

```
Ввод [1]: # Read the file
import pandas as pd
titanik_data = pd.read_csv('titanik.csv')

# Check sums by Sex
all_female_passengers_count = titanik_data.loc[titanik_data['Sex'] == 'female'].shape[0]
all_male_passengers_count = titanik_data.loc[titanik_data['Sex'] == 'male'].shape[0]
any_gender_count = titanik_data.shape[0]

print(f"All female passengers = {all_female_passengers_count}")
print(f"All male passengers = {all_male_passengers_count}")
print(f"All passengers = {any_gender_count}")

if (all_female_passengers_count + all_male_passengers_count) == any_gender_count:
    print('Every passenger has a gender')
else:
    print(f'Number of passenger without gender = {any_gender_count - all_female_passengers_count - all_male_passengers_

All female passengers = 314
All male passengers = 577
All passengers = 891
Every passenger has a gender
```

В dataset находятся данные о 891-м пассажире. Среди них: 314 женщин и 577 мужчин.

```
Ввод [2]: all_female_passengers_survived = titanic_data.loc[(titanic_data['Sex'] == 'female') & (titanic_data['Survived'] == 1)].  
print(f"Survived female passengers = {all_female_passengers_survived}")  
  
all_male_passengers_survived = titanic_data.loc[(titanic_data['Sex'] == 'male') & (titanic_data['Survived'] == 1)].shape  
print(f"Survived male passengers = {all_male_passengers_survived}")  
  
Survived female passengers = 233  
Survived male passengers = 109
```

Выжившие пассажиры: 233 женщины и 109 мужчин.

```
Ввод [3]: # Check sums by Pclass
first_class_count = titanik_data[titanik_data.Pclass == 1].shape[0]
second_class_count = titanik_data[titanik_data.Pclass == 2].shape[0]
third_class_count = titanik_data[titanik_data.Pclass == 3].shape[0]
any_class_count = titanik_data.shape[0]
print(f"First class = {first_class_count}")
print(f"Second class = {second_class_count}")
print(f"Third class = {third_class_count}")
print(f"All passengers = {any_class_count}")
if first_class_count + second_class_count + third_class_count == any_class_count:
    print('Every passenger has a class')
else:
    print(f'Number of passenger without class = {any_class_count - first_class_count - second_class_count - third_class_count}')

First class = 216
Second class = 184
Third class = 491
All passengers = 891
Every passenger has a class
```

216 пассажиров находились в первом классе, 184 - во втором и 491 - в третьем.

```
Ввод [4]: # Check sums by Age
empty_age_count = titanik_data[titanik_data.Age.isnull()].shape[0]
number_age_count = titanik_data[titanik_data.Age.notnull()].shape[0]
any_age_count = titanik_data.shape[0]
print(f"Passengers with empty Age = {empty_age_count}")
print(f"Passengers with defined Age = {number_age_count}")
print(f"All passengers = {any_age_count}")
```

```
Passengers with empty Age = 177
Passengers with defined Age = 714
All passengers = 891
```

В dataset для 714 пассажиров указан возраст. Для 177 пассажиров данные о возрасте отсутствуют.

```
Ввод [5]: titanic_data[['PassengerId', 'Name', 'Ticket', 'Sex', 'Survived', 'Pclass', 'Embarked']].nunique()
```

```
Out[5]: PassengerId      891  
        Name            891  
        Ticket          681  
        Sex              2  
        Survived         2  
        Pclass           3  
        Embarked         3  
        dtype: int64
```

Уникальность данных:

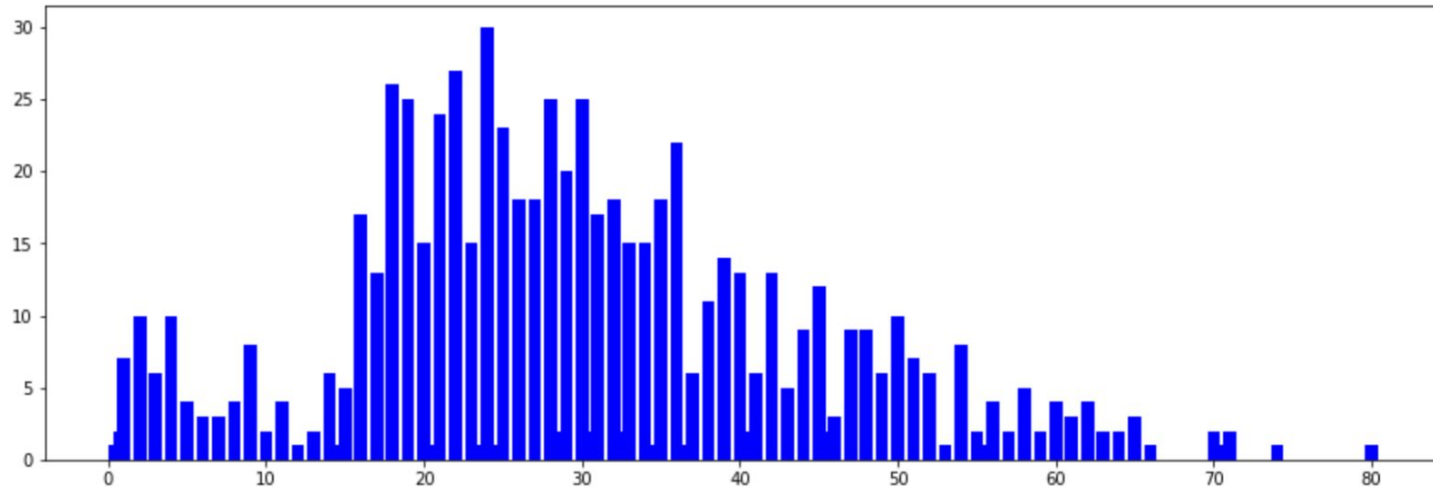
Нет повторяющихся ID пассажиров, полных имен.

Номера билетов повторяются. В этих случаях, скорее всего, родитель и ребенок были зарегистрированы по одному билету.

В данных 2 пола, 2 статуса выживания, 3 класса пассажира - данные корректны.

Также можно видеть, что у нас есть три порта посадки.

```
Ввод [15]: all_people = []  
for i in range(0, (len(dead))):  
    all_people.append(dead[i] + survived[i])  
f = plt.figure()  
f.set_figwidth(15)  
f.set_figheight(5)  
plt.bar(age, all_people, color='b')  
plt.show()
```



Распределение пассажиров по возрасту. Видно, что основная масса пассажиров была в возрасте от 15 до 40 лет.

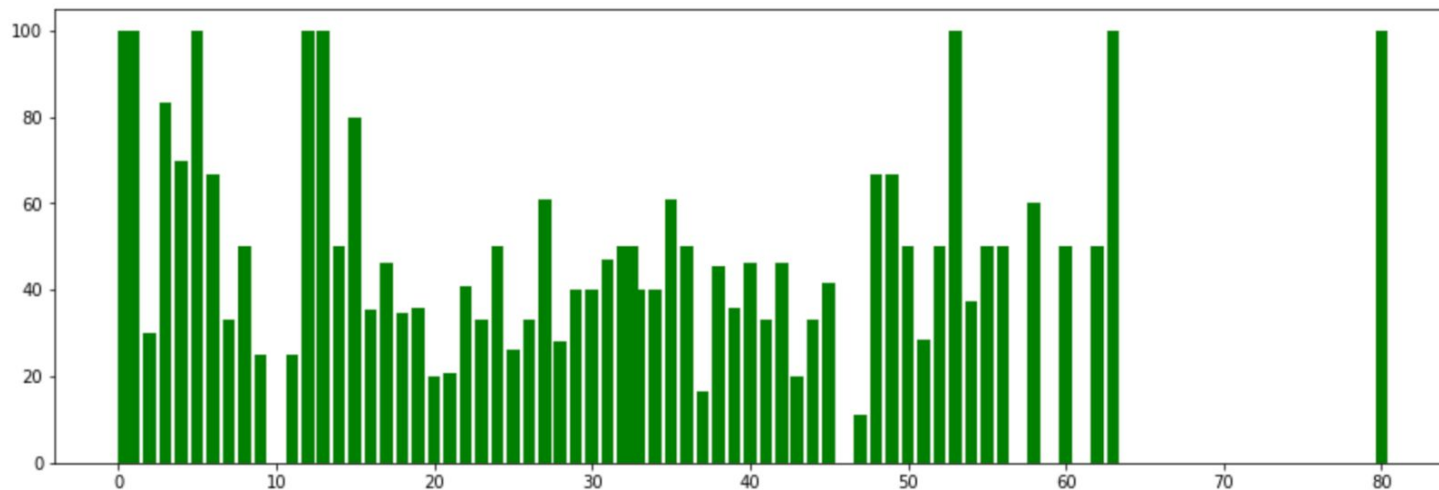
```
Ввод [6]: passengers_died = titanik_data[titanik_data.Survived == 0].shape[0]
passengers_survived = titanik_data[titanik_data.Survived == 1].shape[0]
survival_ratio = passengers_survived / (passengers_died + passengers_survived)
print(f"Survived = {round(survival_ratio*100)}%")
```

Survived = 38%

Общий процент выживших составляет 38%.

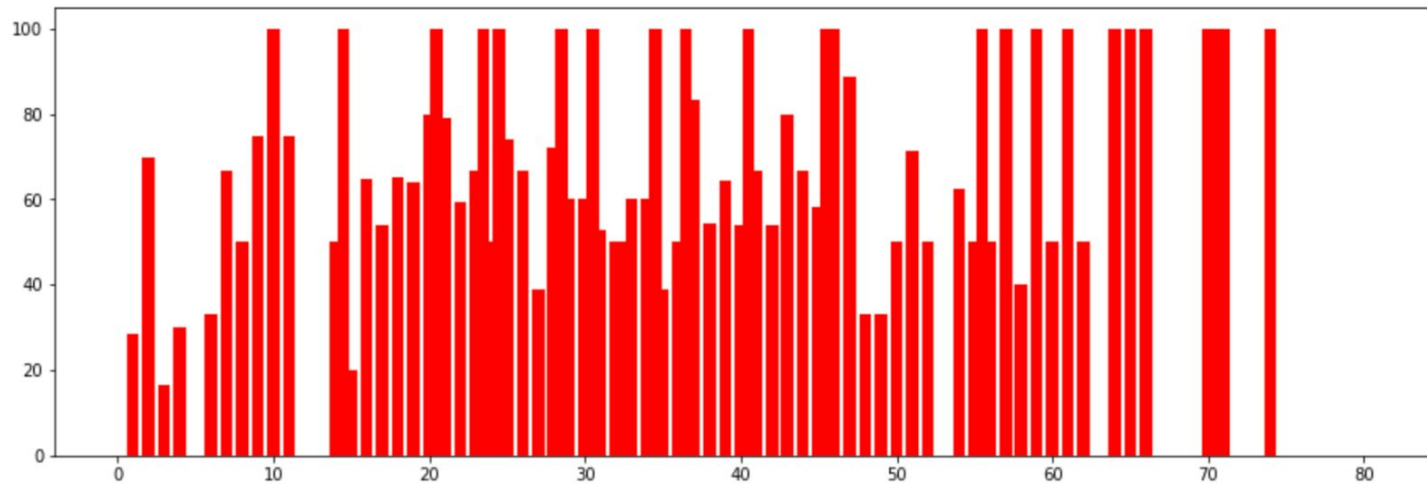
Ввод [13]: `import matplotlib.pyplot as plt`

```
f = plt.figure()
f.set_figwidth(15)
f.set_figheight(5)
plt.bar(age, surv_ratio, color='g')
plt.show()
```



Распределение выживших в зависимости от возраста. Высокая выживаемость среди маленьких детей. Их могли сажать в шлюпки вместе с матерями среди одних из первых.


```
Ввод [14]: dead_ration = []  
for i in range(0, (len(dead))):  
    dead_ration.append(round(100 * dead[i] / (dead[i] + survived[i]), 2))  
f = plt.figure()  
f.set_figwidth(15)  
f.set_figheight(5)  
plt.bar(age, dead_ration, color='r')  
plt.show()
```



Распределение погибших в зависимости от возраста.

```
Ввод [16]: third_class_surv = titanic_data.loc[(titanic_data['Pclass'] == 3) & (titanic_data['Survived'] == 1)].shape[0]
third_class_all = titanic_data.loc[(titanic_data['Pclass'] == 3)].shape[0]
print(f"All 3rd class = {third_class_all}")
print(f"Survived 3rd class = {round(100 * third_class_surv/third_class_all)}%")
```

```
All 3rd class = 491
Survived 3rd class = 24%
```

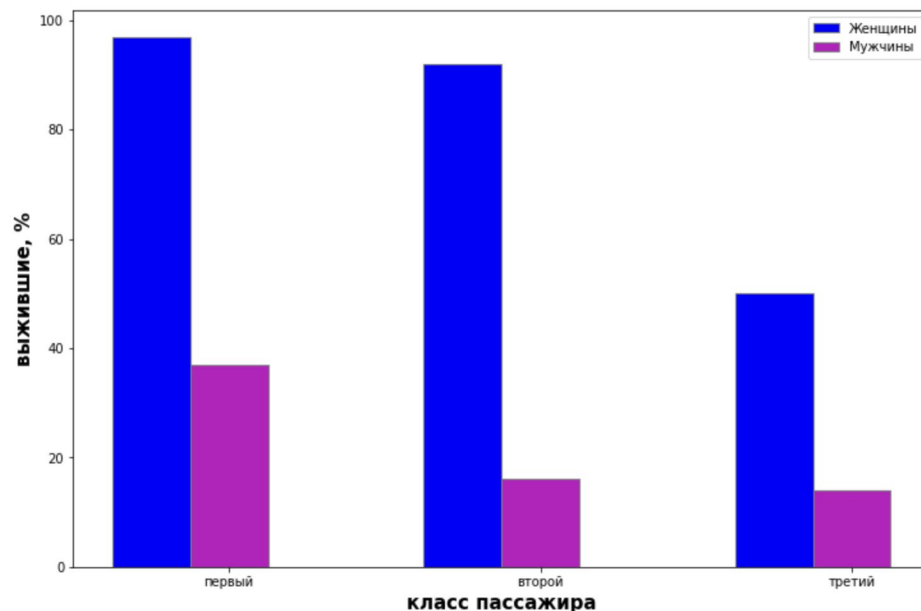
```
Ввод [17]: second_class_surv = titanic_data.loc[(titanic_data['Pclass'] == 2) & (titanic_data['Survived'] == 1)].shape[0]
second_class_all = titanic_data.loc[(titanic_data['Pclass'] == 2)].shape[0]
print(f"All 2nd class = {second_class_all}")
print(f"Survived 2nd class = {round(100 * second_class_surv/second_class_all)}%")
```

```
All 2nd class = 184
Survived 2nd class = 47%
```

```
Ввод [18]: first_class_surv = titanic_data.loc[(titanic_data['Pclass'] == 1) & (titanic_data['Survived'] == 1)].shape[0]
first_class_all = titanic_data.loc[(titanic_data['Pclass'] == 1)].shape[0]
print(f"All 1st class = {first_class_all}")
print(f"Survived 1st class = {round(100 * first_class_surv/first_class_all)}%")
```

```
All 1st class = 216
Survived 1st class = 63%
```

Процент выживших в зависимости от класса пассажира. Наблюдается рост выживаемости в сторону первого класса. Связано с тем, что третий класс, например, находился на самых нижних палубах и выбраться оттуда практически не представлялось возможным.



ПЕРВЫЙ КЛАСС женщины: 97%, мужчины: 37%
ВТОРОЙ КЛАСС женщины: 92%, мужчины: 16%
ТРЕТИЙ КЛАСС женщины: 50%, мужчины: 14%

Процент выживаемости в зависимости от класса пассажира с разделением на мужчин и женщин. Видим, что женщин в процентном соотношении спаслось намного больше. Практически все женщины из 1-го и 2-го классов спаслись. Большая часть мужчин погибла.

Также наблюдается снижение процента выживших в сторону третьего класса в обеих группах.