**Module 7 Portfolio Milestone**

Brooke Dietrich

Final Research Paper - Netflix

MIS581: Capstone: Business Intelligence and Data Analytics

Dr. Steve Chung

May 7, 2023

**Abstract**

This research study examines the TV shows and movies that Netflix has available on its platform. Netflix is a company that specializes in providing TV and movie services all over the world. The company operates on a subscription basis where the user pays a monthly subscription fee. The user has access to hundreds of TV shows, movies, and Netflix original content, where new titles are added frequently as old ones are removed. Although Netflix has found a niche in the market, the market is becoming more saturated with competitors.

The author analyzed a Netflix dataset using various methods to determine the various types of movies and TV shows Netflix offers and how as a company they can better understand the content they provide. If Netflix can better understand the content they provide, then they can adapt their marketing strategy to help to increase revenue. The author utilized predictive and descriptive statistics including correlation and regression analysis to determine the relationships among variables within the dataset.

After analyzing the data, the author was able to draw conclusions. Netflix offers more than twice as many movies as TV shows on their platform with the majority of the content falling into the TV-MA rating category. The author also determined that there is little correlation between the release date of a TV show or movie and the date that Netflix adds the TV show or movie to their platform.

Netflix can use these results to better understand the content they provide, which would help them market their product to specific demographics. With the combination of streaming data, Netflix can use the authors findings to target specific demographic regions to increase customer satisfaction rates and also to target new potential customers.

**Final Research Paper - Netflix**

Netflix is a well-known company that specializes in providing television and movie streaming services all over the globe. “The service that established itself distributing films on DVD by mail in the United States is now mostly aptly categorized as a global video service” (Lobato et al., 2020). As streaming services become more popular in the market, it is important for Netflix to differentiate itself from its competitors. To differentiate itself, the company must understand customer wants and needs to maintain lasting relationships. The company must also continue to create unique content that will bring new customers to the platform.

Netflix operates on a subscription basis where the user pays a monthly subscription fee to use their services. The subscription fee varies based on the level of package the user purchases. Once a user purchases their subscription, they have access to an on-screen menu that allows them to browse through hundreds of TV shows, movies, and Netflix original content. The streaming platform offers a wide range of programming options, and new titles are added frequently while old ones are removed. They also offer content from a vast variety of genres such as comedy, romance, animation, action, and documentaries. Not only does Netflix offer content from other creators, but they also create their own original content. These original TV shows and movies can only be streamed from Netflix. The platform also allows users a download option at no additional cost. This means that users can download TV shows and movies onto a device, so they are watchable without Wi-Fi or internet.

Overall, the company has adapted its business strategy to meet the changing demands of the market. They have migrated from DVD rentals to a global streaming service. They provide a plethora of TV shows and movie options to viewers as well as original content that can only be streamed on Netflix. To ensure their continued company success, Netflix needs to adapt their strategy to stay ahead of their competition.

**Problem Statement**

Although Netflix has found a niche in the market and has continued to adapt to the changing technological advancements, they face the problem of competition. Netflix needs to determine how to stay ahead in the market by establishing long-term relationships with their customers. To determine market trends, Netflix needs to analyze data gathered on their streaming services to better understand how to market to their subscribers. Netflix also needs to look at data such as release dates and type of service, like a TV show or movie, to determine if these affect their overall viewership or subscription rates. The overall goal of the author is to determine whether Netflix can offer “more” to their customers for “less” by examining the relationships between different variables of Netflix content to see if they have an affect on one another and therefore affect Netflix’s profit and subscription rate.

**Goals and Objectives**

When COVID-19 swept over the world, the outlook on television and movies changed. People were stuck at home with nothing to do, but to watch TV. The goal of the author is to summarize the streaming data Netflix has available on their platform, so that they can make recommendations as to how Netflix can continue to build its brand and customer basis into the future.

The author will analyze a Netflix dataset to determine if viewers prefer TV shows or movies more than the other. The author will also determine the most popular release year for TV shows and movies. The author will also analyze the rating, duration of the TV shows and movies, and the different type of categories they fall into. The author will analyze the different variables and to determine if they have any correlation with one another. The goal is to understand the how Netflix can maintain subscriber relationships. If Netflix can better understand the types of TV shows and movies, they have available on their platform, they can better utilize data analytics to market these TV shows and movies to specific customers.

**Overview of Study**

If Netflix can better understand the products that they have available to their subscribers, then they can build brand loyalty and increase their customer basis in the future. “Most streaming platforms provide users with access to vast repositories of content with only a small fraction familiar to them” (Lamkhede et al., 2019). TV shows and movies are an integral part of our modern-day society, and if Netflix can continue to meet the needs of their customers, they will continue to keep a hold of their market share into the future. “Interestingly, even with fragmentation of the television audience across niche channels and the shift to widespread consumption of time-shifted content, viewers continue to derive a sense of participatory cultural citizenship from TV” (Matrix, 2014).

Advancements in technology now allow subscribers to watch TV shows and movies from anywhere. They can quickly and easily pull up a downloaded movie while on an airplane or while on a family vacation. “This motivates the market players to innovate, develop their products, and provide better service to its customers in order to survive the competition” (Oat, 2013). If Netflix can stay one step ahead of their competitors, they will stay successful. The author will analyze the relationship between the variables within Netflix’s streaming content to better understand the correlation between the variables. Understanding the relationships between the variables can give Netflix an advantage when marketing TV shows and movies to their subscribers. It allows the company to better understand their content and how it relates to their customers.

If Netflix can better understand the content they produce and provide, they can compare it against viewership data. This will ultimately allow the company to see how well the TV shows and movies they produce and provide are consumed by subscribers. This information will allow Netflix to target specific demographics and areas with specific content. Not only will this increase viewership, but it will also increase customer satisfaction, which will keep the customers coming back. This will also help Netflix with their marketing strategies. They can target specific groups of people with TV shows and movies to gain new subscribers. They can run promotions to bring on the new subscribers and then continue to offer unique content that the subscriber enjoys keeping them on as a satisfied customer.

Overall, Netflix needs to analyze their content data and viewership data to stay ahead of their competition. Understanding how they can best deliver content to their current customers and how to target new customers will help to increase the future number of customers and future revenue. Understanding how Netflix can best deliver content and what specific content will also benefit them when they are planning what new content to produce and advertise. The company will want to build a specific budget for building new content and then ultimately decide how they want to market the new content, so that it reaches the intended audience. This requires the company to adapt their business model to formulate a strategic understanding of their customers and the content they are providing.

**Research Questions and Hypotheses**

Before analyzing the data, the author must first create research questions and hypotheses. “Good business questions do not necessarily produce good research, but poorly conceived or constructed questions will likely create problems that affect all subsequent states of a study” (Agee, 2009). Therefore, it is important that the research questions give shape and direction to the business problem. When analyzing the Netflix dataset, it is important for the author to determine the importance of each of the variables. The author believes that the major variables will be the type, the date added and release year, the rating, and the category. These variables describe the TV shows and movies and help to show the company what is provided the most frequently on their platform. The research questions the author will aim to answer include: are more TV shows or movies provided on the platform?, is the date added dependent upon the release year of the TV show or movie?, and what rating do the most TV shows and movies fall into?

Hypothesis testing is a method that can be used to help answer business questions. There will be a null and alternative hypothesis for each business question. In other words, testing the null and alternative hypotheses helps to lead to conclusions that help to answer the business questions. The author asks the question of, are more TV shows or movies provided on the platform? The null and alternative hypotheses are as follows:

H**0**: The mean number of TV shows and movies provided on the platform are not equal

H**a**: The mean number of TV shows and movies provided on the platform are equal

The author also asks, is the date added to Netflix dependent upon the release year of the TV show or movie? The null and alternative hypotheses are as follows:

H**0**: The date added to Netflix does not correlate with the release year of the TV show or movie

H**a**: The date added to Netflix does correlate with the release year of the TV show or movie

The author also asks, what rating do the most TV shows and movies fall into? The null and alternative hypotheses are as follows:

H**0**: The mean number of TV shows and movies do not fall into the TV-14 rating

H**a**: The mean number of TV shows and movies fall into the TV-14 rating

If the company can gather intel into these research questions, they can better understand the products they have available on their platform. This will help them to market their TV shows and movies to specific customers to increase customer satisfaction, and it will also help them in content creation to produce successful shows on their platform.

**Literature Review**

The author analyzed other research on Netflix to provide more insight to their study. A major thing that allows user interaction with the Netflix interface to be seamless is the HTML5 technology. “This allows Netflix engineers to modify features on the user interface seamlessly, i.e. customers will not have to download new software” (Oat, 2013). By using this technology, Netflix can provide the latest technology and innovations to their products. This is why streaming technology, including Netflix, has become so prominent in the industry.

Netflix invests in exclusive content to keep its subscribers interested. Netflix also uses tactics such as “binge” watching. This means that they release all the episodes of a new series at once, so the subscriber could watch the entire show in one sitting if they wanted to. This tactic proves that Netflix relies on customer satisfaction rather than short-term profit that can be obtained by releasing one episode per week as viewers have to maintain their subscription longer to view all the episodes in a series. The author believes that they can utilize this studies information to appeal to the customer satisfaction side of Netflix. Netflix needs to better understand the products that they provide on their platform to continue to satisfy their customers.

Netflix also has unique features that allow a user to search on the platform. “User studies at Netflix have revealed three different mindsets in which members interact with Search, namely Fetch, Find and Explore, even though the aim is to watch something for entertainment” (Lamkhede et al., 2019). Fetch involves the users going to the search bar and typing in the exact TV show or movie they are looking for. Find involves a user who has an idea of what they want to watch, but they do not have a specific item in mind. They may search an actress who appears in several movies to select a movie she stars in. Explore involves broader searches such as a category like horror movies. The author believes this study is useful because it allows them to understand how Netflix utilizes the Search interaction. The author is curious if the user searches in the Search interaction can be tracked. This would allow Netflix to understand how to better market TV shows and movies to their subscribers. It would also allow Netflix to understand what the most popular items subscribers search to watch, and Netflix can look to better provide these on their platform.

Consumption of TV shows and movies today is a lot more convenient than it was years ago. Subscribers have 24/7 access to a wide catalogue of media products, which has resulted in a new era of binge-watching. “Binge watching refers to the practice of watching several episodes of the same television show in one sitting” (Perez, 2020). Netflix has adapted to this new era by adapting their platform to meet the needs of their subscribers. They have included the ‘skip intro’ button, which allows customers to automatically start playing the next episode. This study also analyzed Netflix’s usage of Twitter and Instagram for marketing, and the types of ads they posted and how responsive people were. The author believes that Netflix’s adaptation to rapidly changing technology will help to maintain Netflix’s success. Customers are always looking for faster and easier ways to gain content, and Netflix is providing this in a timely manner to its subscribers. Also, the analysis of social media usage for marketing purposes is interesting to the author. If Netflix wants to reach new customers, they need to target them with their ads. The author believes if Netflix better understands their content, they can make their marketing ads more effective.

When it comes to deciding what to watch, humans are naturally bad at making decisions. The study discusses how a Netflix subscriber would typically lose interest after about a minute of searching through titles, so Netflix set out to find a solution. “Our recommender system is not one algorithm, but rather a collection of different algorithms serving different use cases that come together to create the complete Netflix experience” (Gomez-Uribe et al., 2015). The Netflix recommendation algorithm can help users find popular shows or movies they may want to watch. The author believes the research done within this article ties nicely with their hypotheses. The author believes that Netflix can create an even more personalized experience and create more content that will keep users subscribed. On the other hand, Netflix can use this data to market specific content to users on social media or on other platforms.

The author discussed four case studies surrounding various topics involving Netflix. The author believes that each of these studies supports their research and can enhance their research. The author was sure to include several different aspects of the Netflix experience and how they are affected not only by the content they produce and viewership, but also by how and what they market on social media. The author also researched how Netflix’s platform and search can be improved to help the average subscriber.

**Research Design**

**Methodology**

The author collected the Netflix dataset from Kaggle.com. The Netflix dataset provides data on over eight thousand TV shows and movies that can be streamed on Netflix. The author will utilize a quantitative approach to analyze the dataset. Each TV show and movie is given a unique ID, so that it can be distinguished easily. This variable is considered a nominal variable. The type which describes whether the item is a TV show or movie is also provided. This variable called the ‘Type’ is a binary variable since there are only the two choices of TV show or movie. The title, director, cast, and country are also all given in the dataset, and these variables are all nominal variables. The date added onto Netflix is given, and these dates range from 2008 to 2021. Along with the data added, the release year of the TV show or movie is provided, and these dates range from 1925 to 2021. The date added along with the release year are both ordinal variables. The rating of the TV show or movie along with the duration are also provided, and this is also an ordinal variable. “Ordinal variables need present no special impediment to sound substantive research” (Winship et al., 1984). The category that the TV show or movie falls into is also given along with a description of the TV show or movie. Both are considered nominal variables. Overall, the dataset provides information on a list of TV shows and movies that Netflix subscribers have access to, and the author needs to analyze the relationships between the variables to understand how the company can use this information to maintain customer relationships as well as build new relationships.

**Methods**

The author plans to use various methods to analyze the dataset. Statistical tests are important when analyzing datasets as they can provide insight that otherwise would not be known. “Descriptive statistics are used to summarize data in an organized manner by describing the relationship between variables in a sample or population” (Kaur et al., 2018). The author will utilize SAS Studio to perform the summary statistics analysis on the Netflix dataset. The summary statistics will give an overall view of the variables the author chooses to analyze. The author will also use Jupyter Notebook to create visualizations of the frequency of variables to help to visualize the answers to the research questions.

Predictive analytics is an “analytics technique that uses statistical methodologies and forecasting to know what will likely to happen in the future” (Roy et al., 2022). The author will use SAS Studio to perform correlation analysis between the variables in the Netflix dataset. This analysis will determine if there is a high correlation between variables within the dataset, which can show the author how reliable the variables are on one another. The author will also utilize linear regression along with a t-test evaluation to analyze the relationship between the variables in the Netflix dataset. These tools and techniques will be used to portray the story that the dataset is telling. It will also allow viewers to visualize the data in a quick and easy manner, so they are able to make comparisons within the data to draw conclusions.

**Limitations**

Netflix is a large company that collects data on its customers as well as viewership data from its platform. Although this is the case, it is difficult to gain access to this data. The author believes their research could be taken one step further if they had access viewership data. Since the author does not have access to Netflix’s viewership data, they will make recommendations based on the data they do have, so that Netflix can apply the knowledge and research towards other data they may have.

The other limitation the data has is that it only provides data through 2021. It would be nice to have data through current day to see if trends are remaining the same. Although present day data is not available, the author believes 2021 is recent enough to make assumptions about today using this historical data.

**Ethical Considerations**

When analyzing and interpreting the dataset, the author needs to be conscious of the ethical considerations for the industry. First, the author needs to be sure the data is reliable and accurate because if they report information on TV shows and movies that is inaccurate, this could cause producers to become upset. This means the author needs to either keep anonymity and confidentiality or gain voluntary consent from those who the data is about. In this case, the collected data does not affect Netflix subscribers first-hand, but if the company wants to make changes within the company, it will affect their subscribers.

When dealing with personal and sensitive information, it is important to be upfront and transparent. Netflix currently uses algorithms within its platform to collect data on its customers. Netflix uses this data to suggest TV shows and movies to customers and to highlight popular TV shows and movies that are watched the most frequently. The goal of the author is to bring even further insight into the business to increase customer satisfaction and increase the future customer base. It is important when dealing with personal information to be sure it is secure and safe and that customers are made aware of how their personal information is being used. By providing transparency, the company can not only build customer trust, but they also shield themselves from legal issues.

**Findings**

The author wanted to analyze the Netflix dataset to determine if Netflix provides more TV shows or movies to its subscribers. Figure 1 below shows in the form of a bar graph that Netflix provides significantly more movies than TV shows.

Figure 1

*Frequency of Type Bar Graph*

Chart, bar chart

Description automatically generated

The author decided to run another test to provide a second visualization for the number of movies versus TV shows on Netflix. Figure 2 below shows the data in a pie chart. The author had previously stated the null hypothesis that the number of TV shows and movies provided on the platform are not equal. By analyzing Figure 1 and Figure 2, the author concludes that Netflix provides over two times the number of movies than TV shows. Therefore, the author will fail to reject the hypothesis that the mean number of TV shows and movies provided on the platform are not equal.

Figure 2

*Frequency of Type Pie Chart*

Chart, pie chart

Description automatically generated

Next, the author analyzed the ratings for the TV shows and movies in the dataset. The author created the visualization of the bar chart in Figure 3 below to show the rating with the highest average number of TV shows and movies. The author had previously stated the null hypothesis that the highest mean number of TV shows and movies do not fall into the TV-14 rating. According to the authors findings, the highest mean number of TV shows and movies fall into the TV-MA rating. Therefore, the author will fail to reject the hypothesis that the highest mean number of TV shows and movies do not fall into the TV-14 rating.

Figure 3

*Frequency of Ratings*

*Chart, histogram

Description automatically generated*

Next, the author analyzed the relationship between the release year and date added variables. The author analyzed correlation coefficients as well as R-Square values. The author had previously stated the null hypothesis that the date added to Netflix does not correlate with the release year of the TV show or movie. The correlation coefficient is the value of 0.17, which is shown in Figure 4 below. Since the value is close to zero, this indicates that little linear relationship exists between the two variables.

Figure 4

*Correlation between Release Year and Date Added in Python*

*Table

Description automatically generated with medium confidence*

The author also analyzed the R-Square value for the release year and date added variables. This is shown in Figure 5 below. The R-Square value of 0.3636 indicates that the independent variable (the release year) is not explaining much of the variation of the dependent variable (the date added). This shows the author that Netflix does not always add a TV show or movie to their platform the same year it is released, and therefore Netflix provides a lot of older content for its audiences.

Figure 5

*R-Square Value for Release Year versus Date Added*

*Table

Description automatically generated*

The author performed a t-test for the date added variable. The p-value of <.0001 is less than the alpha of 0.05, which means that the author will fail to reject the null hypothesis.

Figure 6

*T-Test for Date Added*

*Graphical user interface, application

Description automatically generated*

The author analyzed the R-Square value for the release year versus the rating of the TV show or movie. This is shown in Figure 6 below. The R-Square value of 0.1016 indicates that the independent variable does not explain much of the variation of the dependent variable.

Figure 7

*R-Square Value for Release Year vs Rating*

*Table

Description automatically generated*

These predictive statistics tests indicate that the two variables have little correlation between the two and therefore, the author will fail to reject the null hypothesis the date added to Netflix does not correlate with the release year of the TV show or movie**.**

**Conclusion**

The author analyzed a Netflix dataset to provide more insight on the products Netflix offers on its platform. The author believes if Netflix can better understand the TV shows and movies and the relationships between the variables, then they can better market them and understand what content to provide next. Capitalizing on marketing opportunities as well as new technological trends can help Netflix to stay ahead of the competition. Not only can Netflix gain new subscribers by offering unique content, but they can make sure the right people are watching the right content to maintain customer satisfaction in the long run. To do this, the author analyzed the variables in the Netflix dataset to determine their relationships. This will allow Netflix to use this data along with viewership data to bring more insight into how the company runs their marketing strategies and in turn bring in more customers and more revenue in the long run.

**Recommendations**

The author analyzed a Netflix dataset to provide more insight on the products the company offers on its platform. The author recommends further research on the topic. The author believes that Netflix can utilize the findings to determine how to adapt their marketing strategy into the future. Further analysis and questioning should be done by the Netflix team to determine the reasoning behind how they choose TV shows and movies for their platform. The author would like to know why Netflix provides more movies on their platform than TV shows. Viewership data would be beneficial to add to the authors research to answer some of these questions. If attention spans of viewers are getting shorter, then Netflix should offer more TV shows on their platform since their duration is shorter.

Overall, the author believes that Netflix can use the findings to not only change the content they have on their platform, but who they market it to. Demographic data would be useful to determine where most people who would view TV-MA rating TV shows and movies live. Since Netflix offers the most content with this rating, they should target this demographic specifically with subscriptions. The author also believes Netflix could survey their customers to gain feedback. From their analysis, the author determined that Netflix does not always add content the same year it is released. Netflix as a company should determine if subscribers prefer newer content or older content. This can be done to increase subscription rates as well as revenue, and the more that Netflix can understand about the content they have available on their platform, the better they can maintain customer satisfaction rates.

**References**

Agee, J. (2009). Developing Qualitative Research Questions: A Reflective Process. *International*

*Journal of Qualitative Studies in Education 431-47.*

Gomez-Uribe, C. & Hunt, N. (2015). The Netflix Recommender System: Algorithms, Business

Value, and Innovation. *ACM Transactions on Management Information Systems 1-19.*

Kaur, P., Stoltzful, J., & Yellapu, V. (2018). Descriptive Statistics. *International Journal of*

*Academic Medicine.*

Lakehead, S. & Das, S. (2019). Challenges in Search on Streaming Services: Netflix Case

Study. *In Proceedings of the 42nd International ACM SIGR Conference on Research and*

*Development in Information Retrieval 1371-1374.*

Lamkhede, S. & Das, S. (2019). Challenges in Search on Streaming Services: Netflix Case

Study. *Proceedings of the 42nd International ACM SIGIR Conference on Research and*

*Development Retrieval.*

Lobato, R. & Lotz, A. (2020). Imagining Global Video: The Challenge of Netflix. *JCMS:*

*Journal of Cinema and Media Studies 132-136.*

Matrix, S. (2014). The Netflix Effect: Teens, Bing Watching, and On-Demand Digital Marketing

Trends. *Jeunesse: Young People, Texts, Cultures 119-138.*

Oat, E. (2013). Analysis of Netflix Architecture and Business Model. *Dostupno*

Perez, Elisa Paz. (2020). Next Episode, Next Market: Social Media Marketing for SVOD: The

Case of Netflix. *GLOBAL*.

Roy, D., Srivastava, R., Jat, M., & Karaca, M.S. (2022). A Complete Overview of Analytics

Techniques: Descriptive, Predictive, and Prescriptive. *Decision Intelligence Analytics and*

*The Implementation of Strategic Business Management 15-30*.

Winship, C. & Mare, R.D. (1984). Regression Models with Ordinal Variables. *American*

*Sociological Review 512-525.*