

Popescu Ionut-Teodor 311CC

Proiect-Pclp3

Partea I

La acest proiect am reusit sa fac 8 din cele 10 cerinte,cele pe care nu am stiut sa le fac fiind cerinta 3 ,cea cu histogramele si cerinta 8,cea in care trebuia sa completam valorile lipsa cu media altor pasageri.

Pentru cerinta 1 nu e mult de explicat,am dat df.info pentru informatii despre tipul coloanelor.Apoi am folosit .isnull pentru a vedea unde sunt valori lipsa iar apoi .sum pentru a le numara pentru fiecare coloana.La fel am facut si pentru numarul de linii duplicate.

La cerinta 2 am folosit .valuecounts pe coloane care numara cate intrari diferite am,ceea ce e bine pt coloane la care vreau sa vad cate persoane am de fiecare tip.Am calculate cat reprezinta din total iar apoi am inmultit cu 100 pentru a afla procentajul.Templateul de grafic 1 am luat de pe net si din lab ,si l am modificat cu datele din cerinta.

La cerinta 4 am facut numarul de coloane care au valori lipsa, si am calculate procentul din total,iar apoi am facut cu .groupby pentru coloanal Survived si dupa am luat coalnele de interes si am aplicat o functie generica lambda ,in care transform in 1 si 0 valorile,iar apoi fac suma cu .sum dupa care impart la numarul total de oameni si transform in procentaj cu *100.

Cerinta 5 consta in impartirea pe grupe de varsta.Am folosit o functie separate pentru a calcula cati oameni avem pt fiecare categorie de varsta.Apoi calculam distributia cu .value_counts si aflam proportia si facem *100 pt procentaj.

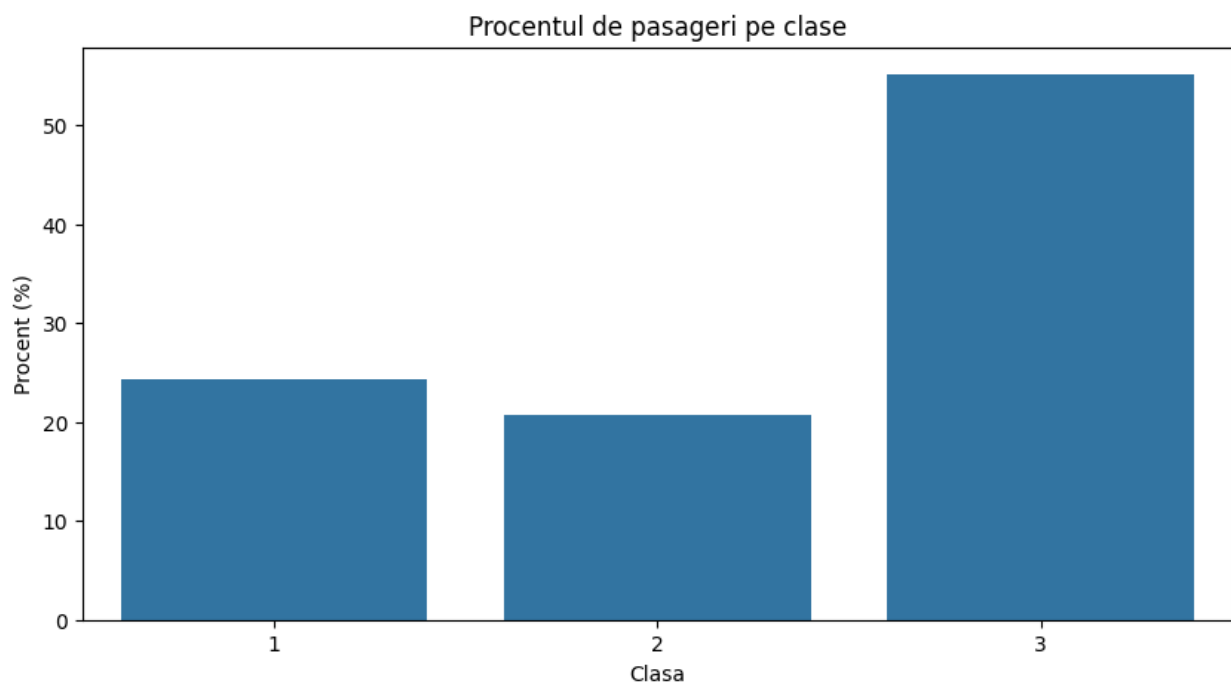
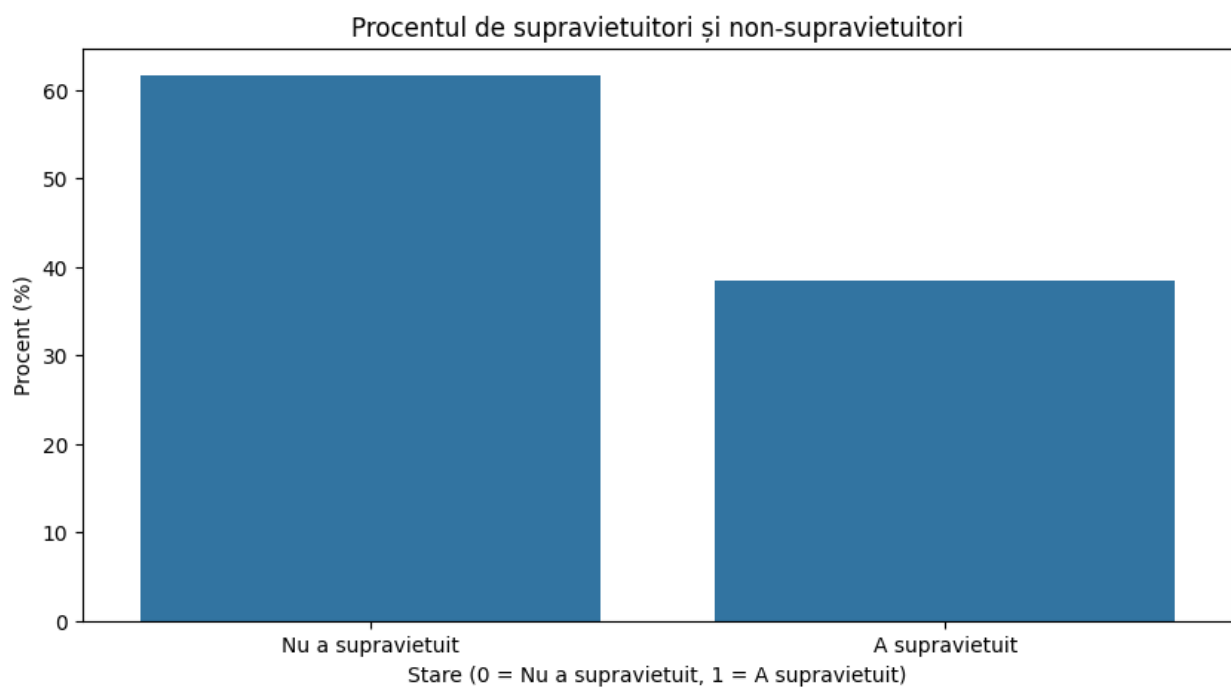
La cerinta 6 am preluat de la 5 datasetul ,iar apoi am adaugat conditia ca sexul sa fie male,dupa care am folosit groupby si sum pentru a vedea exact nr de barbati care au supravietuit din fiecare categorie de varsta,iar apoi am folosit metoda de mai sus ,luand cati barbati au

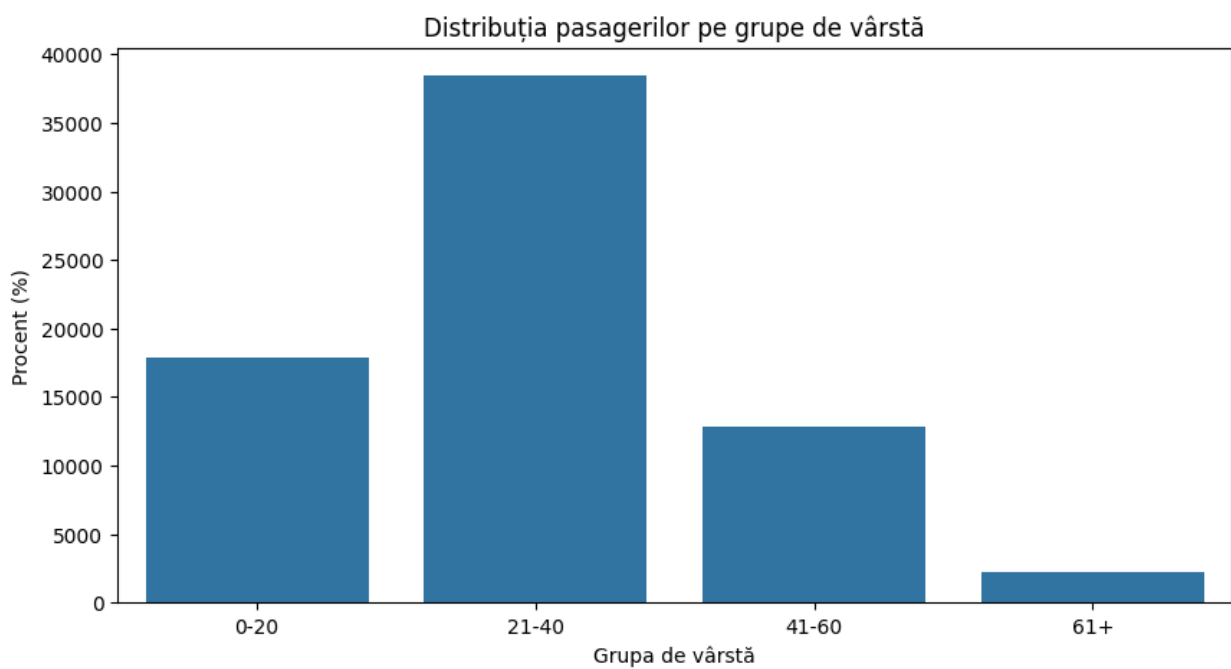
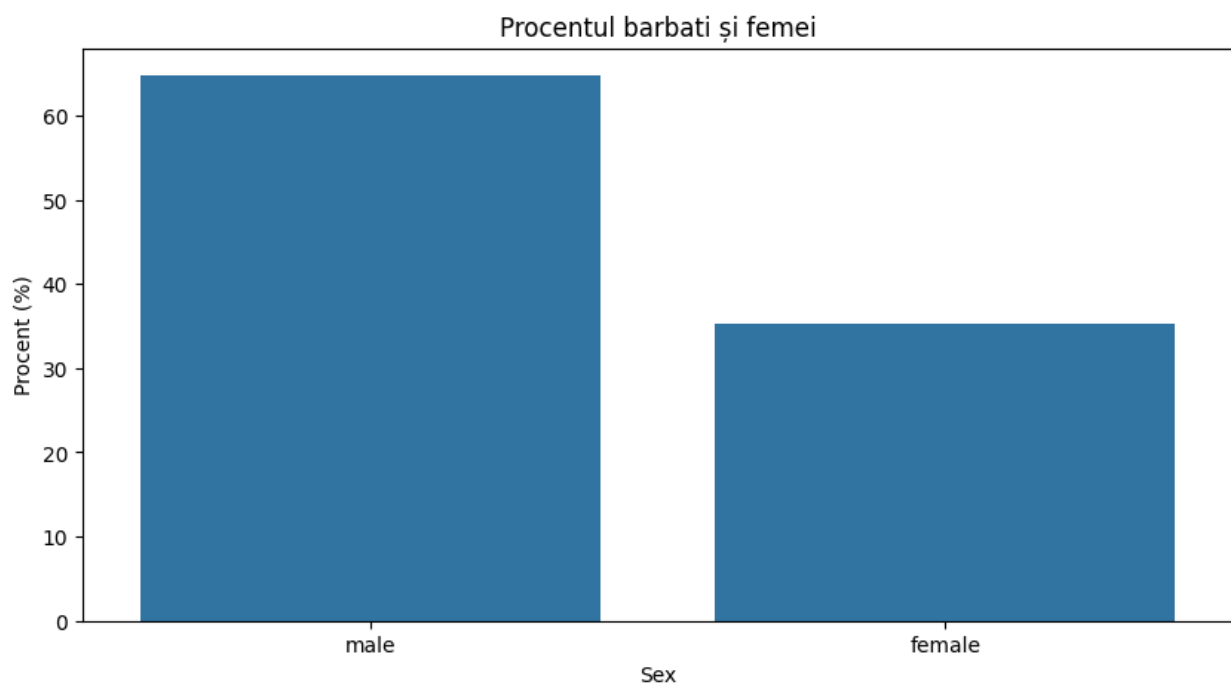
supraviețuit și împartim la câți avem în total pentru fiecare categorie de vârstă și am $\cdot 100$ pentru procentaje.

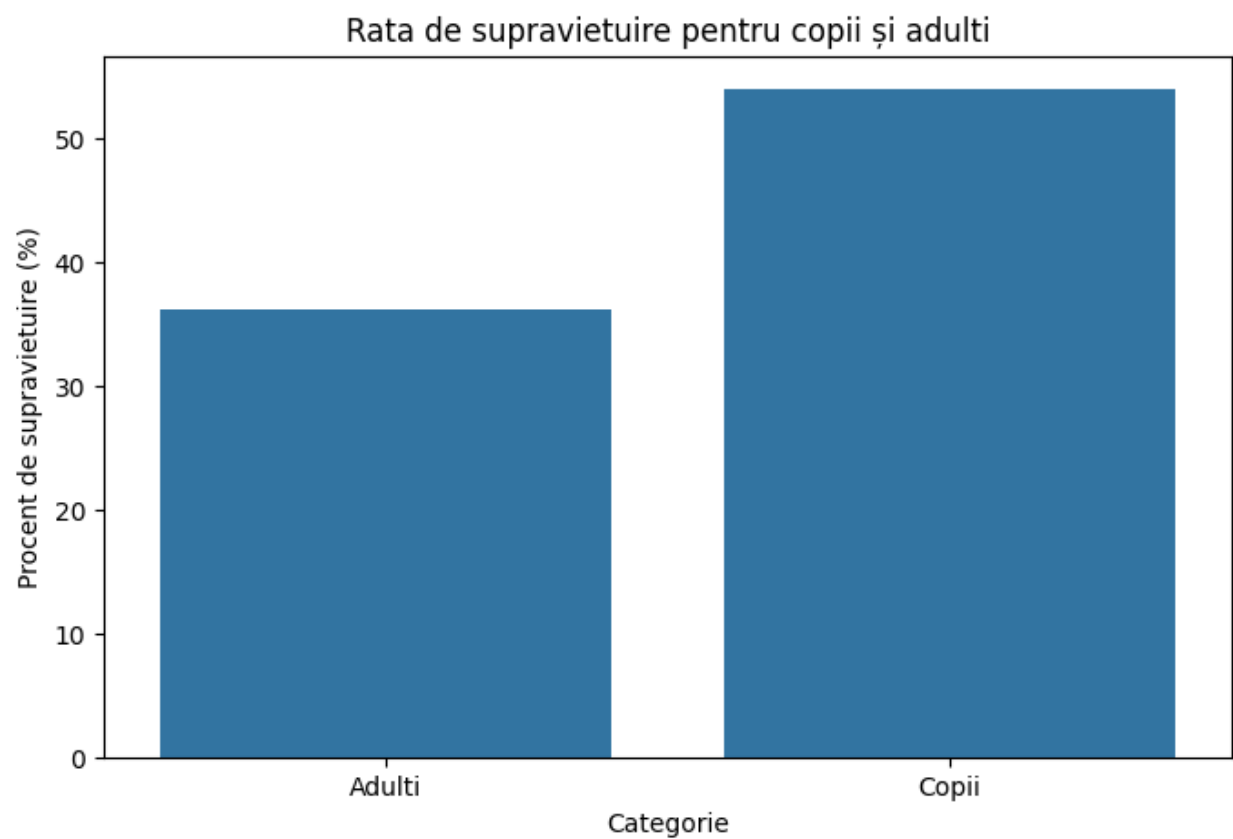
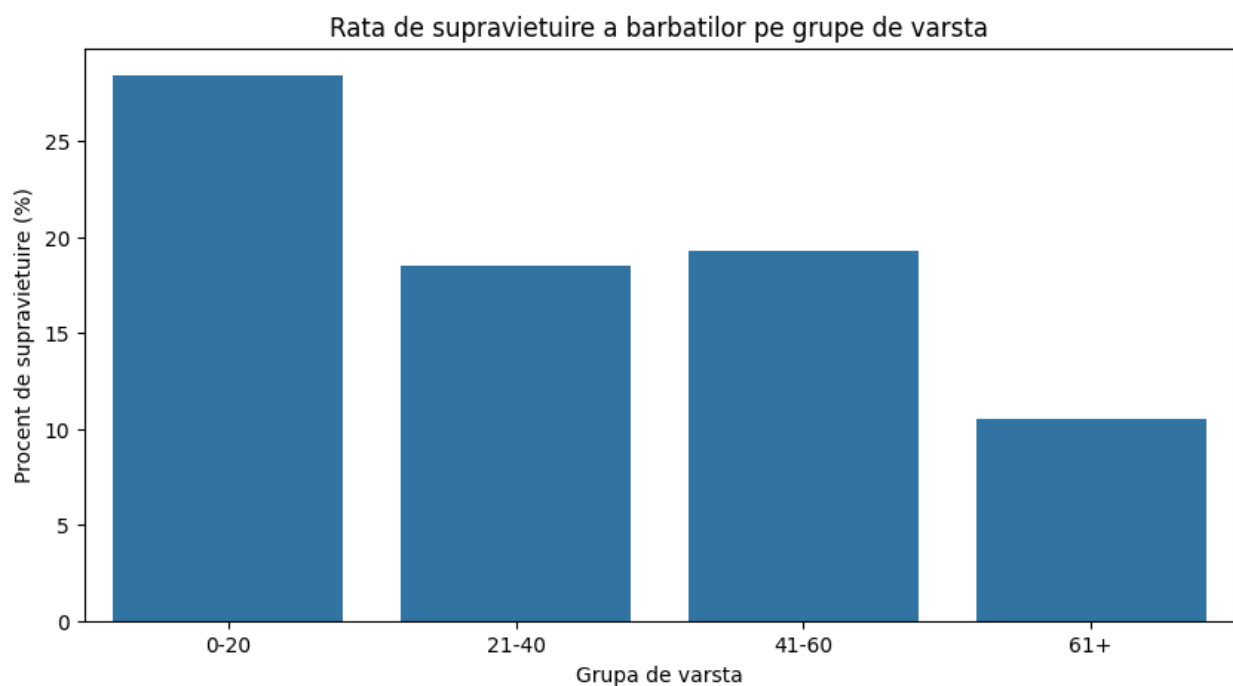
La cerința 7 am aflat câți copii sunt prin creerea unei noi coloane care are toate elementele de pe coloana `age` care sunt mici decât 18. Facem suma pentru valorile diferite de 0 și apoi procentajul copiilor. Pentru diagrama am calculat proporțiile prin copiilor prin adunarea valorilor din coloana `survived` în funcție de coloana `copii`, după care am făcut pentru `survival_rates[true]`, adică copii, procentajul, și pentru false, pentru adulți.

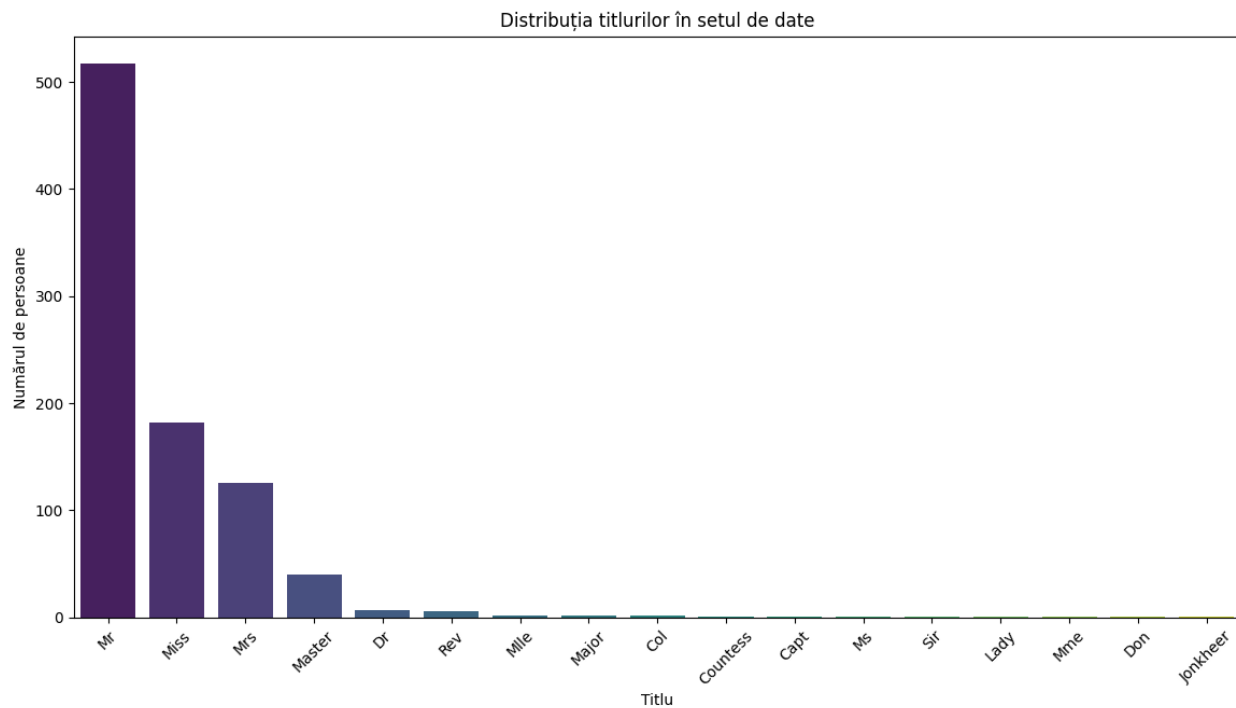
În cerința 9 am folosit regex ca să extrag titlurile din nume, iar apoi am făcut o tabelă de corelație pentru a vedea cum sunt titlurile în funcție de sex. Următoarea parte a fost dificilă: am făcut un dicționar în care am pus titlurile în corelație cu sexul pe care ar trebui să îl aibă, după care am făcut o coloană nouă pentru care am aplicat dicționarul și am pus sexul respective. Apoi am făcut o altă coloană pentru a verifica corectitudinea titlurilor și al sexului sau dacă de exemplu avem doctor care poate avea ambele genuri am folosit `.isnull` pentru a returna true pe coloana `incorrect`. Apoi a trebuit să fac media și astfel am văzut câte titluri din fiecare am și împartim numărul de titluri corecte la numărul total din fiecare, apoi am $\cdot 100$ pt procent.

La cerința 10 am făcut o coloană nouă în care faceam suma coloanelor în care aveam info despre familia sau relațiile pasagerilor. Am luat primele 100 de înregistrări și m-am documentat cum să fac diagrama respectivă.









Relația dintre tarif, clasă și supraviețuire pentru pasagerii singuri și non-singuri

