Braviceanu – Badea Clarisse 312CB – PARTEA 1

# README

**Taskul 1:**

Ne folosim de libraria Pandas pentru a citii datele din fisierul de tip CSV din folderul titanic si le stocam intr-un data frame numit df.

df = pd.read\_csv(*r*'titanic\train.csv')

Pe prima linie este afisat numele taskului urmat de numarul de coloane calculat cu ajutorl lui shape[1].

for i in range(0,df.shape[1]):

    print(df.columns[i],*end*="  ")

Ciclam prin fiecare coloana si afisam numele sau separate prin spatiu, iar pe urmatoarea line se afiseaza in aceeasi ordine tipul de date (df.dtypes.iloc[i] ). Rezultatul avand formatul

Nume coloana

Tip coloana

Urmeaza sa ne folosim de isna() (is not a) pentru a determina elementele cu valori lipsa .

Numarul de linii duplicate se calculeaza prin diferenta dintre numarul total de linii si nr de linii dupa eliminarea celor duplicate cu ajutorul lui drop\_duplicates

print("Numarul de linii dublicate este " + str(len(df.to\_string()) - len(df.drop\_duplicates().to\_string())))

**Taskul 2:**

Procentul supravietuitorilor se calculeaza prin selectarea coloanei Survived si folosindu ne de mean() calculam media iar inmultita cu 100 ne da procentul

survivors = int(df["Survived"].mean()\*100)

Cel al mortilor se scade din 100% pe cei vii

deads = 100 - survivors

Asemanator se calculeaza si femeile cu barbatii

Pentru charturi se foloseste pie

graph, pies = plt.subplots(1, 3)

pies[0].pie(data[0:2], *labels*=data\_names[0:2], *autopct*='%1.1f%%')

pies[1].pie(data[2:4], *labels*=data\_names[2:4], *autopct*='%1.1f%%')

pies[2].pie(data[4:], *labels*=data\_names[4:], *autopct*='%1.1f%%')

plt.show()

luand datele dfin vectorul de data\_names ca labeluri si datele din data.

**Taskul 3:**