

# Netflix project

Sarah Sahli - Jedha Bootcamp

# Agenda

- Problem statement
- Dataset
- Methodology
- Dominant genre on the platform
- Top production countries
- Dominant Genre per Country
- Catalogue seasonality
- Conclusion & Next steps

# Problem statement



1

Netflix wants to understand the current state of their catalogue and has provided a csv file with the shows released on their platform.



2

The goal is to analyze the Netflix show data to provide insights and answers to the questions posed.



3

Key questions to answer: Which genres are most prevalent? How many countries' content does Netflix produce? Which country has the most shows? What is the dominant genre per country? Is there seasonality in catalogue release?

# Dataset Summary

**Jan 2008 -  
Sep 2021**

Data Period


**8807**

Row Count


**12**

Column Count


# Methodology



The process began with data cleaning and processing before moving on to exploratory data analysis using Plotly.



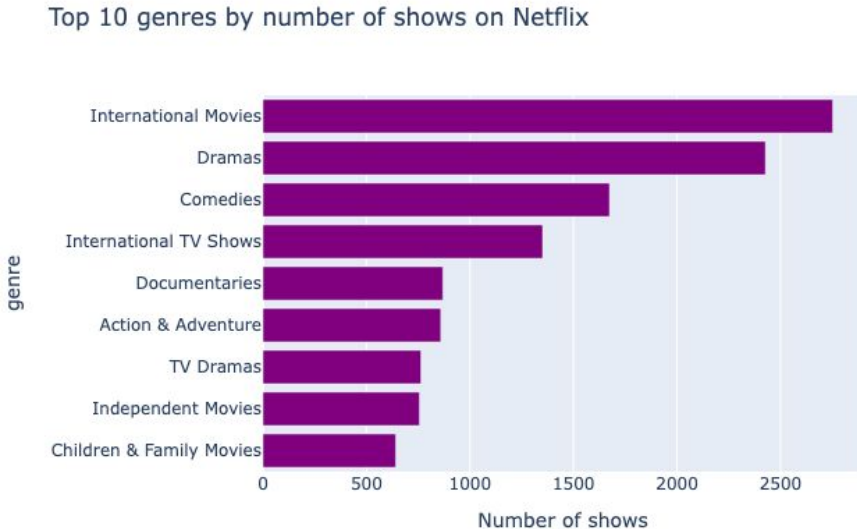
Importing the csv file revealed incorrect data types in certain columns, requiring conversion for proper manipulation.



Columns such as "country" and "listed\_in" contained multiple values and needed to be broken down for analysis purposes, particularly for question-specific insights.

# Dominant genre on the platform

## Distribution of content genres

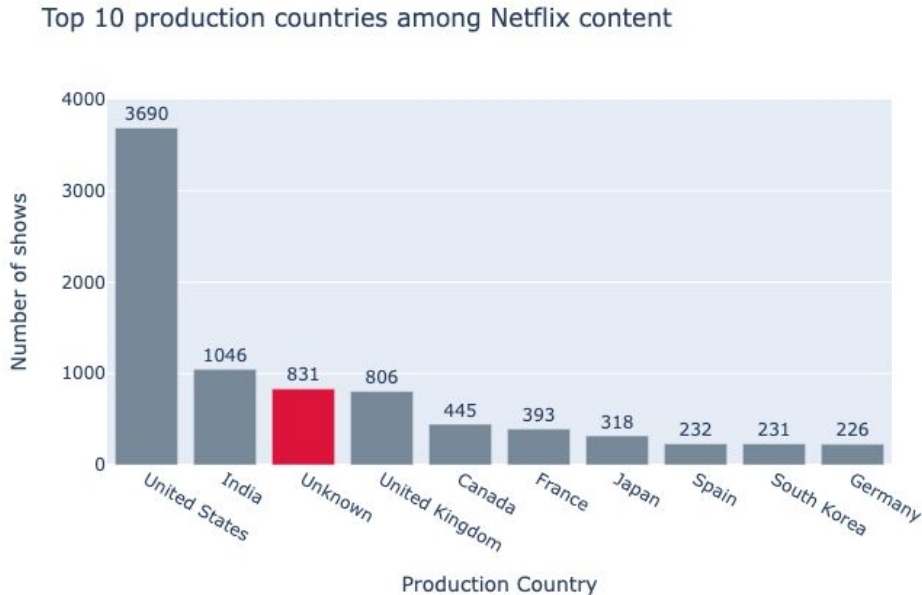


## Key Insights

- The most recurrent genre is "International Movies" (non-US movies) followed by "Dramas" and "Comedies".
- This trend suggests that Netflix is focused on providing relevant content for their non-US customer base.
- There are 42 unique genres on the platform that could possibly be regrouped for a more streamlined categorization.

# Top production countries

## Distribution of Content by Production Country



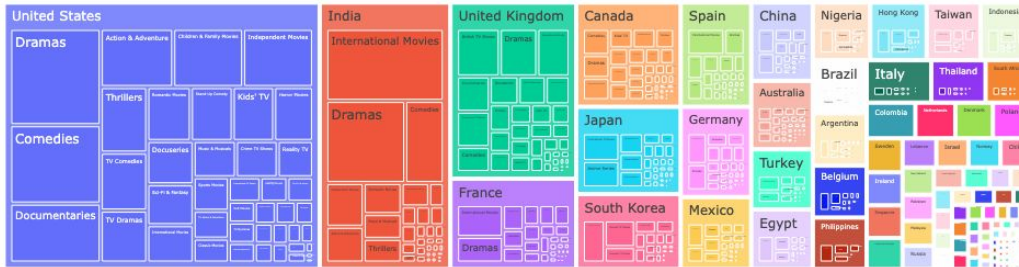
## Insights on Production Countries

- The US is the top production country for Netflix content, followed by India, reflecting their status as top movie producers globally.
- The presence of 831 shows with unknown production country in the dataset is a significant factor that could impact the ranking of other countries.
- The rest of the countries are primarily Western countries, with the exception of South Korea, due to the increased popularity of K-dramas worldwide.

# Dominant Genre per Country

## Dominant genre per country

Dominant Genre per Country



## Insights on Country-Specific Genres

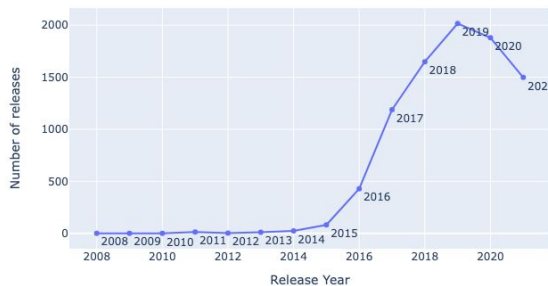
- The dominant genre per country reveals country-specific genres, such as "British shows" for the UK and "Anime" for Japan.
- This data reflects the dominant genre per production country, not the dominant genre among watchers in the country.
- It would be interesting to see which genres are most famous in each country, taking into account the actual viewership data.



# Catalogue seasonality - Charts

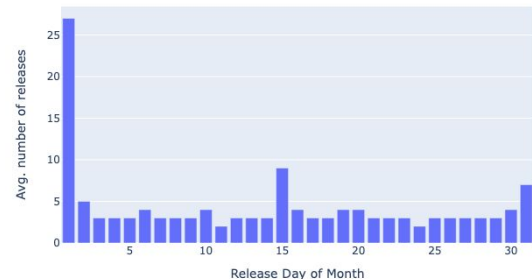
## Yearly Seasonality

Total number of releases per year



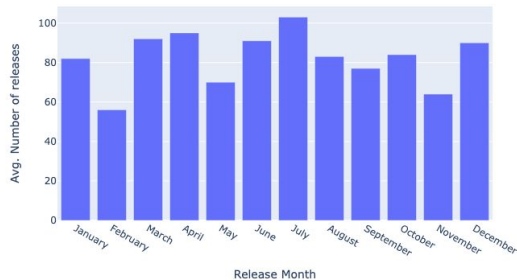
## Day of Month Seasonality

Average number of releases per day of month



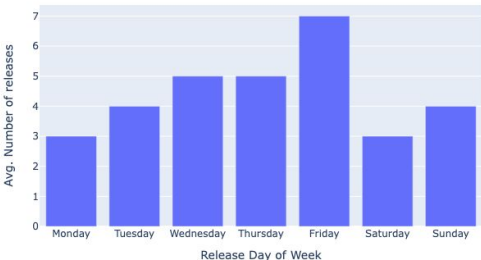
## Monthly Seasonality

Average number of releases per month



## Day of Week Seasonality

Average number of releases per day of week



# Catalogue seasonality - Insights



## Covid Impact

The number of shows added to the platform declined in 2020 due to the Covid-19 crisis, which made it more difficult to shoot new movies/shows.



## Monthly Peak

Netflix tends to add the majority of new shows to the platform on the 1st of the month.



## Fridays Release

Most releases are done on Fridays. This could be a strategy from Netflix to get people to watch new content during the week-end.

# Conclusion & Next steps

## Genre Analysis

Netflix has a high number of genres on the platform (42) and a next step could be to look at how to regroup some of these genres together to decrease the number of genres, making the analysis easier.

## NLP Project Opportunity

A potential NLP project could analyze the 'description' column to identify the most common keywords in show descriptions.

## Explore Other Columns

We could explore the other column such the 'director', 'cast', 'rating' and 'description' to gather other insights.