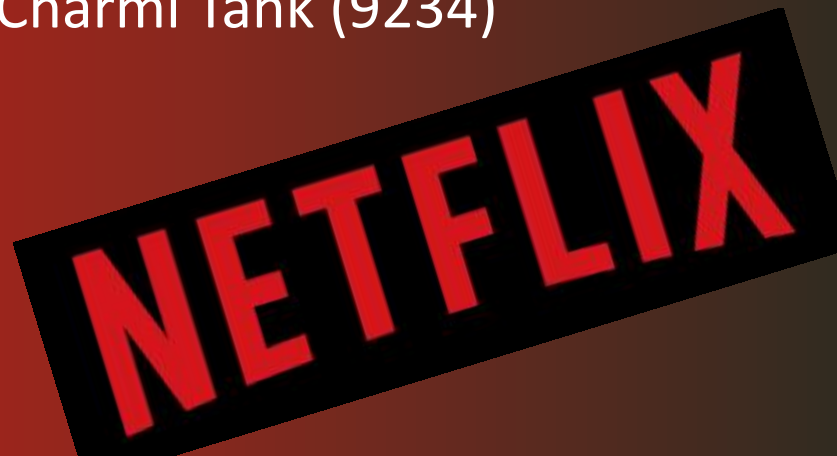
Unveiling Data-Driven Strategies for Streaming

Atharva Pawar (9427)

Aditya Vyas (9238)

Charmi Tank (9234)



1.Unreeling Netflix: Understanding and Improving Multi-CDN Movie Delivery

The paper provides a measurement study of Netflix's architecture and service strategy, revealing its use of data centers and Content Delivery Networks (CDNs) for content distribution.

The authors propose a measurement-based adaptive CDN selection strategy and a multiple-CDN-based video delivery strategy, demonstrating their potential in significantly increasing user's average bandwidth

Methodology

THE CURRENT SYSTEM:

Static assignment: Netflix assigns a single CDN to a user for a long time, even though bandwidth on CDNs can fluctuate.

Backup reliance: The other two CDNs are mainly backups, only used when the primary CDN can't deliver even low-quality video.

ALTERNATE VIDEO DELIVERY STRATEGIES

A. Room for improvement

B. Measurement based CDN selection

C. Using multiple CDNs simultaneously

Findings

The findings of the study can be beneficial for future bandwidth-demanding services, such as 3D movies, by increasing video delivery bandwidth and enabling the provision of more satisfying experiences.

The paper highlights the potential for new video delivery strategies and the utilization of multiple CDNs to improve user experience and accommodate evolving services.

2 Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms

Methodology

- The study conducted an exploratory analysis of data obtained from Flixable, which is a search engine that lists the content available on Netflix.
- A dataset of 7,787 unique records was analyzed to highlight essential information about the content available on this platform.
- Implemented a recommendation system using the TF-IDF and Cosine similarity algorithms, which are models widely used in Natural Language Processing (NLP).
- The exploratory analysis has revealed interesting data on the current trends of the content delivered on Netflix.
- Despite the limitation of the recommendation system in its current state, it looks promising when additional features are taken into consideration.

NETFLIX

FLIXABLE

Findings

Unequal Content Distribution: The US, India, and UK hold over 75% of Netflix content, with Africa having significantly less.

Movies Still Dominate (for now): Currently, most content is movies (69%), but TV shows are catching up (31%).

Shifting Trends: Since 2018, TV shows are being added faster than movies. In 2020, TV shows even surpassed movies for the first time.

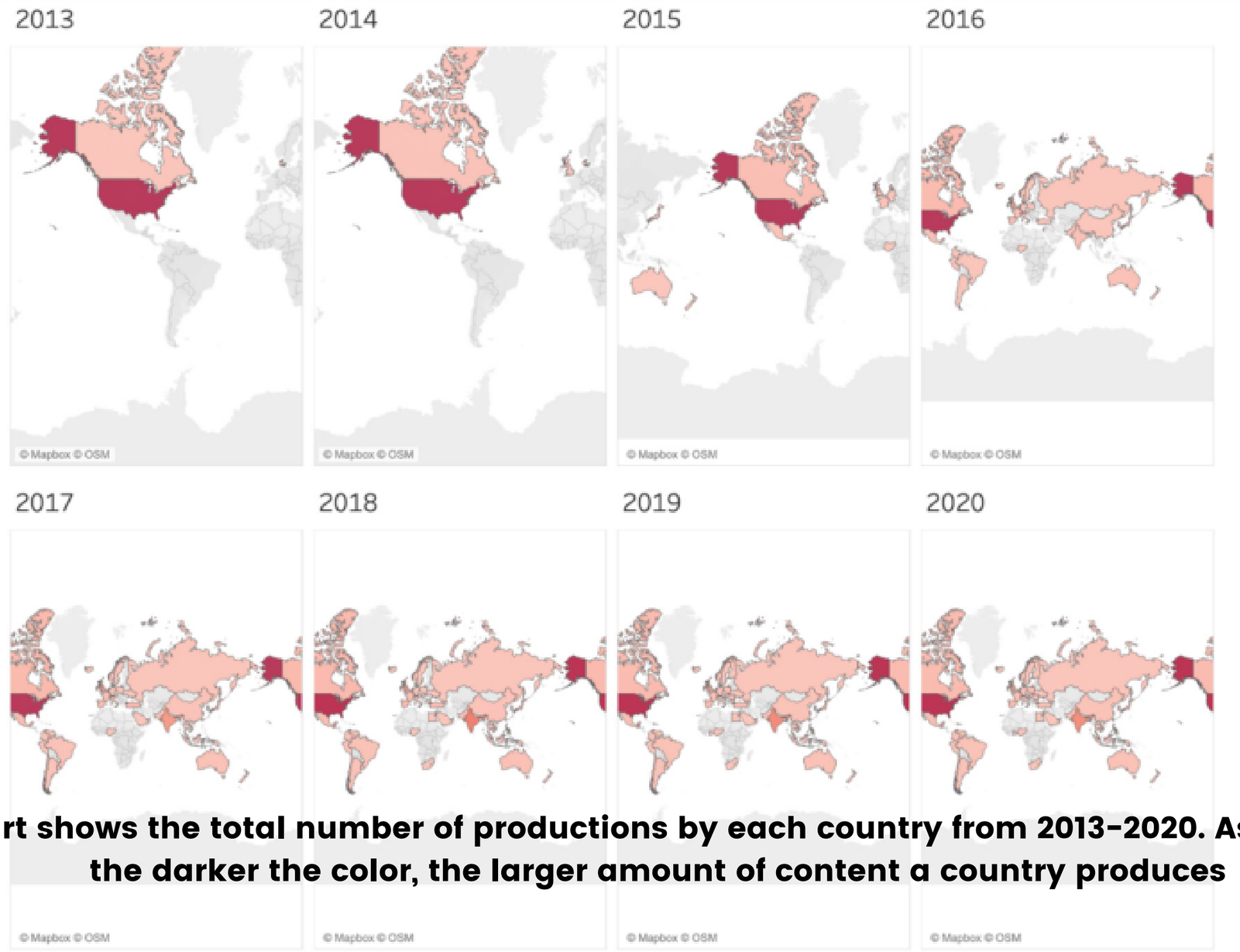
Possible COVID Influence: The pandemic's stay-at-home orders might have increased demand for serialized content like TV shows.

Long-Term Trend or Anomaly?: The gap between movies and TV shows is shrinking. The 2020 shift might be a continuation of a pre-existing trend, not just a COVID effect.

3. Understanding the Development of Netflix during Recent Years through Data Visualization

- The objective of this paper is to use the techniques of data visualization to tell the story of how Netflix developed in recent years,including its global layout, financial performance etc.

- By usingTableau to create different tables and charts, this paper is able to show the trend of Netflix’s development and invoke insights about its current operation, especially under the global pandemic.



This chart shows the total number of productions by each country from 2013-2020. As we can see, the darker the color, the larger amount of content a country produces

Fig 1. Expansion of Netflix's content globally

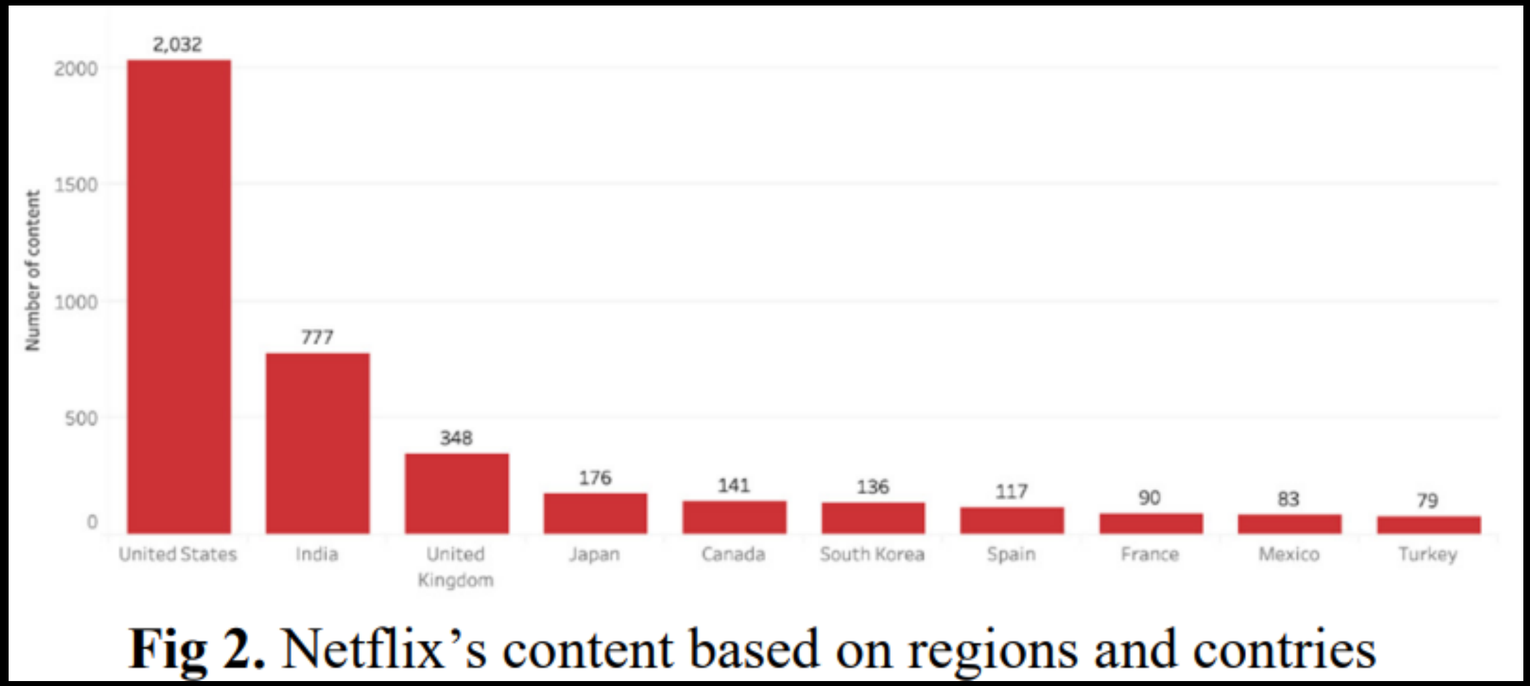
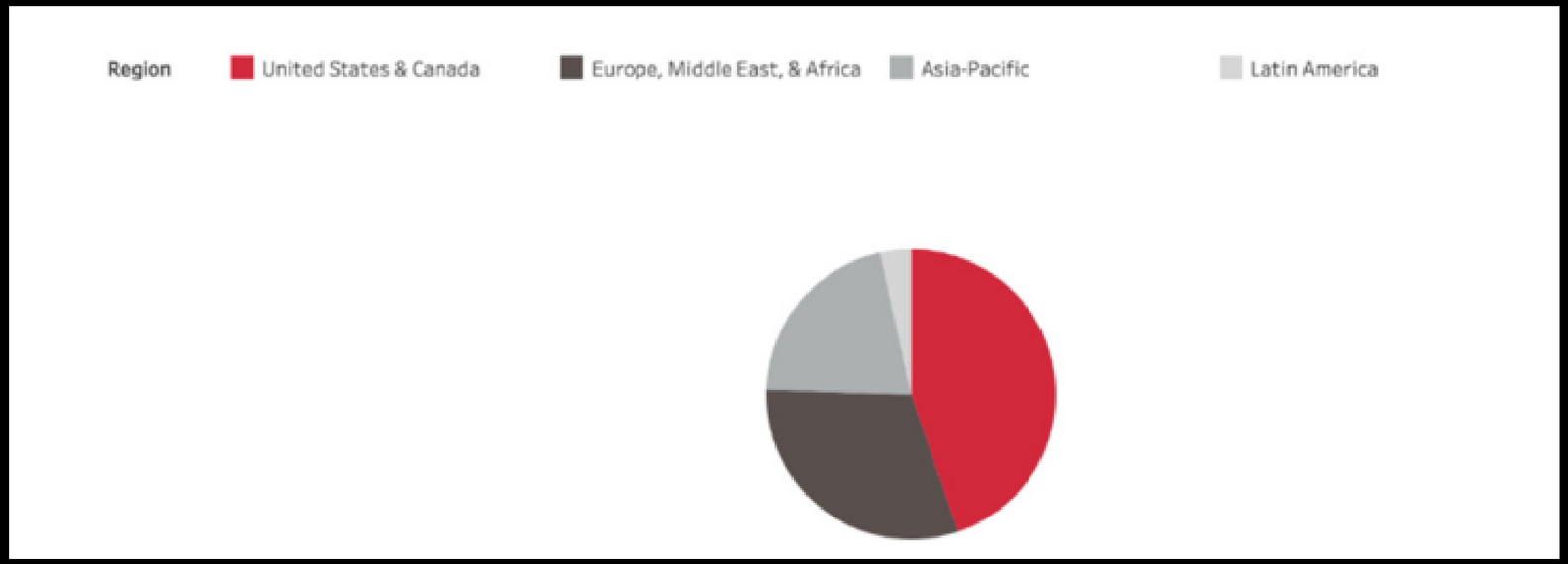


Fig 2. Netflix's content based on regions and contries

The bar chart shows the top 10 countries with the largest amount of production.



The pie chart shows the distribution of different areas' production

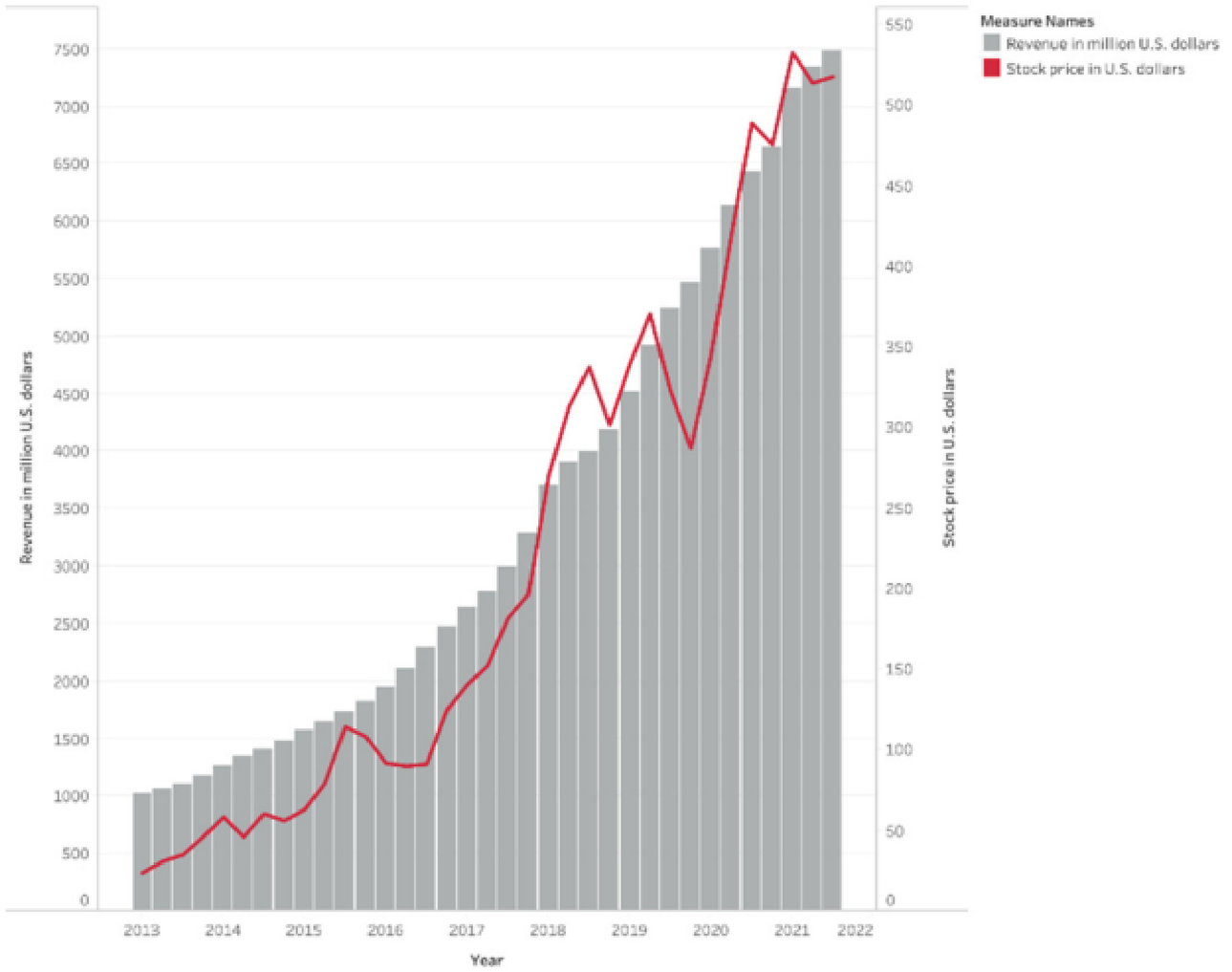


Fig 3. Netflix's financial performance

This chart shows Netflix's revenue and its stock price from 2013 to 2021 , with the stock price shown as line chart and revenue shown as bar chart

Findings

Positive Growth Trend: Data analysis across various aspects (expansion, finances) suggests a promising future for Netflix.

Resilience During Pandemic: The study indicates Netflix wasn't significantly impacted by the COVID-19 pandemic.

Content Quality and Diversity: Moving forward, maintaining high-quality original content and offering a diverse selection will be crucial for attracting a wider audience.

Importance of Regional Preferences: Understanding subscriber preferences in different regions is essential for improving recommendation systems and user experience.

Continuous Audience Analysis: Continued analysis of user habits and preferences can help Netflix retain and expand its subscriber base.

4. Content Based Recommendation System on Netflix Data

Methodology

- **Focus on Text-Based Features: Due to limitations in the dataset, the study prioritizes show descriptions, release years, genres, and ratings for recommendations. Viewer-level data like watch history is not included.**
- **Recommendation Goal: The system aims to suggest five shows similar to a list of shows a viewer has already watched.**
- **Data Preprocessing: Textual data from director, description, genre, and rating is combined into a single variable.**
- **Text is converted to lowercase and stripped of punctuation and stop words.**
- **TF-IDF Vectorizer: This technique is used to convert the preprocessed text into numerical features, considering both word frequency within a show description and overall frequency across all shows.**
- **Cosine Similarity: This metric assesses the similarity between shows based on their TF-IDF vectors. Shows with higher cosine similarity share more textual characteristics and are considered more similar recommendations.**

5. Data analysis and prediction of Netflix stock

Methodology

- Linear regression,
- Random forest regression
- LSTM

Findings

- Random forest regression has the most variation in its predictions.
- Linear regression and LSTM have lower variation in their predictions compared to random forest.
- Random forest regression achieved the highest fitting accuracy among the three models.
- Looking beyond the visual chart, the researchers compared additional performance metrics:
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
- Across all three metrics, random forest regression yielded the lowest errors, indicating it outperformed the other two models in this specific dataset.

Strategies of translating swear words into Arabic: a case study of a parallel corpus of Netflix English-Arabic movie subtitles

https://www.nature.com/articles/s41599-023-01506-3

Findings:

- Analyzed 1564 English swear words in study.
- Identified three main translation strategies: omission, softening, swear-to-non-swear.
- Omission most common strategy across genres.
- 'Shit' most frequently omitted swear word.
- Emphasized importance of cultural understanding in translation.

Strategy:

- Omission
- Softening
- Swear-to-non-swear

English swear words:

Shit, Damn, Hell, Crap, Bloody, etc

Tools used:

SketchEngine :

Linguistic corpus analysis tool aiding language research and comprehension.

Application:

- Subtitling Industry
- Cross-Cultural Communication
- Language Teaching and Learning
- Content Localization
- Research and Corpus Linguistics

Objective:

Investigates translation of English swear words to Arabic in Netflix subtitles. Emphasizes cultural sensitivity, identifies common strategies, and implications.

Results:

- Most Common Translation Strategies
- Quality of Translation
- Comparison with Satellite TV Channels
- Frequency of Translation Strategies
- Comparative Analysis

Limitations:

- Data Source
- Genre Limitation
- Sample Size
- Cultural Context
- Generalizability

Big Data Analytics in the Entertainment Industry: Audience Behavior Analysis, Content Recommendation, and Revenue Maximization

https://researchberg.com/index.php/rcba/article/view/142

Findings:

- Identifies demand regions, targeting marketing and distribution efforts.
- Analyzes viewing patterns, enhancing future promotional content.
- Measures view duration, understanding captivating aspects of content.
- Personalizes recommendations, optimizes resource allocation, maximizing revenue.

Strategy:

- Data Collection and Analysis
- Market Data Analysis
- Audience Engagement Metrics
- Content Recommendation Systems
- Geographic Analysis
- Iterative Optimization

Tools used:

- Apache Hadoop, Apache Spark
- Hootsuite, Sprout Social, Brandwatch: Monitoring social media conversations, sentiments, and trends.
- Tableau, Power BI, Google Data Studio: Creating visual representations of data insights
- Nielsen, Comscore, Kantar: Providing market data and consumer behavior insights.
- RapidMiner, KNIME, Weka: Extracting patterns and insights from large datasets.

Algorithm:

- Collaborative Filtering
- Content-based Filtering
- Hybrid Approaches
- Geographic Analysis Algorithms
- Data Mining Algorithms

Objective:

Entertainment utilizes big data for audience insights, personalized recommendations, and revenue optimization strategies, enhancing content creation and marketing efforts.

Results:

- Presents findings of data analysis in study.
- Includes insights on viewer preferences, engagement metrics.
- Examines market trends and consumer behavior.
- Discusses effectiveness of recommendation systems, revenue strategies.

Limitations:

- Considers analysis technique limitations.
- Examines sample size constraints.
- Notes findings' generalizability limitations.
- Acknowledges potential biases in study.

Netflix original series, global audiences and discourses of streaming success

https://journals.sagepub.com/doi/abs/10.1177/0196859211072446

Findings:

- Streaming success discourses remain elusive despite advanced analytics tools.
- Netflix faces challenges defining "television audience" due to its evolution.
- Industry discourses reflect shift in post-network television landscape.
- Netflix originals like Fauda, La Casa de Papel analyzed.

Strategy:

- Analyzed industry discourses on streaming success.
- Focused on Netflix's global originals.
- Examined reception in television landscape.
- Investigated implications for streaming success discourse.

Tools used:

- Secondary data analysis
- Industry trade publications
- Media and communication research sources

Social Media Platform:

Twitter was chosen as the social media platform for the study due to its real-time nature, allowing for immediate access to industry discussions and trends.

Application:

- Defines streaming success parameters.
- Identifies audience engagement challenges.
- Guides platform strategies, like Netflix.

Objective:

Explore Netflix's streaming success discourse, highlight challenges in audience comprehension, utilizing industry discussions and specific original content analysis.

Results:

- Identify challenges in defining success in streaming era.
- Highlight disconnect between industry discourses and audience behaviors.
- Explore evolving nature of popular television in global streaming.
- Reveal complexities shaping perceptions of success in streaming landscape.

Limitations:

- Relies on secondary data.
- Biases from secondary sources.
- Incomplete information from sources.
- Challenges capturing audience nuances.

The Analysis of User Intention to SubscribeNetflixUsing UTAUT Framework

https://gemapublisher.com/index.php/jiste/article/view/10/14

Findings:

- Factors influencing subscription renewal: content, performance, price, habit, ethics.
- Hedonic motivation, habit, price value influence user satisfaction strongly.
- Satisfaction, habit, morals, ethics impact continuance intention significantly.
- Significant increase in Indonesian streaming application users observed over years.

Strategy:

- Quantitative approach with hypothesis testing.
- Experimental research design used.
- Analyzed impact of independent variables.
- Focus on user behavior towards streaming subscriptions.

Tools used:

- Microsoft Word 2016 for writing reports.
- Microsoft Excel for classifying questionnaire data.
- SmartPLS for data processing.
- Draw.io for creating supporting images.
- Mendeley Desktop for managing references.

Research Methodology:

- Utilized UTAUT2 model for analysis framework.
- Employed quantitative method for data collection.
- Data gathered via questionnaires and surveys.

Application:

- Improve user understanding.
- Enhance customer satisfaction.
- Optimize renewal and expansion strategies.

Objective:

Research explores factors shaping user loyalty in Indonesian streaming applications, informing expansion decisions.

Results:

- Key findings include factors influencing subscription renewal intentions.
- Impact of variables like content, performance, and ethics examined.
- Relationships between satisfaction, habit, and continuance intention explored.
- Results elucidate complex dynamics of user subscription behavior factors.

Limitations:

- Sample size constraints noted.
- Data collection challenges addressed.
- Generalizability limitations acknowledged.

Thank you


THANK YOU
i'm SORRY