

Университет ИТМО

Практическая работа №4
по дисциплине «Визуализация и моделирование»

Автор: Севумян Наталья Ренатовна

Поток: ВИМ 1.1

Группа: К3221

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Описательная статистика по нормализованным данным

Гипотезы из предыдущей практической работы:

1. Зависимость рейтинга от количества проголосовавших - экспоненциальная.
2. Фильмы в жанре анимация более популярны, поэтому у сериалов и меньше общий рейтинг, потому как анимационных сериалов меньше.
3. Фильмы, добавленные на платформу гораздо позже их создания, имеют меньший рейтинг.
4. Контент с одним из жанров Family имеет высокие оценки.
5. Чем больше фильмов или сериалов снял определенный режиссер, тем популярнее контент.

Проверка гипотез

Гипотеза 1

Для проверки этой гипотезы была использована библиотека Plotly. Построен график зависимости рейтинга согласно IMDb от количества проголосовавших (рисунок 1).

Рост рейтинга от количества проголосовавших

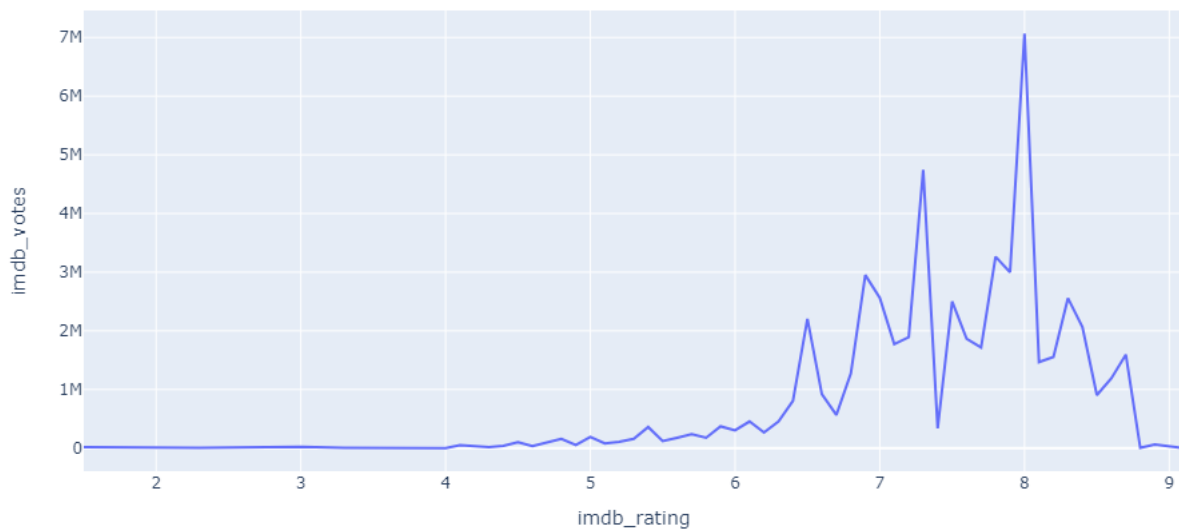


Рис. 1: Зависимость рейтинга от количества проголосовавших

Рейтинг 7.4 имеют всего 27 фильмов:

"Alice in Wonderland"(Алиса в стране чудес)

"Beezy Bear"

"Chip an' Dale"(Чип и Дейл)

"Corn Chips"(Кукурузные чипсы)

"Deep Blue"(Глубина)
"Raven's Home"(Дом Рэйвен)
'Bears',
"Disneynature: Wings of Life"(Скрытая красота: История любви, которая питает Землю)
"Diving with Dolphins"
"Don's Fountain of Youth"(Фонтан молодости Дона)
"Easter Island Underworld"(Под островом Пасхи)
"Float"
"Jungle Cat"(Кот джунглей)
"The New Yoda Chronicles: Clash of the Skywalkers"(Новые хроники Йоды: Тактика Скайуокера)
"MARS: Inside SpaceX"
"Marvel 75 Years: From Pulp to Pop!"(Документальный фильм к 75-летию Marvel)
"Guardians of the Galaxy"(Стражи галактики)
"Phineas and Ferb the Movie: Across the 2nd Dimension"(Финес и Ферб: Покорение 2-го измерения)
"Pinocchio"(Пиноккио)
"Pollyanna"(Поллианна)
"Queen of Katwe"(Королева Катве)
"Santa's Workshop"(Мастерская Санта Клауса)
"Smart Guy"(Умник)
"The Color of Friendship"(Цвет дружбы)
"The Greatest Game Ever Played"(Триумф)
"The Muppets"(Маппеты)
"Winter Storage"(Запас на зиму)

imdb_id	title	type	rated	year	released_at	added_at	runtime	genre	director	writer	actors	language	awards
tt0043274	Alice in Wonderland	0	G	1951	1951-09-14	2019-11-12	75	Animation, Adventure, Family, Fantasy, Musical	Clyde Geronimi, Wilfred Jackson, Hamilton Luske	Lewis Carroll (adaptation), Winston Hibler (st...	Kathryn Beaumont, Ed Wynn, Richard Haydn, Ster...	English	Nominated for 1 Oscar. Another 1 nomination.
tt0047867	Beezy Bear	0	Unrated	1955	1955-09-02	2019-11-12	7	Animation, Comedy, Family	Jack Hannah	David Detiege (story), Al Bertino (story)	James MacDonald, Clarence Nash, Bill Thompson	English	NaN
tt0039263	Chip an' Dale	0	Approved	1947	1947-11-28	2019-11-12	7	Animation, Short, Comedy, Family	Jack Hannah	Dick Kinney, Bob North	Dessie Flynn, James MacDonald, Clarence Nash	English	NaN
tt0043428	Corn Chips	0	Approved	1951	1951-03-23	2019-11-12	7	Family, Animation, Comedy	Jack Hannah	Bill Berg (story), Nick George (story)	Dessie Flynn, James MacDonald, Clarence Nash	English	NaN
tt0365109	Deep Blue	0	G	2003	2004-01-29	2019-11-12	83	Documentary	Andy Byatt, Alastair Fothergill	David Attenborough	Michael Gambon, David Attenborough, Pierce Bro...	English	NaN

Рис. 2: Контент с рейтингом 7.4

Видно, что от рейтинга 6 и выше зависимость похожа на экспоненциальную, но после 8 резко падает. Можно сказать, что гипотеза подтвердилась частично.

Гипотеза 2

Для проверки второй гипотезы сначала визуализировано отношение общего количества фильмов и сериалов к количеству фильмов и сериалов с жанром "Анимация" (рисунки 3 и 4).

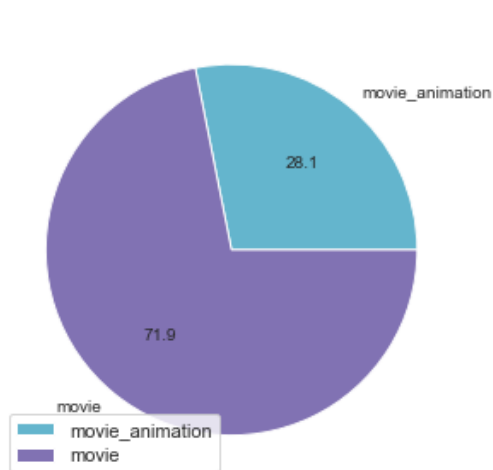


Рис. 3: Фильмы

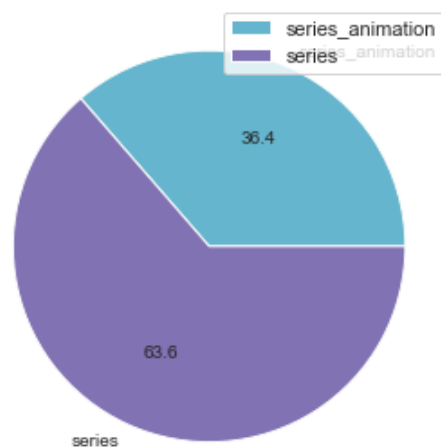


Рис. 4: Сериалы

Теперь, используя библиотеку matplotlib, визуализирую зависимость рейтинга от количества проголосовавших для фильмов и сериалов с жанром анимация и без него.

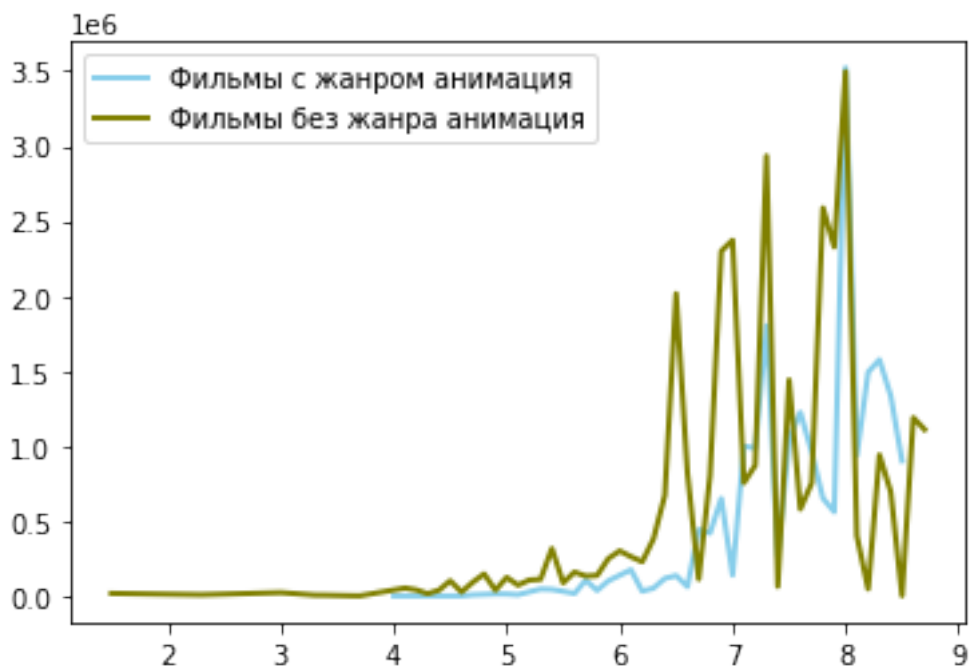


Рис. 5: Фильмы



Рис. 6: Сериалы

Из графиков видно, что сериалы с жанром анимация популярнее, чем без него. Однако для фильмов такая зависимость не работает. К тому же, соотношение сериалов с жанром анимация в процентном соотношении к общему количеству больше, чем фильмов. Можно сделать вывод, что гипотеза не подтвердилась, и популярность не зависит от присутствия жанра "Анимация".

Гипотеза 3

Для проверки третьей гипотезы в датафрейм добавлен еще один столбец "dif" в котором записана разница между датой релиза и датой загрузки на платформу в днях. Затем с помощью модуля seaborn построена диаграмма рассеяния для зависимости рейтинга от разницы в днях (рисунок 7).

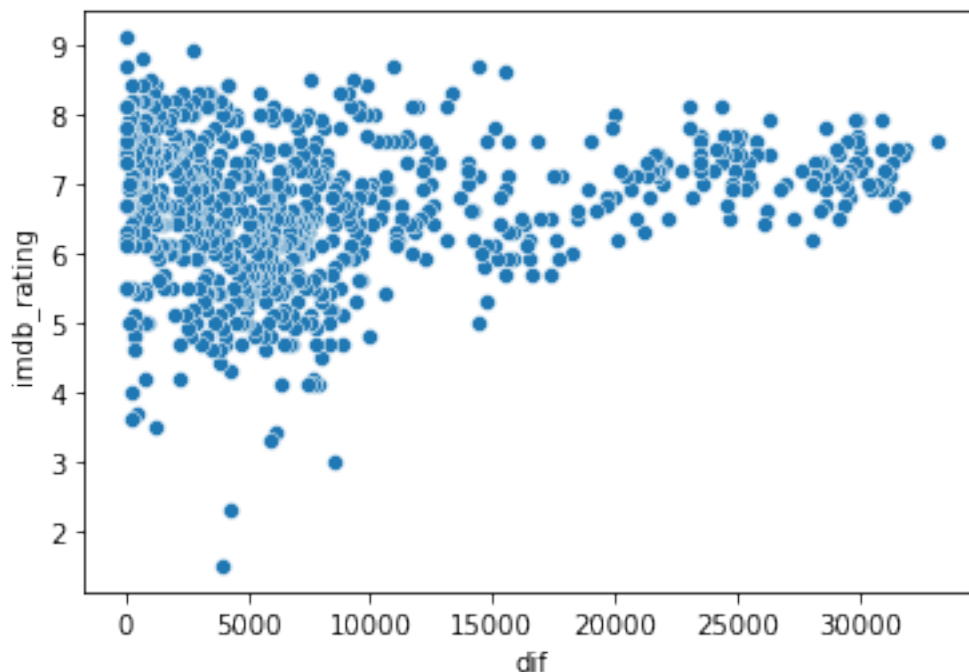


Рис. 7: Зависимость рейтинга от разницы в количестве дней

По графику видно, что гипотеза не подтвердилась. Да, самый высокий рейтинг у фильмов с минимальной разницей, но и фильмы с большой разницей в днях имеют довольно высокие оценки (на среднем уровне).

Гипотеза 4

С помощью модуля seaborn были визуализированы графики, показывающие количество контента по рейтингу с жанром "Family" и без него (рисунки 8 и 9).

По этим графикам видно, что рейтинг фильмов с этим жанром выше. Можно сделать общую визуализацию двух зависимостей (рисунок 10).

Добавила легенду к графику.

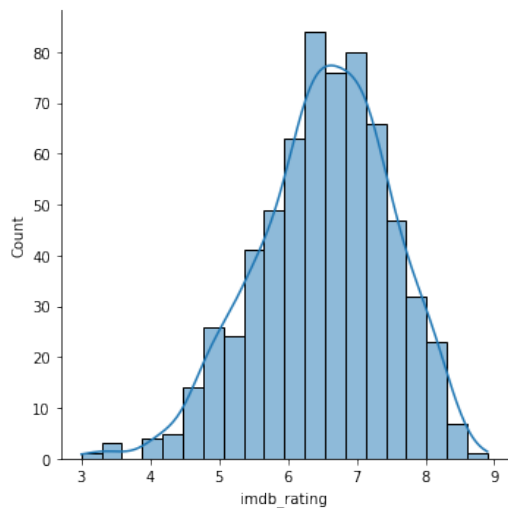


Рис. 8: Фильмы с жанром Family

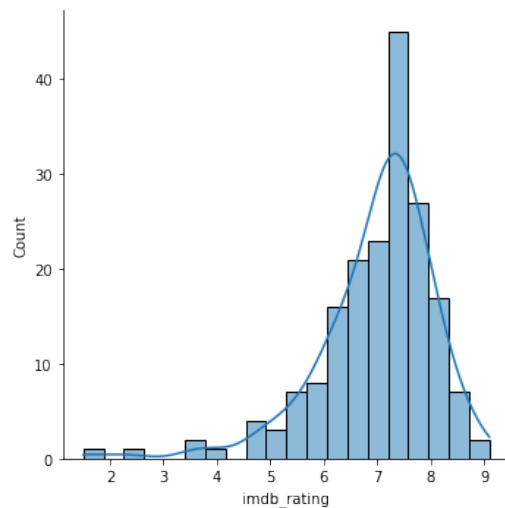


Рис. 9: Фильмы без жанра Family

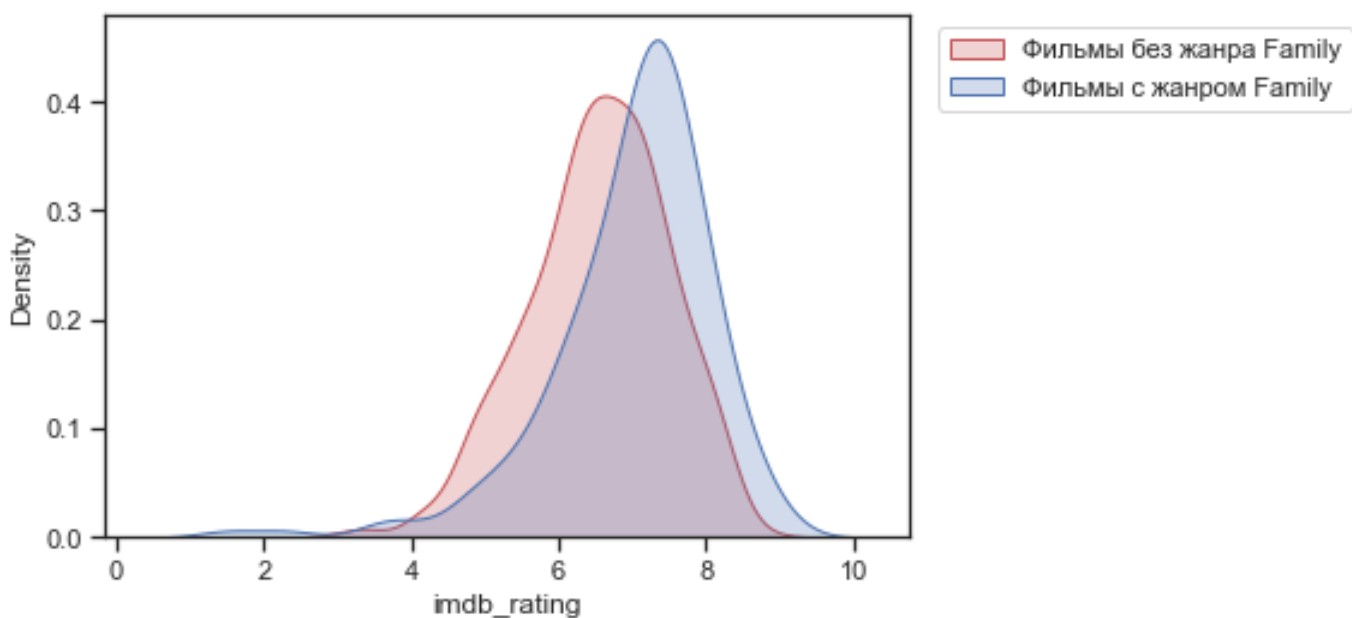


Рис. 10: Контент с жанром Family и без него

Гипотеза подтвердилась.

Гипотеза 5

Для проверки последней гипотезы был сформирован словарь с режиссерами, количество фильмов которых превышает 5. С помощью библиотеки Plotly был визуализирован средний рейтинг по фильмам этих режиссеров (рисунок 11).

Отсортировала значения по возрастанию среднего рейтинга.

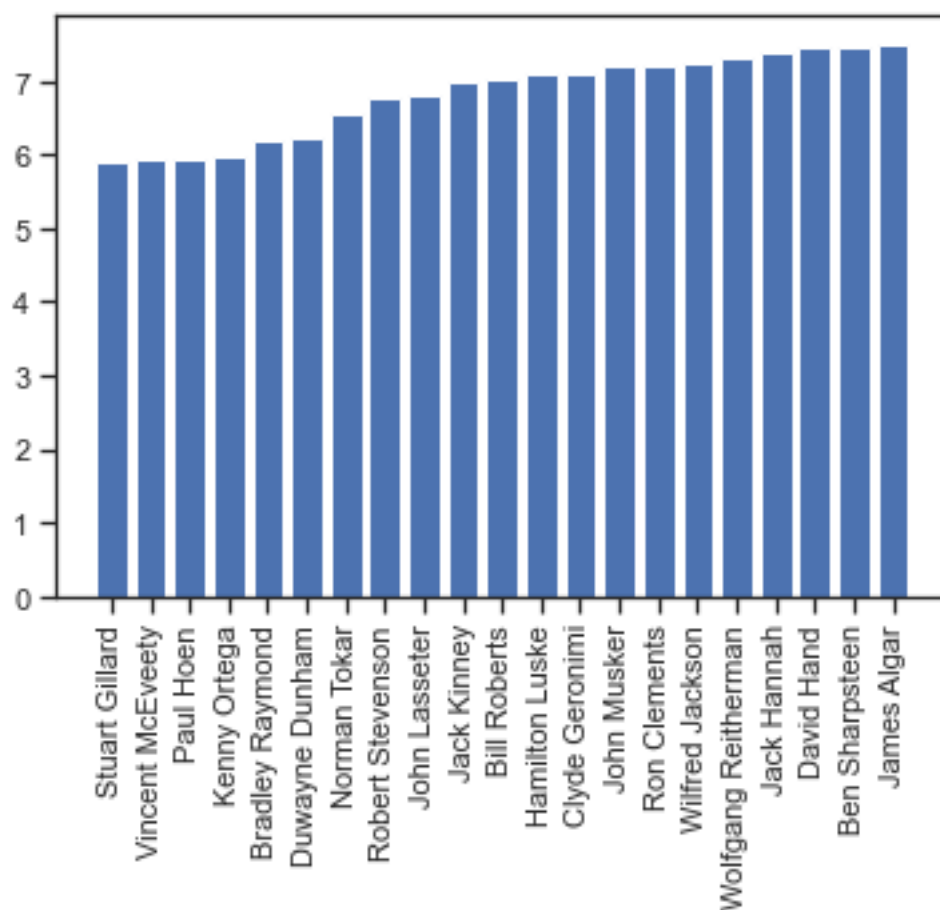


Рис. 11: Средний рейтинг фильмов режиссеров

По графику видно, что средний рейтинг у всех этих режиссеров довольно высок, что делает гипотезу верной.