

АНАЛИЗ ДАННЫХ 1000 **NETFLIX** SHOWS

Анализ выполнен:
Есаков **Никита Сергеевич**
Шиндерук **Виктория Игоревна**
Бадмаева **Алия Азатовна**

ИССЛЕДОВАНИЕ ДАННЫХ НА ДУБЛИКАТЫ

Подборка 5 шоу с максимальным количеством дублей

Большинство полных дубликатов - шоу, относящиеся к 2017 году

13 Reasons Why

2017 - год выхода первого сезона сериала

500

Количество полных
дубликатов / уникальных
значений в данных

Girlboss

2017 - год выхода сериала

Prison Break

2017 - год выхода пятого сезона сериала

New Girl

2017 - год выхода шестого сезона сериала

The Vampire Diaries

2017 - год выхода восьмого сезона сериала



Количество дубликатов



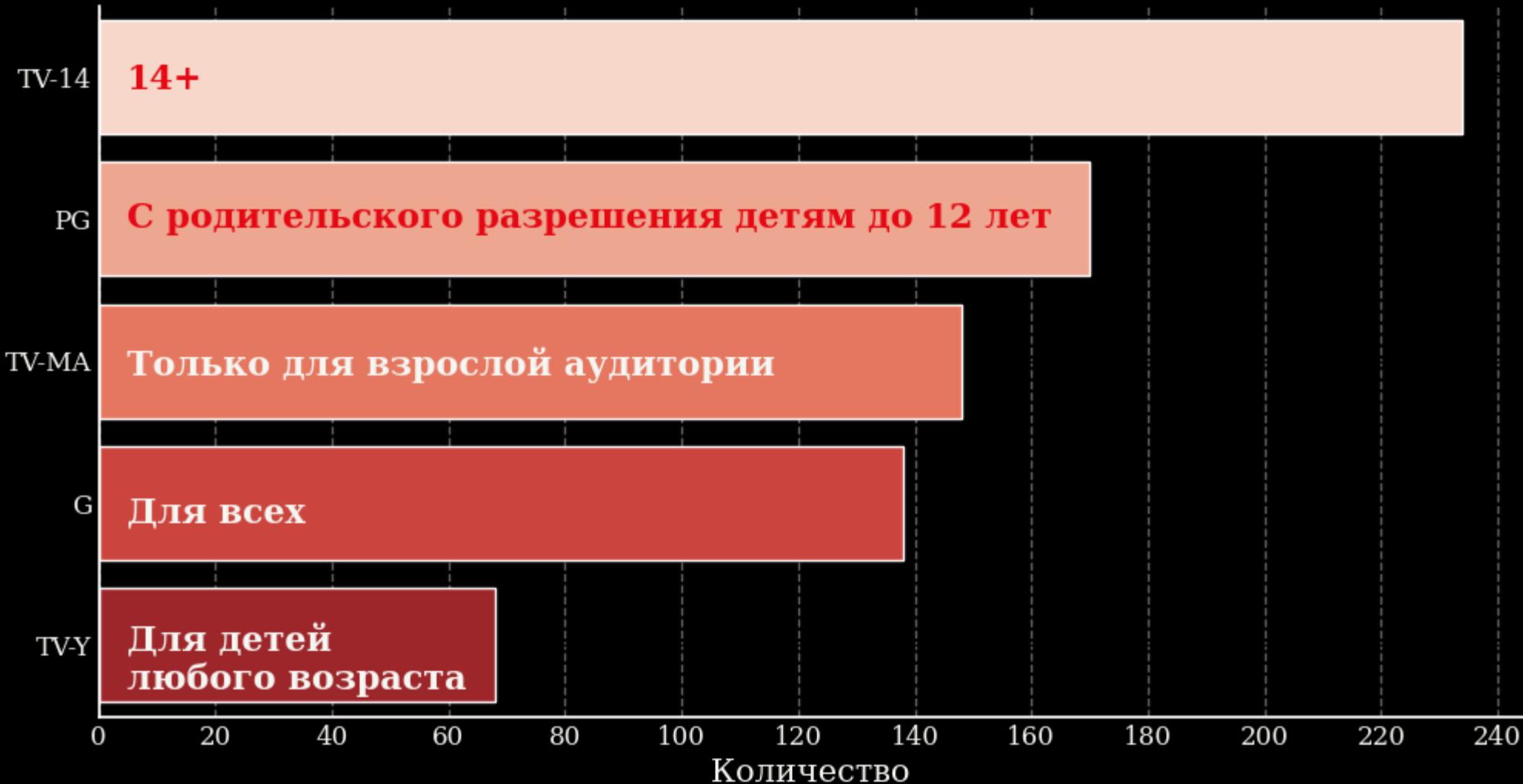
ЧАСТО ВСТРЕЧАЮЩИЕСЯ НАЗВАНИЯ ШОУ

A close-up portrait of a young man with short brown hair, smiling broadly. He is wearing a light blue and white plaid shirt. The background is dark.

Chase Willden

ИССЛЕДОВАНИЕ ДАННЫХ НА ПОВТОРЫ

Шоу по возрастным категориям с дублями

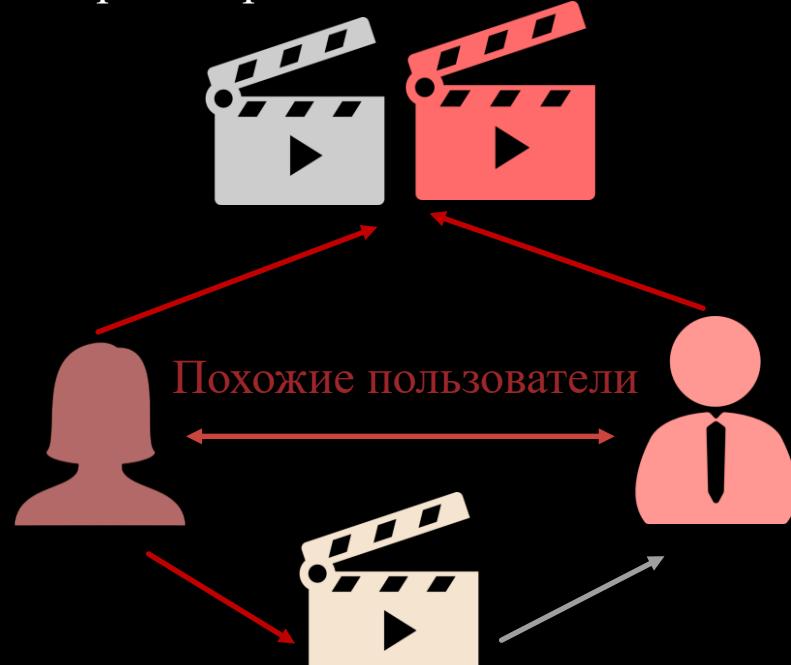


Группы с большим количеством шоу совпадают с группами изначальных шоу, используемых для формирования данных: TV-14, PG, TV-MA, G

АЛГОРИТМЫ РЕКОМЕНДАЦИЙ NETFLIX

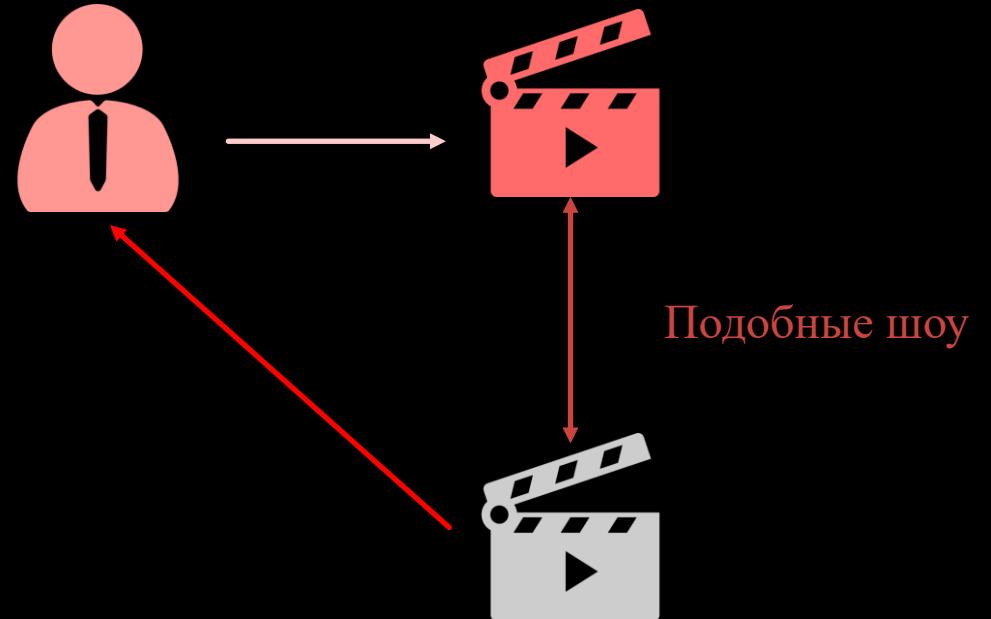
Совместная (коллаборативная) фильтрация

Просмотрено обоими пользователями



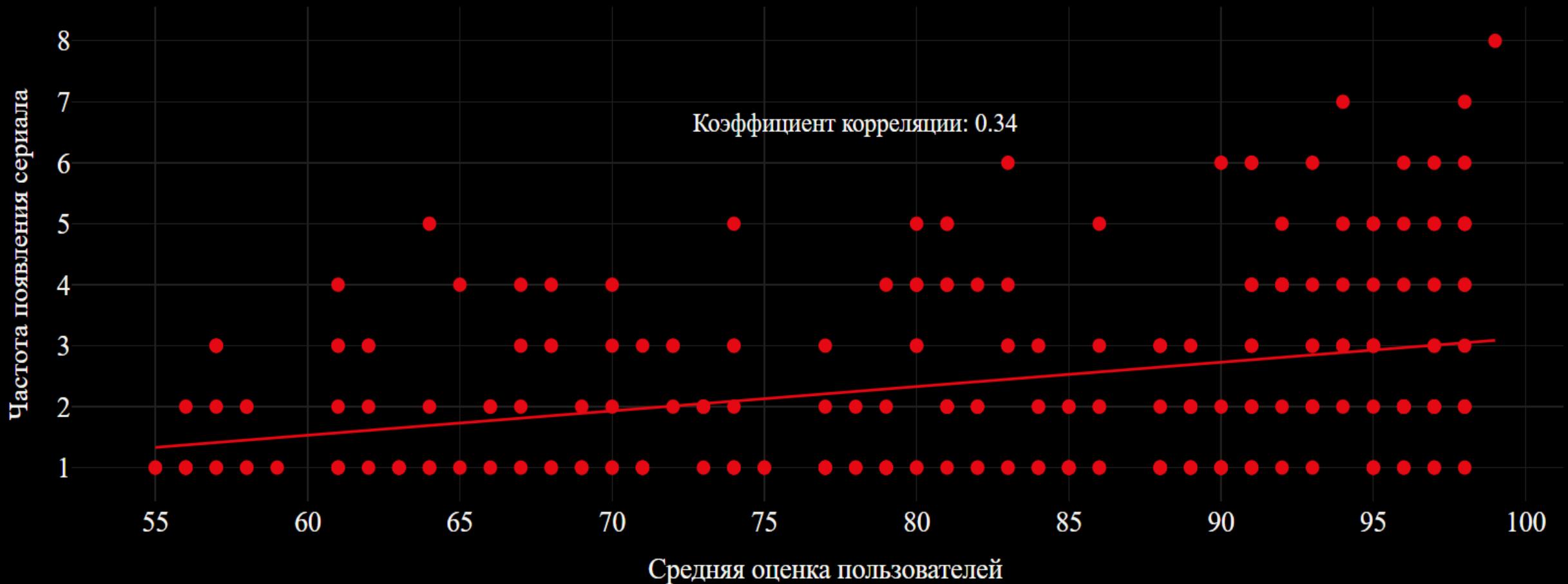
Фильтрация, основанная на контенте

Просмотренные зрителем



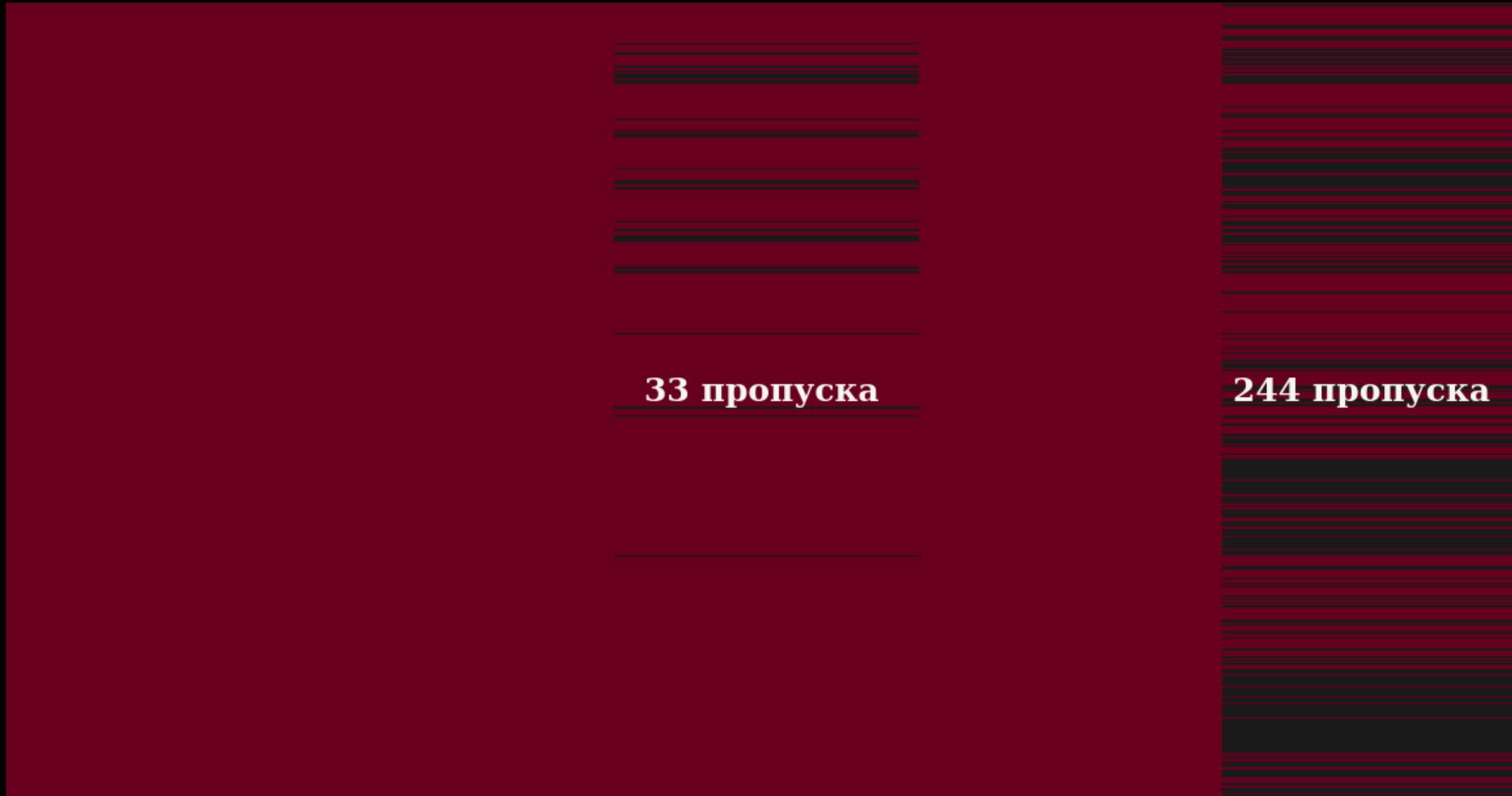
ИССЛЕДОВАНИЕ НА ЗАВИСИМОСТЬ ПРИЗНАКОВ

Зависимость частоты появления сериала от его средней оценки с трендовой линией



АНАЛИЗ ДАННЫХ НА ПРОПУСКИ ПОСЛЕ ОЧИСТКИ ОТ ДУБЛЕЙ

Пропуски по признакам после удаления дубликатов



Большинство пропусков в колонке ‘user rating score’ – оценка зрителя

title

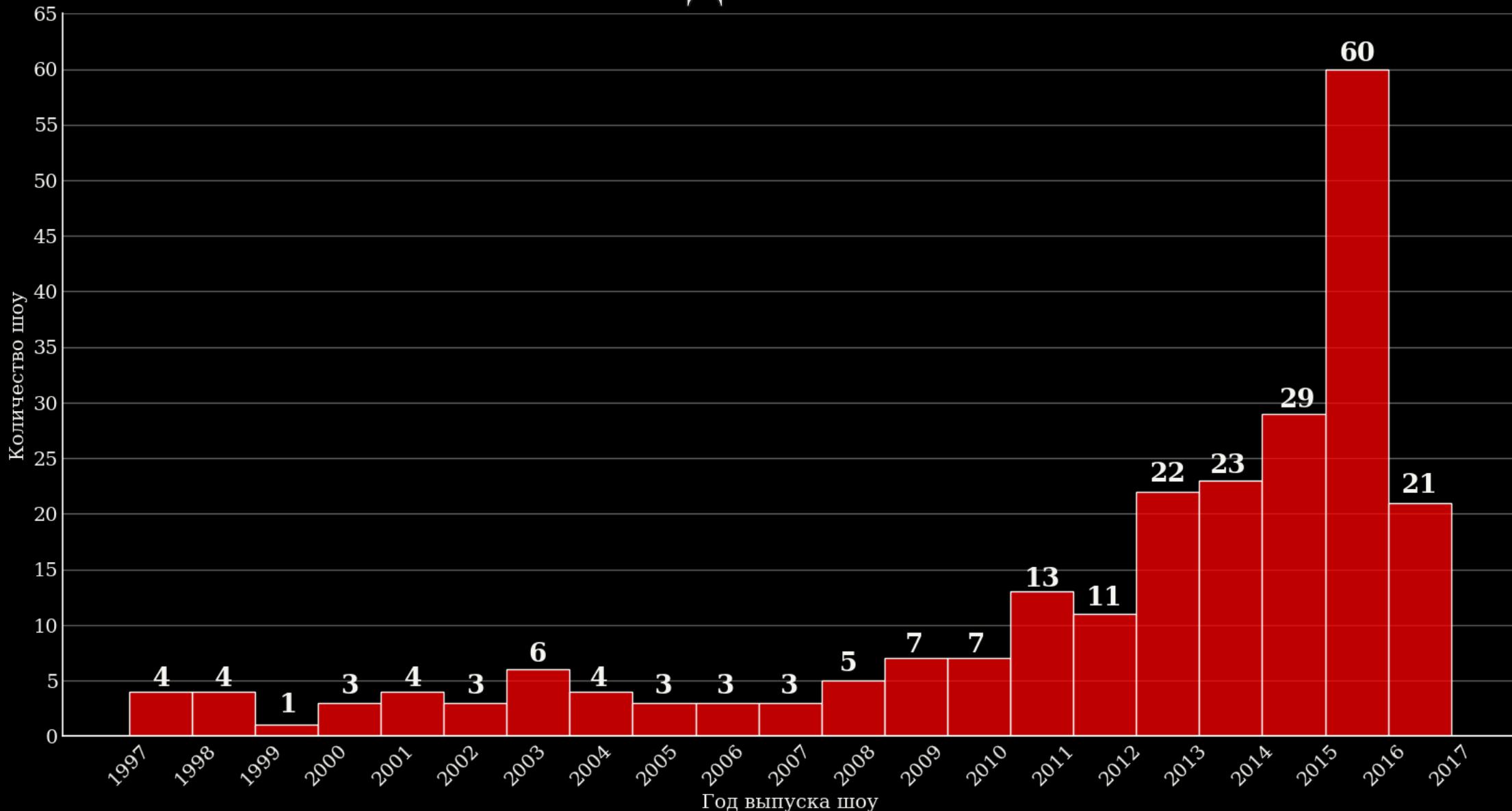
rating

ratingLevel
Признаки

release year

user rating score

ЧАСТОТА ПОЯВЛЕНИЯ ПРОПУСКОВ В ЗАВИСИМОСТИ ОТ ГОДА ВЫПУСКА ШОУ

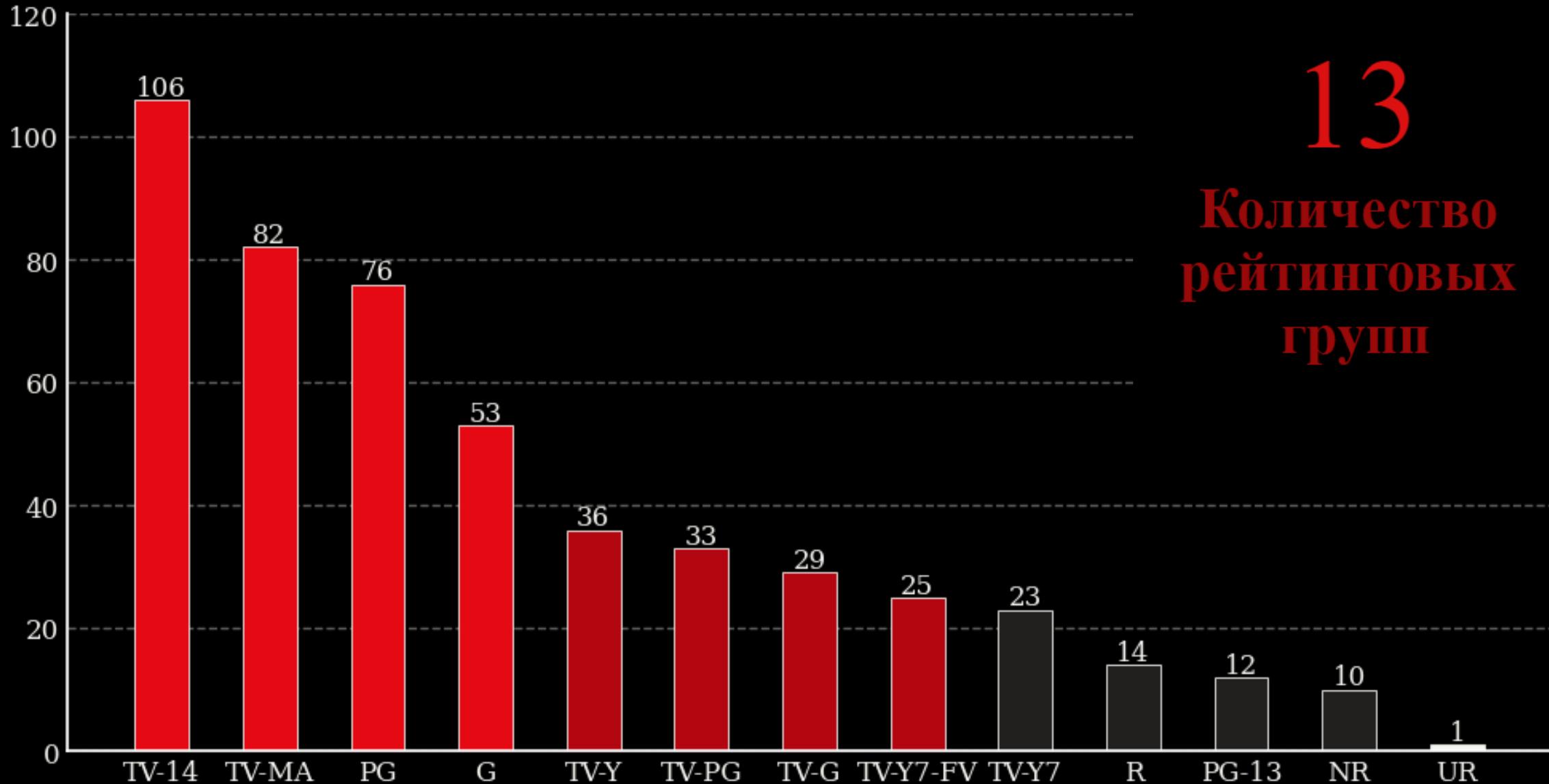


КОЛИЧЕСТВО ВЫПУСКОВ ШОУ ПО ГОДАМ



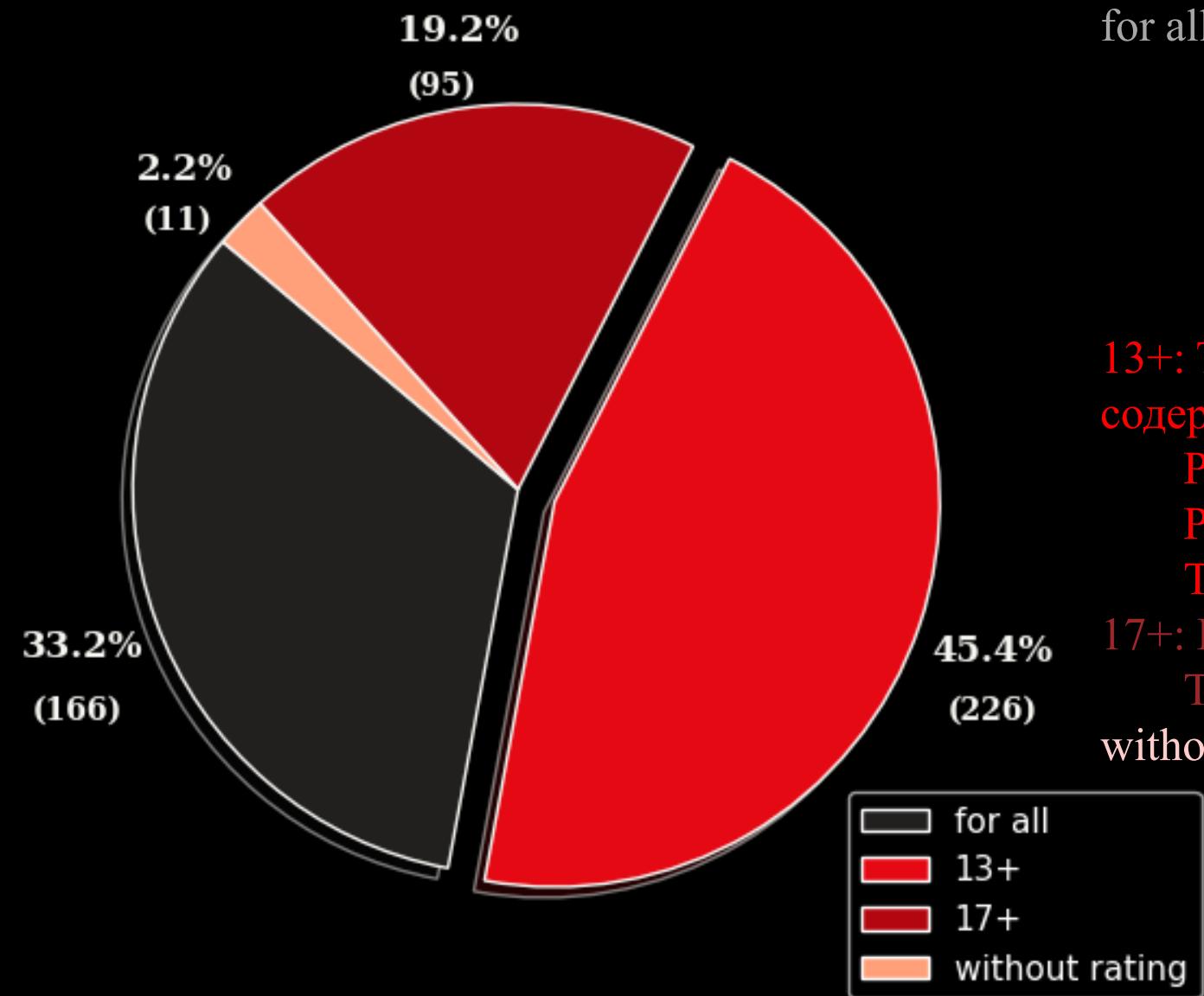
Количество шоу по возрастной категории (рейтингу)

Самое большое количество шоу с возрастной категорией 14+ (TV-14)



13
Количество
рейтинговых
групп

КОЛИЧЕСТВО ШОУ ПО ВОЗРАСТАМ



for all: G - для всех возрастов

TV-G - для всех возрастов, включая детей

TV-Y7- 7 +

TV-Y7-FV - для 7+, возможны сцены причинения вреда фантастическим персонажам

TV-Y - для дошкольного возраста

13+: TV-PG - родители должны следить за содержимым.

PG – содержит материал, не подходящий для детей

PG-13 - 13+

TV-14 - 14+

17+: R - 17+, просмотр со взрослым

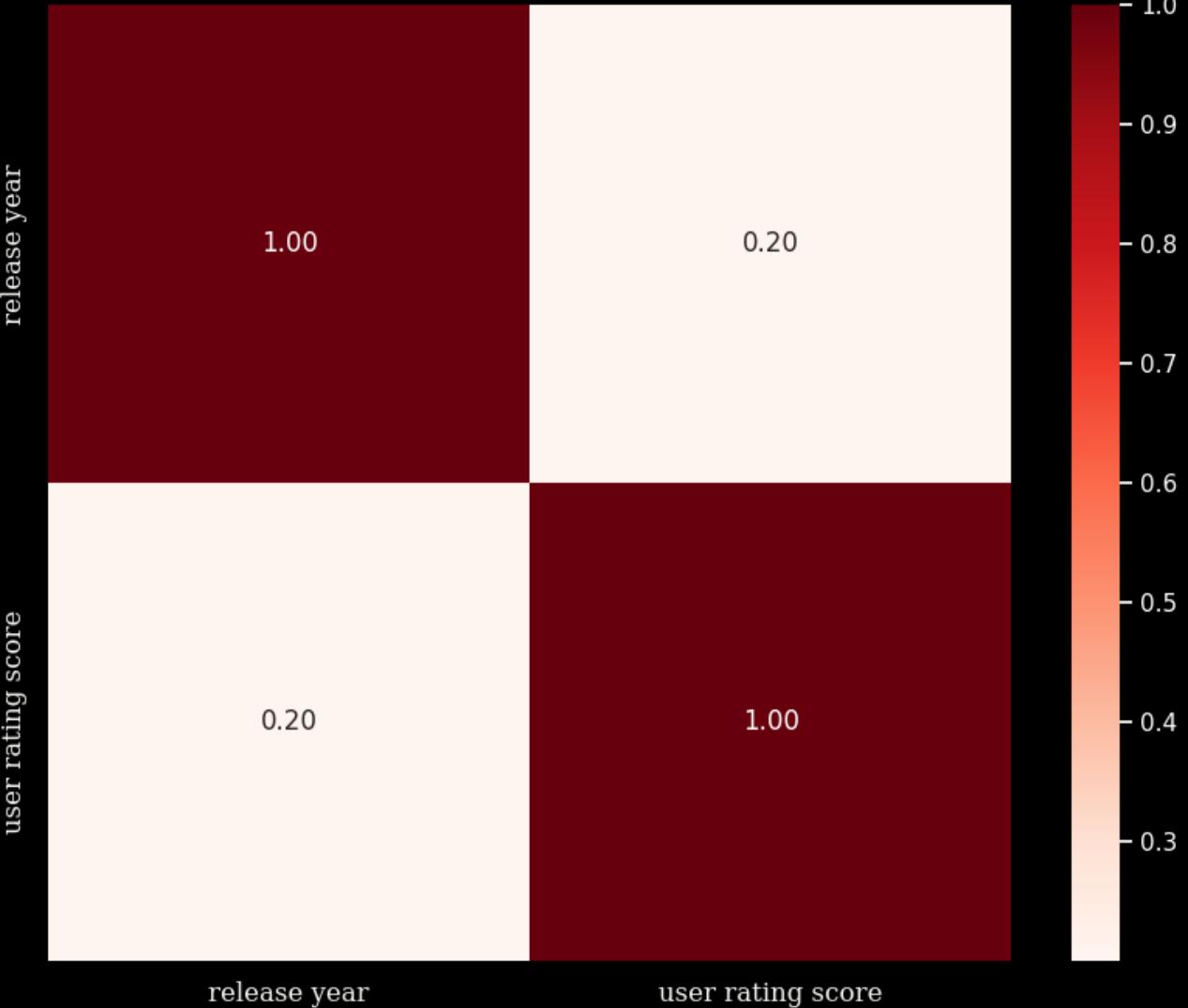
TV-MA - только для взрослых

without rating: NR – без рейтинга

UR – шоу без официальной оценки

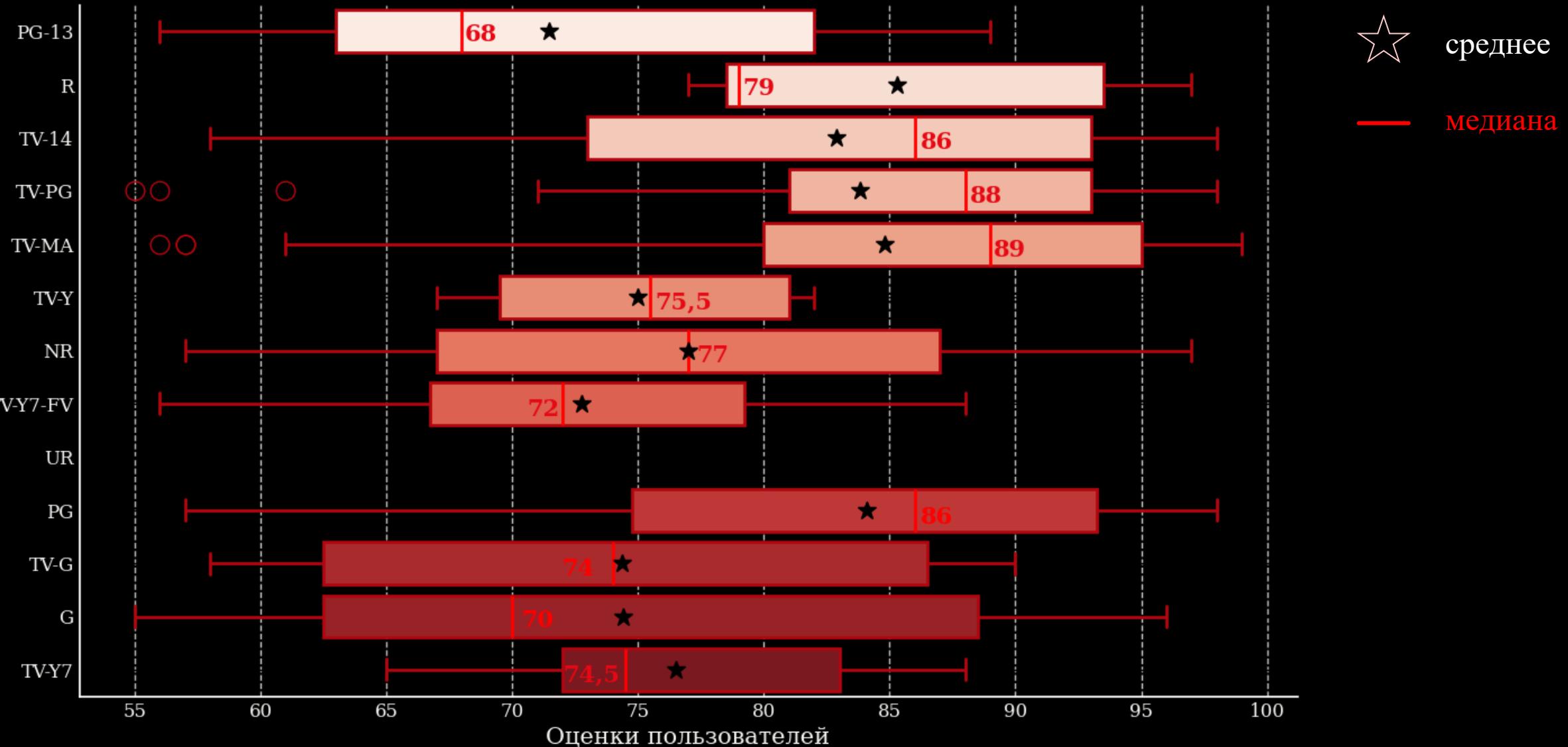
ЗАВИСИМОСТЬ ПРИЗНАКОВ

**Матрица корреляции года выпуска
от оценок пользователей**



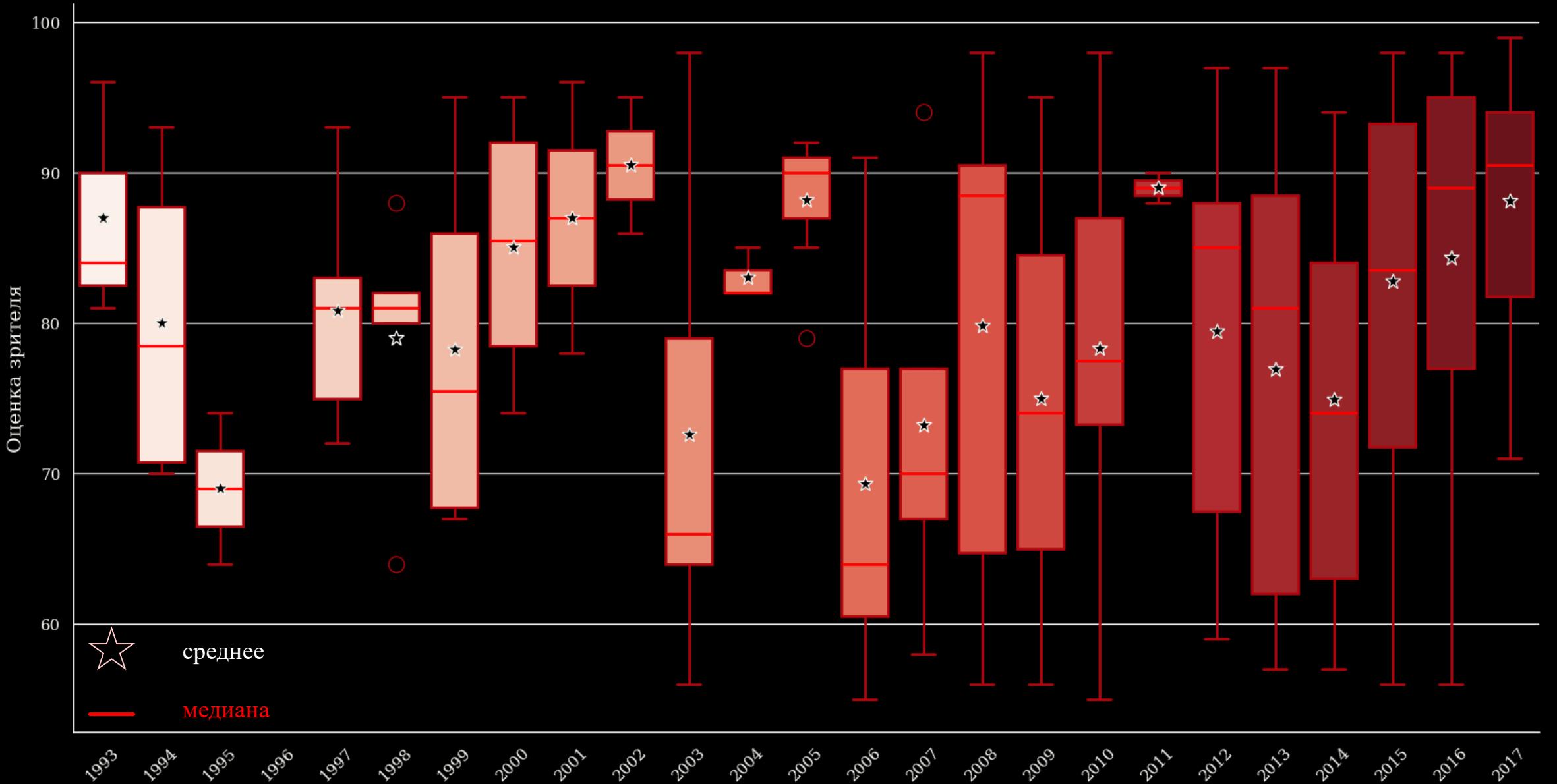
На графике видно, что для числовых переменных – год выпуска и оценка зрителя корреляция слабая - близка к нулю.

ИНФОРМАЦИЯ О РАСПРЕДЕЛЕНИИ ОЦЕНОК ПОЛЬЗОВАТЕЛЕЙ ПО ВОЗРАСТНЫМ КАТЕГОРИЯМ

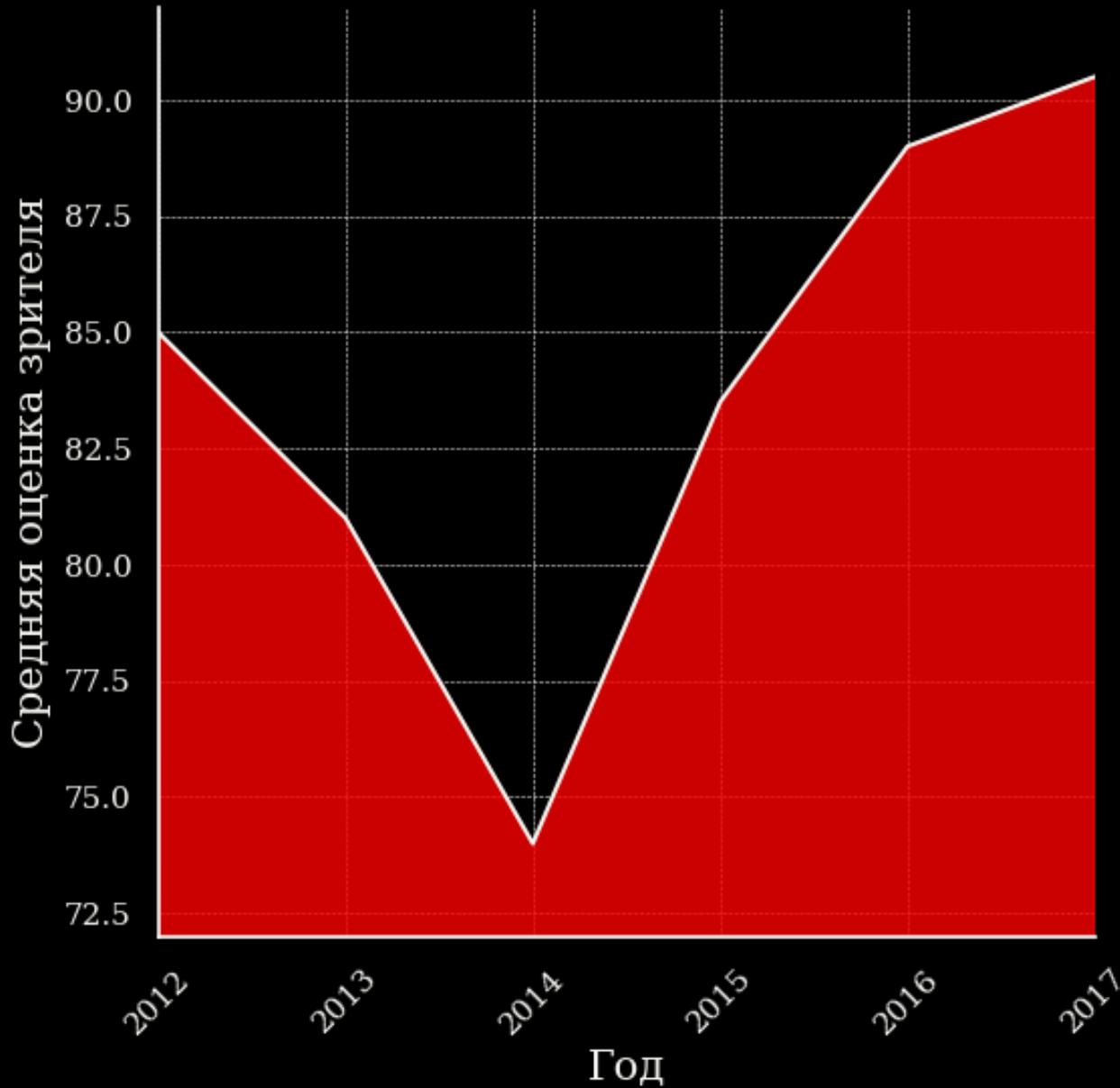




ОЦЕНКИ ЗРИТЕЛЕЙ ПО ГОДУ



УСЕЧЕННЫЙ АНАЛИЗ ОЦЕНКИ ЗРИТЕЛЕЙ



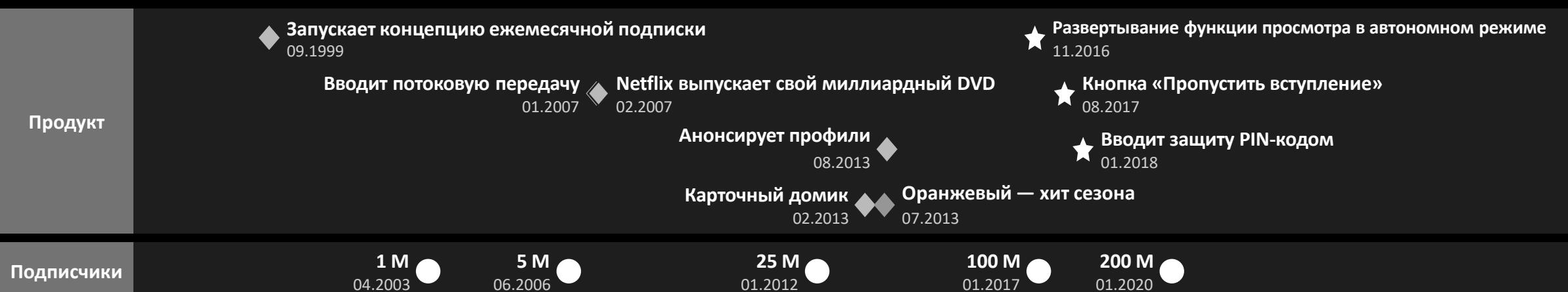
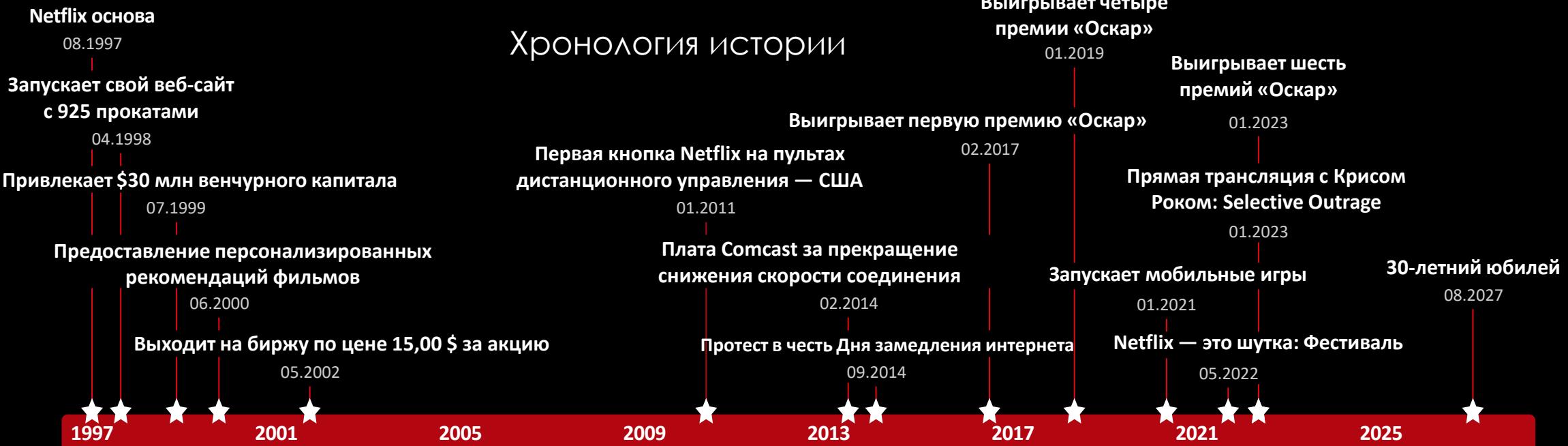
Исследуя пользовательские оценки по годам, была обнаружена просадка в 2014 году.

Предполагаемые причины просадки:

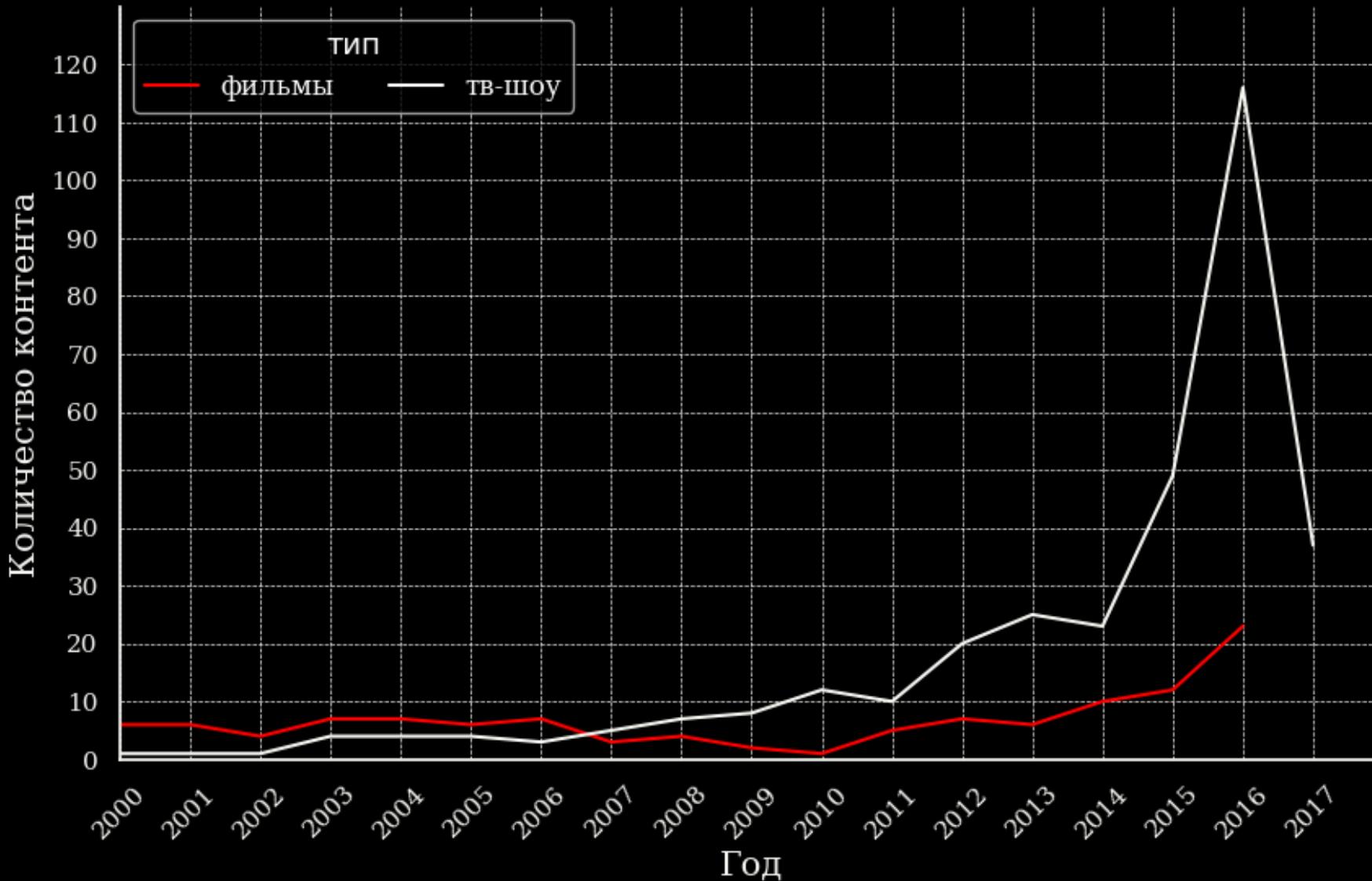
- Мировой кризис 2014 и, как следствие, нестабильность акций компаний
- Популярность других киносервисов (развитие конкурентов)

Несмотря на это, после 2014 наблюдается рост, возможно, связанный с удачной ставкой на производство оригинального контента.

NETFLIX



КОЛИЧЕСТВО ВЫПУСКАЕМЫХ ШОУ И ФИЛЬМОВ ПО ГОДАМ



В рамках проведения разведочного анализа данных было решено создать новый признак - разделение на ТВ-шоу и киноленты

ИСТОЧНИКИ НОВЫХ ПРИЗНАКОВ ДЛЯ
ОБОГАЩЕНИЯ СУЩЕСТВУЮЩИХ ДАННЫХ

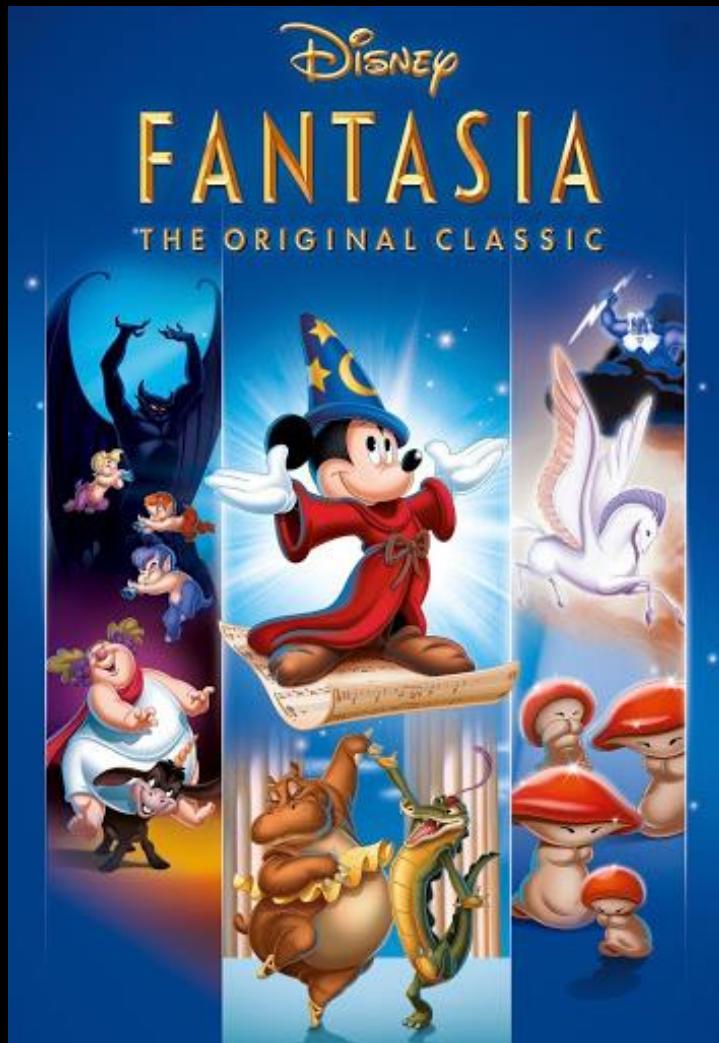
kaggle GitHub



data.world

В ДАННЫХ ПО ГОДУ ВЫПУСКА КАРТИН НЕТ ВЫБРОСОВ

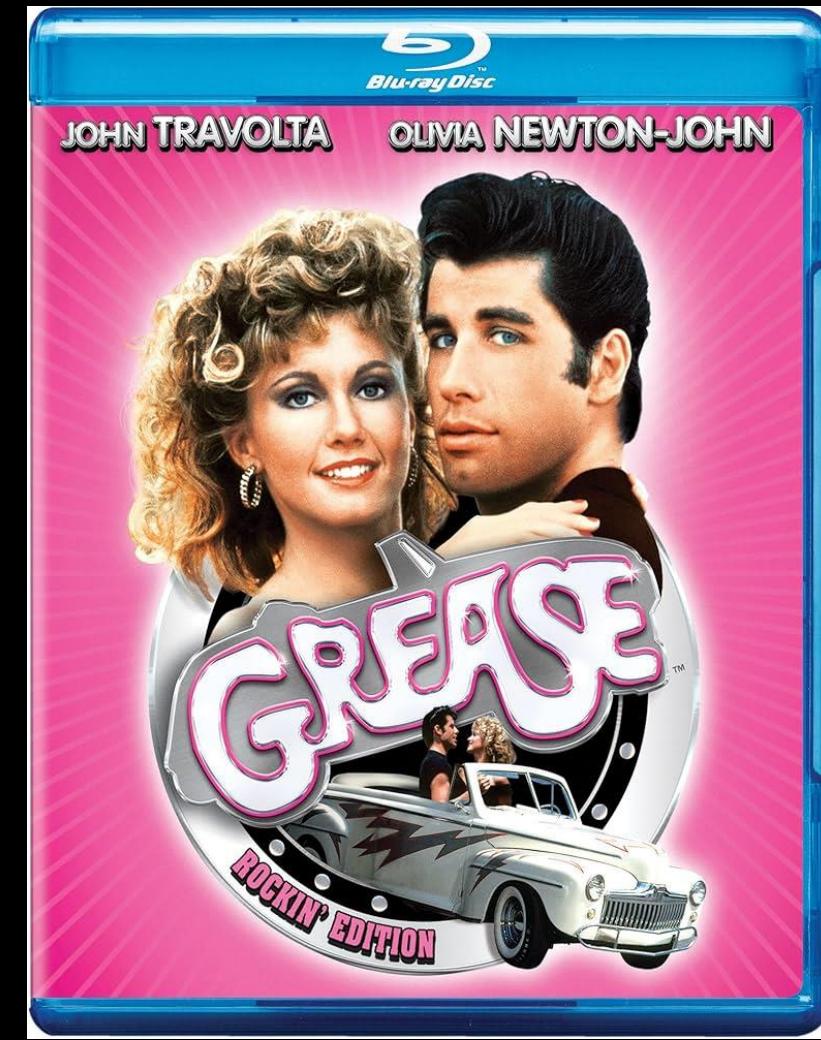
1940 G



1976 G



1978 PG





ЧАСТО ВСТРЕЧАЕМЫЕ В ДАННЫХ АКТЕРЫ



ЧАСТО ВСТРЕЧАЕМЫЕ КИНОКОМПАНИИ

Кинокомпания





ЧАСТО ВСТРЕЧАЕМЫЕ ДИРЕКТОРА

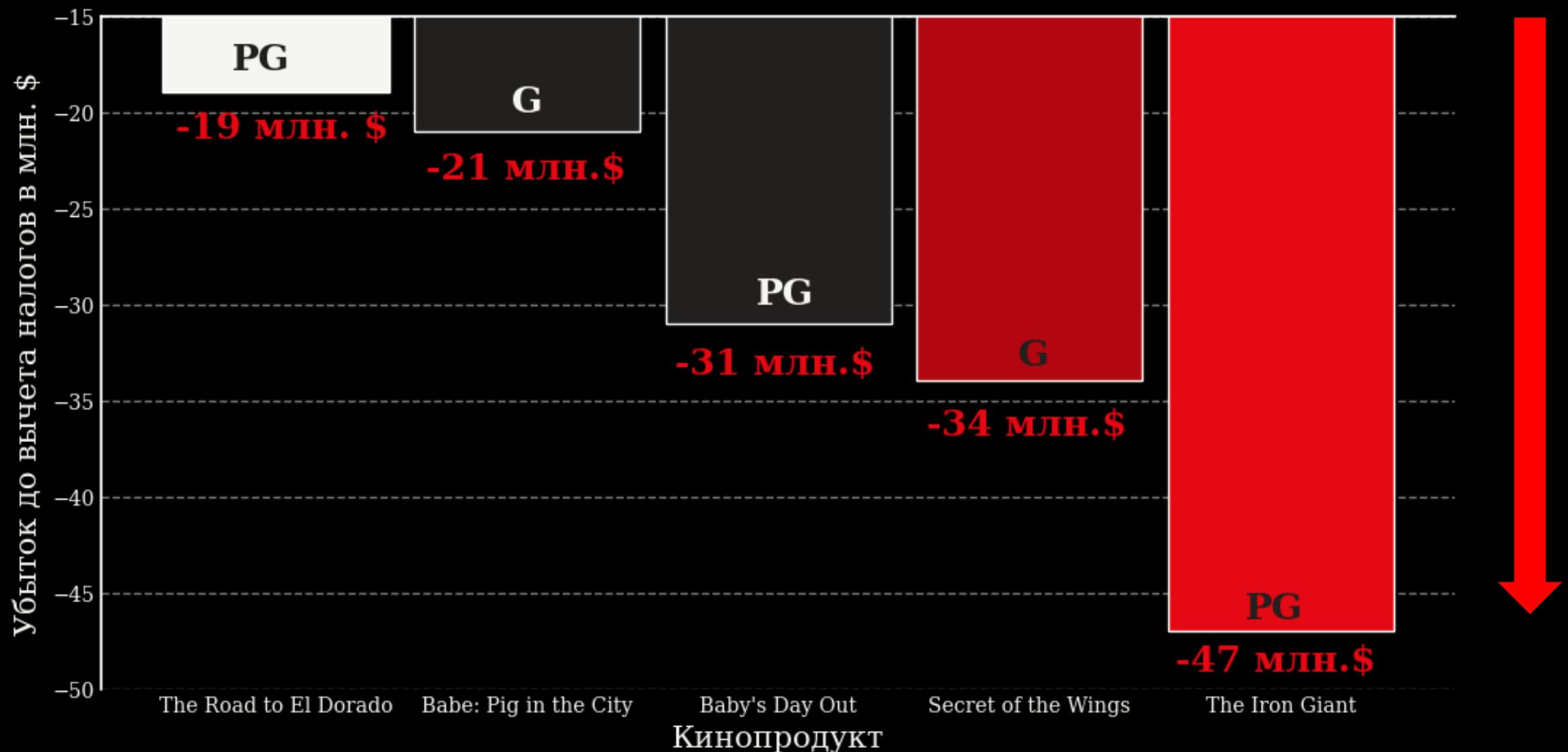




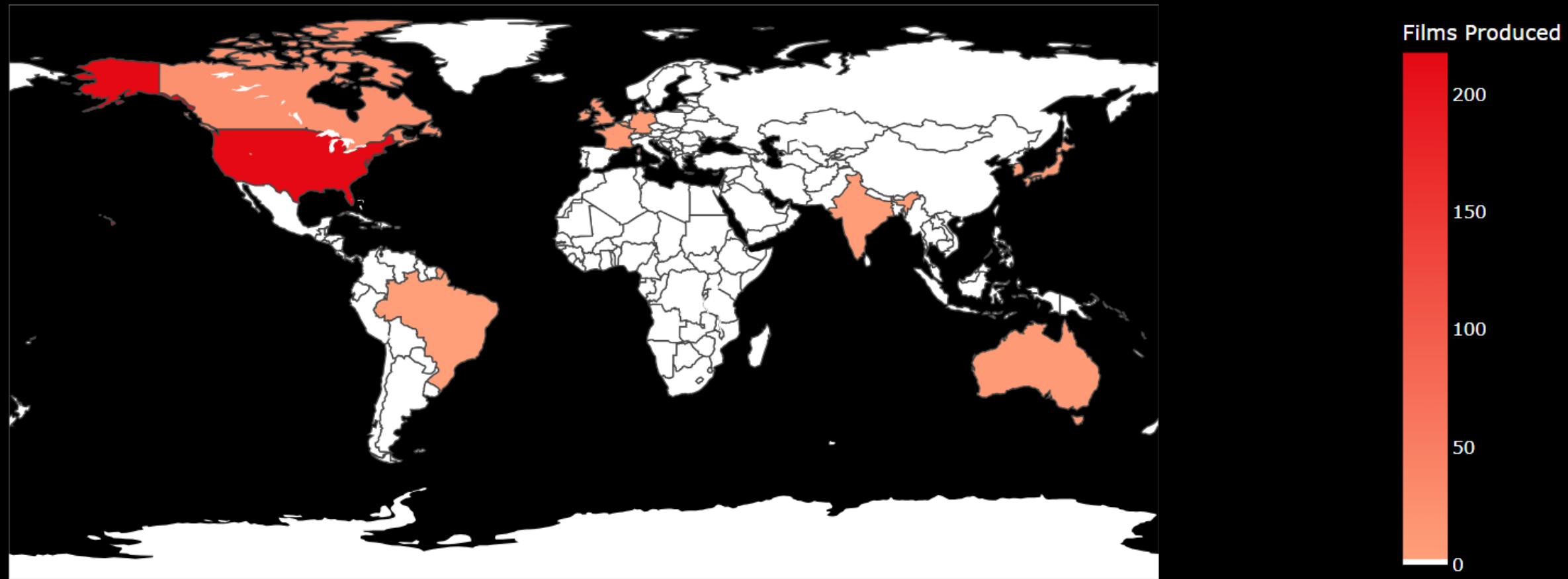
ТОП-5 ПРИБЫЛЬНЫХ ФИЛЬМОВ



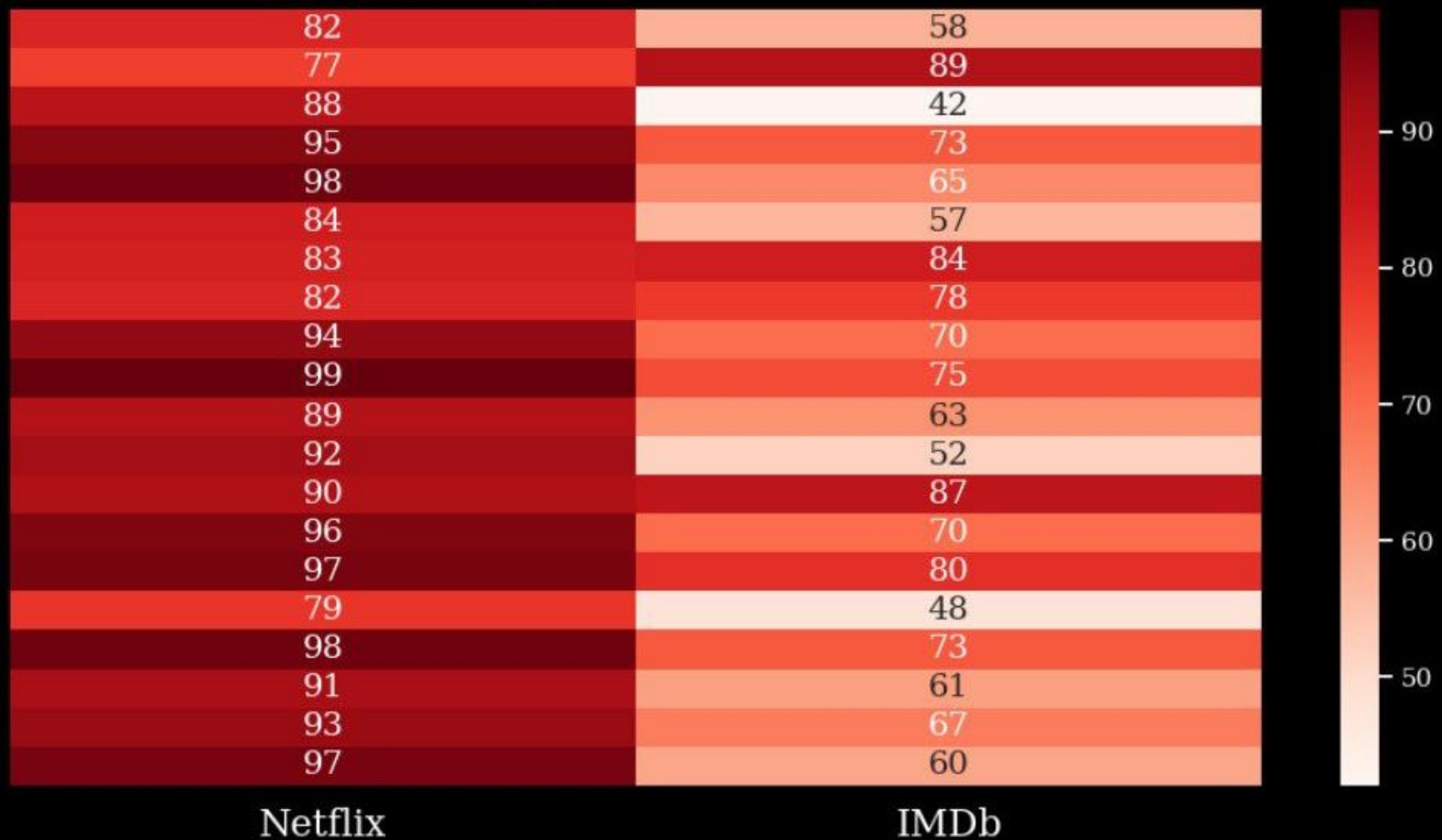
ТОП-5 УБЫТОЧНЫХ ФИЛЬМОВ



КАРТА СТРАН ВЫПУСКА КИНОПРОДУКТОВ



СРАВНЕНИЕ ОЦЕНОК ПО ШКАЛАМ NETFLIX И IMDB



ВЫВОДЫ

- 1. В ДАННЫХ МНОГО ДУБЛЕЙ, ЧТО СВЯЗАНО С АЛГОРИТМОМ РЕКОМЕНДАЦИЙ НЕТФЛИКС.
- 2. БОЛЬШЕ ВСЕГО РЕКОМЕНДАЦИЙ ПО ВОЗРАСТНЫМ КАТЕГОРИЯМ ПРЕДСТАВЛЕНО ИСХОДЯ ИЗ ВКУСА АВТОРА
- 3. В РЕКОМЕНДАЦИЯХ БОЛЬШЕ ВСЕГО ШОУ 2016 ГОДА.
- 4. АВТОРУ ИНТЕРЕСНЫ ШОУ КОМПАНИИ ДИСНЕЙ.
- 5. В ПОДБОРКЕ ЕСТЬ ЧАСТО ВСТРЕЧАЕМЫЕ АКТЕРЫ.
- 6. АВТОРЫ ДАННЫХ НРАВЯТСЯ КАК УБЫТОЧНЫЕ, ТАК И ПРИБЫЛЬНЫЕ ШОУ.
- 7. БОЛЬШЕ ВСЕГО ШОУ ПРОИЗВЕДЕНЫ В США
- 8. В НАБОРЕ ДАННЫХ ЕСТЬ КАК ФИЛЬМЫ, ТАК И ТВ-ШОУ
- 9. ЭТОТ ДАТАСЕТ ЯВЛЯЕТСЯ ИНДИВИДУАЛЬНОЙ ПОДБОРКОЙ ПО ВКУСАМ ЕГО АВТОРА, ПОЭТОМУ НЕ МОЖЕТ ЯВЛЯТЬСЯ РЕКОМЕНДАЦИЕЙ ДЛЯ ВЫПУСКА НОВЫХ ШОУ
- 10. БЛАГОДАРЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЕ NETFLIX, ПРОДУКТЫ НА ПЛАТФОРМЕ РЕКОМЕНДУЮТСЯ ИМЕННО ЦЕЛЕВОЙ ГРУППЕ ПОЛЬЗОВАТЕЛЕЙ, БЛАГОДАРЯ ЧЕМУ ОЦЕНКА ПРОДУКТА ВЫШЕ, ЧЕМ ПО РЕЙТИНГУ IMDB