

Pandas Project

Kazikhanov
Zhan
DataScience



01 – Проект

02 – Обзор

03 – Вопросы

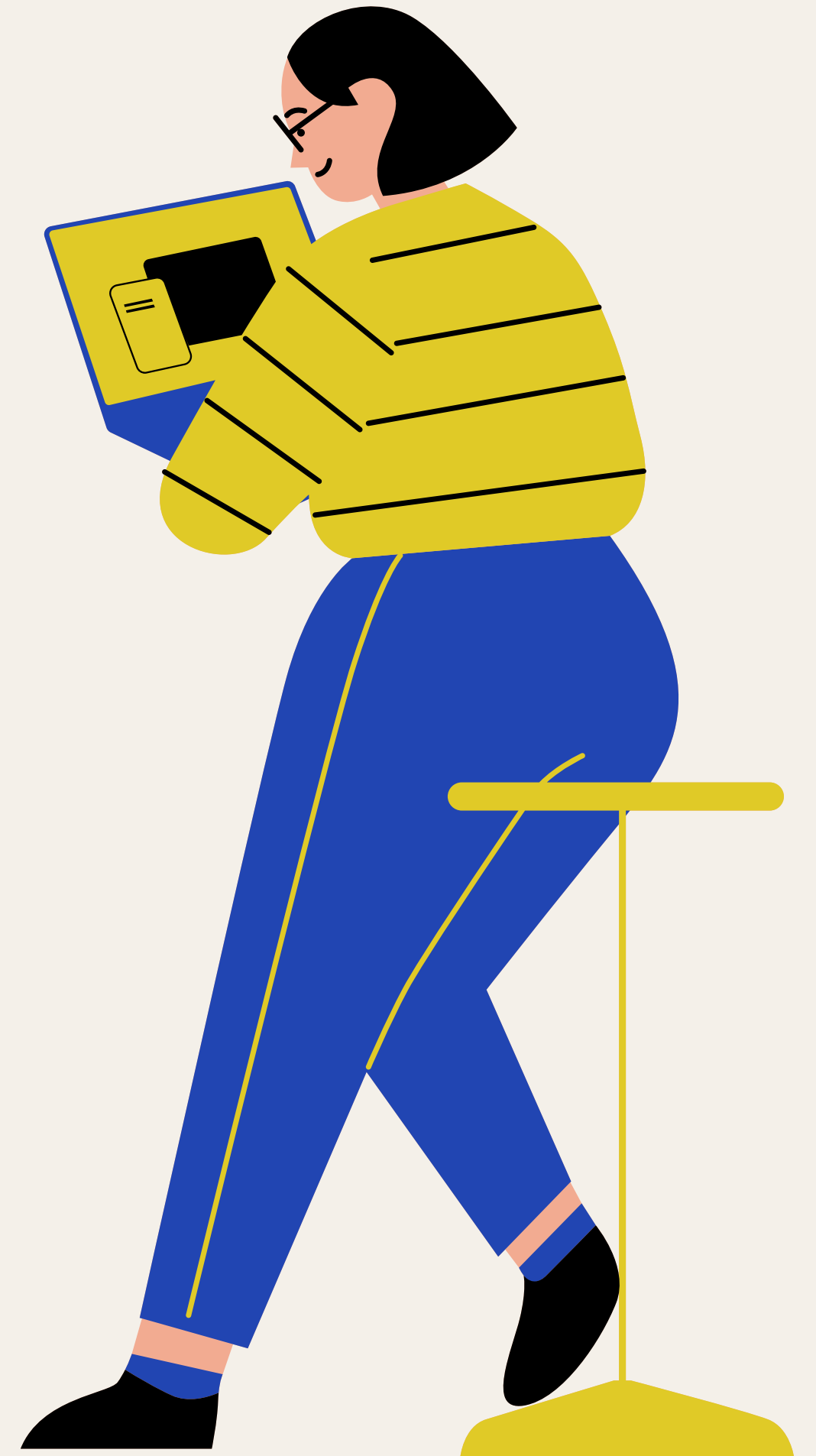
04 – Подготовка

05 – Анализ

06 – Заключение

Data

Analysis



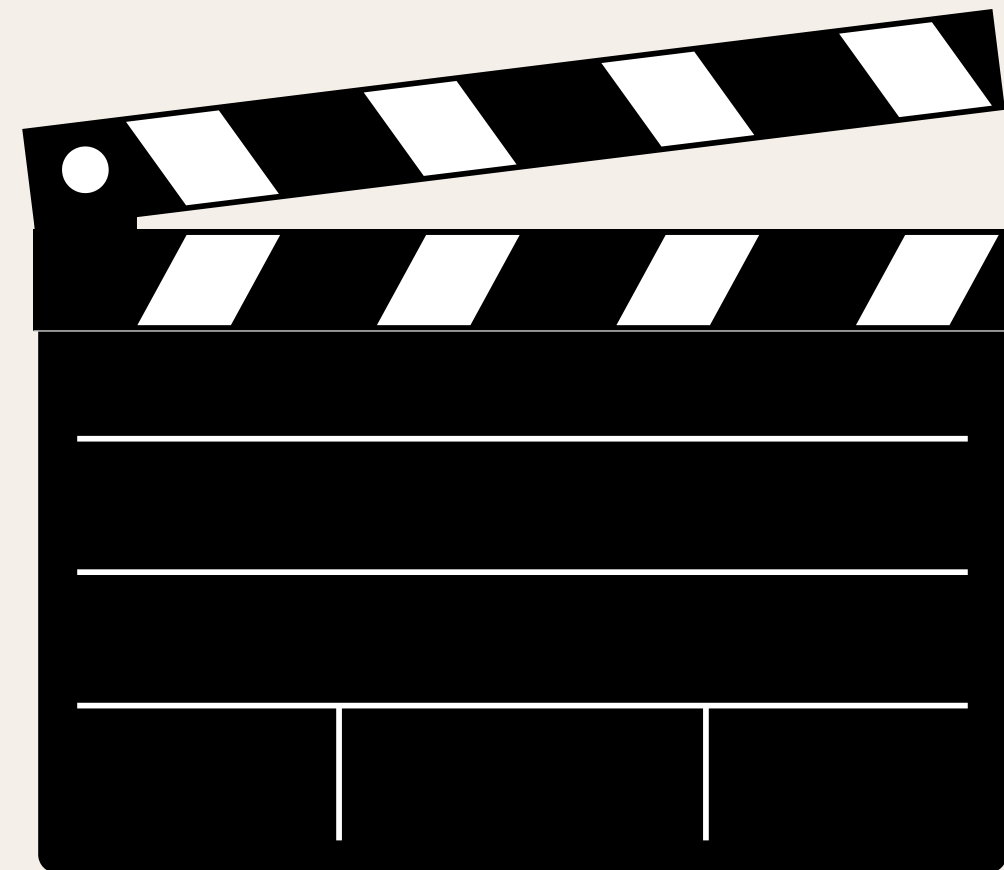
01 – Проект

Выбранный набор данных называется "Netflix IMDB Scores", и содержит в себе подробную информацию о фильмах и сериалах на стриминговой платформе Netflix и фокусируется на соответствующие оценки и голоса на сайте IMDb.

Благодаря этому набору данных, можно узнать как распределяются оценки на те или иные фильмы.

Ссылка:

kaggle.com/datasets/the-devastator/netflix-imdb-scores



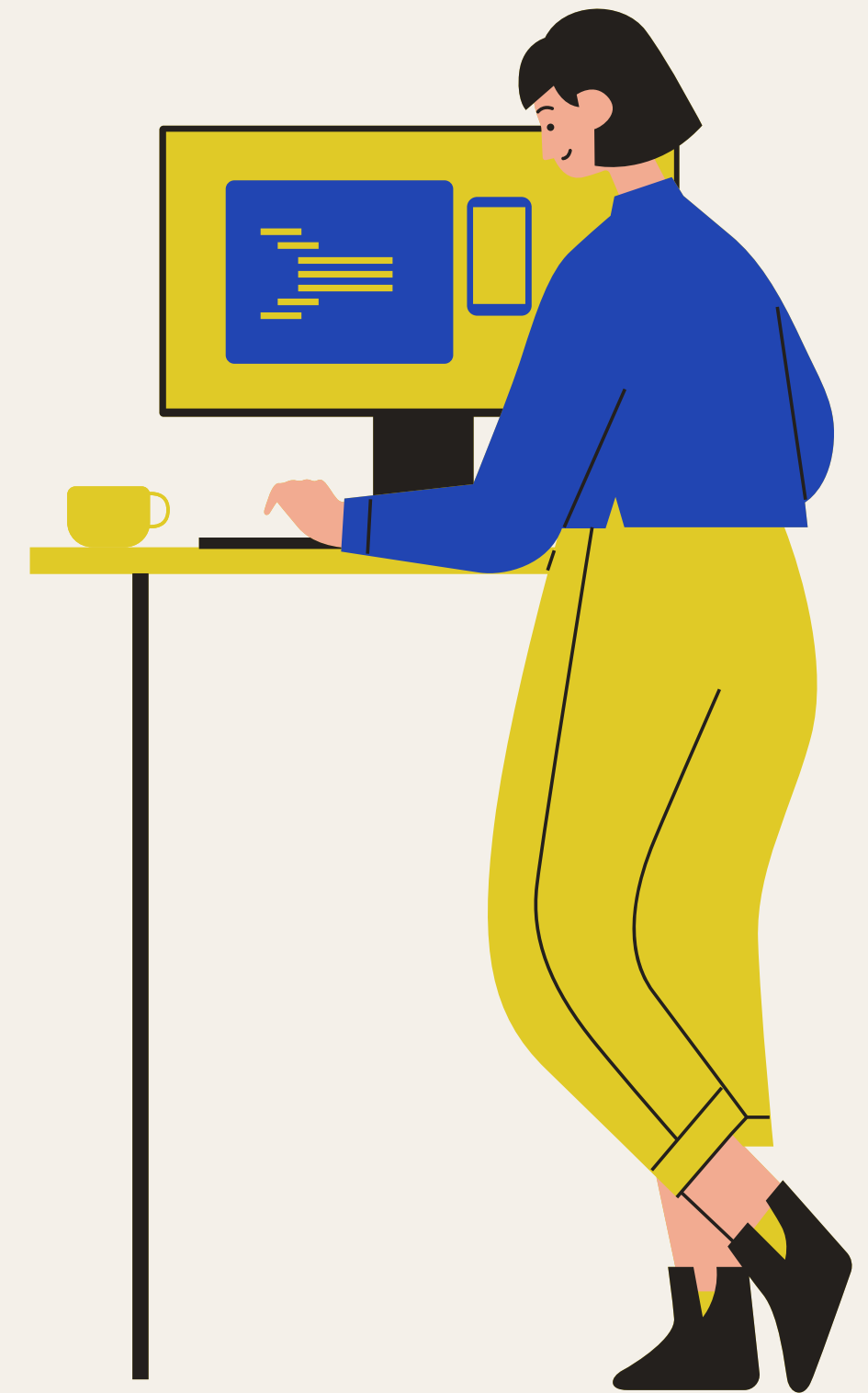
Data

Analysis

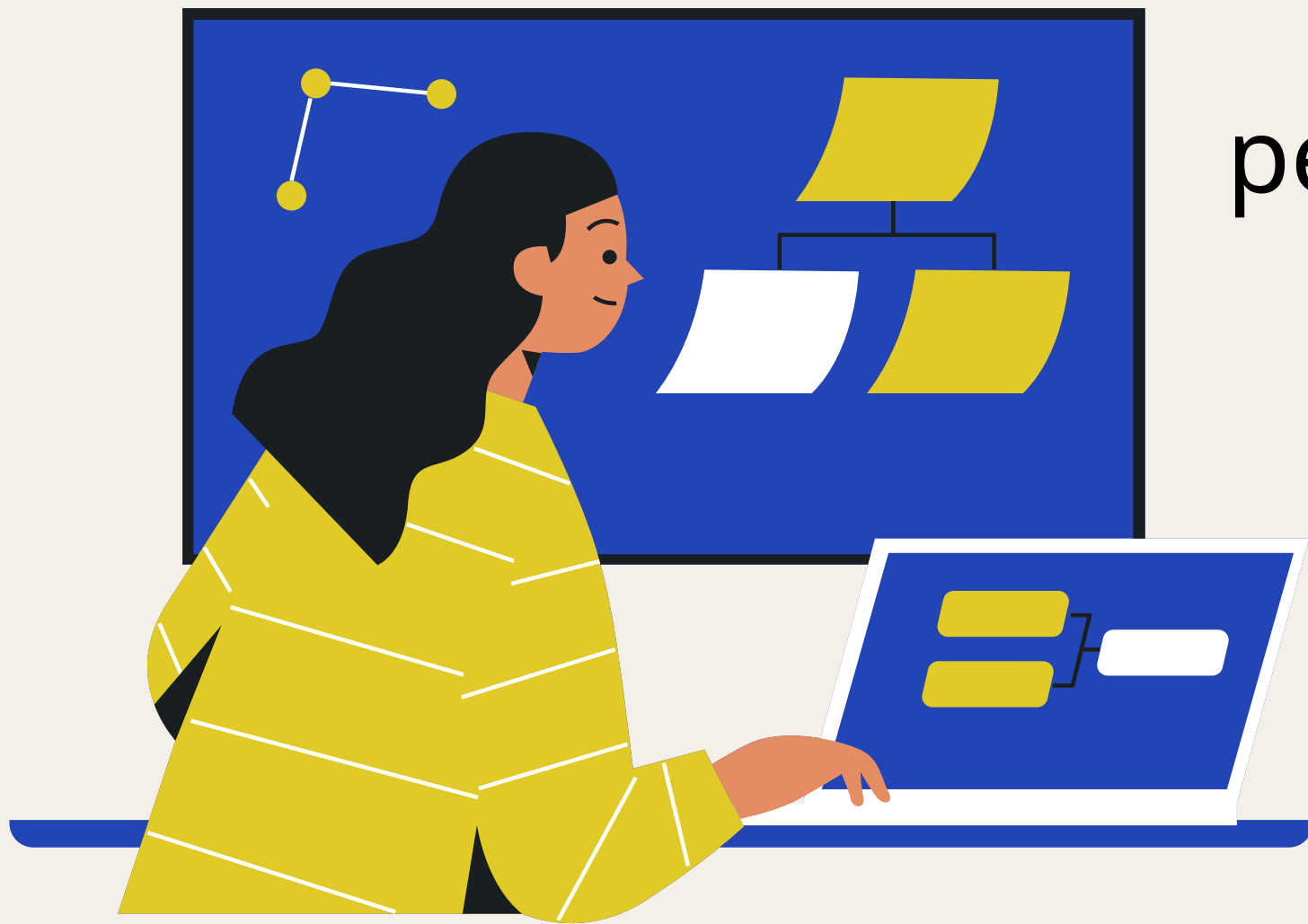
index – число, номер ряда
id – объект, номер фильма/сериала.
title – объект, название фильма/сериала.
type – объект, это фильм или сериал.
description – объект, аннотация.
release year – число, год выхода.
age_certification – объект, возрастной рейтинг.
runtime – число, длительность.
imdb_id – объект, номер в IMDb.
imdb_score – число, оценка в IMDb.
imdb_votes – число, количество голосов.

02 – Обзор

*Набор Данных имеет
11 атрибутов*



ОЗ – Вопросы



Data

Analysis

- 1) Как изменялись средние оценки контента на Netflix сквозь время?
- 2) Насколько большая разница между рейтингами лучших фильмов и сериалов на Netflix?
- 3) Может ли длительность картины влиять на оценку?
- 4) Как ежегодно менялось количество контента выпускаемого на Netflix?
- 5) Каким образом распределились голоса зрителей если опираться на возрастной рейтинг картин?

О4 – Подготовка

PLAN, PACK

PREPARE

В наборе данных имеются пропущенные значения, которые могут помешать объективному анализу данных.

Description – 5

Age_Certification – 2285

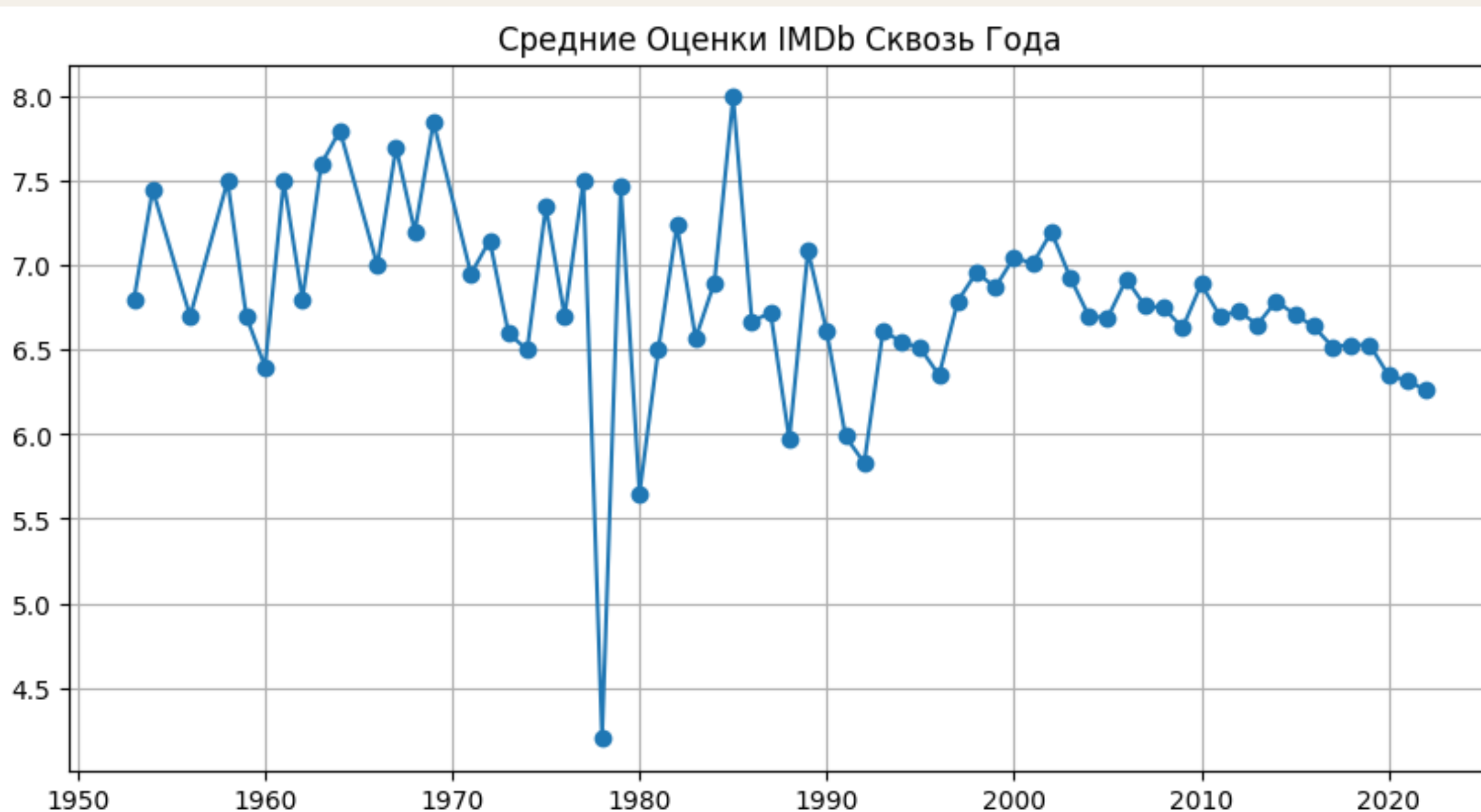
IMDb_Votes – 16

- Удаляем пропущенные значения описаний и оценок
- Но, удаление пропущенных значений возрастного рейтинга приведет к большой потере данных
- Поэтому, создаем 2 новых набора данных:
- 1) `prep-data` – набор данных с пропущенными значениями возрастного рейтинга
- 2) `prep_data_specific` – набор данных без пропущенных значений

О5 – Анализ

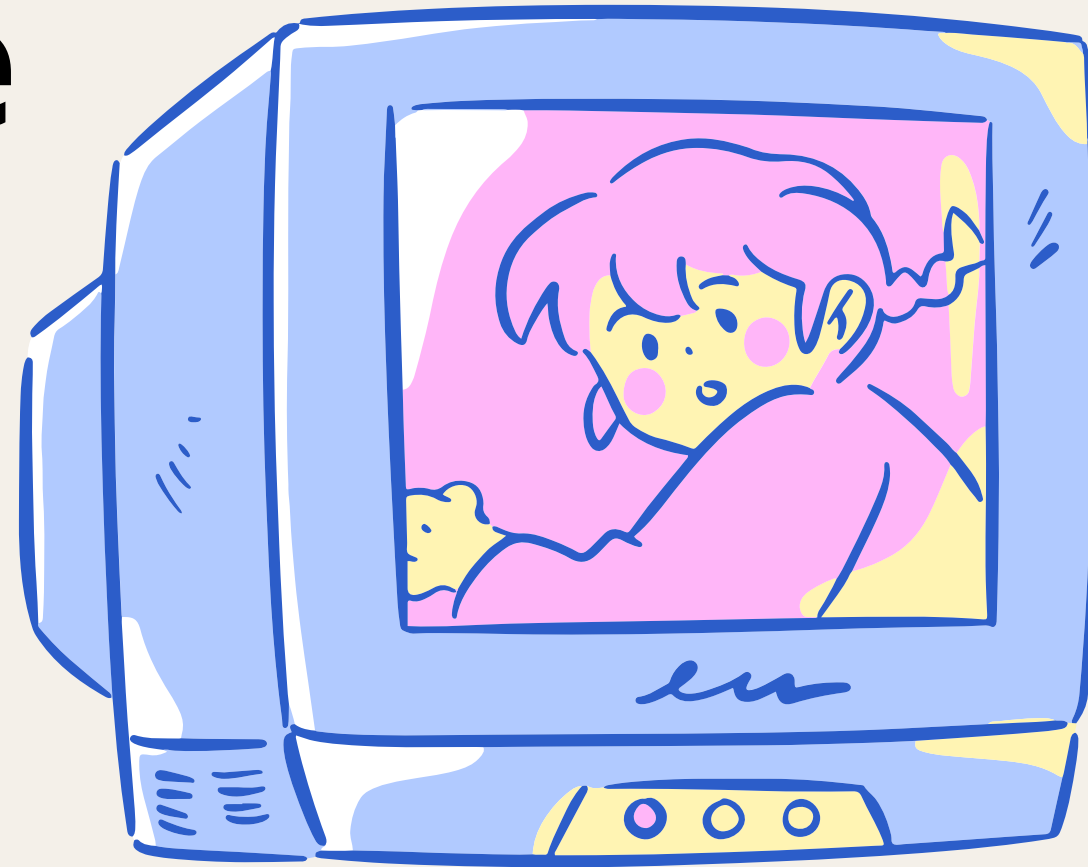
I Наблюдение

Худшие оценки получили фильмы и сериалы, которые были произведены в 1978 году со средним баллом меньше 4. В 1985 году, напротив, на свет вышли картины с лучшими оценками в среднем получая ровно 8 баллов. В целом, фильмы и сериалы XX века стабильно получали высокие баллы по сравнению с картинами XXI века, которые идут на спад в последние годы.



II Наблюдение

Лучший фильм:
No Longer Kids
1979 год
IMDb 9.0
Голоса – 943



Лучший сериал:
Khawwatir
2005 год
IMDb 9.6
Голоса – 3046

6.27 СРЕДНИЙ
РЕЙТИНГ
ФИЛЬМОВ



СРЕДНИЙ
РЕЙТИНГ
СЕРИАЛОВ 7.02

III Наблюдение

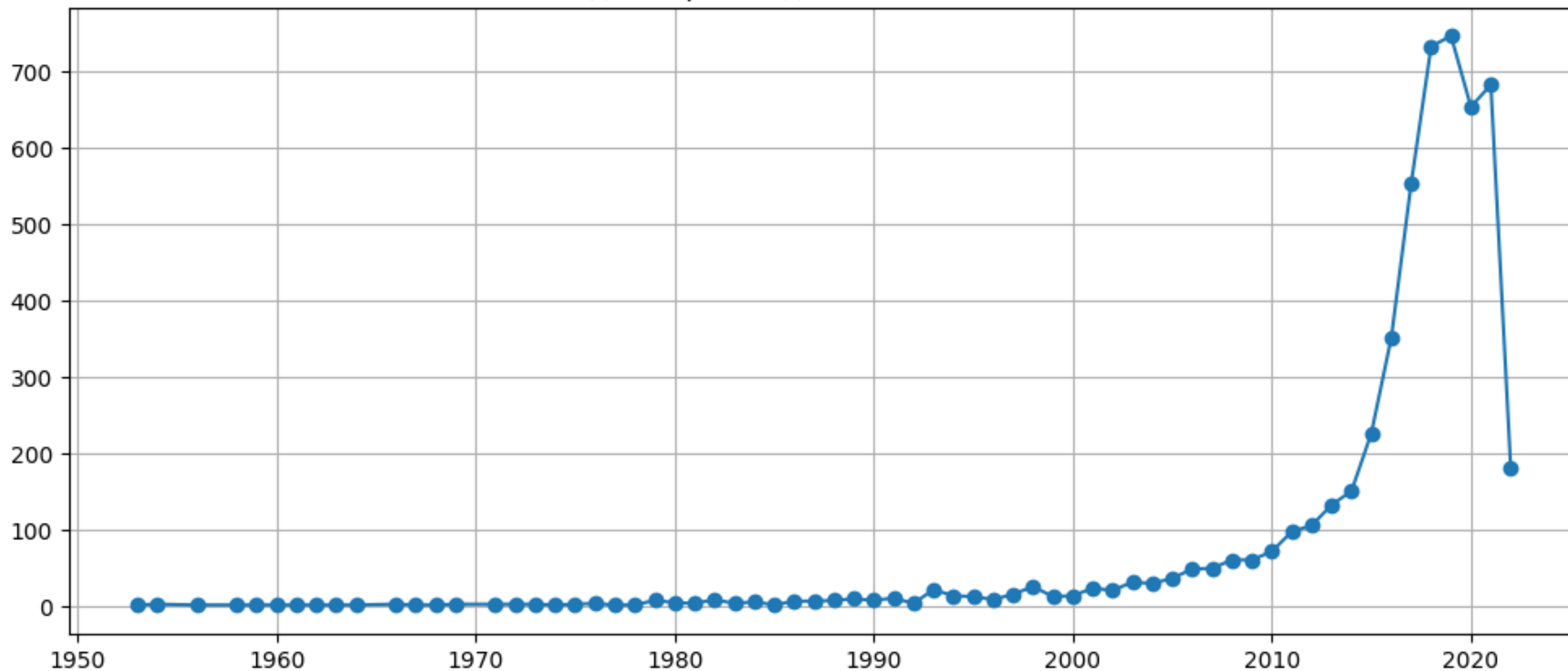
- Длительность большинства контента колеблется около 100 минут
- Такие фильмы были оценены в среднем от 5.6 до 7.2
- Лучший рейтинг (9.0) - 240 минут
- Худший рейтинг (4.5) - 190 минут
- Фильмы от 160 минут до 185 минут имеют самый высокий балл



IV Наблюдение

- Netflix был основан в 1990-х.
- Время спустя, количество контента на стриминговом сервисе начало постепенно расти
- Сервис достиг пика в 2019-2020, когда произвел более 700 фильмов и сериалов
- После 2020, Netflix резко убавил обороты, это вероятно связано со всемирной пандемией.

Ежегодное производство контента на Netflix



V Наблюдение

- Контент с рейтингом R ушел далеко вперед со средним количеством голосов составляющий почти 80000
- Самое наименьшее количество голосов имеет жанр TV-Y(для всех детей)
- Самые популярные возрастные рейтинги: PG-13 (13+), PG (15+), R (18+)



06 – Заключение

Data

Analysis

Можно сделать вывод что большинству зрителей могут понравится:

Сериалы с возрастным рейтингом R / PG-13, с длительностью эпизода около 40-50 минут

Фильмы с возрастным рейтингом R / PG-13 с хронометражем около 2.5-3 часов.

Спасибо
за ваше
внимание!

