

Proiect PCLP3

Partea I Aceasta ne-a propus să rezolvăm cerințe pentru a ne putea obișnui cu diferitele instrumente de manipulare a informației.

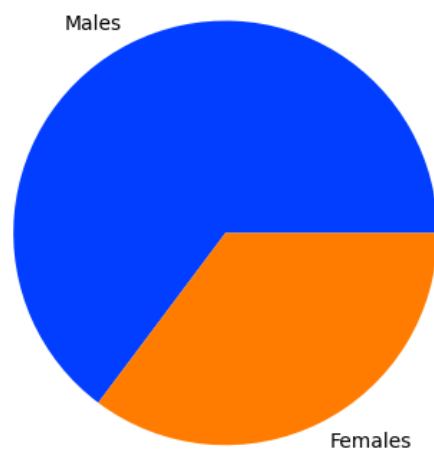
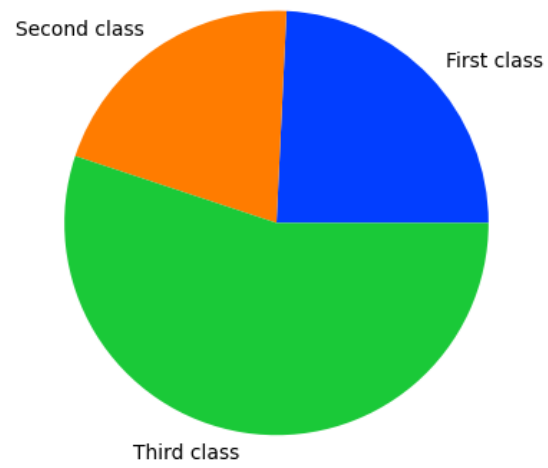
Cerința I

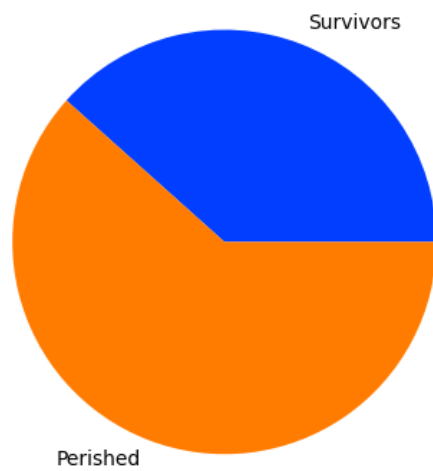
Am folosit label-urile oferite de pandas pentru a afișa informațiile cerute (ex. columns), cât și metoda duplicated.

- The number of columns is: 11
- The number of rows is: 891
- The data types of the columns are: Survived int64 Pclass int64 Name object Sex object Age float64 SibSp int64 Parch int64 Ticket object Fare float64 Cabin object Embarked object
- The number of null values in each column is: Survived 0 Pclass 0 Name 0 Sex 0 Age 177 SibSp 0 Parch 0 Ticket 0 Fare 0 Cabin 687 Embarked 2
- The number of duplicated rows is: 0

Cerința II

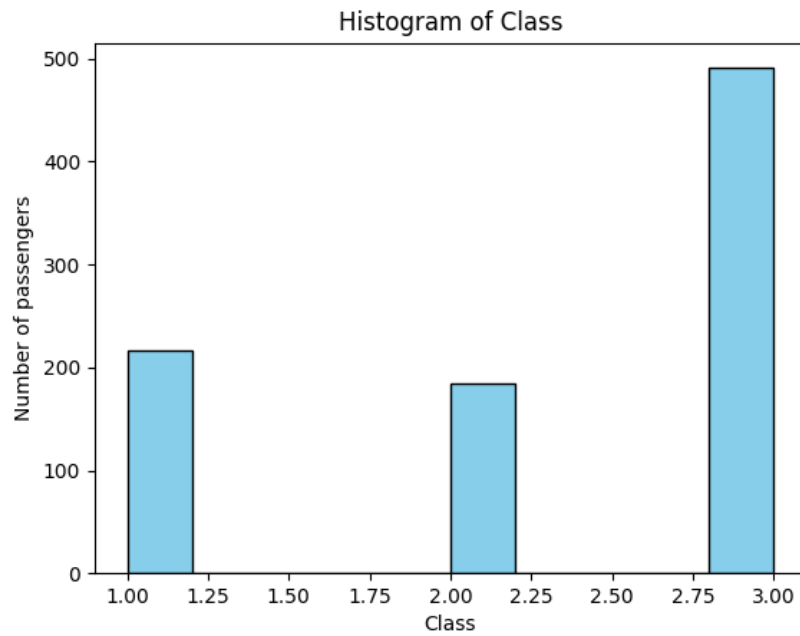
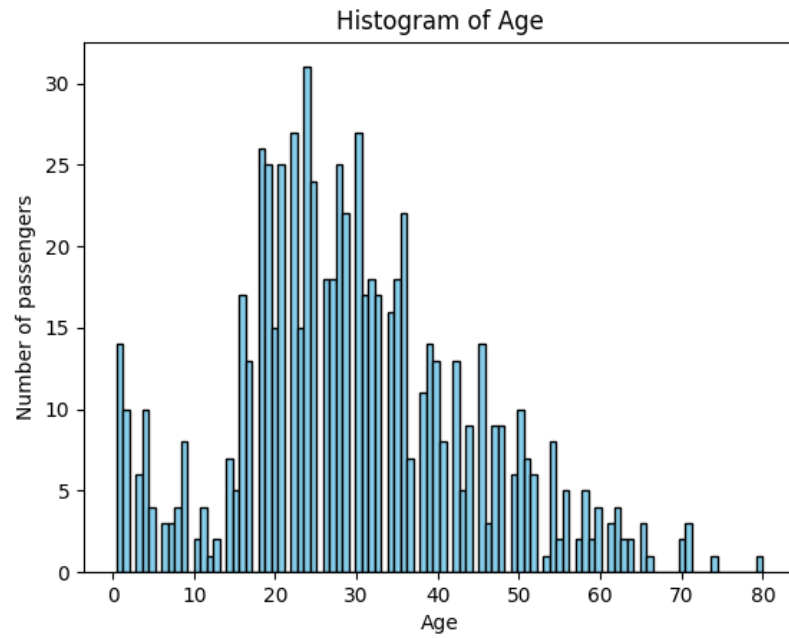
Am ales pentru reprezentarea graficului un pie chart, deoarece valorile constituie procente care se însumează la 100%. Am folosit seaborn pentru crearea unei palete de culori pe care apoi am folosit-o în generarea propriu-zisă a graficului utilizând matplotlib.

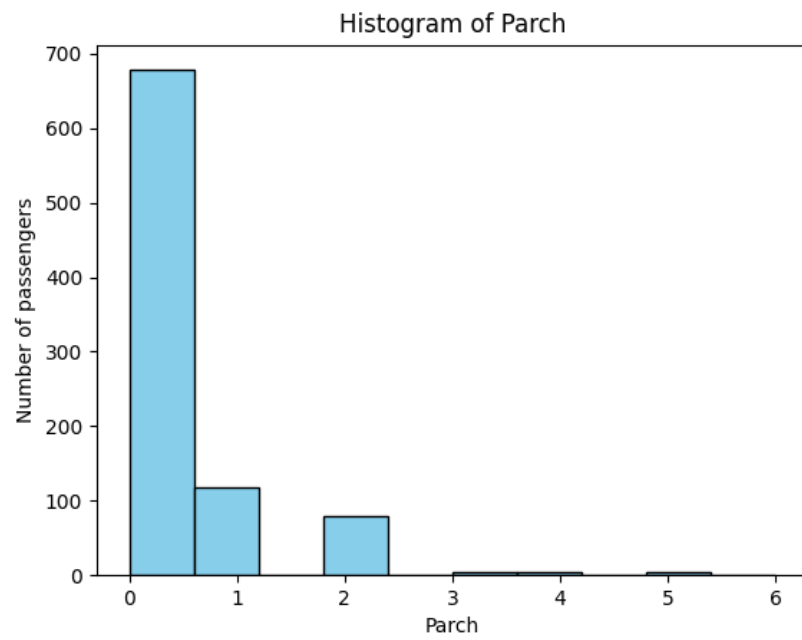
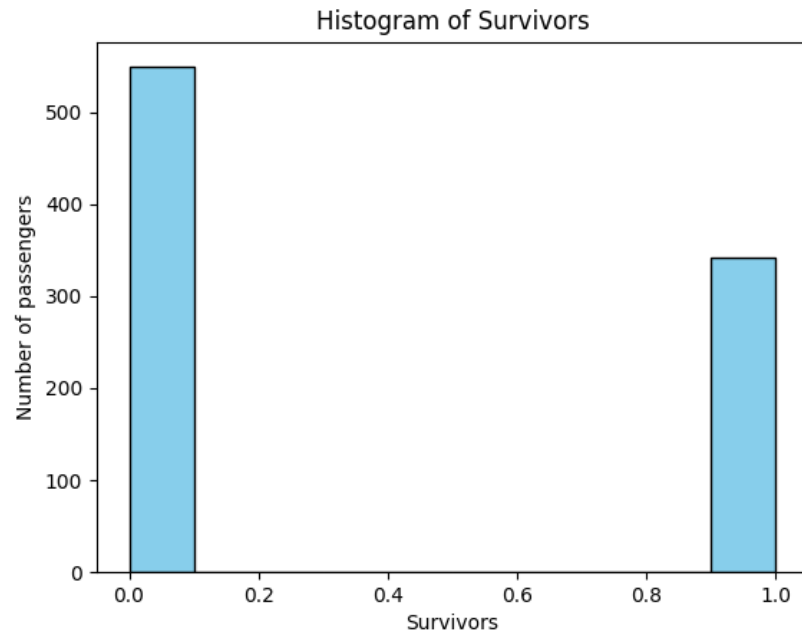


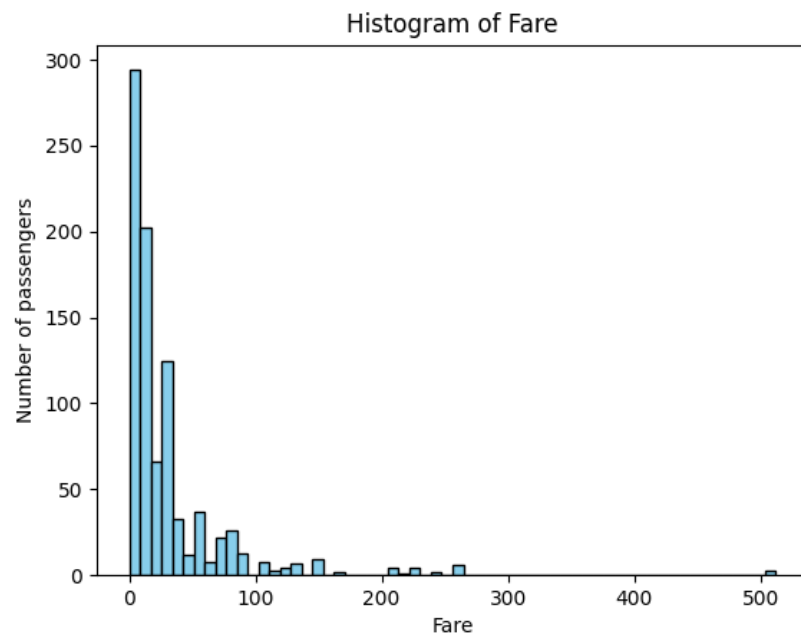
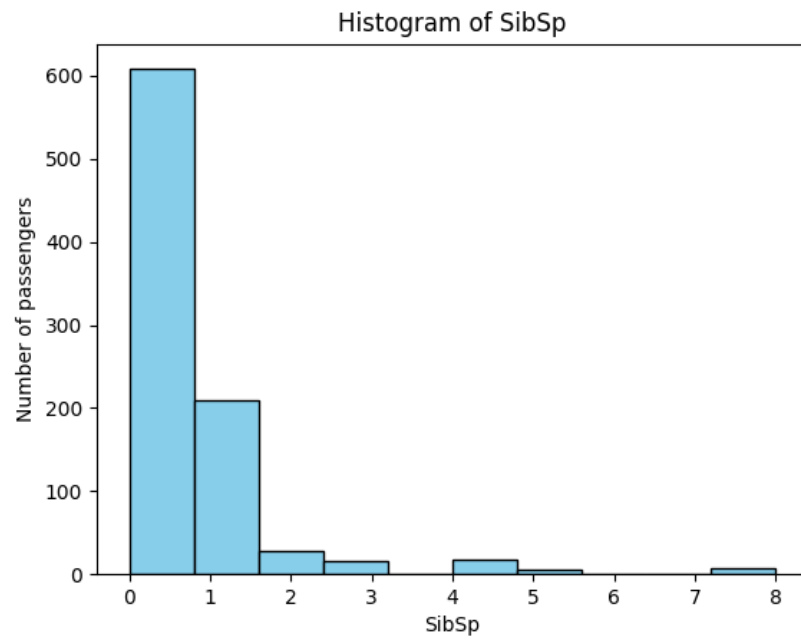


Cerința III

Pentru realizarea acestei cerințe am început prin a-mi construi setul de date pe care să îl reprezint ulterior sub forma un histograme folosindu-mă de funcția `hist()` oferită de biblioteca `matplotlib`.







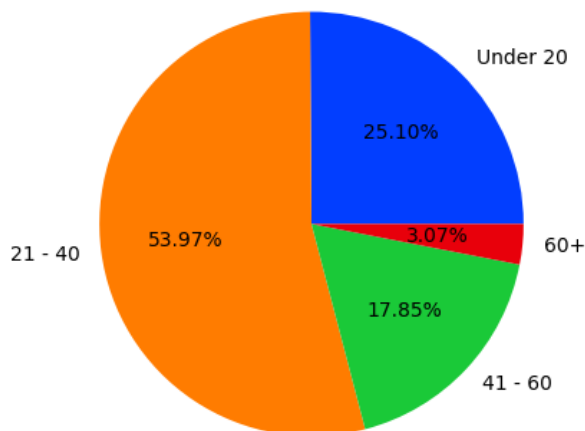
Cerința IV

Pentru realizarea sarcinii propuse am făcut folos de metoda `isnull` din biblioteca `pandas` care returnează o listă booleană a fiecărui element nul din coloană, pe care apoi i-am însumat.

- These are the columns containing null values and their number: Age: 177 Cabin: 687 Embarked: 2
- The percentages of null values: Age: 19.87% Cabin: 77.10% Embarked: 0.22%
- Percentages for survivors: Age: 15.20% Cabin: 60.23% Embarked: 0.58%
- Percentages for the perished: Age: 22.77% Cabin: 87.61%

Cerința V

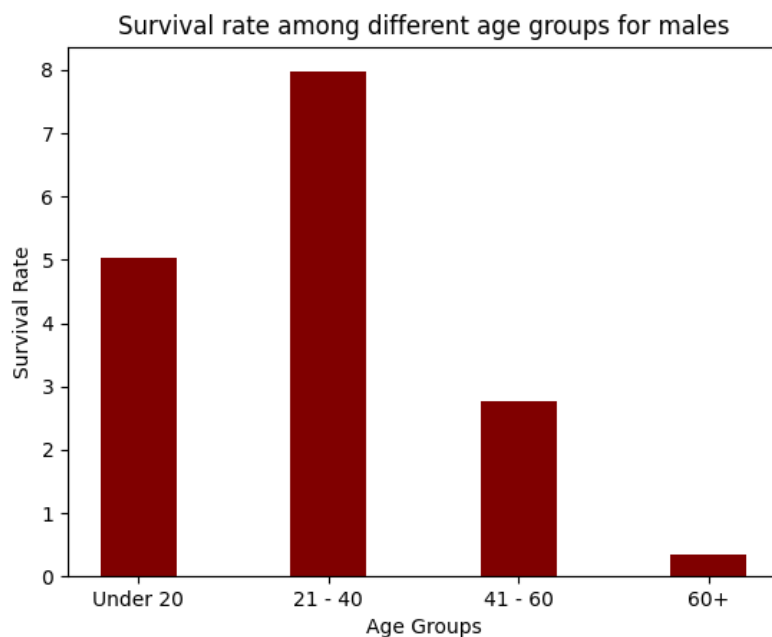
Am început rezolvarea cerinței V prin a identifica numărul precis al pasagerilor care aparțin uneia din cele 4 categorii. Am folosit funcția `len`, care indică numărul de linii într-un data frame. Am folosit `matplotlib` pentru construirea pie chart-ului. În final, am salvat data frame-ul rezultat după adăugarea coloanei de „Age Group” în format `csv`.



Cerința VI

Analog cerinței V, am determinat valorile precise corespunzătoare fiecărei categorii. Apoi am calculat procentele corespunzătoare, pe care le-am reunit într-o listă. Am realizat un bar plot care să reprezinte rezultatele obținute

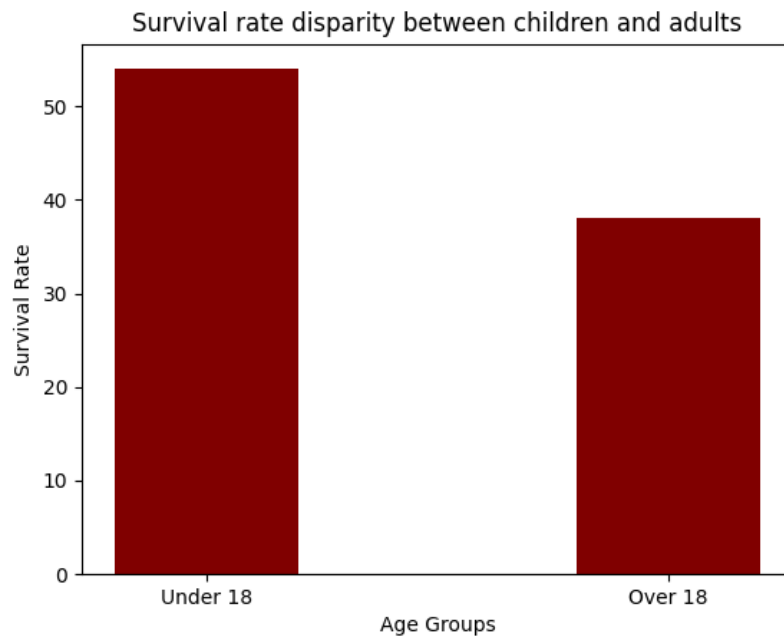
- Under 20: 29 survived, 862 dead
- 21 - 40: 46 survived, 845 dead
- 41 - 60: 16 survived, 875 dead
- 60+: 2 survived, 889 dead



Cerința VII

După ce am extras valorile precise ale fiecărei categorii relevante, anume a copiilor și a adulților, am calculat procentele cerute și le-am reprezentat într-un bar plot, folosindu-mă de biblioteca matplotlib.

- This is the percentage of children on board: 12.68%

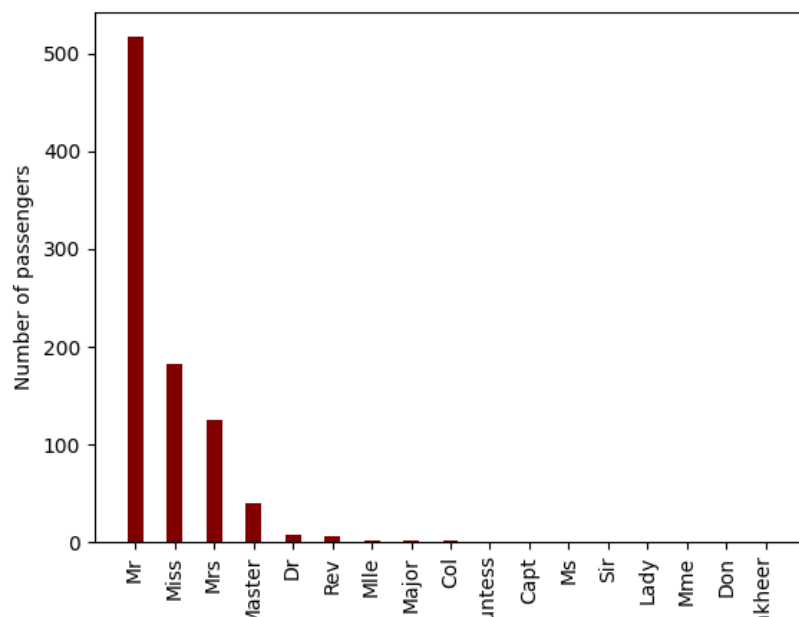


Cerința VIII

Am început rezolvarea cerinței prin segregarea indivizilor din setul de date astfel: în cei care au supraviețuit și cei care au pierit. Următorul pas a fost să identific în cele două mulțimi coloanele cu valori numerice și pe cele fără. După, am calculat valoarea medie pentru coloane numerice, iar pentru cele non-numerice am stabilit care este elementul cu cele mai multe apariții în coloana respectivă. Am folosit utilitarul fillna din pandas pentru completarea data frame-ului. Am salvat într-un fișier .csv data frame-ul completat.

Cerința IX

Cerința am început a o rezolva prin aplicarea comenzii split pe coloana de nume pentru a putea găsi toate titlurile de noblețe din setul de date. După, am construit un dicționar care să permită asocierea dintre titlurile de noblețe și sexul căruia corespund. Am construit două noi coloane în care să evidențiez titlul persoanei și sexul titlului. Apoi am comparat sexul persoanei cu cel corespunzător titlului purtat.



Cerința X

M-am folosit de coloanele SibSp și Parch pentru a putea stabili dacă un călător era, sau nu, singur. Am folosit matplotlib pentru realizarea histogramei. Funcția head mi-a permis salvarea primilor 100 de pasageri. Apoi am introdus corespunzător fiecare dată în utilitarul catplot.

