NETFLIX TOP 10 - USA

Presented by

Rucha Patil
Viktor Spasic
 Ci Song

TABLE OF CONTENTS

01

INTRODUCTION

Let us tell you about how we got our dataset and why did we choose it?

Our Datasets:

1. **Primary: Netflix Daily Top 10**
   - Published by Prasert Kanawattanchai on Kaggle
2. **Supplementary: Netflix Movies and TV shows**
   - Published by Shivam Bansalon on Kaggle

About Dataset:

1. **Source**: Kaggle, an online crowd-sourced platform used by data scientists globally
2. **Why choose?:** Netflix, a trending topic of discussion
3. **Purpose:** To understand the Netflix consumer viewership behavior in the time of Covid crisis.

SOURCE COMPARISON

Dataset	Original Source	Kaggle Source	Reasoning
Primary: Netflix Daily Top 10	The_numbers.com, the largest freely-available database of movie business information	A subset that contain top 10 Netflix titles for the period of April 2020 to Mid- March 2022	We chose Kaggle to focus on Covid period exclusively
Supplementary: Netflix Movies and TV shows	Flixable, a search engine company that lists information of movies and shows on all streaming platforms	A subset that contains all Netflix movies and shows data added until September 25, 2021	We will use kaggle for data until Sept 2021 and also refer to Flexible for any additional 2022 specific data.

02

DATA OVERVIEW

A peek inside what these datasets include

NETFLIX DAILY TOP 10

7100 rows, 10 columns and 5 data types in the primary dataset

Netflix Daily Top 10

Column	Datatype	Description	Units	Missing Value%
As of	Date	The date of record for daily rank	Date	0%
Rank	Int	Daily rank for top 10 shows each day	Hours spent by subscribers on daily basis	0%
Year to Date Rank	Int	The rank for each show from the beginning of the year to record date	Hours spent by subscribers on daily basis	12%
Last Week Rank	String	The rank for each show from last week	Hours spent by subscribers on weekly basis	0%
Title	String	Name of the show	/	0%
Netflix Exclusive	String	Whether the show is only available on Netflix or not	/	35.2%
Netflix Release Date	Date	The date when the show is released on Netflix	Date	0%
Days in Top 10	Int	Days when the show is staying on the list of top 10	Day	0%
Viewership Score	Int	Points assigned from 1 to 10 based on the daily rank 1 to 10	Point	0%

NETFLIX MOVIES AND TV SHOWS

8807 rows, 12 columns and 3 data types in the supplementary dataset

Netflix Titles

Column	Datatype	Description	Units	Missing Value%
Type	String	The date of record for daily rank	/	0%
Title	String	Daily rank for top 10 shows each day	/	0%
Director	String	The rank for each show from the beginning of the year to record date	/	37.1%
Cast	String	The rank for each show from last week	/	9.4%
Country	String	Which country is the show made	/	9.4%
Date_added	Date	The date when the show is available on Netflix	Date	0.1%
Release_year	Date	Name of the show	Year	0%
Rating	String	Whether the show is only available on Netflix or not	/	0%
Duration	String	The date when the show is released on Netflix	/	0%
Listed_in	String	Days when the show is staying on the list of top 10	/	0%
Description	String	Points assigned from 1 to 10 based on the daily rank 1 to 10	/	0%

03

DATA PREPARATION

Steps taken to clean, format and organize our data

DATA SCOPE & CLEAN

DROP COLUMNS

Drop irrelevant columns

FORMAT DATA TYPE

Format data to make it readable and standard

REMOVE DUPLICATES

Remove any duplicate rows to avoid double counting

RENAME COLUMNS

Rename columns for making it user friendly

CREATE COLUMNS

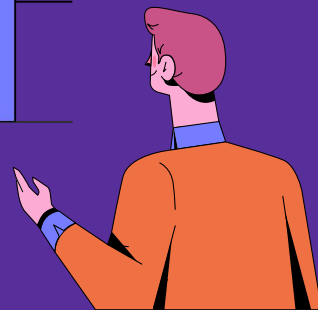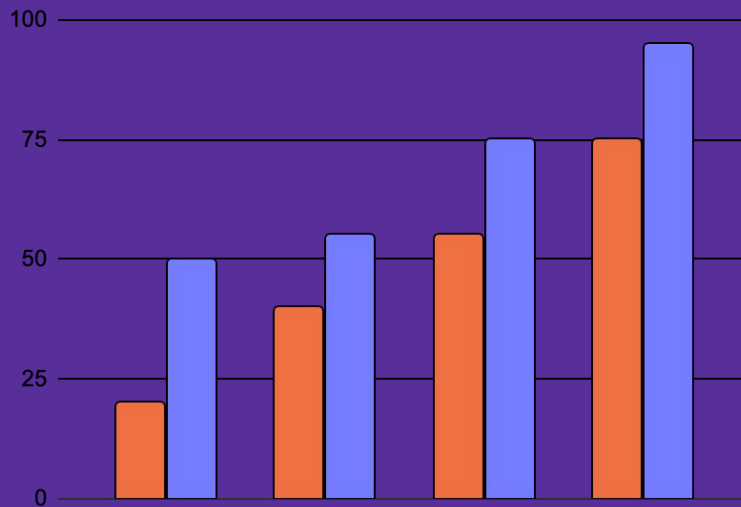Create new columns derived from existing columns for cleaner values

JOIN DATASETS

Join datasets to create analytics using both datasets
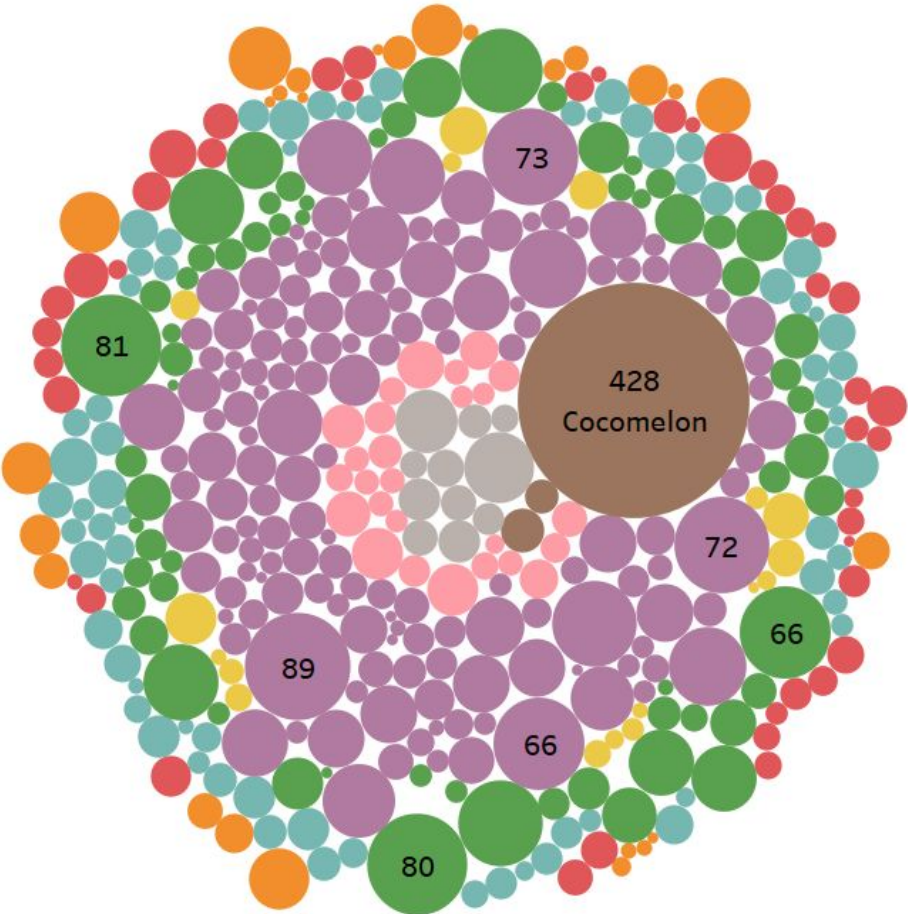
04

DATA VISUALIZATION

Insights from Visual Analytics

Netflix Program Type Popularity

As.of
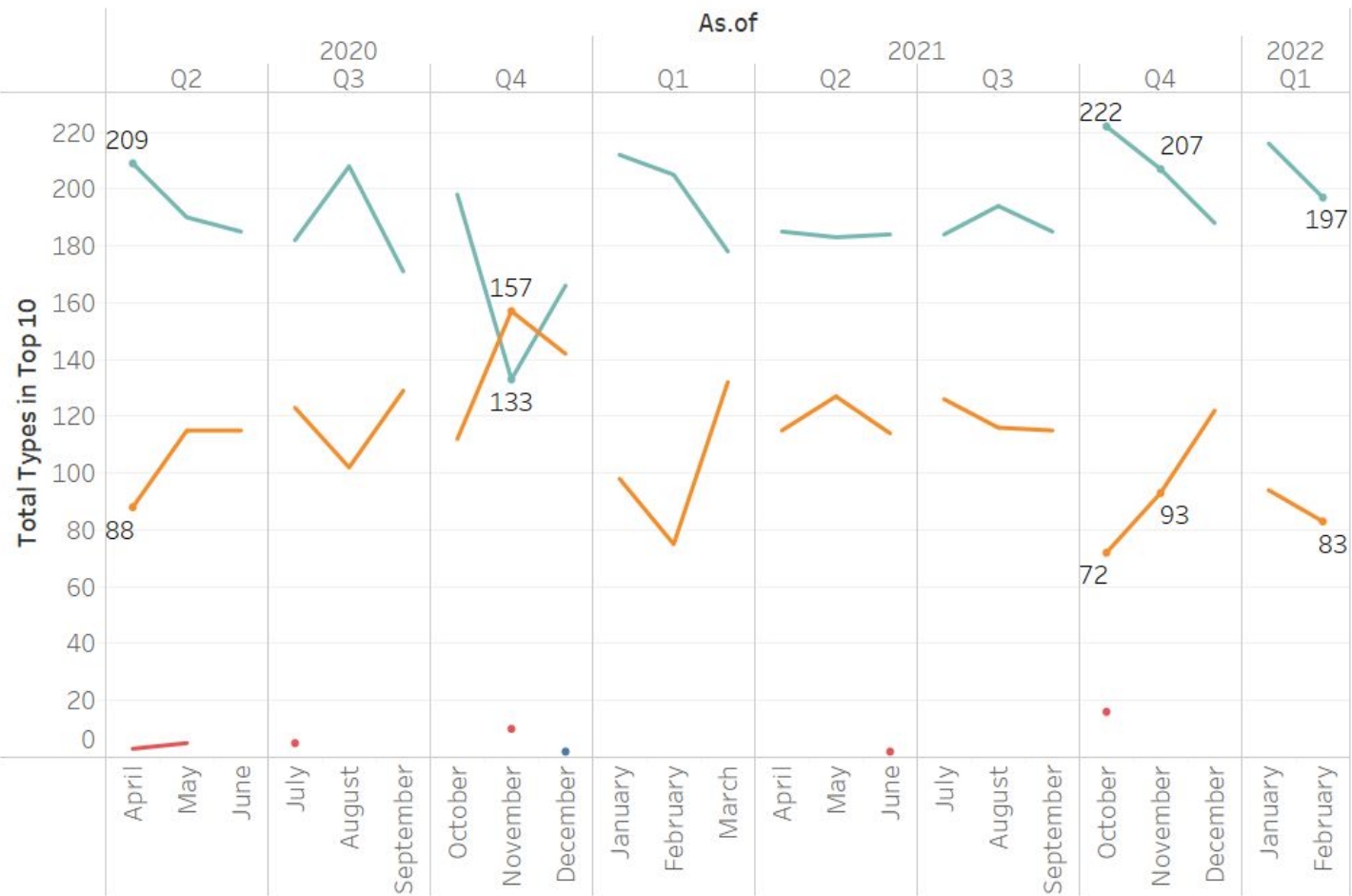
Netflix Program Type
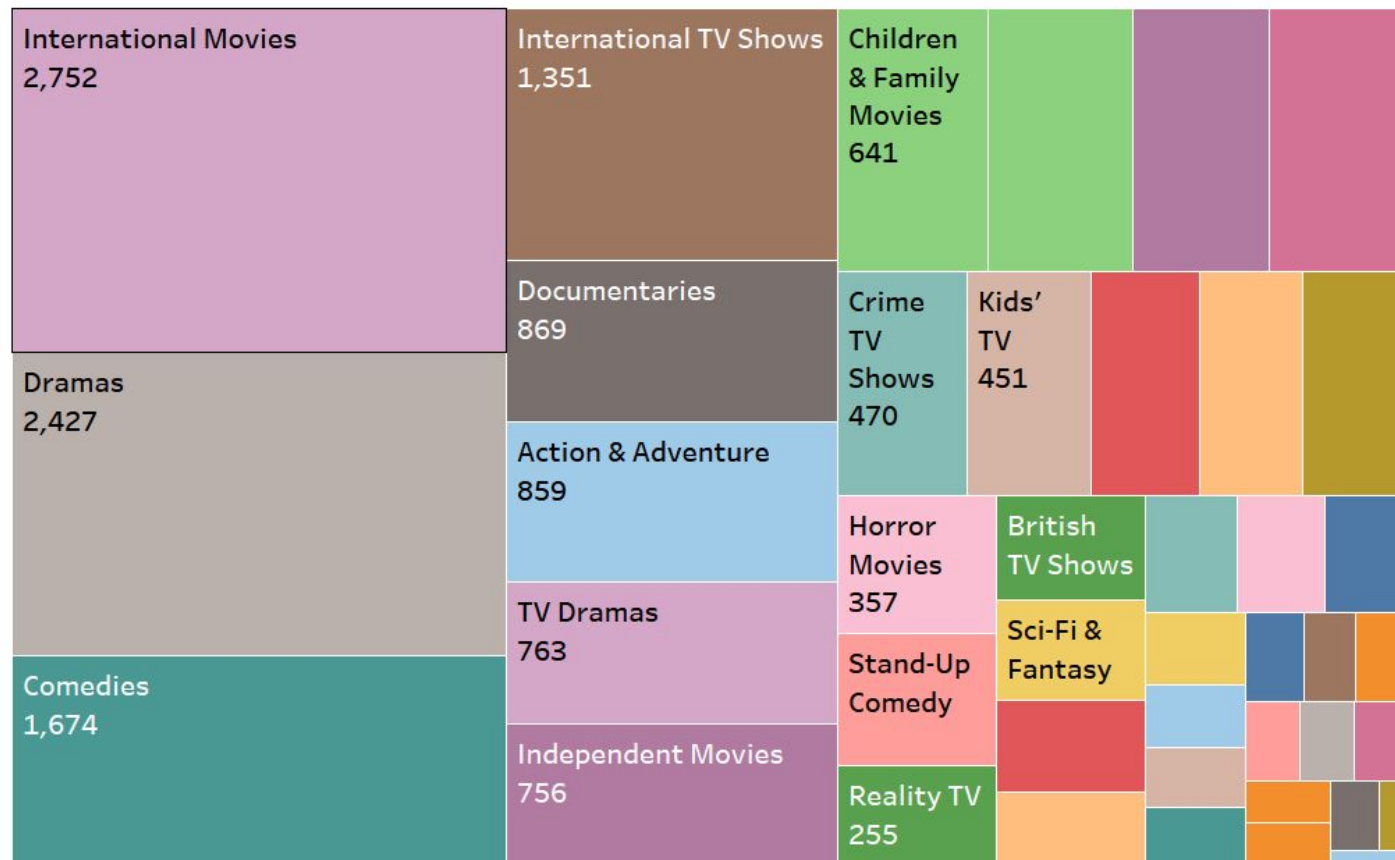- Concert/Perf...
- Movie
- Stand-Up Comedy
- TV Show

Y-axis: Total Types in Top 10

Data labels visible: 209, 88, 157, 133, 222, 207, 197, 72, 93, 83

Days in Top 10

Rating
- TV-Y7
- TV-Y
- TV-PG
- TV-MA
- TV-G
- TV-14
- R
- PG-13
- PG

Genres Distribution on Netlix USA

Genre	Count
International Movies	2,752
Dramas	2,427
Comedies	1,674
International TV Shows	1,351
Documentaries	869
Action & Adventure	859
TV Dramas	763
Independent Movies	756
Children & Family Movies	641
Crime TV Shows	470
Kids' TV	451
Horror Movies	357
British TV Shows	
Stand-Up Comedy	
Sci-Fi & Fantasy	
Reality TV	255

MAX(Genres)

- Action & Adventure
- Anime Features
- Anime Series
- British TV Shows
- Children & Family ..
- Classic & Cult TV
- Classic Movies
- Comedies
- Crime TV Shows
- Cult Movies
- Documentaries
- Docuseries
- Dramas
- Faith & Spirituality
- Horror Movies
- Independent Movies
- International Movi..
- International TV S..
- Kids' TV
- Korean TV Shows
- LGBTQ Movies
- Movies
- Music & Musicals
- Reality TV
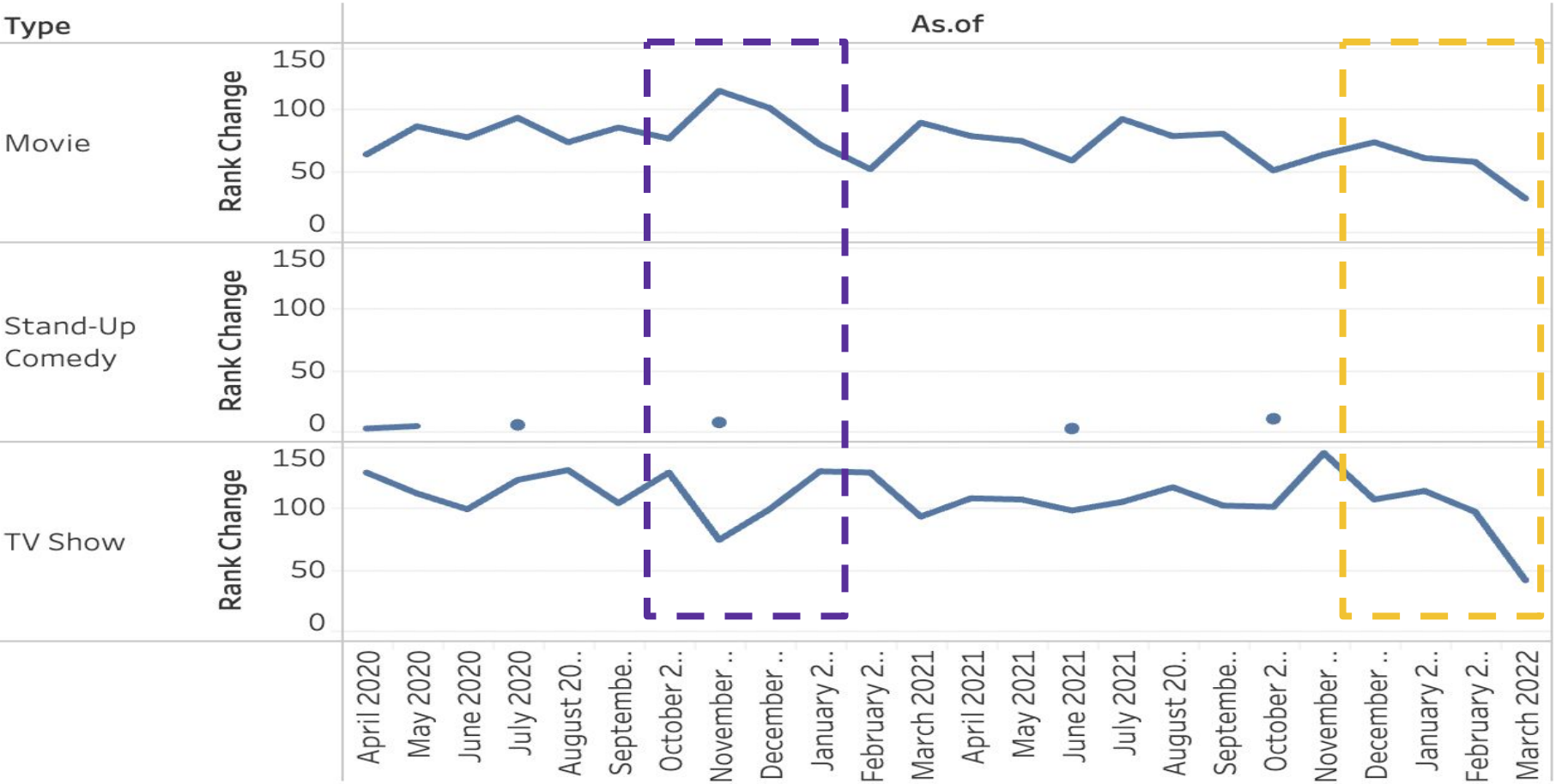- Romantic Movies
- Romantic TV Shows
- Sci-Fi & Fantasy

Monthly count of comedies/horrors in top 10
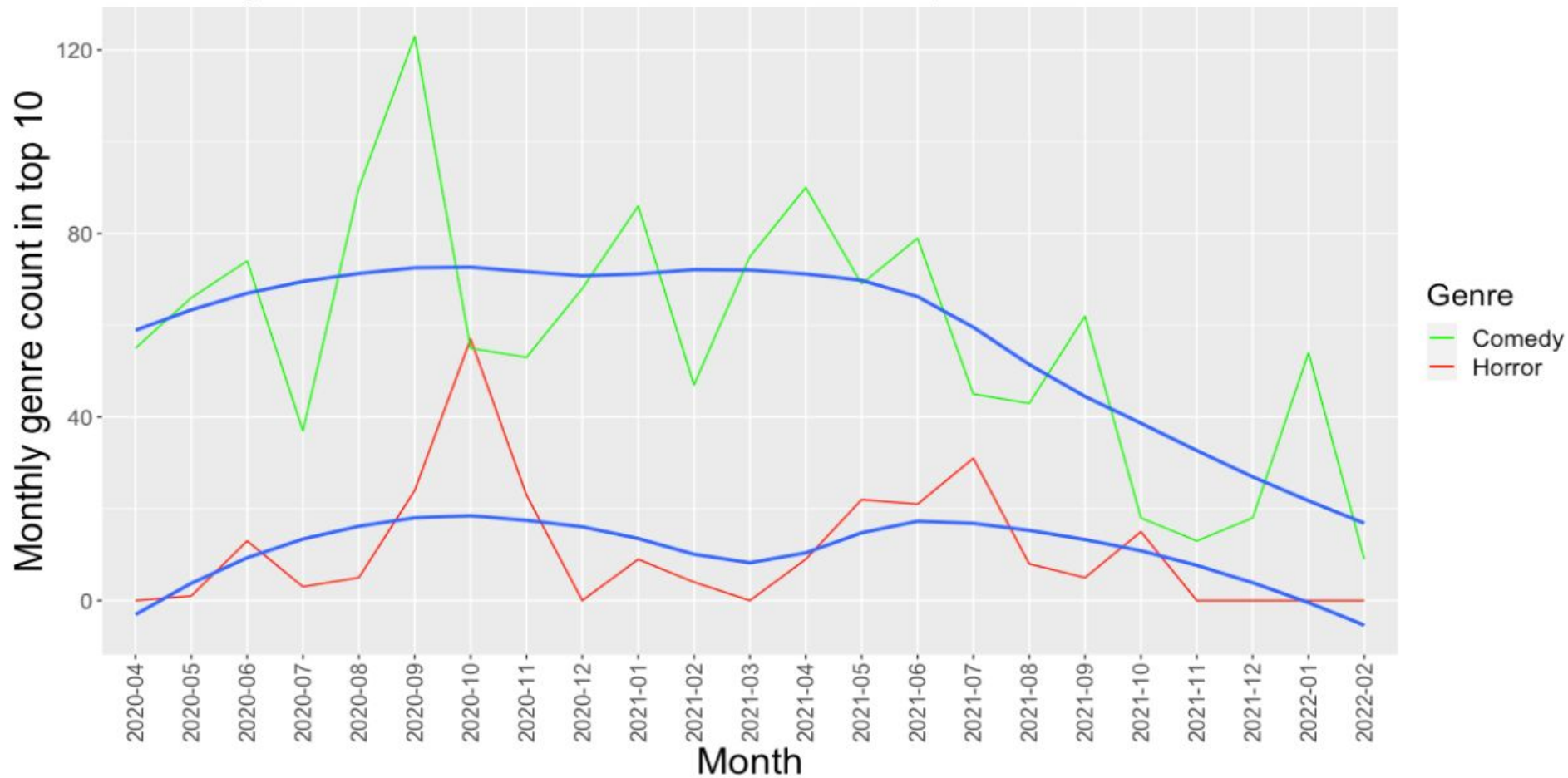
Monthly count of Netflix releases from January 2018
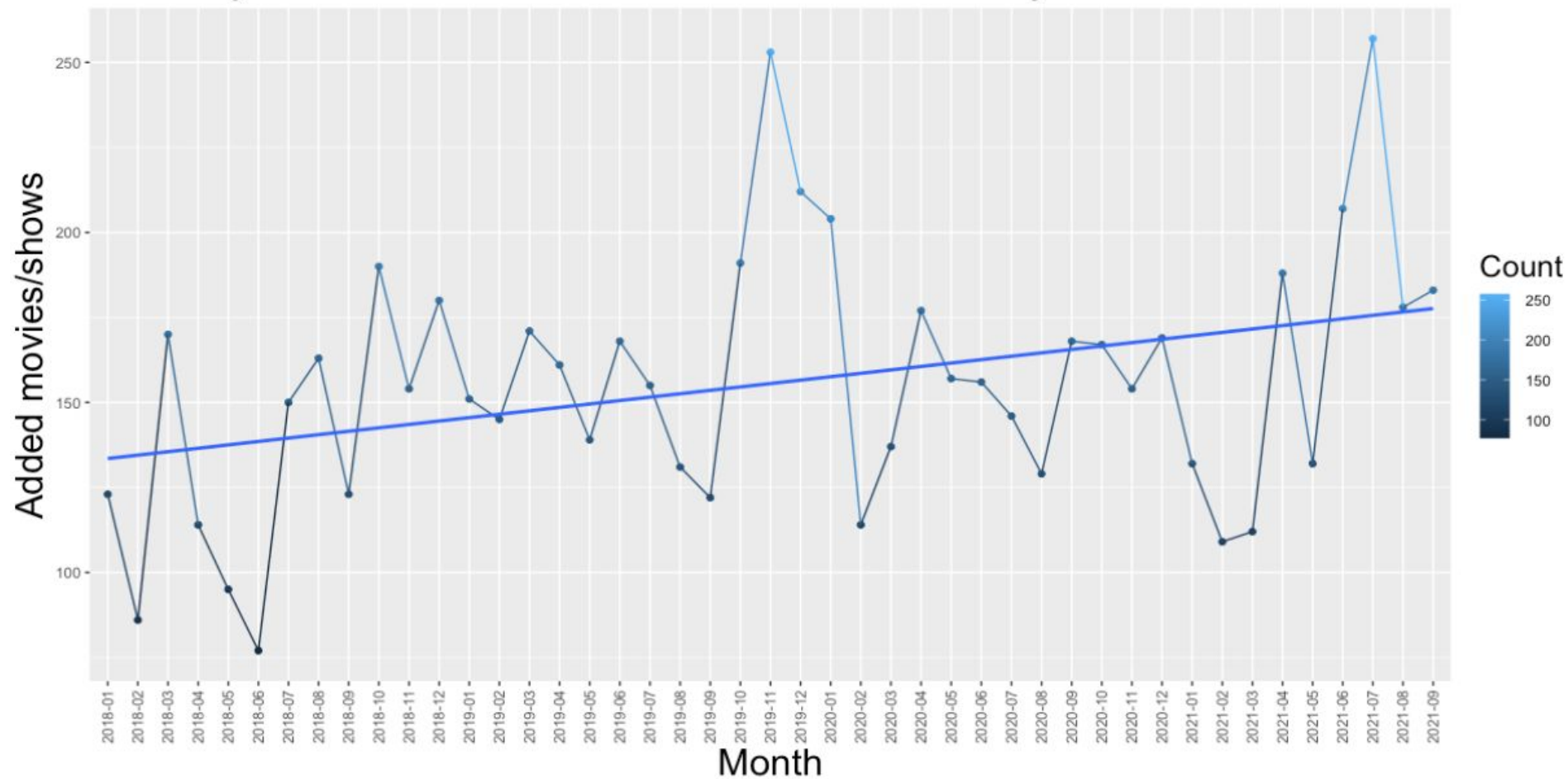
Rank Change from Apr 2020 - Mar 2022

05

CONCLUSION

Summary, Dataset usability and potential future applications

Summary of Findings:

- **Release more content to accommodate younger audience during national emergencies as the parents will need help with entertaining their children**

- **Release more comedies to keep the mood light**

- **Time spent watching horrors are correlated with comedies**

- **There is an implication that people are leaving Netflix -**
  - More Competition
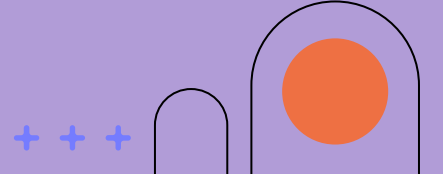 - Covid Decline
  - Boredom

Benefits:

- Covers COVID timeline
- Descriptive data on shows and movies gives us more detailed insight on patterns
- Ranking data, based on hours spent by subscribers, is a good indicator of gain in popularity of digital entertainment needs during a specific time period.
- Data is easily accessible and user friendly, making it easy to fact check and accept or oppose our findings.

Limitations:

- Limits to top 10 daily ranked on Netflix
- Geographical limitation to USA
- Cannot analyze how well the movies and shows did outside the Netflix platform such as in the theaters,different streaming platforms, etc.
- Bias in top 10 trends due to external factors such as dropping of a show or a movie from the platform due to contracting, etc. reasons.
- Missing data/ incomplete data
- Lack of standardization of string values causing difficulty in handling the data

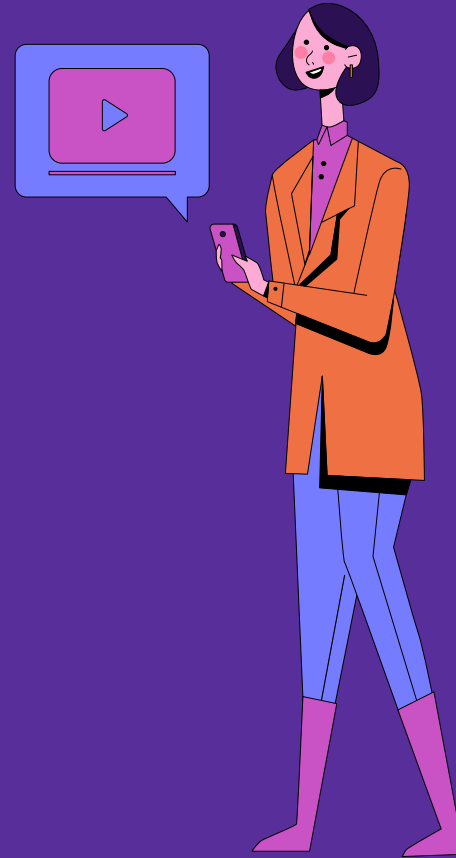# Future goals and applications:

To understand:
- How will consumers potentially behave in next home confiding catastrophe in terms of entertainment needs and how will they behave in a normal setting after a catastrophe?
- How to budget/invest for future movie/show releases?
- Which entertainment types/categories need more marketing to make up for the production expenses during covid?
- Future competitive analysis of streaming platform performance?

THANKS!

Do you have any questions?

REFERENCES

Datasets:

Main: https://www.kaggle.com/datasets/prasertk/netflix-daily-top-10-in-us

Supplementary: https://www.kaggle.com/datasets/shivamb/netflix-shows?resource=download
https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores

Netflix News:

https://www.forbes.com/sites/jonathanponciano/2020/10/19/netflix-earnings-5-numbers-growth-continues-during-the-coronavirus-pandemic/?sh=7cdeb05d225e

https://time.com/6198354/netflix-subscribers-loss-second-quarter-future/

https://www.nytimes.com/interactive/2021/us/covid-cases.html

https://www.bbc.com/news/business-62226912

https://9to5mac.com/2022/09/22/netflix-subscribers-planning-to-leave/