

Министерство науки и высшего образования Российской
Федерации Санкт-Петербургский политехнический университет
Петра Великого Высшая школа прикладной математики и
вычислительной физики

Курсовой проект
по курсу «Большие данные»
Анализ данных Фильмы и телешоу Netflix

Выполнили

студент гр. 5040102/30201

Преподаватель

К.Ш. Зокиров

И.В. Никифоров

Санкт-Петербург - 2023

ВВЕДЕНИЕ

Netflix — одна из самых популярных платформ потокового мультимедиа и видео. На их платформе доступно более 8000 фильмов и телешоу, а по состоянию на середину 2021 года у них более 200 миллионов подписчиков по всему миру. Этот табличный набор данных состоит из списков всех фильмов и телешоу, доступных на Netflix, а также таких подробностей, как актерский состав, режиссеры, рейтинги, год выпуска, продолжительность и т. д.

Данная работа представляет собой анализ данных о фильмах и телешоу популярной платформы видеостриминга Netflix. Netflix, основанная в 1997 году, стала одной из самых успешных и востребованных платформ для просмотра фильмов и сериалов. Ее библиотека содержит тысячи произведений различных жанров, что делает ее идеальным полем для исследования и анализа данных.

В рамках проекта используются различные методы анализа данных, включая статистические инструменты и визуализацию данных. Основные вопросы, на которые мы попытаемся ответить, касаются предпочтений зрителей в отношении жанров, популярности фильмов и телешоу, а также специфических особенностей каждого географического региона.

ПОСТАНОВКА ЗАДАЧИ И АКТУАЛЬНОСТЬ

Задачей проекта является анализ данных о фильмах и телешоу на платформе Netflix с целью выявления основных тенденций, предпочтений зрителей и анализ контента. Конкретные задачи исследования могут включать:

1. Распределение рейтингов контента на Netflix, представленных на платформе. Это позволит более глубоко понять предпочтения зрителей и сделать выводы о популярности конкретных жанров.
2. Также рассмотреть режиссеров фильмов на Netflix и определить топ-5 самых продуктивных и успешных режиссеров на платформе.
3. Провести анализ актеров по их участию в фильмах на Netflix и определить топ-5 наиболее активных и популярных актеров на платформе.
4. Изучение тенденций в контенте, выпускаемом на платформе Netflix, с течением времени, анализ изменений в объеме контента и предпочтений зрителей по жанрам.
5. Исследовать, как меняется настроение контента на платформе Netflix с течением времени, выявить тренды и распределение жанров по настроению.

ОБЗОР СУЩЕСТВУЮЩИХ ИССЛЕДОВАНИЙ

При подготовке данного отчета проведен анализ предыдущих исследований, посвященных анализу данных о фильмах и телешоу на платформе Netflix. Это позволило обобщить и систематизировать существующие знания и выделить основные тенденции и результаты, полученные в предыдущих исследованиях.

Одно из исследований, проведенное на основе данных Netflix, предоставило информацию о предпочтениях зрителей в разных странах. Была выявлена разница в предпочтениях жанров в разных регионах. Например, в США наиболее популярными жанрами оказались драма и комедия, в то время как в Японии идология и анимация оказались на верхушке списка. Этот результат подтверждает важность учета культурных особенностей в процессе создания и распространения контента.

Другое исследование сосредоточилось на оценке популярности различных телешоу и фильмов на платформе. На основе данных об использовании и оценках пользователей были выделены самые популярные произведения. Это позволило выявить тренды во вкусах и интересах аудитории, что может быть полезным для создания и продвижения новых контентных проектов.

Есть еще исследования сфокусировавшиеся на анализе длительности фильмов и телешоу на Netflix. Были изучены средняя продолжительность разных жанров и сегментов, что помогло понять предпочтения зрителей в отношении продолжительности контента. Результаты этого исследования могут быть использованы для оптимизации продолжительности фильмов и телешоу, чтобы удовлетворить ожидания аудитории.

РЕАЛИЗАЦИЯ ПРОЕКТА

В работе используются данные 18.10.2021г. Разработка приложения производится на языке Python. В качестве среды разработки выступает программный продукт Visual Studio Code. Для анализатора объявлений были использованы следующие библиотеки:

- numpy
- pandas
- plotly
- textblob

АРХИТЕКТУРА ПРОЕКТА

Результаты

На рисунке (1) показано распределение контента на Netflix

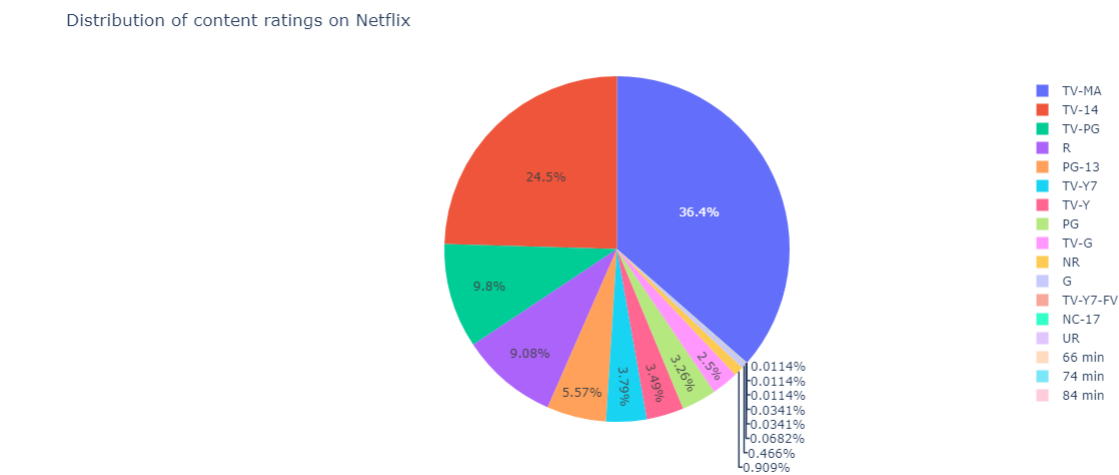


Рис. 1: Распределение контента

На рисунке (2) показано топ 5 лучших продуктивных режиссеров на Netflix

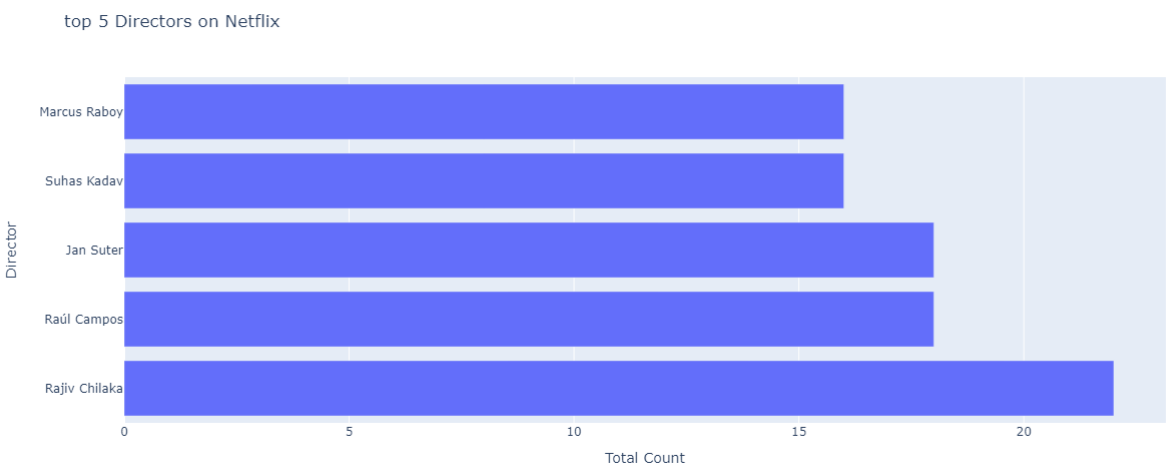


Рис. 2: 5 лучших режиссеров Netflix

В основе диаграммы используется горизонтальная шкала, где по оси Y указываются имена режиссеров. Для каждого режиссера отображается строка, длина которого соответствует количеству его фильмов. Таким образом, более длинная строка указывает на режиссера, который снял больше фильмов.

На рисунке (3) также показано топ 5 лучших продуктивных и успешных актеров на Netflix

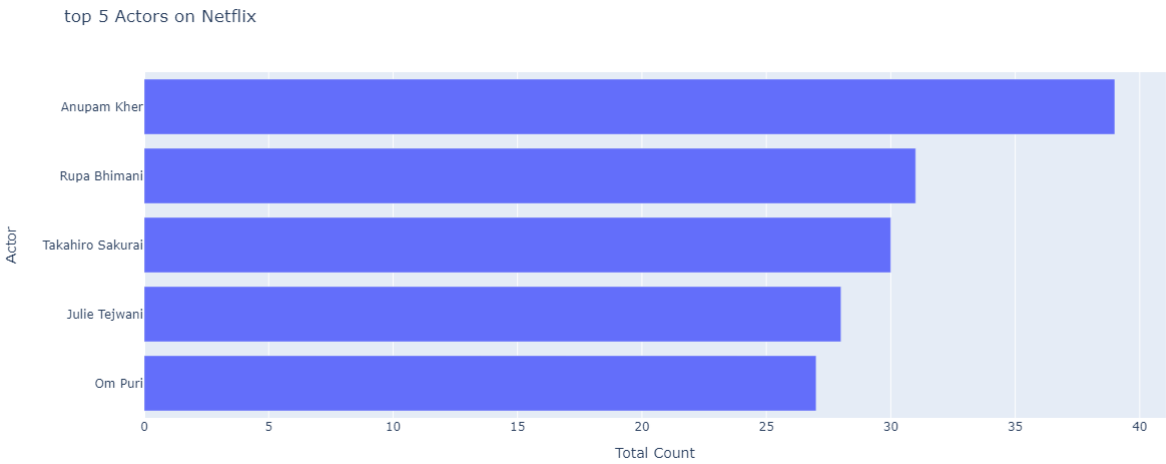


Рис. 3: 5 лучших актеров Netflix

На рисунке (4) показано Тенденция контента, выпускаемого на Netflix каждый год

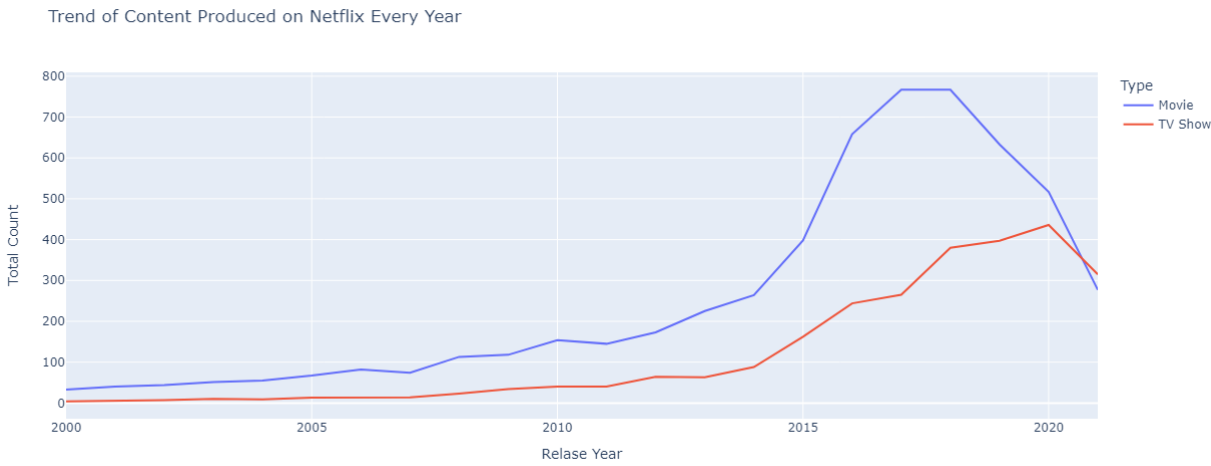


Рис. 4: Тенденция контента

Диаграмма "Тенденция контента, выпускаемого на Netflix каждый год" является инструментом визуализации данных, который отображает изменения в контенте, который выпускается на платформе Netflix с течением времени.

На данной диаграмме ось X представляет годы, а ось Y указывает на количество выпущенных контентов в соответствующем году. Для каждого года используется точка или столбец, размер которого отражает объем выпущенного контента. Таким образом, данная диаграмма позволяет наглядно проанализировать изменения в объеме контента на Netflix по годам.

На рисунке (5) показано Анализ настроения контента на Netflix

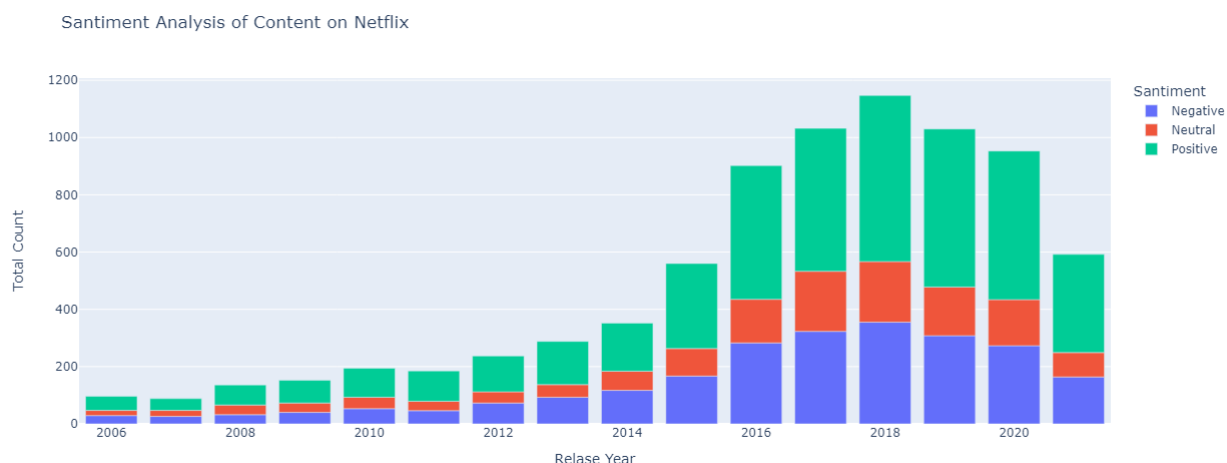


Рис. 5: Анализ контента

Анализ настроения контента на Netflix будет представлять ось X, представляющую годы, и ось Y, указывающую на показатель настроения, который будет отображаться в виде количества положительных, отрицательных и нейтральных отзывов, положительные отзывы обозначены голубым цветом, отрицательные — красным, а нейтральные — синим.

На рисунке (6) показано Диаграмма распределения контента в разных странах на Netflix

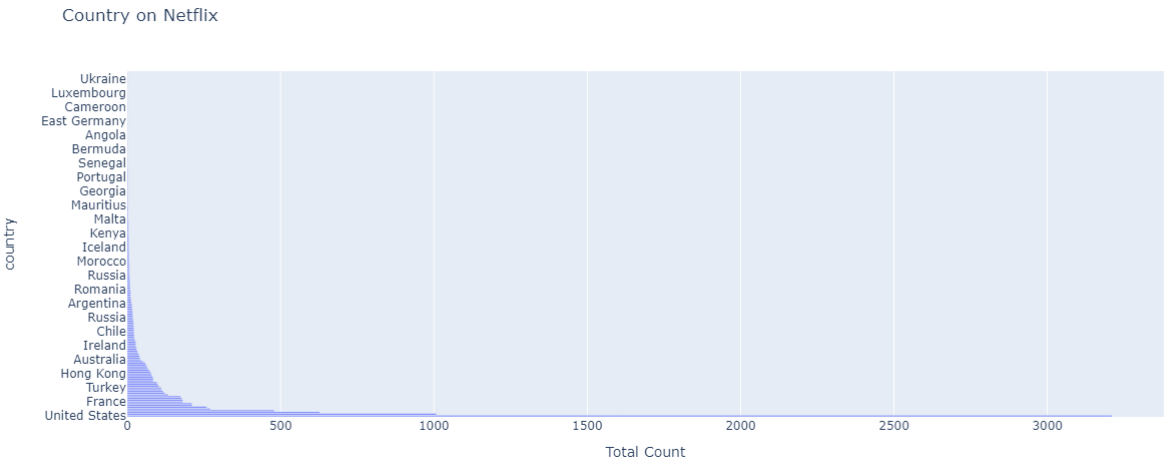


Рис. 5: Диаграмма распределения контента