

Proiect PCLP3

Exercitiul 1

Acest script incarca un dataset din fisierul “train.csv” si analizeaza structura acestuia. Determina si afiseaza numarul de coloane, tipurile de date din fiecare coloana, numarul de valori lipsa, numarul total de linii si numarul de linii duplicate.

Exercitiul 2

Acest script incarca un dataset dintr-un fisierul “train.csv” si calculeaza procentele de supravietuire, distributia pe clase si distributia pe gen pentru pasagerii Titanicului. Datele sunt vizualizate prin trei grafice: procentul de supravietuitori si non-supravietuitori, procentul pasagerilor pentru fiecare clasa si procentul barbatilor si femeilor.

Exercitiul 3

Acest script incarca un dataset dintr-un fisierul “train.csv” si genereaza histograme pentru fiecare coloana numerica din dataset. Histogramele sunt afisate intr-un singur grafic pentru o analiza vizuala facila a distributiilor valorilor numerice din dataset.

Exercitiul 4

Acest script incarca un dataset dintr-un fisierul “train.csv” si identifica coloanele cu valori lipsa, calculand numarul si proportia valorilor lipsa pentru fiecare coloana. De asemenea, calculeaza procentul valorilor lipsa pentru fiecare dintre cele doua clase din coloana "Survived" si afiseaza rezultatele.

Exercitiul 5

Acest script incarca un dataset dintr-un fisierul “train.csv” si creeaza categorii de varsta pentru pasageri. Genereaza un grafic pentru a evidentia numarul de pasageri pentru fiecare categorie de varsta si afiseaza aceste rezultate in consola.

Exercitiul 6

Acest script incarca un dataset dintr-un fisierul “train.csv” si creeaza categorii de varsta pentru pasageri. Calculeaza si vizualizeaza procentul de barbati care au supravietuit in fiecare categorie de varsta, afisand rezultatele intr-un grafic de tip bara.

Exercitiul 7

Acest script incarca un dataset dintr-un fisierul “train.csv” si clasifica pasagerii in copii si adulti pe baza varstei (< 18 ani pentru copii). Calculeaza si vizualizeaza procentul de supravietuitori pentru fiecare grupa de varsta (copii si adulti) intr-un grafic de tip bara.

Exercitiul 8

Acest script incarca un dataset dintr-un fisierul “train.csv” si completeaza valorile lipsa. Pentru coloana Age, valorile lipsa sunt completate cu media varstei pasagerilor din aceeasi clasa (Pclass) si cu aceeasi stare de supravietuire (Survived). Pentru coloanele categorice, valorile lipsa sunt completate cu cea mai frecventa valoare (modul). Output-ul este salvat intr-un nou fisier.

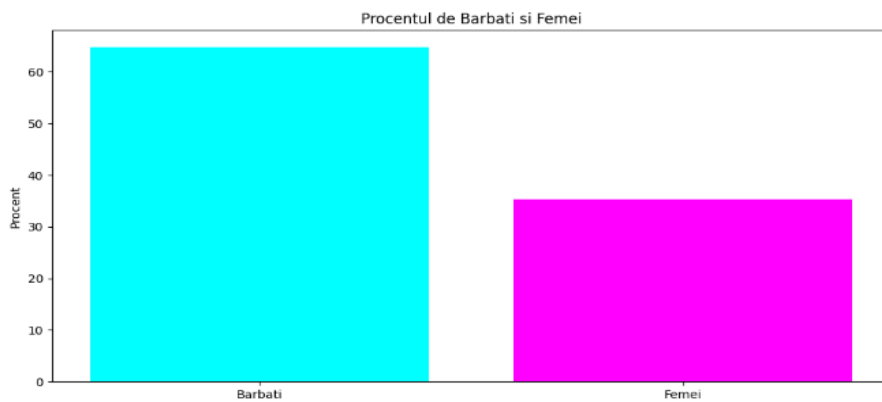
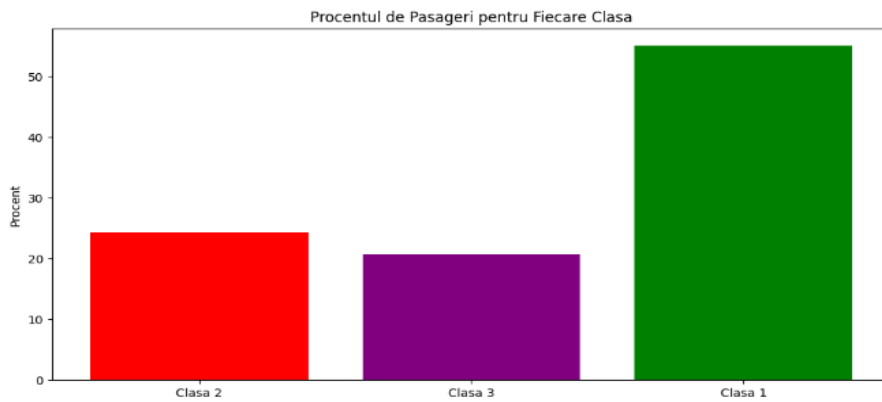
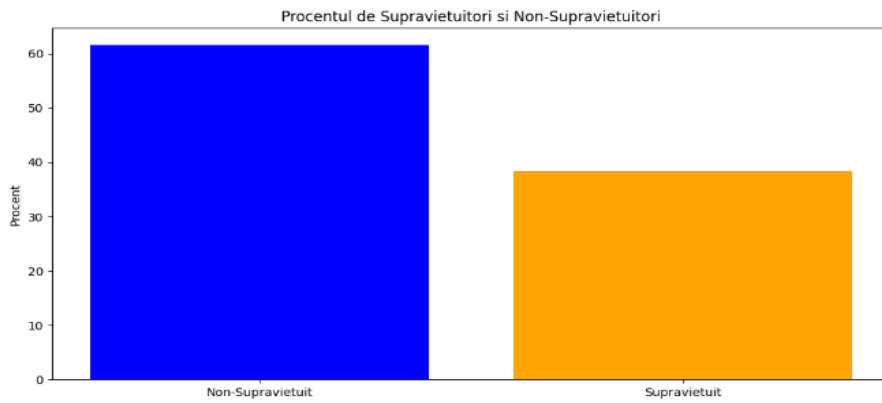
Exercitiul 9

Acest script incarca un dataset dintr-un fisierul “train.csv” si extrage titlurile din coloana Name. Verifica daca titlurile de noblete corespund cu sexul persoanei si vizualizeaza distributia titlurilor in functie de sex intr-un grafic de tip bara suprapus.

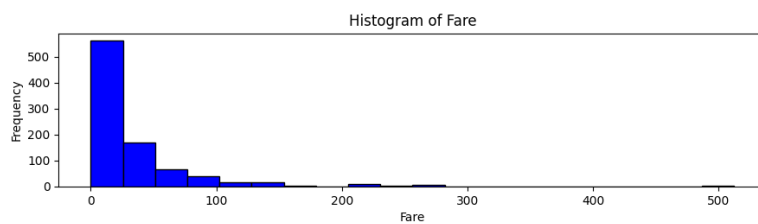
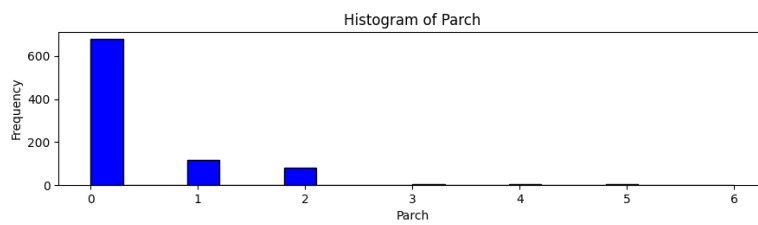
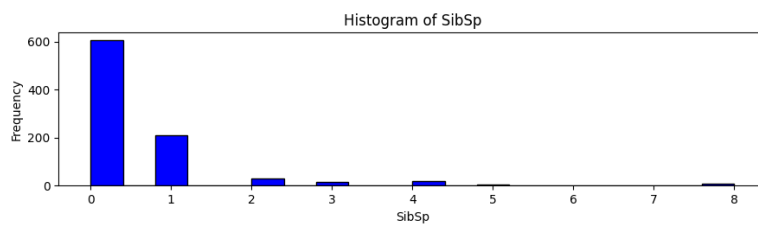
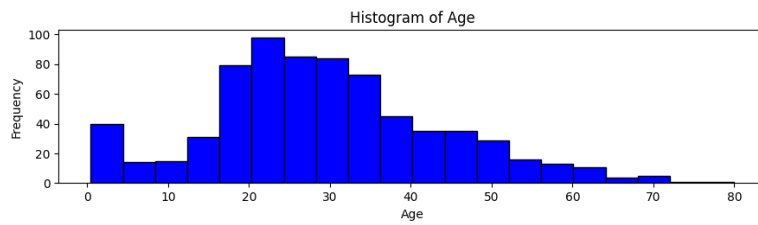
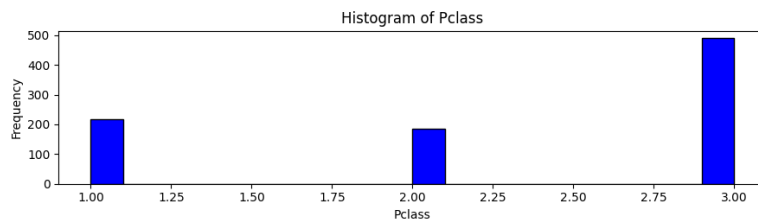
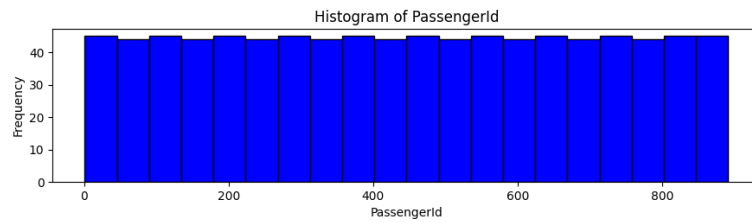
Exercitiul 10

Acest script incarca un dataset dintr-un fisierul “train.csv” si adauga o coloana pentru a identifica pasagerii singuri (fara rude). Selecteaza primele 100 de inregistrari si genereaza un grafic de tip catplot pentru a investiga relatia dintre tarif, clasa si supravietuire, segmentat in functie de pasagerii singuri si cei cu rude.

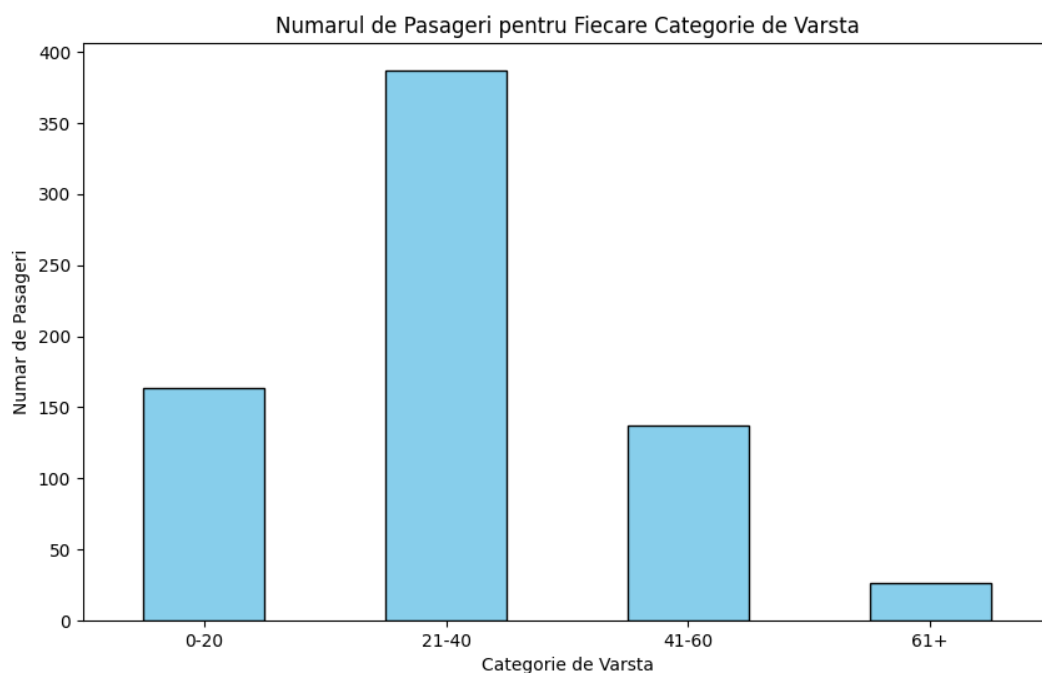
Exercitiul 2 grafic



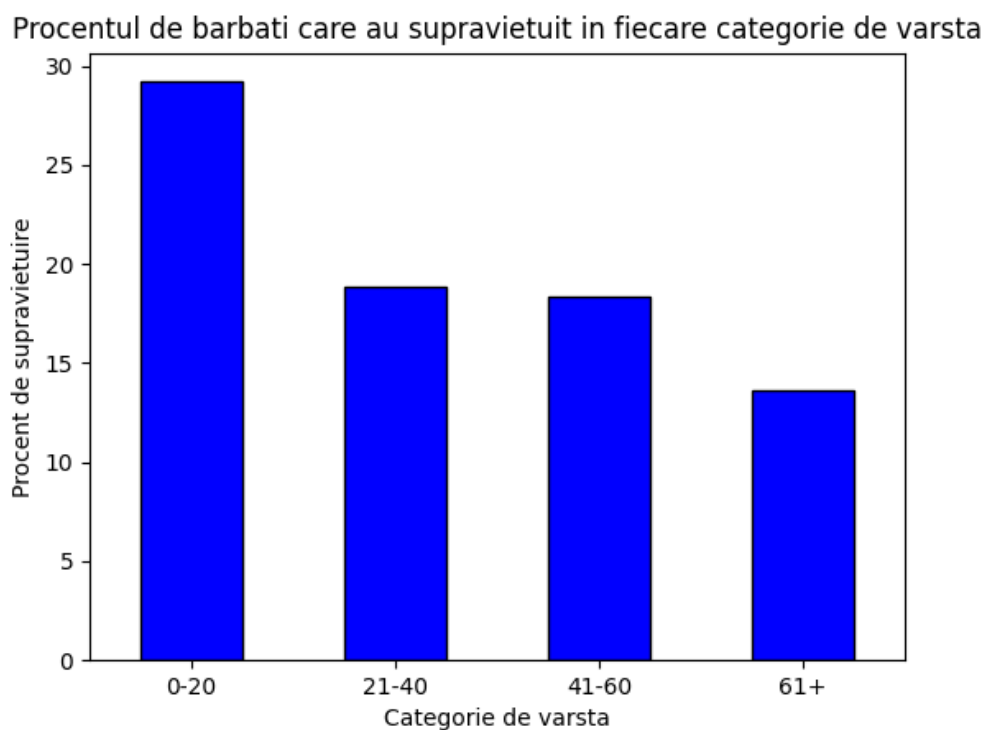
Exercitiul 3 grafic



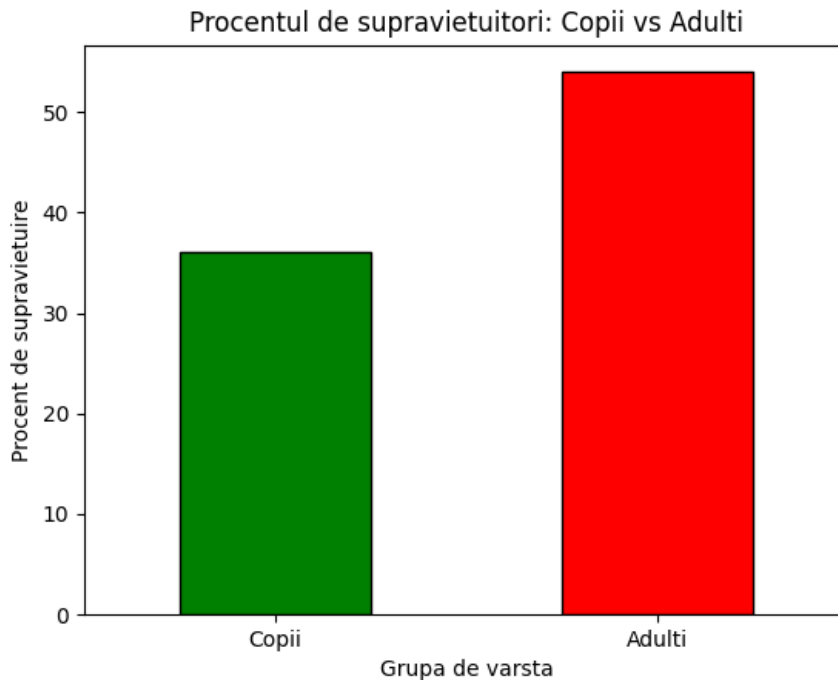
Exercitiul 5 grafic



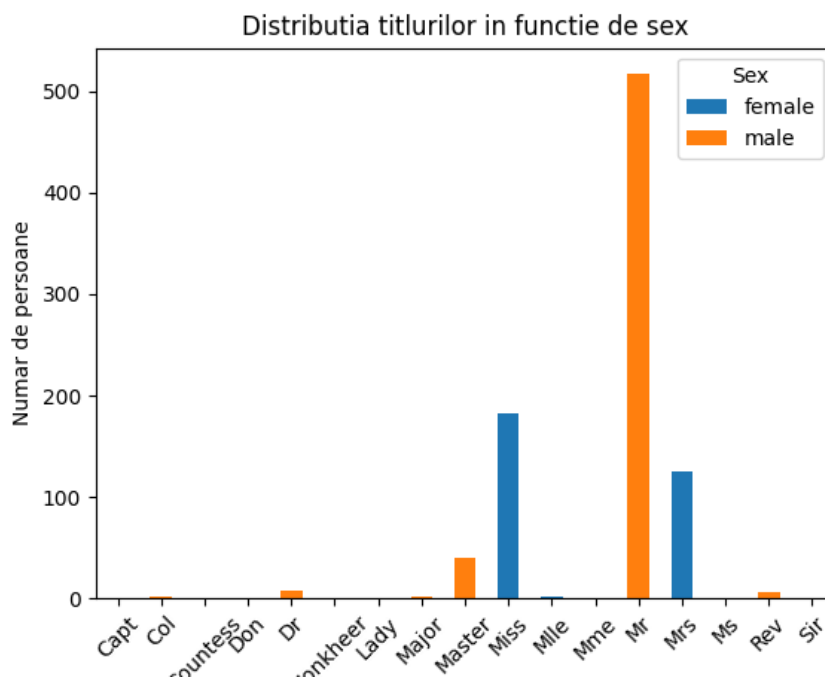
Exercitiul 6 grafic



Exercitiul 7 grafic



Exercitiul 9 grafic



Exercitiul 10 grafic

