

Университет ИТМО

Практическая работа №4
по дисциплине «Визуализация и моделирование»

Автор: Ефимов Павел Леонидович

Поток: ВИМ 1.1

Группа: К3220

Факультет: ИКТ

Преподаватель: Чернышева А.В.

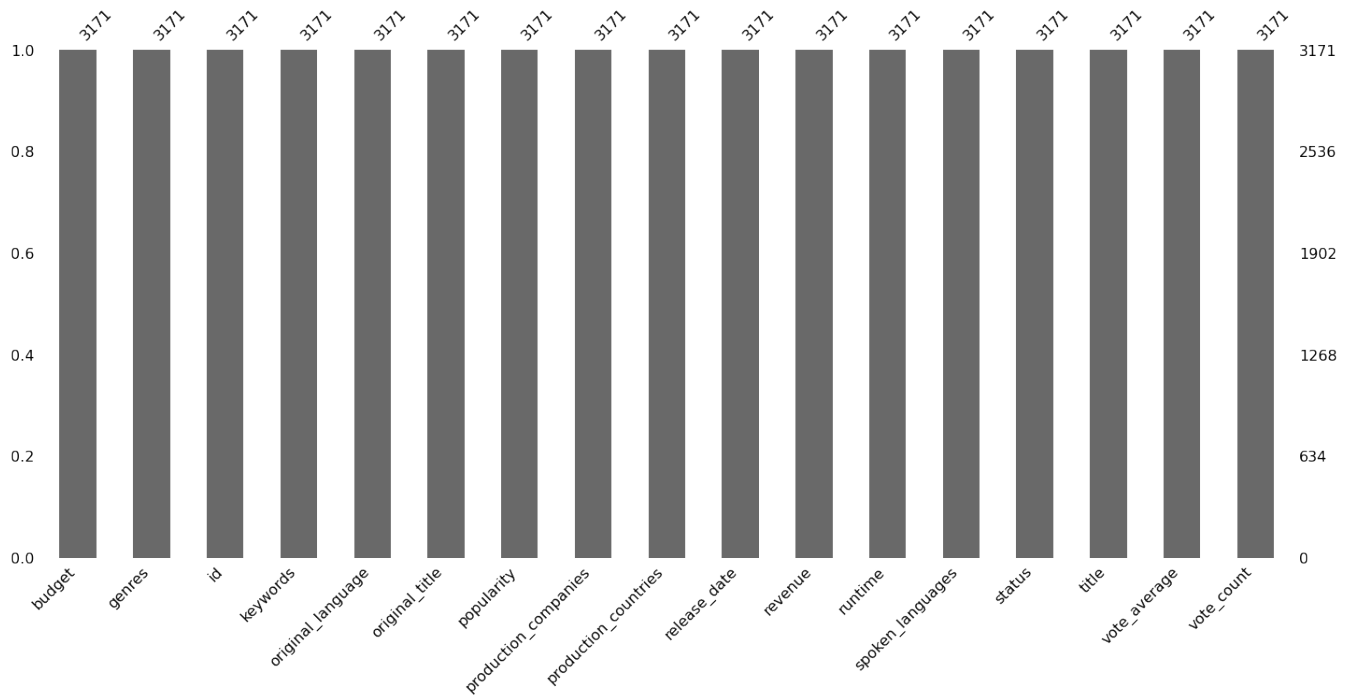
Санкт-Петербург, 2021 г.

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.

1 Проверка датасета

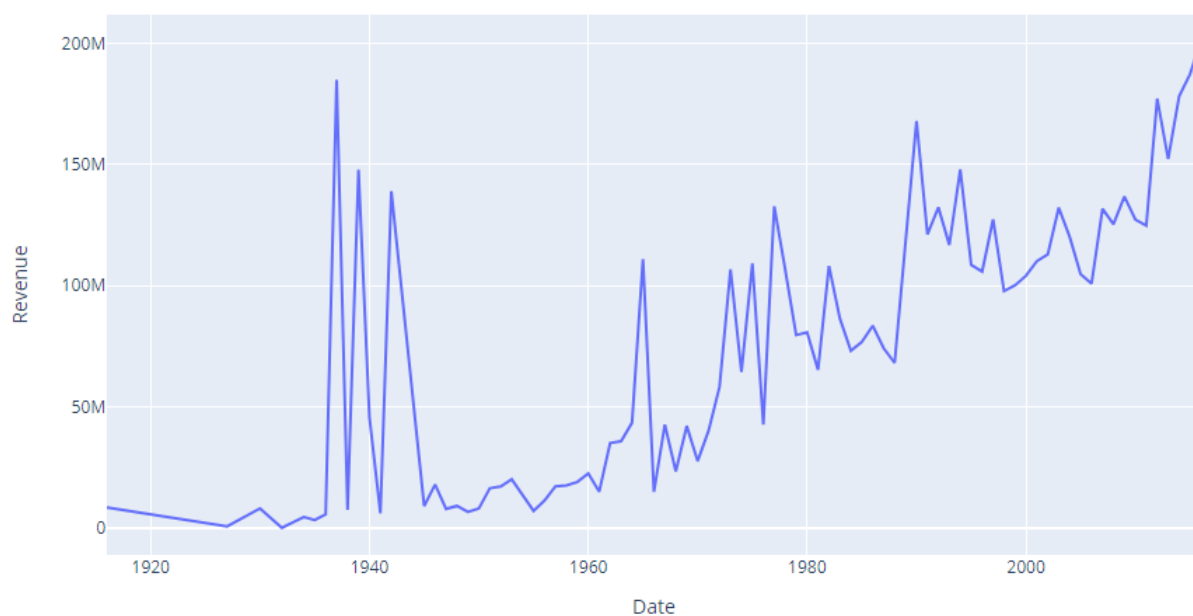
Как видно на графике, нет пустых данных в датасете



2 Гипотезы

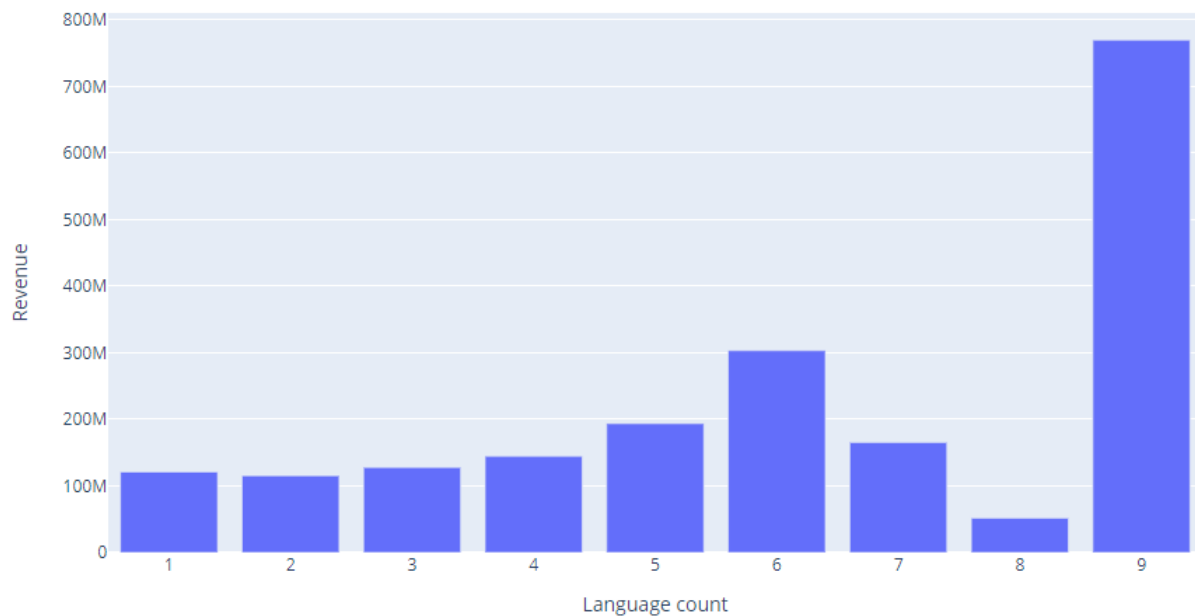
2.1 Фильмы вышедшие позже, зарабатывают больше

В целом наблюдается рост дохода у фильмов вышедших позже, исключением являются 35-43 года.



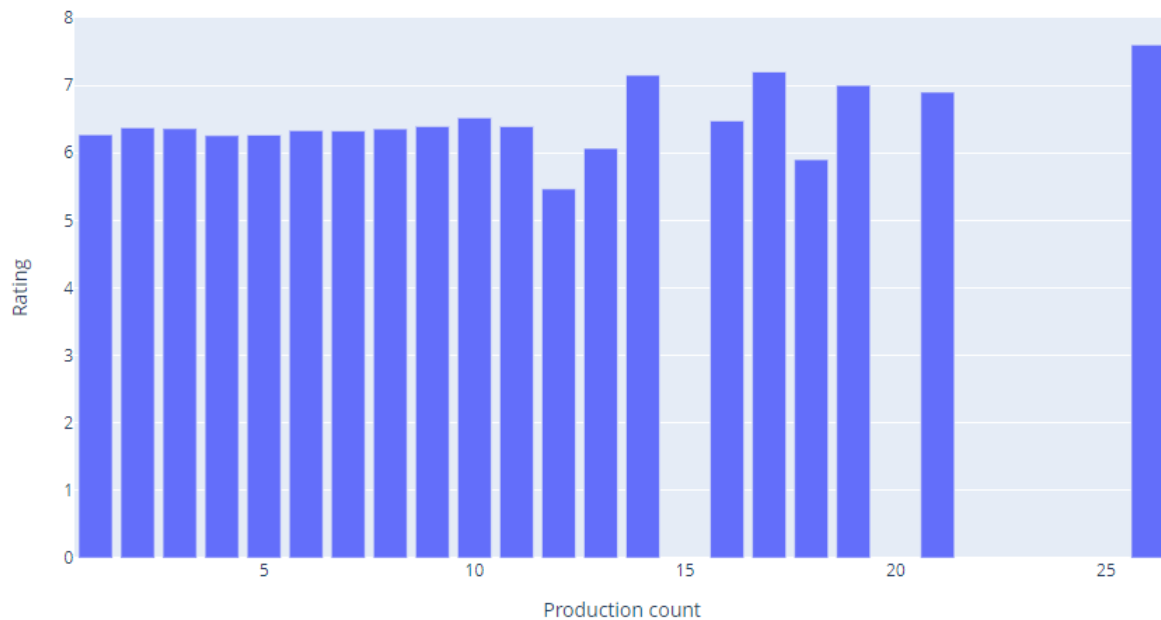
2.2 Если фильм переведен на большее количество языков, тем больше он заработал

Фильмы переведенные на 6 и 9 языков заработали больше всего, но нельзя утверждать гипотезу, т.к. фильмы переведенные на 7 и 8 языков заработали меньше



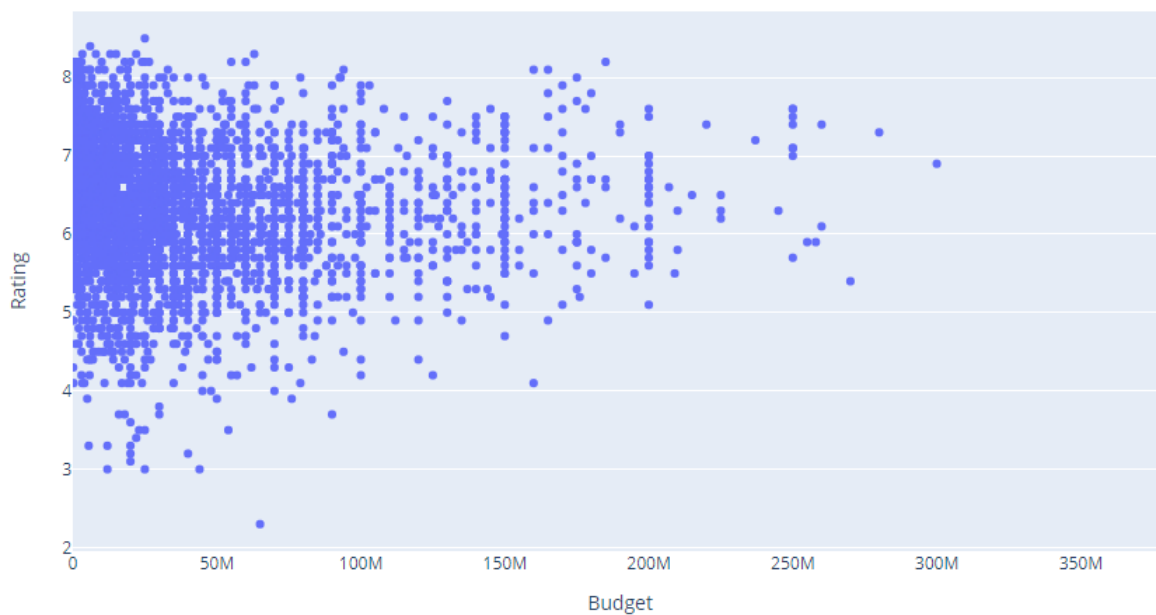
2.3 Чем больше компаний участвовало в процессе съемки фильма, тем больше его рейтинг

Нет никакой зависимости между количеством компаний участвовавших в съемке фильма и его рейтингом



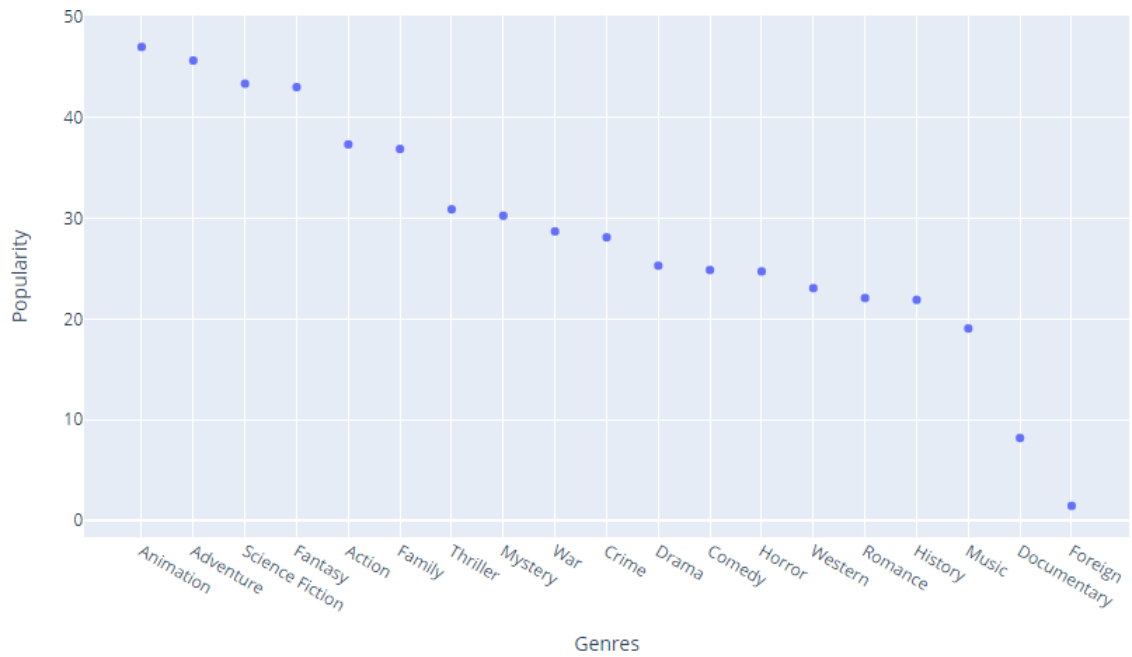
2.4 Большой бюджет фильма обеспечит большой рейтинг

Рейтинг фильма зависит только по нижней ступени фильма, чем больше бюджет, тем меньше шансов, что фильма получит низкую оценку



2.5 Самый популярный жанр это Science Fiction

Как видно на графике, самый популярный жанр Animation, а Science Fiction занимает лишь 3 место



Посмотреть на реализацию можно по ссылке: <https://colab.research.google.com>