Social Media Analysis Using Python

Netflix Data Analysis

Table Of Contents

- Aim

- Introduction

- Methodology & Data Description

- Implementation

- Results

- Conclusion

Aim

This project aims to analyze a Netflix dataset to understand trends in the platform's content offerings. Key goals include:

1. Analyzing genre distribution to identify popular types of content.

2. Examining the countries most frequently producing content for Netflix.

3. Tracking the number of releases over time to detect shifts in release volume.

4. Exploring ratings and duration to determine characteristics of popular shows.

5. Investigating monthly release frequency patterns.

6. Identifying the most prolific directors by the number of titles.

By uncovering these insights, we hope to reveal patterns in Netflix's content strategy and viewer preferences over time.

Introduction

In recent years, Netflix has transformed the entertainment industry, offering a vast library of movies, series, and documentaries to a global audience. With an extensive variety of genres, languages, and formats, Netflix has revolutionized how we consume media. Understanding the trends in Netflix's content can provide valuable insights into viewer preferences, popular genres, and how Netflix's content strategy has evolved over time.

This project aims to analyze a dataset of Netflix shows to uncover patterns and insights within its offerings. Specifically, we'll examine the distribution of genres, the primary countries producing content, and the frequency of releases over time. Additionally, we'll analyze ratings, show duration, monthly release trends, and the productivity of directors based on the number of titles.

By delving into these aspects, we hope to identify how Netflix has tailored its content over time to meet audience demands, potentially revealing broader trends in the

streaming industry. This analysis will offer a data-driven perspective on the strategies that have made Netflix a leader in digital streaming

.

Methodology

1. Data Loading and Cleaning :

  - Load the dataset and review columns such as category, title, director, cast, country, release date, rating, duration, type, and description.

  - Clean the data by handling missing or null values, standardizing formats, and removing duplicates.

2. Exploratory Data Analysis (EDA) :

  - Genre and Category Analysis : Investigate the distribution of content types and genres to understand popular categories.

  - Country and Cast Analysis : Identify key content-producing countries and analyze frequently featured cast members.

  - Release Patterns : Examine the timeline of releases by year and month to detect trends in release frequency.

  - Ratings and Duration : Analyze ratings and duration to explore what audiences may find engaging.

  - Director Contributions : Determine the most prolific directors based on title counts.

3. Visualization :

  - Use visualizations like bar charts, line graphs, and word clouds to illustrate genre popularity, release patterns, and prominent directors or cast members.

Data Description:

The dataset provides detailed information on Netflix shows, including:

- **Category** : Genre or content classification.

- **Title** : Name of the show.

- **Director** : Name of the director.

- **Country** : Country of origin.

- **Release Date** : Date when the show was released on Netflix.

- **Rating** : Age rating (e.g., TV-MA, PG-13).

- **Duration** : Length of content (runtime for movies, seasons for TV shows).

- **Type** : Type of content (movie or TV show).

- **Description** : Brief summary of the content.

This methodology and dataset structure will guide the analysis, providing insights into Netflix's content evolution and viewer preferences.

Implementation

Based on your code snippets, here's a complete implementation outline for your Netflix dataset analysis project. This will include loading the data, cleaning it, and conducting various analyses to visualize trends in Netflix's content.

```python
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt
```

• **Load the dataset**

df = pd.read_csv(r"C:\Users\tvais\OneDrive\Desktop\modified_dataset.csv")

• **Display the first few rows and basic information about the dataset**

print(df.head())

print(df.info()

• **Data Cleaning**

• **Check for missing values and handle them if necessary**

print(df.isnull().sum())

• **Assuming 'Release_Date' and other essential columns should not have missing values**

df.dropna(subset=['Title', 'Release_Date'], inplace=True)

• **Convert 'Release_Date' to datetime format**

df['Release_Date'] = pd.to_datetime(df['Release_Date'])

• **1. Category Distribution**

category_counts = df['Category'].value_counts()

plt.figure(figsize=(8, 8))

plt.pie(category_counts, - Cast : Main actors featured in the show.

labels=category_counts.index, autopct='%1.1f%%', startangle=140)

plt.title("Category Distribution")

plt.show()

• **2. Top 10 Genre Distribution**

df["Type"].value_counts().head(10).plot(kind="bar", title="Top 10 Genre Distribution")

plt.xlabel("Type")

```
plt.ylabel("Count")
```

```
plt.show()
```

• 3. Top 10 Content Producing Countries

```
df["Country"].value_counts().head(10).plot(kind="bar", title="Top 10 Content Producing
Countries")
```

```
plt.xlabel("Country")
```

```
plt.ylabel("Count")
```

```
plt.show()
```

• 4. Number of Releases Over Time

```
df["Year"] = df["Release_Date"].dt.year
```

```
df["Year"].value_counts().sort_index().plot(kind="line", title="Number of Releases Over
Time")
```

```
plt.xlabel("Year")
```

```
plt.ylabel("Number of Releases")
```

```
plt.show()
```

• 5. Rating Distribution

```
df["Rating"].value_counts().plot(kind="bar", title="Rating Distribution")
```

```
plt.xlabel("Rating")
```

```
plt.ylabel("Count")
```

```
plt.show()
```

• 6. Monthly Release Frequency

```
df['release_month'] = df['Release_Date'].dt.month
```

```
monthly_releases = df['release_month'].value_counts().sort_index()
```

```
monthly_releases.plot(kind='line', title="Monthly Release Frequency", figsize=(12, 6))

plt.xlabel("Month")

plt.ylabel("Number of Releases")

plt.show()
```

• **7. Top 10 Directors by Number of Titles**

```
plt.figure(figsize=(10, 6))

top_directors = df["Director"].value_counts().head(10)

top_directors.plot(kind='bar')

plt.title("Top 10 Directors by Number of Titles")

plt.xlabel("Director")

plt.ylabel("Number of Titles")

plt.legend(["Count of Titles"], title="Legend")

plt.show()
```

Explanation of Each Step:

1. Data Loading : The dataset is loaded from a specified path using pd.read_csv().

2. Data Cleaning : The script checks for and drops missing values in essential columns (e.g., Title, Release_Date). It also converts the Release_Date column to a datetime format.

3. Category Distribution : A pie chart visualizes the distribution of categories in the dataset.

4. Top 10 Genre Distribution : A bar chart shows the most frequent genres available in the dataset.

5. Top 10 Content Producing Countries : Another bar chart visualizes the top countries producing content for Netflix.

6. Number of Releases Over Time : A line chart illustrates the trend of releases over the years.

**7. Rating Distribution** : A bar chart displays the distribution of ratings across shows.

**8. Monthly Release Frequency** : A line chart visualizes the frequency of releases by month.

9. Top 10 Directors : Finally, a bar chart highlights the directors with the most titles in the dataset.
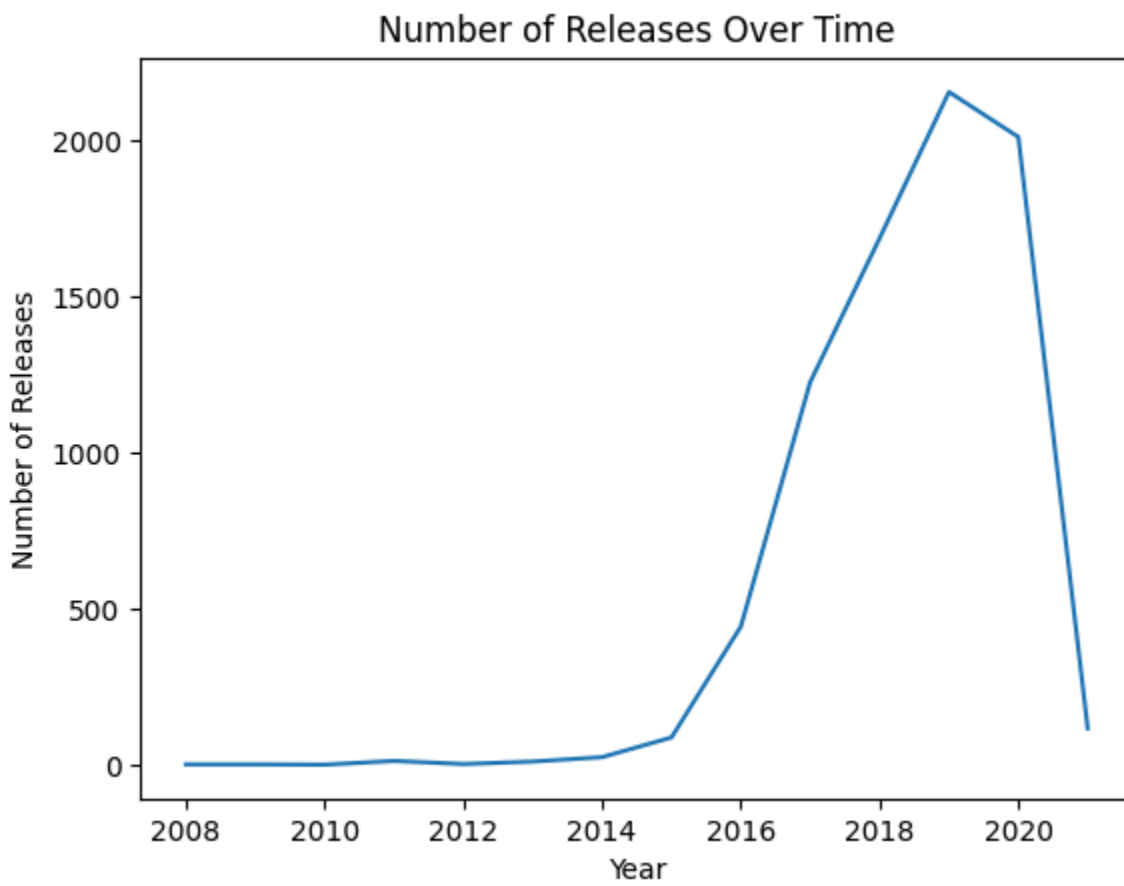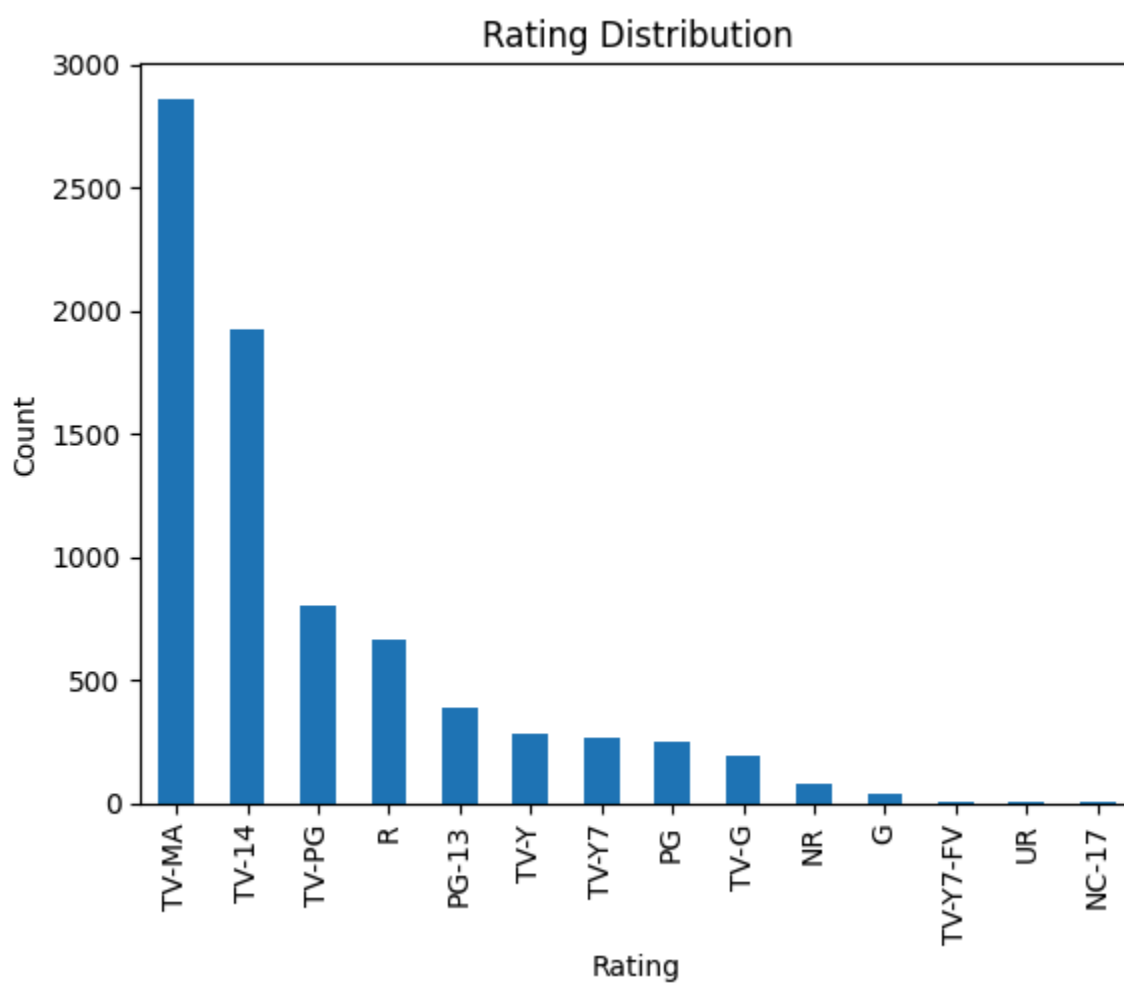
Result

Category Distribution

Top 10 Genre Distribution

Top 10 Content Producing Countries

Number of Releases Over Time

Rating Distribution

Top 10 Directors by Number of Titles
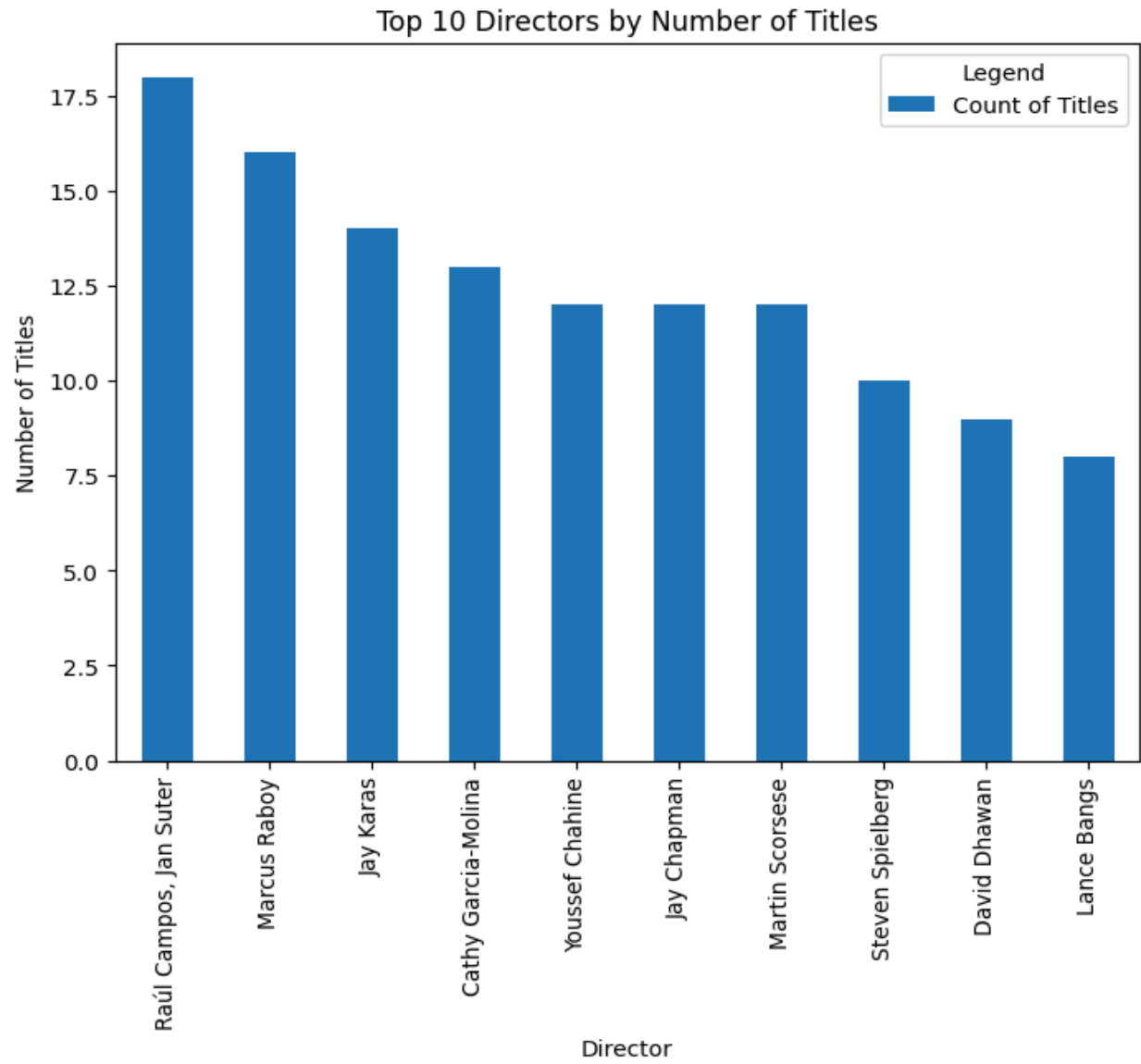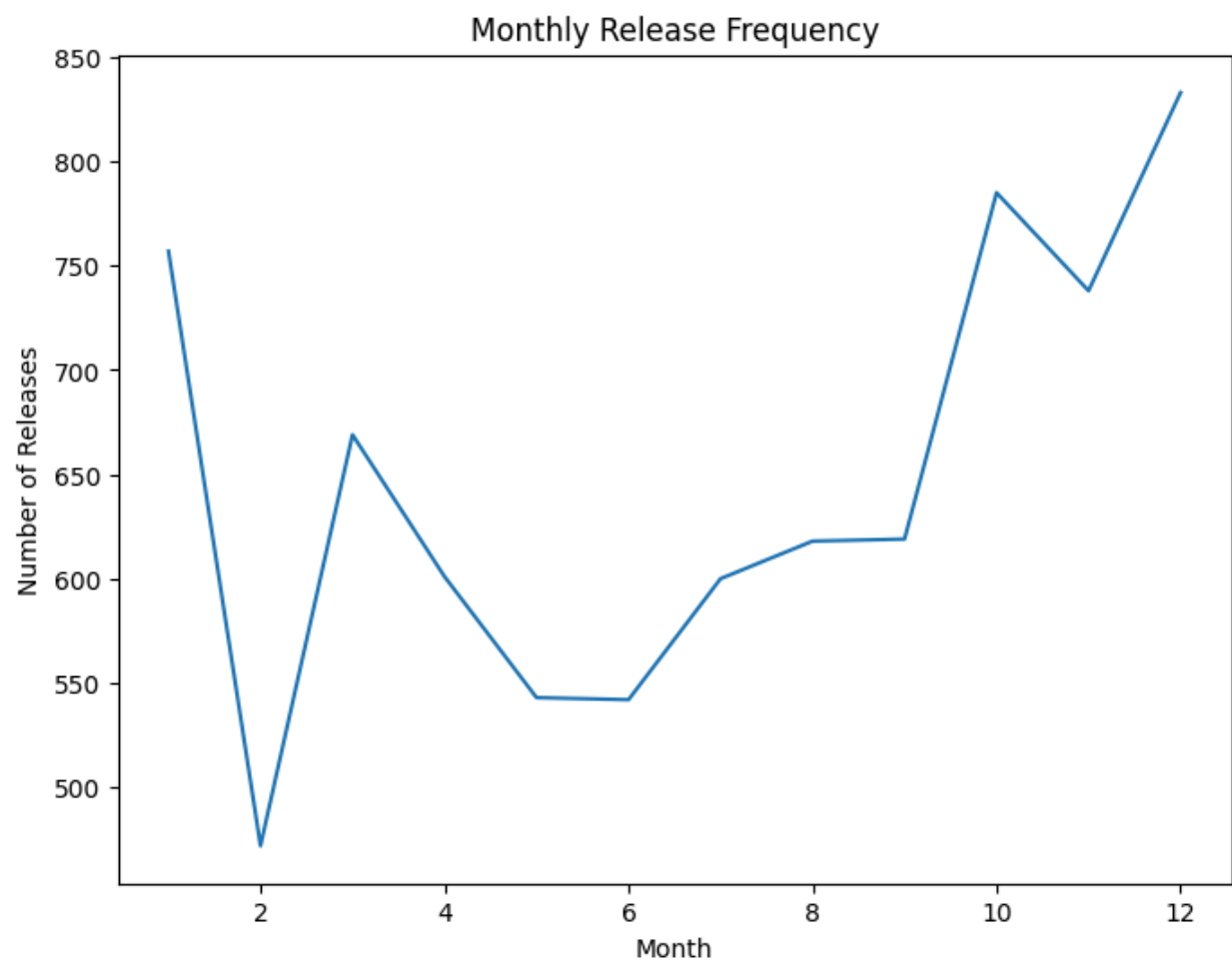
Monthly Release Frequency
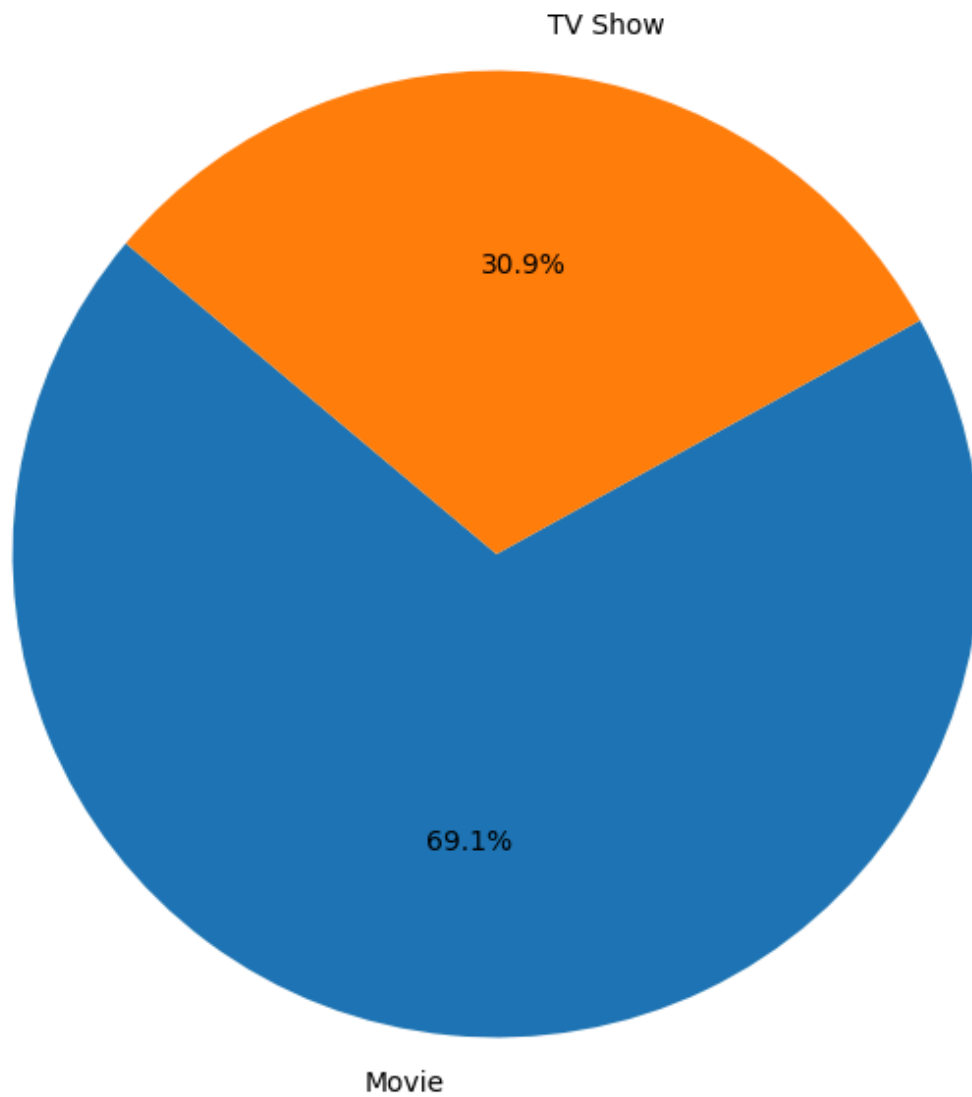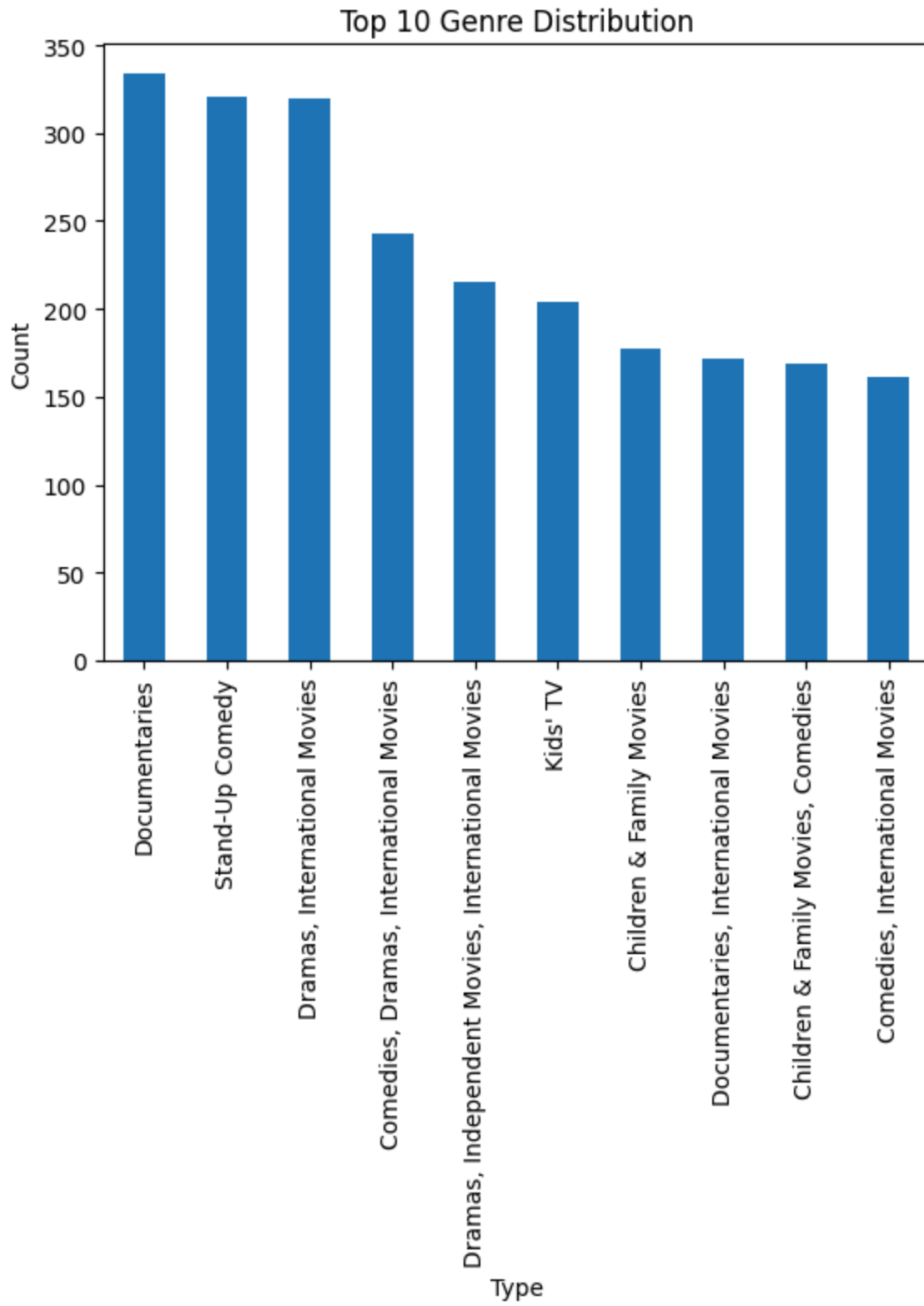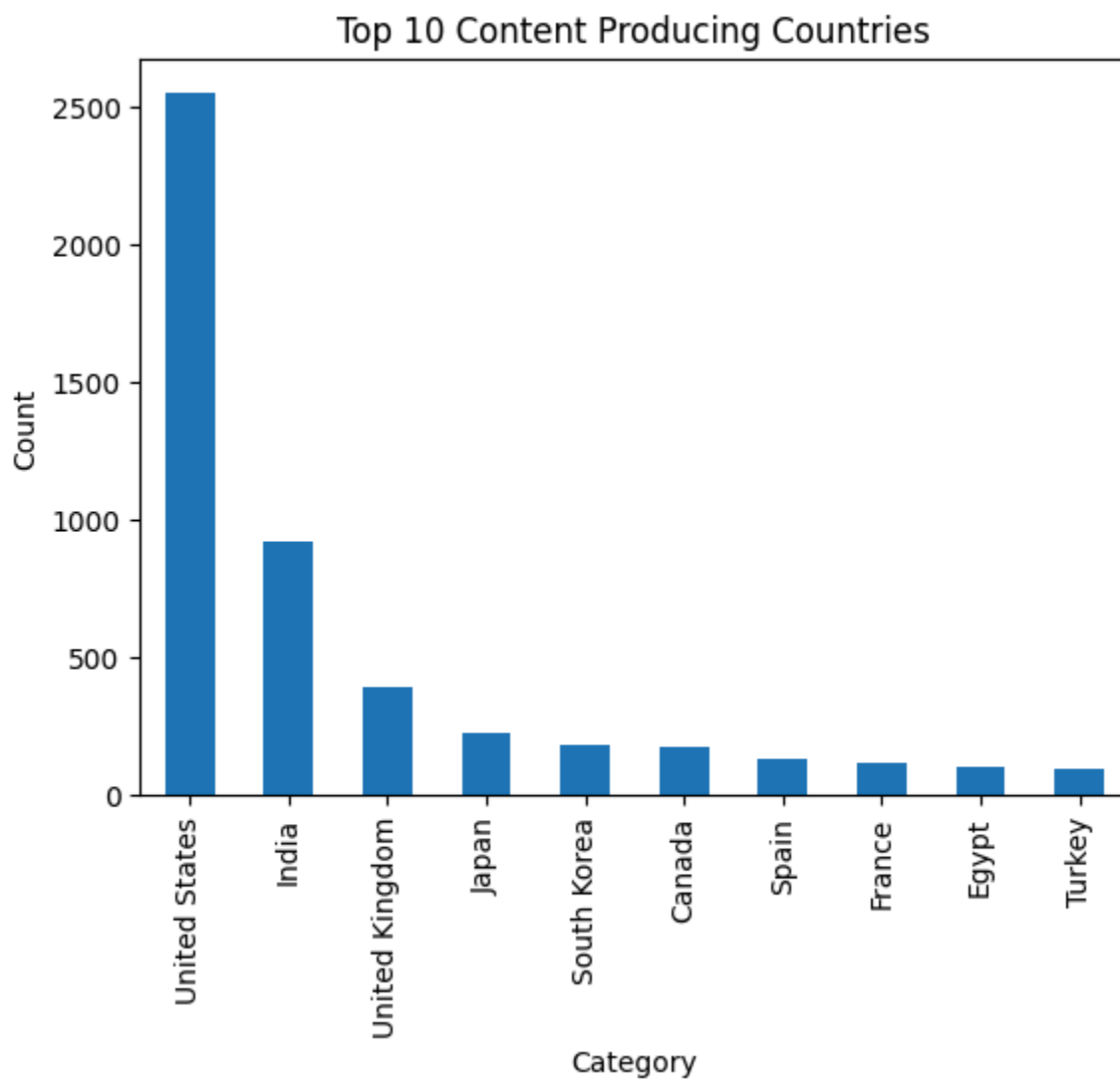
Conclusion

This project analyzed a dataset of Netflix shows to uncover trends and insights about content offerings and viewer preferences. Through various analyses, we explored key aspects such as genre distribution, country of origin, release patterns, ratings, and director contributions.

Key findings include:

1. Genre Trends : We identified which genres are most popular in Netflix's library, giving insights into what viewers prefer.

2. Content Production : The analysis revealed the top countries producing content, showing Netflix's global reach and diverse programming.

3. Release Patterns : We observed an increase in the number of releases over time, indicating Netflix's response to growing demand for streaming.

4. Ratings Analysis : Examining ratings helped us understand audience engagement and the perceived quality of shows.

5. Director Contributions : Identifying top directors highlighted key figures contributing to Netflix's success.

Overall, this analysis provides valuable insights into Netflix's content strategy and viewer preferences. Future research could explore deeper questions, such as the

impact of specific genres on viewer engagement or audience demographics, to enhance our understanding of the streaming industry.