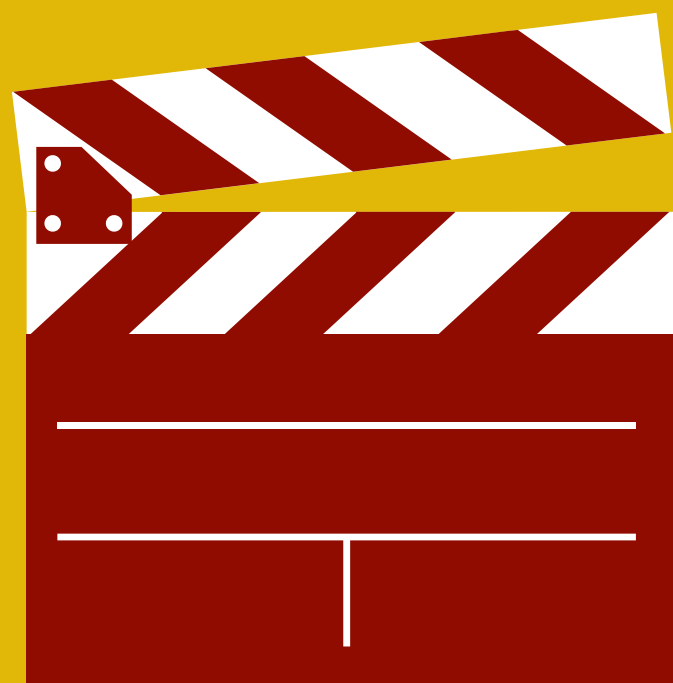


КЕЙС №6



САМЫЕ ПОПУЛЯРНЫЕ ФИЛЬМЫ IMDB

Цель:

Написать код для проверки гипотезы: "Фильмы которые длятся дольше 2.5 часов популярнее фильмов которые длятся меньше 2.5 часов"



Разработка кода:

```
import pandas as pd
import matplotlib.pyplot as plt

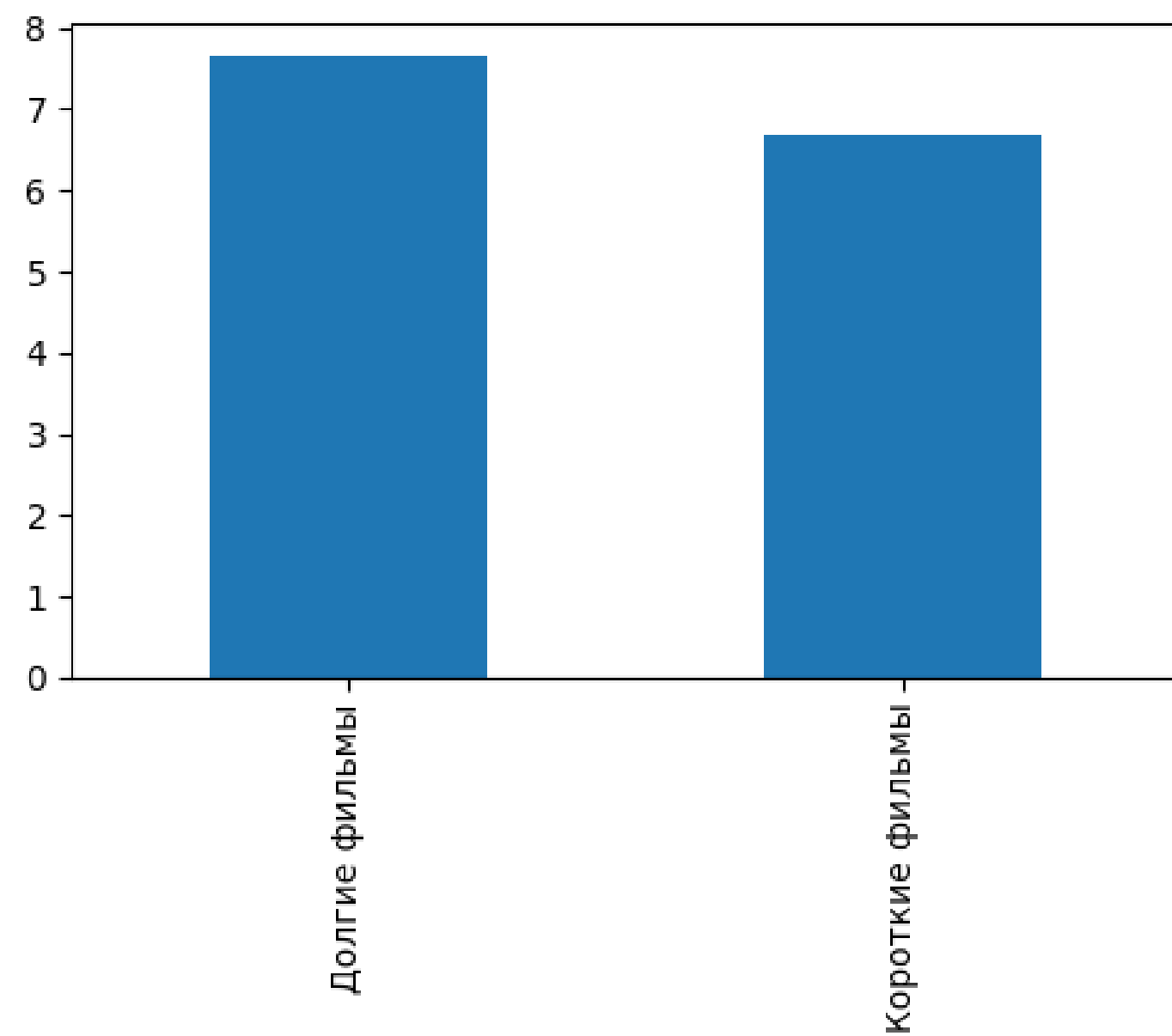
df = pd.read_csv('IMDB-Movie-Data.csv')

df['Runtime (Minutes)'].fillna(0, inplace = True)
df['Rating'].fillna(0, inplace = True)

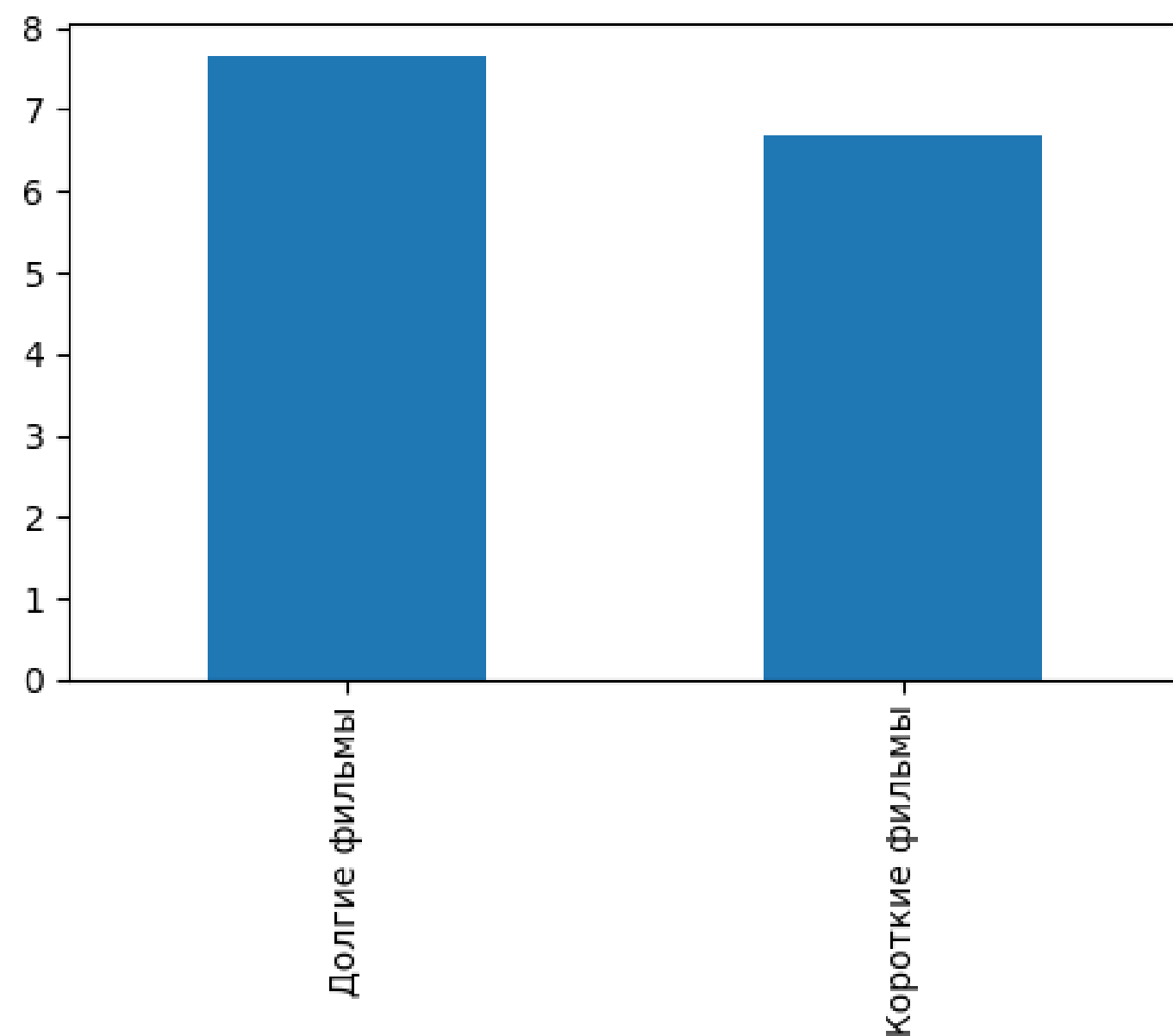
rt1 = df[(df['Runtime (Minutes)'] > 150)]['Rating'].mean()
rt2 = df[(df['Runtime (Minutes)'] < 150)]['Rating'].mean()

s = pd.Series(data = [rt1, rt2],
index = ['Долгие фильмы', 'Короткие фильмы'])
s.plot(kind = 'bar')
plt.show()
```

РЕЗУЛЬТАТ:



РЕЗУЛЬТАТ:



Как мы можем
видеть по графику, у
фильмов дольше 2.5
часов рейтинг выше
чем у фильмов
меньше 2.5 часов



ВЫВОД

НАША ГИПОТЕЗА БЫЛА ДОКАЗАНА



СПАСИБО ЗА ВНИМАНИЕ!

КОНЕЦ