Câu 4: Vẽ biểu đồ histogram của cột vé (Fare)

Bạn nhận xét gì về biểu đồ vừa vẽ

'''

Gợi ý:

- 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
...

# Sắp xếp lại dữ liệu df theo cột Fare giảm dần, và hiển thị các dòng dữ liệu đầu của df
df = titanic.Fare.sort_values(ascending = False)
df.head()

    679    512.3292
    258    512.3292
    737    512.3292
    341    263.0000
    438    263.0000
    Name: Fare, dtype: float64

# Tạo binsVal = np.arange(0,600,10)
binsVal = np.arange(0,600,10)

# Vẽ histogram với dữ liệu là df, bins = binsVal
plt.hist(df, bins = binsVal)

# Thiết lập các thông tin xlabel, ylabel, title
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.title('Fare Payed Histogram')

plt.show()

```

