

Cîrstian Daniel 311CB

Readme – Proiect final PCLP3 – Partea 1

Pentru rezolvarea proiectului am construit 11 fisiere .py, unul pentru main, si 10 pentru fiecare task in parte, astfel am modularizat codul si a fost mai usor sa fac debugging.

De asemenea, proiectul a fost realizat folosind github, am create un repository privat, pe care il va vedea asistentul de laborator la prezentarea proiectului. Fiecare commit a fost semnat. Profilul meu de github: <https://github.com/DaniGM32>.

Pentru rezolvarea primei parti a acestui proiect m-am folosit de cateva biblioteci foarte cunoscute din acest limbaj de programare

- Pandas -> pentru prelucrarea datelor dintr-un set de date de format csv
- Matplotlib -> pentru graficele pe care le voi prezenta ulterior
- Numpy -> pentru prelucrari matematice, in special operatii

Fiecare dintre aceste biblioteci au functii deja implementate, foarte utile in rezolvarea cerintelor, spre exemplu functii precum isnull() care verifica daca elementele de pe o coloana sunt diferite de 0 sau nu, sau groupby(), o functie foarte utila cand vreau sa selectez elemente care respecta o anumita caracteristica de pe o anumita coloana.

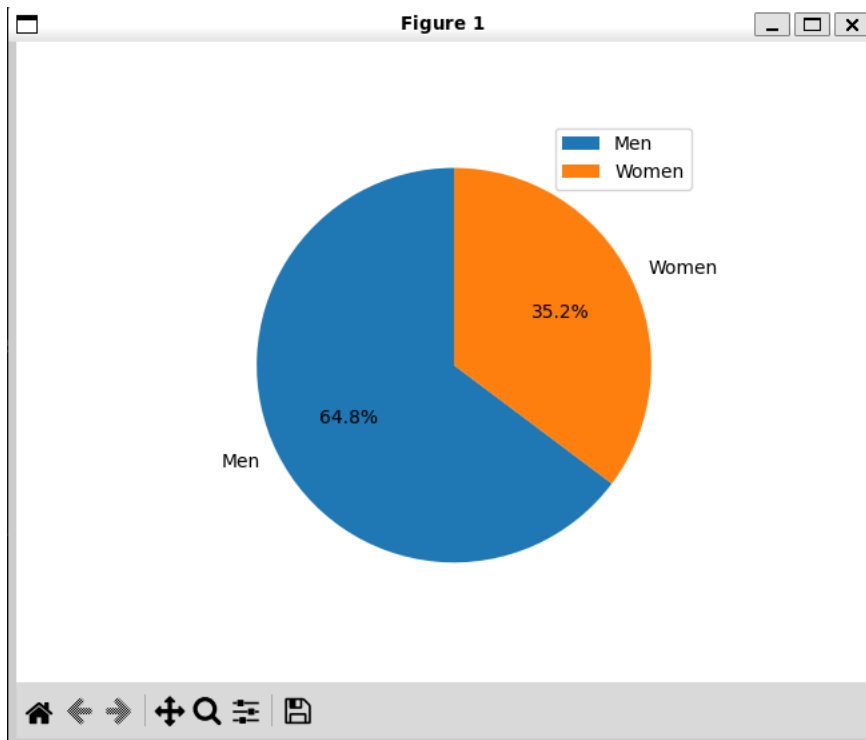
Inainte de a prezenta graficele, mentionez ca am avut dificultati in realizarea graficelor, intrucat lucrez folosind wsl, raspunsul l-am gasit pe stackoverflow, site-ul unde am gasit si multe din functiile pe care le-am mentionat mai sus, alaturi de exemple. Pentru a afisa aceste grafice a trebuit sa instalez mai mult extensii precum PyQt6 si PySide2, dar si sa adaug aceste linii de cod la inceputul programului:

```
matplotlib.use('Agg')  
plt.switch_backend('TkAgg')
```

Mai departe voi prezenta graficele si histogramele obtinute in urma rezolvarii taskurilor si voi explica pe scurt rezolvarea acestora

Pentru taskul 1 am folosit functii predefinite din python, care corespund clasei de data_set (sau data frame)

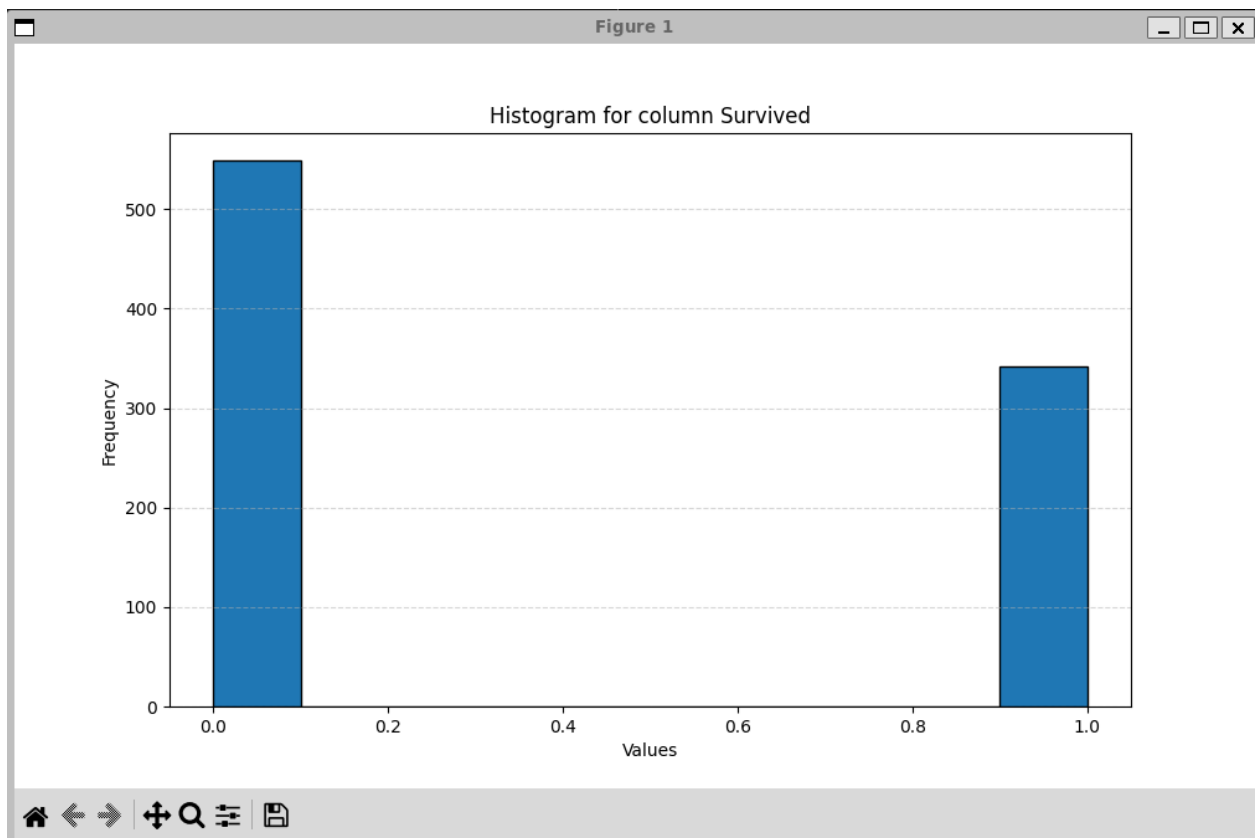
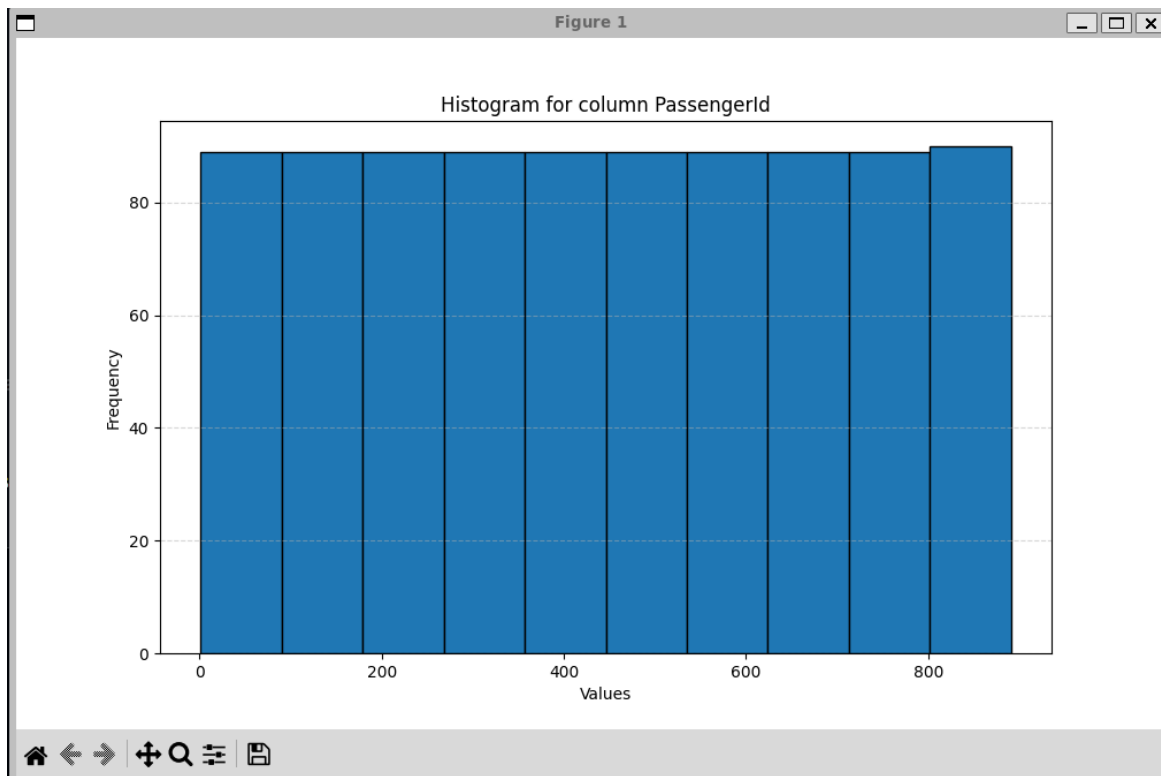
Grafic tip pie pentru taskul 2:

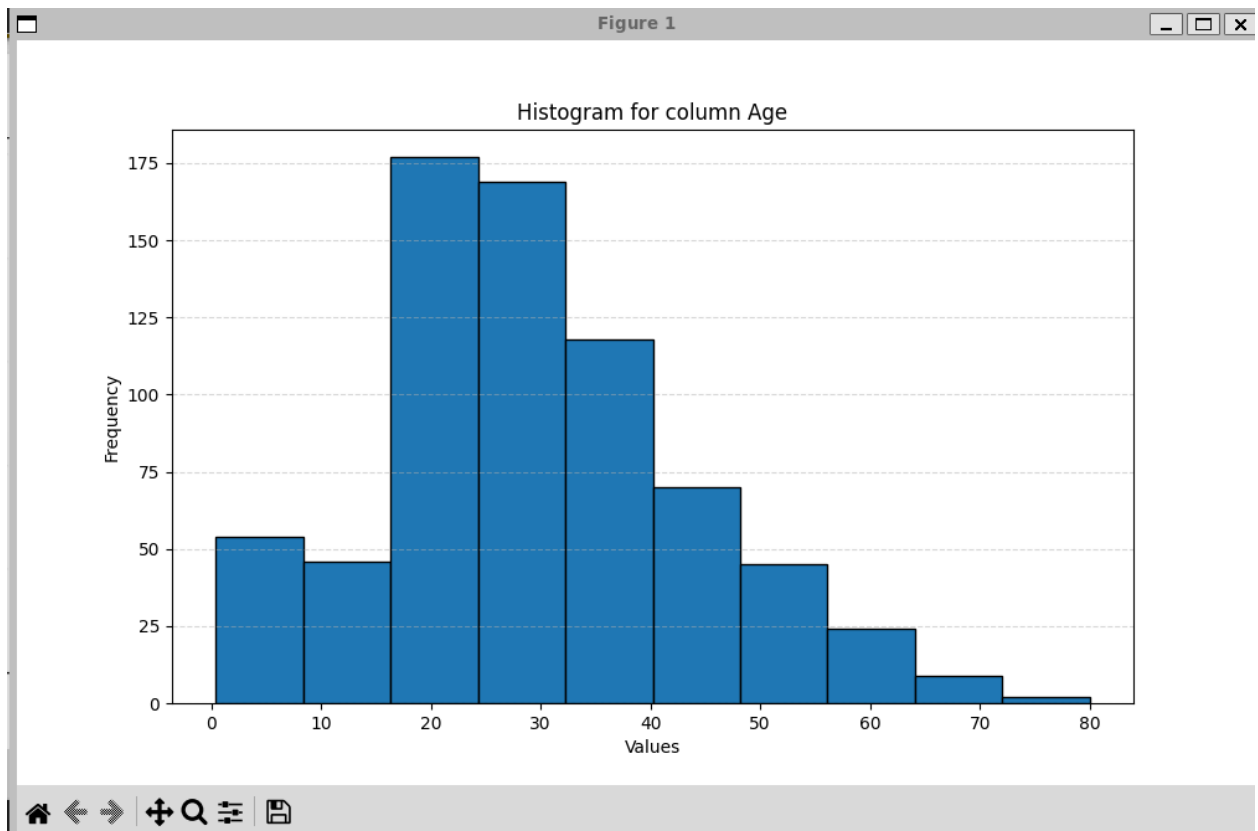
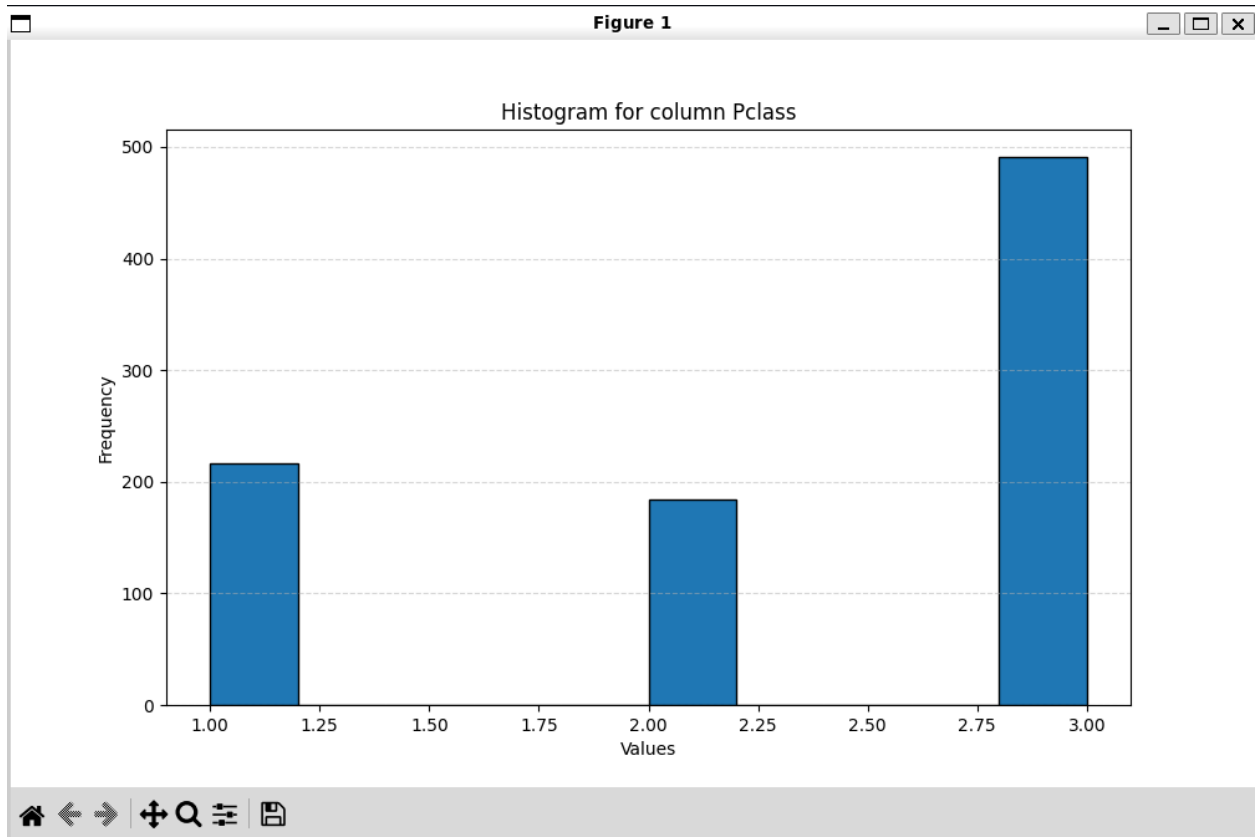


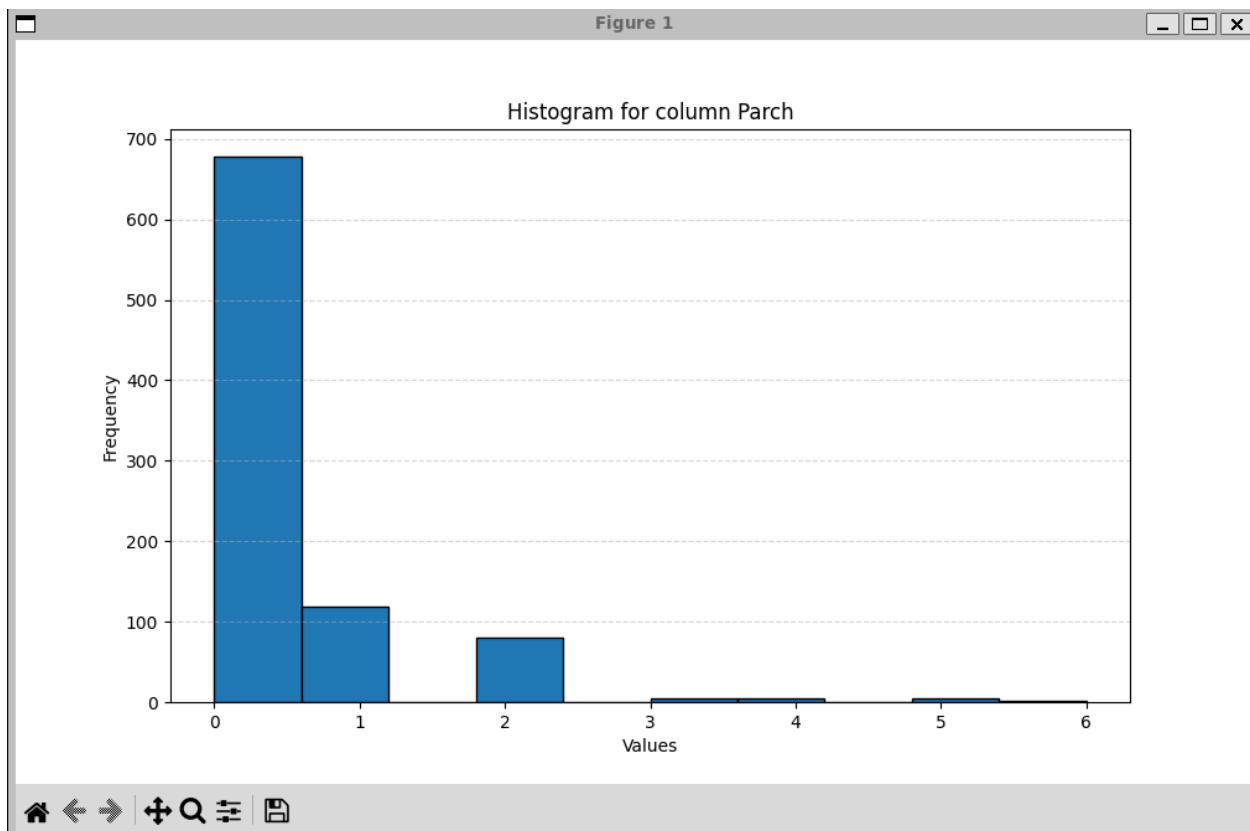
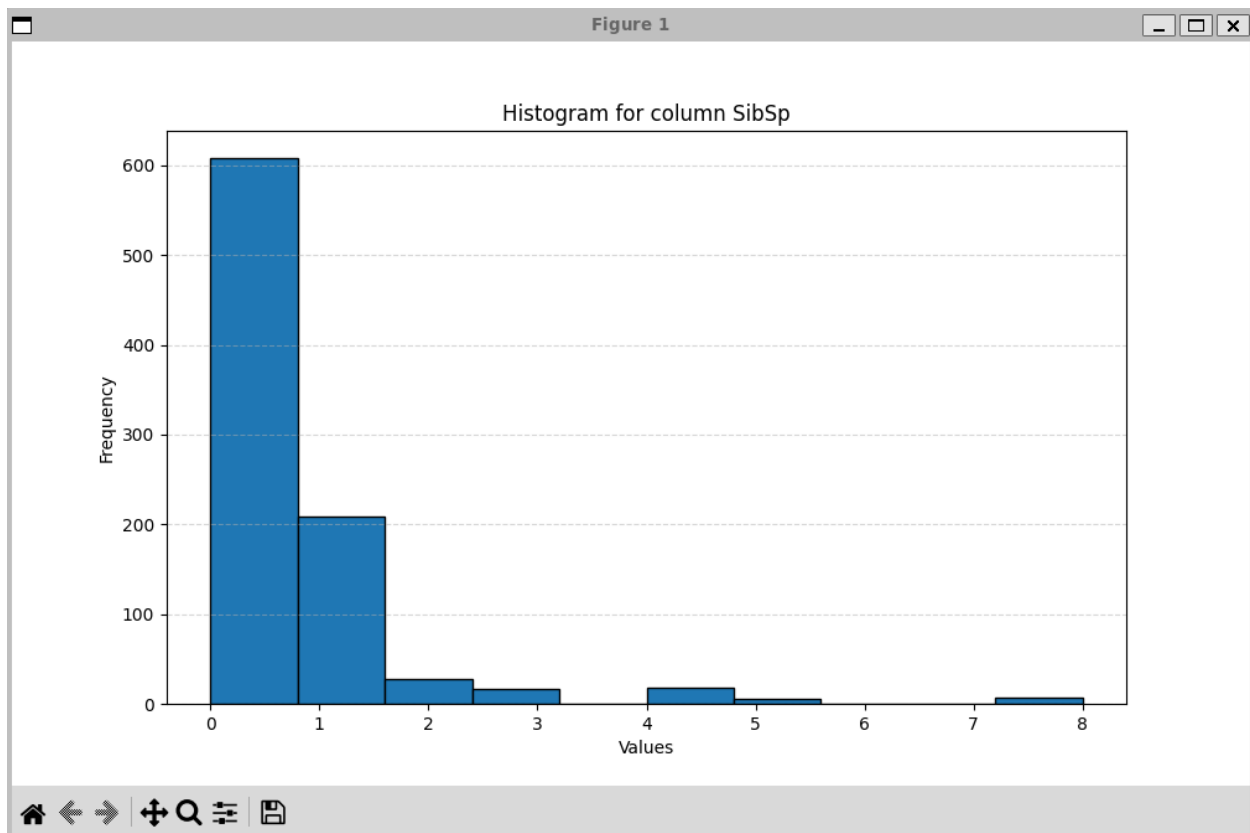
- Procentajul de barbati, respectiv femei de la bordul navei
- Am folosit groupby pentru a extrage coloanele dorite si am calculat procentajul rotunjit cu 3 zecimale la fiecare categorie

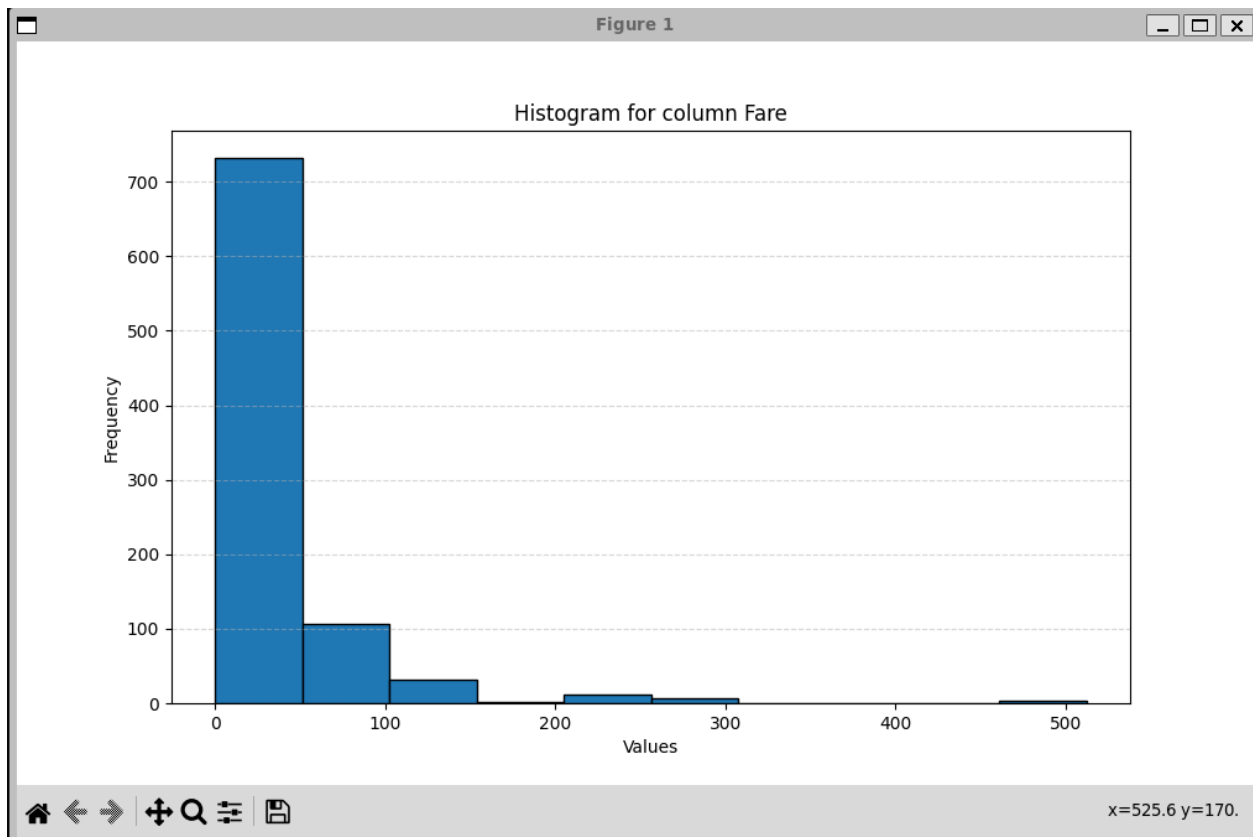
Histogramamele pentru taskul 3 (pentru fiecare coloana numerica).

Am extras coloanele care au valori numerice folosind functia `select_dtypes` si am construit graficele aferente:





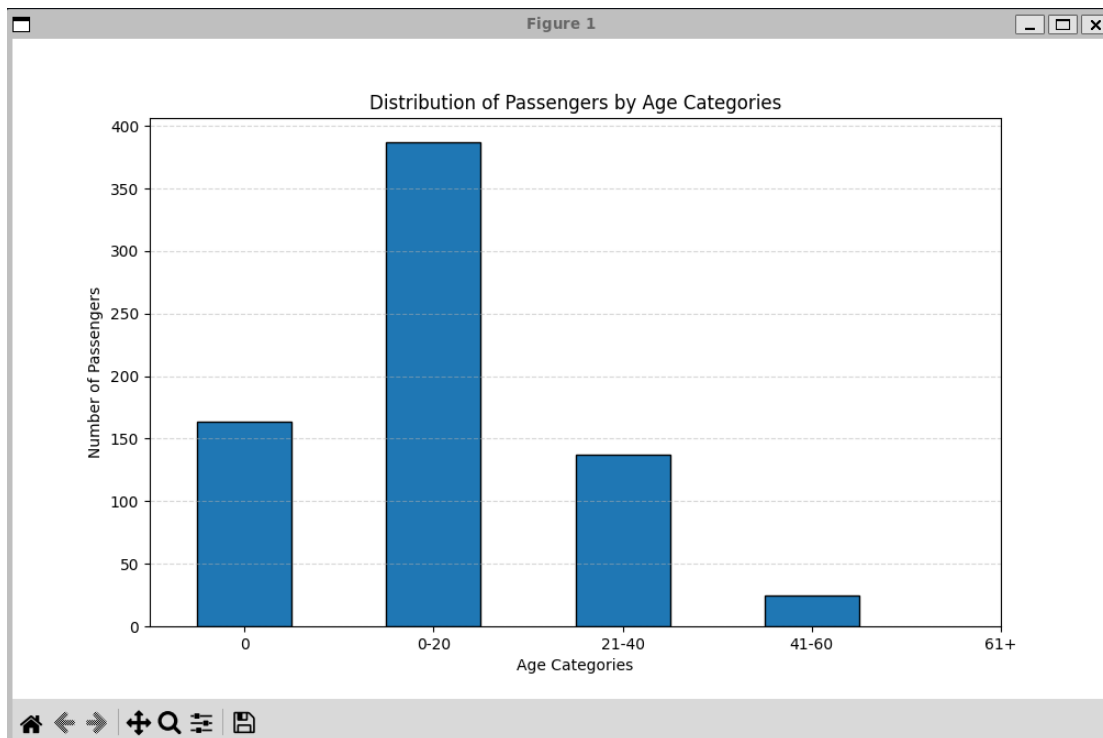




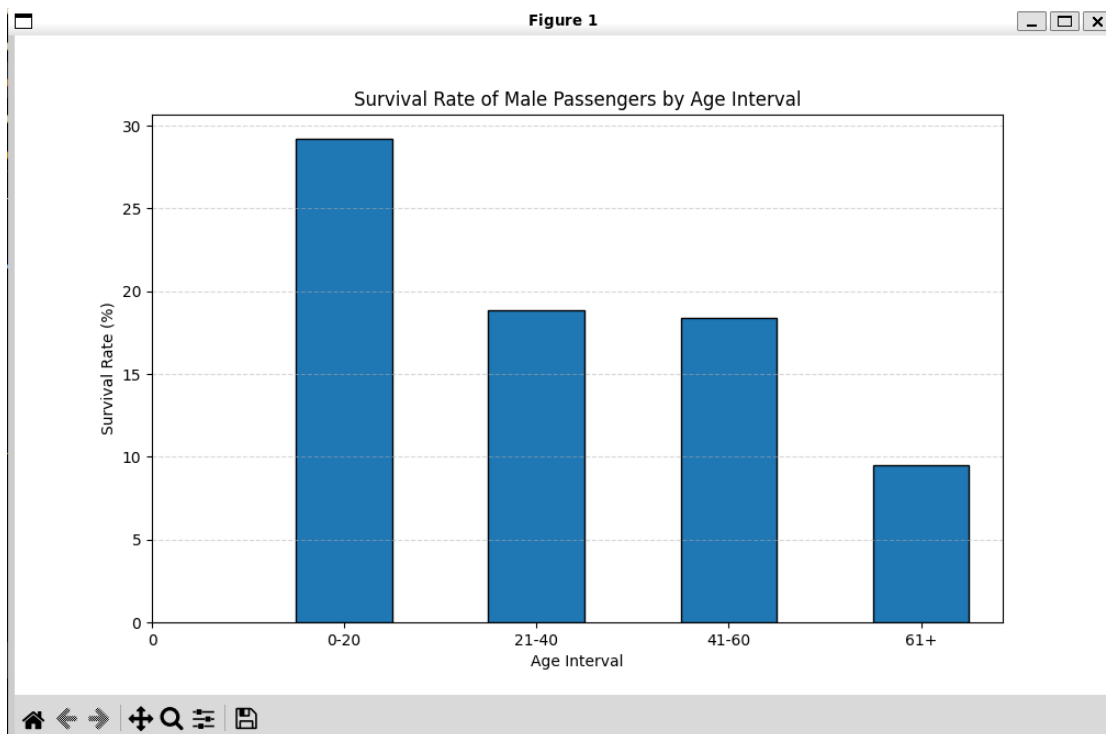
Explicatii task 4:

La fel ca la alte functii am intors rezultatul intr-un dictionar, m-am folosit de functii predefinite ca `isnull()` si `sum()` apelate impreuna pentru a determina nr de persoane care au supravietuit sau care au murit, dar si nr de valori lipsa in privinta asta

Graficul pe categorii de varsta de la taskul 5 (pe cele 4 intervale) :

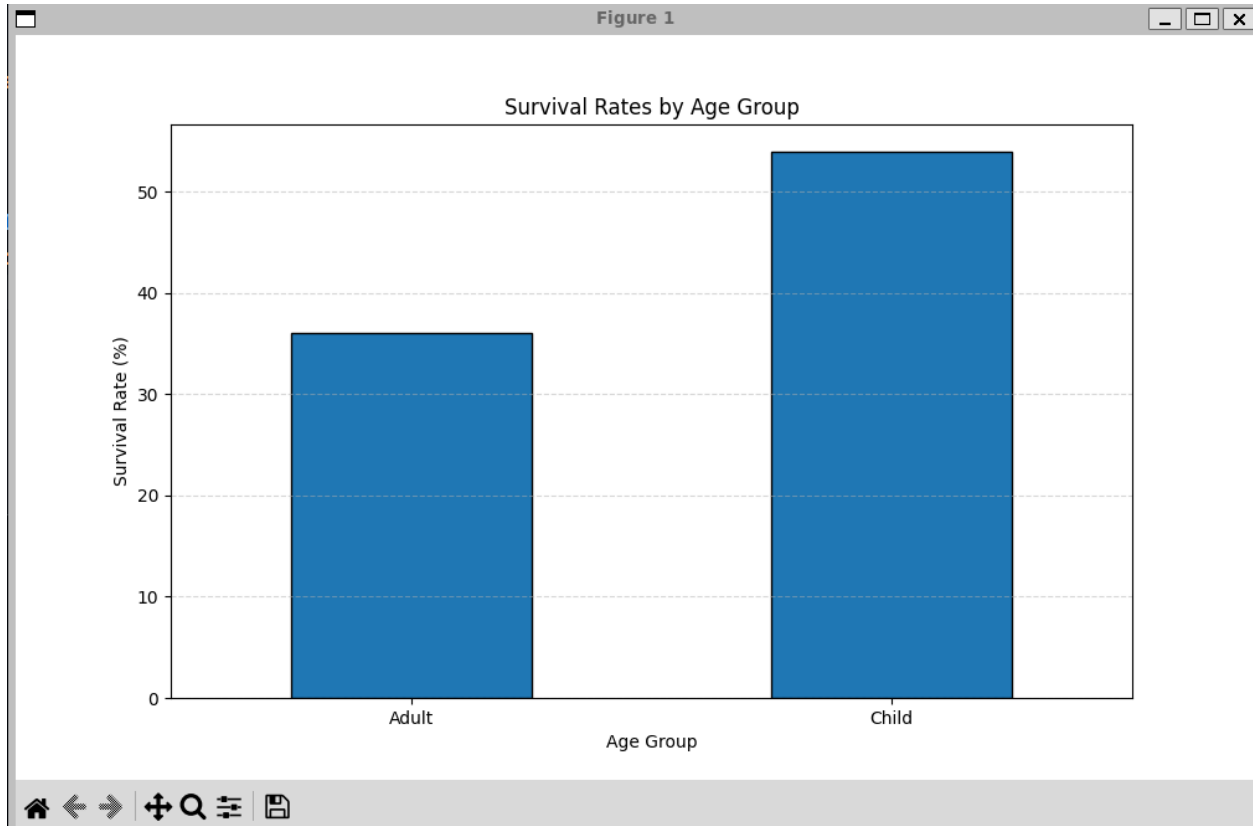


M-am folosit de functia cut din pandas pentru a delimita intervalela
Graficul cu rata de supravietuire pe intervale de varsta de la taskul 6:



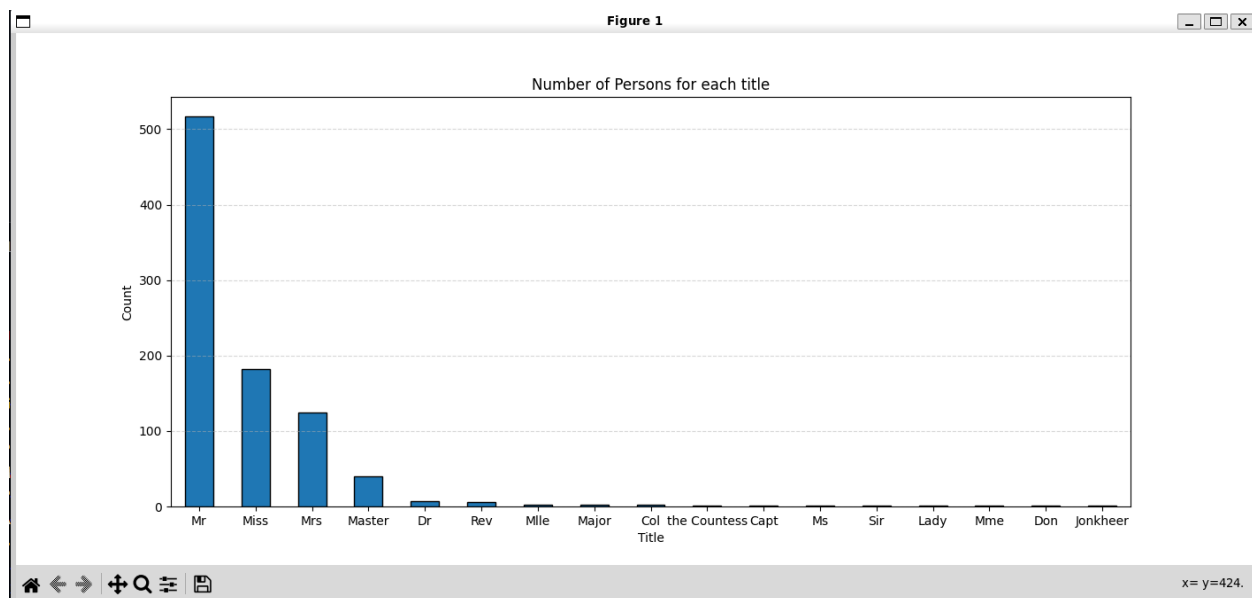
Am adaugat o noua coloana la data_set (Age_Interval), pe care am prelucrat-o dupa ca
sa obtin graficul de mai sus

Rata de supravietuire a copiilor fata de cea a adultilor (task 7):

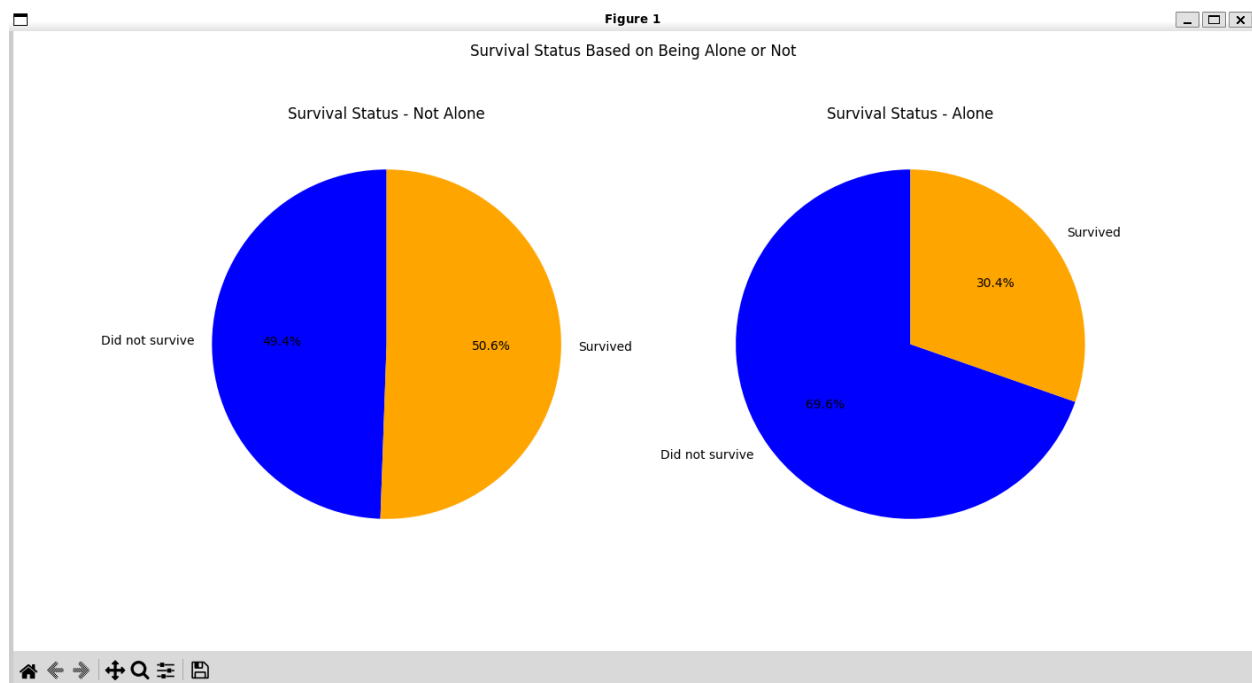


Am facut acelasi lucru ca la taskul anterior doar ca am aplicat o functie lambda pe coloana Age ca sa imparta persoanele in 2 categorii: adulti si copii

Graficul de la taskul 9:



Pentru taskul 10 am analizat daca faptul ca pasagerii erau insotiti sau nu de familie a avut vreo influenta sau nu asupra ratei de supravietuire, am realizat 2 grafice in paralel:



De asemenea, folosind biblioteca seaborn, mai exact functia catplot(), am realizat corelatia dintre tariful de clasa si starea de supravietuire pentru primele 100 de inregistrari:

