

Assignment 1 Python Library Part

Total: 100 points + α

Problem Definition

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

"what sorts of people were more likely to survive?"

Description

- Survival = (0=사망, 1=생존)
- Pclass = Ticket class (1=1등석, 2=2등석, 3=3등석)
- Sex
- Age
- sibsp = Titanic 호에 동승한 자매 또는 배우자의 수
- parch = Titanic 호에 동승한 부모 또는 자식의 수
- ticket
- fare = 승객 요금
- cabin = 방 호수
- embarked = 탑승지 (C=Cherbourg, Q=Queenstown, S=Southampton)

Before solving the problem...

Basically, using only 'numpy, pandas and matplotlib' is enough to solve the problem, but you can use other libraries.

If you fail to comply with the following, you will be deducted.

1) In the case of graph visualization, the x-axis, y-axis and the title should be stated exactly as in the example. You can freely choose anything else. 2) When writing the code, write the number of the corresponding problem (a, b, c ...) as an annotation.

When submitting a assignment, you only need to submit a Python code (.ipynb).

Make sure it runs before submitting the code.

(a) (10pt) 데이터를 불러오고, 최소 3개 이상의 함수를 이용해 데이터를 살펴보아라.

(b) (20pt) 1) 각 변수별로 NaN 값을 몇 개 갖고 있는지 확인하고, 2) 'PassengerID', 'Name', 'Ticket' 열을 데이터에서 제거하여라. 3) 'Age' 열에서 NaN 값이 존재하면 제거하여라. 4) 'Pclass' data type 을 string 타입으로 변환하여라.

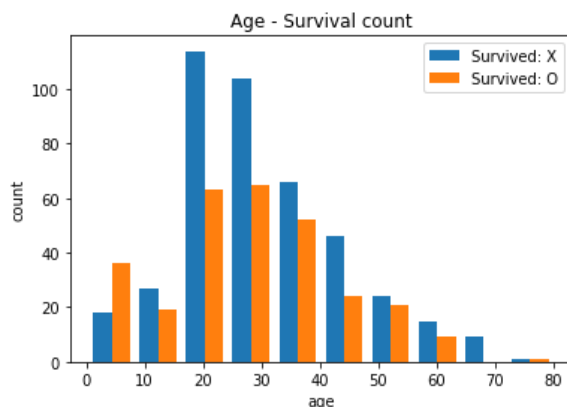
(c) (10pt) 성별로 생존율의 평균을 구하여 출력하여라.

(d) (10pt) 나이를 구간별로 나눌 수 있는 (10,20,30대...) 함수를 정의하고, 나이 구간 별로 평균 생존율을 구하되, 오름차순으로 정렬하여 출력하여라.

(주의: 나이를 구간별로 구분한 데이터를 원본의 'Age' 열에 덮어씌우지 말 것)

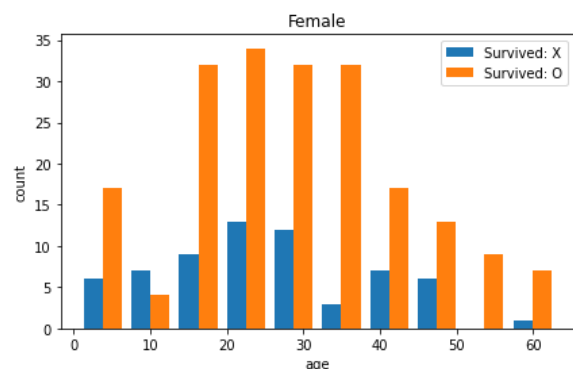
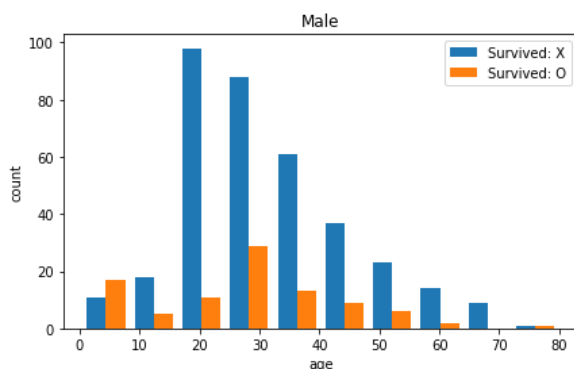
(e) (15pt) 사망자와 생존자 각각을 나이별 분포로 나타낸 히스토그램을 그려라.

예시)



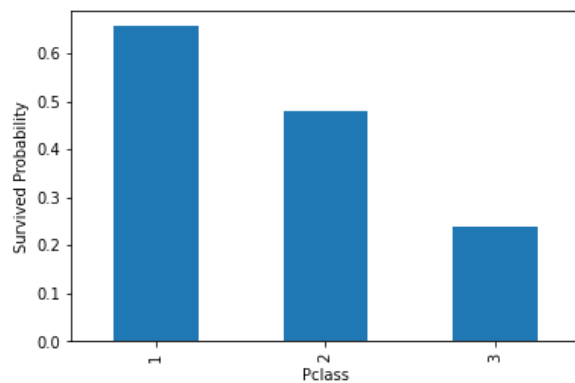
(f) (15pt) 사망자와 생존자 각각을 나이별 분포로 나타내되, 성별에 따라 나타낸 히스토그램을 각각 그려라.

예시)



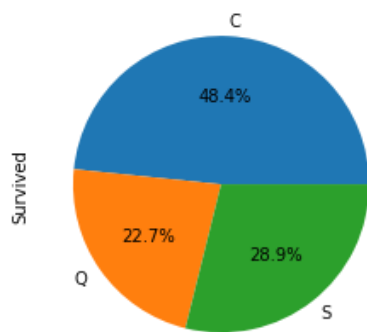
(g) (10pt) Pclass 별로 생존자 평균을 나타내는 그래프를 그려라.

예시)



(h) (10pt) Embarked (탑승지) 별로 생존자 평균을 나타내는 그래프를 그려라.

예시)



Bonus) (maximum 10pt) 위의 분석 (a~h) 으로부터 알 수 있는 사실을 서술하고 (코드에 주석으로 표시하고 서술할 것), 이 외에도 생존에 영향을 미칠 수 있는 변수들이 무엇일지에 대해 분석해보아라 (단, 시각화한 그래프 하나 이상을 포함 해야함).