ASSIGNMENT

Analysis of Netflix Content

Jagruth B

DATASET DESCRIPTION

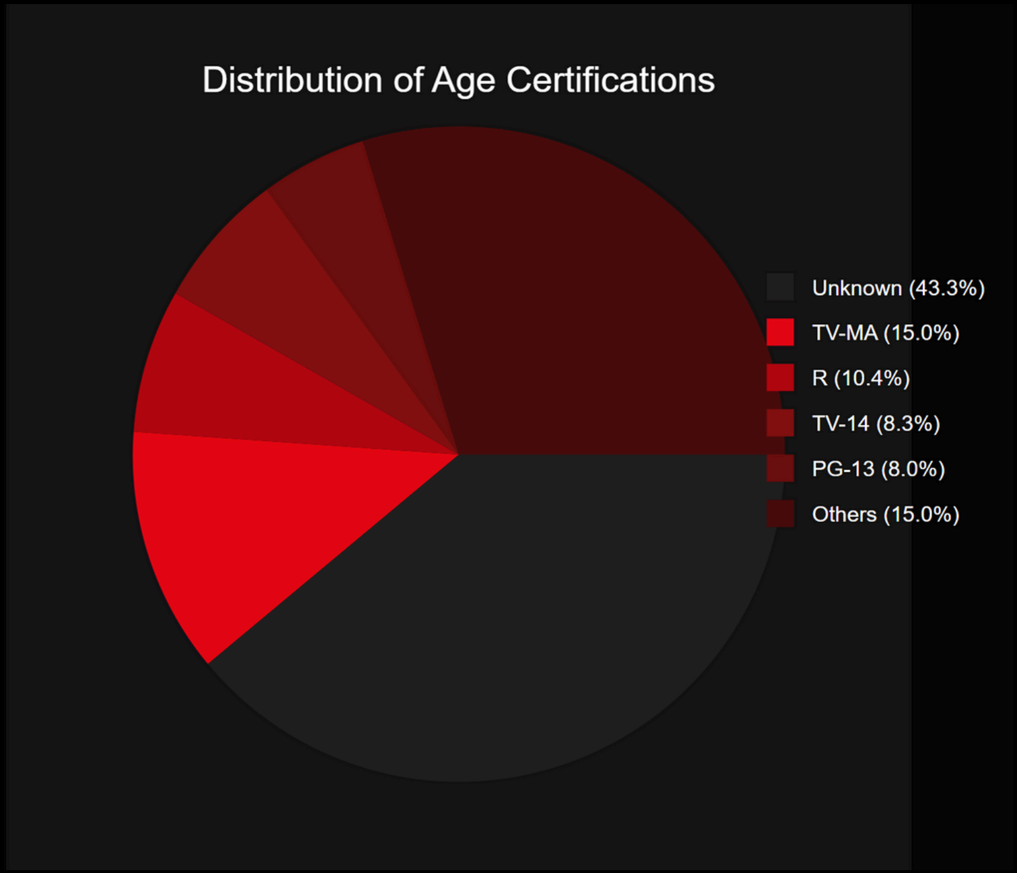
The dataset comprises Netflix content information merged with IMDb ratings data, containing 5,283 unique entries. Each entry includes:

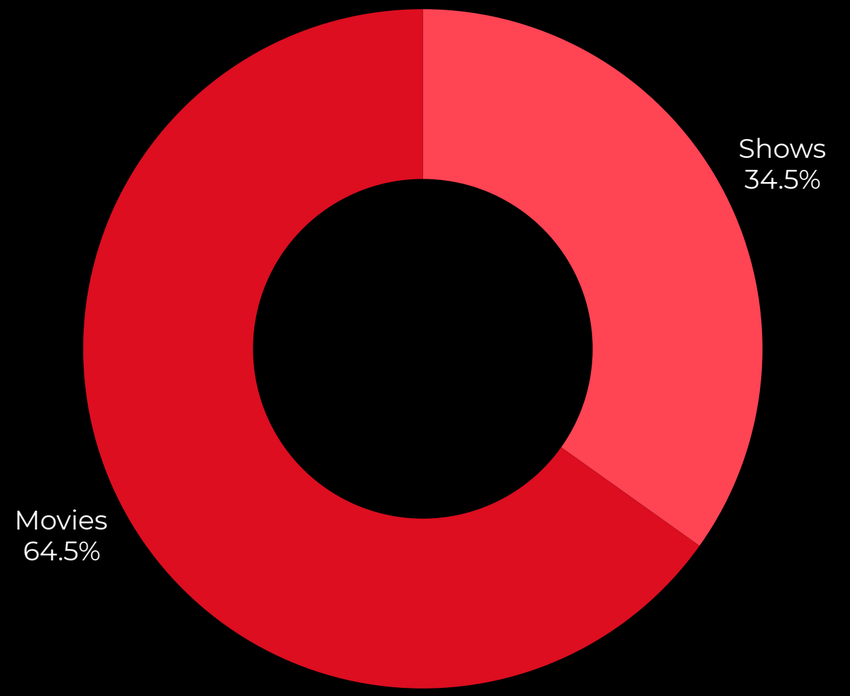
- Title: Name of the movie or TV show
- Type: Classification as either MOVIE or SHOW
- Description: Plot summary or content description
- Release Year: Year of release (ranging from 1953 to 2022)
- Age Certification: Content rating (e.g., R, PG-13, TV-MA)
- Runtime: Duration in minutes
- IMDb Score: Rating from 1 to 10
- IMDb Votes: Number of user votes received

DATA CLEANING PROCESS

- Cleaning Steps Implemented
  - Identified 2,285 missing values in age certification
  - Found 5 missing descriptions
  - Detected 16 missing IMDb votes
  - Applied appropriate filling strategies:
    - Age certification: Filled with 'Unknown'
    - Descriptions: Replaced with 'No Data'
    - IMDb votes: Used median imputation based on content type and certification

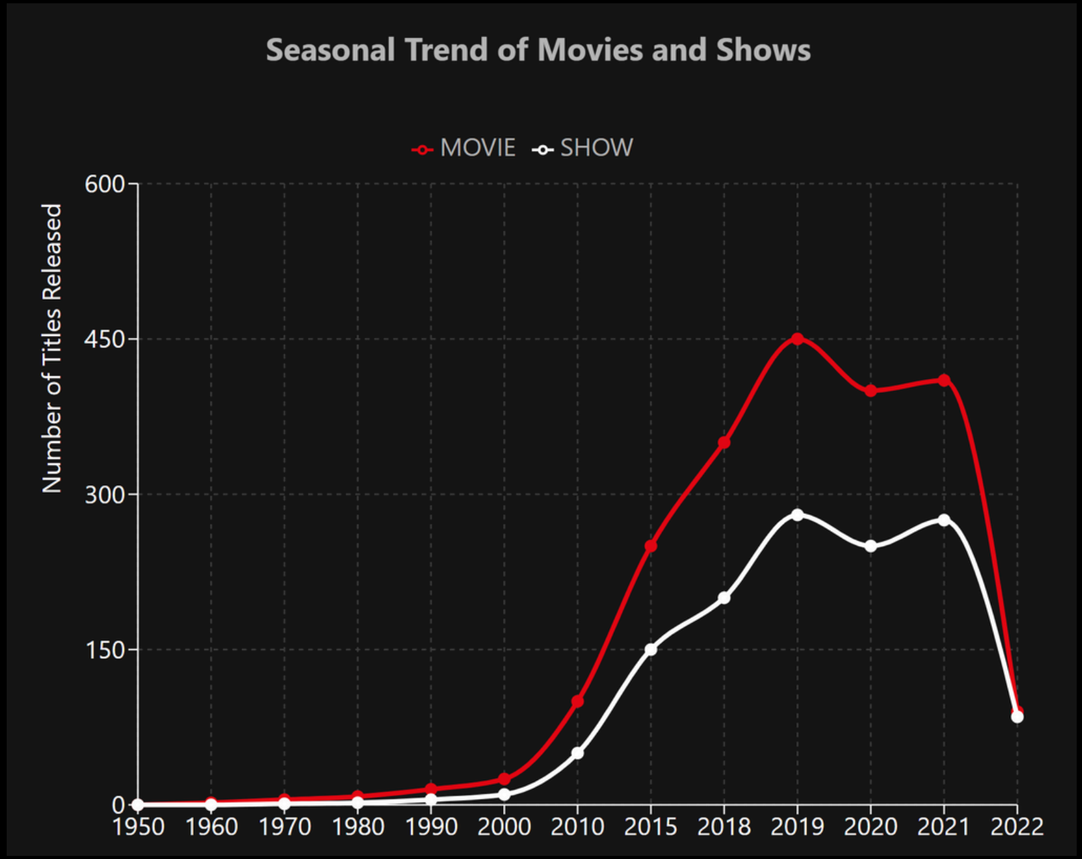
TRENDS & INSIGHTS



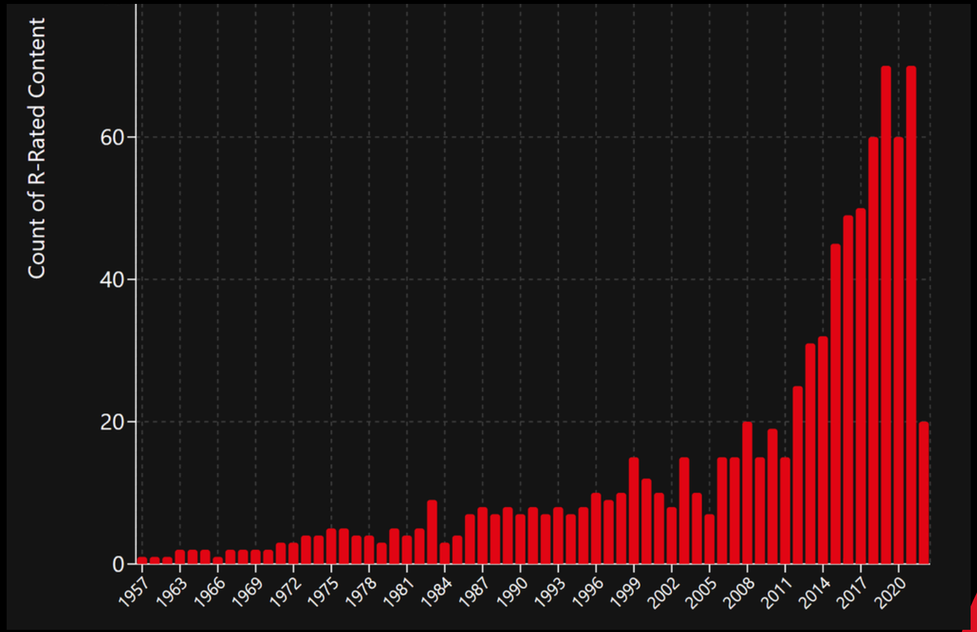


- **Content split shows a dominance of movies (64.5%) over TV shows (35.5%)**
- **Age certification analysis reveals that TV-MA accounts for the highest proportion of content at 15.0%, followed by R-rated at 10.4%, TV-14 at 8.3%, and PG-13 at 8.0%, indicating a significant focus on mature audiences.**

TRENDS & INSIGHTS

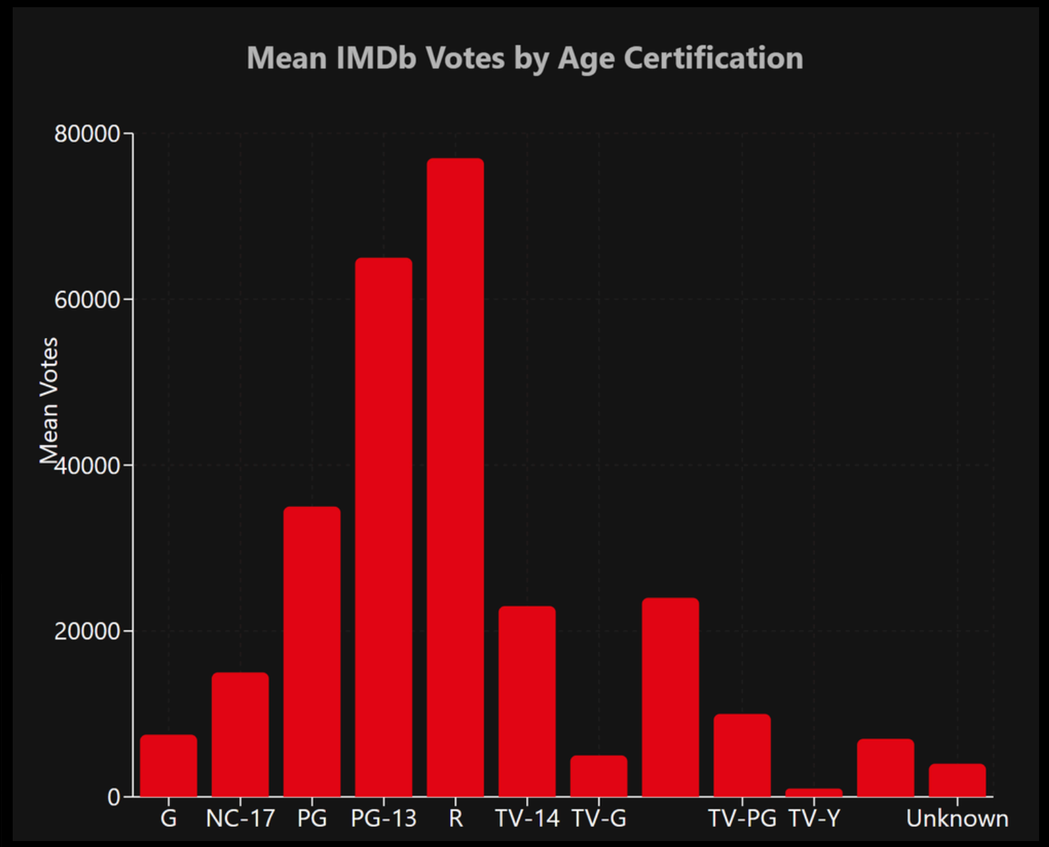


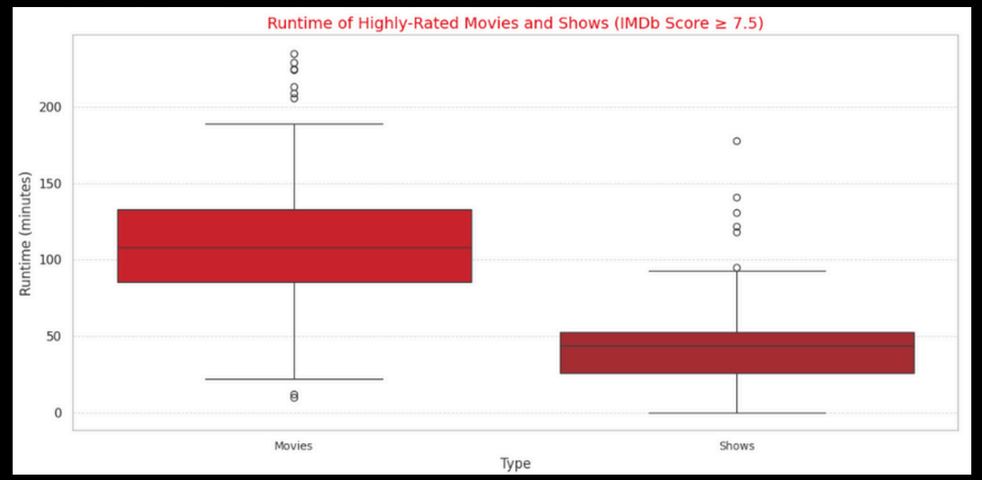
- **Netflix's content releases show dramatic growth post-2010, with movies peaking at 450 annual releases in 2019 and shows reaching 280, highlighting their aggressive expansion in both formats while maintaining a higher ratio of movies to shows.**
- **R-rated content on Netflix shows steady growth from 1957-2000 (1-15 titles/year), followed by rapid expansion 2010-2019 (reaching 70 titles)**

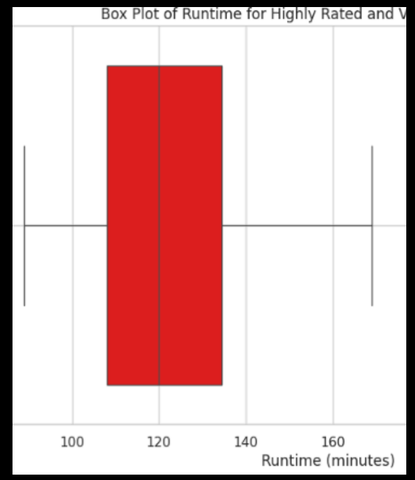


TRENDS & INSIGHTS

- **The boxplot reveals distinct runtime patterns: Movies rated above 7.5 cluster around 100-150 minutes, while highly-rated shows maintain shorter 30-50 minute episodes, optimizing viewer engagement while maintaining quality storytelling.**
- **R-rated content leads with 77K mean votes, demonstrating strongest viewer engagement, followed by PG-13 with 65K votes.**

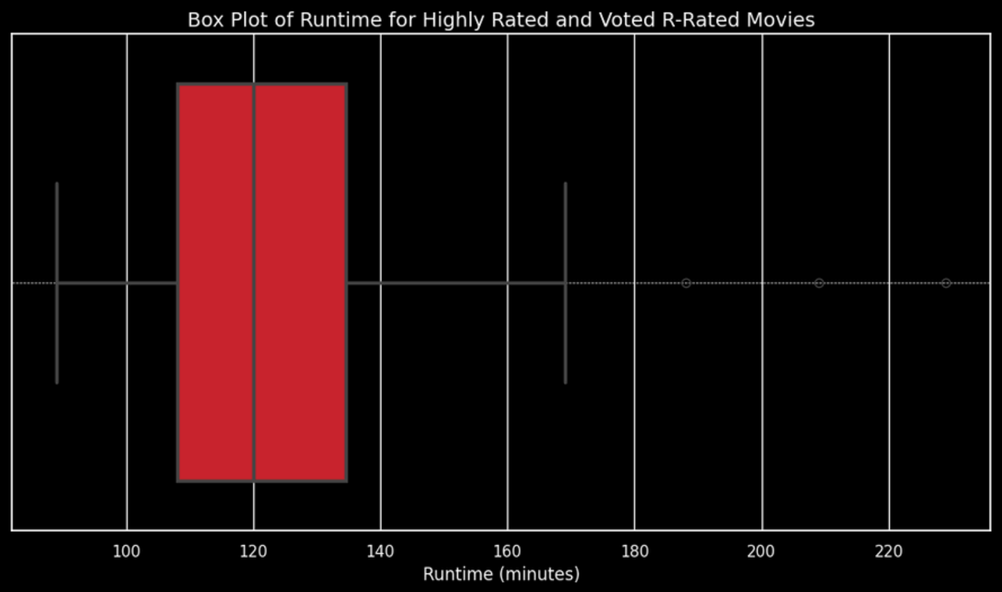


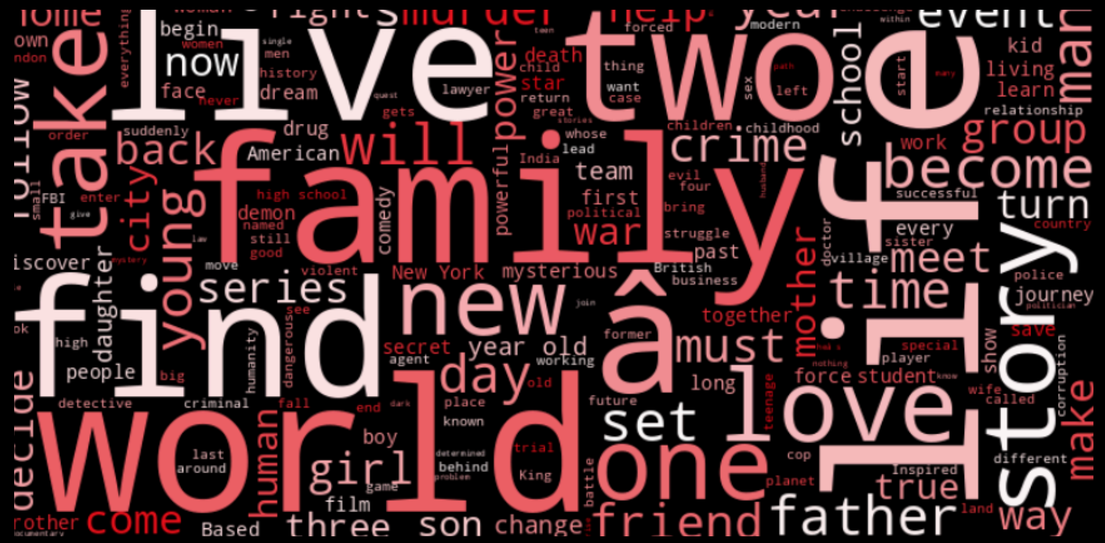




TRENDS & INSIGHTS

- Successful R-rated movies have an optimal runtime between 120-140 minutes, with the median around 130 minutes, the concentrated distribution suggests audiences strongly prefer R-rated content within this 2-hour sweet spot.
- The word cloud reveals prominent themes in highly-rated content across age certifications:
  - Core themes focus on "family" relationships (father, son, friend) and human connections
  - Strong emphasis on temporal elements ("time", "new", "day") suggesting dynamic storytelling
  - Action/conflict elements ("murder", "fights", "war") balanced with emotional content ("love", "help")
  - This visualization suggests successful content combines personal narratives with action elements while maintaining strong character relationships.





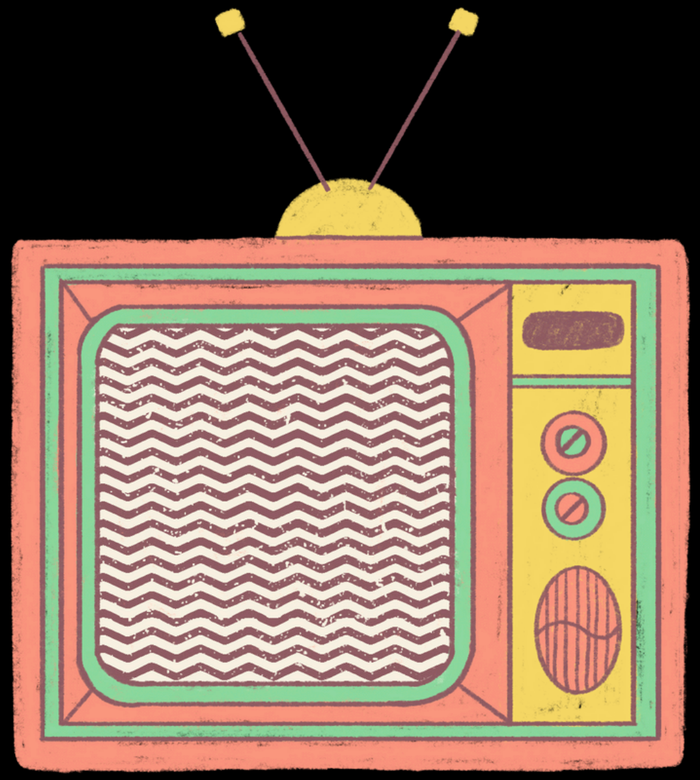
KEY HYPOTHESIS

- Mature Content Evolution Hypothesis:
  - The surge in R-rated content (70+ titles by 2019) reflects growing audience sophistication and demand for complex narratives
  - Higher engagement metrics for R-rated content (77K mean votes) suggest mature audiences are more actively involved
- Runtime Optimization Hypothesis:
  - 120-140 minute sweet spot for R-rated movies maximizes storytelling impact while maintaining audience engagement
  - Shorter runtimes (<100 minutes) may inadequately develop mature themes
  - Outliers beyond 180 minutes show diminishing returns

- Quality-Rating Correlation Hypothesis:
  - TV shows (TV-14, TV-MA) consistently outperform movies in ratings
  - Average rating of 7.26 for TV-14 content suggests structured episodic format better serves complex narratives
  - R-rated movies' average 6.32 rating indicates room for quality improvement
- Content Theme Impact Hypothesis:
  - Success correlates with balanced mixing of action elements (fights, murder) with emotional depth (relationships, personal growth)
  - Family/relationship themes remain central across ratings, suggesting universal appeal

RECOMMENDATION FOR R RATED CONTENT

- Recognizing the recent trend towards R-rated content, we should start developing R-rated shows that align with audience preferences, typically ranging from 30 to 50 minutes per episode, as this format has gained traction and resonates well with viewers seeking deeper narratives and mature themes
- To enhance viewer engagement among r rated movies, we recommend a target runtime of 120-140 minutes, ensuring tight storytelling while avoiding excessive lengths unless narratively essential.
- Our content strategy should prioritize character-driven narratives that blend action with emotional depth, focusing on themes such as complex family dynamics, personal transformation, moral conflicts, and contemporary social issues.
- We propose releasing 5-6 R-rated titles per quarter to maintain a steady flow of content while emphasizing quality over quantity during peak engagement periods.



THANK YOU

JAGRUTH BACHALAKURA
https://www.linkedin.com/in/jagruthb/