**Анализ шоу с Нетфликса**

Савинова Анастасия 1

*1 – Технический университет Молдовы*

e-mail: anastasia.savinova@iis.utm.md

**Аннотация**

Стриминговый сервис Нетфликс является одной из самых известных платформ для просмотра фильмов и сериалов. В данном исследовании были проанализированы данные об около 8000 фильмов и сериалов с Нетфликса. Данные наглядно отображены на различных графиках. Также была реализована модель для предсказания количества выходящих фильмов по жанрам.

**Введение**

Целью данного исследования является изучение разнообразия шоу с Нетфликса и выявление наиболее успешных из них.

В данной работе продемонстрирована разница в количестве между фильмами и сериалами, число вышедших фильмов по годам, 10 стран, выпустившие наибольшее количество фильмов и сериалов, распределение рейтингов и среднюю продолжительность всех фильмов, выпущенных в один год, распределенную по годам. Также была построена модель, предсказывающая число выходящих фильмов различных жанров по годам.

**Материалы и Методы**

*Анализ*

Весь анализ, а именно, датасет, графики, презентация и прочее находятся в свободном доступе на платформе GitHub. Репозиторий: “AyanoNakahara/ AnalysisNetflixShows” [1].

*Датасет*

Выбранный датасет называется “NetFlix Shows”, находится в свободном доступе на известной научной платформе «Kaggle» и содержит подробную информацию об около 9000 шоу с Нетфликса [2].

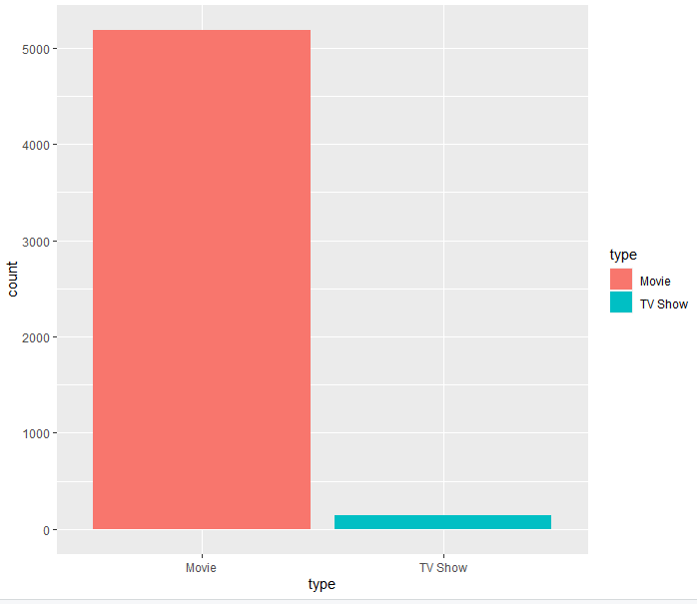
*Инструменты для анализа*

Главным инструментом для анализа выступил язык и среда для статистических вычислений и графики “R”. Были использованы такие библиотеки данного языка, как “tidyverse”, “ggplot” и “dplyr”.

Основной анализ был выполнен с помощью техники EDA (Exploratory Data Analysis), посредством которой мы создаем наглядно отражающие предоставленную информацию графики. Модель для предсказания количества выходящих фильмов по жанрам была создана с помощью линейной регрессии.

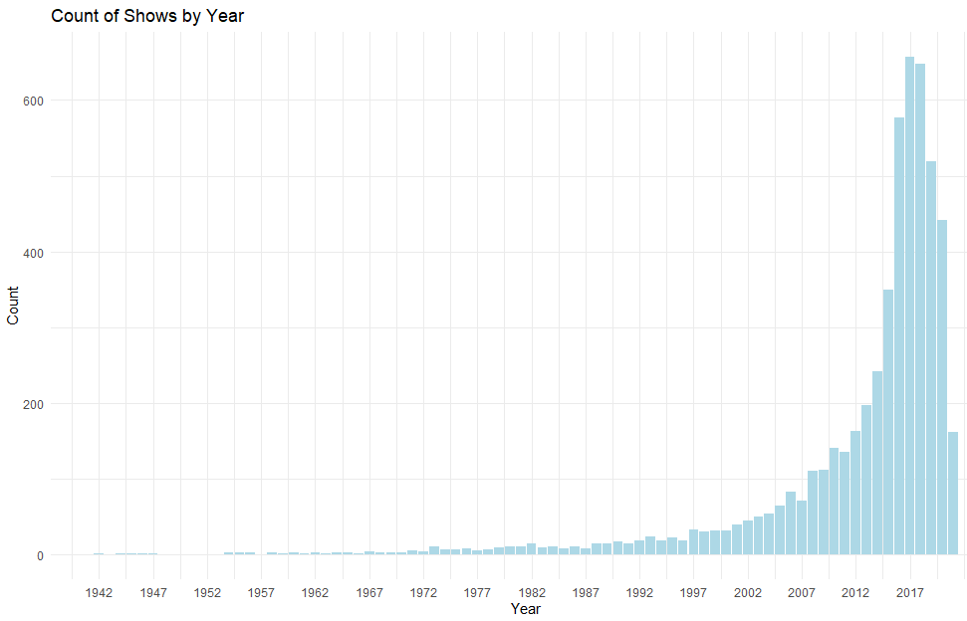
**Результаты**

Согласно информации, представленной в выбранном датасете, мы можем увидеть, что число фильмов на Нетфликсе во много раз превышает число сериалов.



**Рисунок 1.** Разница между количеством фильмов и сериалов

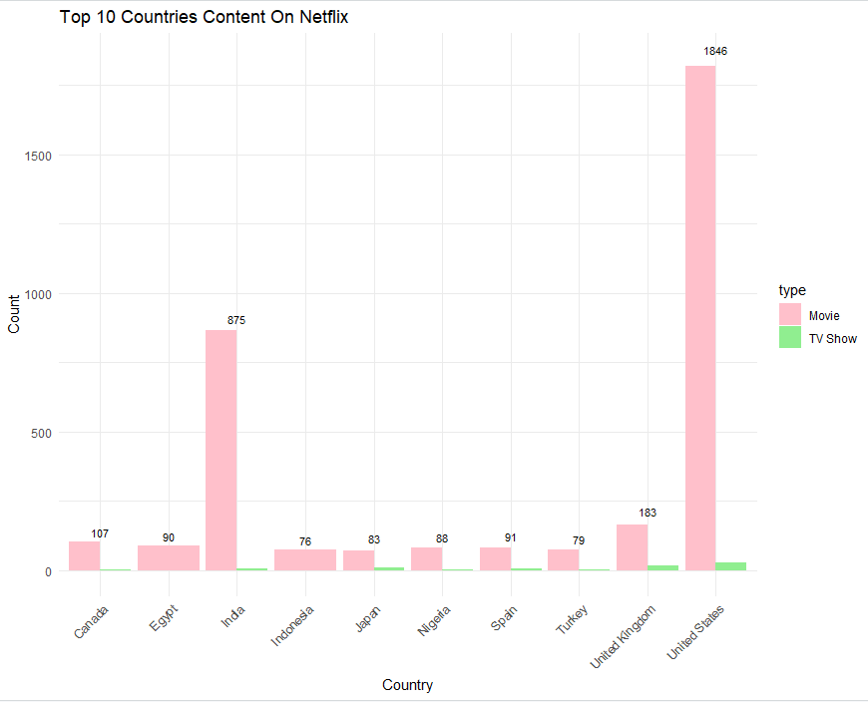
Далее нам предоставляется информация о количестве шоу с Нетфликса по году выпуска. Больше всего фильмов и сериалов, находящихся на Нетфликсе, было выпущено в 2017 году. Также видно, что шоу более ранних годов выпуска значительно меньше, чем, например, шоу, которые были выпущены в последнее десятилетие.



**Рисунок 2.** Количество шоу по годам

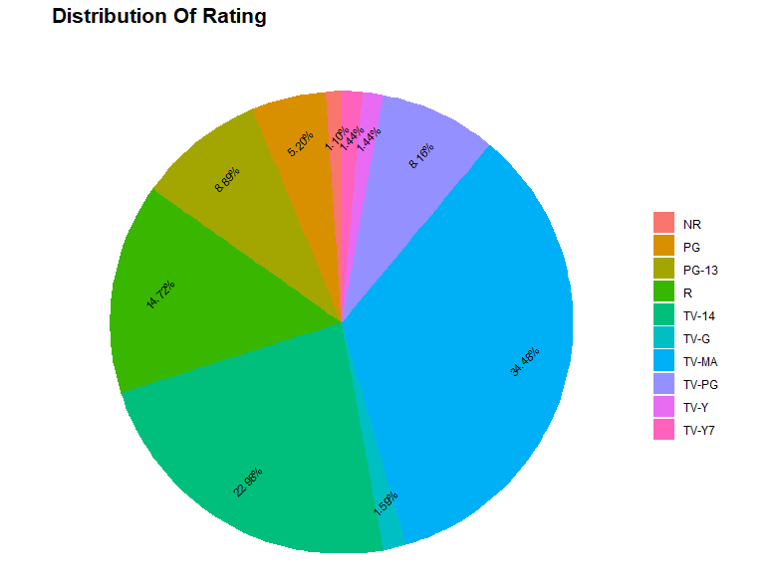
В топ 10 стран, выпустивших больше всего фильмов и сериалов, входят: США, Индия, Великобритания, Канада, Испания, Египет, Нигерия, Япония, Турция и Индонезия.

Больше всего фильмов выпустили Соединенные Штаты Америки, на втором же месте расположилась Индия. Рекордсменом по сериалам также является США, но на втором месте расположилась уже Великобритания.



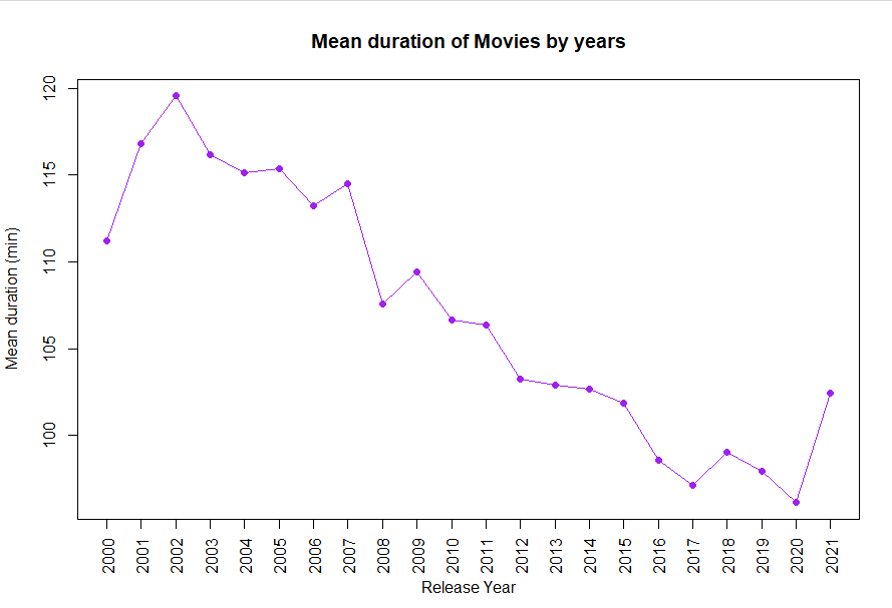
**Рисунок 3.** Топ 10 стран выпуска

Самым распространенным рейтингом является TV-MA, а за ним TV-14. Самым редким рейтингом стал NR, а следом за ним расположились TV-Y и TV-Y7.



**Рисунок 4.** Распределение рейтингов

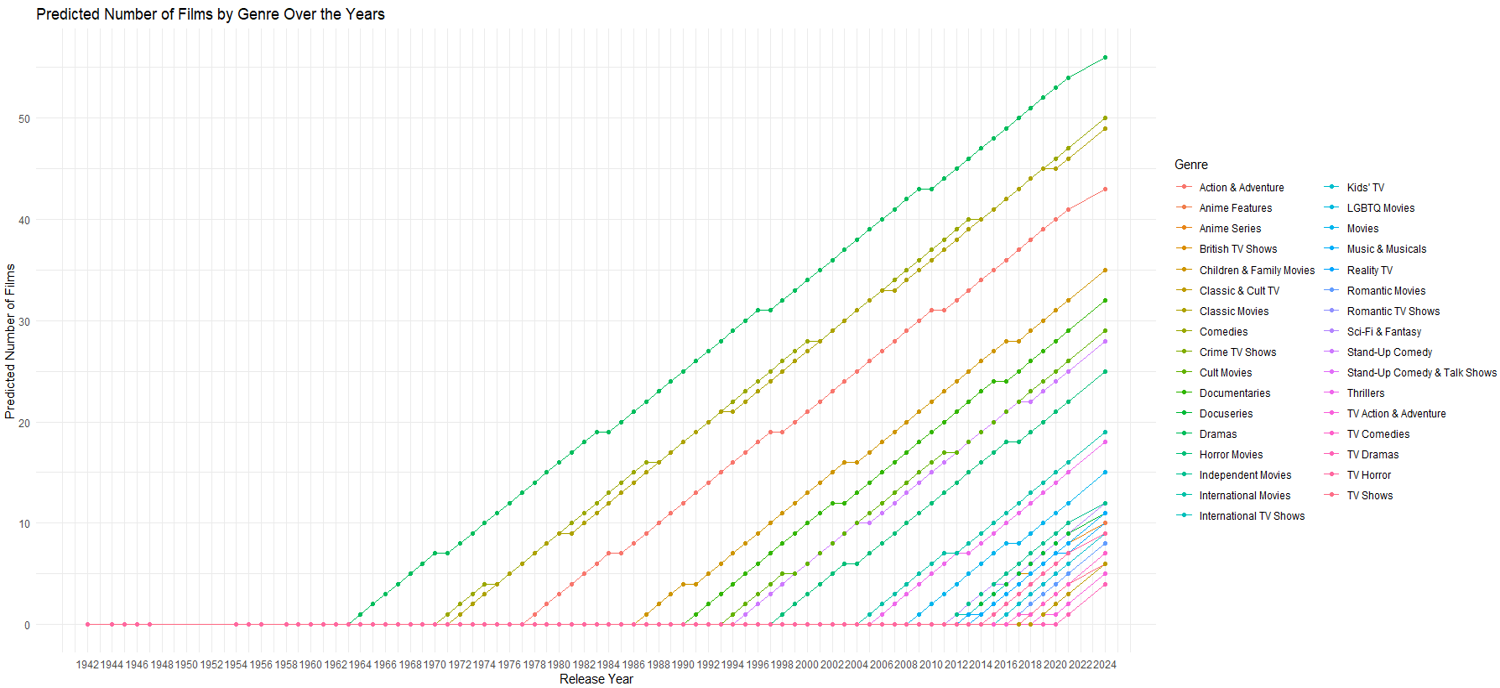
Наибольшая средняя продолжительность фильмов приходится на 2002 год (почти 120 минут, то есть 2 часа), а наименьшая же приходится на 2020 год.



**Рисунок 5.** Средняя продолжительность фильмов по годам

С использованием предоставленных данных была построена модель линейной регрессии для предсказания количества выходящих фильмов разных жанров по годам. Модель представляет собой y = wx + b, где x – это год, а y – количество фильмов.

Затем мы можем пронаблюдать, как с каждым годом количество фильмов повышается с выходом новых фильмов, и попробовать предсказать, как увеличится число фильмов в последующем году.



**Рисунок 6.** График, отображающий предсказание повышения количества фильмов к 2024 году.

**Заключение**

Проанализировав и изучив информацию о шоу с Нетфликса, мы выяснили, что количество фильмов превышает количество сериалов, пик выхода шоу приходится на 2017 год, больше всего фильмов и сериалов выпускает США, а самым распространенным рейтингом является TV-MA.

Мы проследили за увеличением количества фильмов различных жанров по годам. Благодаря соответствующей модели мы также смогли попытаться предсказать повышения количества фильмов к 2024 году.

**Библиография**

1. Савинова Анастасия, GitHub Репозиторий AnalysisNetflixShows, <https://github.com/AyanoNakahara/AnalysisNetflixShows>. [18.12.2023]
2. Dataset: NetFlix Shows,

<https://www.kaggle.com/datasets/infamouscoder/dataset-netflix-shows/>. [18.10.23]