Content Analysis Report - Netflix

Objective

The objective of this content report and this project as a whole is to access, analyze & assess the data of Netflix's global catalogue of Movies & TV shows across multiple languages to uncover patterns and trends that could be used to inform Netflix's strategy moving forward.

The personal objective of this project was to deepen my understanding of fundamental techniques in data extraction, manipulation, and presentation as I begin my first foray into the world of Data science.

This project is for me to practice using sqlite3, pandas, matplotlib & seaborn libraries in Python.

Tools & Resources Used

The database used here was sourced from Kaggle.

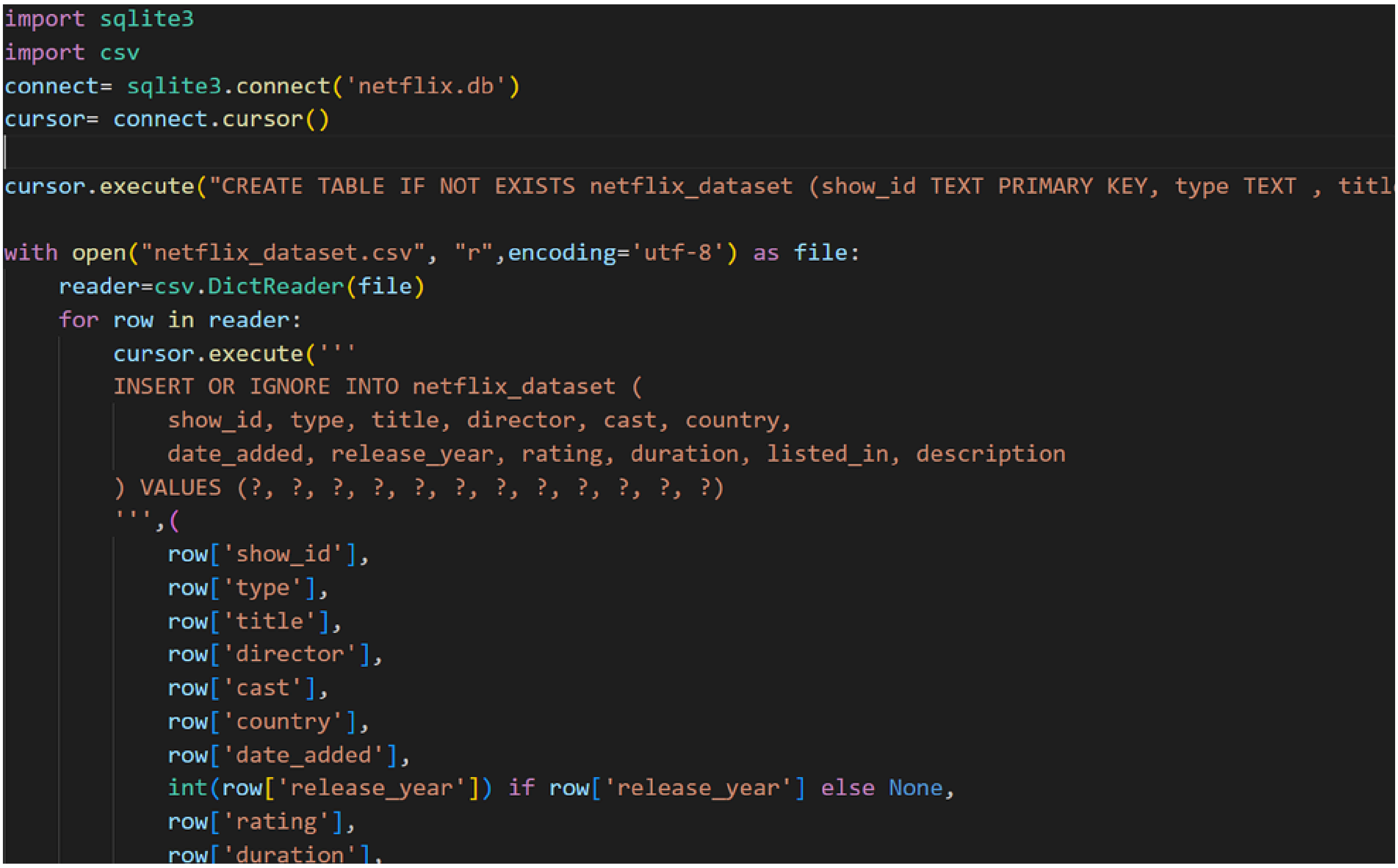
SQL though Sqlite3 was used for viewing, changing & comparing data within the given CSV file.

The python libraries pandas, matplotlib & seaborn were used in conjunction for the loading, then the visualization of the data in the form of vertical & horizontal barcharts, Histogram + Kernel density line and line charts.

This report was written by me alone using Canva Docs. All images are screenshots of the working code (partial) / graphs generated though matplotlib & seaborn.

Data Preprocessing

Table was created and the data was loaded into it using sqlite3 module in python.

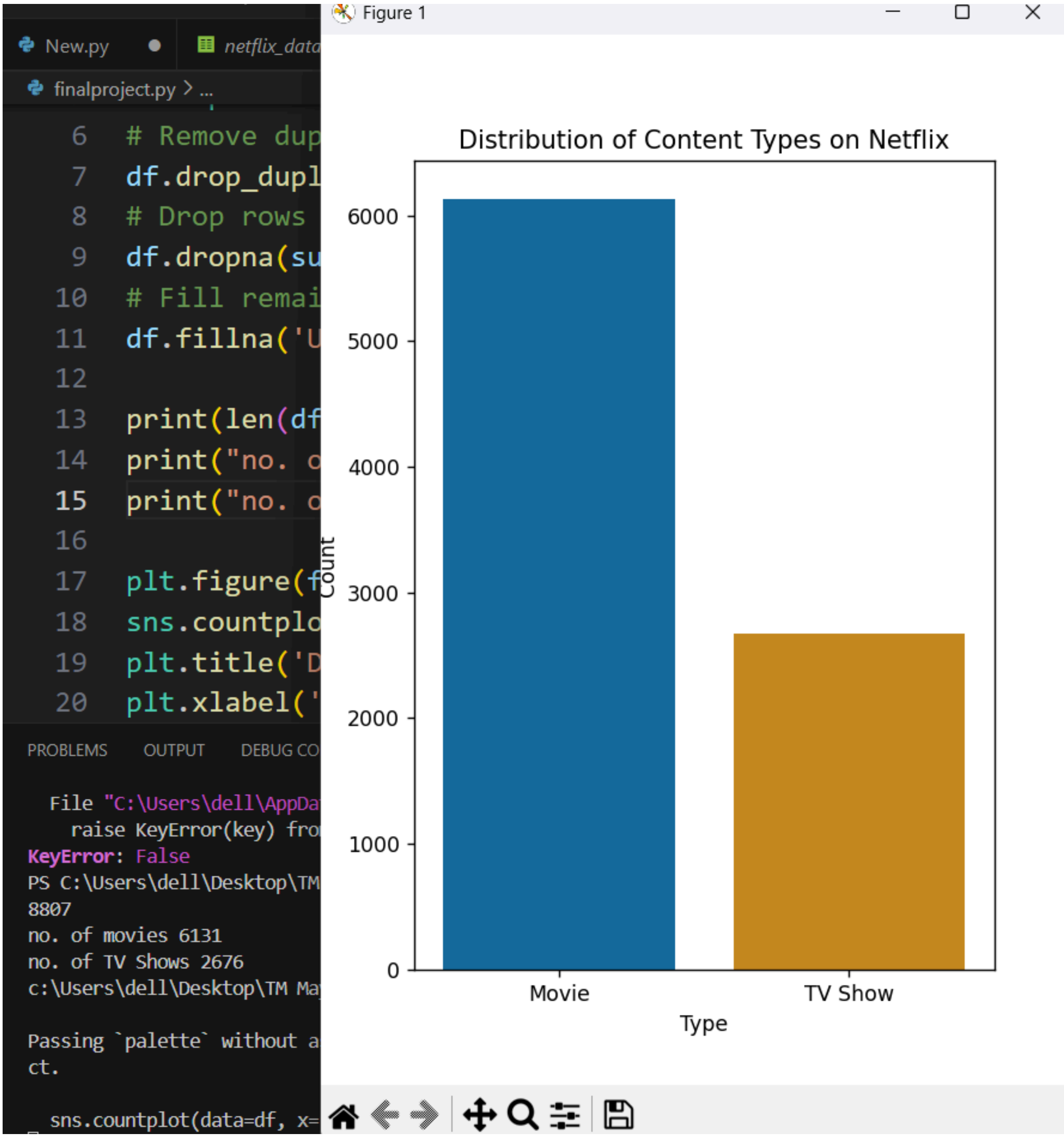


```
import sqlite3
import csv
connect= sqlite3.connect('netflix.db')
cursor= connect.cursor()

cursor.execute("CREATE TABLE IF NOT EXISTS netflix_dataset (show_id TEXT PRIMARY KEY, type TEXT , titl

with open("netflix_dataset.csv", "r",encoding='utf-8') as file:
    reader=csv.DictReader(file)
    for row in reader:
        cursor.execute('''
        INSERT OR IGNORE INTO netflix_dataset (
            show_id, type, title, director, cast, country,
            date_added, release_year, rating, duration, listed_in, description
        ) VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?)
        ''',(
            row['show_id'],
            row['type'],
            row['title'],
            row['director'],
            row['cast'],
            row['country'],
            row['date_added'],
            int(row['release_year']) if row['release_year'] else None,
            row['rating'],
            row['duration'],
```

Using pandas, Dataset was loaded, duplicates were removed, some entries missing fundamental information crucial for this analysis like title, release year and type (Movie/TV Show) were removed entirely. Empty values in the remaining entries were given the value ‘unknown’.

General Overview of Netflix’s Content Library

There seems to be 8807 total entries in the library. 6131 of these entries are movies while 2676 entries are TV Shows.



```
# Remove dup
df.drop_dupl
# Drop rows
df.dropna(su
# Fill remai
df.fillna('U

print(len(df
print("no. o
print("no. o

plt.figure(f
sns.countplo
plt.title('D
plt.xlabel('
```

```
  File "C:\Users\dell\AppDa
    raise KeyError(key) fro
KeyError: False
PS C:\Users\dell\Desktop\TM
8807
no. of movies 6131
no. of TV Shows 2676
c:\Users\dell\Desktop\TM Ma

Passing `palette` without a
ct.

  sns.countplot(data=df, x=
```

Figure 1: Distribution of content types on Netflix.

It is clear from the data that the vast majority of the content on Netflix was produced in the US; that is, 3690 titles were produced in the US (Including international co-productions). Of those 2818 were solely produced in the us. India is the second largest producer but far behind the us with 1046 titles with 972 of them being produced in India alone. 831 titles have no country listed as their place of origin.

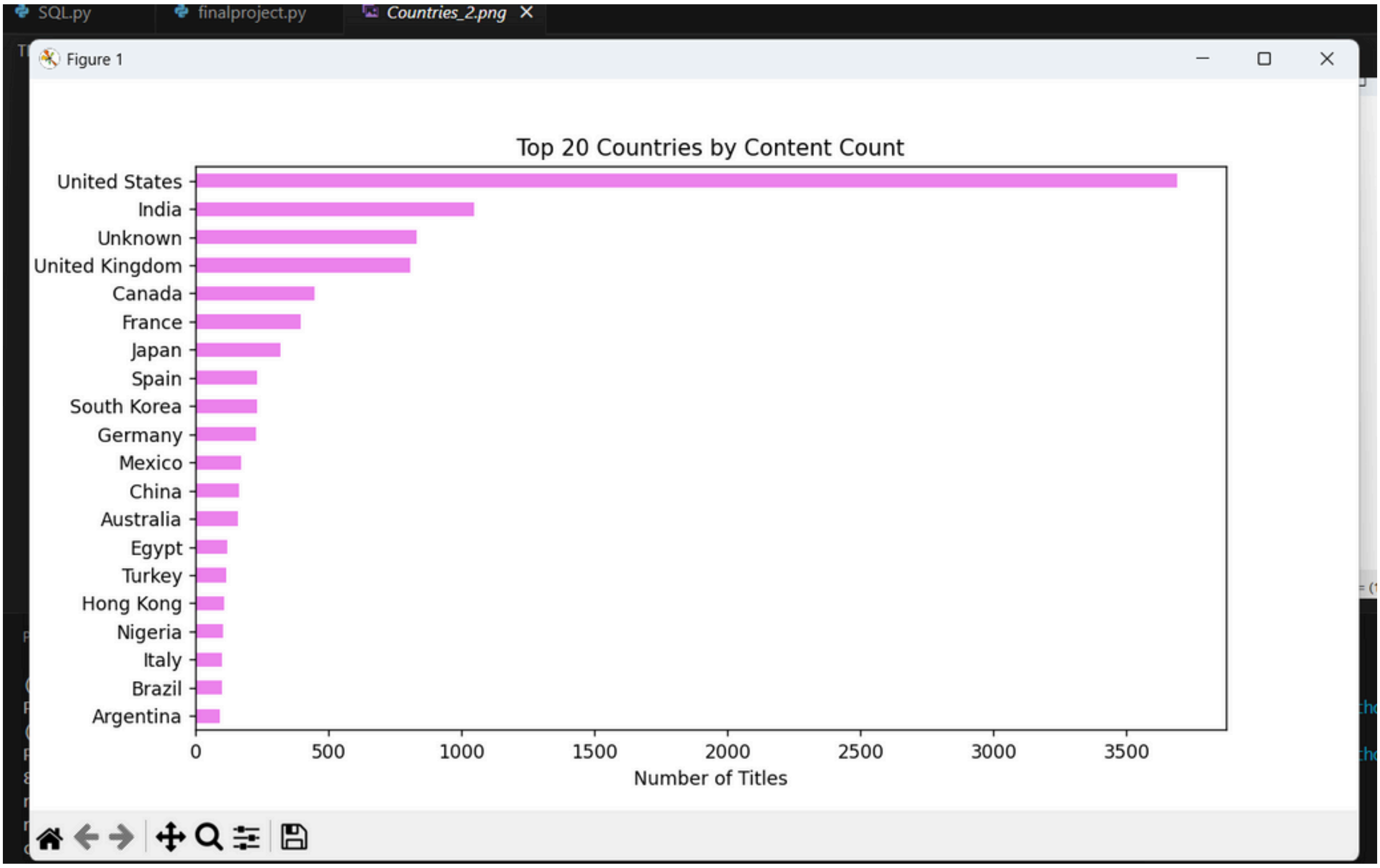


Figure 2: Top 20 Countries by content count

The most popular genre of content appears to be International movies with 2752 titles. Close behind, Drama is the second most popular genre with 2427. Comedies have 1674 titles.

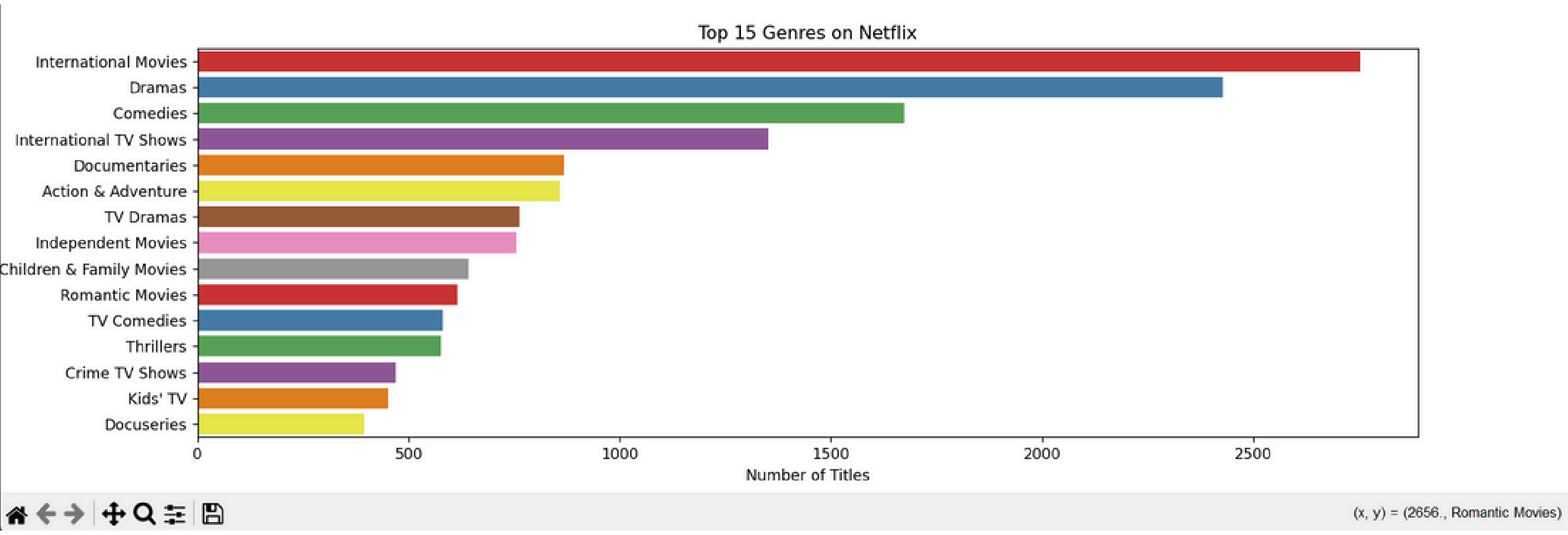
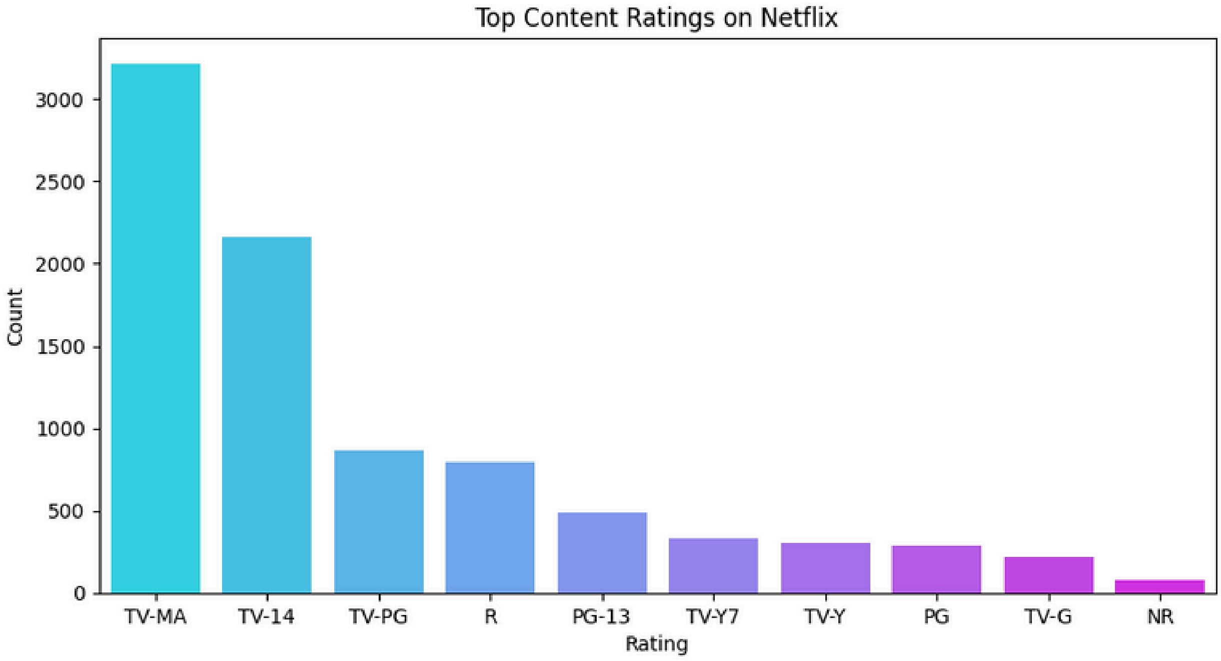


Figure 3: Top 15 Genres on Netflix



Temporal Analysis

The year with the most amount of Titles released is 2018 with Netflix hosting 1147 Titles. The earliest title was released in 1925. The graph shows a very steep increase in the no. of titles produced after the year 2010 that suddenly drops precipitously around the year 2020. This can be safely attributed to the covid-19 pandemic affecting productions all around the world.

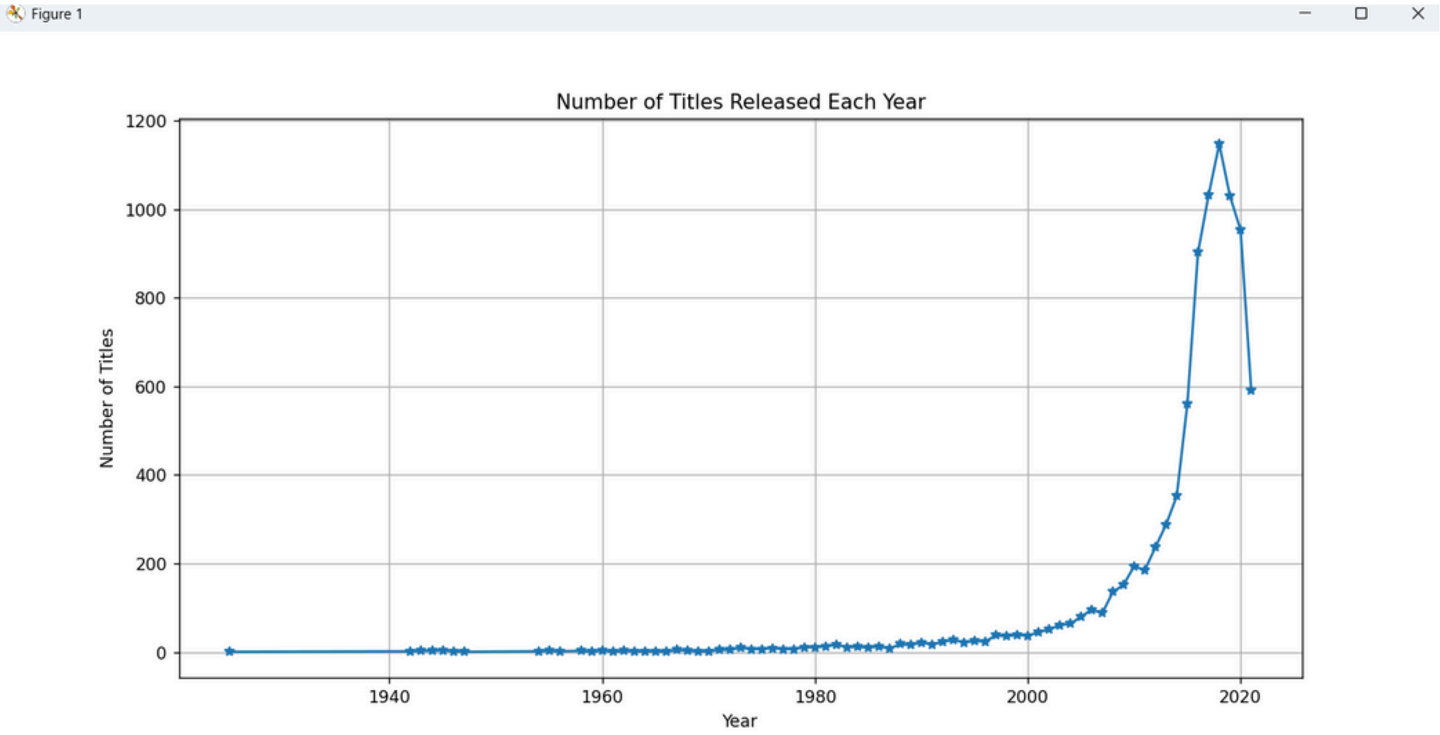


Figure 4: No. of Titles released each year

Looking at the bell curve graph, It can be concluded that on average, the duration of movies on Netflix is around 100 minutes. Most movies seem to be between 75 to 130 mins. long.

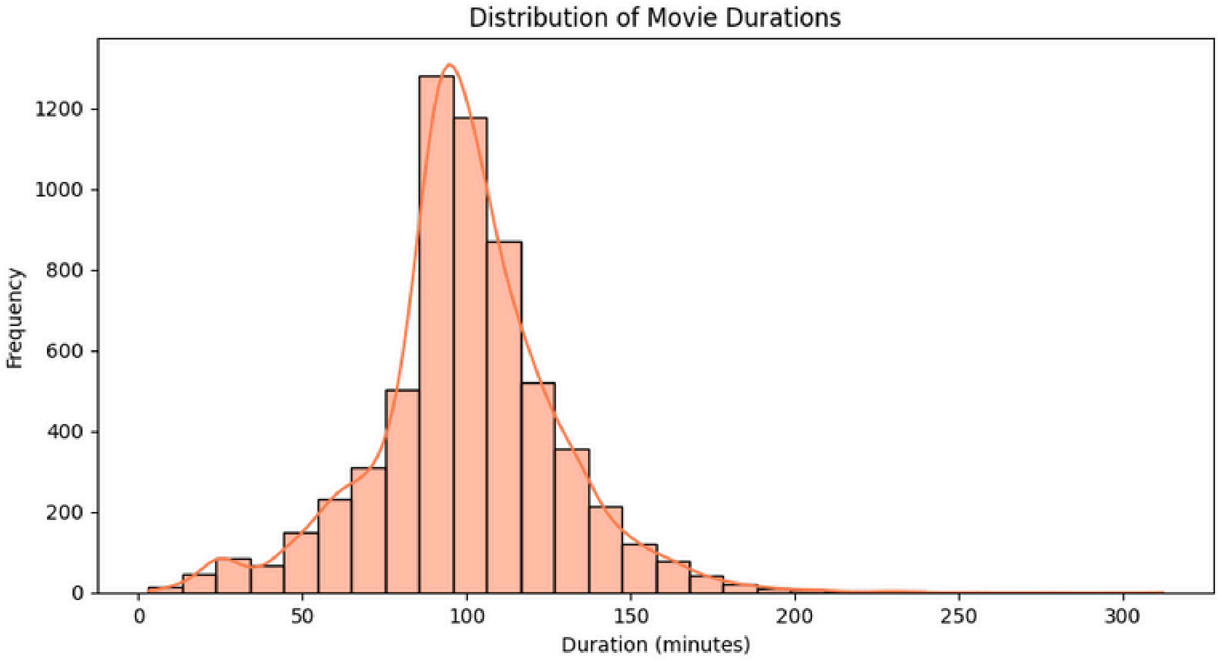


Figure 5: Graph of Probability density function of movie durations

Details Breakdown

Most content on Netflix is rated either TV-MA (intended for adults) orTV-14 (suitable for ages 14+) with 3207 & 2160 occurrences each. Given the relatively lower no of content geared toward kids, this shows that Netflix is more geared towards adults & primarily wants to attract a more mature clientele.

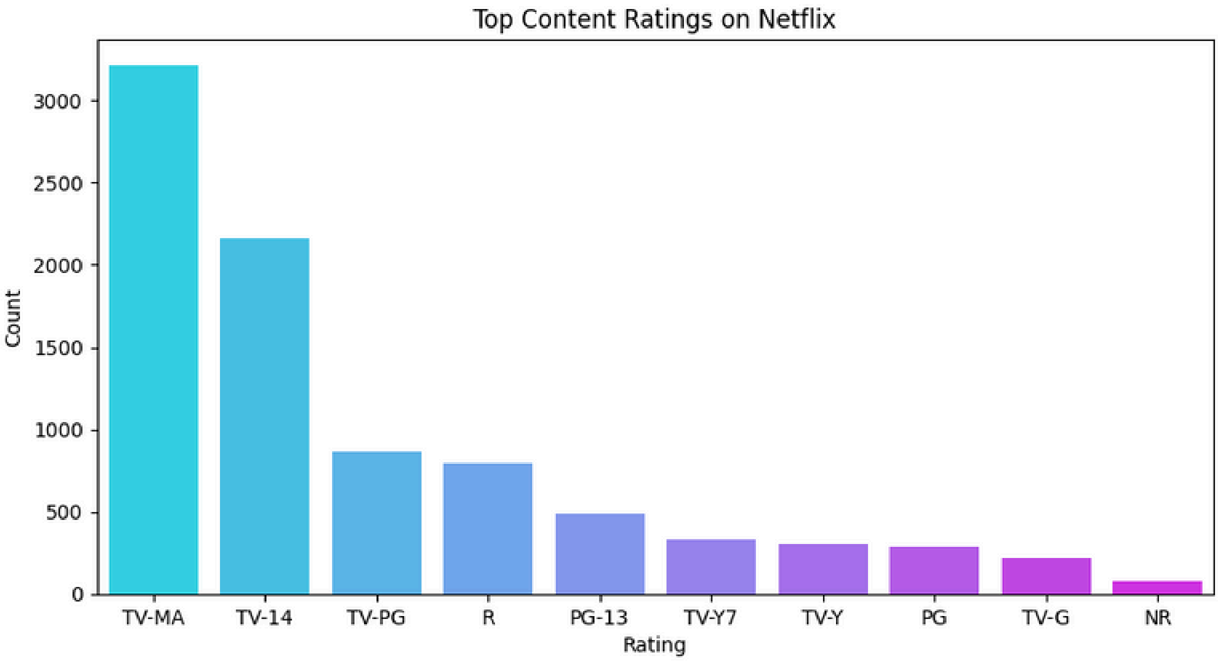


Figure 6: Most common ratings on Netflix

The data shows that the vast majority of the TV Shows hosted on Netflix only last for one season (1793 entries). This reflects a worldwide preference for shorter duration content. As seen earlier, There are more movies than TV Shows on Netflix.

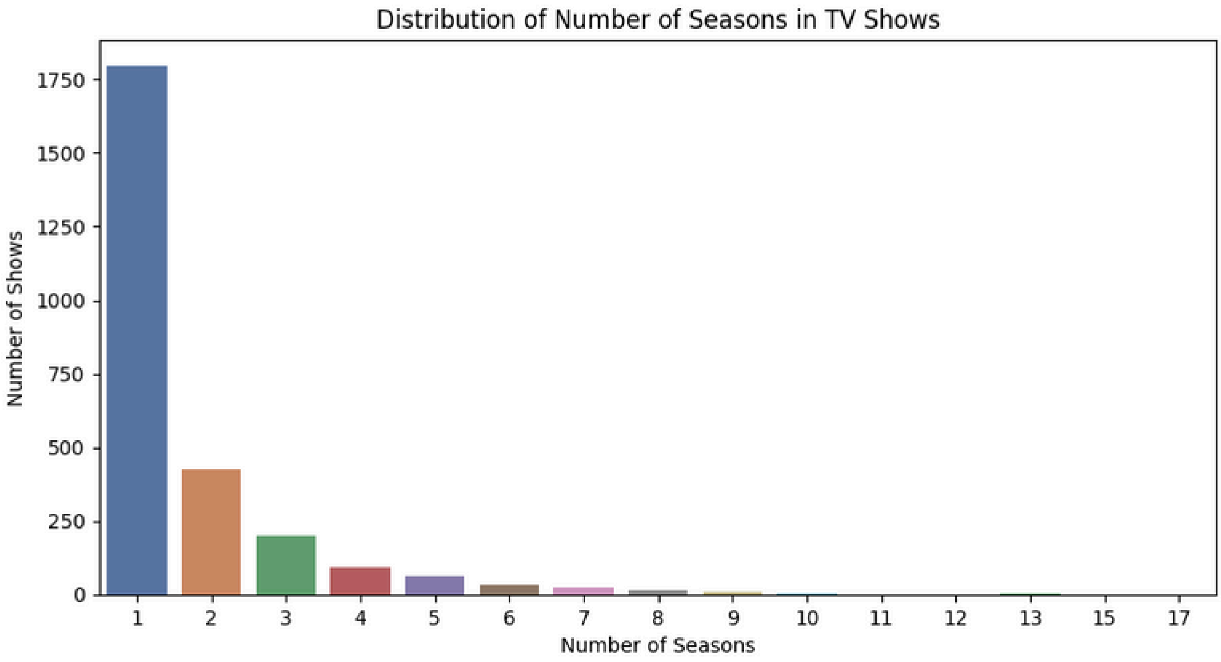


Figure 7: No. of seasons in TV shows

Despite there being more content from the United States on Netflix than from any other country by a long shot, the top ten most frequent cast members are dominated by Indian actors & actresses. 8 of the top ten spots are occupied by Indians while the other 2 spots are claimed by Japanese cast members primarily appearing in anime.

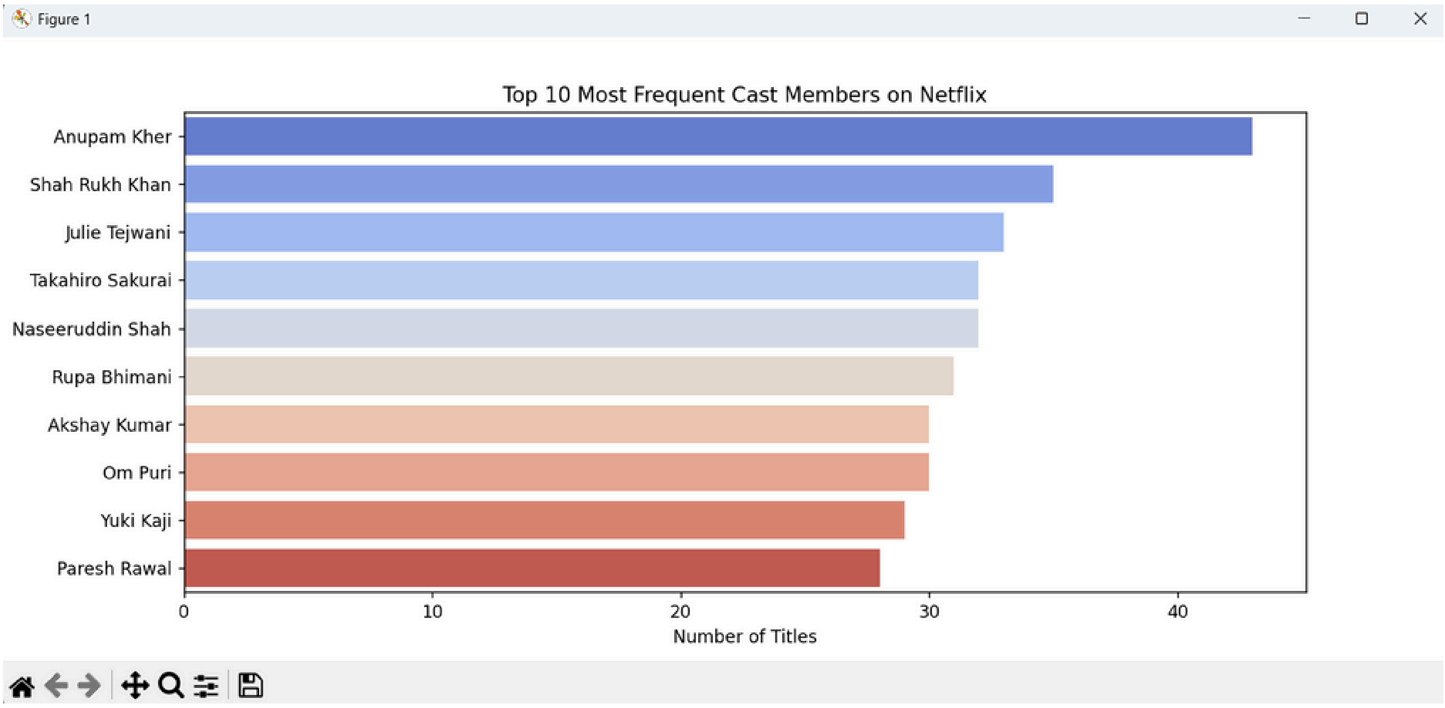


Figure 8: Top 10 most frequent Cast members on Netflix

Finally, The most popular director is Rajiv Chilaka with 22 Movies followed closely by Jan Suter with 21 movies.

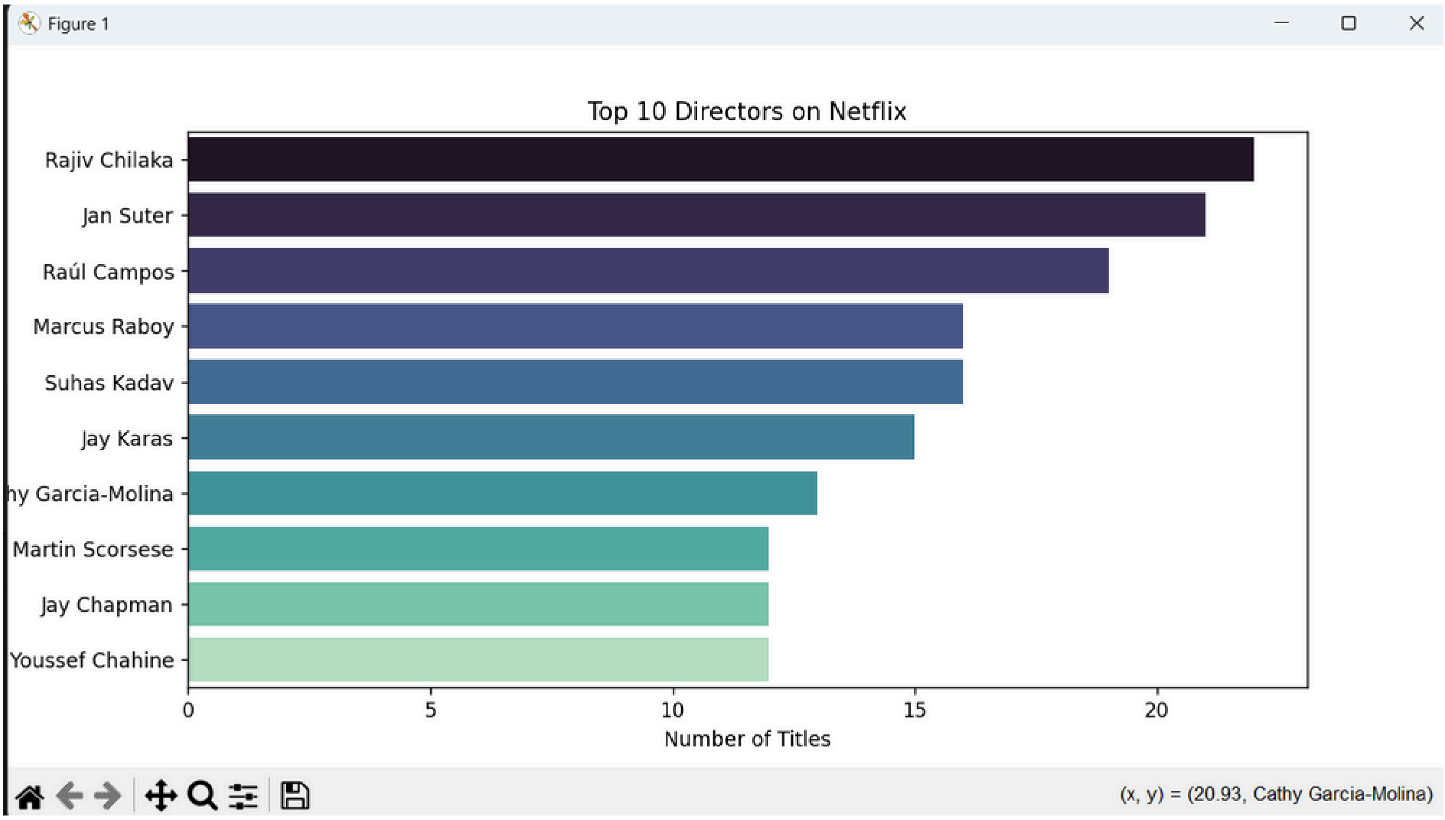


Figure 9: Top 10 Directors on Netflix

Key Insights

In summation, here are the key insights that were extracted.

This analysis of the Netflix dataset reveals several patterns in the platform's content offerings:

- Movies form a larger portion of the content compared to TV Shows.
- The most common genres include International Movies, Dramas, and Comedies, reflecting a global and diverse audience base.
- A significant number of titles were added to Netflix in recent years, with a peak around 2017–2020.
- TV Shows tend to be limited-run series, with most having 1–2 seasons.
- Movie durations are typically concentrated between 80 and 120 minutes, indicating a preference for feature-length content.

Limitations of the Dataset

No viewership or popularity metrics (e.g., watch counts, ratings, or user reviews), which limits our ability to assess actual audience preferences.

Inconsistent or missing values in fields like director, cast, and country can skew our analysis.

Suggestions for Improvement

Include data from various sources: Allows for a more clearer, unbiased view of content popularity, ratings, or critical reception (IMDb, TMDB, Google Trends etc.)

Language: Add language metadata to assess language diversity on the platform.

X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X--X

Thank you

Prepared by:

Celia Victor

On:

31st May 2025

For:

DS & ML Internship at Techmaghi