Data Selection

I chose the dataset from Netflix to explore in this report. I love to watch movies very much, and I'm a big fan of Netflix since it produces tons of excellent films. I'm interested in exploring more from it.

After reviewing the documentation of the dataset, I came up with the following four initial questions to visualize:

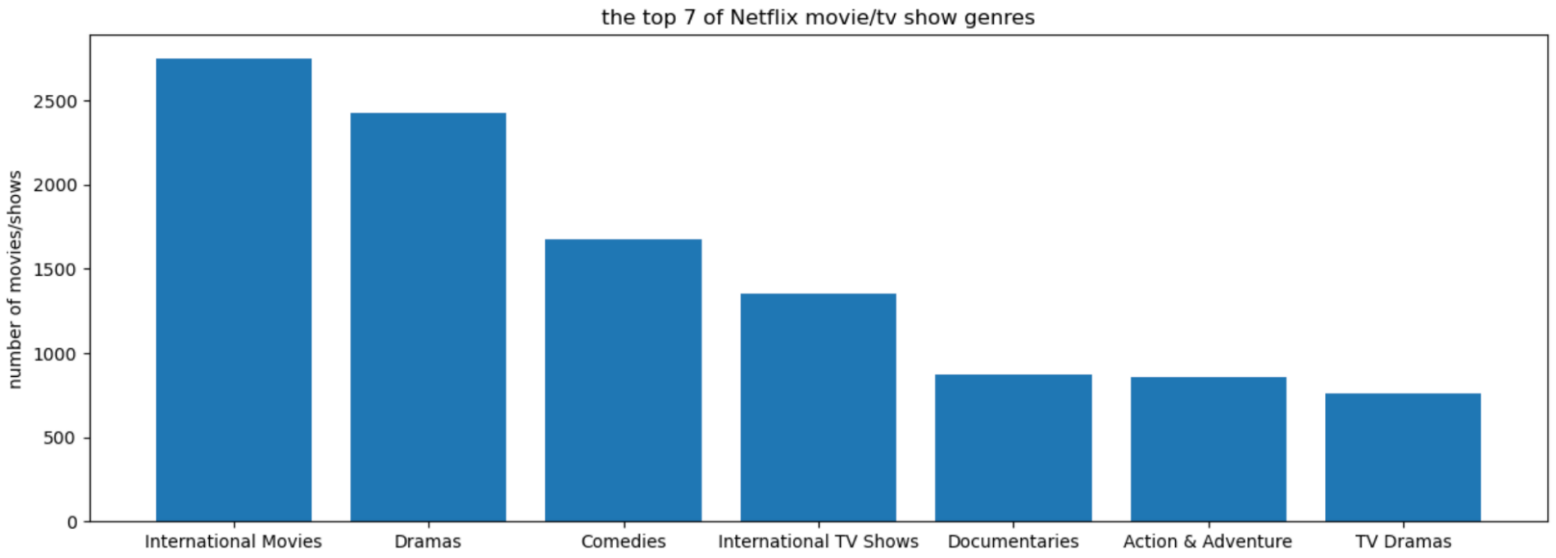
1. What are the top 7 Netflix movie/tv show genres? - 柱状图
2. Which 5 countries have the most movies on Netflix? - 饼图
3. How many movies have been published on Netflix each year since 2000? - 线形图
4. Among the movies of the United States, who are the top 5 most-seen celebrities? - 柱状图

Exploratory Visual Analysis

- **First Phase**
  - **What variables does the dataset contain?**
    show_id; type; title; director; cast; country; date_added; release_year; rating; duration
  - **What type of variable does the dataset contain? (e.g., nominal, ordinal, discrete).**
    show_id: nominal variable
    type: nominal variable
    title: nominal variable
    director: nominal variable
    cast: nominal variable
    country: nominal variable
    date_added: continuous variable
    release_year: continuous variable
    rating: continuous variable
    duration: nominal variable
  - **How are they distributed?**
    I tried to find out the distribution of some data columns in this dataset, so I wrote down several questions about the distribution in the previous phase.
  - **Are there any relationships among the variables?**
    There are some relationships among the variables, such as how the release year is related to the distribution of the countries of newly published movies/TVs.

Second Phase - Investigating Questions

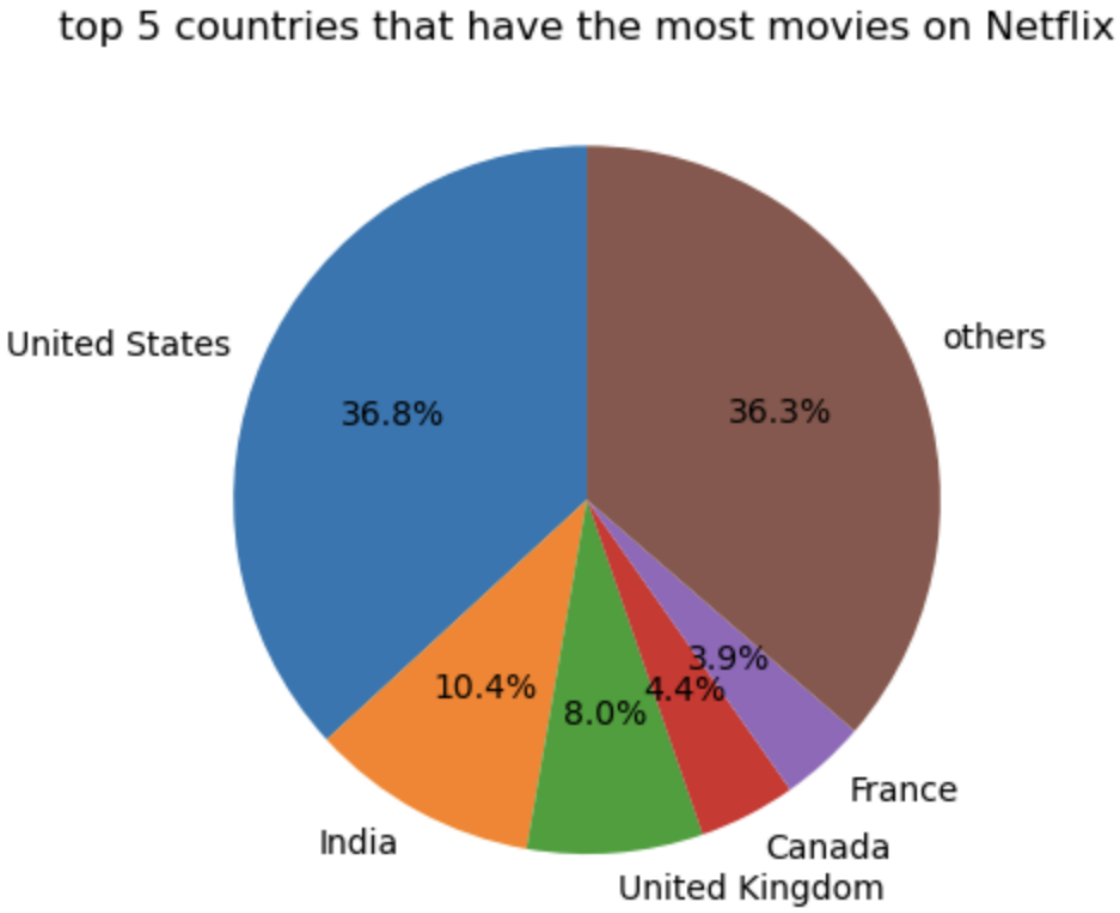
1. What are the top 7 Netflix movie/tv show genres?



Insights

From the visualization, we can see that the top seven movies/TV show types are International movies, Dramas, Comedies, International TV shows, Documentaries, Actions, and TV dramas. International movies and TV shows dominate this graph since they appeared two times, and international movies are at the top. It would be better if specific countries among these international movies/TVs can be given in this dataset, and we can tell more about what genres users like.

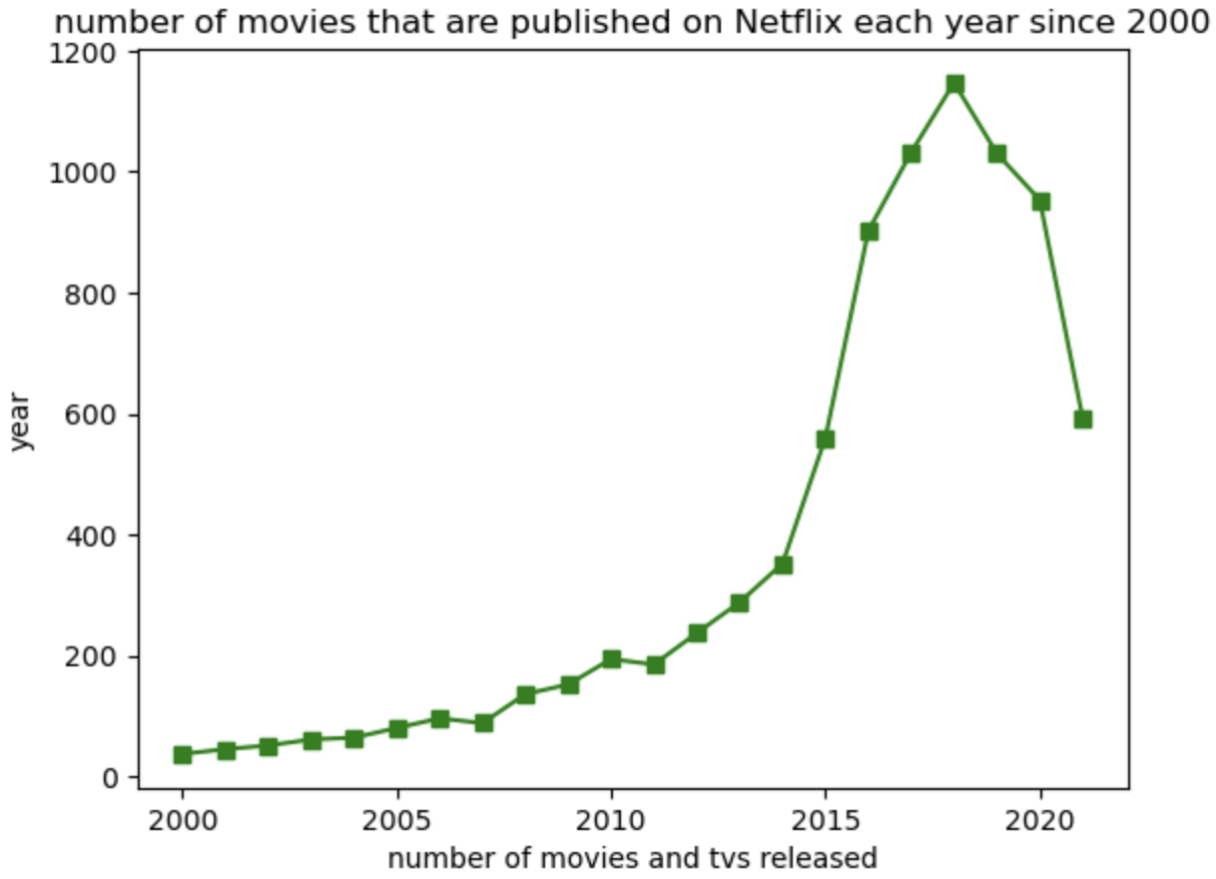
2. Which 5 countries have the most movies on Netflix?



Insights

The United States has the greatest number of movies/TVs on Netflix, and it might be because that Netflix is a domestic company. What is surprising to me is that India is in second place in this rank, and I never thought that India has so much cultural exportation.

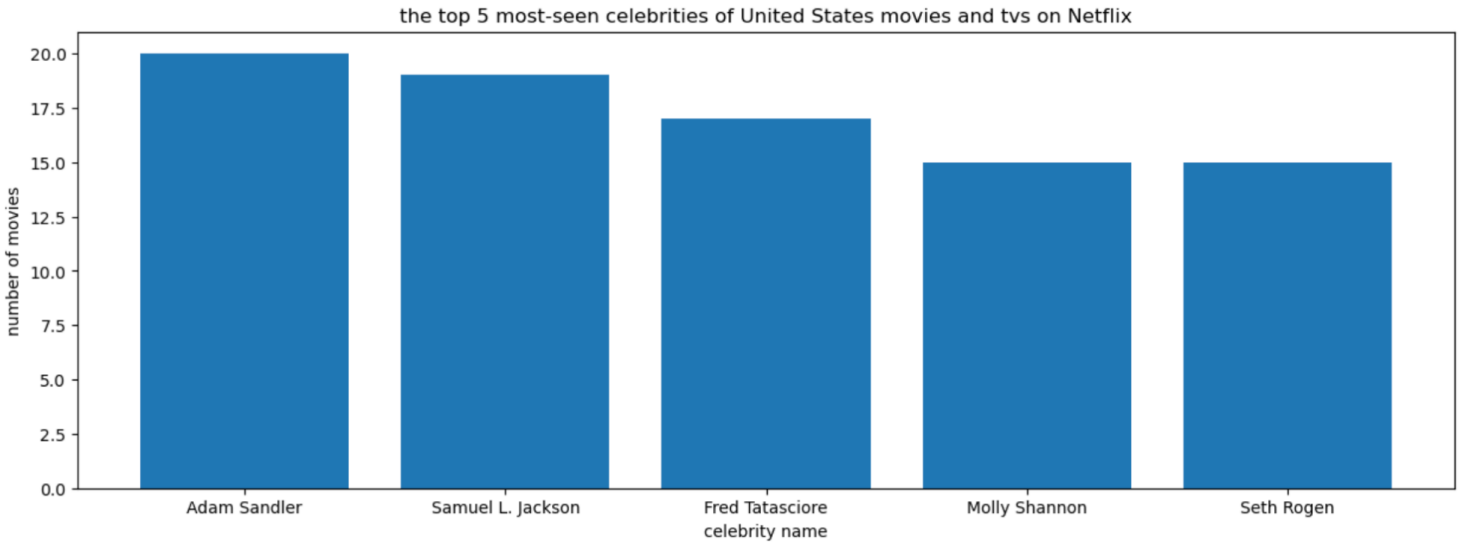
3. How many movies have been published on Netflix each year since 2000?



Insights

We can tell from the graph that between 2000 and 2014 the increasing ratio of movies/TVs on Netflix is relatively slow, but it burst to start from 2015 and finally started to decrease in 2017. This graph indicates the fluctuation of the potential commercial value of Netflix by showing the boom and bust of fresh movies/TVs.

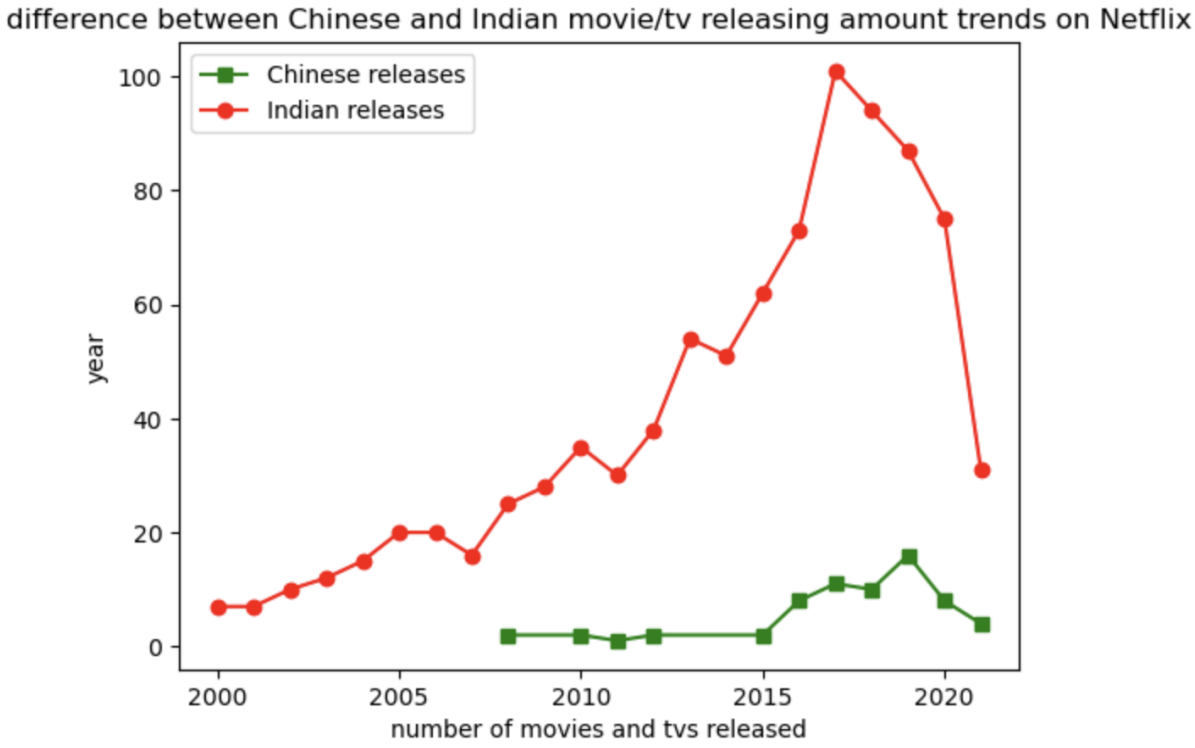
4. Among the movies of the United States, who are the top 5 most-seen celebrities?



Insights

In this graph, we can see that the top five celebrities seen on Netflix United States movies/TVs are Adam Sandler, Samuel L. Jackson, Fred Tatasciore, Molly Shannon, and Seth Rogen. This would indicate the popularity of celebrities.

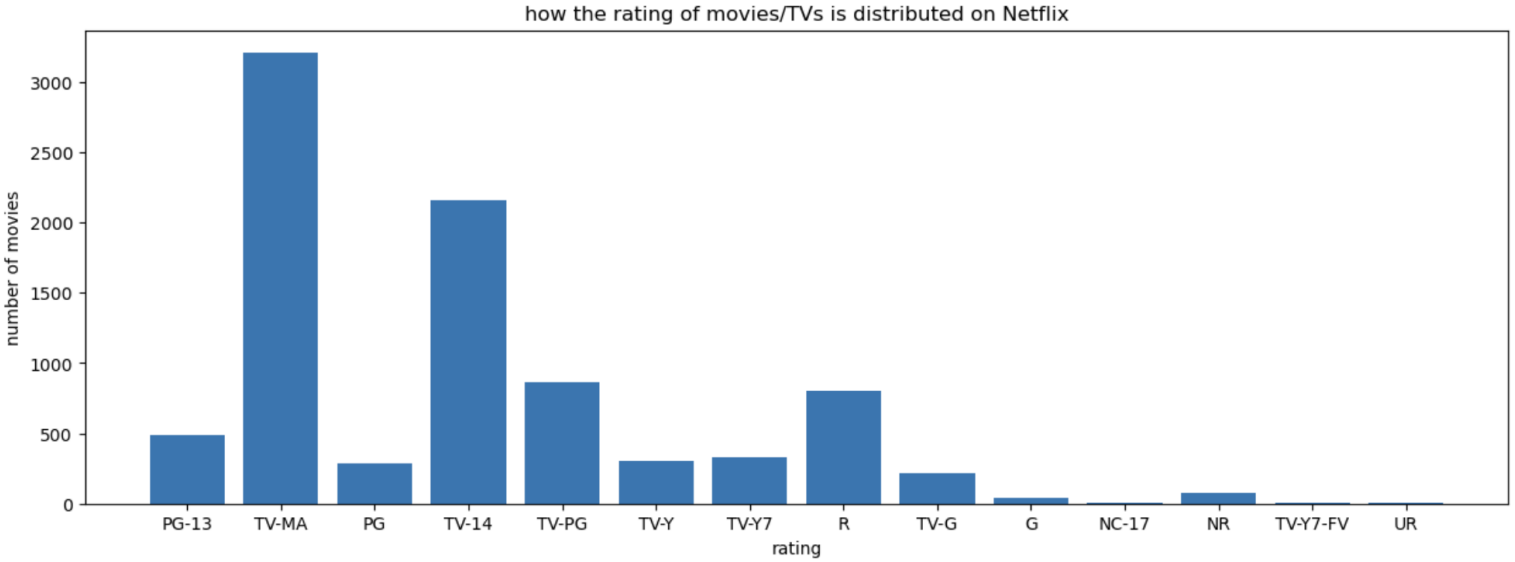
5. What's the difference between Chinese and Indian movie/tv releasing amount trends on Netflix?



Insights

In this graph, we can tell that India has been publishing movies/TV on Netflix since 2000, but China just began to do so in 2007. India has experienced a huge drop in the number of new movies/TVs published starting from 2017, but it still exceeds China to a great extent.

6. How is the rating of movies/TVs distributed on Netflix?



Insights

For the final graph, I came out with the idea to visualize the relationship between the release year and the time duration. However, I failed to project the time duration strings to the scatter plot y-axis. Thus, I chose to visualize the rating distribution of movies/TV on Netflix. From this graph, we can see that the majority of ratings in Netflix movies/TVs are the TV-MA, and the last ones would be the NC-17, TV-Y7-FV, and UR.
TV-MA means that the movies/TVs are not designed for children under 17, and this can somehow indicate the age distribution of Netflix users, and at least 40 percent of the users are adults.