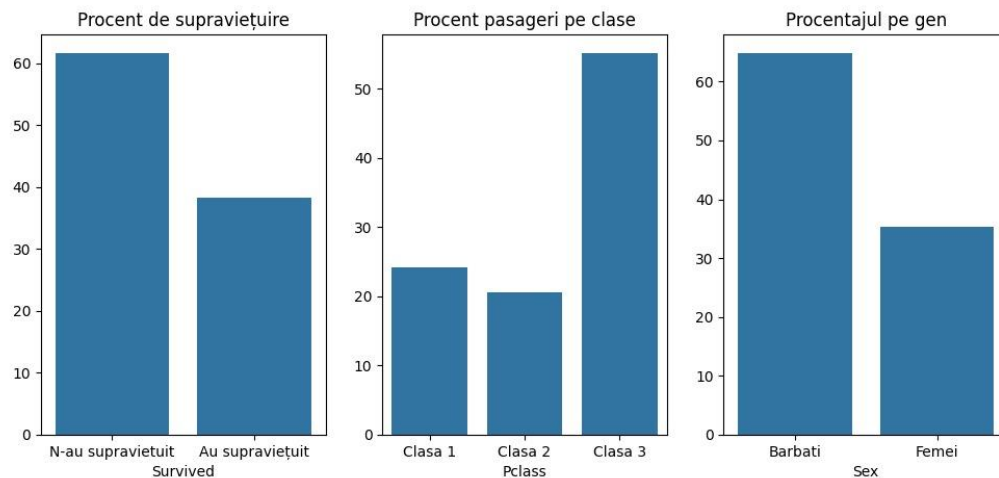


Proiect PCLP3

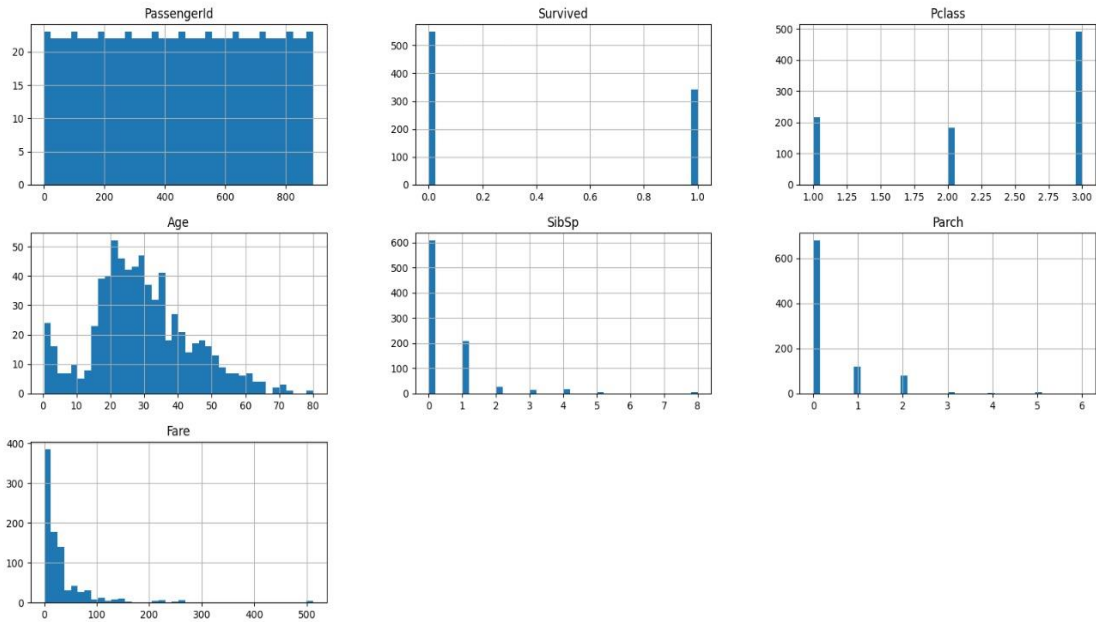
În acest proiect folosesc biblioteci din python precum numpy, pandas, matplotlib și seaborn pentru a modifica și vizualiza date din setul de date Titanic.

Cerinta 1: creez dataframe-ul df citind din fisierul "train.csv" și număr valorile de tip NaN din df, la fel și cu duplicatele

Cerinta 2: iau fiecare coloană de interes, număr câte valori sunt pentru fiecare posibilitate și împart la câte sunt (normalize = True). Înmulțind cu 100 se obține procentul. Acestea sunt graficele:



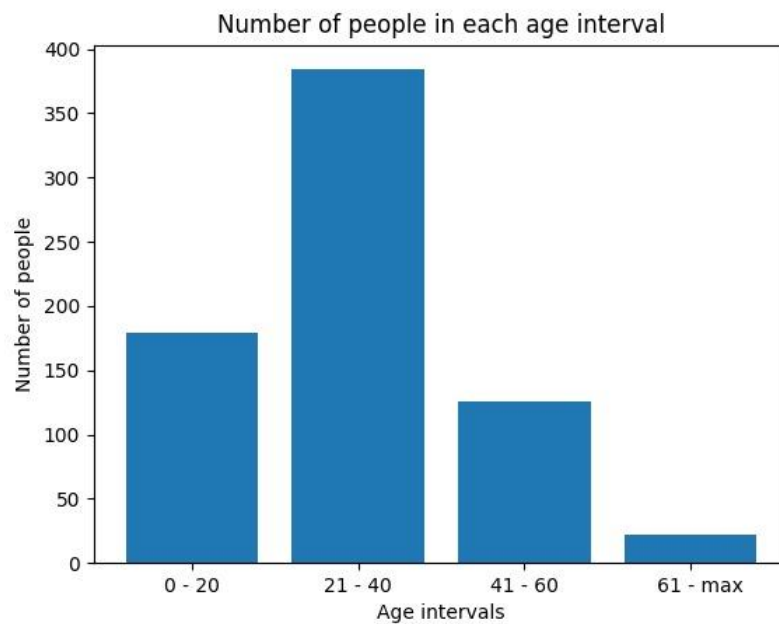
Cerinta 3: preiau coloanele numerice folosind `select_dtypes` si creez histogramele pentru fiecare din ele.



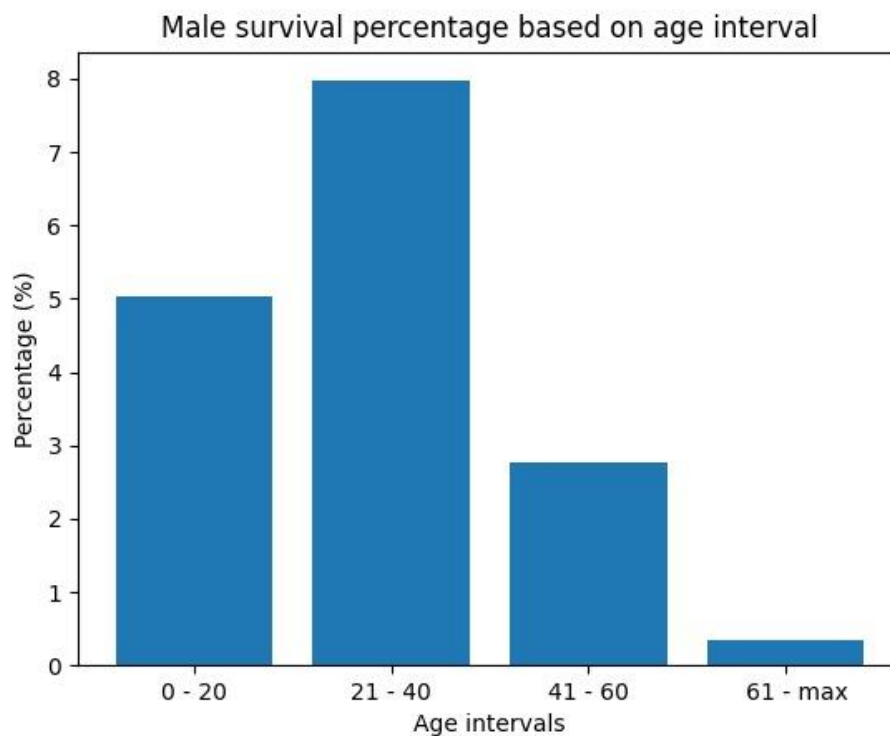
Cerinta 4 : Pentru fiecare coloana ce are valori lipsa, calculez numarul lor si il afisez. Impartind la numarul total de linii si inmultind la 100 obtin procentajul.

Iau o lista ce contine intervale de varsta (tuple-urile cu sfarsitul si inceputul intervalului). Pentru fiecare interval numar cati oameni au varsta din interval si pun valoarea in lista.

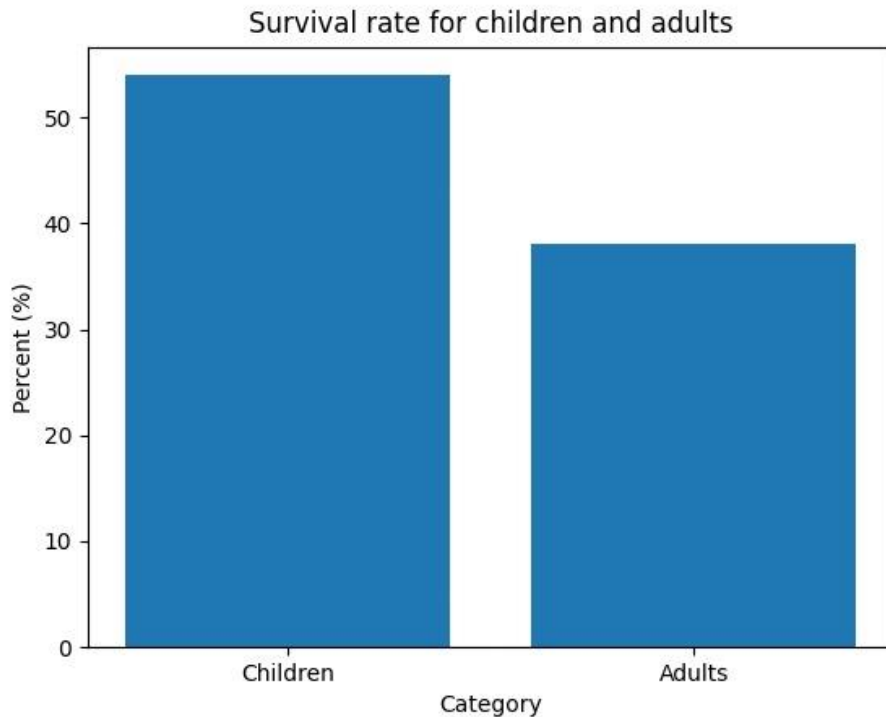
Apoi, trec prin toate intrarile si pun in lista indexul corespunzator intervalului din lista cu range-uri. In final inseres coloana in dataframe.



Cerinta 6 : iau o coloana cu toti barbatii care au supravietuit si pentru fiecare interval de varsta identic ca la cerinta anterioara, calculez numarul si procentajul.

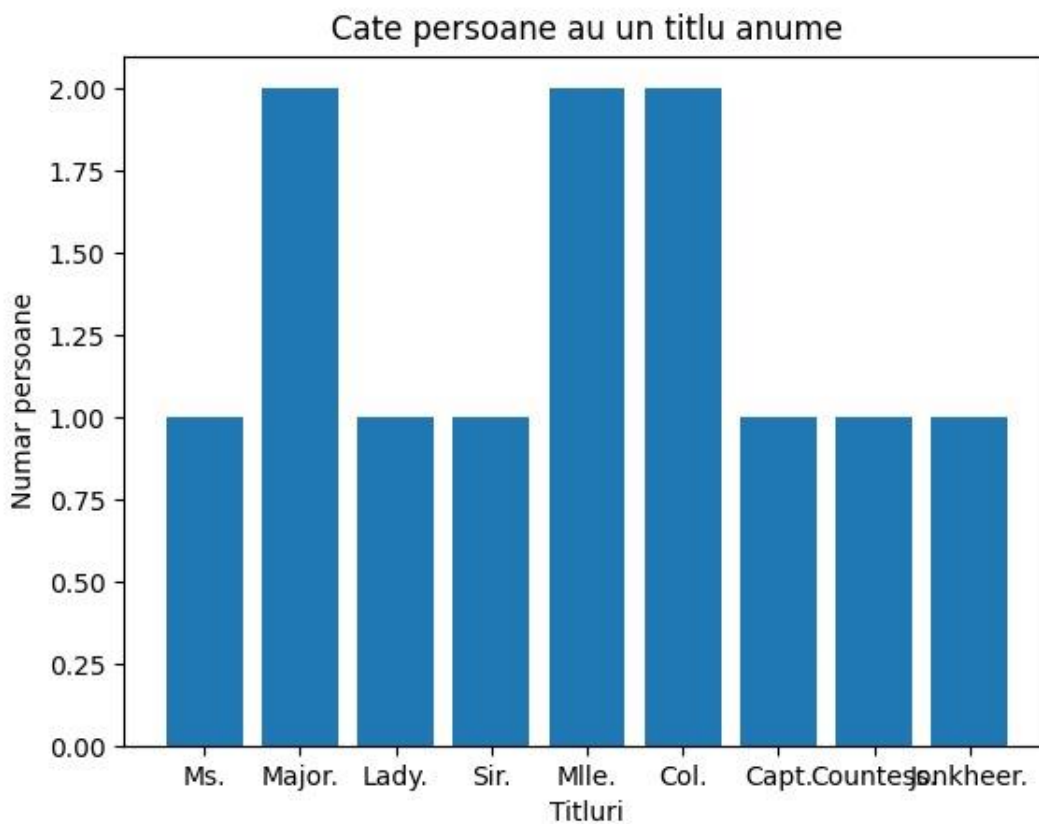
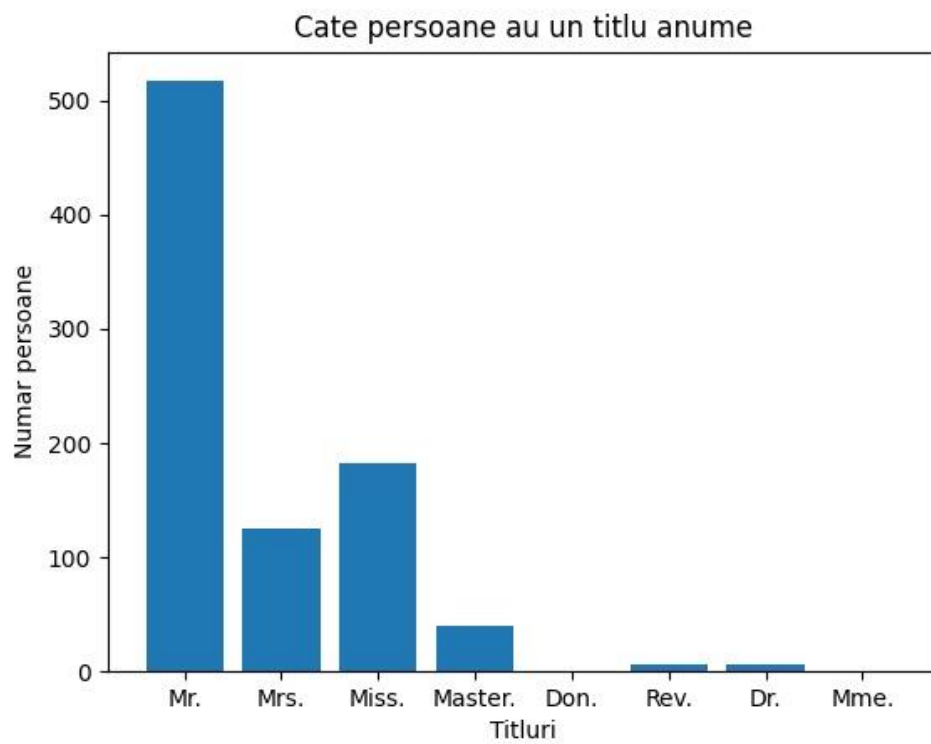


Cerinta 7 : la cerinta 7, calculez procentul de copii si adulti care au supravietuit, graficul :



Cerinta 8 : la cerinta 8 valorile lipsa sunt umplute cu media (pentru "Age") sau cea mai frecventa valoare (pentru "cabin" si "Embarked")

Cerinta 9 : Am extras toate titlurile unice si am construit o lista "genuri" in care pentru orice pozitie avem genul corespunzator titlului de pe aceeaasi pozitie din lista de titluri unice. Initializez o lista cu un contor pentru fiecare titlu si trecand prin tot dataframe-ul, incrementez contorul corespunzator titlului daca corespunde cu genul. Impart afisarea in 2 subplot-uri pentru lizibilitate.



Cerinta 10 : Calculez rata de supravietuire pentru cei care au supravietuit din categoriile : singur, acompaniat si afisez graficul care arata diferente de 2%, foarte mici. Apoi afisez relatia dintre tarif, clasa si starea de supravietuire pentru primele 100 de intrari din dataframe folosind catplot. Starea de a fi singur pe TItanic pare sa nu influenteze sansele de supravietuire, diferenta fiind de 2%. In graficul care arata relatia dintre tarif clasa si daca s-a supravietuit sau pare pare ca majoritatea persoanelor care nu au supravietuit sunt in clasa a 3-a

