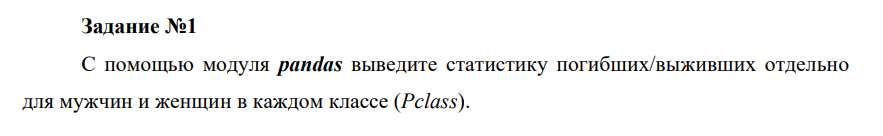
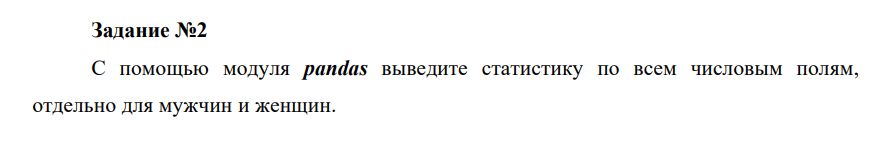
БДЗ   
«Анализ данных на Python»

Выполнил: Арюхин Александр ИВТ-42

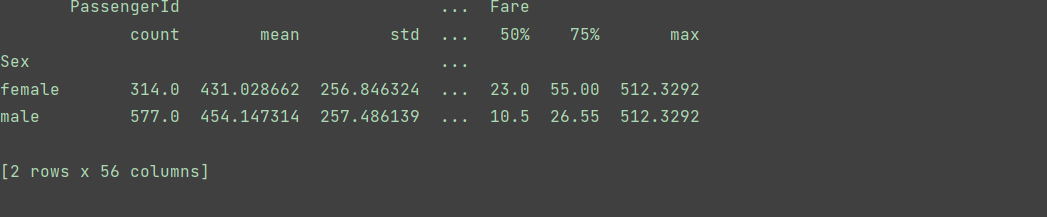


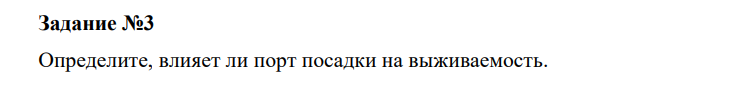
**import** pandas **as** pd  
df **=** pd.read\_csv**('train.csv')**a **=** df.groupby**(['Pclass'**, **'Sex'**, **'Survived'])['Survived']**.agg**('count')**print**(**a**)**



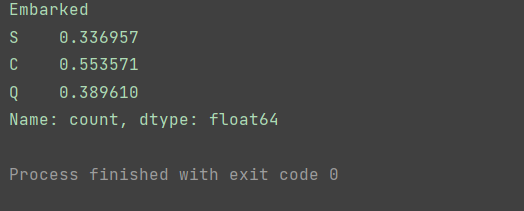


**import** pandas **as** pd  
df **=** pd.read\_csv**('train.csv')**numeric\_columns **=** df.select\_dtypes**(**include**='number')**res **=** df.groupby**('Sex')[**numeric\_columns.columns**]**.describe**()**print**(**res**)**

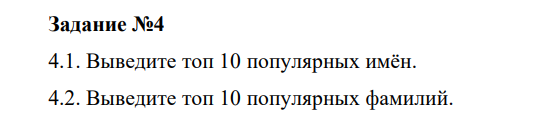




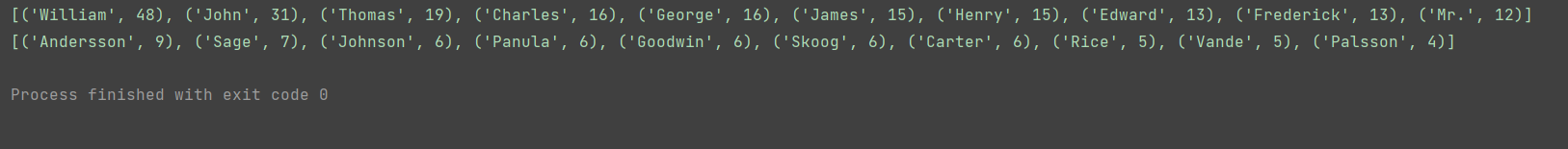
**import** pandas **as** pd  
df **=** pd.read\_csv**('train.csv')**val **=** df**['Embarked']**.value\_counts**()**df\_surv **=** df**[**df**['Survived'] ==** 1**]**val\_surv **=** df\_surv**['Embarked']**.value\_counts**()**print**(**val\_surv **/** val**)**

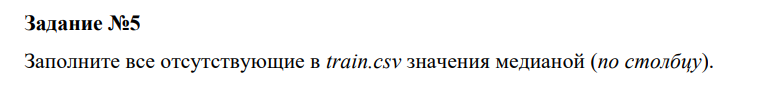


Если исходить только из данных, то можем заметить, что люди, севшие в Cherbourg выжили в бОльшем соотношении.

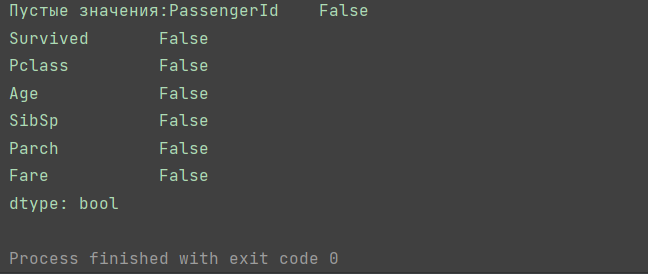


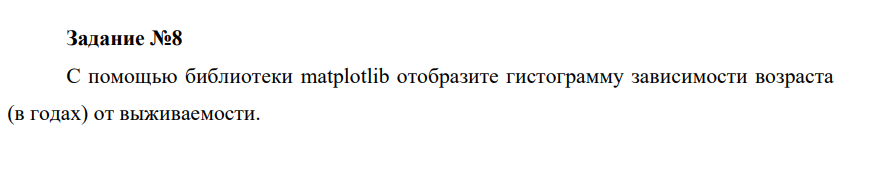
**import** pandas **as** pd  
**from** collections **import** Counter  
df **=** pd.read\_csv**('train.csv')**name **=** df**['Name']**.apply**(lambda** x**:** x.split**())**first\_name **= []**second\_name **= []  
for** i **in** name**:** first\_name.append**(**i**[**2**])** second\_name.append**(**i**[**0**][:-**1**])**print**(**Counter**(**first\_name**)**.most\_common**(**10**))**print**(**Counter**(**second\_name**)**.most\_common**(**10**))**





**import** pandas **as** pd  
df **=** pd.read\_csv**('train.csv')**numeric\_columns **=** df.select\_dtypes**(**include**='number')  
for** column **in** numeric\_columns**:** median\_value **=** numeric\_columns**[**column**]**.median**()** numeric\_columns**[**column**]**.fillna**(**median\_value, inplace**=True)**missing\_values **=** numeric\_columns.isnull**()**.any**()**print**(f'Пустые значения:{**missing\_values**}')**





**import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
df **=** pd.read\_csv**('train.csv')**survived\_age **=** df**[**df**['Survived'] ==** 1**]['Age']**.dropna**()**not\_survived\_age **=** df**[**df**['Survived'] ==** 0**]['Age']**.dropna**()**plt.hist**([**survived\_age, not\_survived\_age**]**, bins**=**20, label**=['Выжившие'**, **'Не выжившие']**, alpha**=**0.7, color**=['green'**, **'red'])**plt.title**('Гистограмма зависимости возраста от выживаемости')**plt.xlabel**('Возраст')**plt.ylabel**('Частота')**plt.legend**()**# Отображение гистограммы  
plt.show**()**

