EDA and Visualization for

NETFLIX

Kejia Wang

2023.8

Content

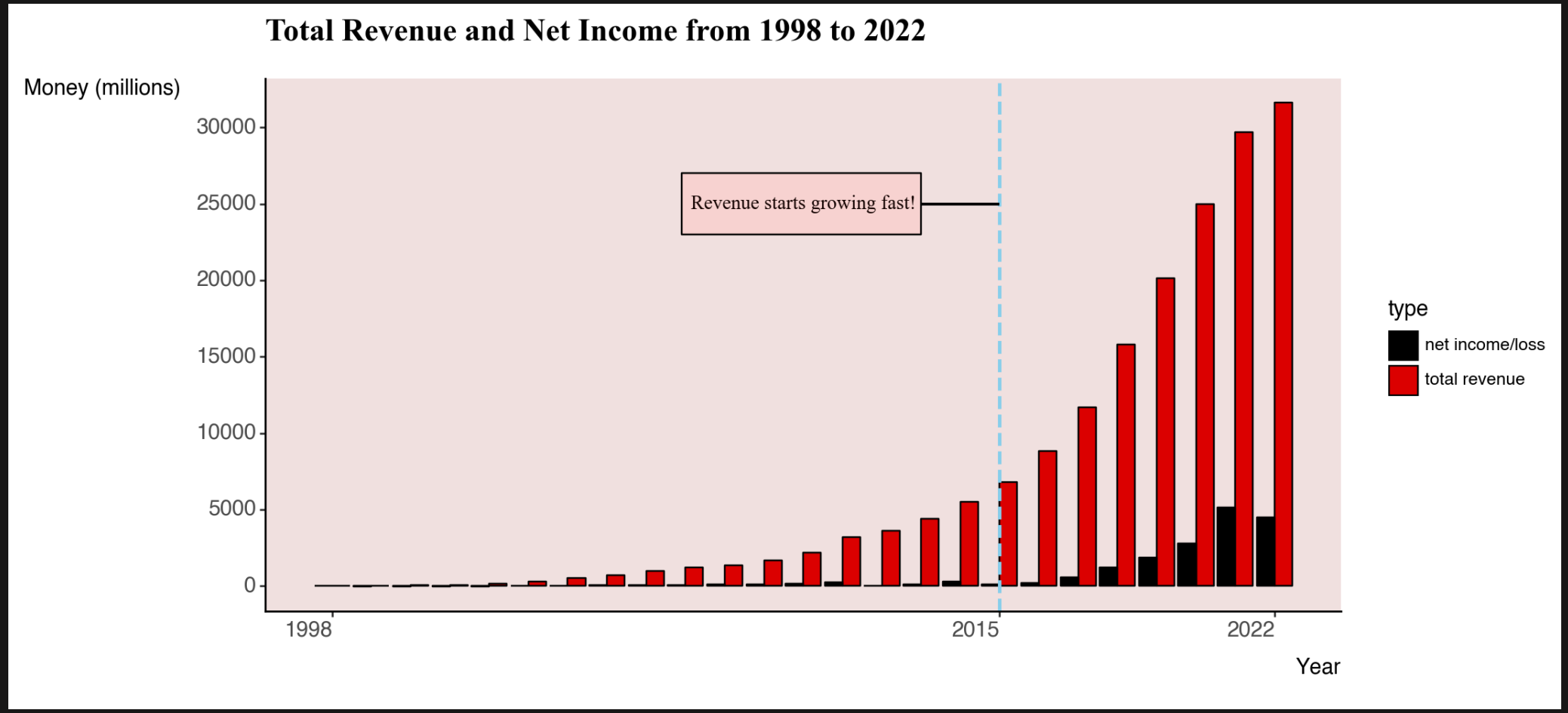
Part 1

Research Background

Understanding the revenue change in Netflix with our dataset

Background: A Sudden Revenue Increase

We looked into the Netflix financial data, and here is what we found…



What We Guess: Globalization

Our Goal:

Identify trends in content added to the Netflix platform that fostered growth and allowed the company to capitalize on towards global expansion in the years following 2015

Our Guess:

They expand and release their content at the right time

Globalization strategy starts its influence

Our Methods:

Use Kaggle data to see whether we can find any supportive evidence

Part 2

Data Summary

Dataset and variables of interest

Our Datasets: Default Data and External Data

Default Data

Listings of movies and tv shows on Netflix

From Kaggle

- **8807** total records in the dataset
- **12** different columns

type: A Movie or TV Show
show_id: Unique ID for every Movie / TV Show
title: Title of the Movie / TV Show
director: Director of the Movie
cast: Actors involved in the movie / show
country: Country where the content produced
date_added: Date it was added on Netflix
release_year: Actual Release year of the content
rating: TV Rating of the movie / show
duration: Total Duration in minutes or seasons
listed_in: Category/Genre of Movie/TV Show
Description: Its description

We dropped all the missing values.

External Data

The revenue and net income data
from 1998 to 2022
from their annual financial report
(Exactly the plot we showed before)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8803 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

Our Attributes: The columns we choose

Our (only) criteria for selecting columns:

- Have something to do with the change in revenue and net income

The columns we focused on

- **Type** (Movie or TV Show)
- **Date_Added** (When Netflix added this title to library)
  - Primary focus on years 2015-2021
- **Listed_In** (Category/Genre of Movie/TV Show)
- **Country** (Location the Movie/Show was produced)
- **Rating**
  - Maturity rating of content
  - Grouped by recommended audience - Adults/Teens/Kids

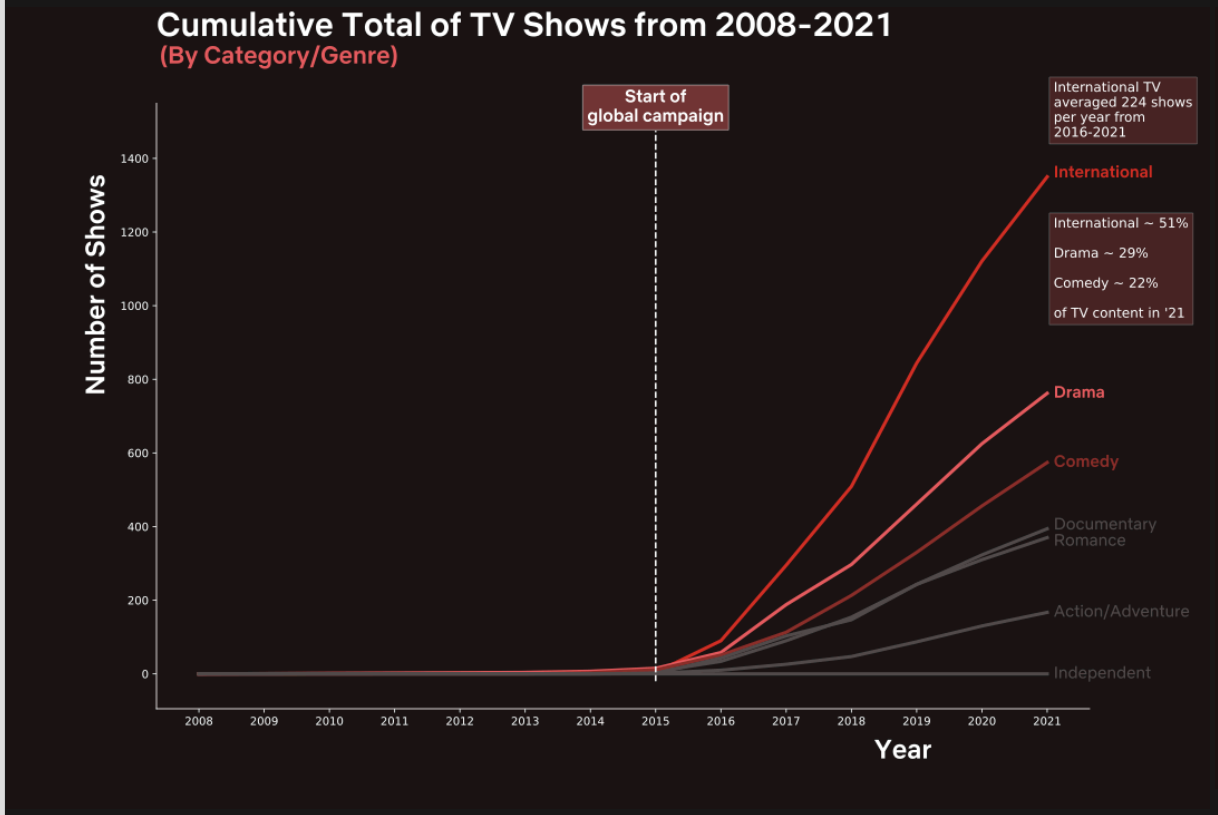
Part 3

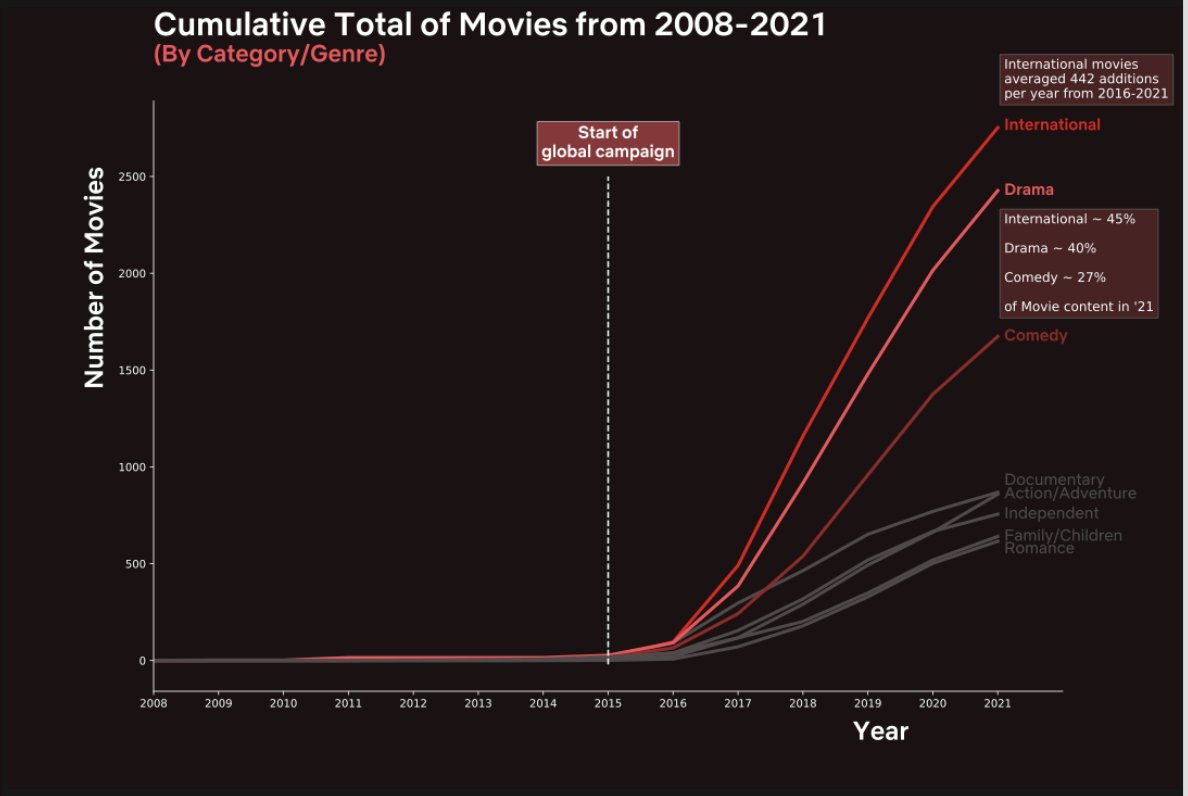
Our Findings

Might be some evidence

Our findings: Genre/Category Growth

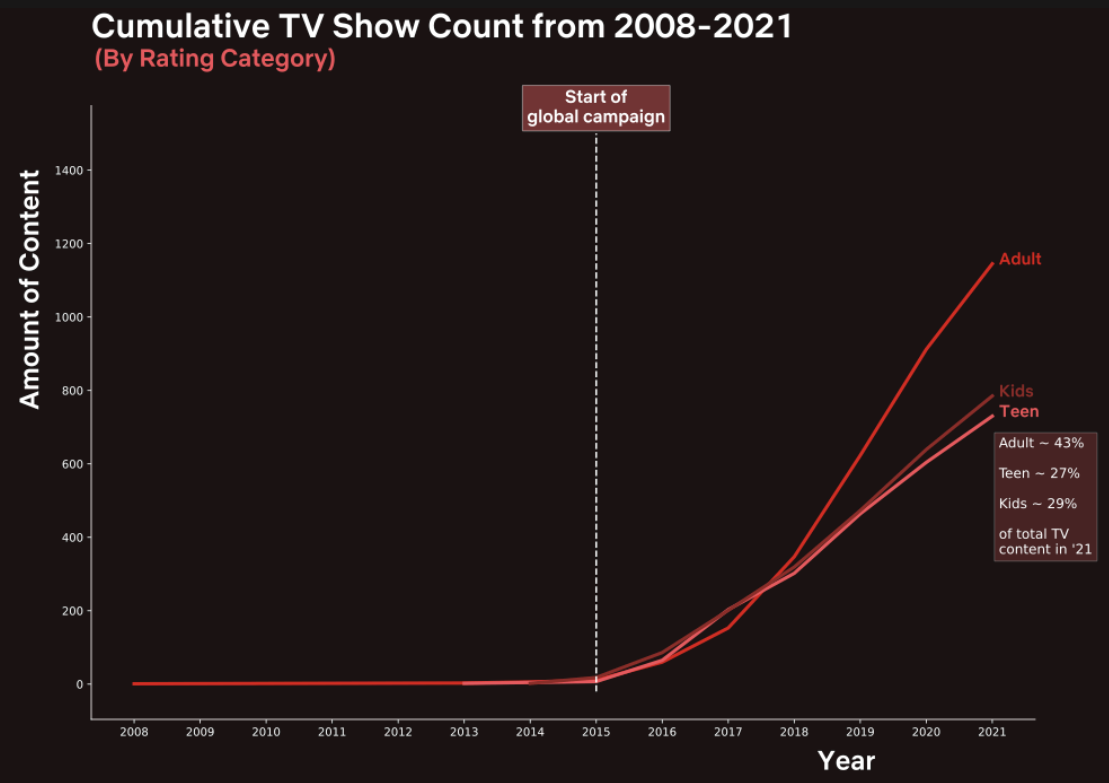
The addition of TV show and movies both increased suddenly around 2015, with ‘International’, ‘Drama’, and ‘Comedy’ as highest performers

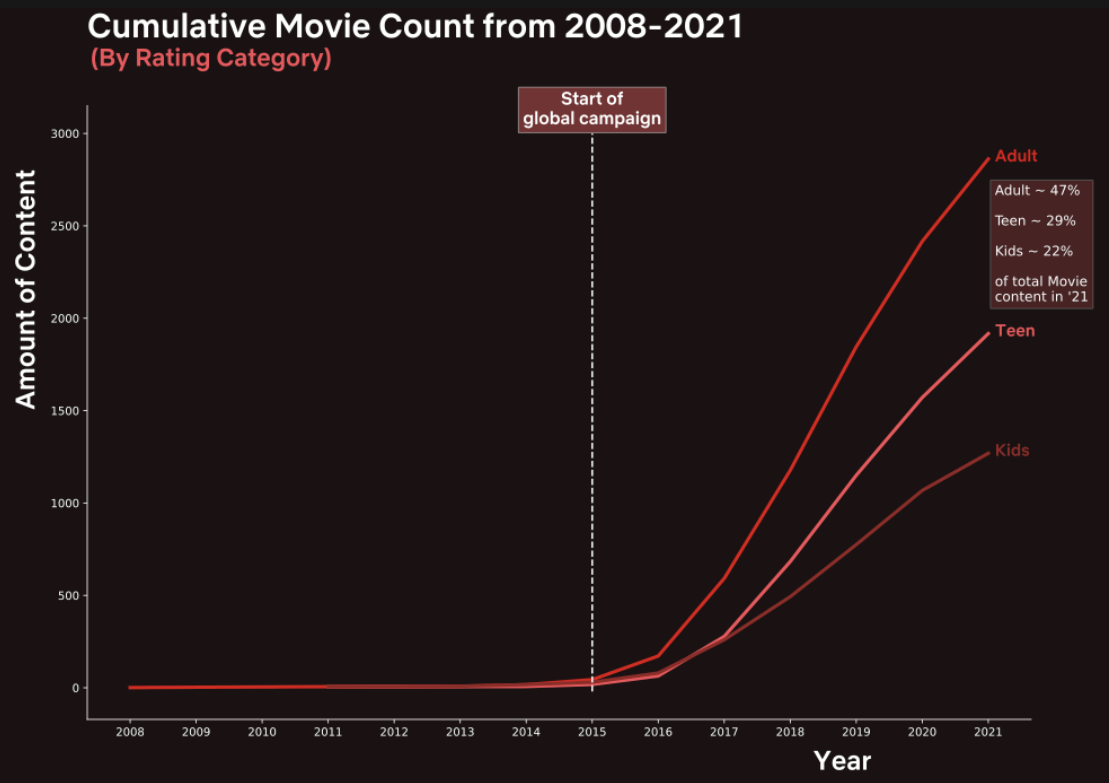




Our findings: Ratings

Adult rated programming highest in both Movies/TV

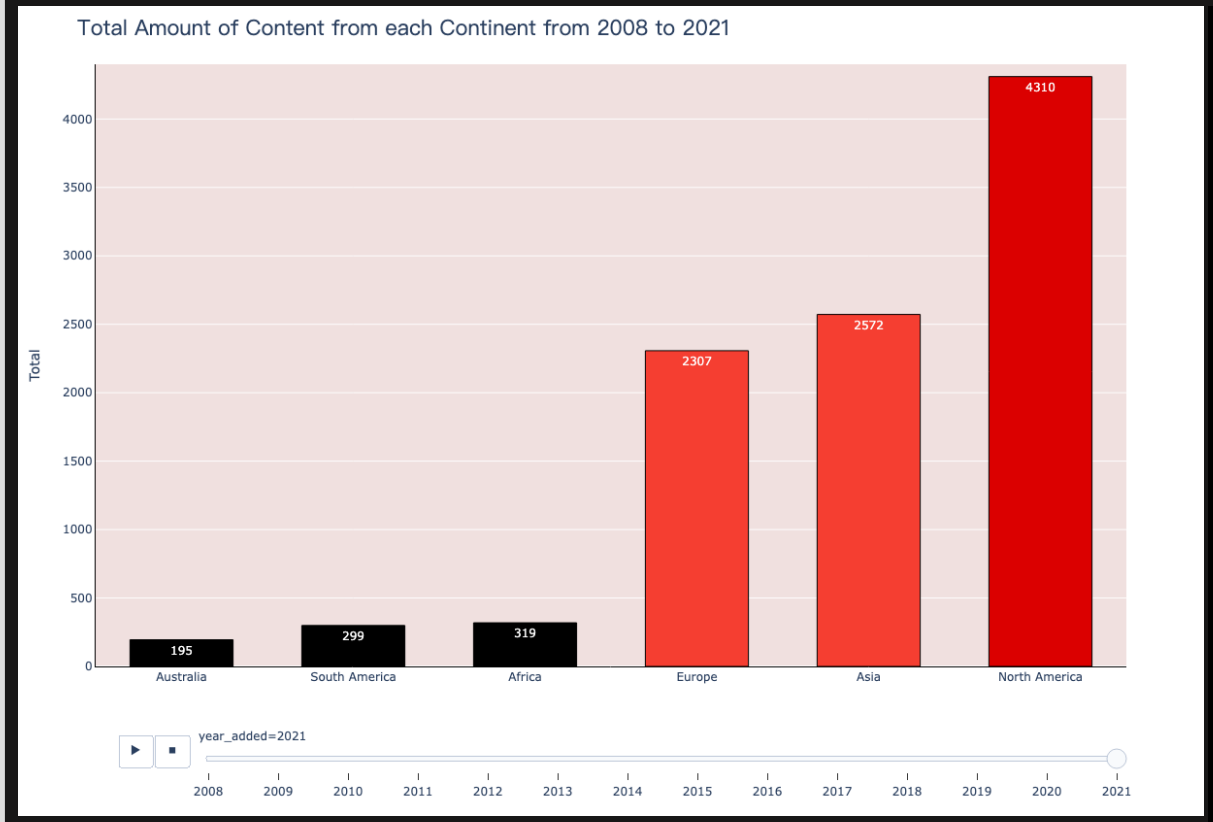


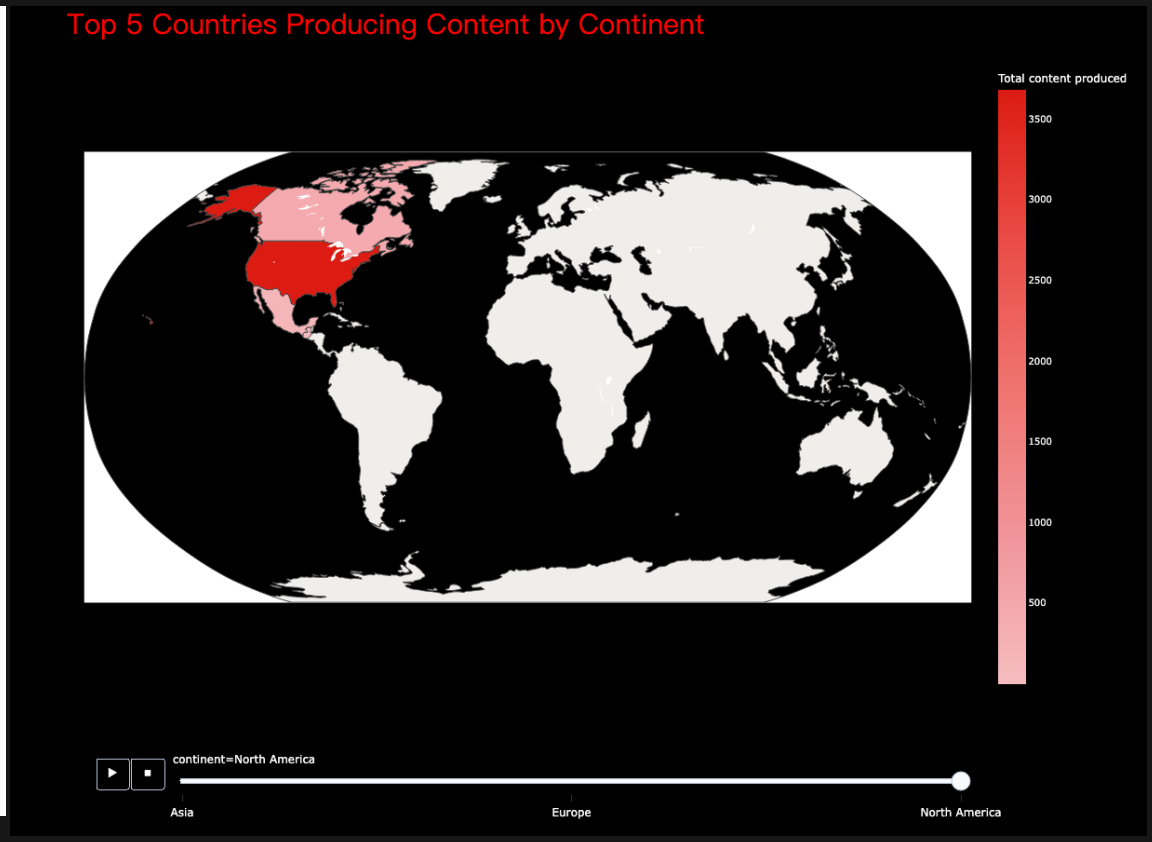


Netflix Rating Guidelines: https://help.netflix.com/en/node/2064

Our findings: Countries

A fast globalization pace





P1: https://gitmebrandonhom.github.io/bar.html
P2: https://gitmebrandonhom.github.io/map.html

Part 4

“What if”s

How we could come to a stronger conclusion if we had more time and data

What If We Have More Time and Data

Our primary goal:

Identify trends in content added to the Netflix Platform that fostered growth and allowed the company to capitalize on global expansion in the years following 2015

Possible next steps:

1

Expand the Dataset:

- Viewing data:

  How many people have actually seen those content

- Rating data and Comments:

  Whether people like and how they think about the content

2

Regression Analysis:

- Revenue Models:

  Estimating the correlation between each element Netflix expanded into and the corresponding revenue growth, and find the most important contributor and find the casual relationship

3

Interviews:

- With Customers/ Netflix Executives and Employees/ Content Creators and Partners/ Industry Experts
- To get qualitative insights to complement the quantitative findings from regression analysis

THANKS

Kejia Wang

2023.8