

Documentatie

Student: Ciupitu Bogdan-Costin 312CD

Partea I

Am declarat la inceput toate bibliotecile specializate ce va urma sa le folosesc in fisierul ipynb. In variabila "data" am citit informatiile din setul de date din fisierul "train.csv". Dupa toate modificarile facute aduse le-am salvat in fisierul "train_modified.csv".

Cerinta 1

Output:

Numarul de coloane: 15

Tipul datelor: PassengerId int64
Survived int64
Pclass int64
Name object
Sex object
Age float64
SibSp int64
Parch int64
Ticket object
Fare float64
Cabin object
Embarked object
AgeCategory category
Title_of_nobility object
Alone bool
dtype: object

Numarul de valori lipsa pentru fiecare coloane: PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 0
Embarked 0
AgeCategory 177

Title_of_nobility 0
Alone 0
dtype: int64

Numarul de linii: 891

Nu exista linii duplicate

Cerinta 2

Output:

Numarul de persoane supravietuite: 342

Numarul de persoane decedate: 549

Numarul total de persoane: 891

Procentul de persoane supravietuite: 38.38383838383838%

Procentul de persoane decedate: 61.61616161616162%

Procentul de persoane din fiecare clasa: Pclass

3 55.106622

1 24.242424

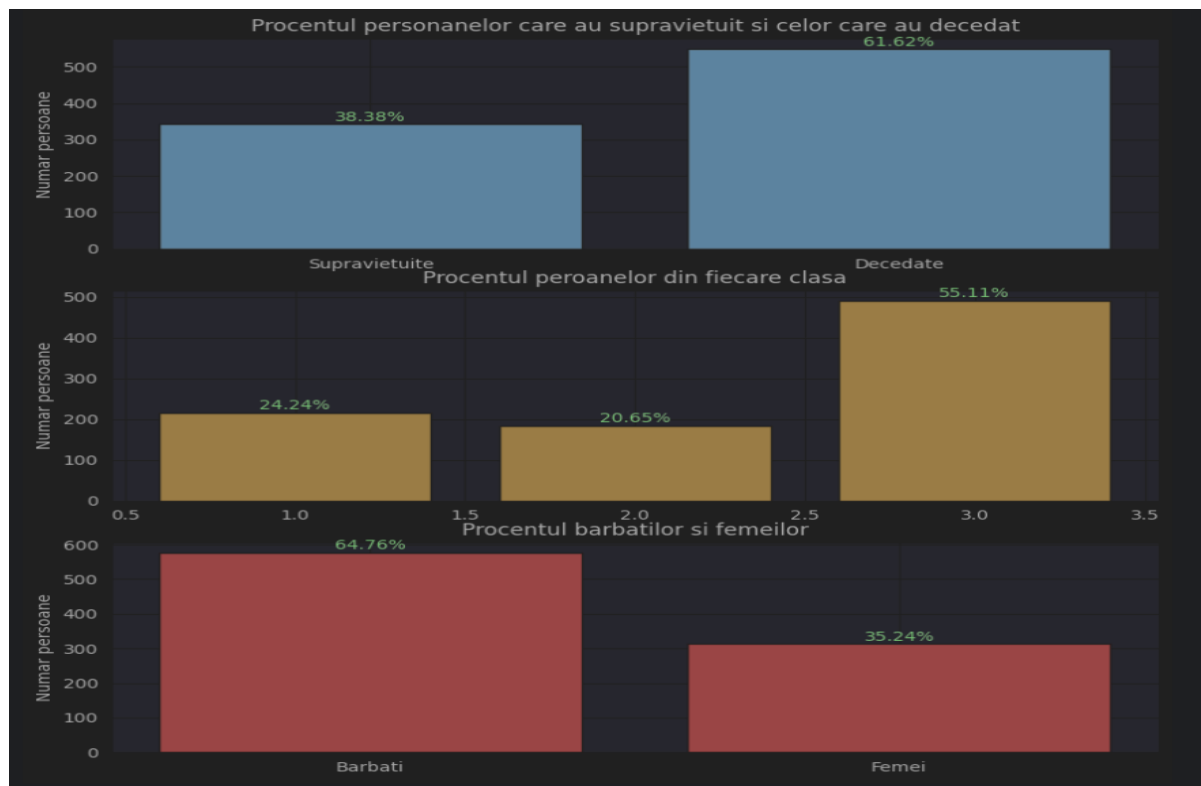
2 20.650954

Name: count, dtype: float64

Procentul de barbati: 64.75869809203144%

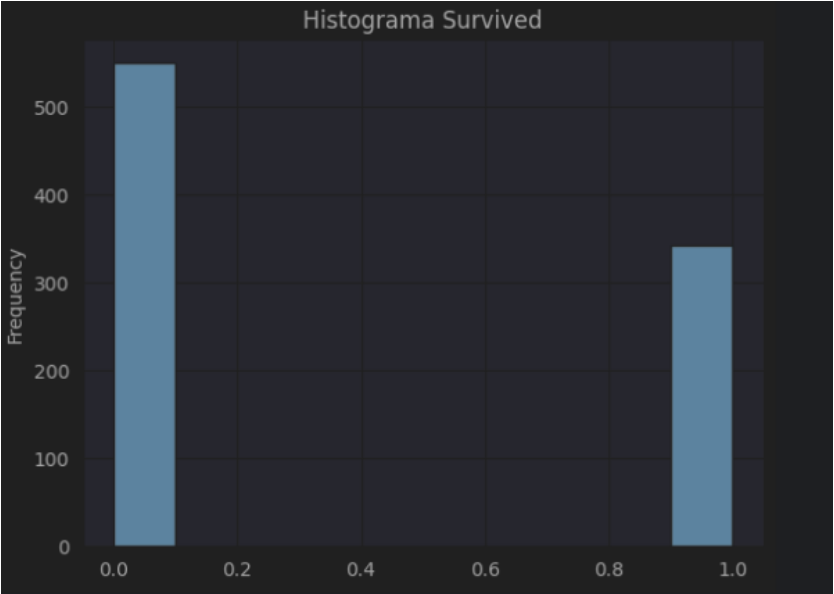
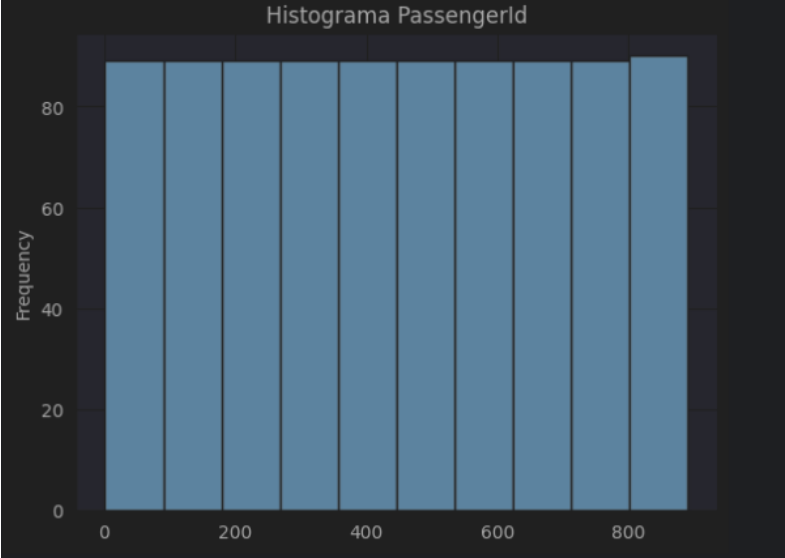
Procentul de femei: 35.24130190796858%

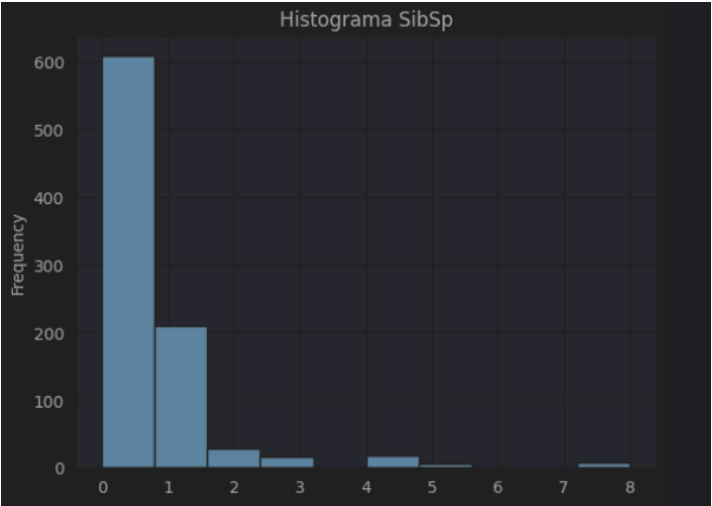
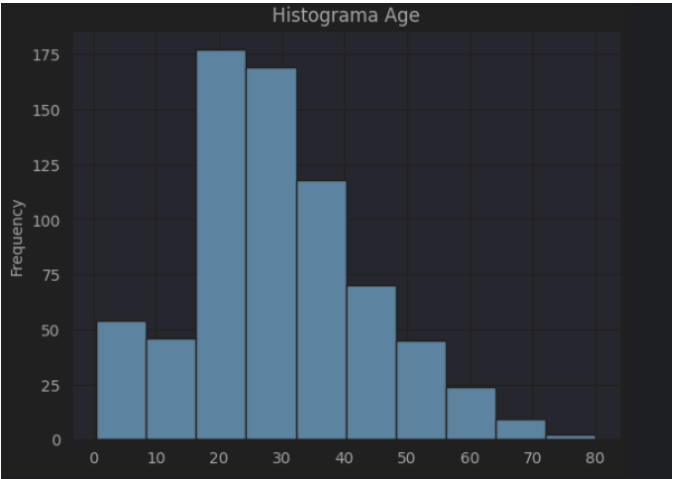
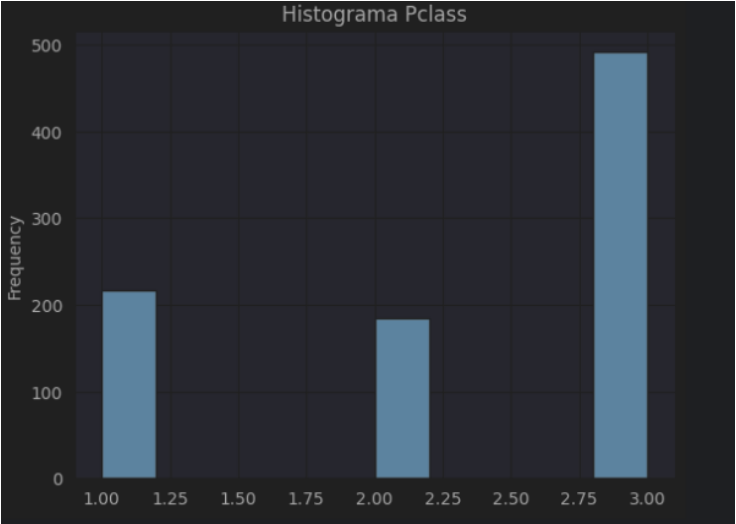
Graficele:



Cerinta 3

Histogramele:





Cerinta 4

Output:

Coloana Age are 177 valori lipsa si procentul de 19.865319865319865%

Coloana Cabin are 687 valori lipsa si procentul de 77.10437710437711%

Coloana Embarked are 2 valori lipsa si procentul de 0.22446689113355783%

Clasa 0:

Coloana Age are 125 valori lipsa si procentul de 22.768670309653917%

Coloana Cabin are 481 valori lipsa si procentul de 87.61384335154827%

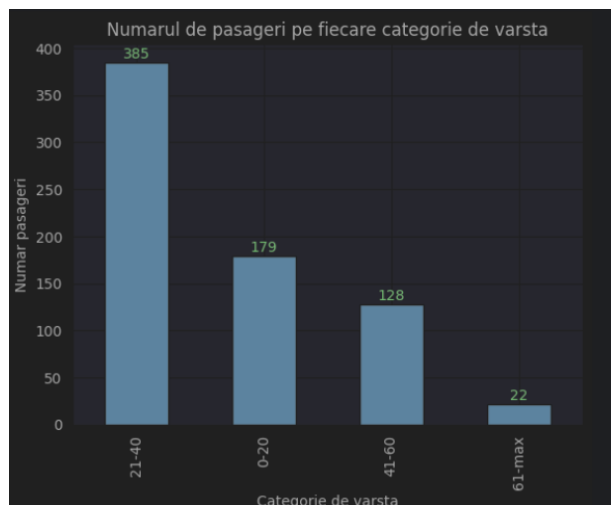
Clasa 1:

Coloana Age are 52 valori lipsa si procentul de 15.204678362573098%

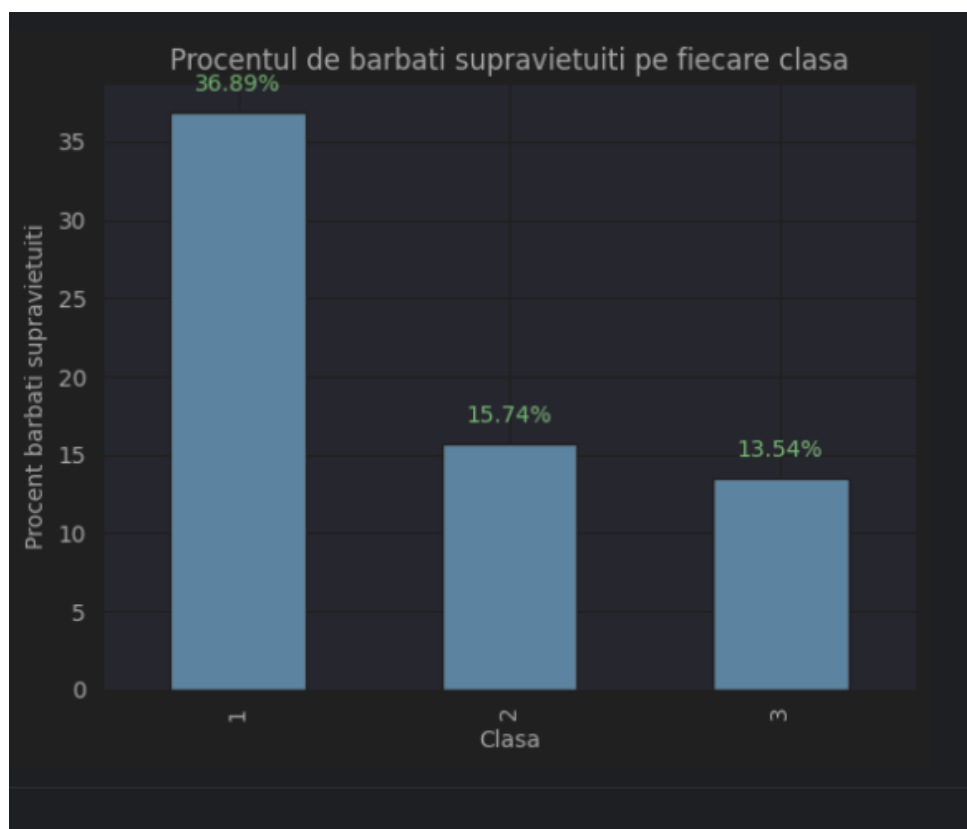
Coloana Cabin are 206 valori lipsa si procentul de 60.23391812865497%

Coloana Embarked are 2 valori lipsa si procentul de 0.5847953216374269%

Cerinta 5



Cerinta 6

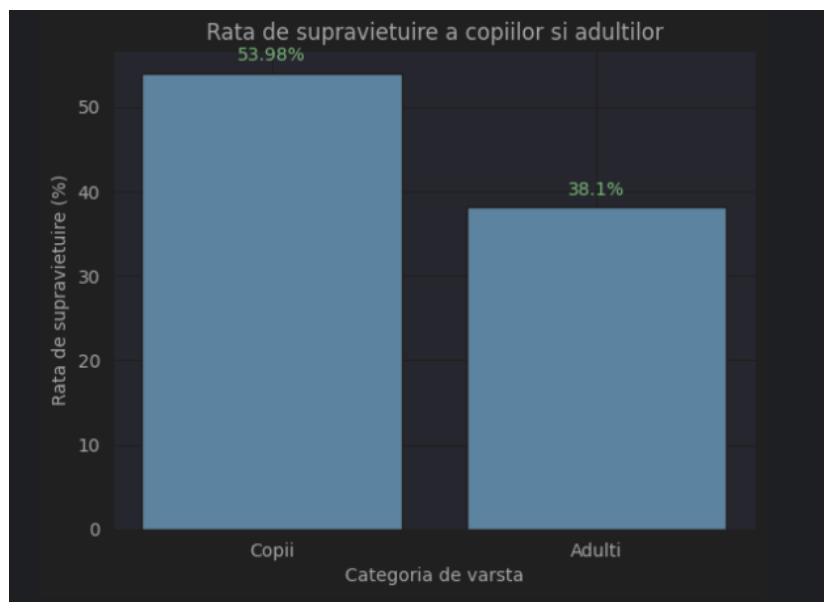


Cerinta 7

Output:

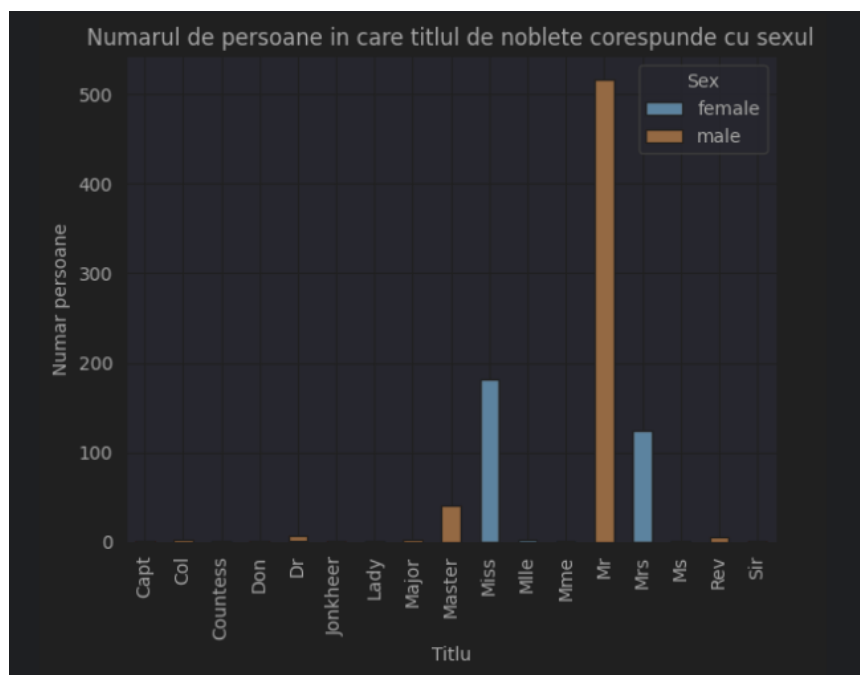
Procentul de copii aflati la bord: 12.682379349046016%

Graficul:

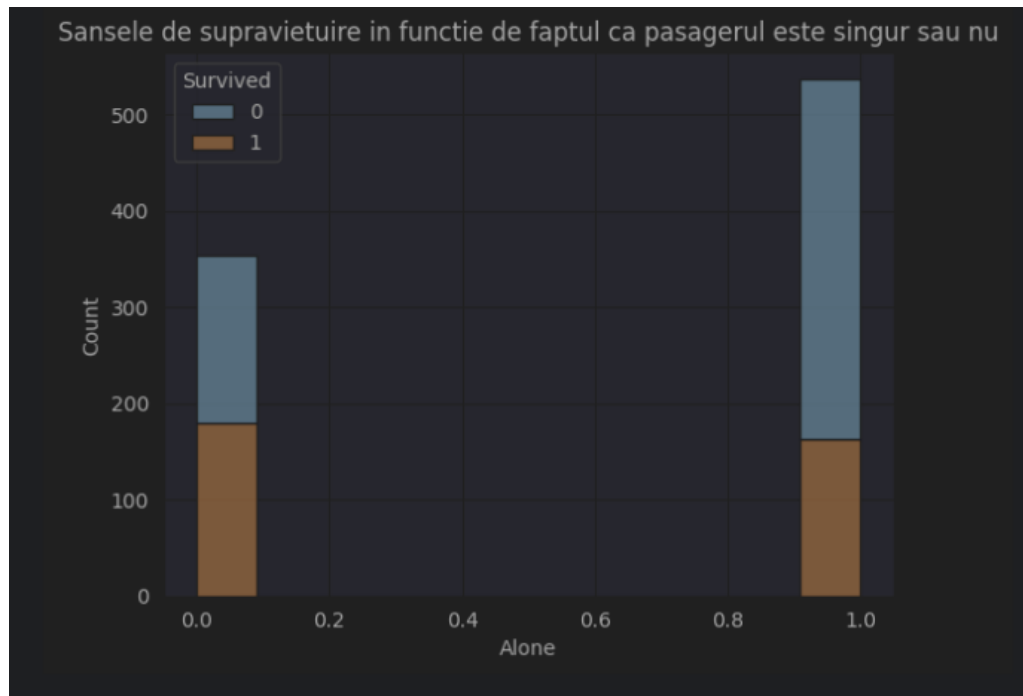


Cerinta 9

Graficul:



Cerinta 10



Relatia dintre tarif, clasa si starea de supravietuire pentru primele 100 de inregistrari

