

Имеется набор данных по киноиндустрии.

Мы выполнили предварительную обработку данных. Проанализируем их, посмотрим на динамику дохода в зависимости от жанра, рейтинга, затрат на кинокартину.

Итоговая визуализация - [Tableau](#)

Импортируем данные

Out[3]:

| | Unnamed: 0 | name | rating | genre | released | score | votes | director | writer | star | country | budget | gross | company | run |
|------|------------|----------------------------|--------|-----------|-----------------------------------|-------|----------|----------------|------------------|--------------------|----------------|------------|------------|--------------------------|-----|
| 4375 | 4375 | The Triplets of Belleville | PG-13 | Animation | February 13, 2004 (United States) | 7.8 | 51000.0 | Sylvain Chomet | Sylvain Chomet | Michèle Caucheteux | France | 9500000.0 | 14776760.0 | Les Armateurs | |
| 6020 | 6020 | A Good Old Fashioned Orgy | R | Comedy | October 27, 2011 (Russia) | 6.1 | 22000.0 | Alex Gregory | Alex Gregory | Jason Sudeikis | United States | NaN | 1378947.0 | Endgame Entertainment | |
| 6303 | 6303 | Stoker | R | Drama | March 1, 2013 (United Kingdom) | 6.8 | 105000.0 | Park Chan-Wook | Wentworth Miller | Mia Wasikowska | United Kingdom | 12000000.0 | 12077441.0 | Fox Searchlight Pictures | |

Для удобства чтения данных, преобразуем таблицу

Out[8]:

| | Название | Рейтинг фильма | Жанр | Рейтинг IMDb | Голоса | Кинорежисер | Режиссер | Звезда | Страна | Бюджет | Доход от фильма | Компания | Продолжительность | Г. выпуск |
|-----|---------------------------------|----------------|---------|--------------|--------|--------------------|------------------|-------------------|--------|--------|-----------------|-----------------------|-------------------|-----------|
| 748 | Fraternity Vacation | R | Комедия | 5.3 | 2200.0 | James Frawley | Lindsay Harrison | Stephen Geoffreys | США | <NA> | 3333306 | New World Pictures | 94.0 | 19 |
| 908 | The Boy Who Could Fly | PG | Драма | 6.5 | 5600.0 | Nick Castle | Nick Castle | Lucy Deakins | США | <NA> | 7177431 | Lorimar Film Partners | 114.0 | 19 |
| 421 | Yor, the Hunter from the Future | PG | Боевик | 4.4 | 2900.0 | Antonio Margheriti | Juan Zanolto | Reb Brown | Италия | <NA> | 2810199 | Diamant Film | 88.0 | 19 |

Таблица отформатирована для дальнейшей работы с ней, экспортируем данные для работы в Tableau

Выполнена визуализация в Tableau public.tableau.com

Посчитав корреляцию между столбцами ниже, посмотрим на диаграмму рассеяния в Tableau (График 1-й)

Корреляция

Out[10]:

| | Рейтинг IMDb | Голоса | Бюджет | Доход от фильма | Продолжительность |
|-------------------|--------------|----------|----------|-----------------|-------------------|
| Рейтинг IMDb | 1.000000 | 0.409182 | 0.076254 | 0.186258 | 0.399451 |
| Голоса | 0.409182 | 1.000000 | 0.442429 | 0.630757 | 0.309212 |
| Бюджет | 0.076254 | 0.442429 | 1.000000 | 0.740395 | 0.320447 |
| Доход от фильма | 0.186258 | 0.630757 | 0.740395 | 1.000000 | 0.245216 |
| Продолжительность | 0.399451 | 0.309212 | 0.320447 | 0.245216 | 1.000000 |

На графике рассеяния (в Tableau) наблюдается взаимосвязь между величинами. Так же имеется высокий коэффициент корреляции между доходом и бюджетом фильмов. Но это не показатель на причинно-следственную связь.

Посмотрим на пять самых популярных жанров и объем дохода по ним.

Out[11]:

Жанр
Боевик 243435855754
Комедия 97175468462
Анимационный 80142045796
Драма 57150648537
Приключение 45916596693
Name: Доход от фильма, dtype: Int64

3-й график в Tableau подтверждает вычисления выше

- по графику можно предположить – анимационный жанр приносит больший доход, чем комедия. Но по количеству значений всё таки выигрывает комедия.

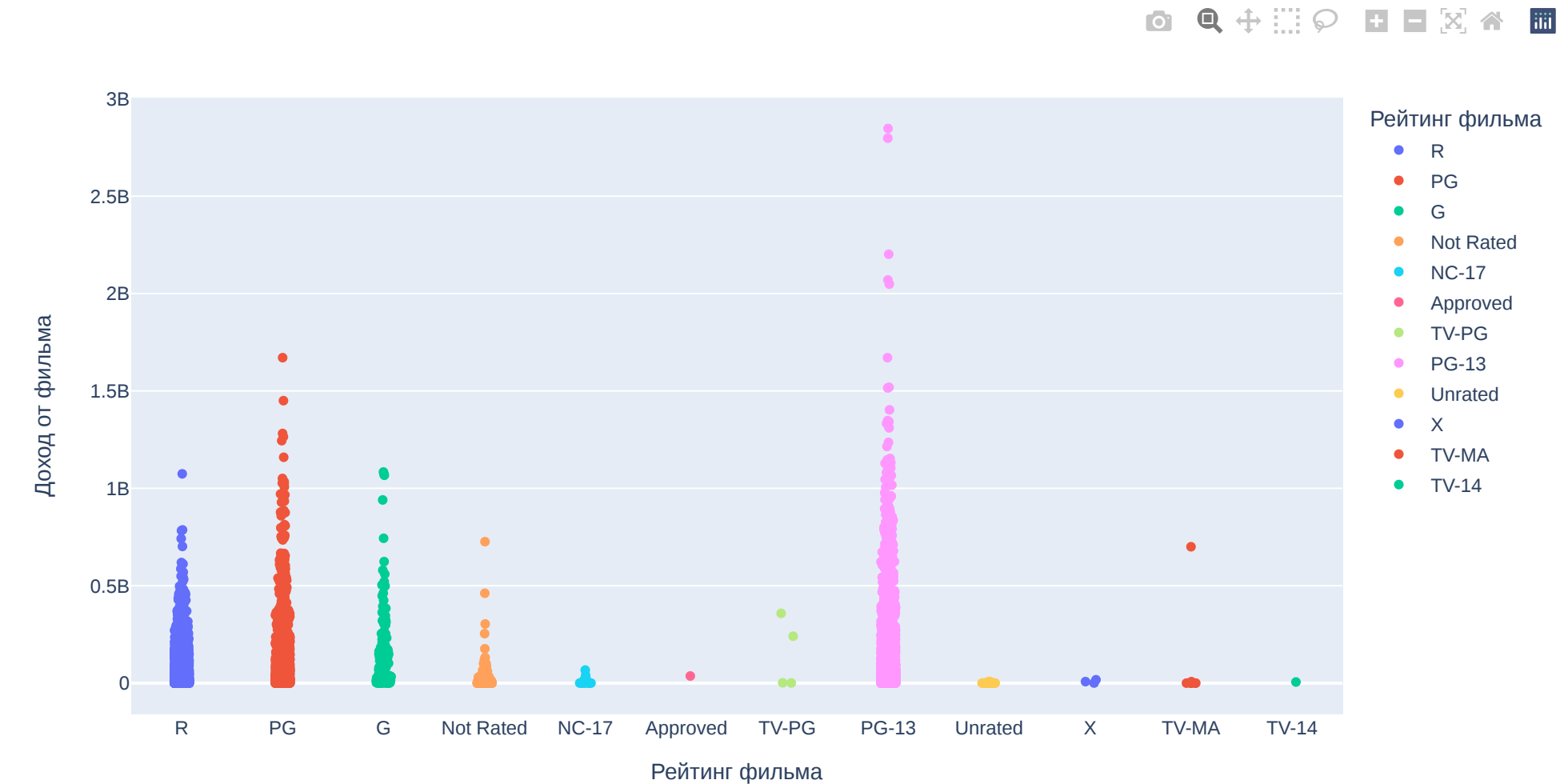
Количество анимационных фильмов - 338

Количество комедий - 2245

Далее посмотрим на динамику дохода по рейтингу

На 2-м графике в Tableau рассматривается доход по рейтингу, по версии IMDb.

И ниже доход по американской рейтинговой системе МРАА, то есть по нашим данным.



Необходимо ориентироваться на рейтинг IMDb начиная со значения 6 и заканчивая на значении 7.8

И по системе МРАА больший спрос имеют фильмы с рейтингом - R, PG, PG-13, G

- Рейтинг G — Нет возрастных ограничений
- Рейтинг PG — Рекомендуется присутствие родителей
- Рейтинг PG-13 — Детям до 13 лет просмотр не желателен
- Рейтинг R — Лицам до 17 лет обязательно присутствие взрослого