



# **NETFLIX ORIGINAL PRODUCTION ANALYSIS**

Team 4  
Emory's Next Top Models



# Inspiration

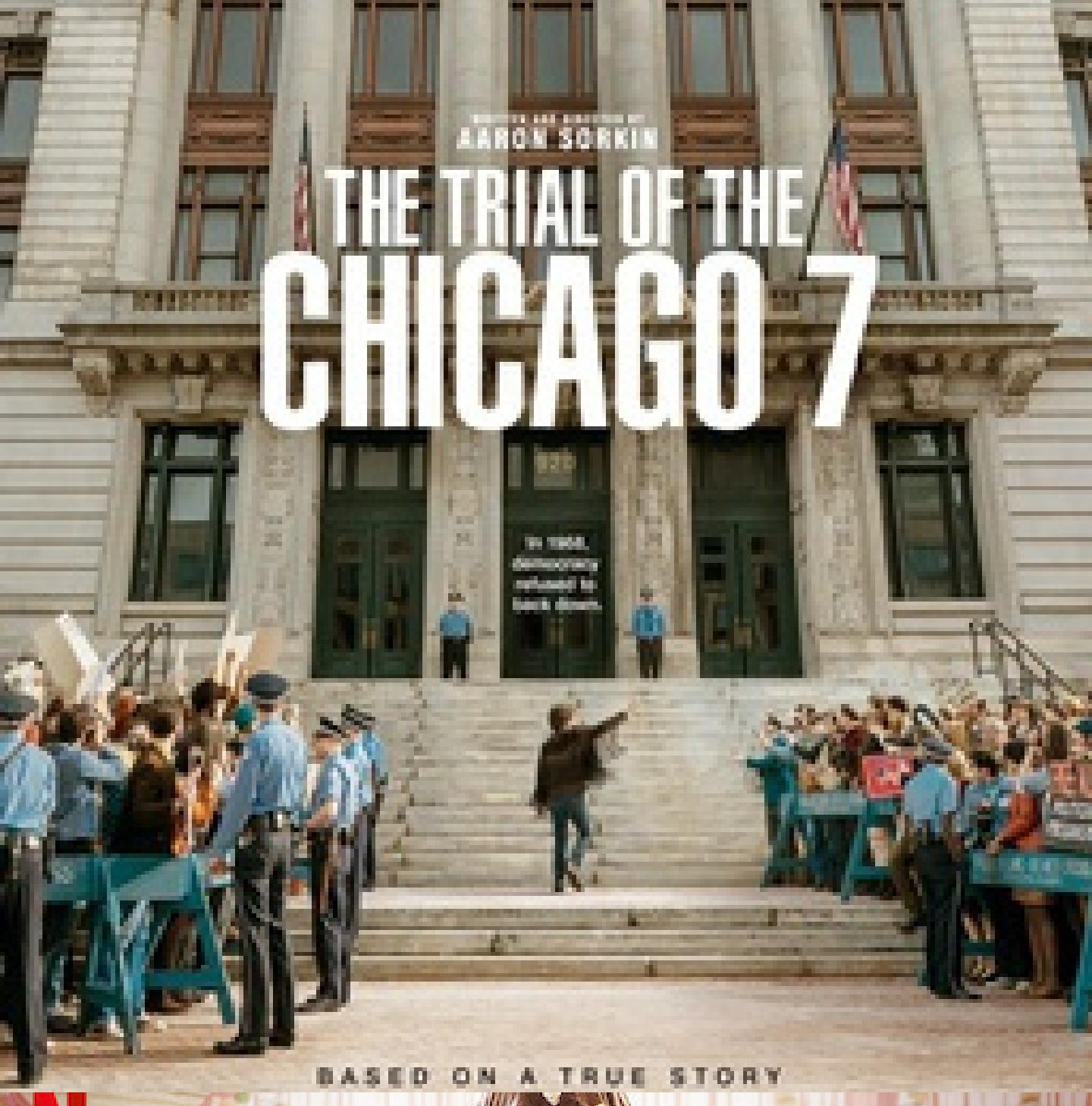
COVID - 19

Streaming War – Competitive Advantage

Netflix Original Series

How can Netflix grow their business and prevent subscriber churn?





# Goals

- Collect data on Netflix Original Productions
- Analyze key players and critical reception
- Analyze viewer trends and popular sentiment using Twitter
- Give recommendations on which productions to renew and which not to





# TOP PICKS FOR YOU



TOP  
10

Neha Bansal

98% Match 1 Season TV-MA 2020

Soapy • Campy • Quirky



TOP  
10

Emmy Fortunato

98% Match Limited Series TV-MA 2020

Cerebral • Intimate • Emotional



TOP  
10

Boping Zhang

98% Match 2 Seasons TV-MA 2019

Relaxing • Feel-Good • Reality TV





# Managing Big Data

- Original movie and TV show data from 3 streaming platforms
- **Data Wrangling:**
  - S3
  - PySpark
  - PySpark-SQL



hulu

Disney+

NETFLIX



# Data Analysis



Netflix: 3681

Hulu: 141

Disney+ : 22

## Movies

Netflix: 2473

Hulu: 61

Disney+: 18

## TV Shows

Netflix: 1208

Hulu: 80

Disney+: 4

Hulu

6.9

Disney+

6.4

Netflix

6.26

Disney+

7.57

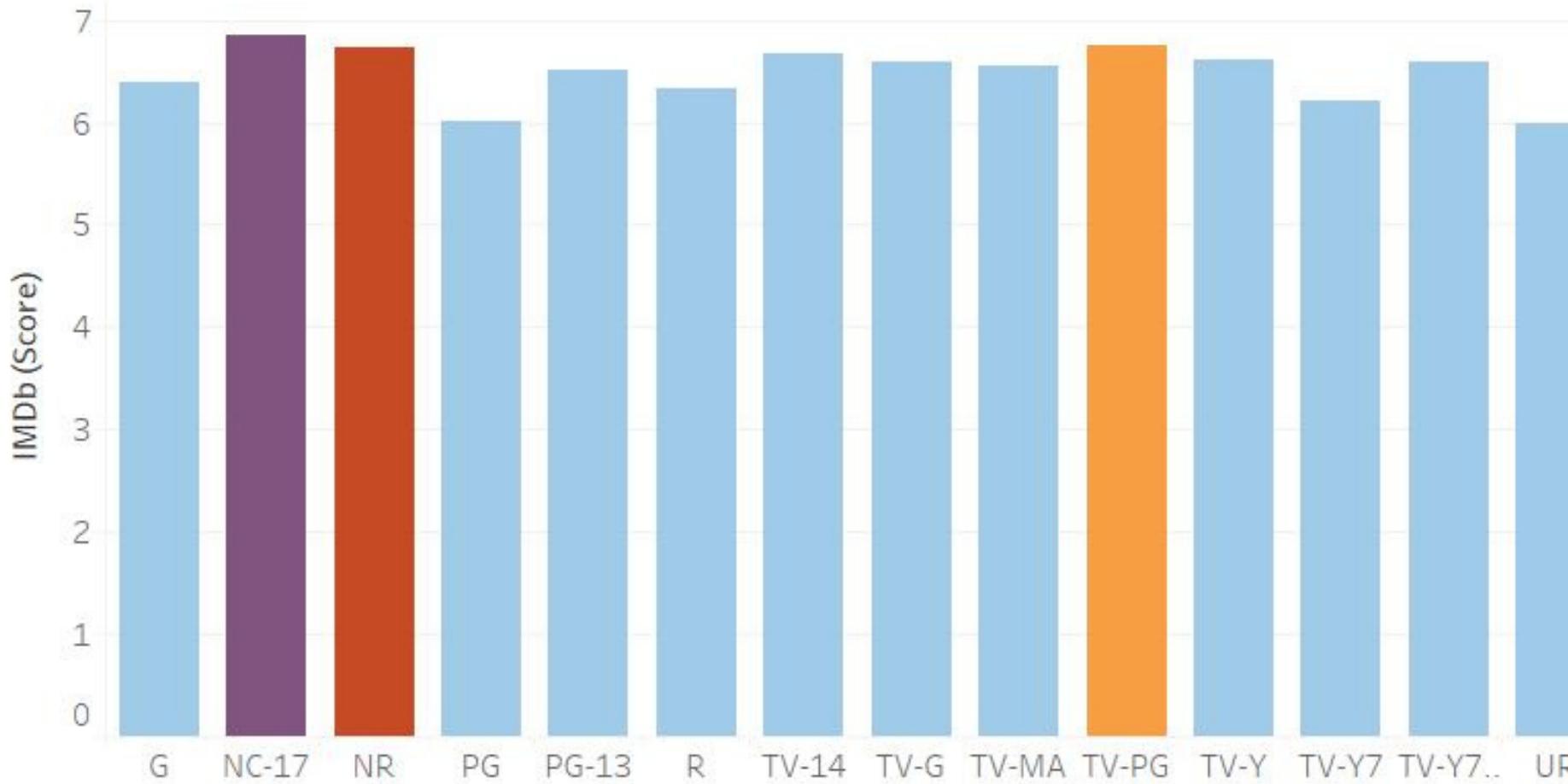
Hulu

7.53

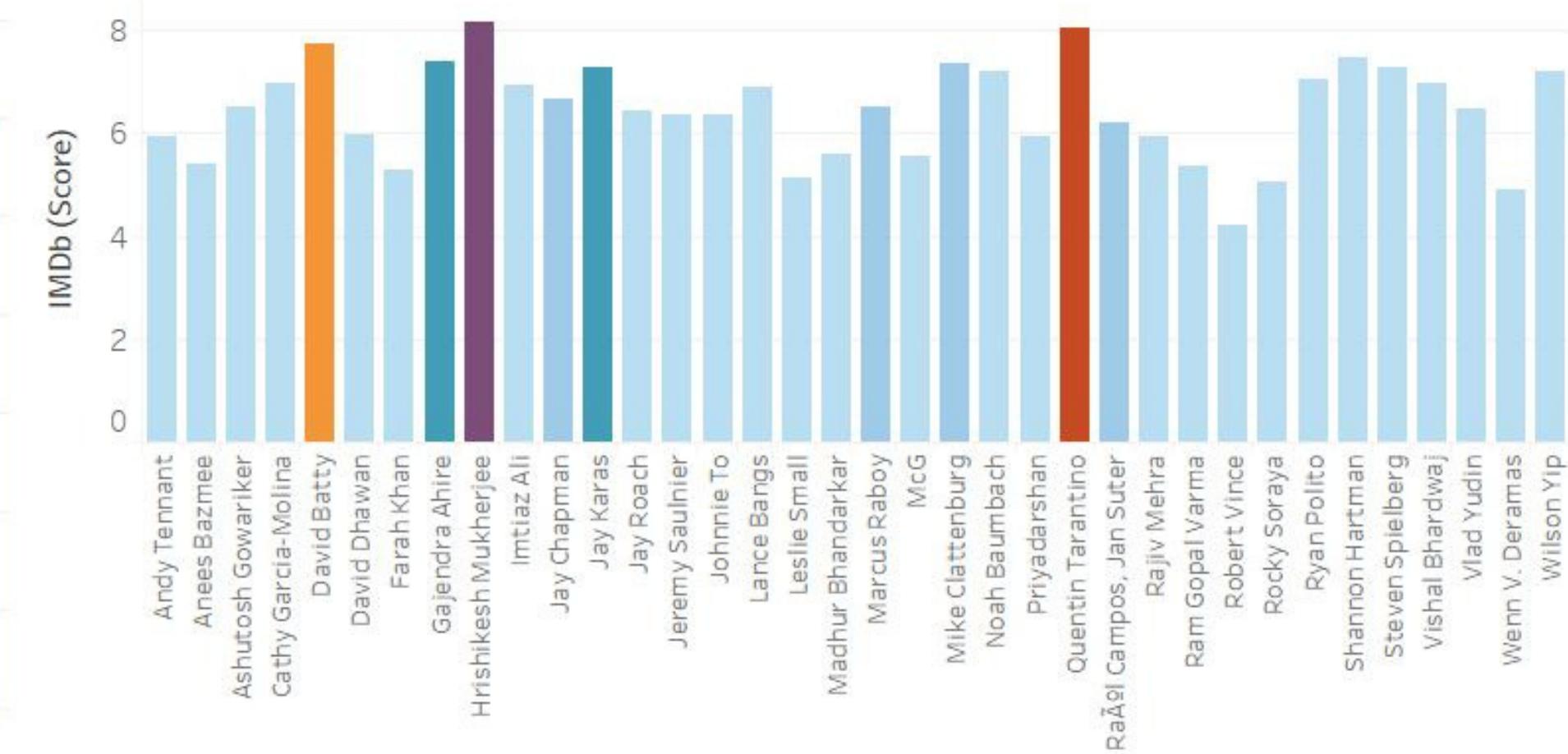
Netflix

7.21

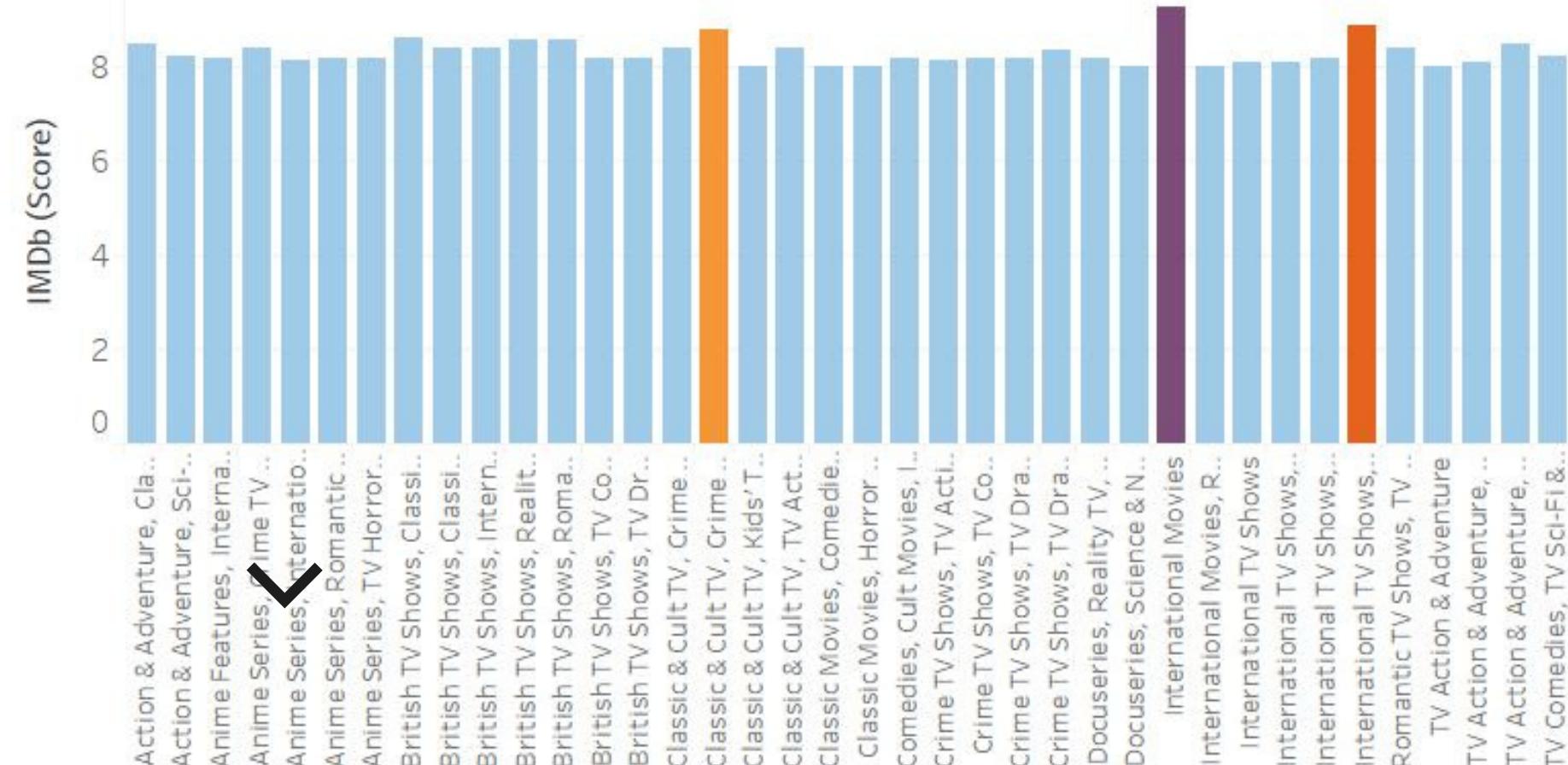
## <Average IMDb(Score) Segmented by Different Rating>



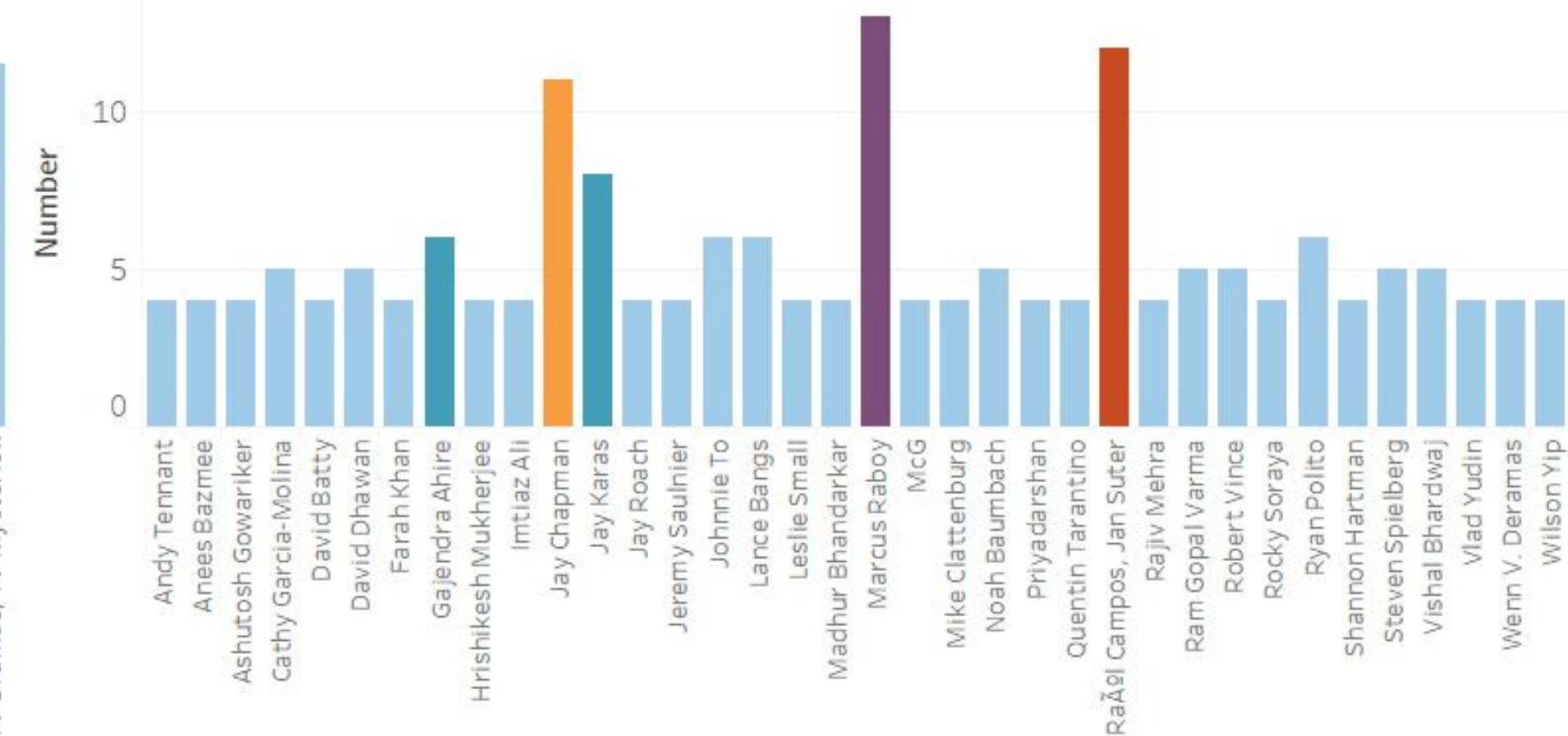
## <Average IMDb(Score) by Each Director>



## <Average IMDb(Score) Segmented by Each Category



## <Total Number of Movies Segmented by Each Director>



# Sentiment Analysis



## Data Sourcing:

```
# SEARCH TWEETS CONTAINING #Netflix
# Define the search term
search = "#Netflix -filter:retweets" # using '-filter:retweets' removes all retweets

# COLLECT TWEETS - 3000 tweet sample from each day for the last week
# using 3000 based on test - no days with more than 3k tweets written in English with hashtag '#Netflix'
# queries broken down by day so that they can be queried within API rate limitations
# Get Tweets for 11/19/20
tweets = tw.Cursor(api.search,
                    q=search,
                    lang="en",
                    since='2020-11-19',
                    until='2020-11-20',
                    tweet_mode='extended').items(3000)

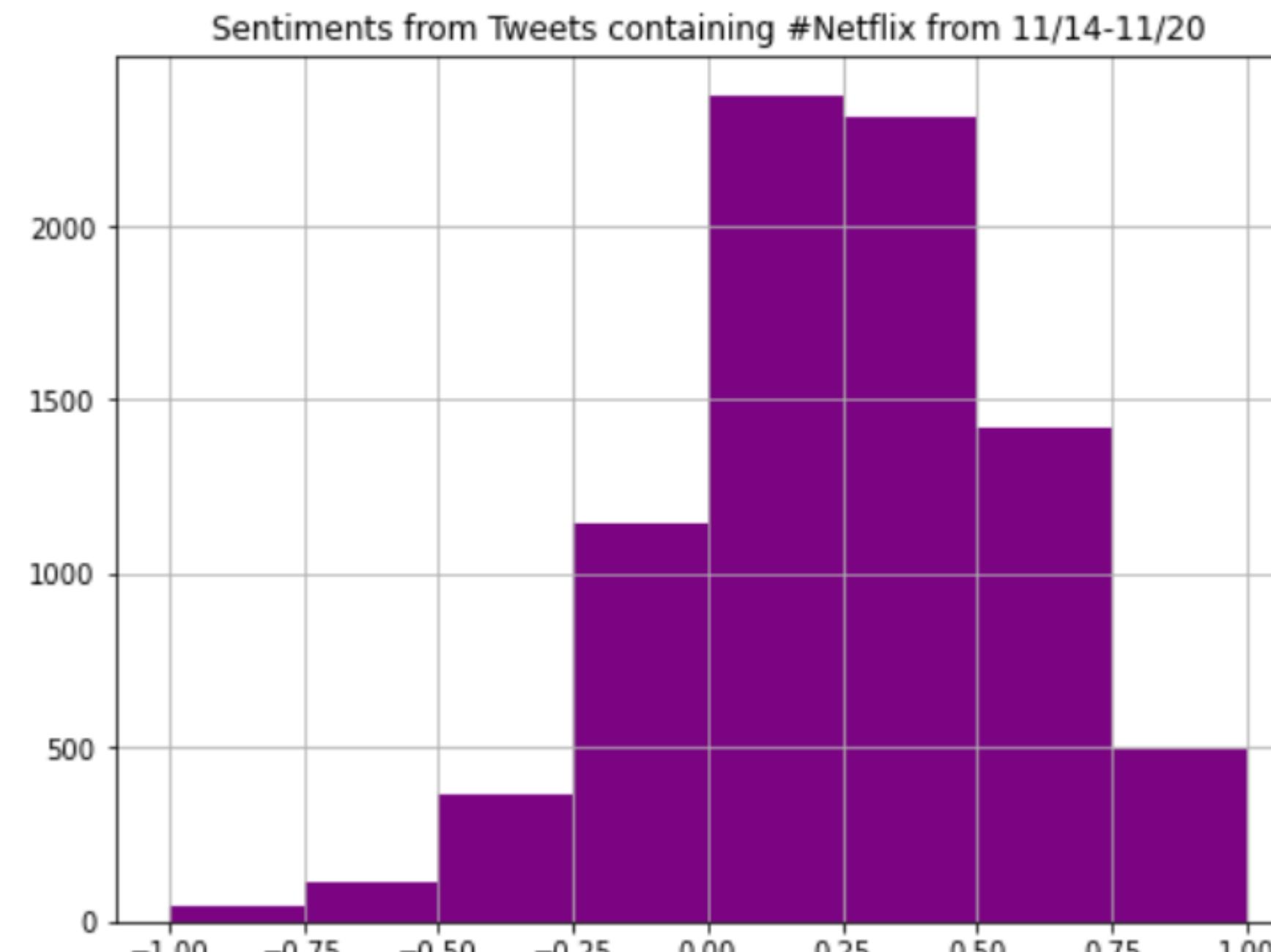
tweet_text = [tweet.full_text for tweet in tweets]
tweet_text[:5]

['Thoughts on Cagayan. #SurvivorCagayan, #Netflix https://t.co/BgQeb4deGM',
 'START-UP | Day 34 | Top 10 list in 17 countries | Top 8 on Netflix in World\nAvailable in 25 countries\n10 episodes\n-EP 11 : 21 Nov\n-EP 12 : 22 Nov\n#스타트업 #BaeSuzy #Suzy #Netflix @Suzy https://t.co/aM0ZUJd7Sr',
 'Nobody: Freestyle Unity: "It's time for Gary ... GARRYYYYY... and Company" #TheParkers #Netflix',
 '#PedroPascal #boydholbrook #taylordooley #WeCanBeHeroes #Netflix #Netflixfilm @PedroPascali ❤️ @boydHolbrook1 @taylordooley in We Can Be Heroes, so powerful, wonderful, awesome & special, I love it so much specially for my sweet love Super Pedro Pascal, can't wait to see more https://t.co/HqnkAmhfqY',
 'My podcast on the beauty and brilliance of the #QueensGambit on #netflix https://t.co/VZ2SLoBFQH']
```



### words count

thecrown	953
amazon	864
good	695
disney	674
love	565
apple	536
crown	476
christmas	425
queens	420
survivorcagayan	398
great	397
gambit	391
thequeensgambit	359
thecrownnetflix	341
savetheorder	328
best	315



## Analysis & Insight:

- Generally positive
- Trending productions:
  - The Crown
  - The Queen's Gambit
  - Survivor
  - The Order
  - Jingle Jangle
- Fierce competition between platforms resulting from cancellation of shows



# Recommendations

## **Content to release now:**

- Michael Steed Productions
- Nostalgic Reality Shows
  - The Amazing Race
- Strong Female Leads
- Christmas Content
- Any combinations of the above



# QUESTIONS?

