UNLEASHING THE POWER OF DATA

How Netflix Dominates Streaming with Analytics and Recommendation Systems

ABSTRACT

Collective intelligence, through the utilisation of data and analytics, is transforming the business landscape by harnessing data and improving personalisation. This approach enhances the user experience, propels innovation, uncovers valuable insights, detects potential risks, and streamlines operations.

Shuaib Suleman

BUS4901 Workplace Research and Analysis Project
Student ID: 22180539
Master of Business Administration in Accounting and Management programme
29 January 2023

Trinity College Zurich

List of Contents

List of Figures **2**

Table of Acronyms **3**

Section 1 ..... **4**

Project Title 4

Project Summary 4

Project Background 4

Section 2 ..... **4**

Project Objective 4

Section 3 ..... **5**

Project Methodology 5

Section 4 ..... **7**

History of Netflix 7

Evolution of the Business Model 8

Current Strategy Concepts 10

How does Digitalisation Impact Strategy Creation? 12

Understand Netflix's Strategy and how New Strategies are Implemented to Address Competition 13

Overview of the Recommendation System and How it Supports the Functional (IT) Strategy 15

Section 5 ..... **15**

Overview of Data Analysis 15

Exploratory Data Analysis 21

Section 6 ..... **29**

Recommender System for Movies 29

Hybrid Recommendation System - Content and Collaborative 33

Section 7 ..... **35**

How Does Netflix Employ Big Data Analytics to Enhance Customer Satisfaction and Recommend Movies? 35

Netflix Recommendation System 35

Section 8 ..... **39**

Conclusion 39

Bibliography **41**

List of Figures

Figure 1 - Relationship of data 16
Figure 2 - Code for the shape of data 17
Figure 3 - Shape of data output 17
Figure 4 - First five records on the Movie file 17
Figure 5 - Code to replace genres with the genre 18
Figure 6 - Code for movies without genre 18
Figure 7 - Output of number of movies without genre 18
Figure 8 - Code to split title and column 18
Figure 9 - Output from the splitting of title column 18
Figure 10 - Scatter and density plot for links.csv 19
Figure 11 - Correlation matrix for links.csv 19
Figure 12 - Output after dropping TMDbld 19
Figure 13 - Output of the summary statistics for the ratings data set 19
Figure 14 - Distribution graph of useId and ratings 20
Figure 15 - Scatter and density plot for ratings.csv 20
Figure 16 - Correlation matrix for ratings.csv 20
Figure 17 - Defining the function to obtain ratings of movies 21
Figure 18 – Summary output for the rating of every movie 21
Figure 19 – Final Joined Dataset 21
Figure 20 – Most famous genres of all time 22
Figure 21 – Movies released 23
Figure 22 – Movies produced and released 24
Figure 23 - Release year distribution per genre 24
Figure 24 - Median genre distribution per year 25
Figure 25 – Median rating distribution over time 26
Figure 26 – Top 5 Genres 26
Figure 27 - Top 10 most popular titles of all time 27
Figure 28 – Worst ten titles of all time 27
Figure 29 - Code for median positive and negative feedback 27
Figure 30 - Median positive and negative feedback 28
Figure 31 - Top 10 years for movie releases per genre 28
Figure 32 - Top 5 tags per genre 29
Figure 33 – Movie rating table 30
Figure 34 – Similarities between users in rating patterns and between movies 31
Figure 35 – Code for the similarity between users 31
Figure 36 – Code for films rated most relative to Jumanji 1995 31
Figure 37 – Output of similar movies like Jumanji 1995 32
Figure 38 – Code for movies similar to The Godfather 1972 32
Figure 39 – Output for movies similar to The Godfather Part II 32
Figure 40 - Schematic diagram of collaborative and content filtering 33

Table of Acronyms

AMC	American Movie Classics
AWS	Amazon Web Services
CBF	Content-Based Filtering
CDN	Content Delivery Network
CF	Collaborative Filtering
EDA	Exploratory Data Analysis
HBO	Home Box Office
HD	High Definition
HF	Hybrid Filtering
IMDbld	Internet Movie Database Build
Inc	Incorporated
IT	Information Technology
MIT	Massachusetts Institute of Technology
MPA	Motion Picture Association
OTT	Over-The-Top
RS	Recommender System
Sci-Fi	Science Fiction
SD	Standard Definition
SVOD	Subscription Video on Demand
TMDbld	The Movie Database Build
TV	Television
UHD	Ultra-High Definition
UNOGS	Unofficial Netflix Online Global Search
URL	Uniform Resource Locator
VRIN	Valuable, Rare, Inimitable, and Non-Substitutable

Section 1

Project Title

Unleashing the Power of Data: How Netflix Dominates the Streaming Game with Analytics and Recommendation Systems

Project Summary

The primary objective of the research is to determine if Netflix uses big data analytics to guide its business strategy, functional approach (Information Technology (IT)), and recommendation system, providing the firm with a competitive edge.

Project Background

Netflix collects massive quantities of data from its diverse user base and employs big data analytics to understand its client better. They then deliver superior services and products to customers by utilising this data and converting it into intelligence. It gathers information such as a user's location, the material seen by the user, the user's interests, the data sought by the user, and the time the user watched the stream *(Anon, 2022)*. Its system generates a tailored recommendation depending on the user's interests based on these characteristics *(Gomez-Uribe & Hunt, 2016*). Netflix has always focused on shifting business demands, transitioning from DVD rental to video on demand and developing original series powered by data analytics, strengthened by one of the best recommendation systems.

Maintaining existing members and expanding the number of new subscribers, increased competition from other streaming providers such as Hulu, Disney, Warner Media, and Amazon, and rising production costs for original content are Netflix's most significant challenges at present (*Lee 2022).* The organisation uses big data analytics to solve these impediments and has made substantial investments in extensive data analytics research, confidently spending over $8 billion on content (*Blog, 2021*). They have a separate organisation known as Netflix Research that focuses primarily on data analytics, including user experience, recommendations, and machine learning (*Anon, 2022*). The organisation invests extensively in data science and analytics for their recommendation algorithms. These recommendation systems comprehend the user and deliver appropriate recommendations (*Gomez-Uribe & Hunt, 2016*).

Customer retention is described as engaging customers and persuading them to use the service or purchase the product. This may appear to be a simple approach, yet it is widely regarded as the most potent tactic any company uses. Netflix has exploited it so well that its subscriber retention rate is astounding and has steadily increased over the years (*Needle, 2021*).

Section 2

Project Objective

The purpose of this project is to investigate the business strategies of Netflix and determine the extent to which big data analytics is used to improve the customer experience. We will also examine whether Netflix's IT strategy creation and recommendation system is employed as a tool to drive transformation and build competitive advantage. Using data examples, our analysis will explore data analysis and fundamental collaborative filtering principles, including user- and item-based methods. We will then draw conclusions based on the established objectives.

We aim to demonstrate that analytics is a crucial factor in the digital age and that developing advanced algorithms and machine learning can impact an organisation's ability to use advanced decision-making tools and gain a competitive advantage in the market. By analysing the data, we hope to gain insight into Netflix's current strategy implementation and understand its performance in achieving strategic objectives.

Additionally, we intend to explore whether effective data management is a strategic axis for the Subscription Video on Demand (SVOD) industry. Using business analytics and big data, companies like Netflix can gain insights, make more accurate predictions, identify risks, and operate efficiently.

The study entails:

- History of Netflix
- Evolution of the business model
- Current strategy and concepts
- Overview of business and data analytics
- Overview of Netflix recommendation system
- Exploratory data analysis on the MovieLens dataset
- Recommendation system analysis
- Explore whether Netflix employs big data analytics to enhance customer experience and satisfaction
- Conclusion

Section 3

Project Methodology

The research is based on a variety of sources available on the internet. We will be working with primary and secondary sources related to the firm's data management: bibliographic reviews, analysis of data published by the firm itself (Netflix Tech Blog and the corporate area of Netflix.com), the information provided by its staff in various discussions forums, and analysis of data provided by the specialised press and case studies. The key references are the websites of Netflix Inc., big data analytics recommendation system blogs, and various conference/journal articles about data analytics and recommendation systems.

Google Scholar is primarily utilised as a search engine to locate relevant literature for case studies. The key argument for using Google Scholar is that it is free and includes many academic publications. It has excellent features, such as citation tracking and individual citation export.

Data is obtained from kaggle.com titled ml-latest-small, an official Netflix data set. This dataset describes MovieLens' 5-star rating and free-text labelling activities. It includes 100836 ratings and 3683 tag applications for 9742 films. Six hundred ten users generated these data between March 29, 1996, and September 24, 2018. The date of creation of this dataset is September 26, 2018 (*Harper & Konstan, 2016).*

Users were randomly selected for inclusion. All selected users had rated a minimum of twenty films. There is no demographic information provided. An id represents each user, and no other information is provided. The files links.csv, movies.csv, ratings.csv, and tags.csv contain the data. The collection consists of movie ID, customer ID, ratings out of five, and rating timestamps. Each movie ID and customer ID represent a unique movie and consumer, respectively.

First, we import the data, clean the missing data, and make the data ready for analysis and model writing. These operations are made using the Python language 3.8.2. version and in Visual Studio code as IDE as an environment. We use Pandas, NumPy, Math, and Re libraries for data cleaning and missing data operations. Since every user does not rate every movie, some attributes have a NaN value. Also, if any data is corrupted, it is cleaned out. To visualise our dataset, we use Tableau, Matplotlib and Seaborn libraries.

As part of our exploratory analysis, we would seek to understand the following:

- The most popular genre of all time
- When were the movies released?
- The rate at which movies have been produced over the period
- Explore the release year distribution per genre
- Release year distribution per rating
- Median genre distribution per year
- Median rating distribution per year
- Top ten most popular titles of all time (based on users' ratings)
- Worst ten titles of all time (based on users' ratings)
- Top five genre categories of all time
- A comparison of the positive and negative feedback across genres through the years
- Top ten years for movie releases per genre
- Top five tags per genre

We intend to use simple examples to illustrate the key concepts in collaborative, content, and hybrid filtering. We will discuss user-based and item-based collaborative filtering algorithms. We will look at movie genres for content-based recommendations and search for similar ones. We will use the cosine similarity metric to find the similarity between the two movies.

Then seek recommendations from the algorithm on the following:

- Recommend movies like "Jumanji (1995)"
- Get movie recommendations for “The God Father (1972)“

In the study, we aim to compare our data analytics and filtering methods on the MovieLens dataset and use the outcomes to determine an appropriate strategy for Netflix. The two strategies will then be compared. Suppose the observed strategy of Netflix is like or consistent with what we think is appropriate. In that case, we will argue that Netflix uses data analytics outcomes to inform its strategy. If the two strategies are significantly different, then we will discuss the possible reasons for the divergence and consider what factors other than data analytics outcomes drive Netflix's strategy.

Sections

This paper has a total of eight sections. Section 1 provides a brief introduction to Netflix. Section 2 explains the objectives of this case study. Section 3 describes the methodology used in this case study. Section 4 describes the history and evolution of the Netflix business model, tying them to strategy concepts. Section 5 explores the sophisticated use of data analytics supported by digitalisation, whilst Section 6 considers the recommendation systems employed. Section 7 provides an overview of how Netflix employs big data analytics to enhance the customer experience, and finally, this paper concludes with Section 8.

Section 4

History of Netflix

Netflix was formed in Scotts Valley, California, on August 29, 1997, by Marc Randolph and Reed Hastings. Hastings and Randolph thought up Netflix while riding in a carpool between their homes in Santa Cruz, California, and Pure Atria's headquarters in Sunnyvale. The company's headquarters are in Los Gatos, California. Netflix's core business consists of online subscription-based streaming services for TV Shows, Original Productions and Movies. Netflix initially offered a per-rental basis for each DVD, but in September 1999, it introduced a monthly subscription option. Early in 2000, the per-rental model was discontinued, allowing the company to focus on a flat-fee unlimited rental model with no due dates, unpaid fines, shipping and handling fees, or per-title rental fees *(Fancey, 2022)*.

Human nature has always been rooted in storytelling. Technical advances that have significantly altered society have made it possible to tell more prosperous and exciting stories. Netflix is situated at the crossroads of digitisation and storytelling.

Netflix announced the Netflix Prize on October 1, 2006, offering $1,000,000 to the first creator of a video-recommendation algorithm that outperforms its current program, Cinematch, by more than 10% in predicting consumer ratings. BellKor's Pragmatic Chaos was awarded the $1,000,000 prize on September 21, 2009. Cinematch, introduced in 2000, is a recommendation system that recommends movies to its customers *(Fancey, 2022)*.

In its early days, Netflix incurred enormous losses, but as internet users grew, the company shifted its business model from DVD rental and sales to online video streaming in 2007. Netflix realised they could minimise the losses experienced from rentals. To achieve this, Netflix had to alter their business model. In addition to streaming movies and television series from other studios, Netflix started producing its own films and television series.

Global growth commenced in 2010, beginning in Canada and expanding to Latin America in 2011, followed by the United Kingdom and other European countries such as Ireland, Denmark, the Netherlands, and Norway from 2012 to 2015. In 2012, Netflix separated its DVD rental service from its online streaming service into a separate subsidiary.

In January 2016, Netflix announced it would begin VPN blocking since it can be used to watch videos from a country where they are unavailable. Orange Is the New Black was renewed for a fifth, sixth and seventh season in February 2016 *(Fancey, 2022)*. It became available worldwide except in China, Syria, North Korea, Kosovo, and Crimea. The result of the VPN block was that people could only watch videos available worldwide. Other videos are hidden from search results, which can be found on the Unofficial Netflix Online Global Search (*uNoGS*) website.

Netflix earned the most Academy Award nominations of any studio, with thirty-six. Sony Pictures Entertainment announced an agreement for Netflix to hold the U.S. pay television window rights to its releases beginning in 2022. Netflix also won more Emmys than any other network or studio, with forty-four wins, tying the record for most Emmys won in a year set by CBS in 1974. As of August 2021, Netflix Originals made up 40% of Netflix's overall library in the United States (*Fancey, 2022*).

Netflix acquired Next Games in March 2022 to expand into gaming. In July 2022, Netflix partnered with Microsoft to launch its advertising-supported subscription plan, and its gaming platform was reported to have an average of 1.7 million users a day, less than 1% of Netflix's subscribers at the time. On July 22, 2022, it was announced that the service had lost about one million customers, bringing the total to 220.7 million. Meanwhile, The Walt Disney Company stated on August 10, 2022, that its Streaming business has a total of 221 million members from Disney+, Hulu, and ESPN+, making Netflix the second largest streaming service provider after Disney at that time *(Richwine & Chmielewski, 2022).*

By September 2022, Netflix had 221.7 million subscribers globally, including 73.3 million in the United States and Canada; 73.0 million in Europe, the Middle East, and Africa; 39.6 million in Latin America; and 34.8 million in the Asia-Pacific region, making them the most prominent media service provider. Netflix is a Motion Picture Association (MPA) member and has significantly distributed independent films (*Company profile* 2022) (*Netflix - study in China,(2022*). It is available globally except for Russia, China, Syria, and North Korea.

Evolution of the Business Model

The overall business model of Netflix Inc. combines several integrated business models. This hybrid organisational structure results from the company's operations, which include streaming entertainment material on demand and producing original content, including movies and television programs. Following the corporation's general strategy for competitive advantage, these business models utilise the Valuable, Rare, Inimitable, and Non-Substitutable (VRIN) analytical methodology to evaluate Netflix's value chain and the accompanying competitive advantages *(Rivera, 2019).* The company is an excellent illustration of how online business modelling enables large-scale, cost-effective, high-efficiency operations.

The business model of Netflix Inc. is consistent with the company's generic competitive advantage strategy (Porter's model) and intense growth tactics (Ansoff Matrix). This alignment contributes to the company's strategic position as a top competitor in the on-demand digital video streaming sector.

Netflix's business model began as a brick-and-mortar DVD store, but it is now the world's leading and most well-known over-the-top (OTT) platform. Its extensive collection is a result of its acquisition of material from distributors and studios. This has been accomplished through direct purchases, revenue-sharing arrangements, and licensing agreements. It pays off, as 80% of Netflix's revenue is generated by licensed material.

Despite increased competition, Netflix Inc.'s fundamental skills sustain corporate success, according to studies by Rivera *(Rivera, 2019).* Within the Valuable, Rare, Inimitable, and Non-Substitutable (VRIN) framework, these competencies represent the resources and skills of the firm, which contribute to its long-term competitive advantage. Netflix's value chain utilises these core competencies to provide customers or subscribers with efficient services.

A firm's VRIN analysis gives a resource-based perspective of the on-demand digital media industry. It identifies the essential tools and capabilities for sustainable competitive advantages *(Rivera, 2019)*. As Netflix's plans evolve, so do the variables contributing to the strategic planning process relevant to core competencies.

Netflix Originals are those programs with the Netflix logo in the upper left corner. According to industry analysts, expenditure on original content is anticipated to expand by up to 50 per cent of the budget. Netflix is now the industry leader in subscription-based content. It utilises an SVOD (Subscription Video on Demand) concept. Subscribers pay a monthly fee to access a vast media collection anywhere. Therefore, subscriptions are Netflix's primary revenue source (*Complete List of Netflix Originals*, 2022).

Netflix Inc. functions as a business model for its video streaming activities. On the same platform, users may access their chosen forms of entertainment. Netflix Inc. produces new films and television episodes utilising the usual pipeline method. The pipeline business model provides the client with direct material output control. It supports the rapid expansion of Netflix *(Pereira, 2020).*

The organisation circumvents intermediaries by distributing its original content to users directly via its streaming service. The corporation employs this business strategy to utilise its strategic advantages and competencies *(Moore, 2019).* Other suppliers of entertainment material can deal directly with Netflix to access global audiences, therefore helping the company's rapid expansion.

Netflix's marketing technique for unlimited streaming is emblematic of the company's broader business strategy. Customers with an unlimited membership get unrestricted access to entertainment material on the website. This business strategy helps Netflix attract and keep users and boosts the success of its expansion strategies.

The organisation has also formed partnerships with PlayStation, X-Box, and several other companies in the gaming industry. It has included collaborations with the gaming industry to supply all its gaming subscribers with video games that meet their specific requirements. Netflix partnered with Microsoft, Apple, and Android while transitioning from a mail-in system to a streaming platform. After joining the web-based network, it has effectively formed relationships with significant data suppliers like Amazon and Google.

Multiple competitors threaten Netflix's market share, so maintaining existing members and expanding the number of new subscribers, increased competition from other streaming providers such as Hulu, Disney, Warner Media, and Amazon, and rising production costs for original content are Netflix's most significant challenges at present. Netflix uses big data analytics to solve these challenges. Netflix has made substantial investments in big data analytics research, spending over $1 billion (*Sabga-Aboud, 2022)*. They have a distinct organisation known as Netflix Research that focuses primarily on data analytics, including user experience, recommendations, and machine learning. They spend extensively on data science and analytics for their recommendation algorithms. These recommendation systems then comprehend the user, improving and delivering enhanced recommendations (*Gomez-Uribe & Hunt, 2016*).

Current Strategy Concepts

Netflix's Generic Competitive Strategy *(Moore, 2019)*

Netflix's generic strategy is a combination of cost leadership and differentiation. While Netflix's business model is not wholly based on cost leadership, it has exploited competitive pricing to expand its global client base.

The general strategy of Netflix supports the viability of its business model by leveraging appropriate competitive advantages. The corporation's business model and competitive position offset external influences from its primary competitors, Walmart, Amazon, Google, Apple, HBO, and Disney, among others. While addressing these competitive factors, Netflix's aggressive expansion methods advance business development. This approach helps to offset external pressures from significant competitors.

The operational efficacy and benefits of the corporation's competitive advantages are ensured by aligning these growth plans with the general strategy and business model. Netflix also has a broad differentiation approach while constructing cost advantages where needed. The company has a competitive advantage through differentiation since it offers movies and television series that other streaming providers do not.

Cost Leadership

Michael E. Porter's approach provides a competitive advantage through reduced costs and, in some cases, reduced selling prices, supporting the general competitive strategy. Netflix Inc.'s generic cost leadership strategy is secondary, and the pricing approach is based on differentiation rather than cost leadership. The company has set competitive prices to guarantee that clients from all segments may use its services. It has introduced three options that provide users access to Netflix content on various devices. In addition to two more expensive plans, the business has introduced a more affordable basic program. This plan permits users to access standard definition (SD) content on a single screen. Users who are prepared to pay more can access content in high definition (HD) and ultra-high definition (UHD) on two to four devices. Therefore, Netflix provides both premium and standard services. Competitive pricing has accelerated the company's customer base expansion; however, cost leadership is a secondary tactic the firm uses.

This specific, conventional strategy matches Netflix's ambitious expansion plans, which stress market penetration. The strategy builds on the business model and the company's value. In addition to being the "retailer" for streaming content, Netflix is now one of the suppliers, giving it even more power.

Netflix's continuing global expansion has been enabled by its data-mining technique, which has helped the company collect data about the kind of content its customers want, the marketing methods they respond to, and how it can better personalise users' libraries *(Brennan, 2021).* Netflix provides additional languages, subtitles, and dubbing to enhance the global experiences of its viewers. These other features distinguish Netflix from its competitors and enable it to attract a larger audience.

Differentiation

This is the primary strategy Netflix employed to accelerate its global expansion. The organisation has varied its offerings based on content quality and genre. In addition to enhancing the entire customer experience, the company has prioritised the creation of a vast amount of unique content unavailable on other platforms. In this regard, the organisation has teamed up with a few well-known suppliers. The organisation invests in technical innovation to provide excellent customer service to its users. It has built an algorithm called Cinematch that makes recommendations to its customers based on their preferences and past activity. This algorithm is recognised for its effectiveness in providing streaming content ideas for the audience.

The differentiation strategy aids in attracting and retaining customers for the company. The firm has also made substantial technological investments, like its recommendation engine and user interface, to improve the viewing experience. As a result, Netflix has become a popular streaming service with many devoted users.

In addition to its content and technology, Netflix has differentiated itself through its business strategy, providing consumers unrestricted access to its material via a subscription-based service. This approach has contributed to the company's financial success by assisting it in building an extensive and loyal consumer base.

Overall, Netflix's approach to distinctiveness has helped the firm stand out in a crowded and competitive industry. Differentiation is founded on developing and selling "distinct items for diverse customer groups" *(Pearce & Robinson, 2015).* Netflix has distinguished itself by positioning itself as an ad-free, on-demand content supplier instead of a live-streaming service.

The organisation has expanded its original content and prioritised the production of high-quality films and television episodes. The company's offering of original content has been a tremendous success. In addition to its original programming, Netflix provides a vast range to fulfil a variety of tastes.

The Extensive Growth Strategies of Netflix (*Moore, 2019)*

Market Penetration and Development

The core of Netflix Inc.'s ambitious expansion strategy is to provide more subscription content to more users on the platform. This growth plan's objective to increase sales and market share is contingent on promoting the current online subscription service and unique content to new audiences, thereby facilitating commercial growth. Its consistent approach to cost leadership contributes to the success of this expansion plan.

Netflix's operational control system is a dashboard-based method in which all the company's user information is centralised and accessible to all employees. This strategy has enabled Netflix to provide individualised content recommendations for each user while collecting data on customer preferences. "Netflix categorises and tags each episode and film on its site with great care. It has created over 75,000 distinct microgenres to organise its content library *(Opheliac, 2015).* This type of information has distinguished Netflix from its rivals *(Opheliac, 2015).* This method has been in existence since the birth of Netflix. According to a Harvard study, the final part of Netflix's early strategy that contributed to its success was its emphasis on technology, specifically website and movie personalisation *(Gloria, 2018), (Netflix – behind the scenes, (2015).*

By constructing a product requiring users to place orders via the website, Netflix conditioned its customers to consider the website an integral component of their Netflix account (*Gloria, 2018).* The use of dashboards by Netflix has significantly increased consumer satisfaction.

Product Improvement and Diversification

Product development is developing new products to increase revenue. This is one of the main strategies used by Netflix, enabling the company to acquire enormous growth worldwide. Netflix primarily focused on generating content for higher popularity and superior user engagement. The brand has a vast collection of original movies and shows. This has enabled Netflix to differentiate its services from others and build a more extensive customer base. With time, the number of users that use online sources for their dose of entertainment has grown, resulting in faster growth of the user base of companies like Netflix. However, apart from the brand's content, Netflix has also focused on using technology for higher user convenience and offering a better user experience. The organisation efficiently leverages its generic competitive advantage to generate new content for existing subscribers through diversification.

Netflix's success also hinges on how its company culture fosters product innovation-related activities. Netflix develops fresh content for existing users efficiently. Netflix's content delivery network (CDN) has been gradually enhanced to serve its consumers better and offer the highest quality streaming worldwide. AWS (Amazon Web Services) is the most cost-effective solution for Netflix to support their massive growth over the past few years compared to its servers. AWS enables Netflix to store its entire catalogue in the cloud. Netflix uses Amazon Web Services (AWS) as its primary server. Still, it maintains numerous worldwide servers so members can receive the same streaming quality internationally *(AWS Innovator: Netflix | Case Studies, Videos and Customer Stories)*. Netflix's machine learning structure enhances video and audio compression and adaptive bitrate selection for a superior streaming service for millions of subscribers using more than five hundred million devices *(Cool, 2019).* This intense expansion plan aims to expand the company's activities beyond its present internet streaming industry.

The company invests extensively in research and development to build user-friendly streaming technology. This has aided in establishing a solid brand image and consumer loyalty, enhancing the company's competitive position. In addition, Netflix's emphasis on data-driven decision-making and customisation algorithms contributes to delivering a highly relevant and tailored viewing experience to its members. This differentiation strategy has enabled the company to stand out in a highly competitive market and maintain its market-leading position in the streaming industry.

Data Analytics

Netflix uses sophisticated data analytics to:

- offer consumers personalised movie and television program suggestions
- before approving unique material, predict its popularity (or not)
- customise marketing materials like videos and thumbnail photos
- optimise production planning and improve overall technical and business decision making

To remain at the forefront of the industry, Netflix utilises data analytics and business intelligence to implement creative strategies. Despite facing increased competition and potential political interference, the company continues to expand its markets and strengthen its brand recognition and reputation to provide exceptional streaming at accessible prices to a global audience.

How does Digitalisation Impact Strategy Creation?

Digital transformation is a process that attempts to improve an entity by inducing significant changes in its attributes via the use of information, computing, communication, and networking technologies *(Vial, 2019)*. It swiftly transforms the landscape of enterprises, industries, and civilisations. At the corporate level, it modifies a company's strategy, organisational structure and procedures, and culture to position it better to withstand the difficulties posed by new market entrants. At the industrial level, digital transformation distinguishes innovative businesses from the competition. At the social level, it affects how individuals interact with and conduct their lives.

A successful business acknowledges that the market does not stay constant, evolves with the market, reaches new customers, and improves its service and products. Netflix is an excellent example of company evolution. Over the years, technology has dramatically changed, and Netflix has continuously evolved and adapted to the market. The most momentous change for Netflix was when they began streaming popular television series. Netflix developed a phenomenally successful business strategy for its renting service. Still, throughout the years, the company has anticipated changes in customer behaviour and adapted its business model appropriately and timeously.

Again, at the height of social media and content marketing, Netflix evolved and began creating its own films and television programs. Today, Netflix is the digital disruption, a level reached by just a few organisations.

Personal advertising now controls our daily lives. More businesses are focusing on content marketing every day. It is the future of marketing, but it is like skating on extremely thin ice; it demands competence, and you may damage your company's worth in the process if you are not careful and precise.

With ambitions to expand further, Netflix has an elevated level of operational and information excellence in website operations, back-end big data analytics, digital transcoding, and streaming operations. As home streaming necessitates an endpoint device, a high-quality network connection, cloud-based content delivery, and expertise in content development, acquisition, management, and distribution, it provides solution leadership in collaboration with partners. Viewers can, for example, watch Netflix streaming on smart televisions with built-in apps or set-top boxes, smartphones, game consoles, tablets, and other specialised devices such as Apple TV units, Microsoft Xboxes, TiVo digital video recorders, or Roku TVs, streaming sticks, or players—cigarette-pack sized devices that connect to Wi-Fi and a TV *(Weinman, 2015).*

This organisation is a role model for fast innovation, employing various mechanisms to enhance its integral components, such as its movie recommendation algorithms and cloud-based entertainment service delivery operations, including cloud-based experimentation and open challenges *(Weinman, 2015).* Finally, Netflix is a collective intelligence model, combining data points from millions of consumers' viewing patterns and tastes to give individualised suggestions to each user at that moment.

Understand Netflix's Strategy and how New Strategies are Implemented to Address Competition

Netflix collects a range of user preference data, including which movies were requested (i.e., added to the queue), searched for but not located, offered but not bought, and which genres are of what level of interest. It may run various A/B testing for merchandising and its effect on consumer behaviour, such as highlighting new movies or old classics first or determining if different icon sizes encourage more viewing.

Behavioural data tends to be more important than user ratings. The rationale is that what users say, claim, or believe is not necessarily true. A user might rate a foreign film highly but prefers watching Dumb and Dumber. Many people claim to enjoy foreign films and documentaries, but actual viewing behaviour proves otherwise. Moreover, viewing behaviour tends to change due to many contextual factors: the kind of movie watched on the family TV during prime time at home on a weeknight might not be the same as the type of movie watched on a mobile device on a Saturday night while on vacation *(Weinman, 2015).*

A significant amount of data comes from freelance Netflix taggers. Through their ratings, reviews, and watching habits, Netflix subscribers provide even more data. But, of course, algorithms built by Netflix engineers and data scientists evaluate all this data to provide recommendations and a user ordering and queuing experience tailored to the individual viewer *(America 2021).*

Data powers the recommendation/personalisation engine with billions of ratings, movie popularity, millions of queue additions every day, millions of search terms every day, user interactions with the menu of movie options, preferences and activities of friends, and user information such as age, gender, location, and language, and data generated outside of Netflix, such as box office receipts and reviews. According to Kevin Slavin of the Massachusetts Institute of Technology (MIT), excessive reliance on algorithms might lead to unanticipated outcomes, especially in the absence of oversight *(Slavin, 2011).* To mitigate this, Netflix gives insight into the consequences of these computations using phrases such as "Because you watched...". This increases the system's credibility, hence boosting client retention.

Despite the sophistication of these algorithms, there are always chances to enhance them by incorporating additional data. In the end, a film is not a single entity but rather one of the most complex human artefacts, designed to elicit a range of emotions such as happiness, anger, sadness, romance, and laughter through the interplay of elements such as a plot, scenes, characters and their arcs, cinematography, and soundtrack *(Weinman, 2015).*

The soundtrack contains loudness, genre, intensity, and danceability. Netflix may utilise this information to make even better recommendations. According to Netflix's research, better recommendations only sometimes result in an upsell because pricing levels include unlimited streaming. On the other hand, better recommendations contribute to higher customer satisfaction, leading to enhanced customer retention and advocacy, higher customer lifetime value, and lower acquisition costs, such as advertising, sign-up logistics, and free trials (*Weinman, 2015).* Because suggestions account for 75% of what users watch on Netflix, optimising the recommendation engine is a major strategic priority. Netflix must have a thousand engineers working on it *(Weinman, 2015).*

The collective intelligence technique employed by Netflix directly impacts the bottom line. According to Netflix, they can directly correlate an improvement in member retention with the personalisation that drives the dashboard selection of rows, the choice of titles within rows, and the ranking of rows and titles inside each row *(Needle, 2021).*

Most important to its success is the high degree of customer satisfaction and retention achieved by its innovative recommendation engine, which achieves a delicate balance of popularity, user behaviour, diversity, originality, personalisation, context, surprise, and delight. Netflix has generated blockbuster hits by employing digital technology to promote collective intelligence.

Overview of the Recommendation System and How it Supports the Functional (IT) Strategy

Netflix was among the earliest adopters of big data analytics. In 2006, Netflix issued a $1 million challenge to anyone who could enhance their existing recommendation system, Cinematch, by 10% *(Bell & Koren, 2007).* The aim was to design an algorithm to anticipate the movie preferences of subscribers based on historical data. Netflix offered a data set containing around one hundred million ratings provided by four hundred and eighty thousand users for 17,000 films. Ratings included the user, movie name, rating date, and rating provided by the user. BellKor's Pragmatic Chaos team comprised a mathematician, data scientist, and engineers from several countries, businesses, and research institutes, including AT&T, Yahoo, and Commendo Research & Consulting GmbH and were granted the prize in 2008 after years of competition*(Team, 2021)*. Numerous teams that competed in this competition have supplied similar solutions to other participants, including e-commerce enterprises. However, the second competition that Netflix launched in 2010 had to be cancelled due to a privacy concern complaint filed against Netflix about the dataset utilised in the contest.

In their recommendation systems, Netflix uses data science and big data analytics. Late in the 1990s, the word recommender system appeared for the first time in information system literature *(Resnick & Varian, 1997).* Recommendation systems continue to be of interest since they are universally applicable to solving the practical challenges of several businesses. Several corporations, such as Amazon and Microsoft, have a commercial recommendation system *(Linden et al., 2003), (Shani & Gunawardana, 2010).*

Section 5

Overview of Data Analysis

Data analysis is analysing, cleaning, manipulating, and modelling data to identify usable information, inform conclusions, and facilitate decision-making. Data analysis encompasses several dimensions and methodologies, including distinct techniques under several titles, and is utilised in various corporate, scientific, and social science disciplines. In the modern corporate environment, data analysis plays a role in making choices more scientific and assisting organisations in operating more efficiently*(Free Encyclopedia, 2023)*.

Our analysis is to determine how we could employ customer retention by improving the customer experience, appealing to them to use the service by watching more content on the platform and keeping them engaged for longer. According to Netflix's analysis, over 75% of customers' viewing habits are influenced by personalised recommendations *(Weinman, 2015).* Several data points are gathered, and a comprehensive profile of each subscriber is developed. The subscriber profile built by Netflix is more extensive than the information or preferences subscribers provide when they first begin using the platform *(Niwate, 2021).* Netflix's capacity to collect and use data has proven successful. It leads to higher client retention year after year. According to the survey, the customer retention rate on Netflix is growing since 75% of customers follow the recommender systems' advice and watch the recommended movie. *(Niwate, 2021)*

Every day, the amount of data acquired by enterprises grows dramatically. Data analytics is the process of gathering, organising, and analysing massive amounts of data that are relevant to the company. Big data analytics refers to analysing and processing this enormous volume of data. Rapid technological advancements have led to the emergence of big data; we are continually confronted with vast quantities of unprocessed data based on an entity's specified parameters and criteria. The most common phase after data collecting is data analysis.

Consequently, data analysis begins with retrieving data from several external-cum-internal sources, followed by an intrinsic critique of the data to identify and get valuable insights that meet the study objectives.

Based on the nature and characteristics of the data, there are two primary data analysis techniques: qualitative and quantitative.

These data analysis approaches can be used individually or in conjunction with other strategies to gain access to some of the best business and intelligence-related insights for making better decisions based on the existing data. We will utilise the obtained data, transform them into insights, results, or visualisations, and recommend movies based on the tastes and interests of users within our dataset.

Data Cleaning and Preparation

Data cleaning and preparation is an important stage in the data analysis process. It is the process of ensuring that data is accurate, consistent, and in a readily analysed format. This stage is critical to the success of any data-driven endeavour and can be the difference between a successful and unsuccessful study.

Data Collection

Data collection and information can be obtained from a variety of sources. Data from diverse sources should be integrated into a single data set. This allows for a more comprehensive data analysis and can lead to better insights and more accurate findings.

Processing and Cleaning of Data

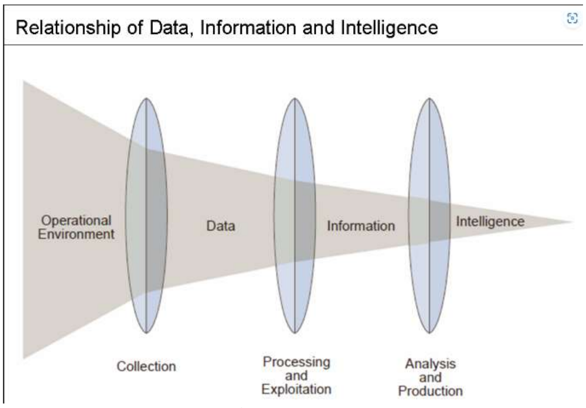


Figure 1 - Relationship of data

When data is intended for use, it must be processed to the extent that analytical requirements are met. This entails arranging the data in human-comprehensible rows and columns. In addition, before the deployment of the data for analysis, it must be cleansed to eliminate any redundant data and minimise the existence of abnormalities.

The difficulty is in transforming these data into relevant information and then utilising that knowledge to obtain intelligence. This involves the application of advanced technologies like artificial intelligence, machine learning, and data analytics. These technologies enable enterprises to handle and analyse vast volumes of data in real-time and derive meaningful insights from them.

Data cleaning and preparation is an important stage in the data analysis process. It is the process of ensuring that data is accurate, consistent, and in a readily analysed format. It also ensures that the data is dependable and can be trusted to produce accurate predictions or discoveries by satisfying data quality standards, consistency, completeness, validation, transformation, integration, security, governance, and documentation.

Import libraries and load the dataset

We import the CSV files(links, movies, ratings, tags), and we Initialize parallel processing for a better performance utilising pandarallel.initialize(). We look to get a high-level sense of the data sets in terms of make-up with the following code.

Input:

```
#checking the dimensions of the dataframe
links.shape, movies.shape, ratings. shape, tags.shape
```

Figure 2 - Code for the shape of data

And output**:**

```
((9742, 3), (9742, 3), (100836, 4), (3683, 4))
```

Figure 3 - Shape of data output

We note that the links and movies dataset has 9742 rows and three columns, the rating dataset has 100 836 rows and four columns, and tags have 3683 rows and four columns.

We printed the first five records of the movie file.

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

Figure 4 - First five records on the Movie file

We take note of the column headings and that there are 9742 records. Add a new column called 'genre' by separating the values in the 'genres' column with the '|' character and selecting the first element from the resultant list. Replace all "(no genres specified)" values in the 'genre' column with "None". The 'genres' column has been removed from the data frame. Subsequently, we replace genres with genre (use the first value from genres) with the following code.

```
# Create a new column 'genre' and set it to the first element of the split 'genres' column
movies_df['genre'] = movies_df['genres'].str.split('|').str[0]

# Replace '(no genres listed)' with 'None' and drop the 'genres' column
movies_df.replace({'genre':'(no genres listed)'}, {'genre':None}, inplace=True)
movies_df.drop(columns='genres', axis=1, inplace=True)
```

Figure 5 - Code to replace genres with the genre

To print movies without genre, we input the following:

```
#Print movies without genre
print("Number of movies without genre:", movies_df['genre'].isnull().sum())
```

Figure 6 - Code for movies without genre

Output:

```
Number of movies without genre: 34
```

Figure 7 - Output of number of movies without genre

We observe thirty-four movies without values in the genre column and place zeros in each column to make the set readable from a quantitative perspective. We then add a new column to the data being the release year and obtain the below-cleaned data as shown, reflecting the output of the data frame reflecting the first and last five records.

We check for movies without the release year and note there are 1094 movies without a release year. We, unfortunately, cannot allocate release years to these rows within our dataset as it will distort the information significantly, so we leave it as is. We extract the movie's release year from the title column of the data frame and store it in a new column, 'release year'.

```
#extract the release year of the movie from the title column of the data
# #frame and store it in a new column 'release year'
def extract_year(title):
    match = re.search(r'\((\d{4})\)', title)
    if match:
        return int(match.group(1))
    return None

movies_df['release_year'] = movies_df['title'].apply(extract_year)
```

Figure 8 - Code to split title and column

	movieId	title	genre	release_year
0	1	Toy Story	Adventure	1995.0
1	2	Jumanji	Adventure	1995.0
2	3	Grumpier Old Men	Comedy	1995.0
3	4	Waiting to Exhale	Comedy	1995.0
4	5	Father of the Bride Part II	Comedy	1995.0
...	...	...	...	...
9737	193581	Black Butler: Book of the Atlantic	Action	2017.0
9738	193583	No Game No Life: Zero	Animation	2017.0
9739	193585	Flint	Drama	2017.0
9740	193587	Bungo Stray Dogs: Dead Apple	Action	2018.0
9741	193609	Andrew Dice Clay: Dice Rules	Comedy	1991.0

Figure 9 - Output from the splitting of title column

Links.CSV

We then look at the links dataset. The below matrix and plots reflect the inverse correlation between IMDbld and TMDbld.

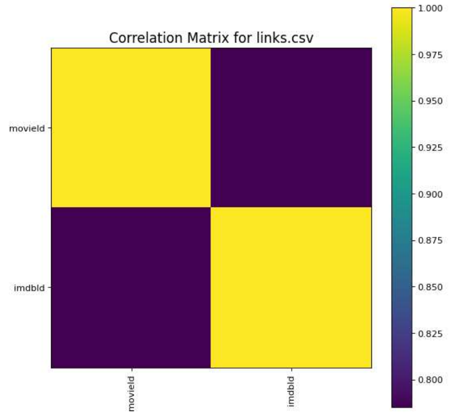


Figure 11 - Correlation matrix for links.csv

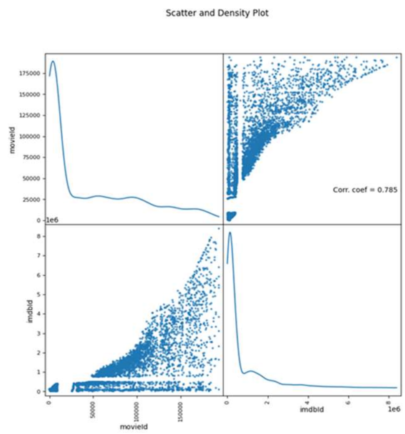


Figure 10 - Scatter and density plot for links.csv

There is also a strong correlation between MovieId and IMDbId. We, therefore, drop the column TMDbId as we do not need the information as it is duplicated by IMDbId and add the IMDbId URL to obtain the output.

	movieId	imdb_url
0	1	https://www.imdb.com/title/tt0114709/
1	2	https://www.imdb.com/title/tt0113497/
2	3	https://www.imdb.com/title/tt0113228/
3	4	https://www.imdb.com/title/tt0114885/
4	5	https://www.imdb.com/title/tt0113041/
...	...	...
9737	193581	https://www.imdb.com/title/tt5476944/
9738	193583	https://www.imdb.com/title/tt5914996/
9739	193585	https://www.imdb.com/title/tt6397426/
9740	193587	https://www.imdb.com/title/tt8391976/
9741	193609	https://www.imdb.com/title/tt0101726/

Figure 12 - Output after dropping TMDbId

Ratings.CSV

The ".describe()" function delivers the summary statistics of the data frame or series. This provides the column's count, mean, median (or 50th percentile), standard deviation, min-max, and percentile values. By default, the described method produces merely a summary of the dataset's numeric characteristics.

	userId	movieId	rating	timestamp
count	100836.000000	100836.000000	100836.000000	1.008360e+05
mean	326.127564	19435.295718	3.501557	1.205946e+09
std	182.618491	35530.987199	1.042529	2.162610e+08
min	1.000000	1.000000	0.500000	8.281246e+08
25%	177.000000	1199.000000	3.000000	1.019124e+09
50%	325.000000	2991.000000	3.500000	1.186087e+09
75%	477.000000	8122.000000	4.000000	1.435994e+09
max	610.000000	193609.000000	5.000000	1.537799e+09

Figure 13 - Output of the summary statistics for the ratings data set

The essential information from the above table reflects those 100 836 records on file. The mean user rating is 3.5 with a standard deviation of one, implying that most users' ratings are within this range of

2,5 - 4,5. The min rating is 0.5, and the maximum is five. It denotes the 25th, the 50th and the 75th percentile of our user ratings.

The distribution graphs that follow for ratings further confirm that most ratings are between 3 and 5 per user:

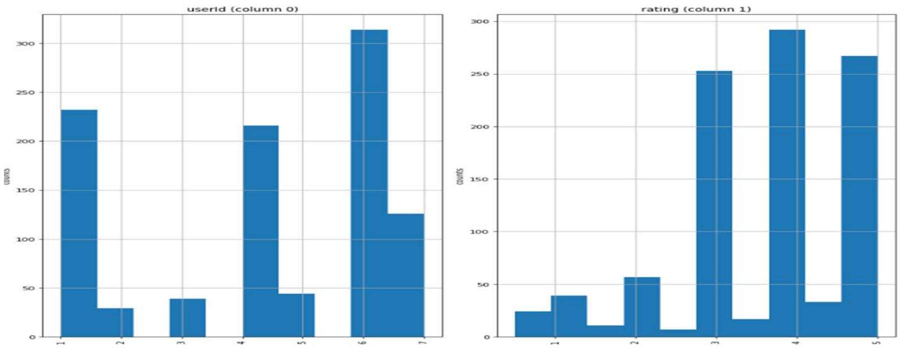


Figure 14 - Distribution graph of useId and ratings

Users rate a movie as average, with a rating of three, implying a general behaviour of not ordering disliked movies. This means that opposed movies are, in most instances, not rated by the user; the user moves on to the next search or movie choice.

The correlation matrix heatmap below reflects accordingly and suggests a stronger correlation between MovieIds and the time stamps, which implies that a more significant proportion of movies are being watched at around a similar time of the day. We use this information to understand which types of movies are being watched at specific times of the day and tailor our original productions to the users' particular preferences.

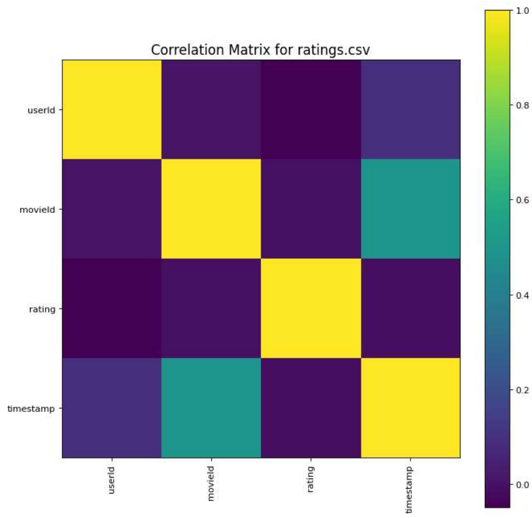


Figure 16 - Correlation matrix for ratings.csv

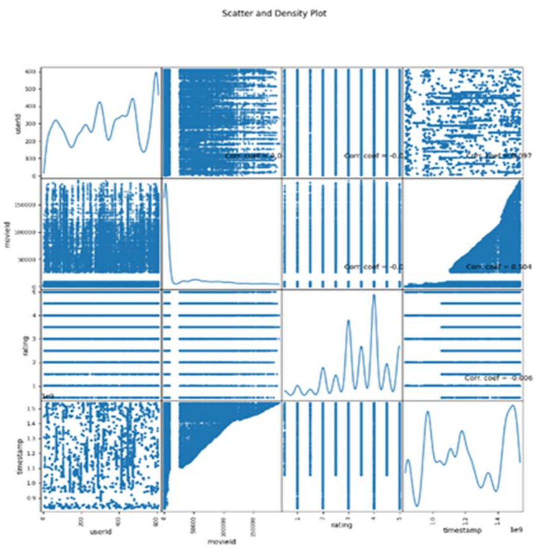


Figure 15 - Scatter and density plot for ratings.csv

We then prepare the ratings.csv file, with 100 836 records and four columns, by dropping the columns userId and timestamp and defining a function to take the data frame with MovieId and rating columns and a list of MovieIds and query the data frame to obtain the mean rating of that movie and put it in a new data frame along with the MovieId with the below code getting the adjacent output.

Input Code:

```
#Indexing MovieId
def get_movie_rating(df, movieIds):
    return df.query('movieId in @movieIds')['rating'].groupby(df['movieId']).mean()

movie_ratings = get_movie_rating(ratings_df, movies_id_list)
movie_ratings_df = movie_ratings.reset_index().rename(columns={'index': 'movieId', 'rating': 'mean_rating'})
```

Figure 17 - Defining the function to obtain ratings of movies

Output:

	movieId	rating
1	1	3.920930
2	2	3.431818
3	3	3.259615
4	4	2.357143
5	5	3.071429
...	...	...
193581	193581	4.000000
193583	193583	3.500000
193585	193585	3.500000
193587	193587	3.500000
193609	193609	4.000000

Figure 18 – Summary output for the rating of every movie

Tag.CSV

Subsequently, prepare the tags.csv file by dropping the user id and timestamp column. Then define the function to obtain the most common tag for every movie in our data set as reflected in the output. Lastly, we create an inner join to join all the cleaned and edited datasets using the MovieId column to join each dataset, obtaining the below result.

	movieId	title	genre	release_year	rating	tag	imdb_url
0	1	Toy Story	Adventure	1995.0	3.920930	pixar	https://www.imdb.com/title/tt0114709/
1	2	Jumanji	Adventure	1995.0	3.431818	fantasy	https://www.imdb.com/title/tt0113497/
2	3	Grumpier Old Men	Comedy	1995.0	3.259615	moldy	https://www.imdb.com/title/tt0113228/
3	4	Waiting to Exhale	Comedy	1995.0	2.357143	None	https://www.imdb.com/title/tt0114885/
4	5	Father of the Bride Part II	Comedy	1995.0	3.071429	pregnancy	https://www.imdb.com/title/tt0113041/
...	...	...	...	...	...	...	...
9737	193581	Black Butler: Book of the Atlantic	Action	2017.0	4.000000	None	https://www.imdb.com/title/tt5476944/
9738	193583	No Game No Life: Zero	Animation	2017.0	3.500000	None	https://www.imdb.com/title/tt5914996/
9739	193585	Flint	Drama	2017.0	3.500000	None	https://www.imdb.com/title/tt6397426/
9740	193587	Bungo Stray Dogs: Dead Apple	Action	2018.0	3.500000	None	https://www.imdb.com/title/tt8391976/
9741	193609	Andrew Dice Clay: Dice Rules	Comedy	1991.0	4.000000	None	https://www.imdb.com/title/tt0101726/

Figure 19 – Final Joined Dataset

Exploratory Data Analysis

Exploratory Data Analysis is a series of methodologies created in 1977 by John Wilder Tukey in his book Exploratory Data Analysis. The guiding principle of this strategy was to study the data before developing a model *(Tukey, 1977).* John Tukey urged statisticians to examine the data and develop hypotheses that could lead to the acquisition of further data or experiments. Today, data scientists and analysts devote most of their time to Data Wrangling and Exploratory Data Analysis (EDA).

Exploratory Data Analysis (EDA) is utilised to glean insights from data. Using statistical graphs and other data visualisation tools, data scientists and analysts attempt to identify various patterns, relationships, and anomalies. EDA includes the following elements:

- Gain the most insights possible from a dataset.
- Reveal the structure beneath
- isolate crucial factors from the dataset
- Determine outliers and anomalies (if any)
- Evaluate underlying assumptions
- Determine the ideal setting for each element

The primary objective of EDA is to identify errors, outliers, and unusual patterns in the data. It enables analysts to better comprehend the data before making assumptions. The results of EDA aid firms in understanding their consumers, expanding their operations, and making appropriate business decisions.

Once the necessary datasets are created with qualities such as an optimum, organised, and human-readable data format, we may proceed with a comprehensive analysis. We have selected EDA as the initial stage in our data analysis process. We utilise several methods to get the most information from our data collection.

What are the most popular genres of all time?

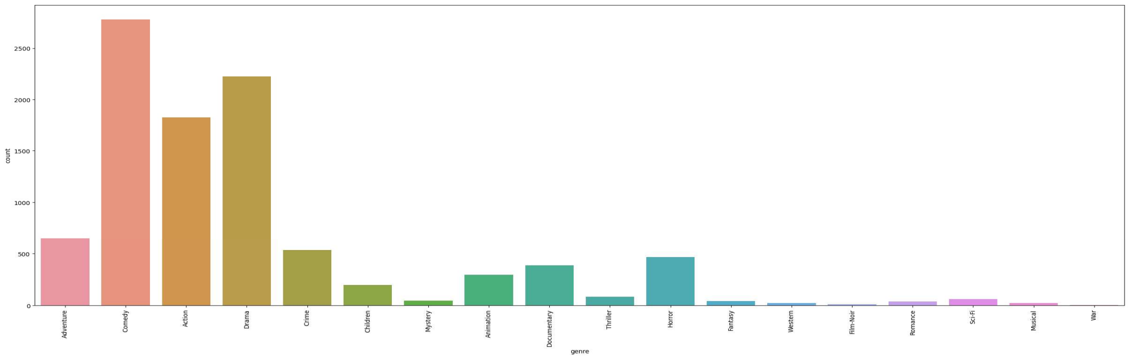


Figure 20 – Most famous genres of all time

As part of our initial analysis, we constructed a bar chart depicted in Figure 20 by genre. This reveals that the most watched movies are Comedies, followed by Drama and Action, whilst the least watched are War and Musical movies. It is likely that since Comedy, Drama and Action films generate the most box office revenue, these are also watched by the users as they resonate with the audience. These categories were more watched than adventure, crime, children, and horror.

This data analysis confirms that most users watch comedy movies as they are likely to be entertained. Users enjoy being considered because it allows them to unwind and escape reality. Many individuals have hectic lives and require time to recover and relax. As a result, people try to escape their frantic lives by viewing entertaining stories through comedy movies, in line with our expected outcomes. Our original movie production genres must align with these themes to increase user retention.

We can utilise the making of movies with the drama genre as they are popular among users because they are excellent at challenging viewers' thinking by introducing them to innovative ideas and humanising characters with opposing viewpoints. TV shows are a safe approach to exposing oneself to perilous situations and putting yourself in the shoes of others. People enjoy an adrenaline rush by watching action movies. It is only temporary tension over which we have total control. We may always pause or stop viewing it at any time. When we watch these movies, our brain is on high alert. It is more stimulated by film than by real-life circumstances.

The outcomes from the initial analysis reflect that our data set does not have significant outliers, and data points fall within the reasonable expectation of our outcomes. We would utilise this outcome to guide us in producing our original content ensuring the most watched genres are supported by the movie themes produced and that we stay away from those most disliked.

When were the movies released?

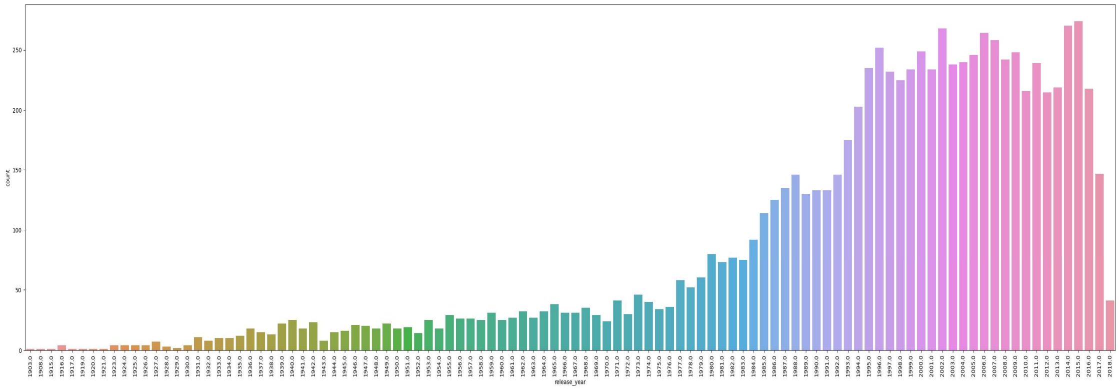


Figure 21 – Movies released

The bar chart of the count by release year reveals that films released experienced a significant increase from 1980 onwards. This is due to the development of video content creation technologies. With the advent of computers and digital technologies, production costs have gradually reduced. Digital tools for video editing were in high demand and accessible, reducing editing costs significantly, thus the increase in movie production.

The above growth in the release year is also attributable to the films of the 1980s, which featured a wide range of genres, including hybrids straddling numerous genres. The tendency to generate ever-larger blockbuster pictures that made more in their opening weeks than any previous film continued, thanks partly to planning releases when consumers had nothing else to pick from. Much of the dependence on these effect-driven blockbusters can be attributed to the Star Wars films at the start of this decade and the innovative visual effects they helped to pioneer. During this decade, the teen comedy subgenre also grew in popularity *(1980s in film, 2022)*.

By monitoring this data, we can understand when there is a decrease in movies released, as noted from 2016 onwards and accordingly schedule and plan our movie releases in periods when only a few other films are being released simultaneously.

When have most of the movies been released?

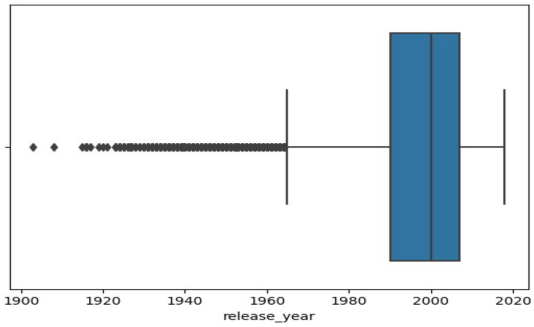


Figure 22 – Movies produced and released

The adjacent box plot further illustrates the rise in the number of films produced. Interestingly, films produced before 1960 are considered outliers given more recent volumes, with most movies, reflected by the interquartile range, created across two decades, 1990-201,0, which is the 25^{th} and 75^{th} percentile, respectively.

We would focus our original movie productions on content and theme based on the most popular genres and content developed between 1990 - 2010. This would be interpreted as a modern-day genre movie most suited to the audience.

What is the release year distribution per genre?

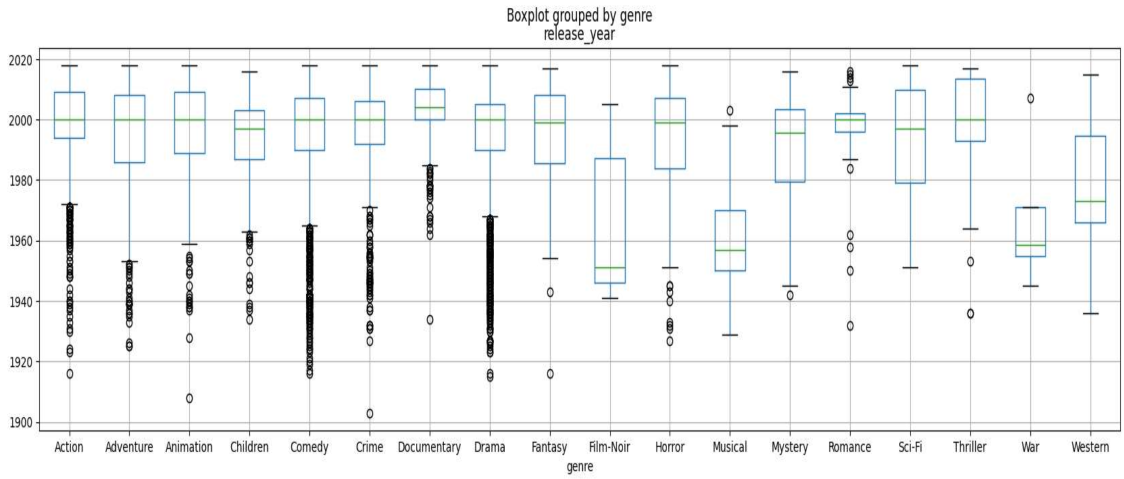


Figure 23 - Release year distribution per genre

We observe the distribution of releases per genre over the period. Most cinema genres have gained popularity during the past two decades. Most of the means for the genres are around the year two thousand, further refining our movie release period. Still, war-themed films and musical genres that were popular forty years ago are no longer.

A modern-day spin on these genres could be considered part of the original content production depending on the budget for such a movie production, given that current releases are sparse. Action, Comedies and Drama are films that always stay in style. Documentary movies have become more popular since 1983, while only a few Musical and War movies have recently been released. The more recent Romance releases are considered out of the ordinary, implying a concentration of such releases around the turn of the millennium.

The outcome guides us in the genres we would focus on within our original content production to have a more successful movie release, satisfying user experience, and appeal to a large user base. We would certainly have to consider more comprehensively the risks associated with producing content with a War or Musical genre, as these themes have since been outdated. There has been a recent concentration on the Thriller genre type of movies being released. If we wanted to play the production and release process with the least risk, we would produce an action, crime or even an animation movie.

What is the median genre distribution per year?

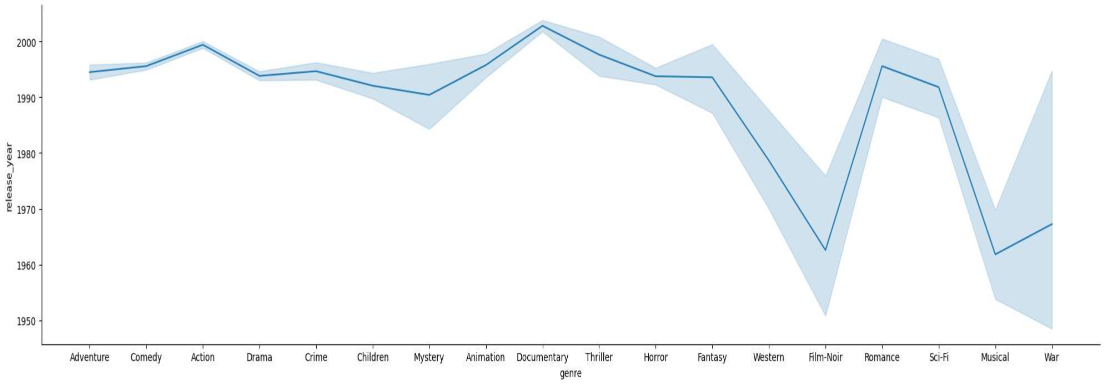


Figure 24 - Median genre distribution per year

Genre is the category in which a film's narrative components are classified. We observe the concentration of genres over time. According to the data, most documentary and Action films are very recent, yet film genres, Musicals and War, go as far back as the 1950s. This could guide us in producing original movie themes that have yet to be recently made in concentration and focus on the lesser-known genres to attract new customers to the platform. We could also cater to genres popular in foreign markets that are lesser known to broaden our user base further.

If a person consumes material that focuses more on women, the user will be shown a trailer for a film that focuses more on female characters. However, the same holds for many other characteristics, such as a person exclusively watching movies directed by filmmakers or starring actors. This report on each user decreases the time needed to research marketing methods, as we already know the customers' preferences and sensitive interests.

Genre films lend themselves to specialisation at every stage of the production process, raising the possibility of maximising revenues via the effective use of existing resources. The standardisation of genres is an industry technique. Producers gain from the genre since they can identify what is now thriving and develop more of the same to capitalise on profit.

What is the median distribution of ratings over time?

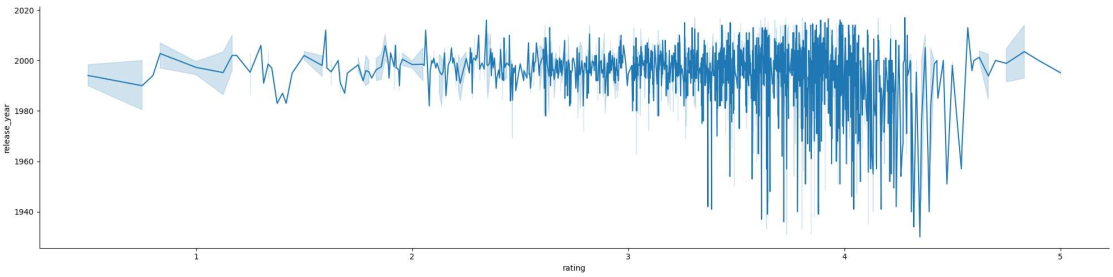


Figure 25 – Median rating distribution over time

The median of a variable is the value in the center of the data set when the values are sorted from least to most significant. It divides the data into two halves, with 50% of the data falling below the median and 50% above it. Most ratings are higher than three for release years back in the 1940s. This means that users deliberately choose the content they watch and rate. There is a concentration of ratings between 3 and 4,5, reaffirming that users tend not to rate movies with poor ratings and watch fewer movies they do not like. We rely on input from the users as one of the conventional data collection methods. After converting the comments into a rating, this is inputted into the system, improving enhancements or recommendations.

What are the Top 5 genre categories of all time?

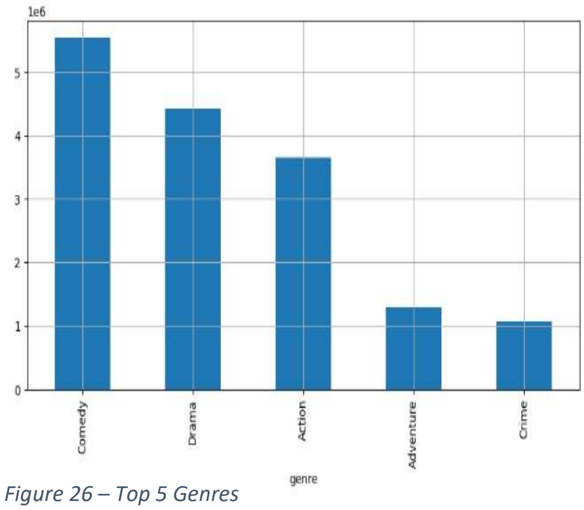
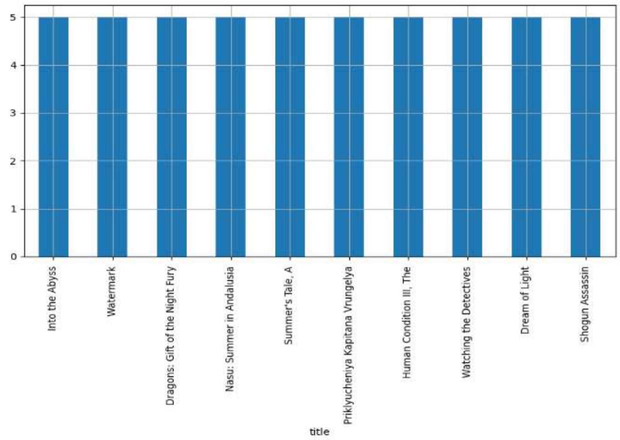


Figure 26 – Top 5 Genres

Film genres are classifications that categorise films according to narrative or stylistic characteristics. The film's genre can influence its characters, setting, story structure, and tone. For instance, action films generally feature fight sequences and slow-motion camera views. The film's genre is a simple categorisation technique for determining how it compares to other films. A large pool of related components groups films together, making it simpler to market them and to describe their overall experience concisely.

Audiences benefit from the genre because it helps them to distinguish distinct narrative styles and kinds, allowing them to choose what they love. As a result of the genre, producers can determine what is most profitable and, thus, produce more of the same to maximise profit.

Top ten most popular titles of all time (based on users' ratings)



A good film's title should suggest its genre. This gives readers and viewers a summary of the film's plot. Why is the title of a film significant? The audience's initial impression of our script is based on its title. Every movie needs a title. The title you choose may establish the tone of your screenplay and offer people a reason to view it (or avoid it). A good title reveals the film's plot and compels the audience to see it.

Figure 27 - Top 10 most popular titles of all time

Worst ten titles of all time (based on users' ratings)

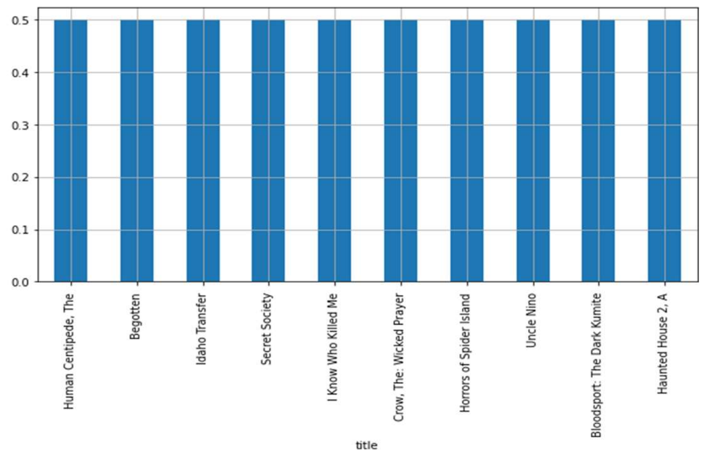


Figure 28 – Worst ten titles of all time

We create a line plot that visualises the median release year of movies for each genre, separated into two groups: those with ratings less than three and those with ratings greater than or equal to 3. We plot the median release year of movies grouped by genre and rating using the group by method, and then the median of the release year column is determined for each group.

```
#create a bar plot to visualize the popularity of different movie genres through the years
fig, ax = plt.subplots(figsize = (20,5))

df.groupby(['rating' > 2.5, 'genre'])['release_year'].median().unstack(level=0).plot(ax = ax, marker = 'o', markersize = 5, linestyle = 'dashed')

ax.set(xlabel = 'Genre',
       ylabel = 'Year',
       title = "Popularity of the genres through the years")

plt.grid(True, alpha = 0.5)
ax.legend(["negative rating", "positive rating"])
```

Figure 29 - Code for median positive and negative feedback

Output:

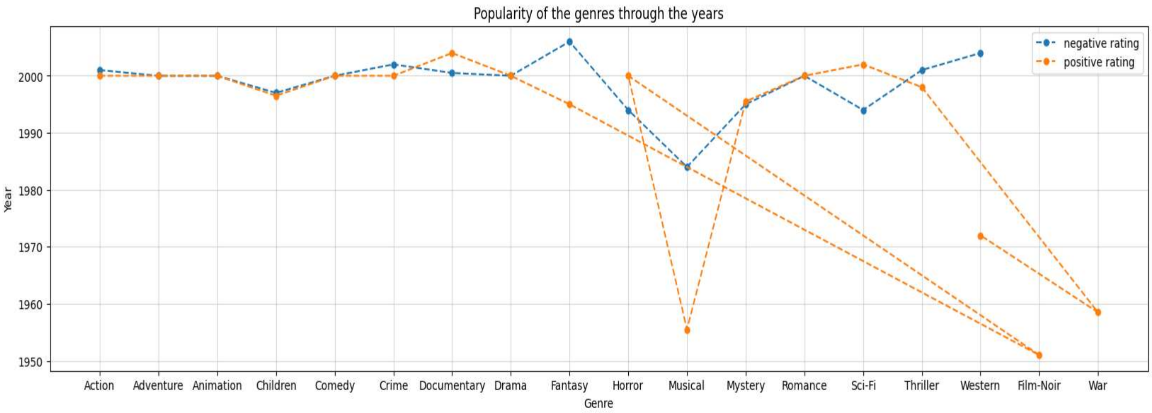


Figure 30 - Median positive and negative feedback

We look at ratings below and above 3. It is hardly unexpected that the most recent worst-rated genres are Fantasy, Science Fiction, given the abundance of such films in recent years, which are frequently remakes or adaptations of ancient masterpieces. Western movies are despised nowadays. Directors ran out of imagination and ideas since "everything has been seen". This guides the production team in knowing precisely when genres are trending and when they are not and tailoring the movie themes accordingly.

Top ten years for movies releases

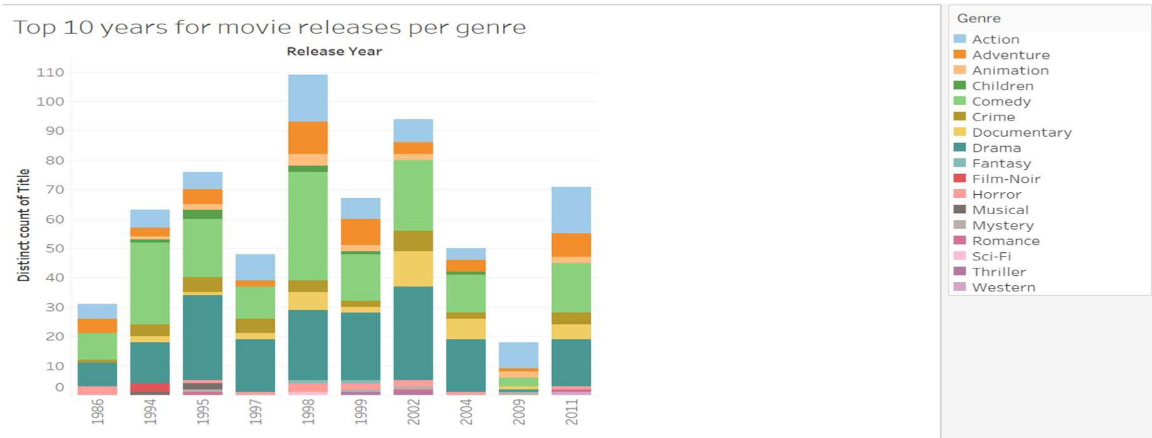


Figure 31 - Top 10 years for movie releases per genre

This bar graph reflects the most successful release years broken down into genres to guide further the possible timing of the product releases and accordingly schedule the production and planning process.

Top five tags per genre

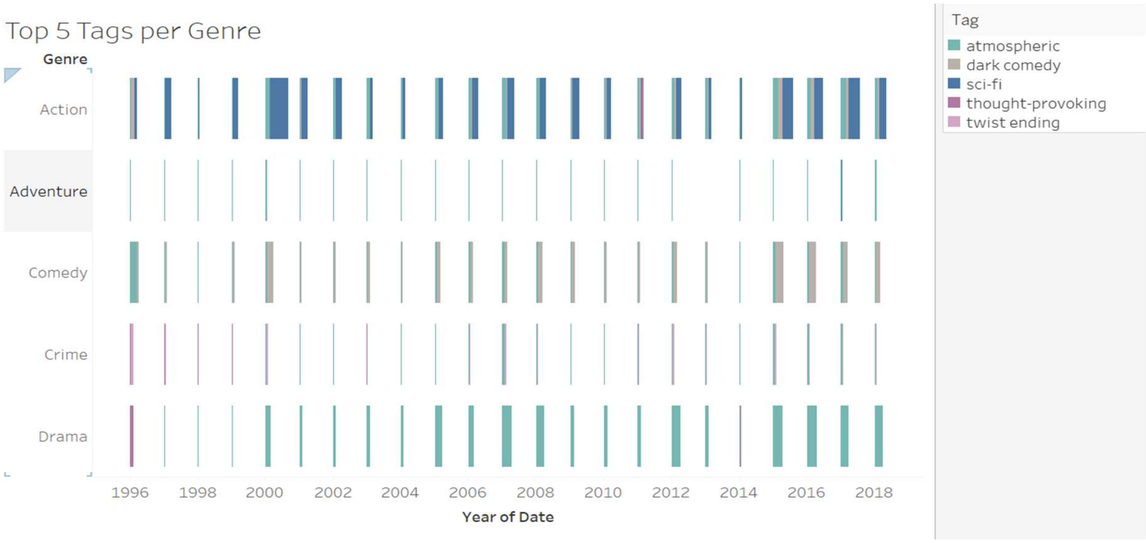


Figure 32 - Top 5 tags per genre

The figure illustrates the most popular tags used amongst the most popular genres. We observe that the tag atmospheric has been used extensively over the years to describe movies in the drama genre. Sci-Fi is the most popular tag used in the Action genre, followed by Dark Comedy within the Comedy genre. This indicates the underlying themes that should be utilised in our original production of movies and uses the popular tags to guide the process further.

The outcomes from the analysis reflect that our data set does not have significant outliers, and data points fall within reasonable expectations. We would utilise the study to guide us in producing our original content ensuring the most watched genres are supported by the movie themes produced and that we stay away from those most disliked. This data is also utilised in the recommendation systems and other algorithms employed through Netflix.

(Ricci et al. 2015) emphasises that recommendation systems are a "great illustration of the widespread application of large-scale data mining." More client information results in more refined recommendations, and vice versa. According to Nicholas Carah, this results in a “closed commercial loop in which culture adapts to its users more than it challenges them” *(Carah, 2016).* In other words, the culture does not influence the data; rather, the data influence the culture. The term for this situation is algorithmic culture *(Striphas, 2015).* These recommendation engines on Netflix contribute to this algorithmic society *(Hallinan & Striphas, 2014).* Netflix is contributing to this culture by developing its highly personalised genres and enhancing the efficacy and accuracy of its recommendation algorithms. In other words, Netflix continuously promotes the growth of this algorithmic culture.

Section 6

Recommender System for Movies

User-based and item-based collaborative filtering methods are discussed. To demonstrate the distinction between these two approaches, imagine we have yet to watch the film “Nomadland” (2020) and wish to determine whether we would enjoy it.

In user-based collaborative filtering, we seek moviegoers whose ratings of films we have viewed are most like our own. We can predict how much we would enjoy “Nomadland” (2020) based on the ratings of the moviegoers who are most like us who have already seen the film.

In item-based collaborative filtering (a technique pioneered by Amazon.com), we would first assess the similarity of all the films we have viewed to “Nomadland” (2020). Then, we estimate “Nomadland” (2020) 's a grade by giving greater weight to our ratings for the films that are most comparable to “Nomadland” (2020).

(Sarwar et al., 2001), in their writing, Item-Based Collaborative Filtering Recommendation Algorithms contain a detailed discussion of user-based collaborative filtering.

Make a movie score table with a user ID, title, and rating. We merge the movies and rating files on the MovieId column.

	movieId	title	genres	userId	rating	timestamp
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1	4.0	964982703
1	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	5	4.0	847434962
2	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	7	4.5	1106635946
3	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	15	2.5	1510577970
4	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	17	4.5	1305696483
...	...	...	...	...	...	...
100831	193581	Black Butler: Book of the Atlantic (2017)	Action|Animation|Comedy|Fantasy	184	4.0	1537109082
100832	193583	No Game No Life: Zero (2017)	Animation|Comedy|Fantasy	184	3.5	1537109545
100833	193585	Flint (2017)	Drama	184	3.5	1537109805
100834	193587	Bungo Stray Dogs: Dead Apple (2018)	Action|Animation	184	3.5	1537110021
100835	193609	Andrew Dice Clay: Dice Rules (1991)	Comedy	331	4.0	1537157606

100836 rows × 6 columns

Figure 33 – Movie rating table

Evaluating User Similarity

Numerous metrics are employed to assess the similarity of user evaluations on an excellent explanation of similarity measurements *(Blattberg et al., 2008).* We shall define the similarity between two users as the correlation between their assessment of all the films they have viewed. Remember that if the correlation between two people's ratings is +1, then if one person ranks a movie higher than average, it is more likely that the other person will do the same. If one person ranks a movie lower than average, it is more likely that the other person will do the same. Alternatively, suppose the correlation between two people's ratings is close to –1. In that case, it is more likely that if one person rates a movie higher than average, the other person will rank it lower than average, and if one person rates a movie lower than average, the other person will rate it higher than average.

We establish similarities between users in their rating patterns and then similarities between movies; refer to the below output.

userId	1	2	3	4	5	6	7	8	9	10
userId										
1	1.000000	0.027283	0.059720	0.194395	0.129080	0.128152	0.158744	0.136968	0.064263	0.016875
2	0.027283	1.000000	0.000000	0.003726	0.016614	0.025333	0.027585	0.027257	0.000000	0.067445
3	0.059720	0.000000	1.000000	0.002251	0.005020	0.003936	0.000000	0.004941	0.000000	0.000000
4	0.194395	0.003726	0.002251	1.000000	0.128659	0.088491	0.115120	0.062969	0.011361	0.031163
5	0.129080	0.016614	0.005020	0.128659	1.000000	0.300349	0.108342	0.429075	0.000000	0.030611
...	...	...	...	...	...	...	...	...	...	...
606	0.164191	0.028429	0.012993	0.200395	0.106435	0.102123	0.200035	0.099388	0.075898	0.088963
607	0.269389	0.012948	0.019247	0.131746	0.152866	0.162182	0.186114	0.185142	0.011844	0.010451
608	0.291097	0.046211	0.021128	0.149858	0.135535	0.178809	0.323541	0.187233	0.100435	0.077424
609	0.093572	0.027565	0.000000	0.032198	0.261232	0.214234	0.090840	0.423993	0.000000	0.021766
610	0.145321	0.102427	0.032119	0.107683	0.060792	0.052668	0.193219	0.078153	0.074399	0.121072

Figure 34 – Similarities between users in rating patterns and between movies

This recommendation system is based on user profiles with comparable characteristics *(Jain et al., 2018).* The recommendation system relies on the two pieces of information listed below to develop a subscriber profile.

- Subscriber preferences
- Subscriber history

If subscriber A watches crime, action, and horror films and subscriber B watches crime, action, and comedy films, then subscriber A will prefer comedy films. At the same time, subscriber B will choose horror films.

Then we recommend movies based on the cosine similarity function, which calculates the cosine similarity between each pair of rows in the rating_df data frame. Cosine similarity is used in this scenario to compare the ratings of various users (rows) in the data frame. The outcome is saved in the user sim variable with the following code:

```
#similarity betweens userID
from sklearn.metrics. pairwise import cosine_similarity
# The cosine_similarity function compares row to row
user_sim = cosine_similarity (rating_df, rating_df) user_sim
```

Figure 35 – Code for the similarity between users

We will look at ten other movies rated closest to “Jumanji (1995)” with the following code:

```
#returning the top 10 most similar movies to "Jumanji (1995)" based on the similarity values
movie_sim_df["Jumanji (1995)"].sort_values(ascending=False)[:10]
```

Figure 36 – Code for films rated most relative to Jumanji 1995

The “Lion King, Mrs Doubt Fire” and “The Mask” are the closest-rated movies to “Jumanji.”

```
title
Jumanji (1995)                                  1.000000
Lion King, The (1994)                           0.588438
Mrs. Doubtfire (1993)                           0.549818
Mask, The (1994)                                0.544981
Jurassic Park (1993)                            0.538046
Home Alone (1990)                               0.524876
Nightmare Before Christmas, The (1993)          0.518161
Aladdin (1992)                                  0.515620
Beauty and the Beast (1991)                     0.507458
Ace Ventura: When Nature Calls (1995)           0.497560
Name: Jumanji (1995), dtype: float64
```

Figure 37 – Output of similar movies like Jumanji 1995

We observe that audience ratings impact important movie performance metrics (i.e., movie revenues and new movie ratings). We, therefore, only need to invest heavily in promotion to maximise the movie's audience, as good ratings from past films support our production method. Viewers’ viewing and rating histories and the collective judgments of movie groups explain audience pleasure. Various components of these ratings describe viewers' new movie ratings as a measure of viewer satisfaction after correcting for movie attributes. In addition, a viewer's movie experiences might drive them to develop a more significant scepticism for ratings over time.

This recommendation system utilises consumer data *(Pazzani & Billsus, 2007).* When a new member first joins the service, the user is asked to enter information such as the genre he prefers and other details. Some of the data acquired from subscribers are listed below.

- It requests that the user rate the material.
- Data from a user's search.
- Sort the material from least to most favourite.
- Choose the better of two options.
- Request that the user create what he loves and hates.
- Analyzing the data from user searches.
- Monitoring the user's viewing history.

We then look at ten other movies rated like “The Godfather (1972)” with the following code:

```
#selecting the row with the title "Godfather, The (1972)"
movie_sim_df["Godfather, The (1972)"].sort_values (ascending=False) [1:11]
```

Figure 38 – Code for movies similar to The Godfather 1972

And obtain the output:

```
title
Godfather: Part II, The (1974)                              0.821773
Goodfellas (1990)                                           0.664841
One Flew Over the Cuckoo's Nest (1975)                      0.620536
Star Wars: Episode IV - A New Hope (1977)                   0.595317
Fargo (1996)                                                0.588614
Star Wars: Episode V - The Empire Strikes Back (1980)       0.586030
Fight Club (1999)                                           0.581279
Reservoir Dogs (1992)                                       0.579059
Pulp Fiction (1994)                                         0.575270
American Beauty (1999)                                      0.575012
Name: Godfather, The (1972), dtype: float64
```

Figure 39 – Output for movies similar to The Godfather Part II

"The Godfather Part 2 (1974)" is the highest-rated similar movie, with the rest of the film being below in terms of rating. Sequels earn more money but receive poorer reviews than originals.

The system creates suggestions from a source based on the features associated with items and the user's information in a content-based recommendation engine. Content-based recommenders approach suggestion as a user-specific classification issue, learning a classifier based on product attributes for the user's likes and dislikes. Offers are created by collaborative recommendation engines based on evaluations supplied by a group of individuals. It finds peer users with comparable rating histories to the present user and produces suggestions for the user. The system requires extra data on the context of item consumption, such as time, mood, and behavioural features, in the context-based recommendation engine. This data might be utilised to improve the suggestions offered to the viewer and improve what could be done without this additional source of information.

We are discussing the item-based strategy outlined in (*Sarwar et al., 2001)*. Companies with a large client base, such as Amazon.com, prefer the item-based method over the user-based approach because the item-based matrix of correlations is more stable over time than the user-based matrix of correlations and hence requires fewer frequent updates.

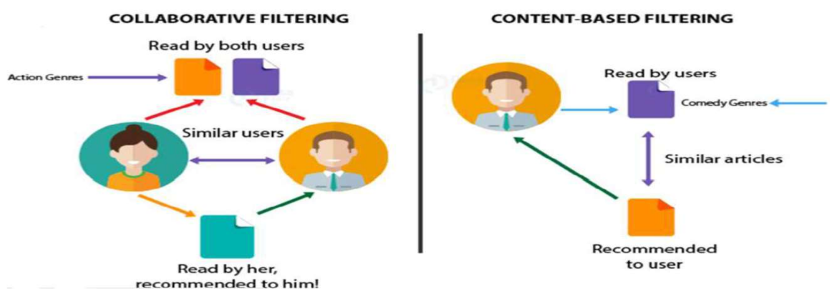


Figure 40 - Schematic diagram of collaborative and content filtering

Hybrid Recommendation System - Content and Collaborative

To build a system that gives more accurate movie recommendations, Netflix employs a combination of collaborative and content-based filtering resulting in a mixed recommendation algorithm, a blended outcome of these approaches. This recommendation system creates a recommendation by combining the subscriber's viewing and searching behaviours with the subscriber's history. The platform makes suggestions by analysing comparable users' viewing and search behaviours (i.e., collaborative filtering) and providing movies that share features with films that a user has already rated highly (content-based filtering). Netflix is an excellent example of a hybrid recommender system in practice.

These days, most recommender systems adopt a hybrid approach, fusing elements of collaborative filtering, content-based filtering, and other methods. The Netflix recommendation system employs the "three-legged stool" idea *(Plummer, 2017).* The first leg is the history of what Netflix users have seen, followed by the taggers. Tags are created by Netflix personnel who are knowledgeable about all aspects of the video, and then lastly, the proprietary machine-learning algorithms combine all the data.

There are "a couple thousand" different taste groups for viewers. These groups determine which suggestions appear at the top of the interface, which genre rows are presented, and how each row is sorted for each viewer. The tags employed by machine learning algorithms are consistent across the world. Nevertheless, a smaller subset of tags is utilised externally, flowing directly into the user interface and varying by nation, language, and cultural environment.

The data that Netflix feeds into its algorithms can be broken down into two types – implicit and explicit. "Explicit data is what you literally tell us: you give a thumbs up to *The Crown*, we get it," Yellin explains. "Implicit data is really behavioural data. You did not explicitly tell us 'I liked *Unbreakable Kimmy Schmidt*'. You just binged on it and watched it in two nights, so we understand that behaviourally. Most useful data is implicit" *(Plummer, 2017).*
Netflix analyses feedback from every visit to the site and continuously retrains its algorithms to increase the accuracy of its forecast of what you are most likely to view. Their data, algorithms, and calculation systems continue to feed into one another to present you with a product that provides an enhanced customer experience.

This principle of operation for recommendation engines can function as an intelligent decision support system that improves decision-making and promotes items and services to enhance sales and customer experiences. These may also increase productivity or establish automatic decision-making procedures.

Listed below are a few hybrid approaches:
- Weighted: In this hybrid approach, two or more recommendation systems make recommendations to users depending on their preferences, and then the scores of the multiple recommendation systems are merged to obtain high accuracy.
- Combination of Features: The result of one or more recommendation systems is delivered as input to the final recommendation system in this hybrid approach.
- Cascade: A hybrid recommendation system with a phased suggestion process depending on priority.
- Switching: In this hybrid technique, switching occurs between recommendation systems based on criteria.
- Feature Augmentation: The result of one recommendation system is offered as input to another recommendation system in this hybrid method. On a higher level, this hybrid recommendation system is order sensitive.
- Mixed: This is the most basic sort of hybrid recommendation system. Diverse types of recommendation systems are merged to generate suggestions in this system.

We use the data analysed to establish the trends that lead viewers to, for example, as stated by Netflix, the Marvel heroes comprising The Defenders, to demonstrate how all this data works together to help viewers discover the latest content. The algorithms accumulate and analyse data points to improve recommendations and provide users with relevant content. Netflix incorporates all its data into its recommendation engines.

Limitations of the Recommendation System

Cold Start: For every new subscriber, more data must be provided to provide recommendations at the start.
Accuracy: one of the primary restrictions since it relates to the recommendation system's capacity to forecast what a user wants more correctly.
Diversity: beneficial in preventing prejudice. There should be a wide range of recommendations provided.

Scalability: Due to a large number of users and content, a massive amount of processing power is required to deliver correct recommendations in the quickest period feasible.
Data Scarcity: Because the database is so extensive, the most active users may have only rated a small portion of the information.

Metrics for Evaluating Recommendation Algorithm Performance

The quality of a recommendation system may be measured using many metrics, such as precision or coverage. The metrics utilised depend on the filtering strategy employed. Accuracy is the proportion of suitable suggestions from the total number of potential recommendations. In contrast, coverage is the proportion of items in the search space for which the system can make recommendations. The accuracy measurements for recommendation filtering systems are separated into statistical and decision support measures *(Gaurav 2021).* Each metric's usefulness depends on the dataset's characteristics and the tasks the recommender system will do.

Section 7

How Does Netflix Employ Big Data Analytics to Enhance Customer Satisfaction and Recommend Movies?

Netflix Recommendation System

It suggests several options to the users based on their interests. Machine learning is incorporated into recommendation systems. It accepts user input and makes recommendations accordingly. In Netflix's recommendation system, user data such as location, user interests, and viewing time are collected when the material is seen and searched by the user. Its system generates a tailored suggestion depending on the subscriber's claims based on these characteristics *(Anon, 2022).*

In most recommendation systems, the subscriber profile is an essential factor. The subscriber profile includes several types of information, such as the subscriber's interests, search query history, system interactions, etc. When a new subscriber account is created or a new profile is added to an existing account, Netflix will ask the subscriber to select a few genres or titles that will serve as the initial criteria for the recommendation system. If the subscriber completes this step, Netflix will populate the user's homepage with the most popular material. Once the user begins to view the content, this will supersede whatever initial preferences the subscriber gave as the subscriber keeps watching. The previously seen information will be utilised to generate more recommendations *(Netflix et al., 2016).*

The subscriber's past viewing habits determine this suggestion algorithm. If the subscriber has previously seen content or movies of the same genre, such as action or comedy, he is likely to do so again. This recommendation system utilises consumer data *(Pazzani & Billsus, 2007).* When new subscribers join the service, they are prompted to give information such as preferred genre and other preferences. Some of the information acquired from subscribers includes the following:

- It asks the user to rate the content
- User Search data
- Rank the content from least favourite to most
- Choose the better of two items

- Ask the user to create what he likes and what he dislikes
- Analyzing the user search data
- Tracking the user viewing history

Recommendation Systems are one of the most valuable business information systems. The typical application of these systems serves two primary functions. One is giving people more personalised recommendations to meet their needs. The second objective is for businesses to maximise profitability by providing precise and rapid client responses. Thus, corporations may design customised portfolios that satisfy their profit maximisation objectives. According to Forbes, Amazon's recommendation engine generates 35 percent of the company's yearly revenue *(Morgan, 2021).* These statistics demonstrate the significance of recommendation systems for the growth of all technological firms; e-commerce, social media, video, and online news platforms have been aggressively installing their recommender systems to assist their clients in selecting items more effectively, which is a win-win approach. Like other recommendation systems, movie recommendation systems utilise various deployment methods.

In the literature, Collaborative Filtering, Content-Based Filtering, and Hybrid Filtering are the most prominent methodologies.

Collaborative Filtering (CF) is one of the most successful strategies based on the premise that users who have previously watched the similar would continue to do so in the future. With this knowledge, CF attempts to group users based on their commonalities. Consequently, an unrated movie or item by a particular user may be deduced by analysing the cluster to which the user belongs (*Resnick & Varian, 1997) (Linden et al., 2003).*

Content-Based Filtering (CBF) pairs items with users to provide valuable recommendations. The primary premise of the CBF is that if a user was interested in an item(s) in the past, they are likely to be interested in it(them) in the future. Regarding movie recommendation, CBF is utilised on several datasets *(Reddy et al., 2018), (Son & Kim, 2017).*

Hybrid Filtering (HF) blends content-based and collaborative filtering to improve the accuracy of suggestions. For example, a single recommender system (RS) can utilise CF to propose comparable items to similar users, and CBF recommends a particular user based on their historic preferences. In the literature, there is additional research employing HF *(Wang et al., 2018) (CHRISTAKOU et al., 2007) (Jain et al., 2018).*

A crucial component is the RS, which assists users in locating movies to view every session. The Netflix recommender system is not a single algorithm but a collection of algorithms that serve particular use cases and work together to produce the Netflix experience (Netflix et al., 2016).

Predicting Viewing Habits of Subscribers

To generate individualised suggestions for its users, Netflix collects massive amounts of user information. The following are some of the items that Netflix monitors:

- When a user rewinds, pauses, or fast-forwards content.
- What time do you watch content?
- What day do you consume material (Netflix discovered that people watch TV shows during the week and movies on weekends.
- If a user stops viewing a piece of material.

- Duration of the user's watch.
- The viewing history of the user.
- What device does he use to watch? (TV, Smartphone etc.)
- User's browsing habits
- User's searching habits
- User reviews and ratings are displayed.

Netflix predicts the watching behaviour of its users using a mix of methods. These strategies include collaborative filtering, which analyses the viewing habits of groups of people to provide suggestions, and content-based filtering, which analyses the features of items to generate recommendations.

Netflix also uses machine learning algorithms to tailor user suggestions. These algorithms examine several data elements, such as the user's watching history, ratings, and interactions with the Netflix site, to provide tailored recommendations *(Opheliac, 2015).*

In addition to these methodologies, Netflix utilises data from additional sources, such as social media and third-party data providers, to increase the accuracy of its forecasts and strengthen its recommendation engine. The company provides customised suggestions based on the above data. Netflix also examines data contained within movies *(Wayne, 2021).* They take numerous "screenshots" to examine "in-the-moment" qualities. Netflix has confirmed that they know when the credits begin to roll.

Finding the Next Hit Television Series with Big Data

Netflix began producing and distributing original content and developing its own TV shows. Big data drives the Netflix approach. Netflix won the bid to create House of Cards over competitors such as Home Box Office (HBO) and American Movie Classics (AMC). Netflix invested $100 million for two seasons and twenty-six episodes without seeing a single program episode. This choice was influenced by analytics of big data. By examining prior viewing data, Netflix was able to establish that viewers of the original House of Cards, initially broadcast in the United Kingdom in the 1990s, were interested in films directed by David Fincher and starring Kevin Spacey *(Havel, 2016).* Netflix was convinced that their program would be a smash hit based on the big data analytics they conducted. Netflix's success rate for original programming is over 80%, far higher than rival networks. The development of original material is determined by utilising big data analytics and other data mining techniques to assess the size of the audience interested in viewing these original contents *(Spangler, 2019).*

Netflix was confident that a remake of the British House of Cards drama starring Spacey and directed by Fincher would have a guaranteed audience based on this information. The expectation was that word-of-mouth from these core viewers would reach a large audience. Episode one of the last season (Season 6, minus Kevin Spacey) was seen by 2,875,000 subscribers during its premiere week. In 2013, when House of Cards premiered, Netflix's stock price was $25 *(Sull, 2015).*

Jonathan Friedland, Chief Communications Officer, said, “Because we have a direct relationship with consumers, we know what people like to watch, and that helps us understand how big the interest is going to be for a given show. It gave us confidence that we could find an audience for a show like House of Cards” *(Car, 2013) (Candeub, 2018).*

Netflix's stock price peaked in 2019 at $421! I am confident that House of Cards' introduction to binge-watching contributed to this astounding gain in Netflix's stock price. In the summer of 2019, around 65-million-member homes watched Season 3 of Stranger Things during its first four weeks of streaming availability. During its first 30 days on Netflix in 2020, 64 million households viewed Tiger King (Sherman, 2020).

Netflix identifies prospective successful television programs with a combination of data-driven processes. These approaches include:

- gathering information on how users interact with its platform, including which episodes they watch, for how long, and how they rate them. This information is utilised to uncover viewing behaviour patterns and identify notably popular or engaging programs.
- utilises machine learning algorithms to examine the data it gathers and generate predictions about which sorts of television programs are likely to be successful. These algorithms can identify trends and patterns in the data that may not be readily evident to human researchers.
- employs external data sources, such as social media and third-party data providers, to improve its understanding of user preferences and uncover potential hit series.

Combining these strategies will enable Netflix to find potential successful TV shows and decide which ones to produce and distribute.

Netflix Personalised Customer Experience

Netflix employs A/B testing to customise the user experience. What is displayed on the platform (e.g., dashboard, photos) is determined by the A/B test data. Each test gives the subscriber two distinct versions and monitors the subscriber's response to each. Netflix makes use of landing cards. Landing cards are graphics or video teasers on the subscriber's homepage. Landing cards help encourage subscribers to see the material. Users are likelier to see the information if the visuals pique their interest *(Khandelwal, 2022).*

Before becoming the default user experience, every product update Netflix considers through extensive A/B testing. Major redesigns enhance the service by making it easier for users to locate the desired material. However, they are only hazardous to implement with rigorous A/B testing, demonstrating that the new experience is superior to the old one. Even the pictures connected with many titles are subjected to A/B testing, which might result in 20% to 30% more page views for that topic *(Tingley et al., 2016).*

Netflix provides a tailored consumer experience using a mix of methods, including:

- employing machine learning algorithms to assess a user's watching history, ratings, and platform interactions to generate tailored recommendations for TV episodes and movies.
- The homepage is personalised for each user depending on their watching history and personal preferences. This allows viewers to view suggestions for programmes and films that are specifically targeted at them.
- allows users to establish numerous profiles inside a single account, each with its personalised watching history and suggestions. This allows each member of a family to have a unique Netflix experience.

- permits users to modify their playback options, including the default language, subtitles, and playback speed. This enables viewers to customise their watching experience according to their tastes.

Using these strategies, Netflix creates a personalised customer experience tailored to each user's interests and preferences. These results highlight the significance of A/B testing. Adhering to an empirical methodology guarantees that product changes are not dictated by the most opinionated and outspoken staff but instead supported by objective facts. This allows users to lead the company strategy toward the experiences they most like.

Section 8

Conclusion

The use of data analytics and recommendation systems is not limited to the film industry. The internet enables mass personalisation, and recommendation algorithms are a significant component of this personalisation. Recommendation systems construct models based on user choices to customise the user experience. Data analytics and recommendation systems are a pillar of these leading companies and require significant volumes of data to enhance improvement giving them a competitive advantage.

In practice, both collaborative filtering and content filtering are utilised. Amazon, Facebook, and Google News are among the businesses that employ collaborative filtering. Companies like Pandora, Rotten Tomatoes, and See This Next use content filtering. In addition, Netflix utilises collaborative and content filtering implying a hybrid approach *(Khan et al. 2021).*

In this digital age, firms frequently offer their clients hundreds of thousands of goods, including movies, music, and Facebook friends. Outstanding recommendation systems may make or ruin these companies. These recommendation systems are based on algorithms that identify comparable consumers or related products.

Netflix is a paradigm for rapid innovation, utilising a range of techniques to improve its business, such as its movie selection algorithms and cloud-based entertainment service delivery operations, as well as cloud-based experimentation and open challenges. The organisation is a collective intelligence model, collecting data points from millions of users' watching behaviours and preferences to provide real-time personalised recommendations to each user *(University 2020).*

The success of a product or service also hinges on how a brand convinces its customers to use the product repeatedly without advertising or promoting it. The more time consumers invest in a product, the more they will appreciate it. The recommendation engine collects more data as consumers view more content, enabling a deeper understanding of consumer behaviour and a more personalised Netflix experience. The enhanced service results in a more captivating experience. It serves as an inhibitor to using the product and service again.

"Binge viewing" is an example of a consumer's time invested. The simultaneous release of all television series episodes caused consumers to spend hours at a time. Automatically loading subsequent episodes prevents us from leaving the television screen. We become more devoted to the service the more information we put into it. The user would have to invest additional time, cognitive effort, data, and social capital if they left Netflix and switched to a rival.

Only if the product or service has a high perceived utility value will a customer continue to utilise it repeatedly. Netflix is determined to provide high-quality entertainment. Customers are directed to films they may have yet to see through Netflix's recommendation engine. This diversity of films entices a large audience. It is one of the most effective methods for engaging users. Variability multiplies the influence of Dopamine rush in the brain, creating a concentrated state, and suppresses the regions of the brain connected with wanting and want, resulting in the development of a desire for the product or service.

Netflix has established an entire research department that relates to their business and engineering teams, going beyond the use of data analytics to promote their business *(Anon, 2022).* They have provided open-source machine learning algorithms and Python frameworks to increase the productivity of data scientists and businesses.

From the recommendations engine to deciding which original shows and films to produce, Netflix has leveraged vast amounts of data to capture its audience and continues to expand. As a customer-focused organisation, they effectively sell their service as a tailored streaming experience for each user. In addition to being a streaming service, they are a data behemoth in the entertainment business.

This project evaluated Netflix's business strategy and determined that the company focuses on optimising the subscriber experience using big data analytics. In addition, the IT strategy generation and recommendation system serve as a transformational and competitive advantage-building tool.

We examined important collaborative filtering principles using data samples and covered user-based and item-based collaborative filtering approaches. Finally, we have concluded that analytics is an indispensable differentiator in the digital age. The development of sophisticated algorithms and machine learning significantly impacts how organisations use advanced decision-making tools and their ability to become more competitive in the market.

The data analysis and recommendation system provide a unique perspective on Netflix's achievement of its strategic objectives. Netflix is an example of collective intelligence since it uses information about millions of users' viewing habits and preferences to tailor its recommendations to each viewer in real time. By leveraging data-driven insights, companies can improve their overall performance, mitigate risks, and remain competitive in a rapidly changing market.

Finally, Netflix employs these technologies to build models based on user preferences to tailor the user experience and gain a competitive advantage. As a pioneer in the industry, the organisation uses a variety of strategies to enhance value, including its movie selection algorithms, cloud-based entertainment service delivery and open-source machine learning algorithms.

The influence of data analytics and recommendation systems on customer engagement and loyalty is highlighted. We discussed how customisation increases consumer engagement and retention. Companies like Netflix can develop a more meaningful relationship with their customers and keep them coming back for more by delivering tailored suggestions. The recommendation system engages consumers and offers a customised viewing experience. As the Subscription Video on Demand (SVOD) industry continues to expand, large-scale data management, business analytics, and the utilisation of big data will be the strategic axis for companies to gain insights, improve predictions, identify risks, and function efficiently and effectively, resulting in an exhilarating customer experience.

Bibliography

1980s in film (2022) *Wikipedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/1980s_in_film (Accessed: October 14, 2022).

America, N. 2021. Why Am I Seeing This? Available at: http://newamerica.org/oti/reports/why-am-i-seeing-this/ (Accessed: 23 January 2023).

Anon, 2022, How Netflix's recommendations System Works. Help Center. Available at: https://help.netflix.com/en/node/100639 (Accessed August 29, 2022).

AWS Innovator: Netflix | Case Studies, Videos and Customer Stories. (no date). Available at: https://aws.amazon.com/solutions/case-studies/netflix/ (Accessed: 24 January 2023).

Bell, R.M. and Koren, Y. (2007) "Lessons from the Netflix Prize Challenge," *ACM SIGKDD Explorations Newsletter*, 9(2), pp. 75–79. Available at: https://doi.org/10.1145/1345448.1345465.

Blattberg, R.C., Kim, B.-D. and Neslin, S.A. (2008) *Database marketing: Analyzing and managing customers*. New York: Springer.

Blog, N.T. 2021. It's All A/Bout Testing. Available at: https://netflixtechblog.com/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15 (Accessed: 10 January 2023).

Brennan, L. (2021) *How netflix expanded to 190 countries in 7 years*, *Harvard Business Review*. Available at: https://hbr.org/2018/10/how-netflix-expanded-to-190-countries-in-7-years (Accessed: November 20, 2022).

Candeub, A. 2018. Netflix's Race-Based Marketing Shows Potential for Anticompetitive Data Abuses. Available at: https://www.forbes.com/sites/washingtonbytes/2018/10/30/netflixs-race-based-marketing-shows-potential-for-anticompetitive-data-abuses/ (Accessed: 23 January 2023).

Carah, N. (2016) "Algorithmic brands: A Decade of brand experiments with Mobile and Social Media," *New Media & Society*, 19(3), pp. 384–400. Available at: https://doi.org/10.1177/1461444815605463.

Carr, D. 2013. Giving Viewers What They Want. Available at: https://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html (Accessed: 23 January 2023).

CHRISTAKOU, C.H.R.I.S.T.I.N.A., VRETTOS, S.P.Y.R.O.S. and STAFYLOPATIS, A.N.D.R.E.A.S. (2007) "A hybrid movie recommender system based on Neural Networks," *International Journal* *on Artificial Intelligence Tools*, 16(05), pp. 771–792. Available at: https://doi.org/10.1142/s0218213007003540.

Company profile (2022) *Netflix*. Available at: https://ir.netflix.net/ir-overview/profile/default.aspx (Accessed: November 20, 2022).

Complete List of Netflix Originals. 2022. Available at: https://www.whats-on-netflix.com/library/originals/ (Accessed: 23 January 2023).

Cool, R. (2019) *Detecting performance anomalies in external firmware deployments - medium*, *Detecting Performance Anomalies in External Firmware Deployments*. Available at: https://netflixtechblog.com/detecting-performance-anomalies-in-external-firmware-deployments-ed41b1bfcf46 (Accessed: November 19, 2022).

Fancey, G. (2022) *Netflix*, *Movie ideas Wiki*. Available at: https://MovieIdeas.fandom.com/wiki/Netflix (Accessed: October 7, 2022).

Free Encyclopedia, W. (2023) Data Analysis, Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Data_analysis (Accessed: January 26, 2023).

Gaurav, gaurav. product 2021. Measuring the performance of recommendation system: Metrics (Practical Guide)— Part IV. Available at: https://medium.com/product-design-data/measuring-the-performance-of-recommendation-system-metrics-practical-guide-part-iv-fd0d9ec7c6d3 (Accessed: 23 January 2023).

Gloria, A. (2018) *The origin of Netflix's success*, *Digital Innovation and Transformation*. Available at: https://d3.harvard.edu/platform-digit/submission/the-origin-of-netflixs-success/ (Accessed: November 18, 2022).

Gomez-Uribe, C.A. and Hunt, N. (2016) "The Netflix Recommender System," *ACM Transactions on Management Information Systems*, 6(4), pp. 1–19. Available at: https://doi.org/10.1145/2843948.

Hallinan, B. and Striphas, T. (2014) "Recommended for you: The netflix prize and the production of algorithmic culture," *New Media & Society*, 18(1), pp. 117–137. Available at: https://doi.org/10.1177/1461444814538646.

Harper, F.M. and Konstan, J.A. (2016) "The movielens datasets," *ACM Transactions on Interactive Intelligent Systems*, 5(4), pp. 1–19. Available at: https://doi.org/10.1145/2827872.

Havel, J. 2016. How "House of Cards" Changed the Content Marketing Game. Available at: https://thehustle.co/netflix-house-of-cards-content-marketing/ (Accessed: 23 January 2023).

Jain, K.N. *et al.* (2018) “Movie recommendation system: Hybrid Information Filtering System,” *Intelligent Computing and Information and Communication*, pp. 677–686. Available at: https://doi.org/10.1007/978-981-10-7245-1_66.

Khan, Z *et al*. (2021). Contextual recommender system for E-commerce applications. Available at: https://www.sciencedirect.com/science/article/abs/pii/S1568494621004737 (Accessed: 23 January 2023).

Khandelwal, A. (2022) *How does Amazon & Netflix personalization work?*, *Blog*. Available at: https://vwo.com/blog/deliver-personalized-recommendations-the-amazon-netflix-way/ (Accessed: November 16, 2022).

Lee, J. 2022. Streaming Services: Breaking Into a Competitive Market | AdRoll. Available at: https://www.adroll.com/blog/streaming-services-breaking-into-a-competitive-market (Accessed: 23 January 2023).

Linden, G., Smith, B. and York, J. (2003) “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Computing*, 7(1), pp. 76–80. Available at: https://doi.org/10.1109/mic.2003.1167344.

Moore, A. (2019) *Netflix's generic strategy, Business Model & Intensive Growth Strategies*, *Rancord Society*. Available at: http://www.rancord.org/netflix-business-model-generic-strategy-intensive-growth-strategies-competitive-advantage#:~:text=Netflix's generic strategy focuses on, to expand multinational streaming operations. (Accessed: October 27, 2022).

Morgan, B. (2021) How Amazon has reorganized around Artificial Intelligence and machine learning, Forbes. Forbes Magazine. Available at: https://www.forbes.com/sites/blakemorgan/2018/07/16/how-amazon-has-re-organized-around-artificial-intelligence-and-machine-learning/?sh=5b4b041b7361 (Accessed: January 23, 2023).

Needle, F. (2021) *How netflix maintains low Churn & High Customer Retention*, *HubSpot Blog*. HubSpot. Available at: https://blog.hubspot.com/service/netflix-customer-service-and-churn-rate (Accessed: August 30, 2022).

Netflix - study in China, (2022) - wiki english (1997) *Study in China 2023*. Available at: https://www.duhoctrungquoc.vn/wiki/en/Netflix (Accessed: October 7, 2022).

Netflix – behind the scenes, (2015) *Technology and Operations Management*. Available at: https://d3.harvard.edu/platform-rctom/submission/netflix-behind-the-scenes/ (Accessed: November 14, 2022).

Netflix *et al,* (2016) *The Netflix Recommender System: Algorithms, business value, and innovation: ACM Transactions on Management Information Systems: Vol 6, no 4*, *ACM Transactions on Management Information Systems*. Available at: https://dl.acm.org/doi/10.1145/2843948 (Accessed: October 9, 2022).

Niwate, S. (2021) *How Netflix uses Data Analytics: A case study - medium*, *How netflix uses Data Analytics: A case study, Medium. Python in Plain English*. Available at: https://python.plainenglish.io/how-netflix-uses-data-analytics-data-science-general-research-case-study-4d525b881038 (Accessed: October 13, 2022).

Opheliac 2015. Netflix – Behind the Scenes - Technology and Operations Management. Available at: https://d3.harvard.edu/platform-rctom/submission/netflix-behind-the-scenes/ (Accessed: 23 January 2023).

Pazzani, M.J. and Billsus, D. (2007) "Content-based recommendation systems," *The Adaptive Web*, pp. 325–341. Available at: https://doi.org/10.1007/978-3-540-72079-9_10.

Pearce, J.A. and Robinson, R.B. (2015) *Strategic management: Planning for domestic & global competition*. New York: McGraw-Hill.

Pereira, D. 2020. Netflix Business Model. Available at: https://businessmodelanalyst.com/netflix-business-model/ (Accessed: 03 January 2023).

Plummer, L. (2017) *This is how Netflix's top-secret recommendation system works*, *WIRED UK*. WIRED UK. Available at: https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like (Accessed: November 17, 2022).

Reddy, S.R.S. *et al.* (2018) "Content-based movie recommendation system using genre correlation," *Smart Intelligent Computing and Applications*, pp. 391–397. Available at: https://doi.org/10.1007/978-981-13-1927-3_42.

Resnick, P. and Varian, H.R. (1997) "Recommender Systems," *Communications of the ACM*, 40(3), pp. 56–58. Available at: https://doi.org/10.1145/245108.245121.

Ricci, F., Rokach, L. and Shapira, B. eds. 2015. Recommender Systems Handbook. *Recommender Systems Handbook* . Available at: http://dx.doi.org/10.1007/978-1-4899-7637-6.

Richwine, L. and Chmielewski, D. 2022. Walt Disney Has Finally Surpassed Netflix In Total Streaming Subscribers. Available at: https://washingtonindependent.com/w/walt-disney-has-finally-surpassed-netflix/ (Accessed: 23 January 2023).

Rivera, A. (2019) *Netflix VRIO/Vrin Analysis & Value Chain Analysis (resource-based view)*, *Rancord Society*. Available at: https://www.rancord.org/netflix-vrio-vrin-analysis-value-chain-analysis-resource-based-view (Accessed: October 18, 2022).

Sarwar, B. *et al.* (2001) “Item-based collaborative filtering recommendation algorithms,” *Proceedings of the tenth international conference on World Wide Web - WWW '01* (Preprint). Available at: https://doi.org/10.1145/371920.372071.

Sabga-Aboud, W. 2022. Why Netflix Will Likely Remain The King Of Streaming (NASDAQ:NFLX). Available at: https://seekingalpha.com/article/4488711-why-netflix-will-remain-the-king-of-streaming (Accessed: 23 January 2023).

Shani, G. and Gunawardana, A. (2010) “Evaluating recommendation systems,” *Recommender Systems Handbook*, pp. 257–297. Available at: https://doi.org/10.1007/978-0-387-85820-3_8.

Sherman, A. 2020. Netflix has replaced broadcast TV as the center of American culture — just look at the viewership numbers. Available at: https://www.cnbc.com/2020/04/21/netflix-massive-viewership-numbers-proves-it-owns-pop-culture.html (Accessed: 5 November 2023).

Slavin, K. 2011. How algorithms shape our world. Available at: https://www.ted.com/talks/kevin_slavin_how_algorithms_shape_our_world/transcript (Accessed: 23 January 2023).

Son, J. and Kim, S.B. (2017) “Content-based filtering for recommendation systems using multiattribute Networks,” *Expert Systems with Applications*, 89, pp. 404–412. Available at: https://doi.org/10.1016/j.eswa.2017.08.008.

Spangler, T. 2019. Netflix Original Content Outscores HBO, Hulu, Amazon on Customer-Satisfaction Survey. Available at: https://variety.com/2019/digital/news/netflix-original-content-consumer-survey-ratings-acsi-1203144795/ (Accessed: 23 January 2023).

Striphas, T. (2015) “Algorithmic culture,” *European Journal of Cultural Studies*, 18(4-5), pp. 395–412. Available at: https://doi.org/10.1177/1367549415577392.

Sull, D. 2015. Netflix’s “House of Cards” secrets: The real story behind Kevin Spacey and Frank Underwood’s meteoric ascent. Available at: https://www.salon.com/2015/04/26/netflixs_house_of_cards_secrets_the_real_story_behind_kevin_spacey_and_frank_underwoods_meteoric_ascent/ (Accessed: 23 January 2023).

Team, D.F. (2021) *Data Science at netflix - A must read case study for aspiring data scientists*, *DataFlair*. Available at: https://data-flair.training/blogs/data-science-at-netflix/ (Accessed: November 3, 2022).

Tingley, M. *et al.* (2016) *It’s all A/bout testing: The Netflix Experimentation Platform*, *It’s All A/Bout Testing: The Netflix Experimentation Platform*. Available at:

https://netflixtechblog.com/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15 (Accessed: November 17, 2022).

Tukey, J.W. (1988) *Exploratory Data Analysis*. Ann Arbor, MI: University Microfilms International.

University, O. 2020. How Netflix Uses Data to Pick Movies and Curate Content | Ohio University. Available at: https://onlinemasters.ohio.edu/blog/netflix-data/ (Accessed: 23 January 2023).

Vial, G. 2019. Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems* 28(2), pp. 118–144. Available at: http://dx.doi.org/10.1016/j.jsis.2019.01.003.

Wang, Y., Wang, M. and Xu, W. (2018) “A sentiment-enhanced hybrid recommender system for movie recommendation: A big data analytics framework,” *Wireless Communications and Mobile Computing*, 2018, pp. 1–9. Available at: https://doi.org/10.1155/2018/8263704.

Wayne, M.L. 2021. Netflix audience data, streaming industry discourse, and the emerging realities of ‘popular’ television. *Media, Culture & Society* 44(2), pp. 193–209. Available at: http://dx.doi.org/10.1177/01634437211022723.

Weinman, J. and Wiersema, F. (2015) *Digital disciplines: Attaining market leadership via the cloud, Big Data, social, mobile, and the internet of things*, *Wiley.com*. Available at: https://www.wiley.com/en-us/Digital+Disciplines%3A+Attaining+Market+Leadership+via+the+Cloud%2C+Big+Data%2C+Social%2C+Mobile%2C+and+the+Internet+of+Things-p-9781118995396 (Accessed: October 12, 2022).