A Project Report

On

"Netflix Data Analytics – Content Trends & Visualization (Python + Power BI)"

Prepared By:

Shravani Nagawade

Email: nagawadeshravani@gmail.com

LinkedIn: linkedin.com/in/shravani-nagawade

GitHub: github.com/Shravani-222

Date of Completion: 1 June 2025

Abstract

The rapid growth of online streaming platforms, Netflix has become a global leader in digital content. This project, titled **"Netflix Data Analytics – Content Trends & Visualization (Python + Power BI)"**, explores patterns in Netflix's publicly available dataset using Python for data preprocessing and exploratory data analysis, and Power BI for interactive dashboards.

The analysis highlights the dominance of movies over TV shows, strong content presence from countries like the U.S. and India, and the popularity of genres like Drama and International Movies. It also reveals evolving content trends over the years, showcasing how data analytics tools can be effectively used to extract valuable business insights and present them through dynamic visual storytelling.

Introduction

Netflix is one of the most popular streaming platforms in the world, offering a wide range of movies, TV shows, and documentaries. With its large content library and global reach, analyzing Netflix data can help uncover interesting trends and patterns.

This project uses Python and Power BI to explore Netflix’s content based on type, country, genre, and release date. The goal is to understand how Netflix adds content over time and what kind of shows or movies are most common. This analysis helps show the power of data in making better decisions and gaining insights from real-world information.

Dataset Overview

The dataset used in this project was sourced from Kaggle and contains information about movies and TV shows available on Netflix. It includes 12 columns and over 8,800 records of content available on the platform up to the year 2021.

- show_id
- type

- title
- director
- cast
- country
- date_added
- release_year
- rating
- duration
- listed_in
- description

Objectives

- To explore and analyze Netflix's content data using Python and Power BI.
- To understand the distribution of content by type (Movies vs. TV Shows).
- To identify trends in content addition over the years.
- To analyze popular genres and country-wise content distribution.
- To visualize key insights through charts and an interactive dashboard.
- To demonstrate the application of data analytics tools in solving real-world problems.

Implementation Of Code

The implementation was done using Python (in Jupyter Notebook) for data cleaning,

1. **Importing Libraries**

```
[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

2. **Loading Dataset:**

```
[2]: df = pd.read_csv(r"C:\Users\ADMIN\Downloads\archive (7)\netflix_titles.csv")
```

```
[3]: df.head()
```

[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...

3. Data Cleaning & Preprocessing

- Checking for null values:

```
[13]: df.isnull().sum()
```

```
[13]: show_id         0
      type            0
      title           0
      director        0
      cast            0
      country         0
      date_added     10
      release_year    0
      rating          4
      duration        3
      listed_in       0
      description     0
      year_added     10
      month_added    10
      dtype: int64
```

- Filling/removing missing values and parsing dates:

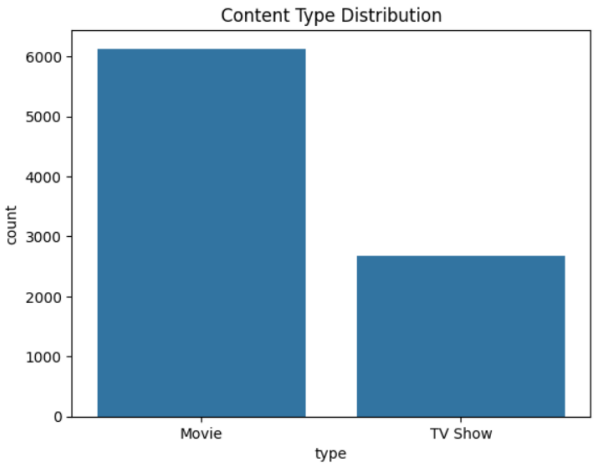
```
[4]: df['director'] = df['director'].fillna('Director Missing')
     df['cast'] = df['cast'].fillna('Cast Missing')
     df['country'] = df['country'].fillna('Country Missing')
     df['date_added'] = pd.to_datetime(df['date_added'].str.strip(), errors='coerce')
```

```
[5]: df['year_added'] = df['date_added'].dt.year
     df['month_added'] = df['date_added'].dt.month_name()
```

4. Exploratory Data Analysis (EDA)

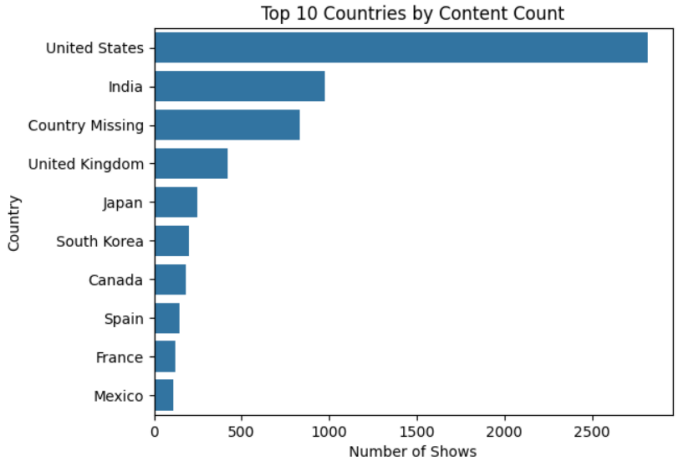
- Content Type Distribution:

[6]:
```
sns.countplot(x='type', data=df)
plt.title("Content Type Distribution")
plt.show()
```



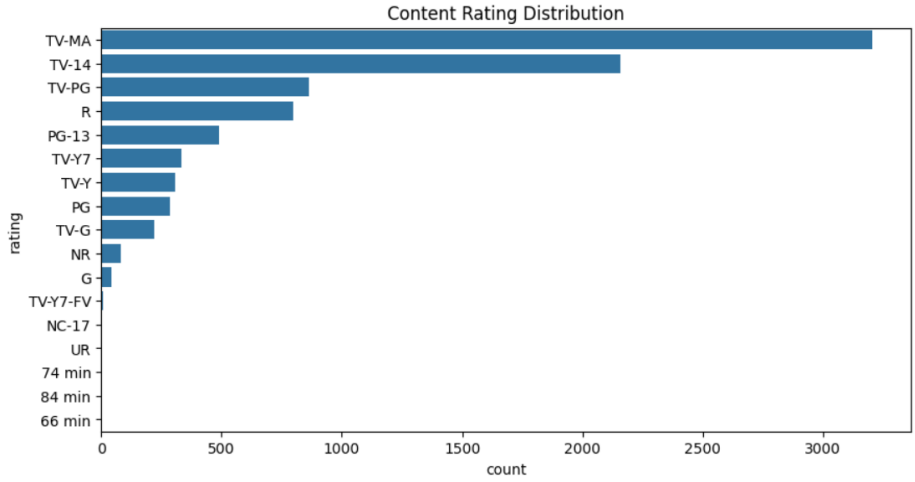
- Top 10 Countries by Content:

[7]:
```
top_countries = df['country'].value_counts().head(10)
sns.barplot(y=top_countries.index, x=top_countries.values)
plt.title("Top 10 Countries by Content Count")
plt.xlabel("Number of Shows")
plt.ylabel("Country")
plt.show()
```



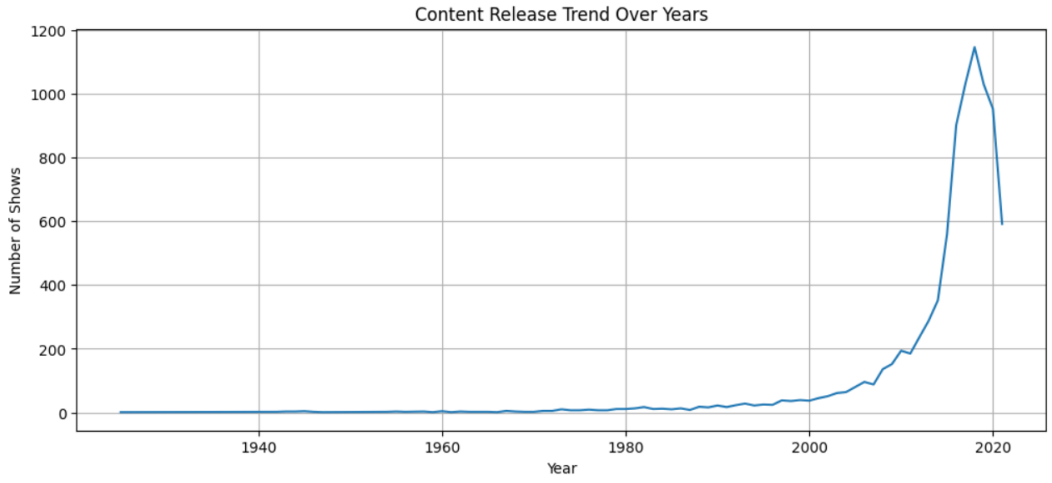
- Content Rating Distribution:

```
[8]: plt.figure(figsize=(10,5))
sns.countplot(y='rating', data=df, order=df['rating'].value_counts().index)
plt.title("Content Rating Distribution")
plt.show()
```



- Content Release Trend Over Years

```
[9]: df['release_year'].value_counts().sort_index().plot(kind='line', figsize=(12,5))
plt.title("Content Release Trend Over Years")
plt.xlabel("Year")
plt.ylabel("Number of Shows")
plt.grid(True)
plt.show()
```



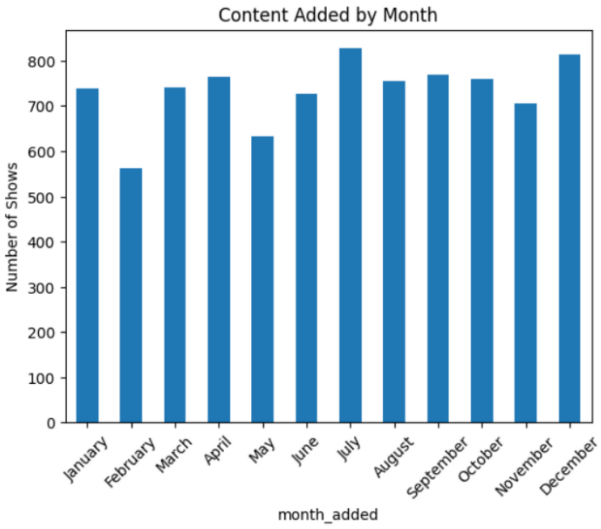
- Content Added by Month:

```
[10]: month_order = ['January', 'February', 'March', 'April', 'May', 'June',
                    'July', 'August', 'September', 'October', 'November', 'December']

df['month_added'] = df['date_added'].dt.month_name()

df['month_added'] = pd.Categorical(df['month_added'], categories=month_order, ordered=True)

df['month_added'].value_counts().loc[month_order].plot(kind='bar')
plt.title("Content Added by Month")
plt.xticks(rotation=45)
plt.ylabel("Number of Shows")
plt.show()
```



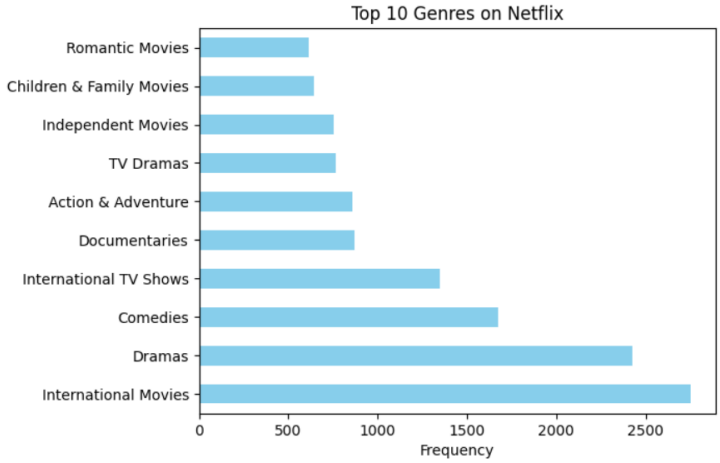
5. Genre (Category) Analysis

- Top 10 Genres on Netflix:

```
[11]: from collections import Counter

all_genres = ','.join(df['listed_in']).split(',')
genre_freq = pd.Series(Counter([g.strip() for g in all_genres])).sort_values(ascending=False).head(10)

genre_freq.plot(kind='barh', color='skyblue')
plt.title("Top 10 Genres on Netflix")
plt.xlabel("Frequency")
plt.show()
```



6. Exporting Cleaned Data (for Power BI)

```
[12]: df.to_csv("C:\\Users\\ADMIN\\Downloads\\archive (7)\\cleaned_netflix_titles.csv", index=False)
```

7. Power BI Dashboard Implementation

After cleaning and exporting the dataset using Python, Power BI was used to create interactive dashboards for better visualization and storytelling.

Steps Followed:

1. Importing Data:
    - The cleaned dataset (cleaned_netflix_titles.csv) was loaded into Power BI.

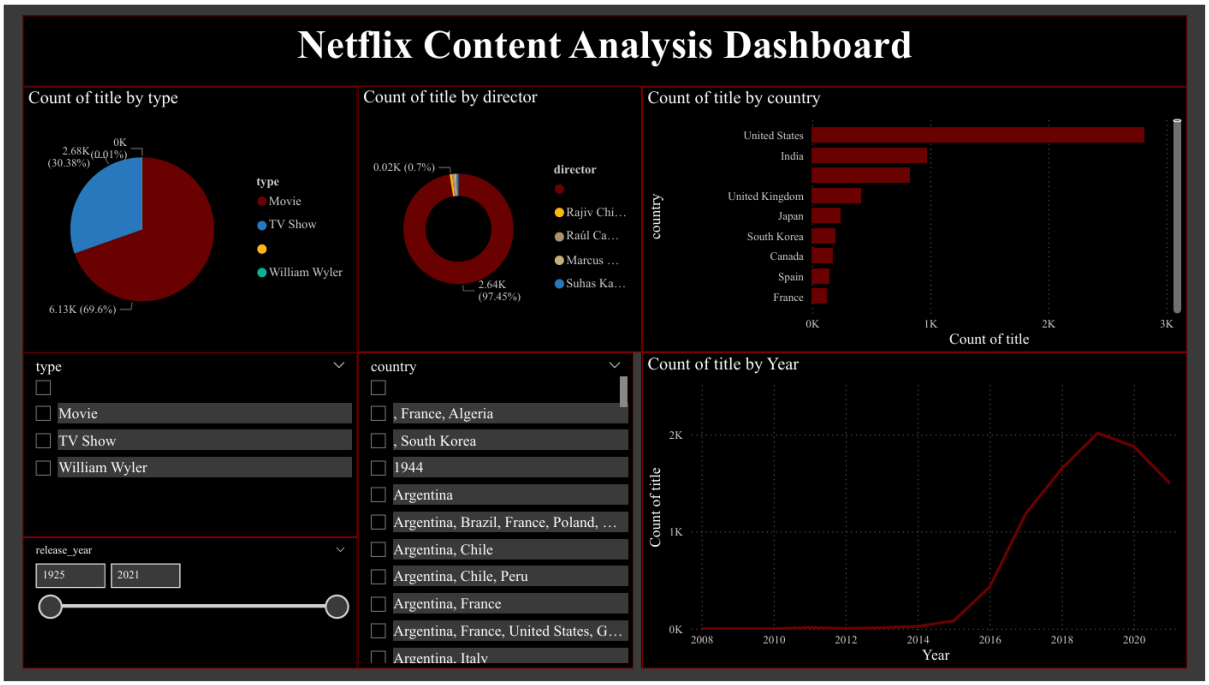
2. Data Transformation:
    - Column types were formatted (e.g., dates, categories).

3. Dashboard Design:
    - Multiple visualizations were created to represent:
        - Total number of movies vs. TV shows
        - Content added over the years
        - Country-wise content availability
        - Most popular genres
        - Ratings distribution

4. Interactivity:
    - Slicers and filters were added to allow users to filter data by:
        - Type (Movie/TV Show)
        - Year of addition
        - Country or genre

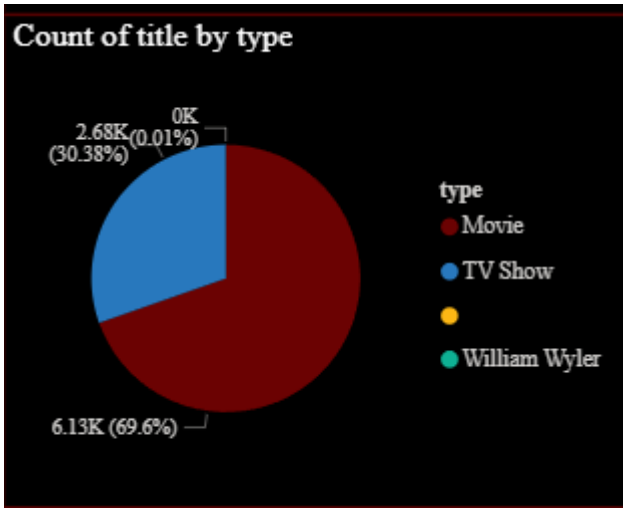
5. Insights Gained:
    - Users can interactively explore how content trends change over time and across regions, helping to understand Netflix's strategy and audience targeting.

- **Power BI Dashboard: Netflix Content Analysis Dashboard**



This dashboard gives a high-level overview of content types, popular countries, genre distribution, and content additions over time.

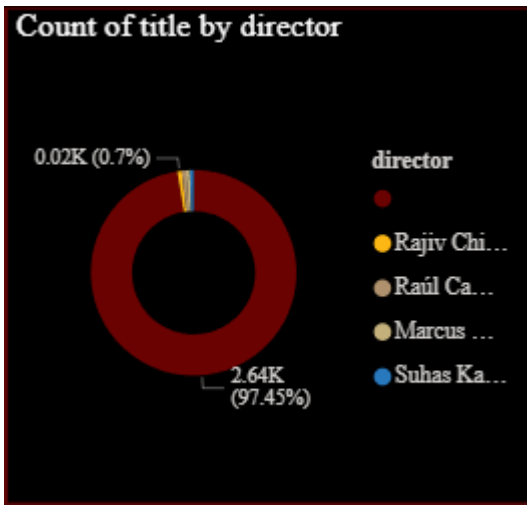
Figure 1: Content Type Distribution – Movies vs TV Shows



Netflix has more movies than TV shows, showing a clear focus on movie content.

Chart Type**:** Pie Chart – used to show the proportion of Movies vs TV Shows in Netflix's catalog.

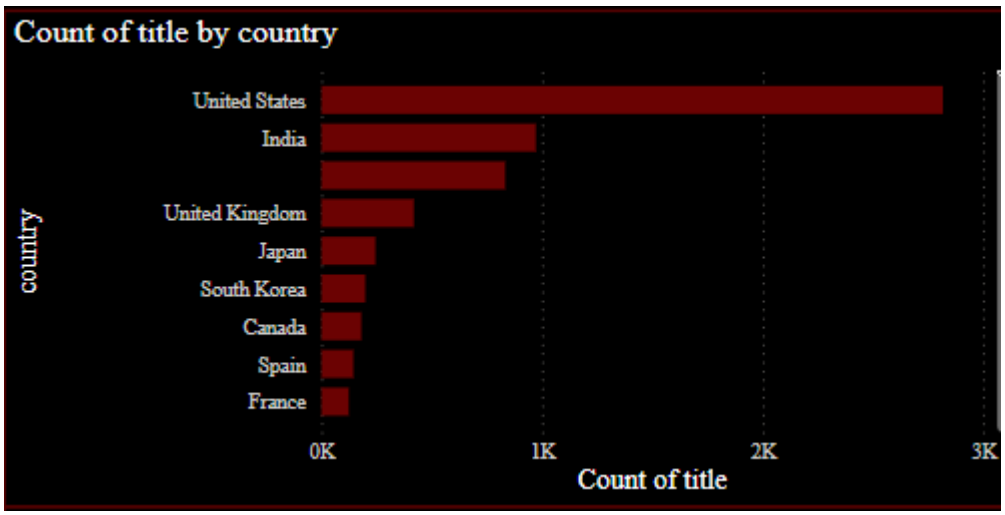
Figure 2: Count of Titles by Director



This donut chart displays the top 5 directors with the highest number of titles on Netflix. It shows who has contributed the most content.

Chart Type: Donut Chart – used to show distribution among the top 5 directors.

Figure 3: Count of Titles by Country



This bar chart highlights the top countries contributing content to Netflix. The United States and India lead in the number of titles.

Chart Type: Bar Chart – used to compare the number of titles produced by each country.

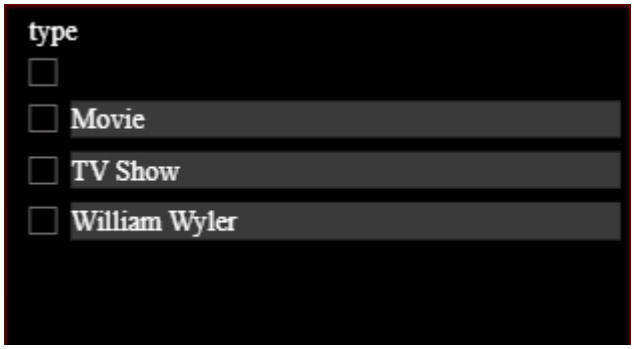
Figure 4: Count of Titles by Year

This line chart shows how many titles were added to Netflix each year. It helps visualize the platform's growth over time.
Chart Type: Line Chart – used to display trends in content addition by year.

Interactive Slicers Used

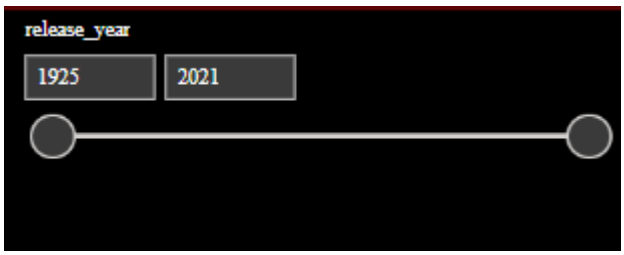
The dashboard includes slicers for Type, Country, and Release Year. These allow users to filter the visuals and explore Netflix content based on specific categories.

- Type – Filter by Movies or TV Shows



- Country – Filter content based on production country

- Release Year – View titles released in a specific year



Results and Findings

- Movies dominate Netflix's content, making up the majority compared to TV shows.
- The United States and India are the top content-producing countries on the platform.
- Drama and International Movies are the most common genres.
- The number of titles added has increased steadily over the years, showing Netflix's global expansion.
- A few directors have contributed multiple titles, with the top 5 having the highest presence.
- Interactive slicers in the dashboard help users explore content trends by type, country, and release year.

Discussion

The analysis of Netflix's dataset revealed several key patterns. Movies make up the majority of the content, suggesting a focus on quick-consumption entertainment. The dominance of countries like the United States and India indicates their strong role in content production.

Popular genres such as Drama and International Movies show user preferences across different regions.

Using Python for preprocessing and Power BI for visualization enabled efficient handling and clear presentation of the data. The slicers (Type, Country, Release Year) added interactive flexibility, helping users drill down into specific insights based on their interests.

Conclusion

This project demonstrated how data analytics tools can turn raw data into meaningful insights. By analyzing Netflix content, we understood platform trends, content distribution, and genre popularity. The interactive dashboard made the insights more accessible and engaging.

Such analysis not only helps viewers understand platform trends but also aids companies in making informed decisions based on viewer behavior and content performance.