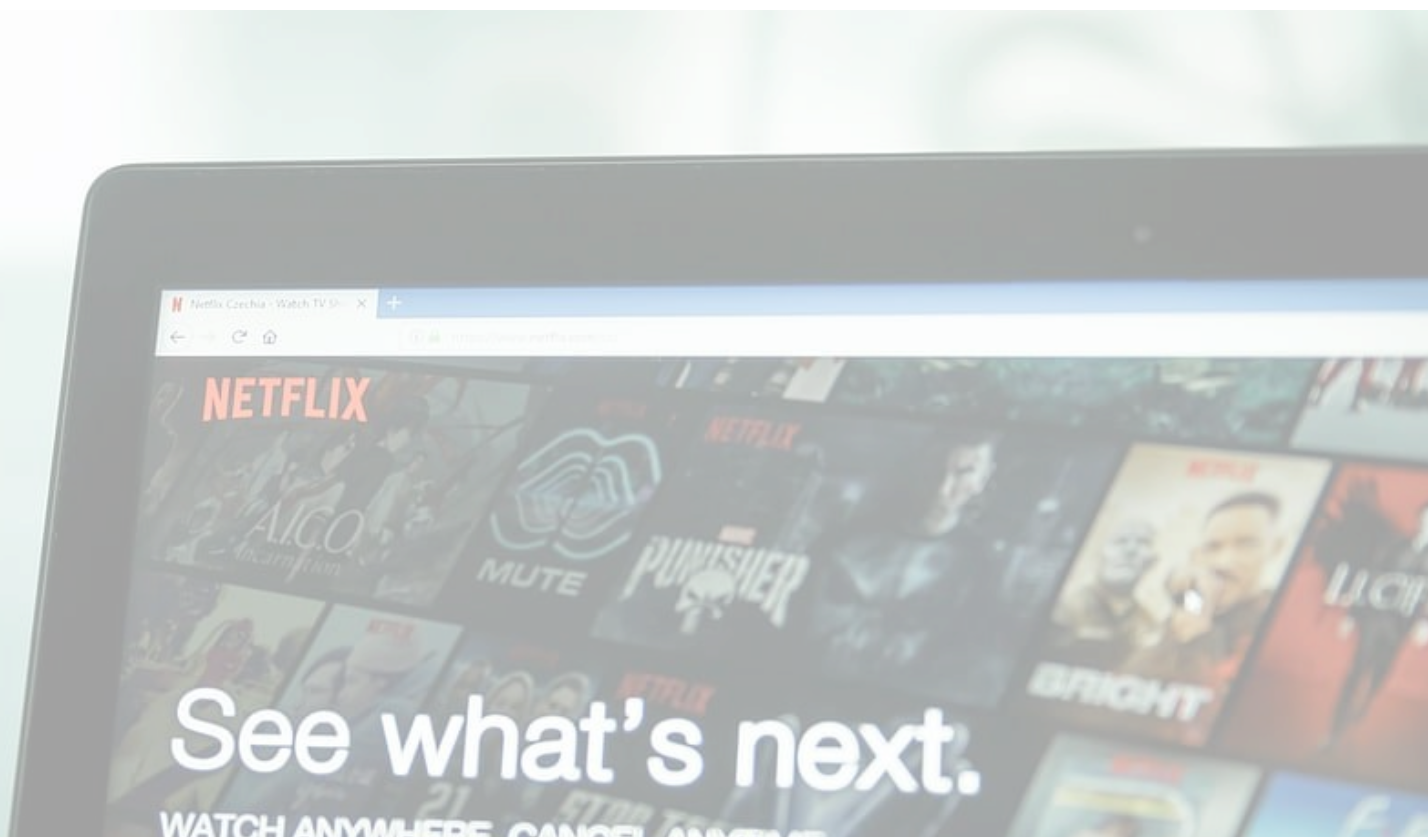


NETFLIX

Business Case **Solution**

By Munmun Mishra



Introduction

- - - - x

Netflix, the epitome of streaming platforms, finds itself at the nexus of art and science in the ever-changing entertainment world. Guided by data-driven insights akin to stars in the night sky, Netflix crafts a symphony of content and global expansion.

As constellations navigate sailors through the dark sea, data guides Netflix's decisions. They analyse viewer preferences and emerging trends, connecting the dots to curate content that transcends boundaries. This approach transforms data into art, creating narratives that captivate a diverse global audience.

Netflix's unique blend of storytelling and analytics ensures it remains a pioneering force in the entertainment industry. By staying attuned to the shifting tastes of its viewers, Netflix not only retains its place in the limelight but also continues to illuminate the path forward for the ever-evolving landscape of digital entertainment.

“

Words(Data) are, in my not-so-humble opinion, our most inexhaustible source of magic(insights).

-Albus Dumbledore, from J.K. Rowling's Harry Potter series.

”



Objectives and Scope

- - - - x

To fulfil Netflix's vision, our analysis sets out to achieve the following objectives:

- Analyse the Netflix dataset, which contains comprehensive information about movies and TV shows on the platform.
- Provide meaningful insights to help Netflix understand the content landscape and how it has evolved.
- Identify relationships between variables in the dataset to inform strategic decisions.

This analysis focuses on the Netflix dataset, a rich source of information about the platform's content offerings. Our scope includes the following aspects:

- **Content Distribution:** We will explore the distribution of content types (movies and TV shows) and provide insights into their prevalence on the platform.
- **Temporal Changes:** We will investigate how the release years of movies and TV shows have evolved and whether there are any notable trends.
- **Relationship Analysis:** We will identify and explore relationships between variables within the dataset to understand factors influencing content availability.



Data Acquisition and Preprocessing

- - - - X

The Netflix dataset was procured from the provided link and loaded into the Google Colab environment, a powerful tool for data analysis. This dataset contains a wealth of information about Netflix's content library, including details on titles, directors, cast, countries of production, release years, ratings, durations, genres, and descriptions.

Before the analysis, we conducted data preprocessing to ensure the dataset was ready for exploration. The following steps were taken:

Data Type Conversions: To enhance data efficiency, the 'Type' and 'Country' columns were converted to the 'category' data type.

```
# Convert 'Type' and 'Country' columns to 'category' data type
netflix_data['type'] = netflix_data['type'].astype('category')
netflix_data['country'] = netflix_data['country'].astype('category')
```

Missing Value Check: We meticulously examined the dataset for missing values and found it was remarkably clean, with no missing entries. This allowed us to proceed with our analysis without imputations or data fillings.

```
# Handle missing data (replace with 0, for example)
netflix_data['duration'].fillna(0, inplace=True)

# Drop duplicates
netflix_data.drop_duplicates(inplace=True)
```



Data Summary

- - - - x

The Netflix dataset is a treasure trove of information, offering a comprehensive view of the platform's vast content library. It comprises a diverse array of attributes, each a unique facet that can potentially reveal valuable insights.

These attributes encompass:

- **'Show_id,'** providing a unique identifier for each content piece
- **'Type',** classifying them into Movies or TV Shows
- **'Title'** denoting the name of the content
- **'Director,'** highlighting the creative minds behind the scenes
- **'Cast'** features the talented actors
- **'Country'** signifies the origin of the production
- **'Date_added'** records the date of introduction onto the platform
- **'Release_year'** divulges the content's birth year
- **'Rating'** illuminates the audience target
- **'Duration'** quantifies the runtime or the number of seasons
- **'Listed_in'** captures the genre
- **'Description'** offers a concise summary

The dataset's meticulous preparation ensures its readiness for profound exploration, promising to unveil the hidden narratives and trends within the world of Netflix's content offerings.

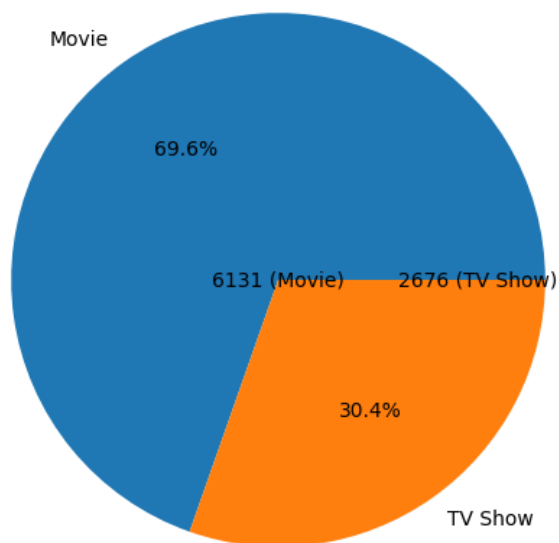


Data Visualisation

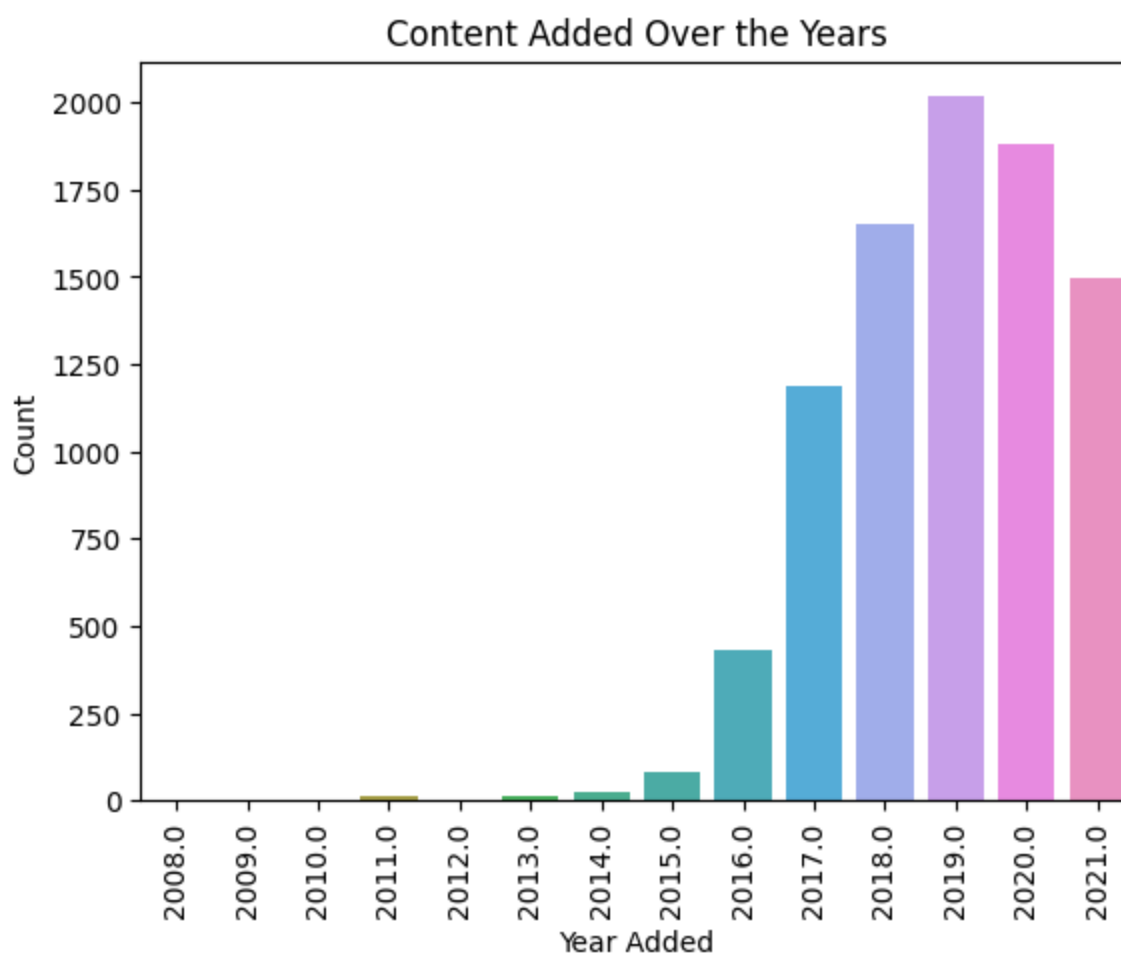
-----X

```
#Visualize the content distribution by type
type_counts = netflix_data['type'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(type_counts, labels=type_counts.index, autopct='%1.1f%%')
plt.title('Distribution of Movies and TV Shows on Netflix')
for i, count in enumerate(type_counts):
    plt.text(0.75 * i, 0, f'{count} ({type_counts.index[i]})',
             ha='center', va='center')
plt.show()
```

Distribution of Movies and TV Shows on Netflix



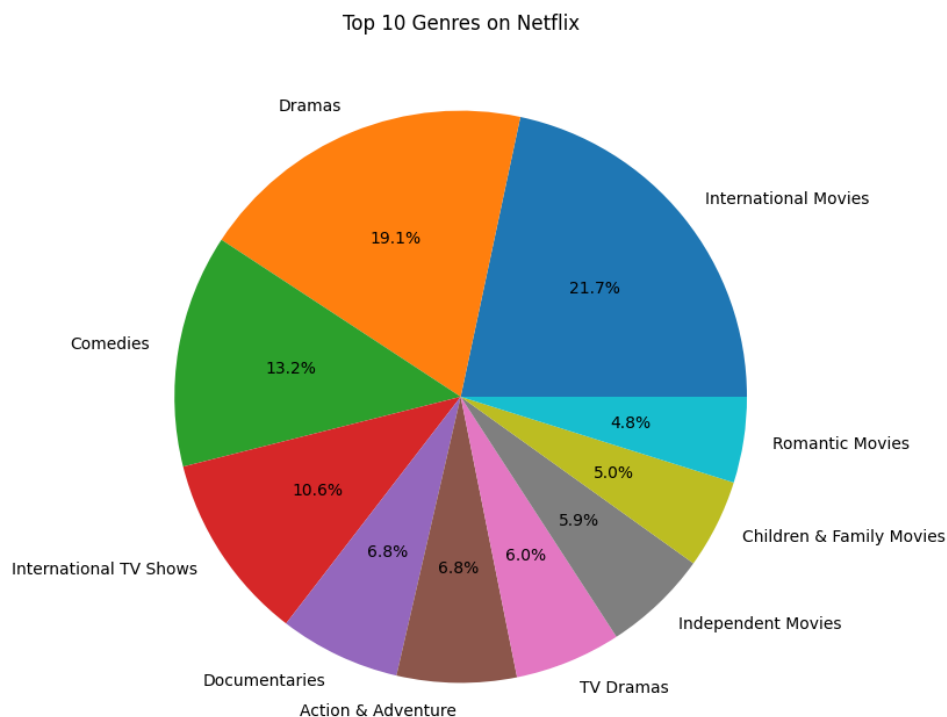
```
#Plot a time series of content added over the years
netflix_data['date_added'] = pd.to_datetime(netflix_data['date_added'])
netflix_data['year_added'] = netflix_data['date_added'].dt.year
sns.countplot(data=netflix_data, x='year_added')
plt.title('Content Added Over the Years')
plt.xlabel('Year Added')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```



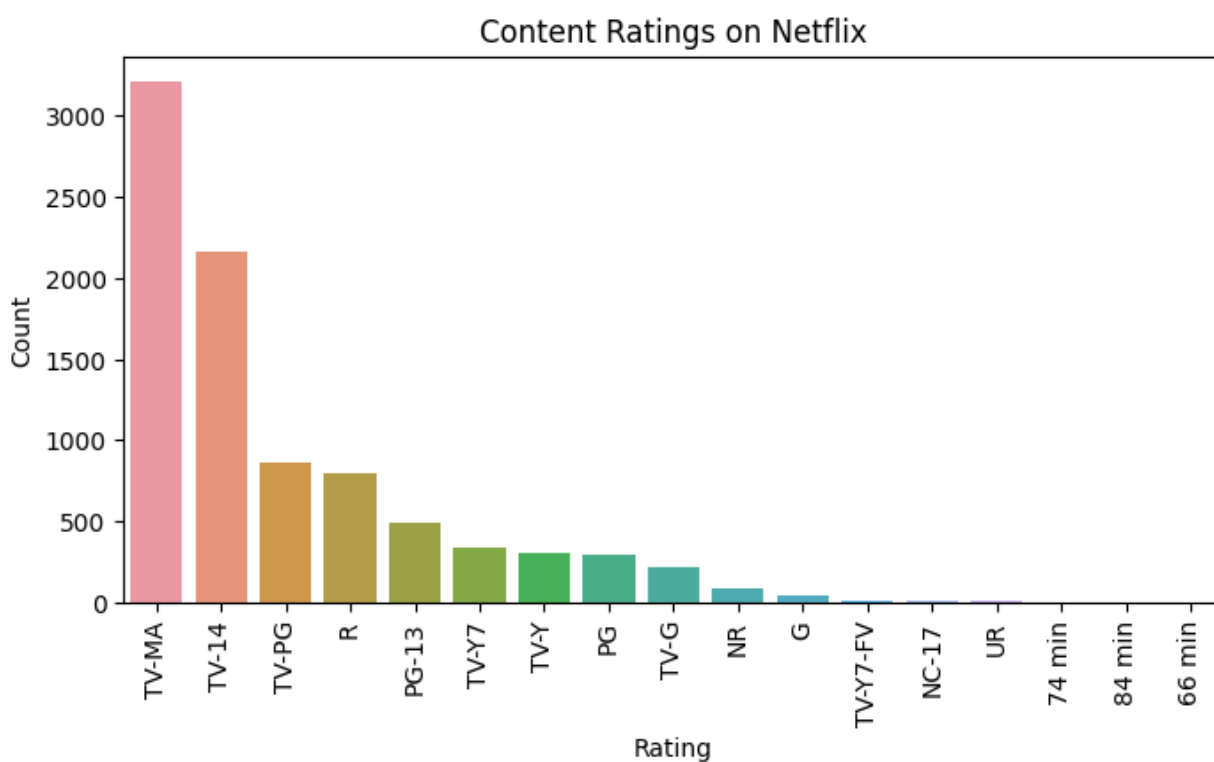
```
# Import the Pandas and Matplotlib libraries
import pandas as pd
import matplotlib.pyplot as plt

# Code 3: Visualize the distribution of content by genre
genre_counts = netflix_data['listed_in'].str.split(',').explode().value_counts()
top_genres = genre_counts.head(10).index

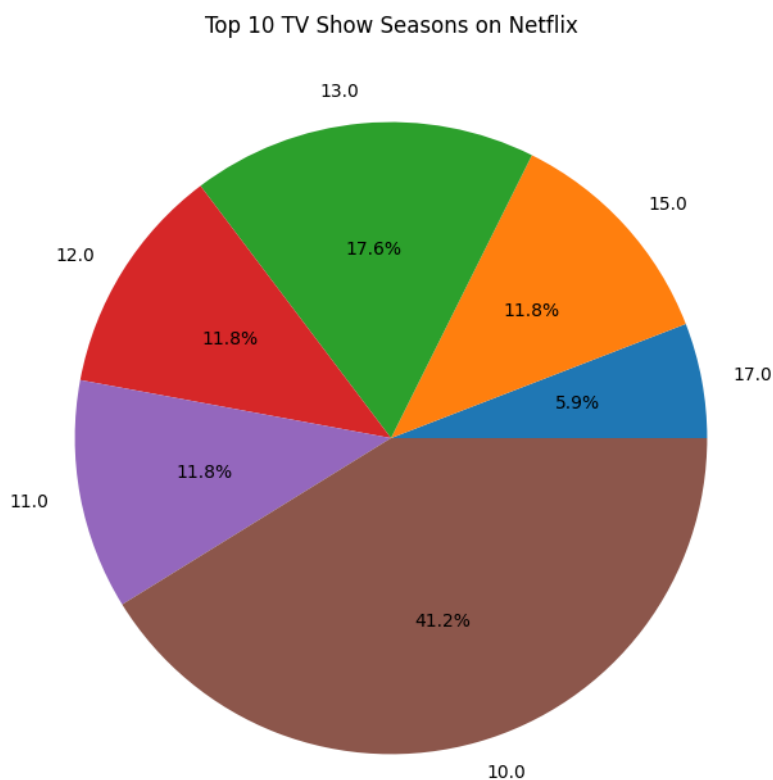
plt.figure(figsize=(8, 8))
plt.pie(genre_counts[top_genres], labels=top_genres, autopct='%1.1f%%')
plt.title('Top 10 Genres on Netflix')
plt.show()
```




```
# Code 4: Visualize the distribution of content by rating
plt.figure(figsize=(8, 4))
sns.countplot(data=netflix_data, x='rating',
order=netflix_data['rating'].value_counts().index)
plt.title('Content Ratings on Netflix')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```



```
#Visualize the distribution of the number of seasons (for TV Shows)
tv_show_durations = netflix_data[netflix_data['type'] == 'TV
Show']['duration']
season_counts =
tv_show_durations.value_counts().sort_index(ascending=False)[:10]
plt.figure(figsize=(8, 8))
plt.pie(season_counts, labels=season_counts.index, autopct='%1.1f%%')
plt.title('Top 10 TV Show Seasons on Netflix')
plt.show()
```

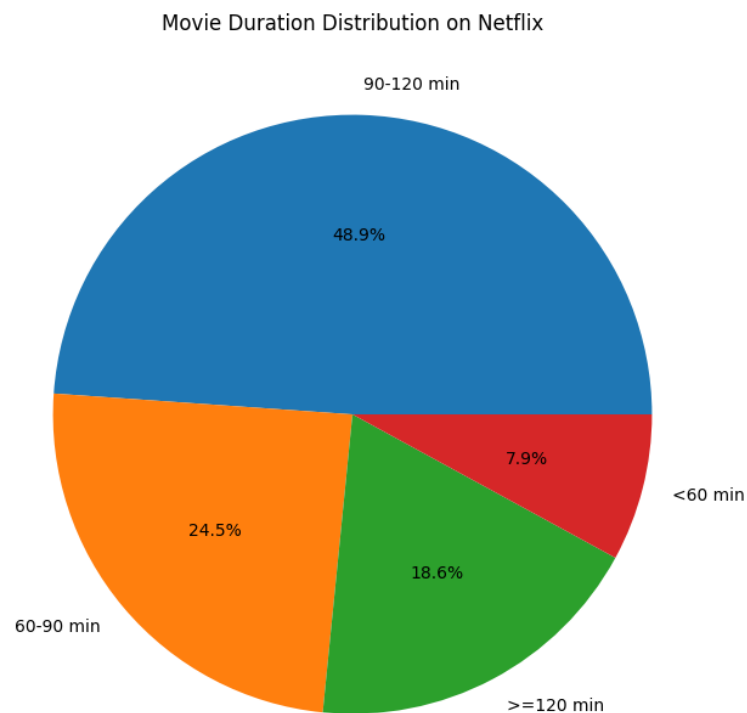


```
# Code 6: Visualize the distribution of movie durations (in minutes)
movie_durations = netflix_data[netflix_data['type'] ==
                              'Movie']['duration']

# Categorize movie durations into different ranges
bins = [0, 60, 90, 120, float('inf')]
labels = ['<60 min', '60-90 min', '90-120 min', '>=120 min']
duration_ranges = pd.cut(movie_durations, bins=bins, labels=labels)

# Count the number of movies in each duration range
duration_counts = duration_ranges.value_counts()

plt.figure(figsize=(8, 8))
plt.pie(duration_counts, labels=duration_counts.index, autopct='%1.1f%%')
plt.title('Movie Duration Distribution on Netflix')
plt.show()
```

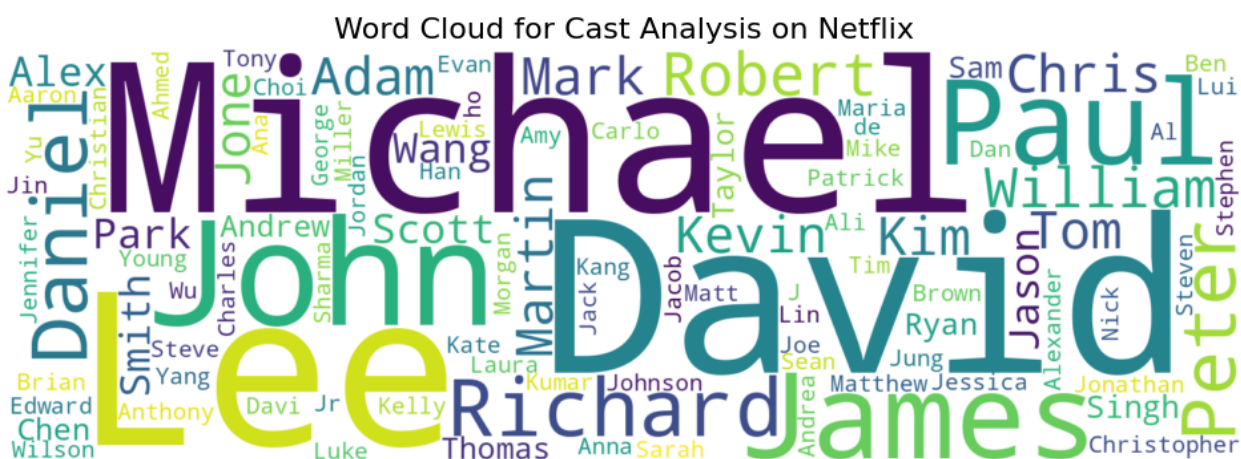


```

# Concatenate all cast members into a single string
all_cast = ' '.join(netflix_data['cast'].dropna())

# Create a word cloud based on cast members
wordcloud = WordCloud(
    width=1500,
    height=500,
    background_color='white',
    colormap='viridis',
    max_words=100, # Maximum number of words to display
).generate(all_cast)
# Display the word cloud
plt.figure(figsize=(12, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Word Cloud for Cast Analysis on Netflix', fontsize=16,
color='black')
plt.show()

```

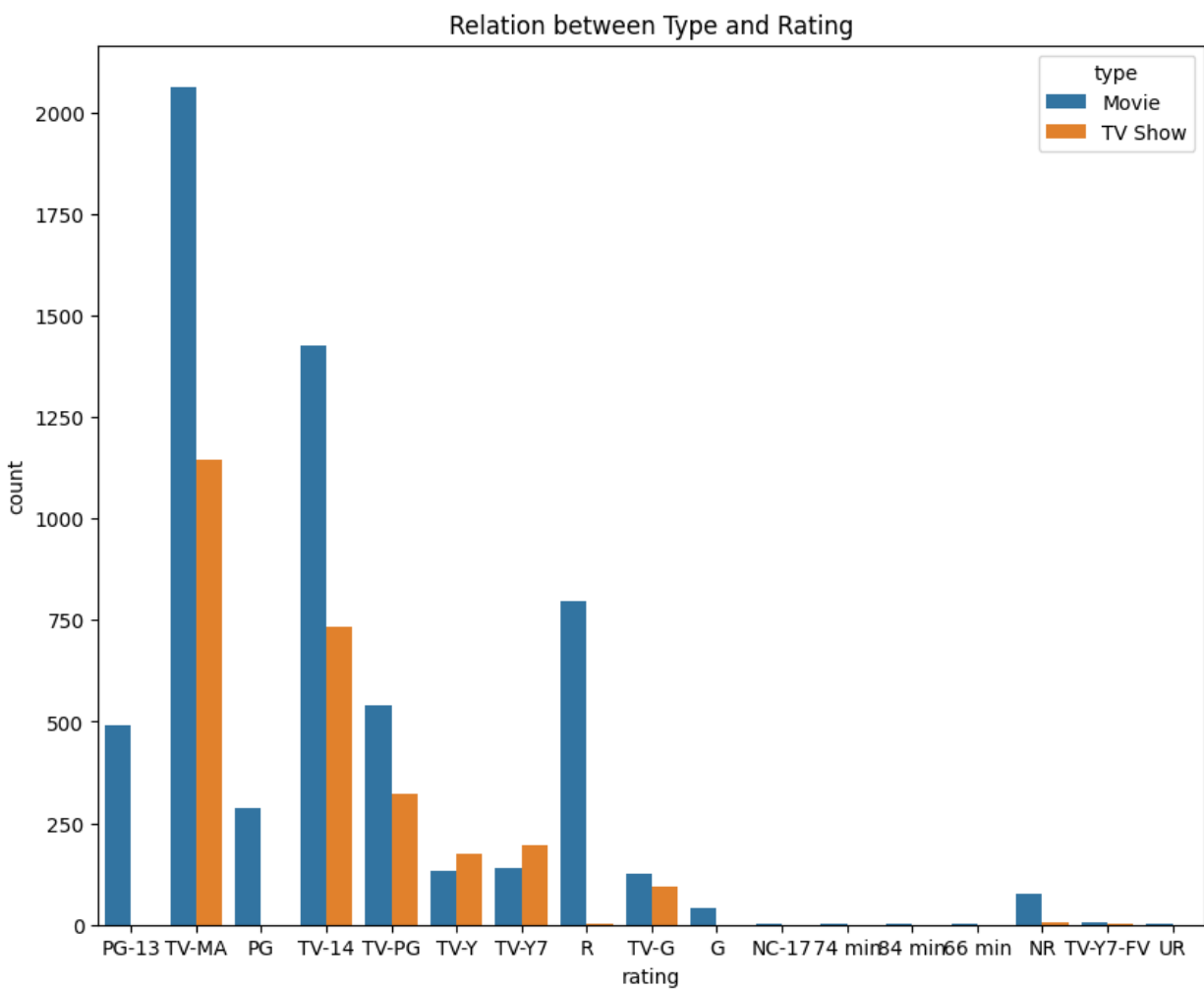


```
# Concatenate all genre categories into a single string
all_categories = ' '.join(netflix_data['listed_in'].dropna())

# Create a word cloud based on genre categories
wordcloud = WordCloud(
    width=800,
    height=400,
    background_color='white',
    colormap='viridis',
    max_words=100, # Maximum number of words to display
).generate(all_categories)

# Display the word cloud
plt.figure(figsize=(12, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Word Cloud for Categories (Genres) Analysis on Netflix',
fontsize=16, color='black')
plt.show()
```

```
plt.figure(figsize=(10, 8))
sns.countplot(x='rating', hue='type', data=netflix_data)
plt.title('Relation between Type and Rating')
plt.show()
```



Data Insights

- - - - X

Based on the provided data visualisations and insights, here's a more detailed interpretation:

Distribution of Movies and TV Shows:

Most content on Netflix consists of movies, accounting for approximately 69.6% of the total, while TV shows make up the remaining 30.4%. This suggests that Netflix primarily focuses on offering a wide variety of movies to its subscribers, with a substantial but smaller selection of TV shows.

Content Added Over the Years:

The data shows that the most content additions occurred in 2019, with over 2000 movies and TV shows added. However, there is a gradual decline in content additions from 2019 to 2021. This trend may indicate that Netflix has been moderating its content additions in recent years, possibly due to changes in content acquisition strategies or other factors.

Top Genres on Netflix:

Within the diverse range of genres available on Netflix, international movies, dramas, and comedies stand out as the top three categories. This suggests that Netflix caters to a broad global audience, offering a mix of content to meet various preferences. International movies' popularity highlights the platform's commitment to providing content from different regions.

Duration of Shows and Movies:

A significant portion of shows (41.2%) on Netflix consists of those with 10 seasons, indicating the popularity of long-running series. Most movies (48.9%) fall into the duration category of 90-120 minutes. This information can help Netflix tailor its content recommendations based on viewer preferences for show length.

Geographical Analysis:

The analysis of content consumption by country reveals that the United States, the United Kingdom, and India are the top consumers of Netflix content. These countries likely have a strong subscriber base and diverse content preferences. This insight can guide Netflix in tailoring its content library and marketing strategies for different regions.



Cast Analysis:

Among the cast members in Netflix content, names like Michael, John, David, and Lee appear frequently. This information can be valuable for understanding which actors are associated with popular content on the platform. Netflix can leverage the popularity of these actors in promotions or consider casting them in new productions to attract viewers.

These insights provide a comprehensive view of the Netflix content landscape and user preferences. Netflix can use this information to make informed decisions about content acquisition, production, and customisation to cater to the tastes and preferences of its diverse subscriber base.



Conclusion

- - - - X

In conclusion, the data visualisations and insights provide valuable information about Netflix's content distribution, growth patterns, genre preferences, content curation, and regional consumption. Netflix's primary focus on movies with a substantial selection of TV shows showcases its commitment to catering to a diverse global audience. The gradual decline in content additions post-2019 suggests a potential shift in content acquisition strategies or other strategic considerations. Understanding the top genres and content duration preferences can aid Netflix in refining its content recommendations and delivery. Lastly, the geographical analysis underscores the importance of tailoring content offerings and marketing strategies to meet the unique preferences and demands of subscribers in different countries. This comprehensive analysis equips Netflix with valuable insights for its success in the highly competitive streaming industry.

