

# Практическая работа №8

## Задание

Выполните задание с применением собственных функций на основе набора данных **films.csv**: Какой актер принес меньше всего прибыли в **2012** году?

Загружаем библиотеки и загружаем файл **csv**

In [2]:

```
import pandas as pd
import numpy as np

df = pd.read_csv('films.csv')
```

Смотрим информацию о датафрейме

все ли столбцы заполнили, в каком формате значения в данных столбцах, как называются заголовки

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1890 entries, 0 to 1889
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   imdb_id               1890 non-null  object
1   popularity             1890 non-null  float64
2   budget                1890 non-null  int64
3   revenue                1890 non-null  int64
4   original_title        1890 non-null  object
5   cast                  1890 non-null  object
6   director              1890 non-null  object
7   tagline                1890 non-null  object
8   overview               1890 non-null  object
9   runtime                1890 non-null  int64
10  genres                 1890 non-null  object
11  production_companies  1890 non-null  object
12  release_date           1890 non-null  object
13  vote_count             1890 non-null  int64
14  vote_average           1890 non-null  float64
15  release_year           1890 non-null  int64
dtypes: float64(2), int64(5), object(9)
memory usage: 236.4+ KB
```

Изменяем старый датафрейм оставляя фильмы только **2012** года

In [4]:

```
df = df.loc[df['release_year'] == 2012]
```

Пишем 2 простенькие функции для расчета прибыли / убытка с фильма и создаем списки с актерами

Добавляем новый столбец в датафрейм с прибылью/убытком за фильм Создаем список всех актеров игравших в фильмах за **2012** год Далее создаем словарь ключ - актер, значение - сумма заработка за **2012** год, до итерации **0** Далее проходимся по циклу в цикле и суммируем для каждого актера денги которые заработал фильм в котором они играли

In [5]:

```
def cost_cast(budget, revenue):
    c = revenue - budget
    return c

def split_actor(cast):
    list_cast = cast.split('|')
    return list_cast

df.loc[:, 'cash'] = list(map(cost_cast, df['budget'], df['revenue']))

list_all_actors = []
for i in df['cast']:
    list_cast = split_actor(i)
    list_all_actors += list_cast
set_1 = set(list_all_actors)
set_1 = list(set_1)

follow = dict.fromkeys(set_1, 0)

for g in range(971, 1081):
    for i in set_1:
        if i in split_actor(df['cast'][g]):
            follow[i] += int(df['cash'][g])
```

Ищем минимальное значение в словаре

с помощью цикла находим кому принадлежит данная сумма ответ **Kirsten Dunst**

In [10]:

```
name = min(follow.values())

for k, v in follow.items():
    if v == name:
        print(k)
```

Kirsten Dunst