

PCLP 2

TITANIC

Pt. I

Cerinta 1:

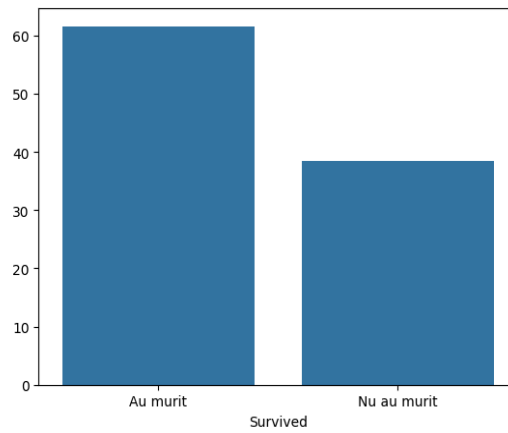
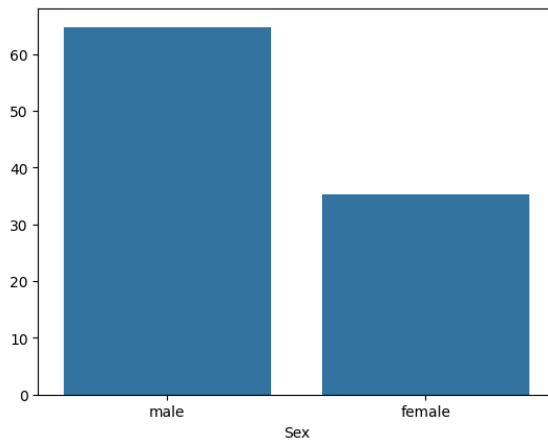
Incepem prin citire setului de date. Apoi aflam numarul de coloane si fiecare tip de data pentru fiecare coloana, numarul de linii, numarul de valori lipsa si daca exista linii duplicate.

```
print(tip_val_coloana)
print(nr_coloane)
print(nr_linii)
print(linii_duplicate)
print(valori_lipsa)
```

```
PassengerId      int64
Survived          int64
Pclass           int64
Name             object
Sex              object
Age              float64
SibSp            int64
Parch            int64
Ticket           object
Fare             float64
Cabin            object
Embarked         object
dtype: object
12
891
0
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin            687
Embarked         2
```

Cerinta 2:

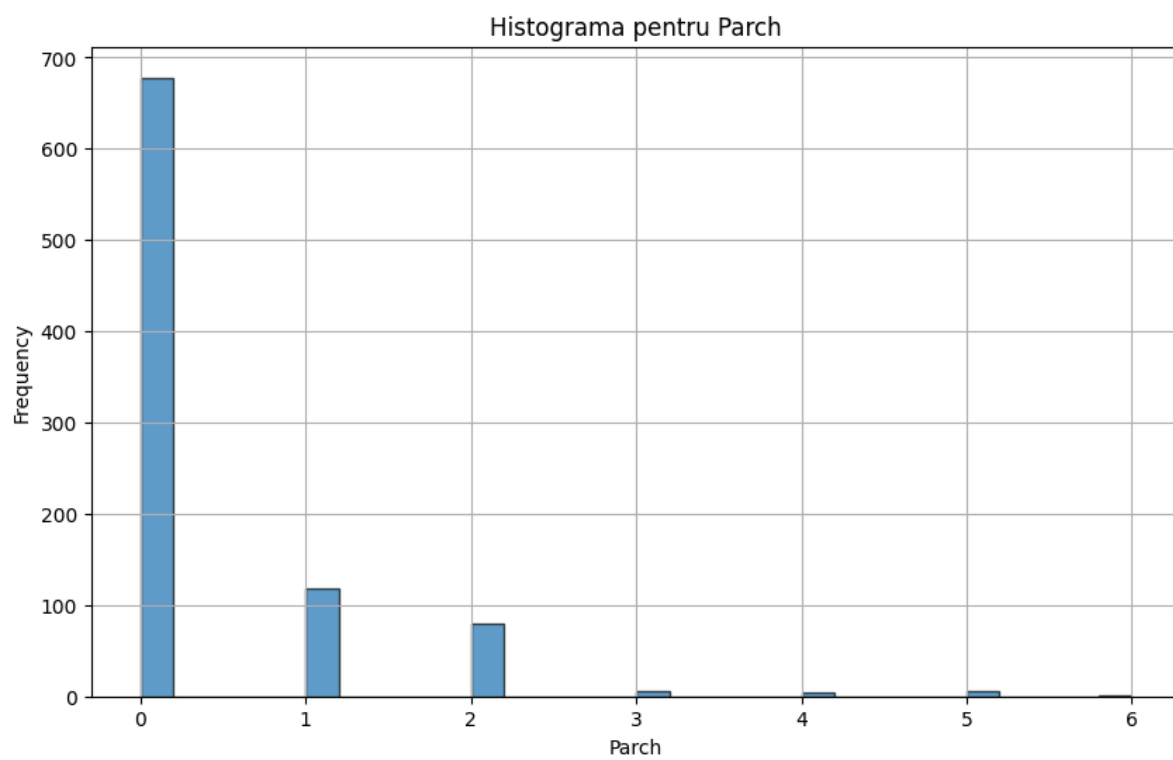
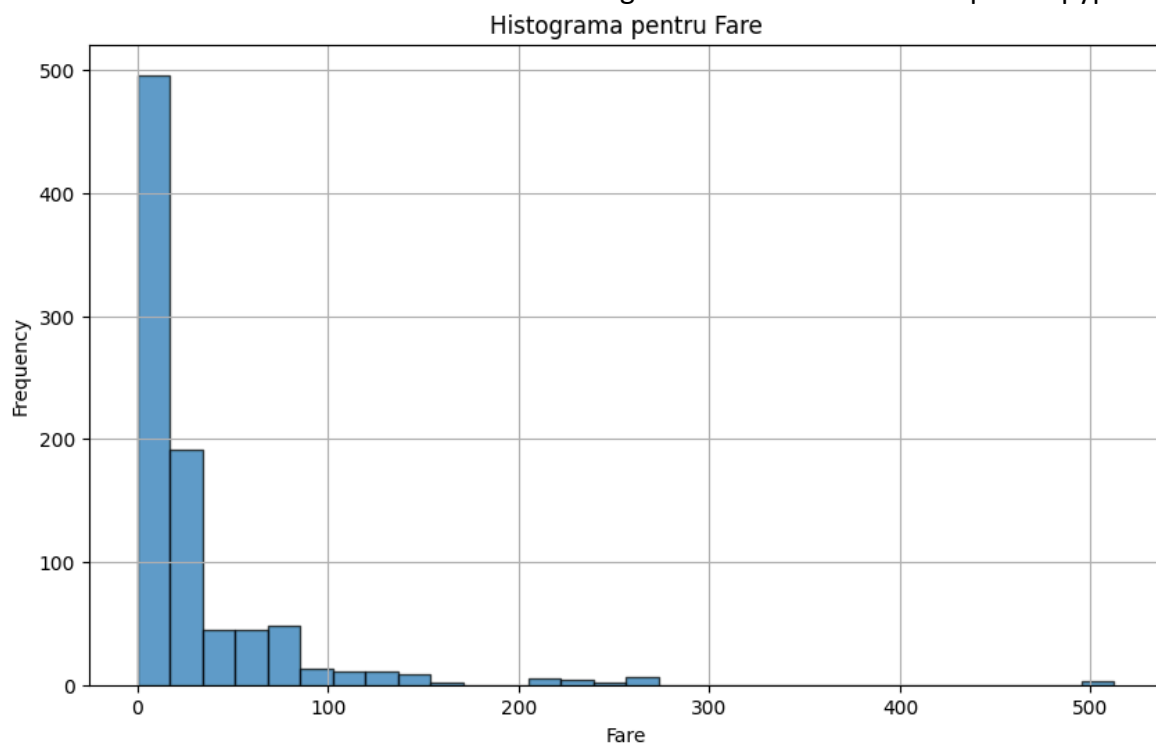
Pentru fiecare data din coloana Survived (0 si 1) si fiecare data din Coloana PClass (1 ,2 si 3) si coloana Sex(male sau female), numaram aparitia si calculam procentul.

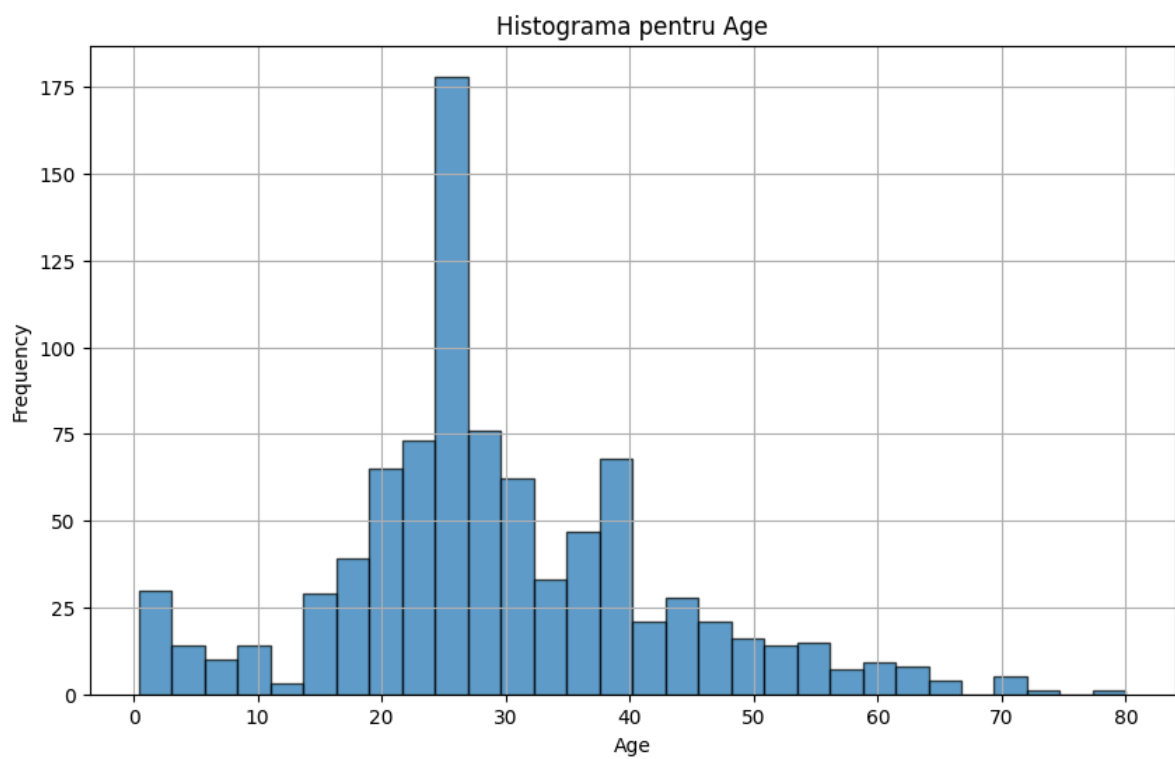
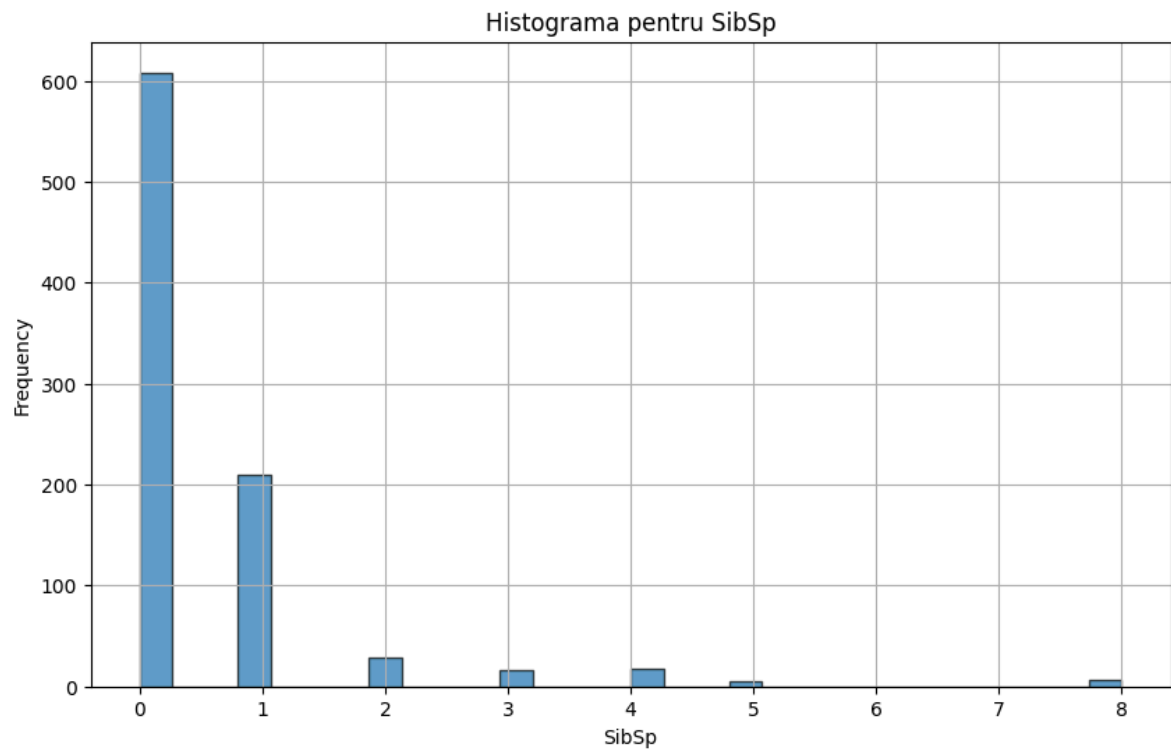


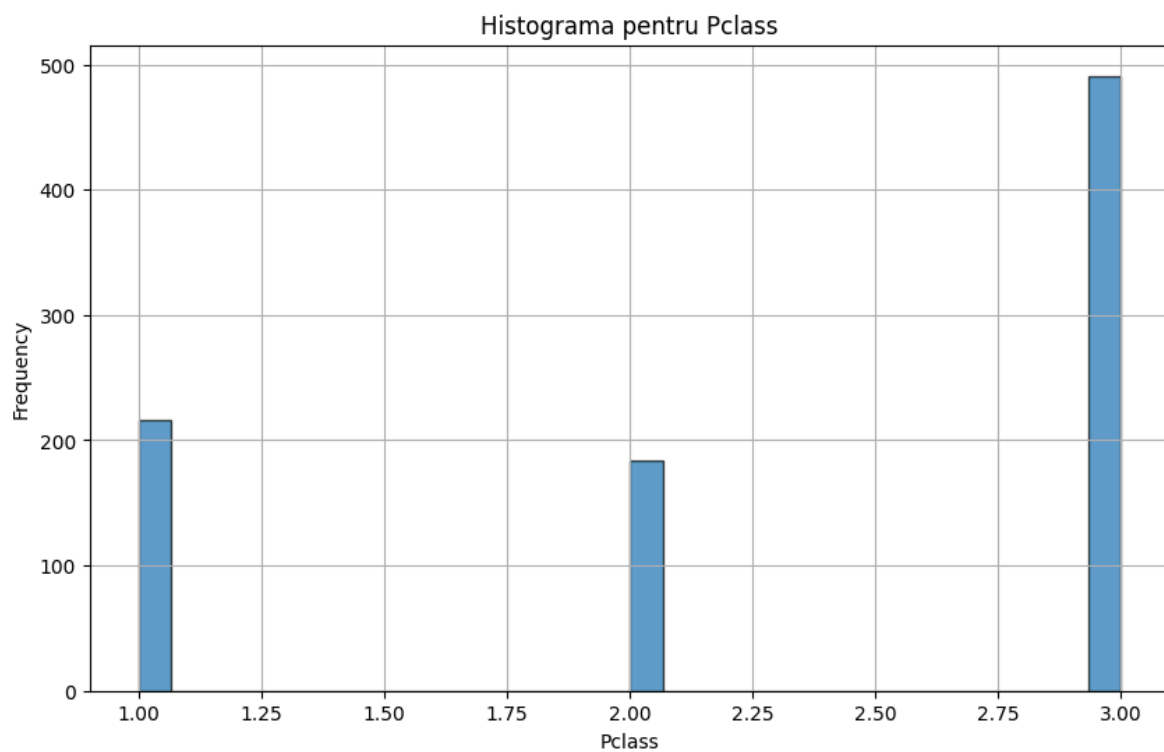
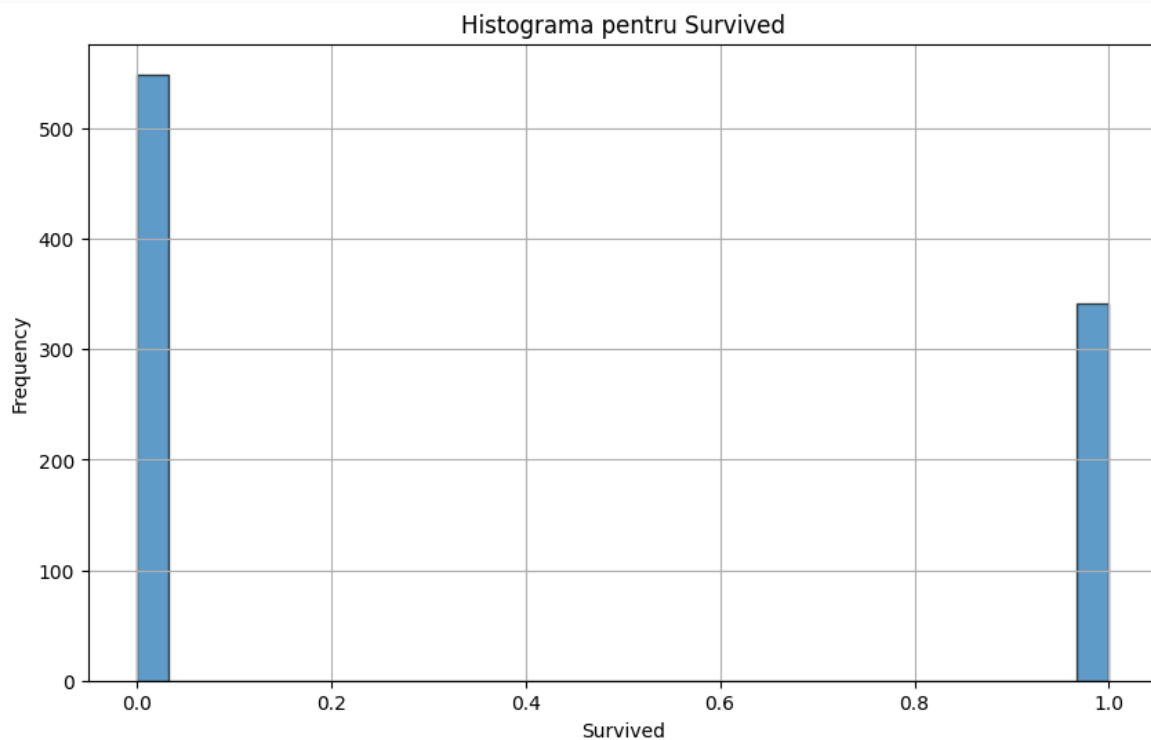
```
Sex
male    64.758698
female  35.241302
Name: proportion, dtype: float64
Pclass
3    55.106622
1    24.242424
2    20.650954
Name: proportion, dtype: float64
Survived
0    61.616162
1    38.383838
Name: proportion, dtype: float64
```

Cerinta 3:

Pentru fiecare coloană numerică realizăm o Histogramă folosind biblioteca matplotlib.pyplot.

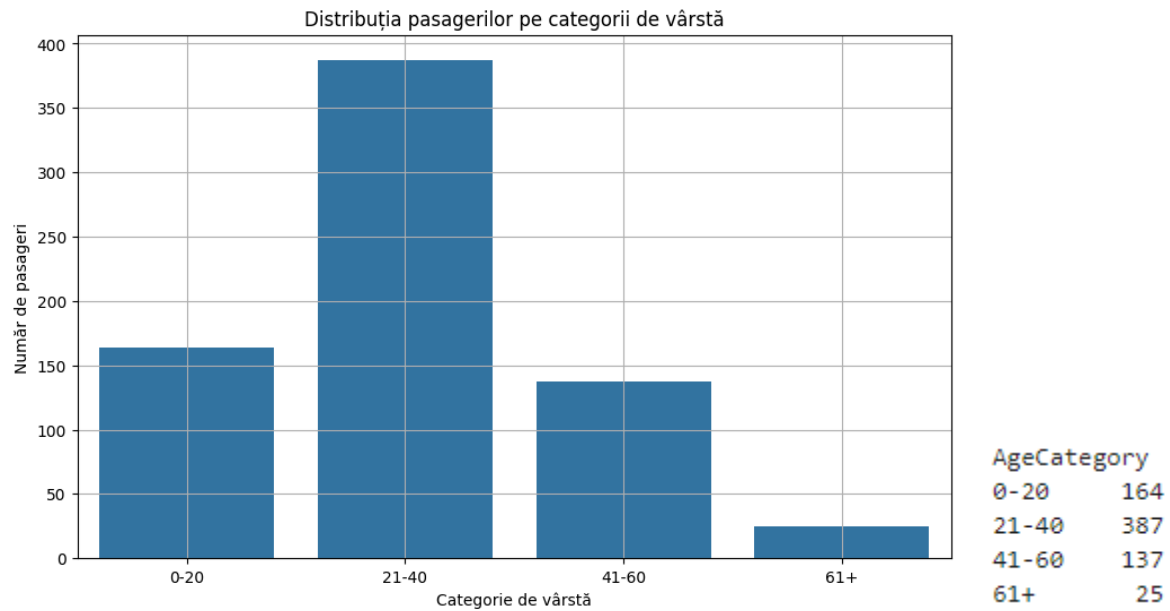






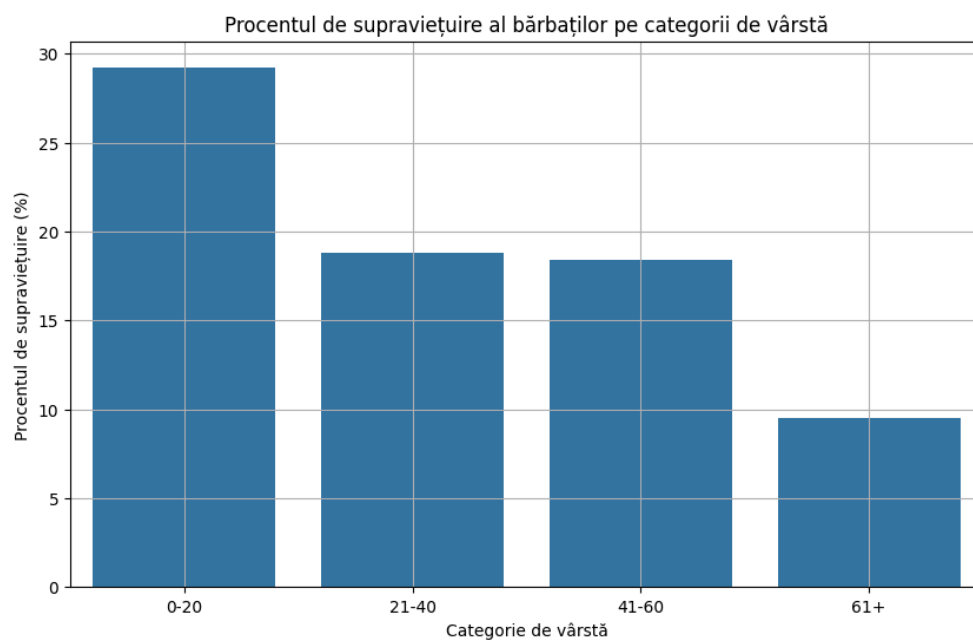
Cerinta 5:

Impartim varsta pasangerilor in categorii si afisam distributia lor intr-un grafic. Afisam si numararea pentru fiecare categorie



Cerinta 6:

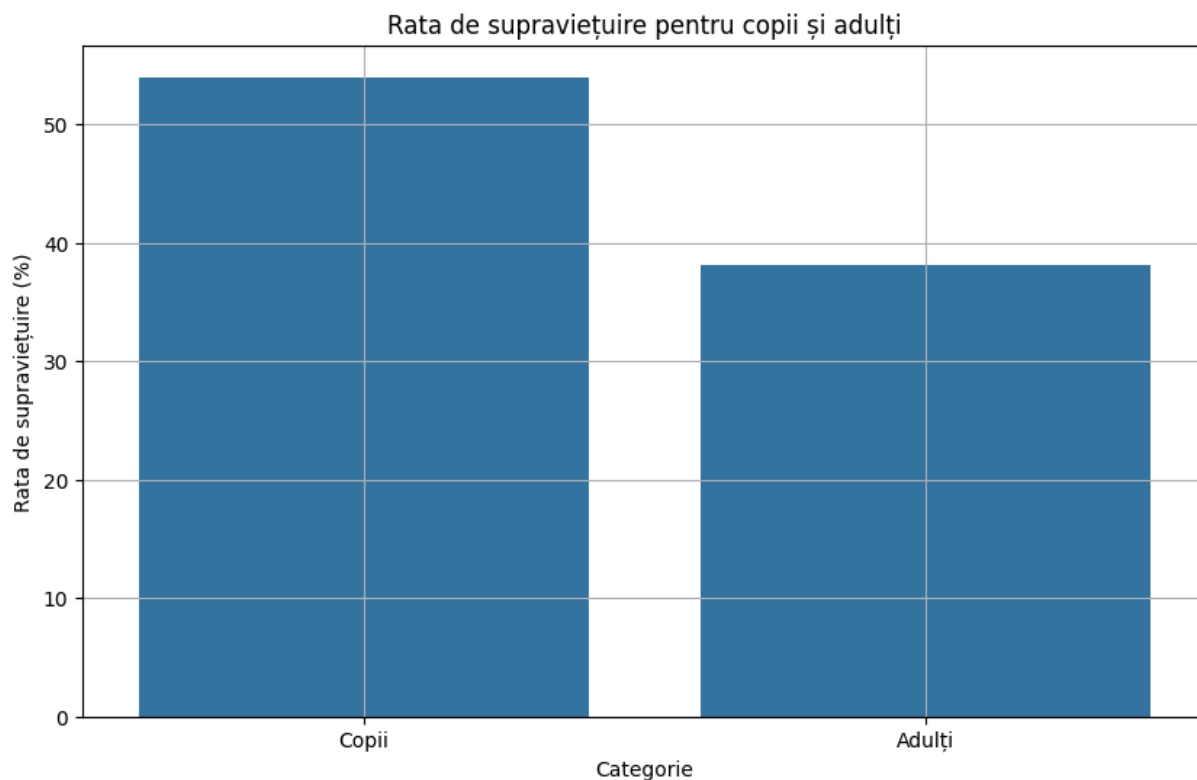
Adaugam coditia sa fie coloanal Sex male



Cerinta 7:

Calculam cati oameni au sub 18 ani si apoi realizam un grafic

Procentul copiilor aflați la bord: 12.68%



Cerinta 8:

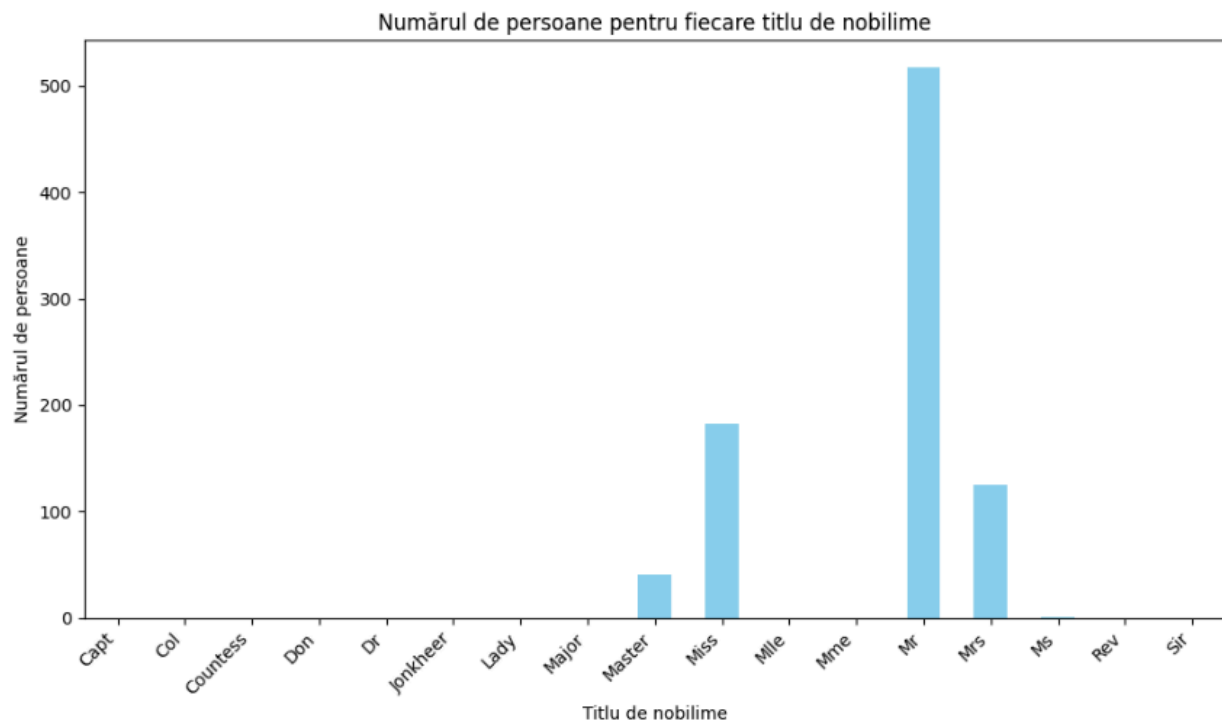
Completam valorile lipsa cu media(valoare int) si realizam un nou fisier cu datele complete

Cerinta 9:

Urmatoarea linie de cod adauga o coloană numită 'Verificare' care evalueaza condițiile pentru a verifica daca titlul de nobilime aparține unui anumit set de titluri ('Mr', 'Mrs', 'Miss', 'Ms', 'Master') pentru bărbați sau ('Mrs', 'Miss', 'Ms') pentru femei și daca sexul corespunde cu titlul.

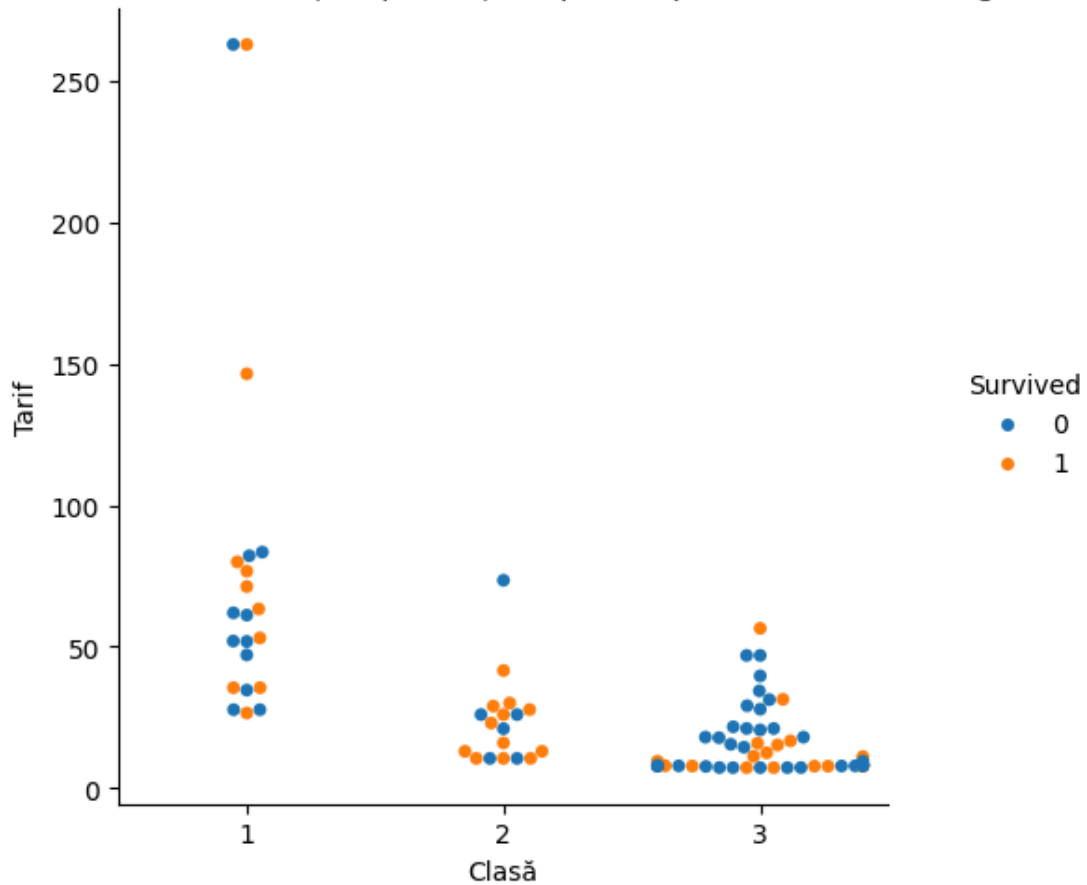
Apoi, numarul de persoane pentru fiecare titlu este calculat folosind metoda groupby() pentru a grupa datele dupa titlu și apoi sum() pentru a numara cate valori sunt True in coloana 'Verificare'

pentru fiecare grup. Aceste numere sunt stocate in variabila 'numar_persoane'



Cerinta 10:

Relația dintre tarif, clasă și supraviețuire pentru primele 100 de înregistrări



Cerinta 5:

Numaram valorile lipsa dinm fiecare coloanal si calculam procentul:

```
Coloana "Age":  
Numărul de valori lipsă: 177  
Procentul de valori lipsă: 19.87%
```

```
-----  
Coloana "Cabin":  
Numărul de valori lipsă: 687  
Procentul de valori lipsă: 77.10%
```

```
-----  
Coloana "Embarked":  
Numărul de valori lipsă: 2  
Procentul de valori lipsă: 0.22%
```

```
-----  
Coloana "AgeCategory":  
Numărul de valori lipsă: 178  
Procentul de valori lipsă: 19.98%
```