

Университет ИТМО

Практическая работа №2
по дисциплине «Визуализация и моделирование»

Автор: Ефимов Павел Леонидови
Поток: ВИМ 1.1
Группа: К3220
Факультет: ИКТ
Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Датасет: <https://www.kaggle.com/tmdb/tmdb-movie-metadata>

Датасет содержит 5 тысяч фильмов с TMDb. В датасете собраны такие данные как: бюджет, компания производитель, дата выхода, прибыль, средний балл.

Название столбца	Данные, хранящиеся в столбце	Тип данных	Шкала
budget	Бюджет фильма	Целое число	Относительная
genres	Жанры	Строка	Номинальная
homepage	Сайт фильма	Строка	Номинальная
id	Номер фильма	Целое число	Номинальная
keywords	Ключевые слова	Строки	Номинальная
original_language	Язык оригинала	Строка	Номинальная
original_title	Оригинальное название	Строка	Номинальная
overview	Описание	Строка	Номинальная
popularity	Популярность	Число с плавающей точкой	Порядковая
production_companies	Компания производитель	Строка	Номинальная
production_countries	Страна производитель	Строка	Номинальная
release_date	Дата выхода	Дата	Номинальная
revenue	Прибыль	Целое число	Относительная
runtime	Время показа на экране	Целое число	Интервальная
spoken_languages	Язык озвучки фильма	Строка	Номинальная
status	Статус	Строка	Номинальная
tagline	Слоган	Строка	Номинальная
title	Название	Строка	Номинальная
vote_average	Средняя оценка	Число с плавающей точкой	Порядковая
vote_count	Количество голосов	Целое число	Порядковая

Вопросы/гипотезы:

1. Какой жанр чаще всего выпускают
2. Какой бюджет у всех фильмов
3. Зависит ли популярность фильма от сайта
4. Зависит ли прибыль фильма от его продолжительности
5. Какие самые популярные студии
6. Есть ли зависимость между рейтингом и прибылью фильма
7. Какое отношение между количеством голосов и популярностью фильма

8. Как зависит средняя оценка фильма и их количество
9. Какая страна произвела больше всего фильмов
10. Какая страна выпускает самые качественные фильмы

1 Ответы на гипотезы

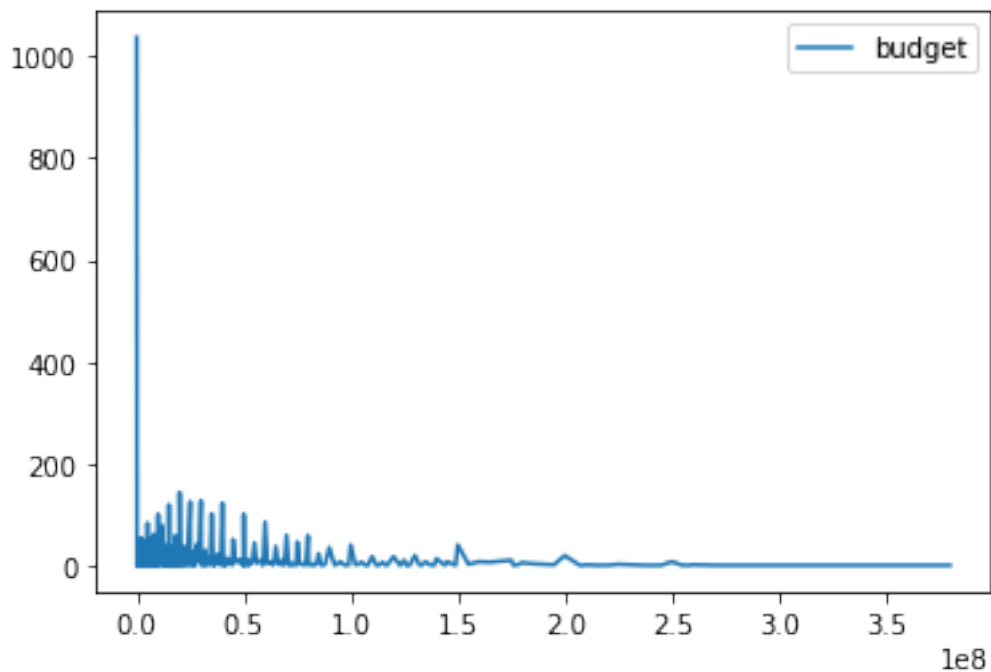
1.1 Какой жанр чаще всего выпускают

Как видно из рисунка, чаще всего выпускают Drama, Comedy и Thriller это означает, что данные жанры пользуются спросом.

	Count
Drama	998
Comedy	877
Action	835
Thriller	727
Adventure	617
Romance	400
Crime	395
Family	373
Science Fiction	351
Fantasy	336
Mystery	203
Animation	194
Horror	175
History	119
War	81
Music	77
Western	37
Documentary	4
Foreign	0

1.2 Какой бюджет у всех фильмов

Как видно из графика, большинство фильмов являются любительскими, так что бюджет на них не нужен. В основном же для фильмов требуется от 0.1 до 1 миллиона долларов.



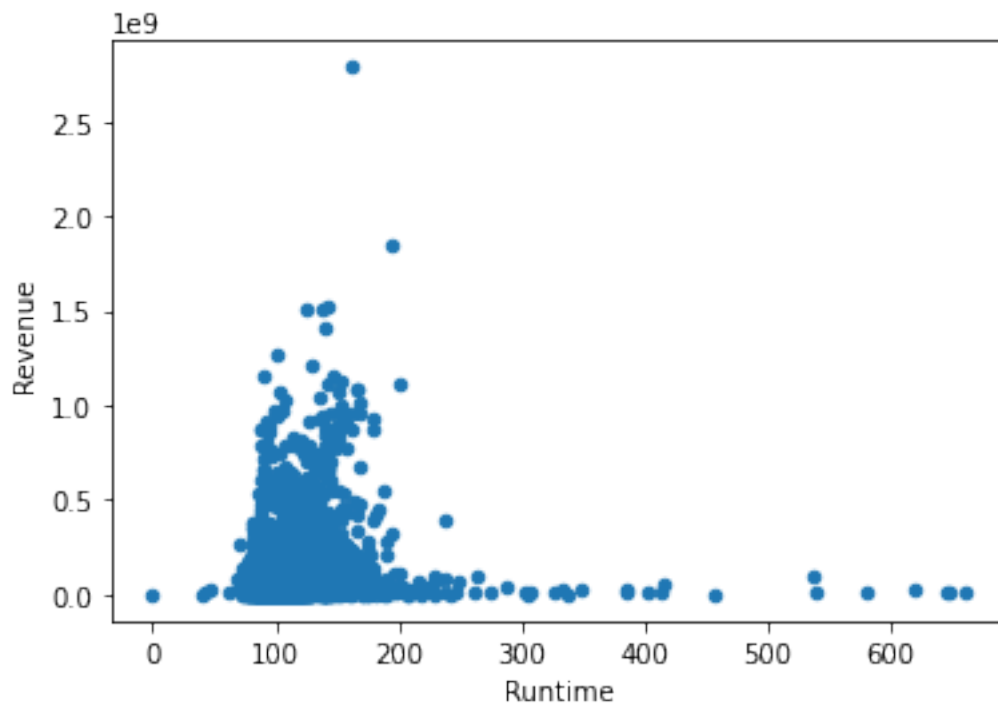
1.3 Зависит ли популярность фильма от сайта

Как видно на таблице, сайты имеющие свой сайт в среднем имеют рейтинг больше.

Average votes	
Exist	6.244685
notExist	6.007700

1.4 Зависит ли прибыль фильма от его продолжительности

Зависимость между прибылью фильма и его продолжительностью не наблюдается, но можно уточнить, что большинство фильмов снимаются в диапазоне 100-200 минут



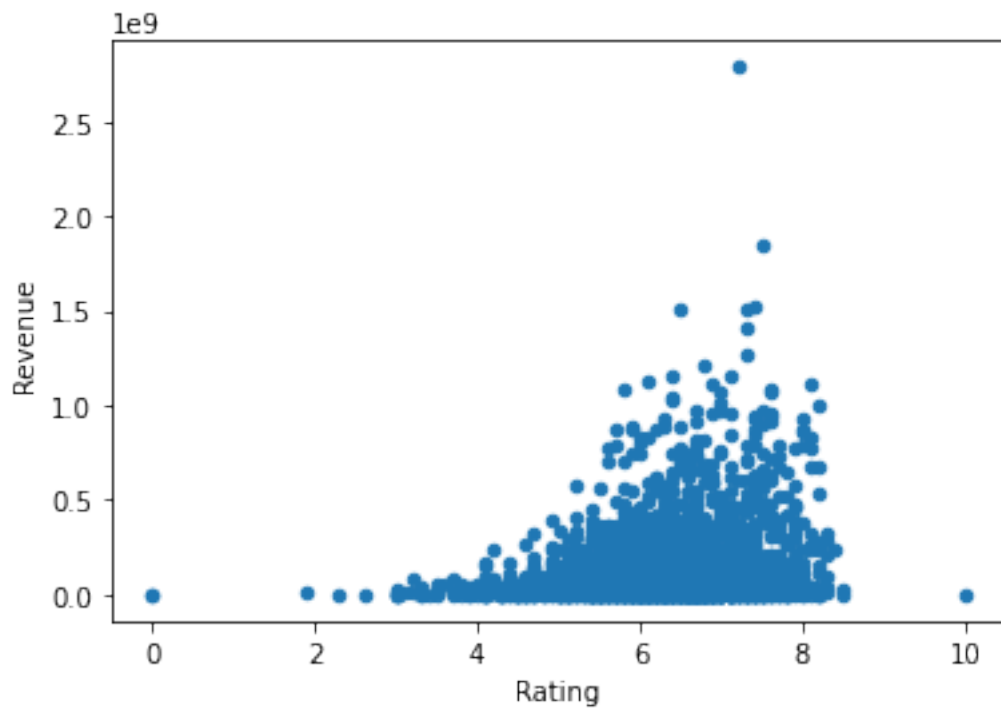
1.5 Какие самые популярные студии

Расчет популярности происходит из суммы популярности каждого фильма деленное на количество всех фильмов. Таким образом The Donners' Company оказалась самой популярной компанией.

	Popularity
The Donners' Company	514.569956
Bulletproof Cupid	481.098624
Kinberg Genre	326.920999
Illumination Entertainment	234.920042
Deluxe Digital Studios	198.372395
...	...
Indiana Production Company	0.008263
Canal Plus Overseas	0.007254
Aztec Musique	0.007254
Illumination Pictures	0.006323
B.Sting Entertainment	0.006069

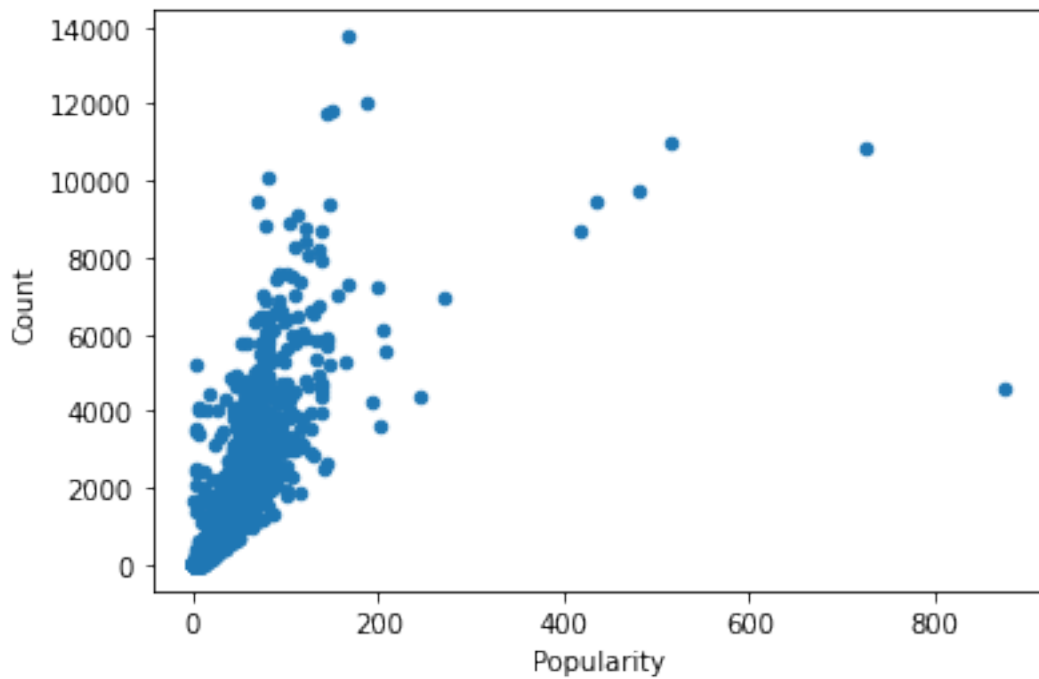
1.6 Есть ли зависимость между рейтингом и прибылью фильма

Есть прямое отношение между прибылью фильма и его рейтингом, самые прибыльные фильмы находятся между рейтингами 6 и 8



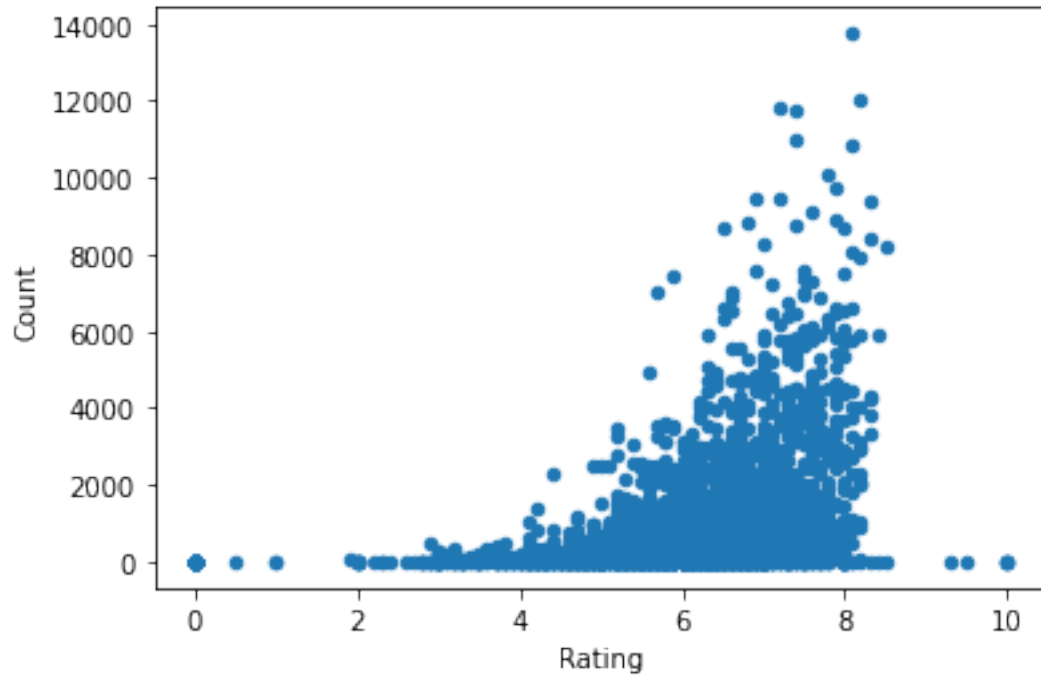
1.7 Какое отношение между количеством голосов и популярностью фильма

Есть отношение между количеством голосов и популярностью фильма, данные значения возрастают по мере увеличения числа голосов и популярности фильма



1.8 Как зависит средняя оценка фильма и их количество

Количество голосов возрастает по мере рейтинга фильма (8 баллов), затем заметно резкое падение, которое вероятно связано с малым количеством фильмов высоких по качеству



1.9 Какая страна произвела больше всего фильмов

Чаще всего фильмы снимают в United States of America, в United Kingdom и Germany уже меньше. Это означает, что United States of America развивает свою киноиндустрию.

	Count
United States of America	3956
United Kingdom	636
Germany	324
France	306
Canada	261
...	...
Monaco	1
Fiji	1
Serbia and Montenegro	1
Turkey	1
Kenya	1

1.10 Какая страна выпускает самые качественные фильмы

Самый высокий рейтинг у Egypt, что возможно связано с высоким качеством фильмов и их малым количеством.

	Rating
Egypt	7.800
Libyan Arab Jamahiriya	7.400
Iran	7.375
Indonesia	7.300
Cameroon	7.300
...	...
Algeria	0.000
Kyrgyz Republic	0.000
Guadaloupe	0.000
Dominican Republic	0.000
Kenya	0.000

Посмотреть на реализацию можно по ссылке: <https://colab.research.google.com/drive/1nJDLp794DkXTnHiMQ6GQI3w78w5?usp=sharing>