# Анализ данных о пассажирах титаника с помощью Pandas

Необходимо дать ответы на следующие вопросы:

1. Какое количество мужчин и женщин ехало на корабле?
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров. Ответ приведите в процентах.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров? Ответ приведите в процентах.
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров.
5. Какое самое популярное женское имя на корабле?

Импортируем необходимые библиотеки:

In [1]:
```
import pandas as pd
```

Прочитаем файл *titanic.csv* и посмотрим на его содержимое:

In [5]:
```
titanic_df = pd.read_csv('titanic.csv')
titanic_df
```

Out[5]:

| | PassengerID | Name | PClass | Age | Sex | Survived | SexCode |
|---|---|---|---|---|---|---|---|
| **0** | 1 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | 1 | 1 |
| **1** | 2 | Allison, Miss Helen Loraine | 1st | 2.00 | female | 0 | 1 |
| **2** | 3 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | 0 | 0 |
| **3** | 4 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | 0 | 1 |
| **4** | 5 | Allison, Master Hudson Trevor | 1st | 0.92 | male | 1 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **1308** | 1309 | Zakarian, Mr Artun | 3rd | 27.00 | male | 0 | 0 |
| **1309** | 1310 | Zakarian, Mr Maprieder | 3rd | 26.00 | male | 0 | 0 |
| **1310** | 1311 | Zenni, Mr Philip | 3rd | 22.00 | male | 0 | 0 |
| **1311** | 1312 | Lievens, Mr Rene | 3rd | 24.00 | male | 0 | 0 |
| **1312** | 1313 | Zimmerman, Leo | 3rd | 29.00 | male | 0 | 0 |

1313 rows × 7 columns

## 1. Количество мужчин и женщин

In [23]:
```
sex_counts = titanic_df['Sex'].value_counts()
print(f"Мужчин: {sex_counts['male']}\nЖенщин: {sex_counts['female']
}")
```

```
Мужчин: 851
Женщин: 462
```

## 2. Доля выживших пассажиров

In [17]:
```
surv_counts = titanic_df['Survived'].value_counts()
surv_percent = 100.0 * surv_counts[1] / surv_counts.sum()
print(f"Доля выживших пассажиров: {surv_percent:0.2f}%")
```

```
Доля выживших пассажиров: 34.27%
```

## 3. Доля пассажиров первого класса

In [22]:
```
pclass_counts = titanic_df['PClass'].value_counts()
pclass_percent = 100.0 * pclass_counts[1] / pclass_counts.sum()
print(f"Доля пассажиров первого класса: {pclass_percent:0.2f}%")
```

```
Доля пассажиров первого класса: 24.52%
```

## 4. Среднее и медиана возраста пассажиров

In [26]:
```
ages = titanic_df['Age'].dropna()
print(f"Среднее возраста пассажиров: {ages.mean():0.2f}\nМедиана во
зраста пассажиров: {ages.median():0.2f}")
```

```
Среднее возраста пассажиров: 30.40
Медиана возраста пассажиров: 28.00
```

## 5. Самое популярное женское имя

Сначала найдём самую популярную фамилию на корабле (так сделало большинство студентов):

In [49]:
```
females_df = titanic_df[titanic_df['Sex'] == 'female']
most_popular = females_df['Name'].apply(lambda x: x.split(',')[0]).
value_counts()
print(f"Самая популярная женская фамилия на корабле: {most_popular.
index[0]}")
```

```
Самая популярная женская фамилия на корабле: Andersson
```

На самом деле в этом наборе данных имя идёт после запятой. Давайте попробуем найти самое популярное женское имя. Сначала определим функцию, которая вытащит имя из строки, для этого будем использовать регулярные выражения.

In [127]:
```
import re

def clean_name(name):

    # Первое слово до запятой - фамилия
    s = re.search('^[^,]+, (.*)', name)
    if s:
        name = s.group(1)

    # Если есть скобки - то имя пассажира в них
    s = re.search('\(([^)]+)\)', name)
    if s:
        name = s.group(1)

    # Удаляем обращения
    name = re.sub('Miss ', '', name)
    name = re.sub('Mrs ', '', name)
    name = re.sub('Ms ', '', name)

    # Берем первое оставшееся слово и удаляем кавычки
    name = name.split(' ')[0].replace('"', '')
    return name
```

С помощью этой функции обработаем записи набора данных:

In [129]:
```
names = females_df['Name'].map(clean_name)
name_counts = names.value_counts()
print(f"Самое популярное женское имя на корабле: {name_counts.index
[0]}")
```

```
Самое популярное женское имя на корабле: Mary
```