

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 setat normalize pe true pentru a afisa proportia fiecărei clase,iar apoi am inmultit cu 100 pentru a afla procentajul.Templateul de grafic l am luat de pe net,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 coalnele de interes si am aplicat o functie generica lambda ,in care transform in 1 si 0 valorile,iar apoi fac media cu .mean si transform in procentaj cu *100.

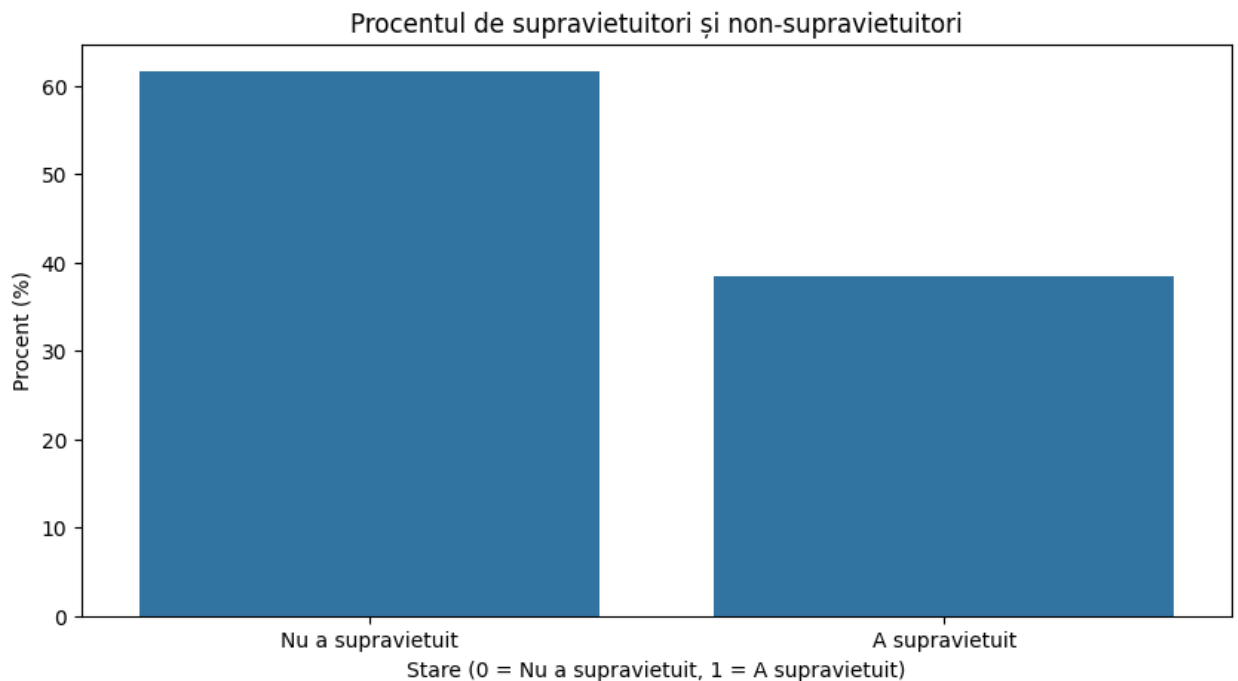
Cerinta 5 consta in impartirea pe grupe de varsta in groups,iar apoi aplicarea pd.cut pentru a segmenta si sorta valorile in functie de intervalele de varsta.Apoi calculam distributia cu .value_counts cu normalize setat pe true si *100 pt procentaj.

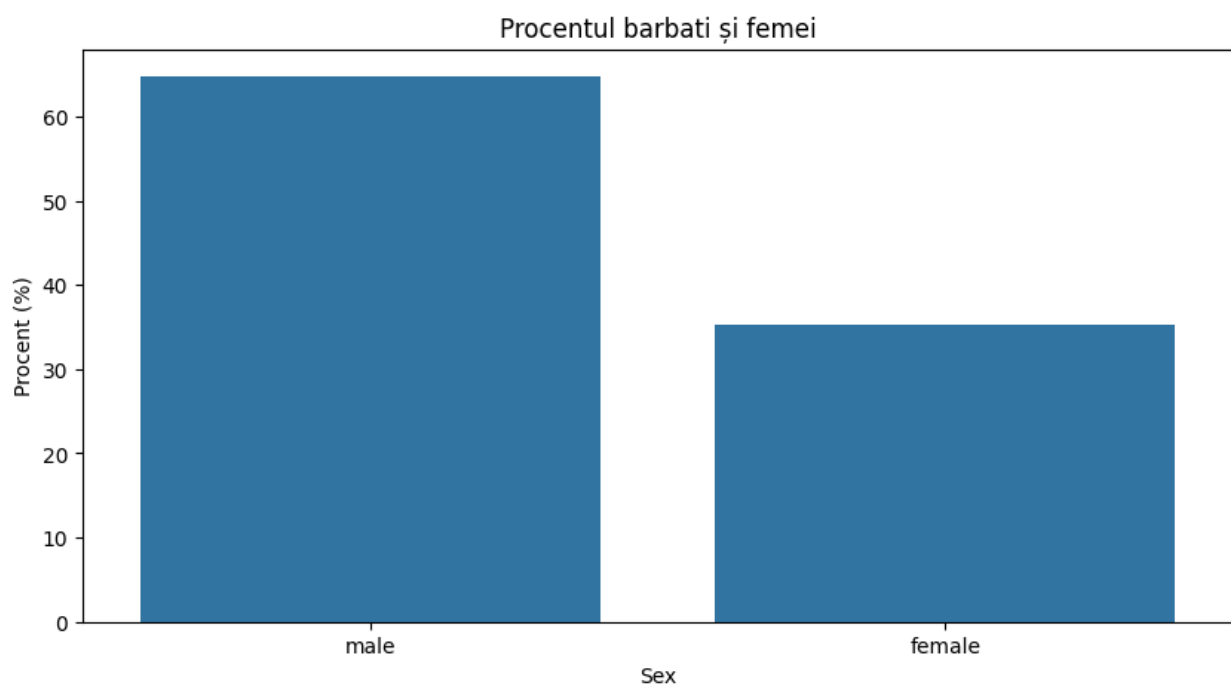
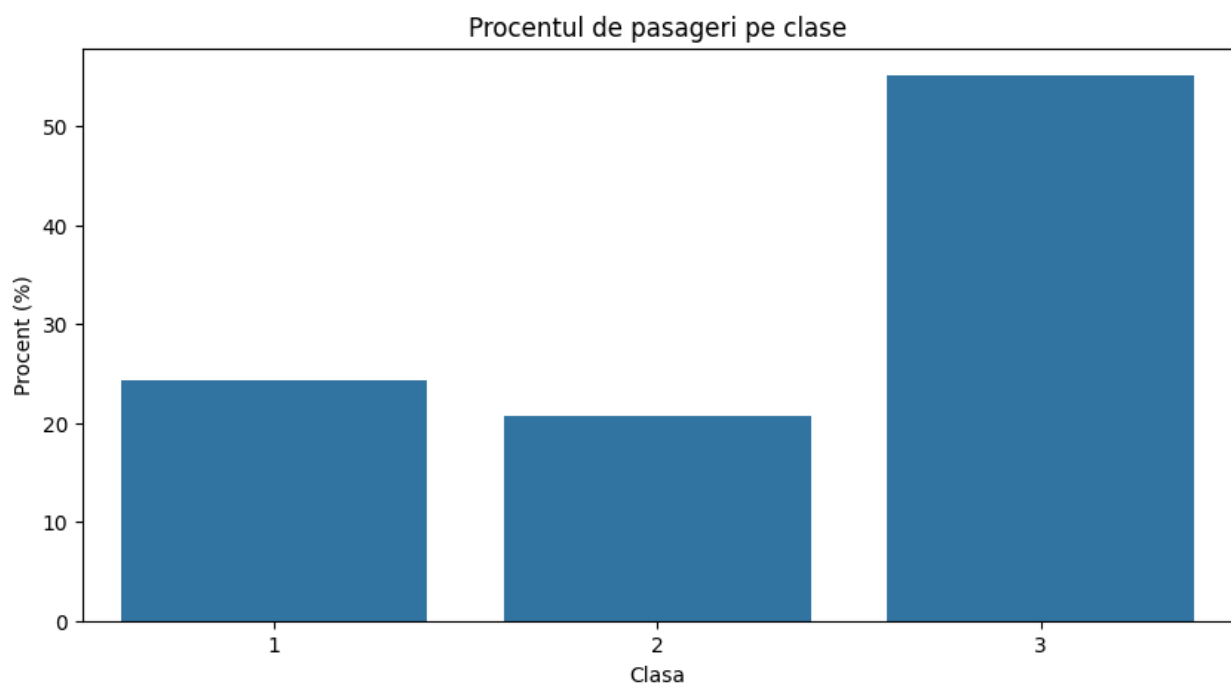
La cerinta 6 am preluat de la 5 datasetul ,iar apoi am adaugat conditia ca sexul sa fie male,dupa care am folosit groupby si sum pentru a vedea exact nr de barbati care au supravietuit din fiecare categorie de varsta,iar apoi am folosit metoda cu .mean()*100 pentru procentaje.

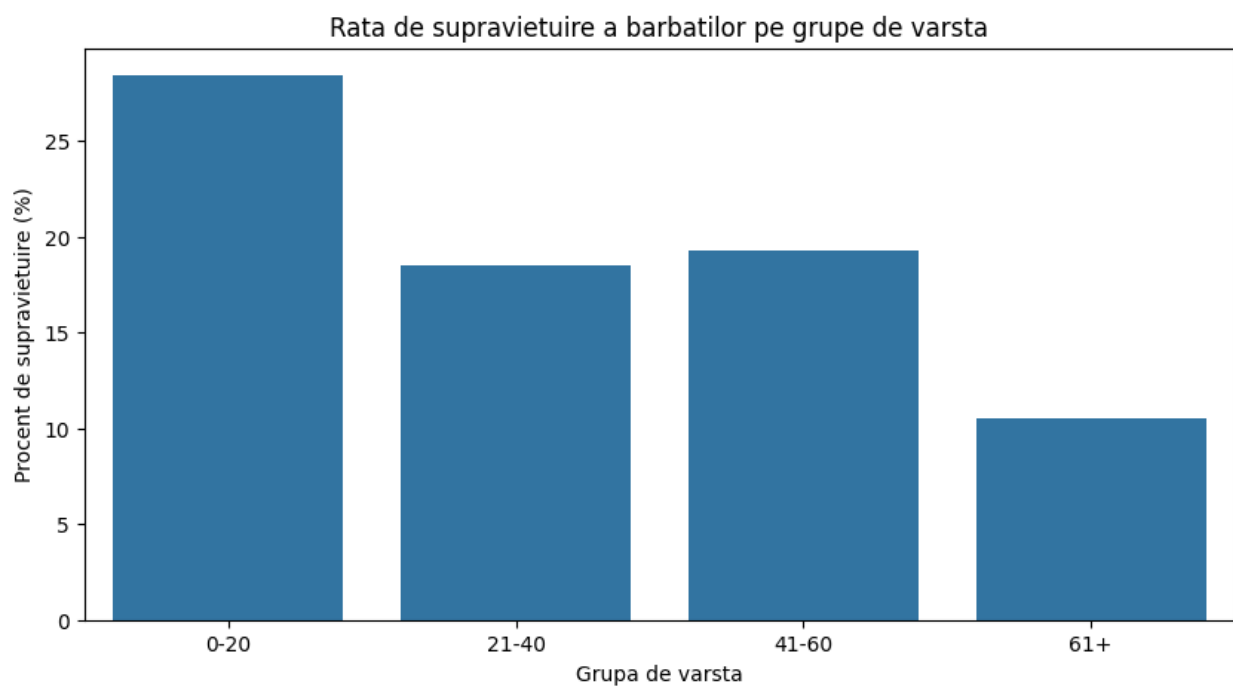
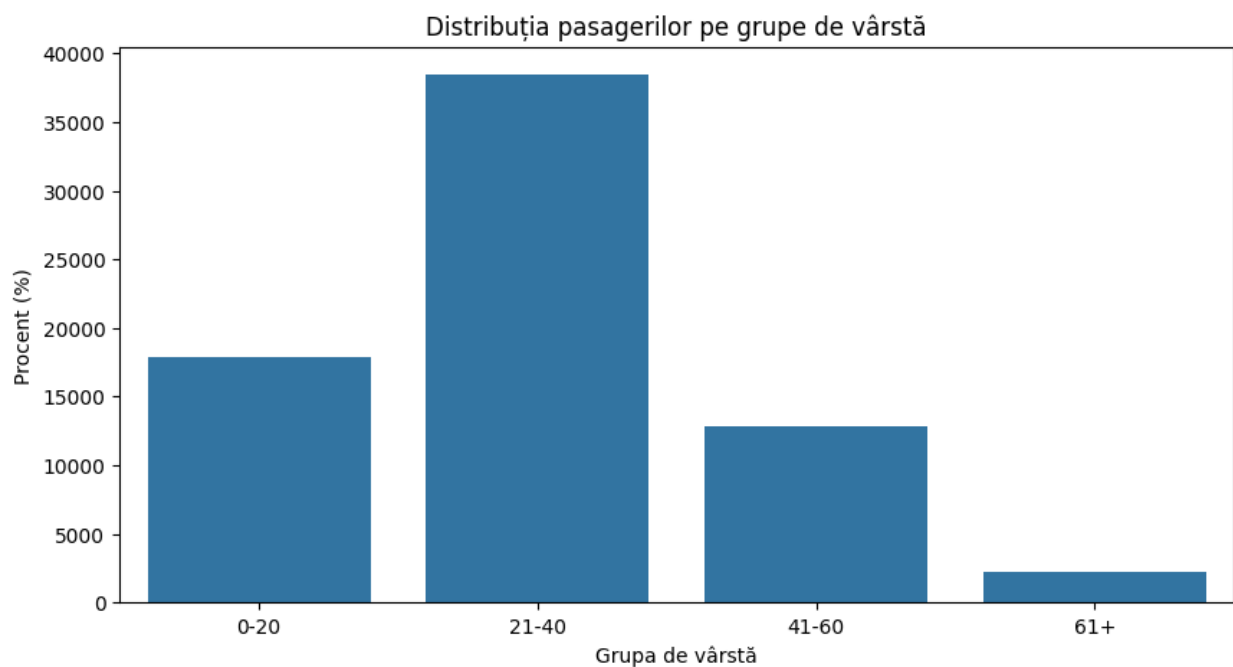
La cerinta 7 am aflat cati copii sunt prin creerea unei noi coloane care are toate elementele de pe coloanal age care sunt mici decat 18.Facem suma pentru valorile diferite de 0 si apoi procentajul copiilor.

In cerinta 9 am folosit regex ca sa extrag titlurile din nume,iar apoi am facut o tabela de corelatie pentru a vedea cum sunt titlurile in functie de sex.

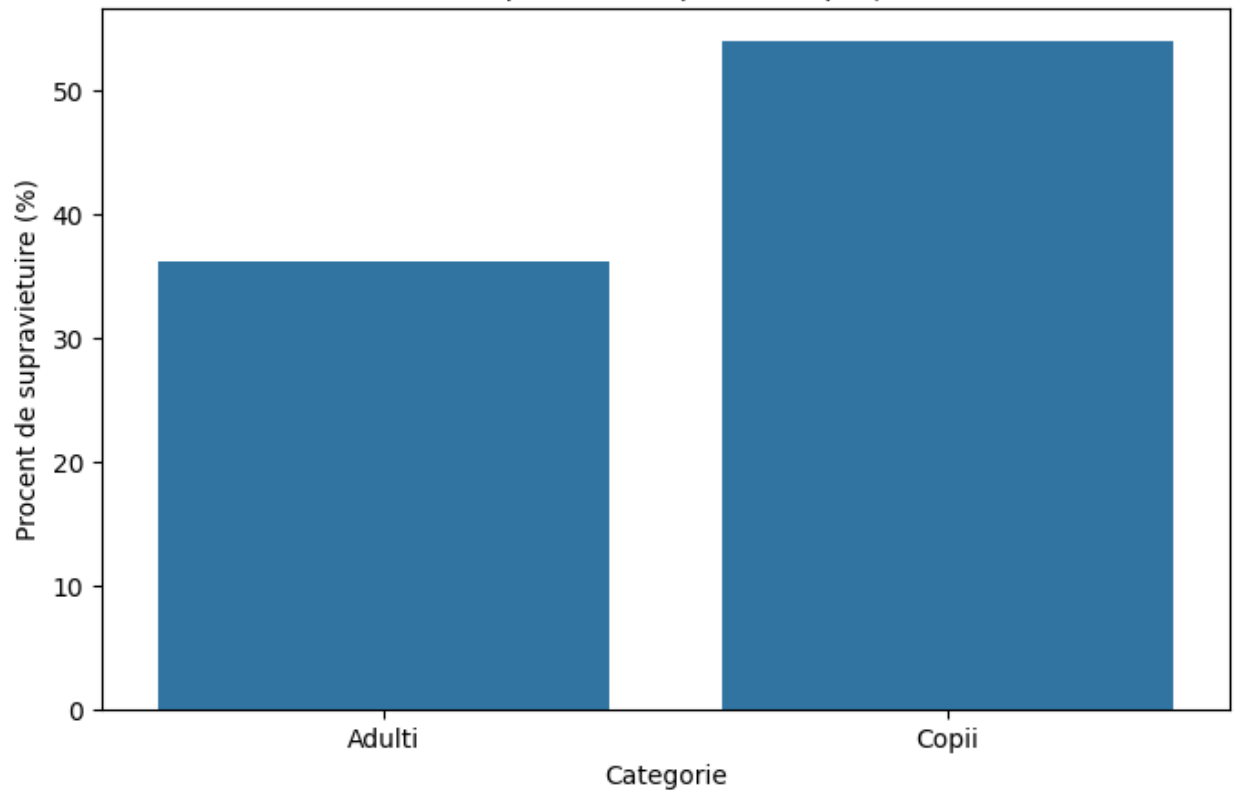
La cerinta 10 am facut o coloanal noua in care faceam suma coloanelor in care aveam inf despre familia sau relatiile pasagerilor.Am luat primele 100 de inregistrari si m am documentat cum sa fac diagrama respective.



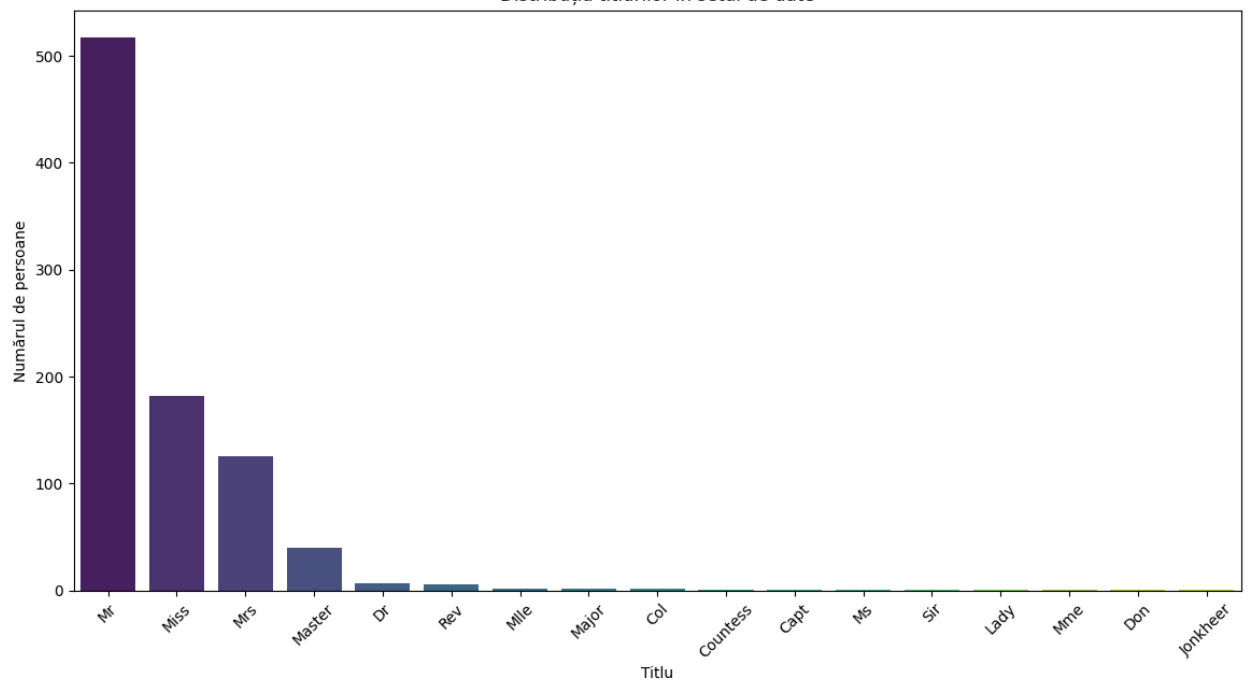




Rata de supraviețuire pentru copii și adulți



Distribuția titlurilor în setul de date



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

