

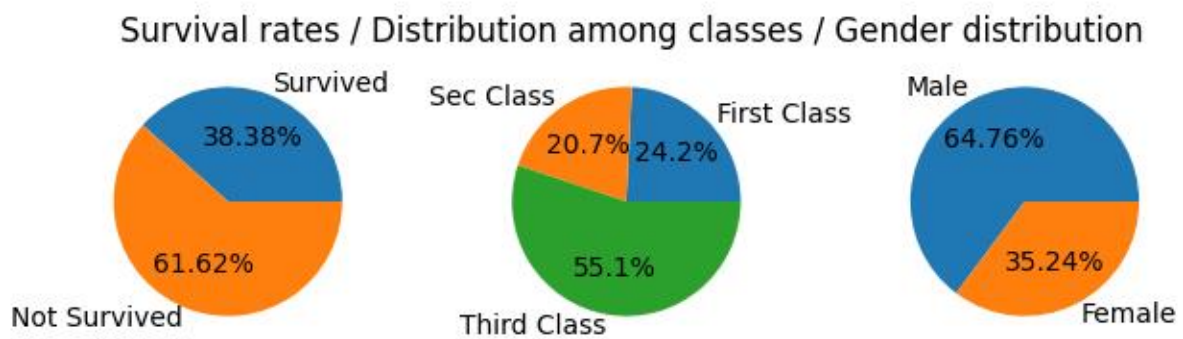
# Proiect Titanic

Detaliile setului de date:

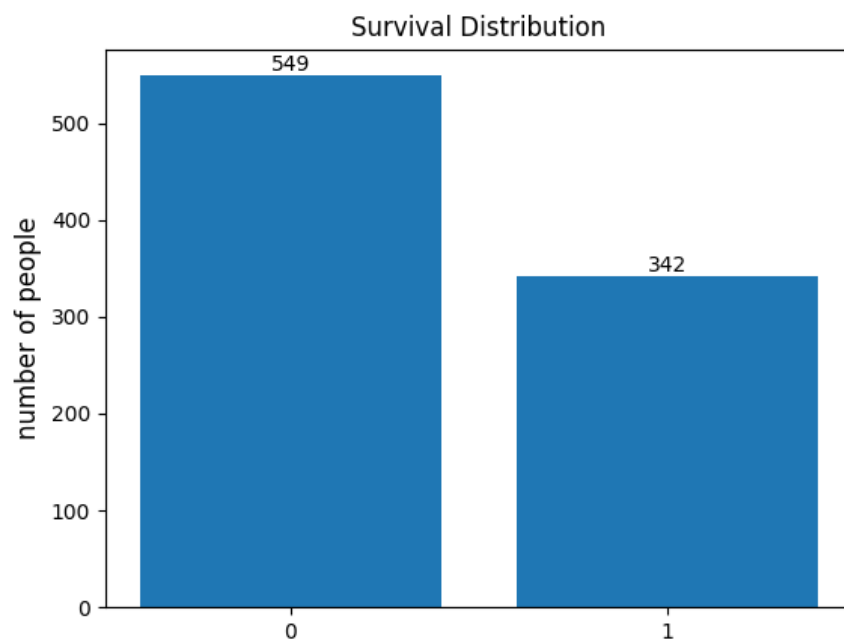
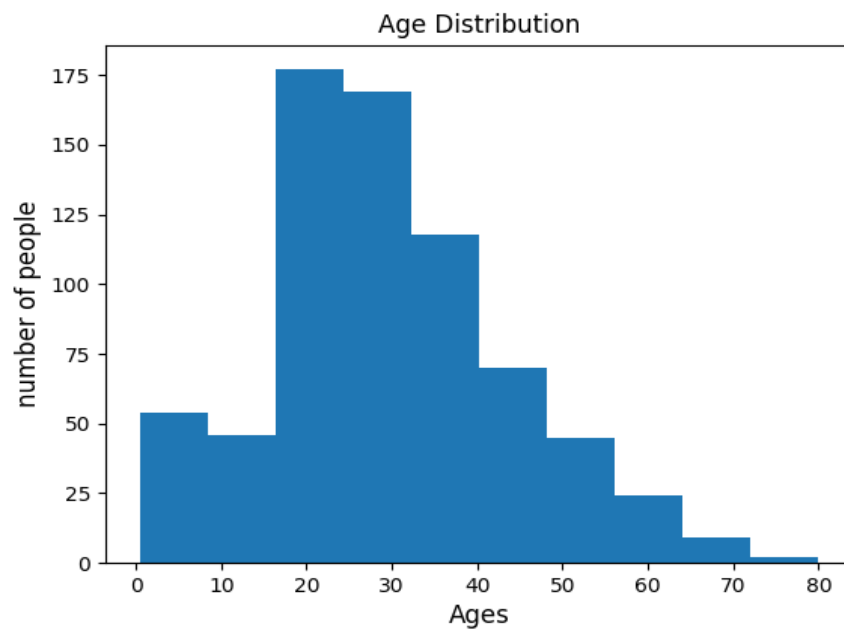
- 12 coloane
- Type pt fiecare coloanal:
  - PassengerId int64
  - Survived int64
  - Pclass int64
  - Name object
  - Sex object
  - Age float64
  - SibSp int64
  - Parch int64
  - Ticket object
  - Fare float64
  - Cabin object
  - Embarked object
- Nr date lipsa pt fiecare coloanal
  - PassengerId 0
  - Survived 0
  - Pclass 0
  - Name 0
  - Sex 0
  - Age 177
  - SibSp 0
  - Parch 0
  - Ticket 0
  - Fare 0
  - Cabin 687
  - Embarked 2
- Nr total linii = 891
- Nu exista linii duplicate

## CERINTA 2:

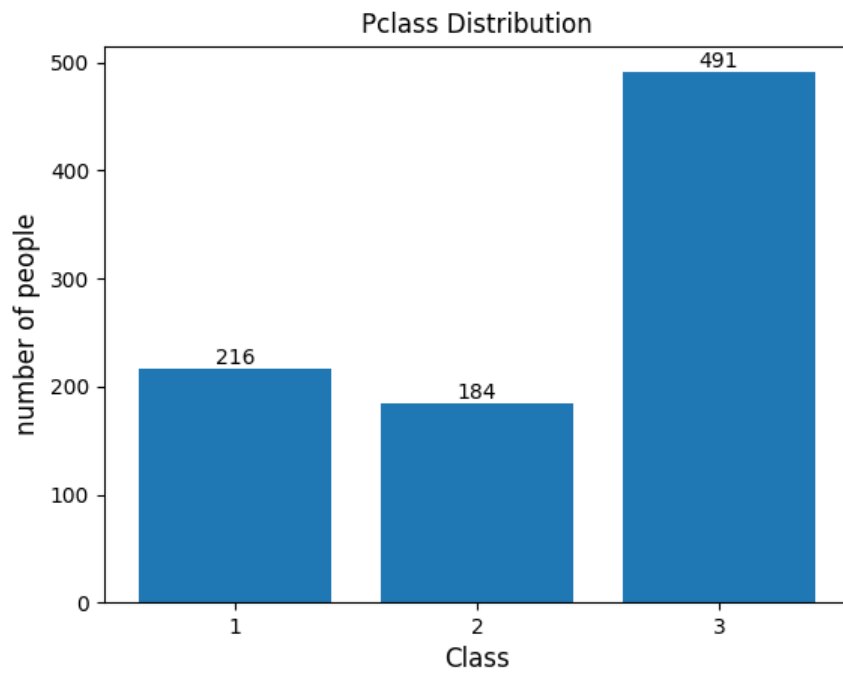
- procent pers ce au supravietuit = 38.38 %
- procent pers ce au NU supravietuit = 61.62 %
- procent pers clasa 1 = 24.24 %
- procent pers clasa 2 = 20.65 %
- procent pers clasa 3 = 55.11 %
- procent barbati = 64.76 %
- procent femei = 35.24 %



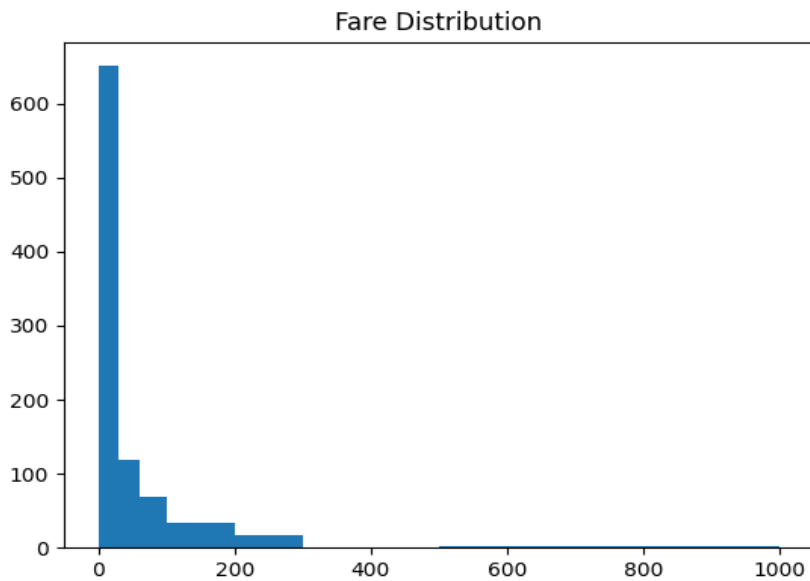
### CERINTA 3:



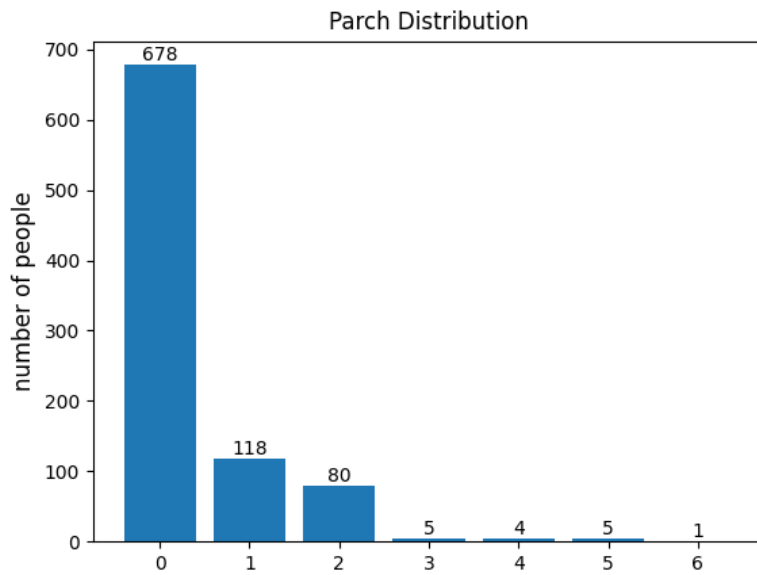
Observam ca majoritatea biletelor au fost cumparate pt clasa a 3 a



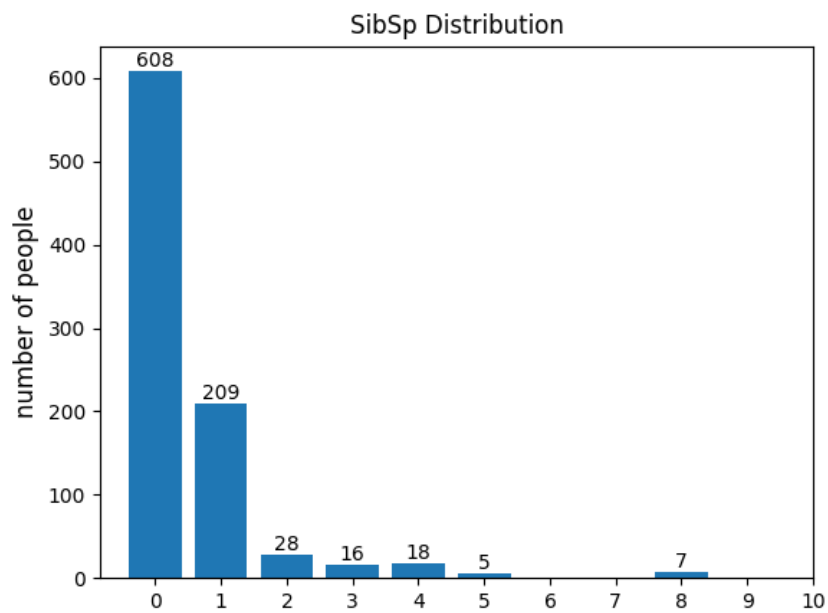
Observam mai jos ca majoritatea biletelor au costat sub 200 in timp ce doar un nr foarte mic au costat de la 500 incolo



Observam ca majoritatea pasagerilor nu aveau nici parinti si nici copii pe vas



Observam ca majoritatea pasagerilor nu aveau nici o ruda pe vas



#### CERINTA 4:

- 19.87 % din datele din col Age lipsesc
  - 77.10 % din datele din col Cabin lipsesc
  - 0.22 % din datele din col Embarked lipsesc
- 
- 5.84 % din datele din col Age lipsesc in cazul pasagerilor ce au supravietuit
  - 23.12 % din datele din col Cabin lipsesc in cazul pasagerilor ce au supravietuit
  - 0.22 % din datele din col Embarked lipsesc in cazul pasagerilor ce au supravietuit
- 
- 14.03 % din datele din col Age lipsesc in cazul pasagerilor ce NU au supravietuit
  - 53.98 % din datele din col Cabin lipsesc in cazul pasagerilor ce NU au supravietuit

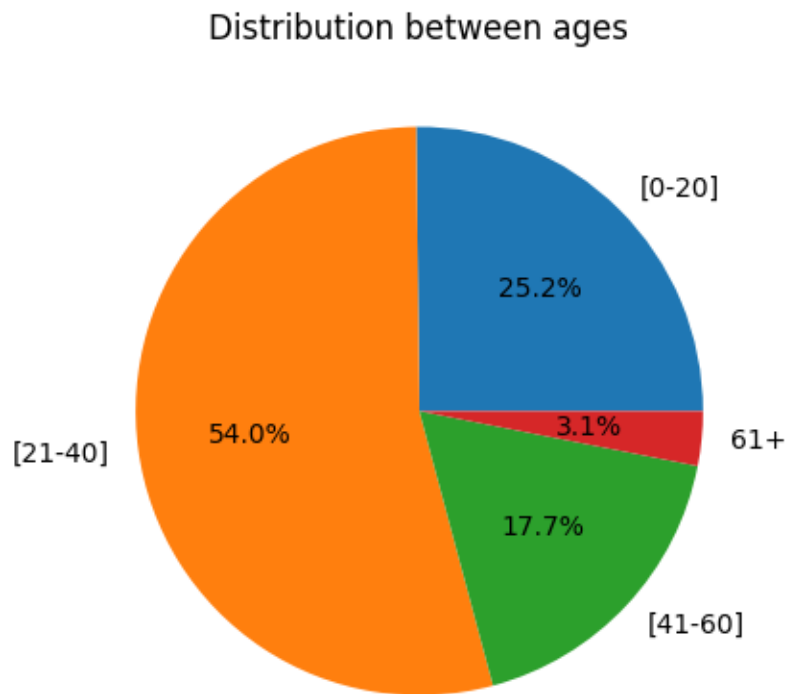
Din aceste procente ne dam seama ca in cazul pasagerilor ce nu au supravietuit lipsesc mai multe date (a fost mai grea recuperarea lor pt ca desigur pasagerul nu mai era in viata si majoritatea pasagerilor nu au avut nici o ruda pe vas, baramite una ce a supravietuit)

## CERINTA 5:

Ne scoatem din data set randurile ce nu contin varsta persoanei

Observam ca erau foarte putini pasageri cu varsta mai mare de 60 ani

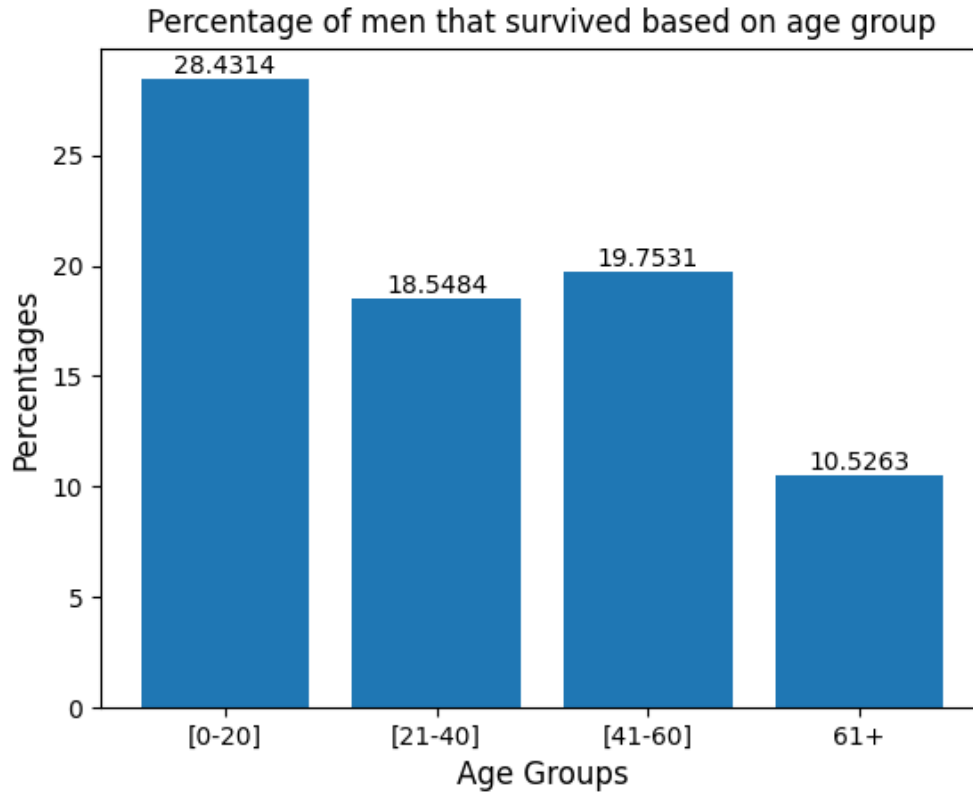
Majoritatea pasagerilor erau intre 21 si 40 de ani lucru ce se potriveste cu faptul ca lumea pleca peste Atlantic sa lucreze



## CERINTA 6:

- au supraviețuit 93 barbati
- din categ 0 - 20 au supraviețuit 29 barbati
- din categ 21 - 40 au supraviețuit 46 barbati
- din categ 41 - 60 au supraviețuit 16 barbati
- din categ 61 + au supraviețuit 2 barbati

Observam ca barbatii tineri au avut prioritate pt barcile de salvare in timp ce cei mai batrani nu

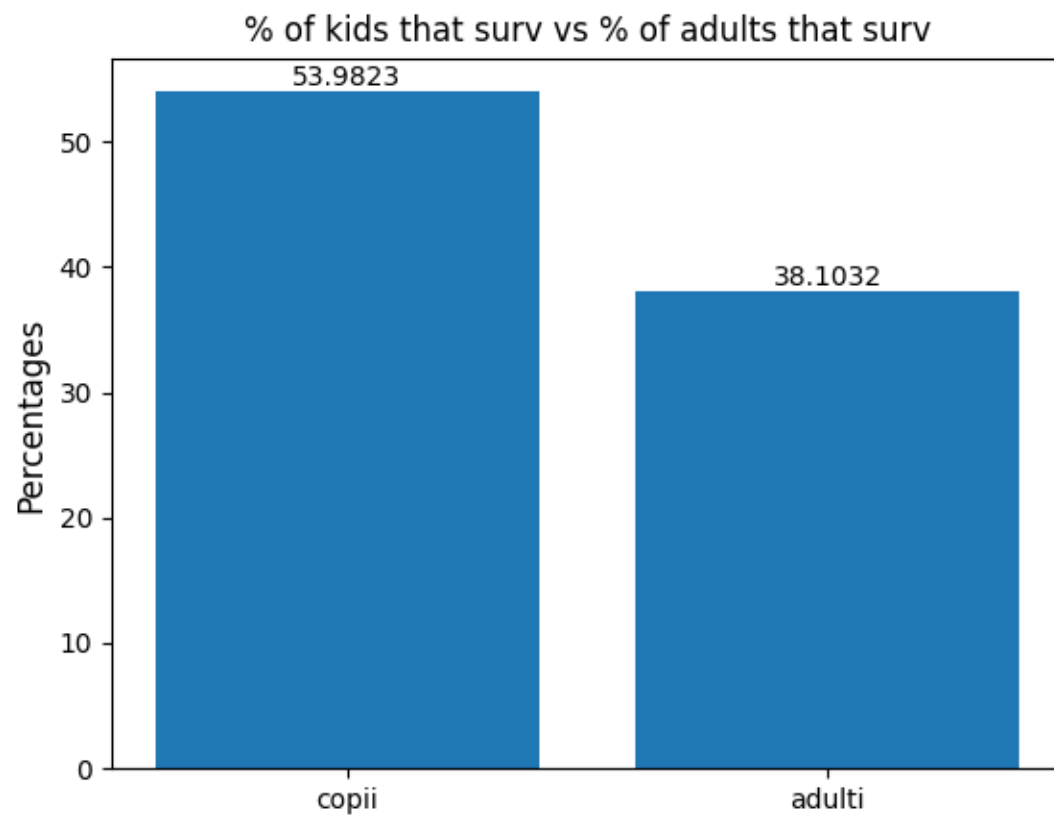




## CERINTA 7:

- copii la bord = 113
- nr copii suprav = 61

Observam ca copiii au avut prioritate mai mare fata de adulti



## CERINTA 8:

Coloana Age are valori lipsa:

Calculam cei ce au supraviețuit și cei ce nu au supraviețuit media de vârstă și o înlocuim pe rânduri corespunzător

Coloana Embarked are valori lipsa:

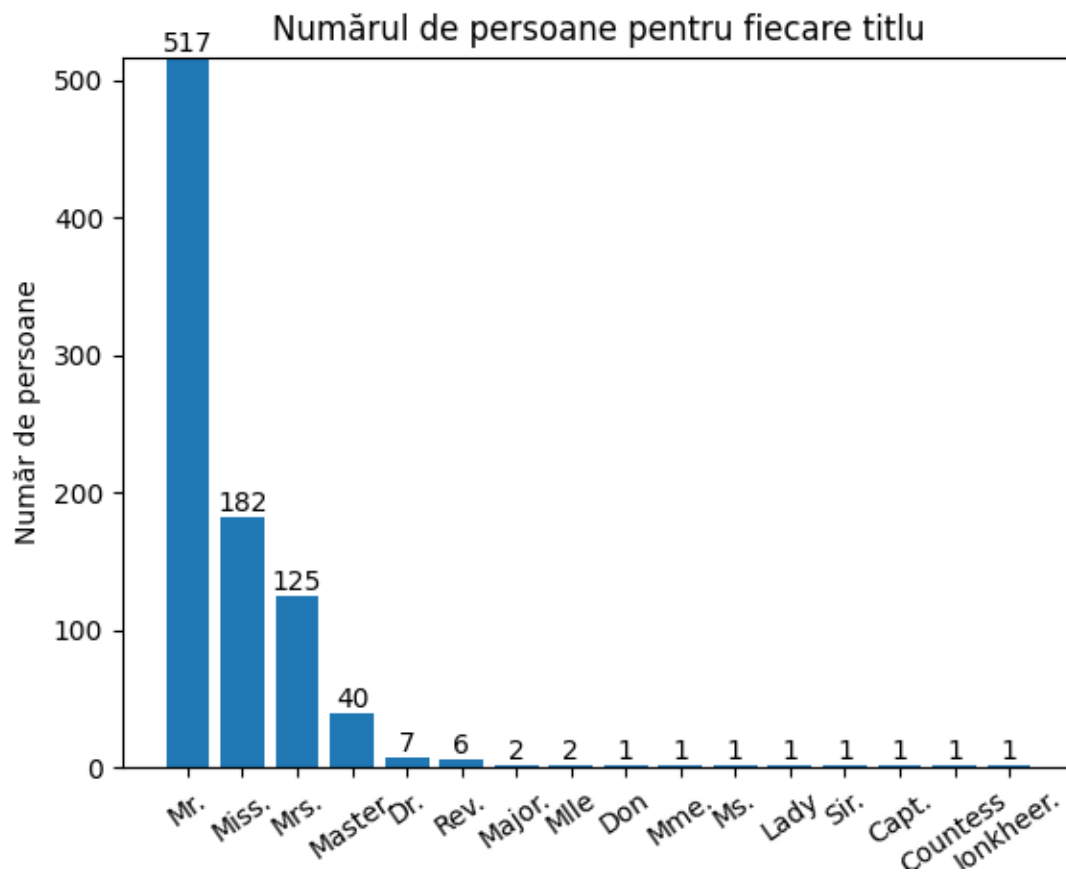
Vedem care e valoarea predominantă pt embarked (are valori discrete nu continue) și înlocuim corespunzător pt cei ce au supraviețuit și cei ce nu

Coloana Cabin are valori lipsa dar nu o putem completa din moment

ce toate cabinele sunt diferite deci nu există o valoare predominantă sau o medie

## CERINTA 9:

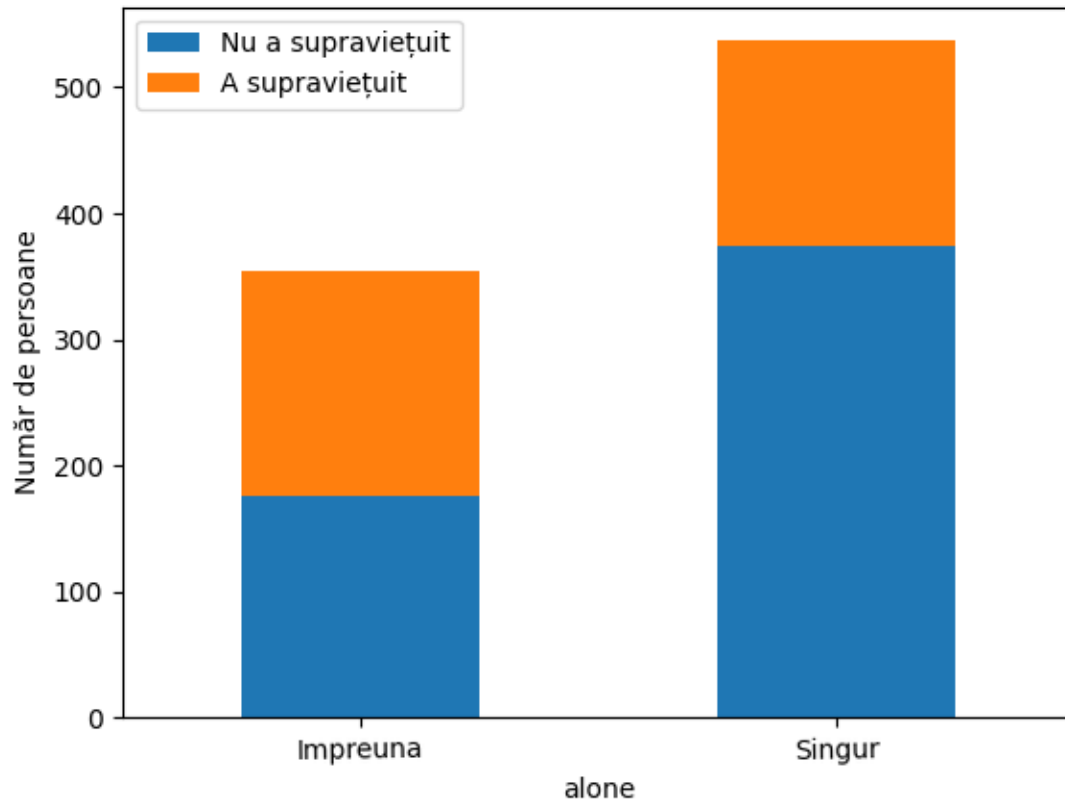
Inca o data observam ca majoritatea pasagerilor erau barbati



Cerinta 10:

Observam mai jos ca pasagerii de mai aveau rude pe vas au supravietuit  
Intr un procent mai mare fata de pasagerii ce nu aveau nici o ruda la  
bord

Influența stării de a fi singur pe Titanic asupra șanselor de supraviețuire



Observam ca majoritatea pasagerilor de la clasa a 2 a nu au supravietuit

De asemenea, tariful platit nu pare a fii un indicator al sansei de supravietuire

