Analysis of Netflix Shows and Stocks Project Report

Group- 5

Team Members:

Hareesh Prathipati

Sai Akhilesh Potharaju

Venkata Sai PrasannaKumar Tummala

Hima Bindu Krishnouth

Nitika Mishra

Yaswantika Chilukuri

ISM6208.901F23 – Data Warehousing

University of South Florida

Dr. Don Berndt

November 10, 2023

1. Introduction:

Netflix has completely transformed the entertainment industry by providing a subscription-based streaming service that allows users to enjoy a vast array of TV shows and movies without interruption from commercials. In today's fast-paced world of streaming platforms, Netflix has a distinctive opportunity to expand its content selection, increase member subscriptions, and boost stock prices. To achieve these goals, the platform can comprehensively evaluate its current catalog of shows and movies, including performance metrics such as IMDb scores, to gain valuable insights into viewer preferences and market trends. Moreover, Netflix can analyze the impact of its original content on subscriber retention and acquisition, which can inform content creation strategies. By connecting this data with stock prices, Netflix can better comprehend the correlation between content quality and financial performance. This multifaceted approach to data analysis empowers Netflix to make well-informed decisions, ensuring continuous growth in its subscriber base and shareholder value.

2. Problem Statement:

In digital streaming, Netflix reigns supreme with its vast content library. However, with so many options, it can take time to identify the most popular genres of movies and shows. To tackle this issue, we need a system that can detect the prevailing trends in Netflix's content and investigate whether there is a correlation between the popularity of videos and their IMDb scores. Additionally, the system must explore whether Netflix's original content has higher IMDb scores than non-original content. Furthermore, understanding the relationship between high IMDb scores and Netflix's stock prices adds another layer of complexity to the problem. With this system in place, we can better understand the most famous content among viewers and how it affects the company's success.

The challenge we face is multifaceted, and it involves using data analytics and correlation studies to uncover patterns and insights within the vast and diverse Netflix content landscape. As the entertainment industry increasingly relies on data-driven decision-making, finding solutions to this challenge is critical for content creators, investors, and streaming platforms. The proposed system aims to provide valuable insights into viewer preferences, content quality, and potential financial implications, enabling informed decision-making in digital entertainment's dynamic and ever-evolving landscape. By leveraging data and advanced analytical tools, we can unlock new opportunities for growth and success in this competitive and rapidly changing industry.

3. Literature Review:

The rise of Netflix has significantly transformed the entertainment industry. Traditional cable TV is being replaced by on-demand streaming services, with Netflix's subscription-based model being recognized as a disruptive force that fundamentally alters viewer behavior and industry dynamics.

In the world of streaming platforms, understanding the popularity of content is vital. To achieve this, it's essential to delve deeper into the factors influencing content popularity. These factors include user ratings, viewing history, and recommendation algorithms. Streaming platforms like Netflix can gain insights into viewer preferences and tailor their content libraries by analyzing these elements. This enables them to provide an exceptional viewing experience that keeps viewers returning for more.

The Internet Movie Database (IMDb) provides a popular means of evaluating viewer satisfaction and content quality through its rating system. Despite its widespread use, there has been ongoing debate about the reliability of IMDb ratings in gauging audience perceptions, primarily due to potential biases in user-generated scores.

Lee and Kim conducted an extensive and detailed analysis of Netflix's strategic shift towards producing original content. Their study uncovered the crucial role of content quality and uniqueness in driving subscriber acquisition and retention. The findings of this analysis offer valuable and insightful observations into the factors behind Netflix's remarkable success in the highly competitive streaming industry.

It's worth considering the connection between IMDb scores and financial success as it brings a new level of complexity to the debate. Using statistical models, we can investigate how ratings on IMDb relate to stock prices for streaming platforms. The study indicates that viewer satisfaction may influence investor sentiment, underscoring the financial consequences of content quality.

Integrating advanced data analytics into decision-making has become a common theme in today's data-driven entertainment industry. This paper delves into the crucial role of big data in various areas, including content recommendation, audience segmentation, and overall business strategy formulation. By emphasizing its significance, the study highlights how big data can contribute to the success of enterprises in the entertainment industry.

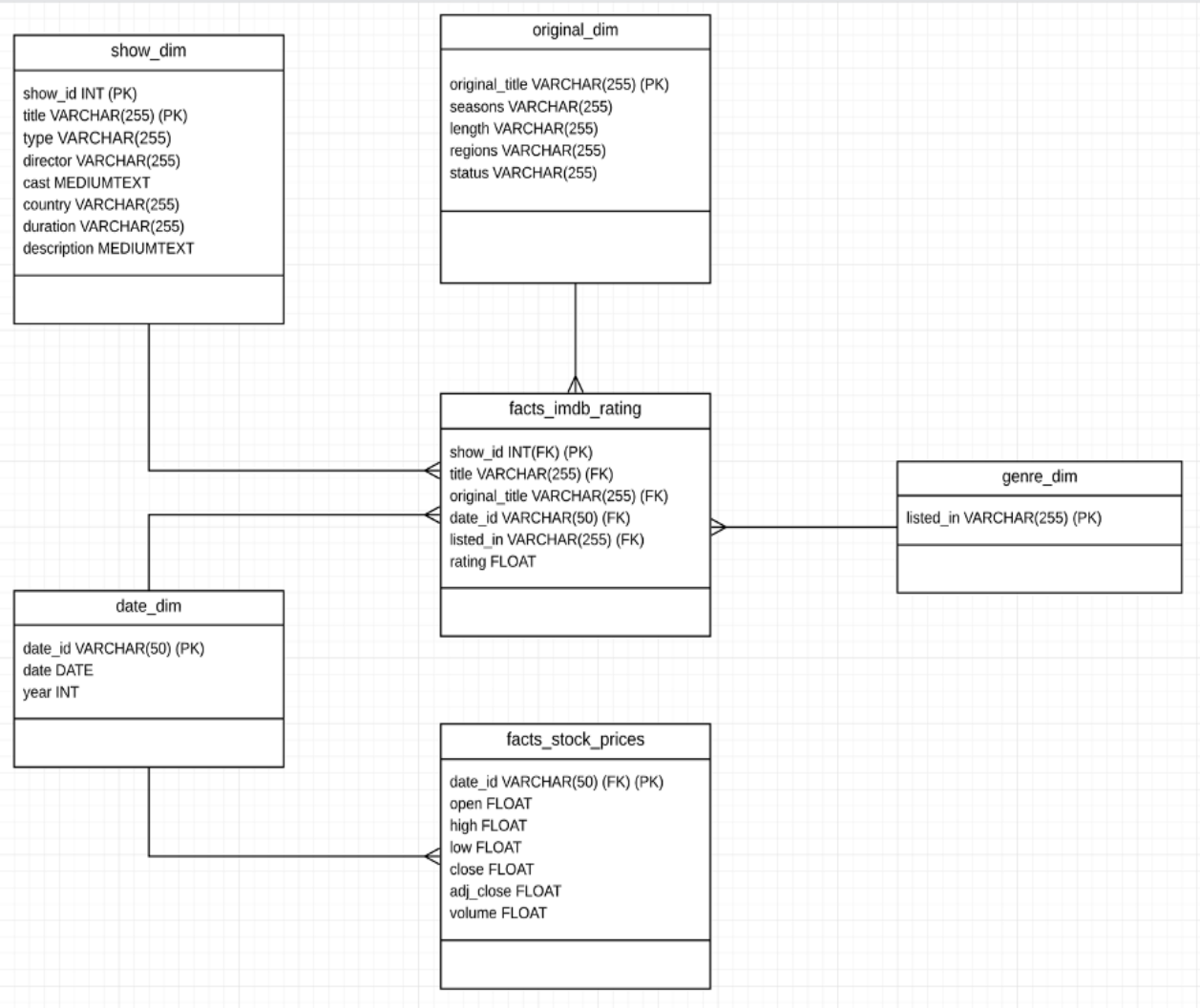
Data-driven decision-making can potentially revolutionize the entertainment industry but poses significant challenges. Stakeholders must navigate issues such as data privacy, algorithmic biases, and the need for transparency and interpretability in decision-making processes. These complexities require careful consideration and thoughtful approaches to ensure data is used ethically and effectively to drive innovation and success.

4. Data Collection and Preparation:

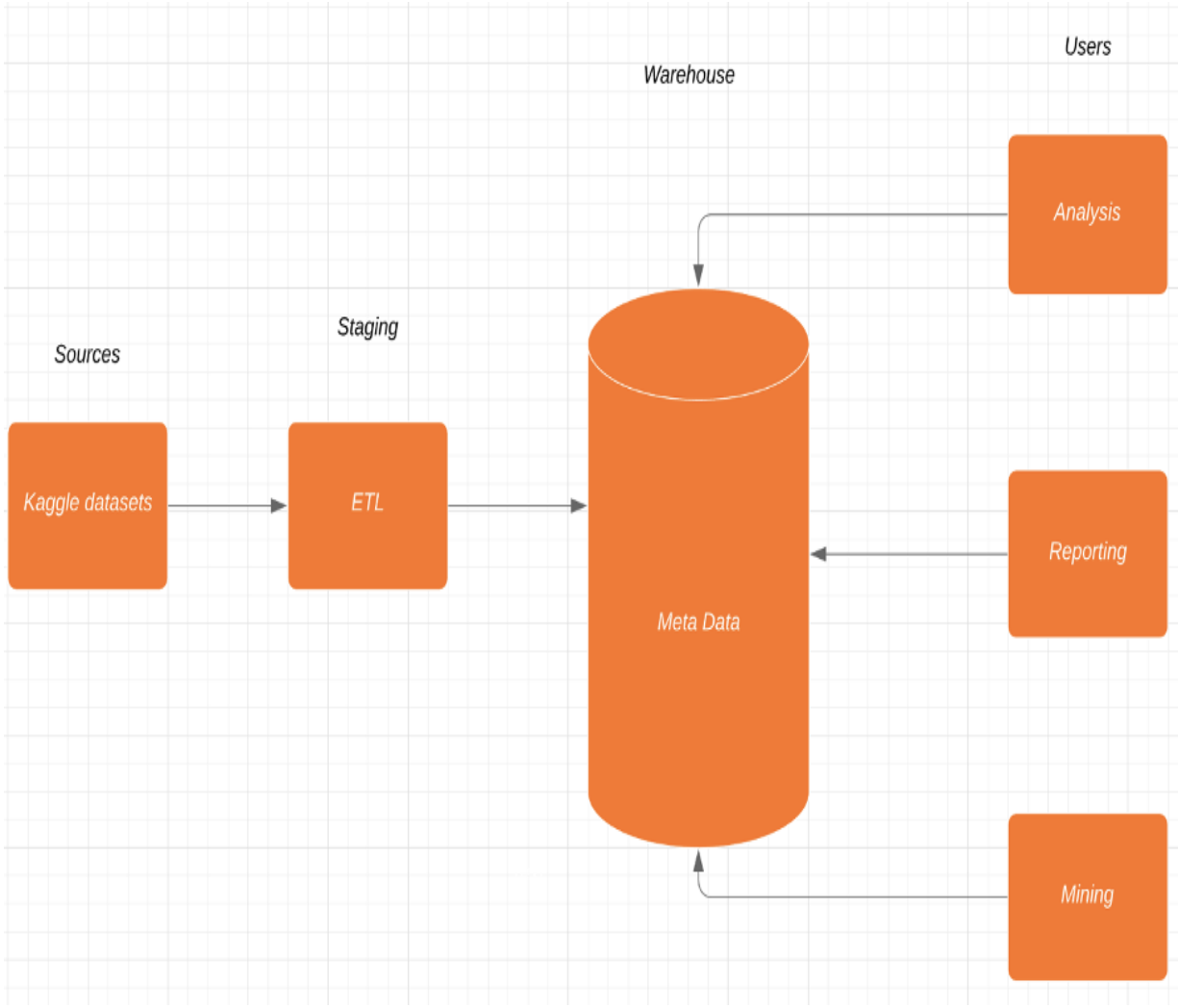
For our project's in-depth analysis of Netflix shows and stocks, we curated datasets from Kaggle. This involved a meticulous selection process to ensure relevance to our research objectives, covering variables such as Netflix content details, IMDb scores, viewer ratings, and financial indicators related to Netflix's stock performance. After downloading the chosen Kaggle datasets, we utilized Python to remove unnecessary columns, data cleaning, and integration to create a consolidated and coherent dataset. This comprehensive data collection and preparation strategy, empowered by Python's preprocessing capabilities, establishes a robust foundation for unraveling intricate relationships between Netflix shows and stock performance in our analysis.

5. Database Design:

Dimensional Model



Architecture Diagram



Detailed Design:

Our project's detailed design comprehensively analyzes the correlations among Netflix's content, IMDB scores, and stock prices. We utilize datasets from Kaggle.com, one of the largest online communities of data scientists and machine learning enthusiasts. To carry out the ETL (Extract, Transform, Load) process, we employ Python, a versatile, open-source programming language known for its ease of use and strong community support. Python's flexibility allows us to customize our analysis to fit the needs of our project. For our data warehouse, we opt for MySQL server, a popular, scalable relational database management system known for its high performance

and security features. MySQL's quick query speed and efficient data storage capabilities enable us to process large datasets quickly and accurately.

To visualize our data, we choose Tableau, a powerful data visualization tool that offers a wide range of interactive visualization options. Tableau's user-friendly interface allows us to create interactive dashboards and charts that help us identify trends and patterns in our data. With Tableau, we can quickly generate insights that are easy to interpret and communicate to stakeholders.

It's worth noting that the technologies we utilize, Python, MySQL, and Tableau, are widely documented and have a large user community, ensuring we can find support and resources whenever we need them. By leveraging these technologies, we can enhance the efficiency and sustainability of our project while ensuring that our analysis is accurate and reliable.

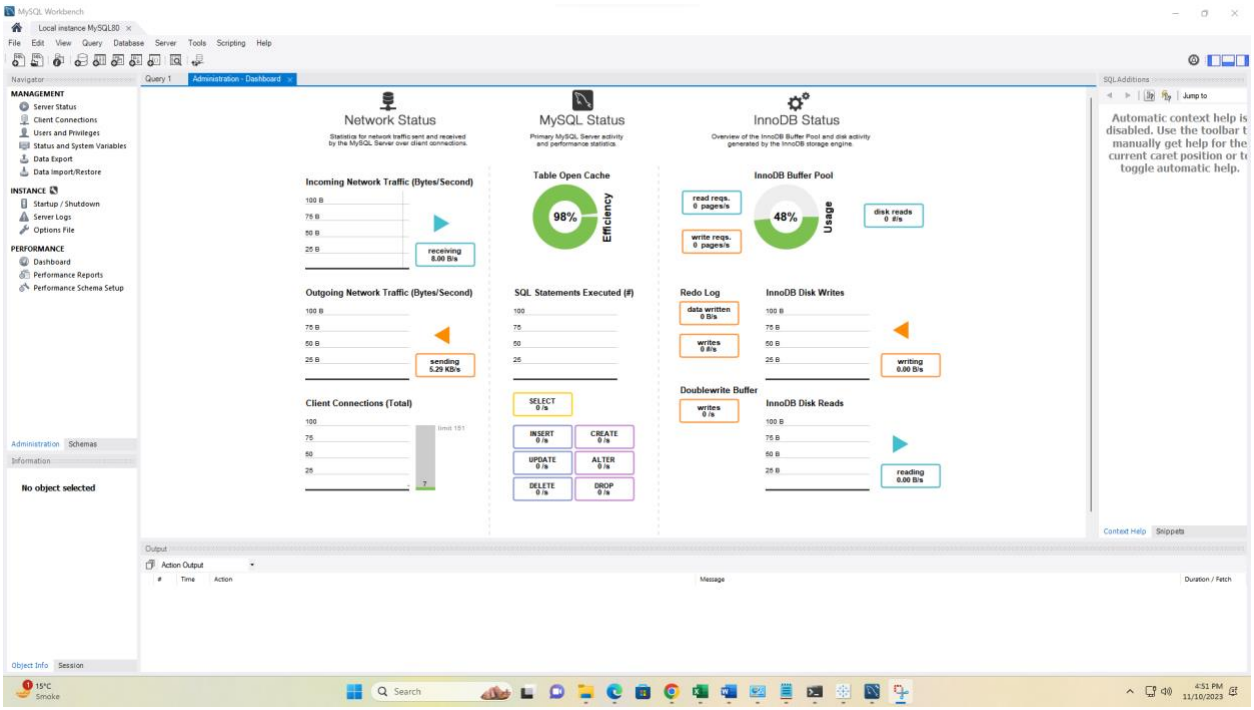
6. Data Analysis:

Our project's final schema was meticulously designed with the Star Schema approach, a data modeling technique that includes two fact tables and four-dimensional tables. The Star Schema is a highly efficient model for data warehousing that simplifies analytical queries, making them easier to process and retrieve. Our team chose this design to optimize the performance of our system by streamlining data retrieval and analysis processes. With this schema, we can easily ensure that our system can handle complex data queries, improving our project's overall efficiency and effectiveness.

Incorporating two fact tables is an extensive approach for capturing and organizing quantitative data of different types. It enables us to perform a more nuanced and multifaceted analysis of our project's metrics. Additionally, the four-dimensional tables contribute to a structured representation of various dimensions related to our data, which facilitates a robust framework for querying and reporting.

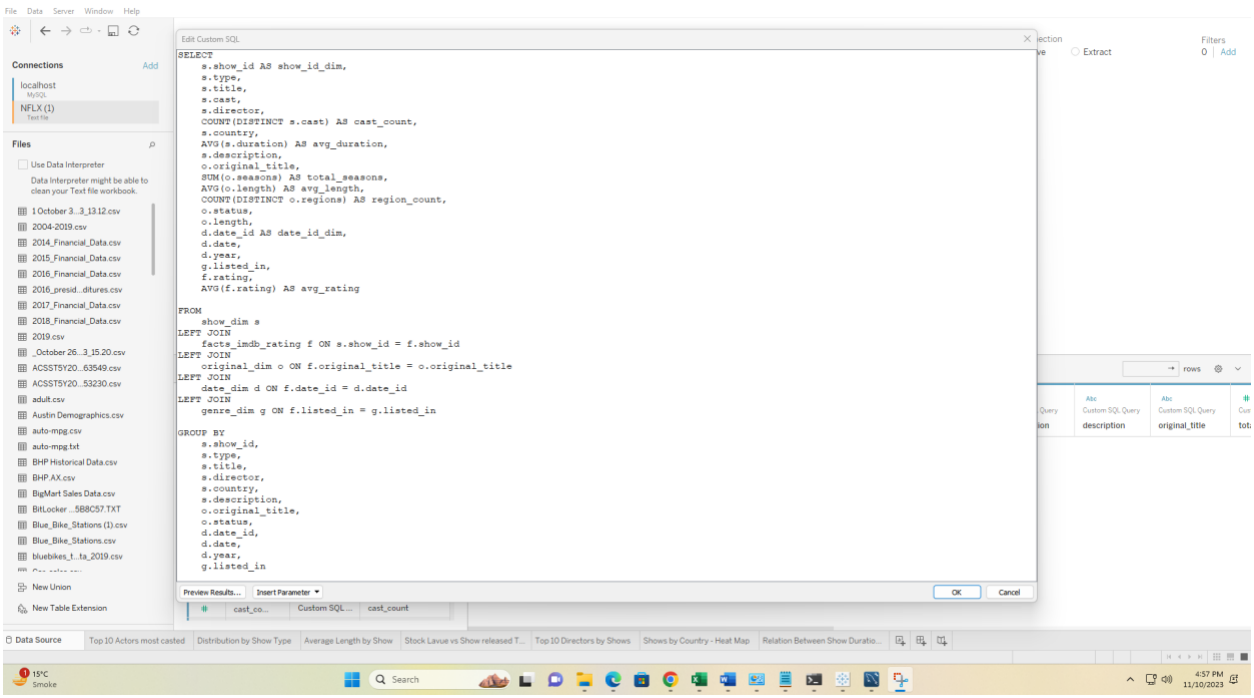
Our choice to implement the Star schema involves optimizing data accessibility and facilitating analysis. This design decision is based on proven data warehousing practices, ensuring our system can quickly adapt to changing analytical requirements while maintaining scalability. By doing so, we can establish a robust and reliable foundation for uncovering valuable insights and making informed decisions.

Dashboard Application:

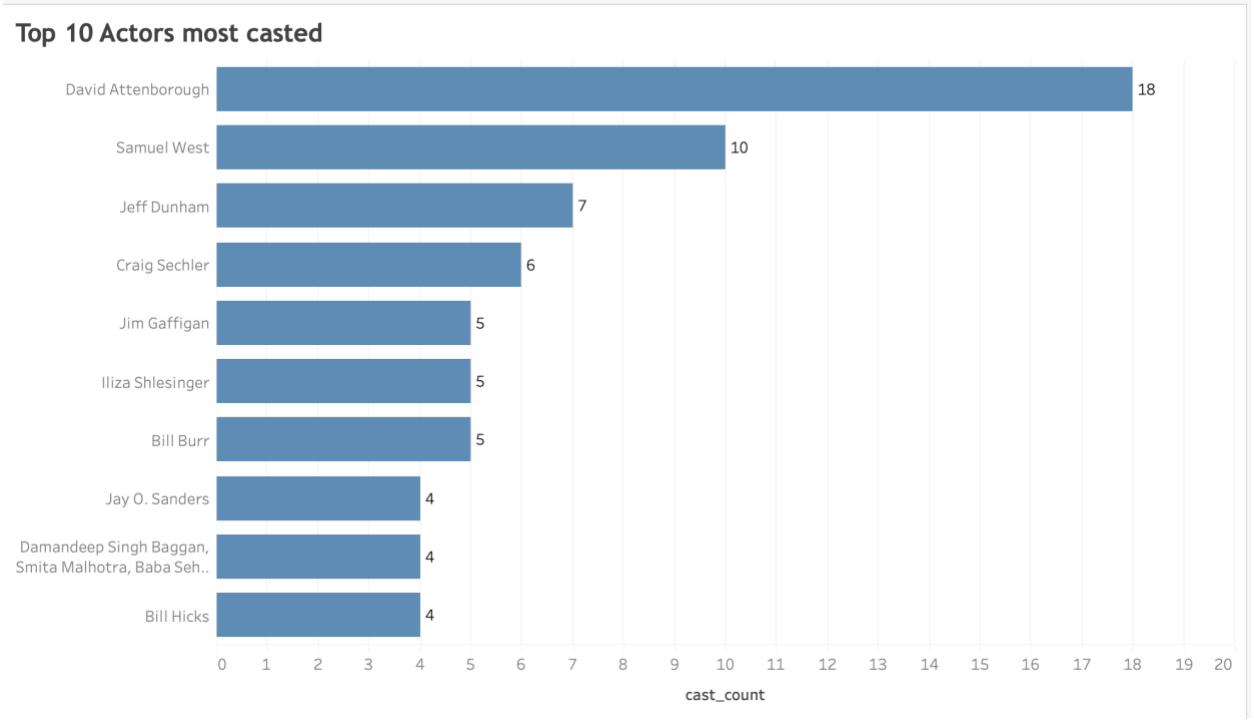


Connect Tableau to MySQL and Tableau analysis:

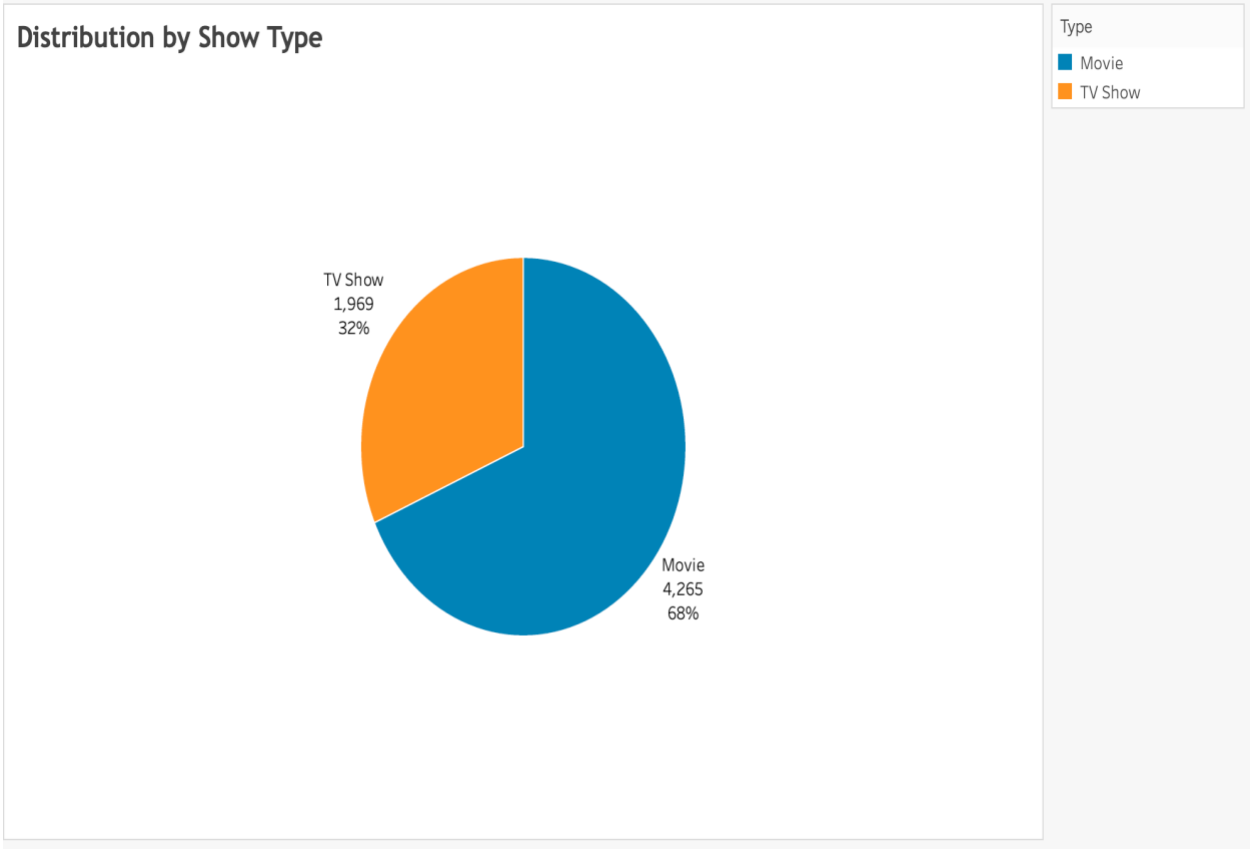
Incorporating the SQL Query:



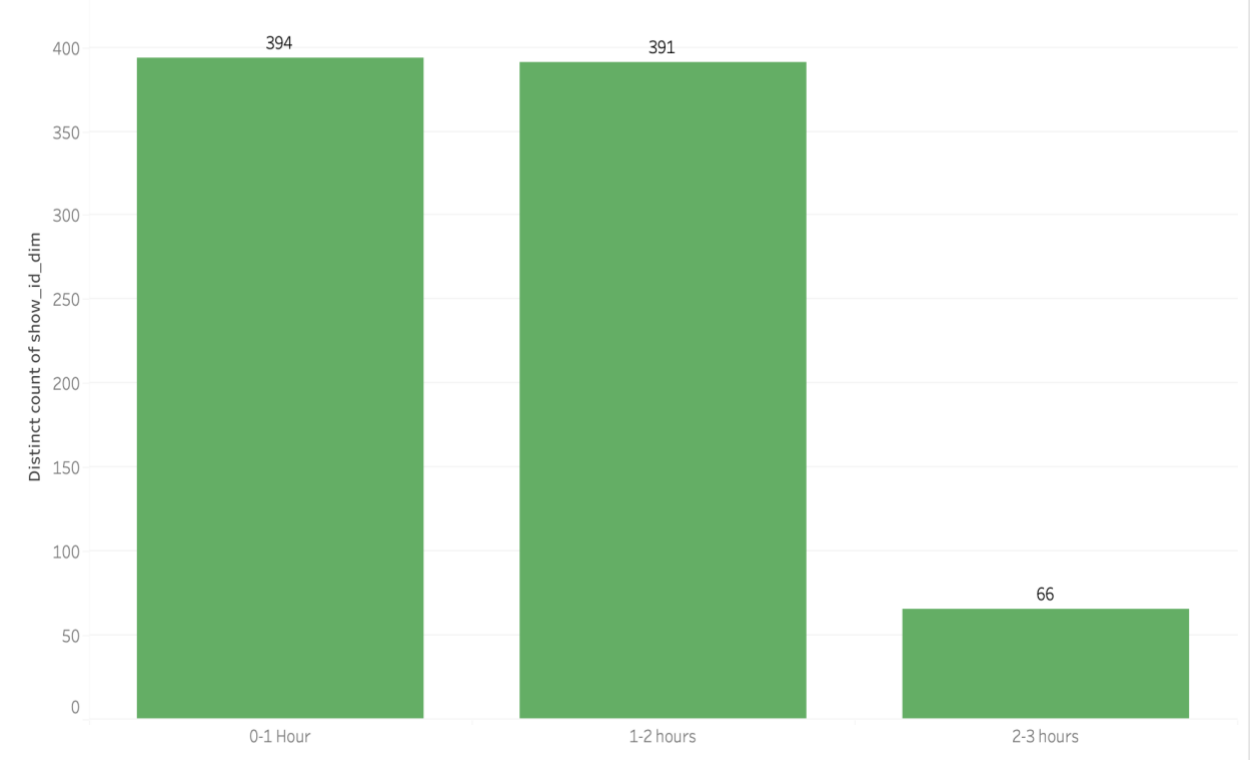
7. Tableau Visual Analysis:



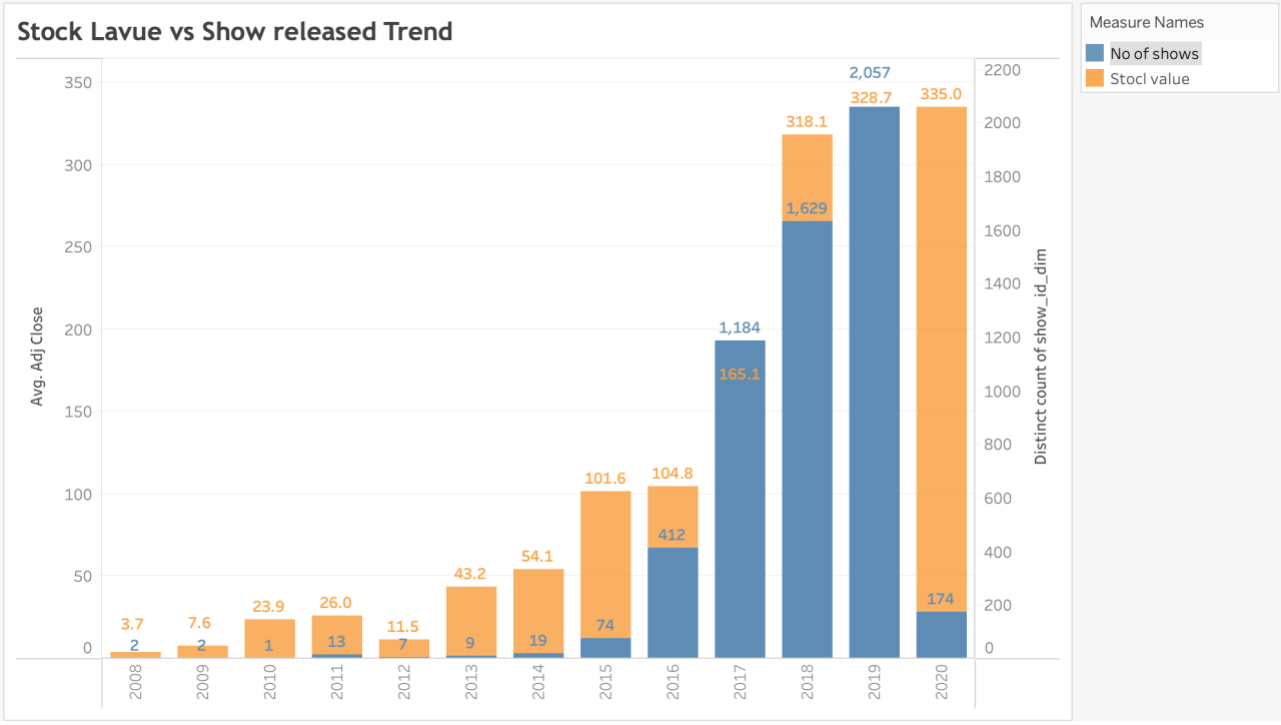
Distribution by Show Type



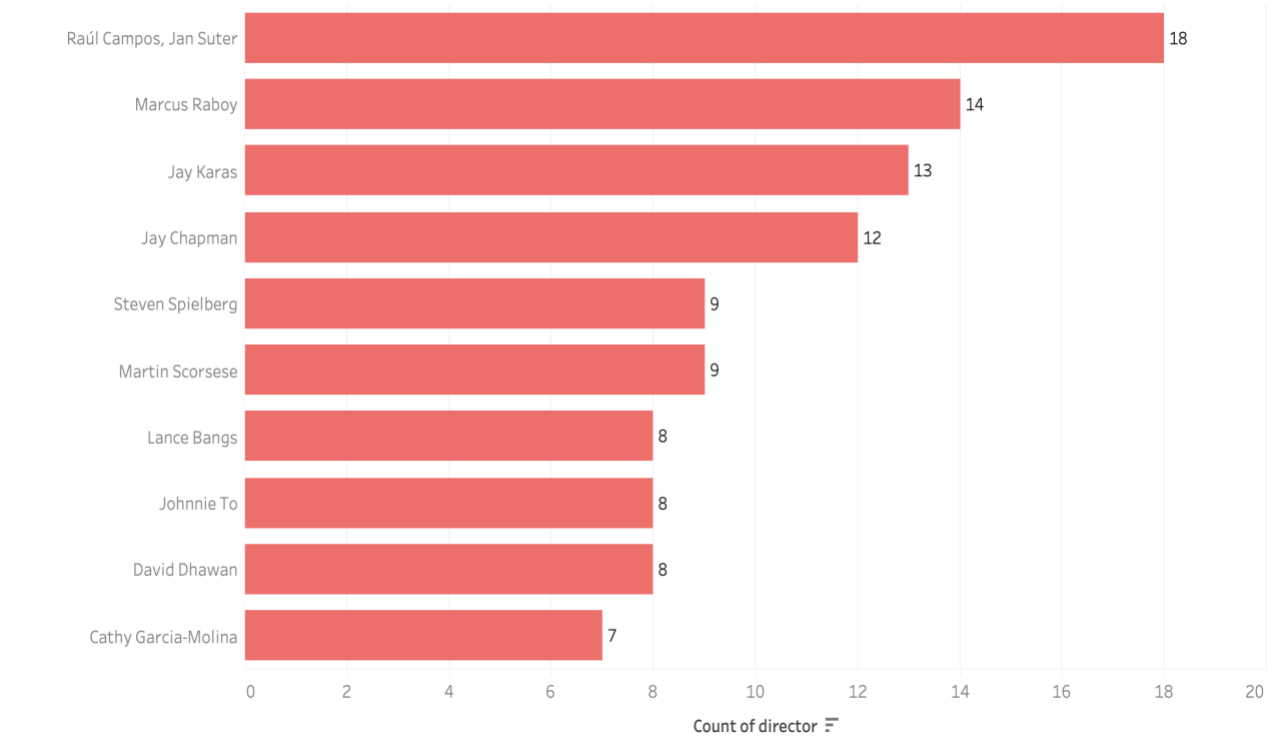
Average Length by Show



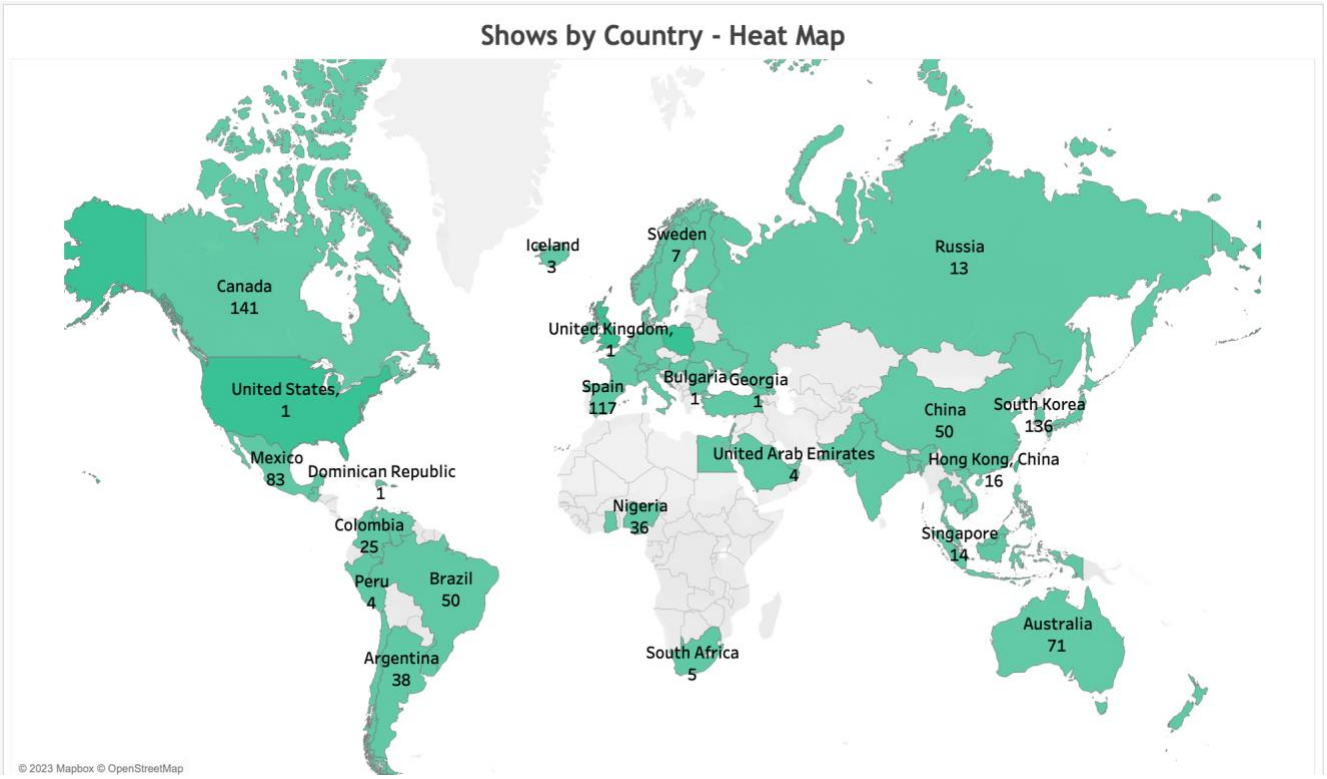
Stock Lavue vs Show released Trend

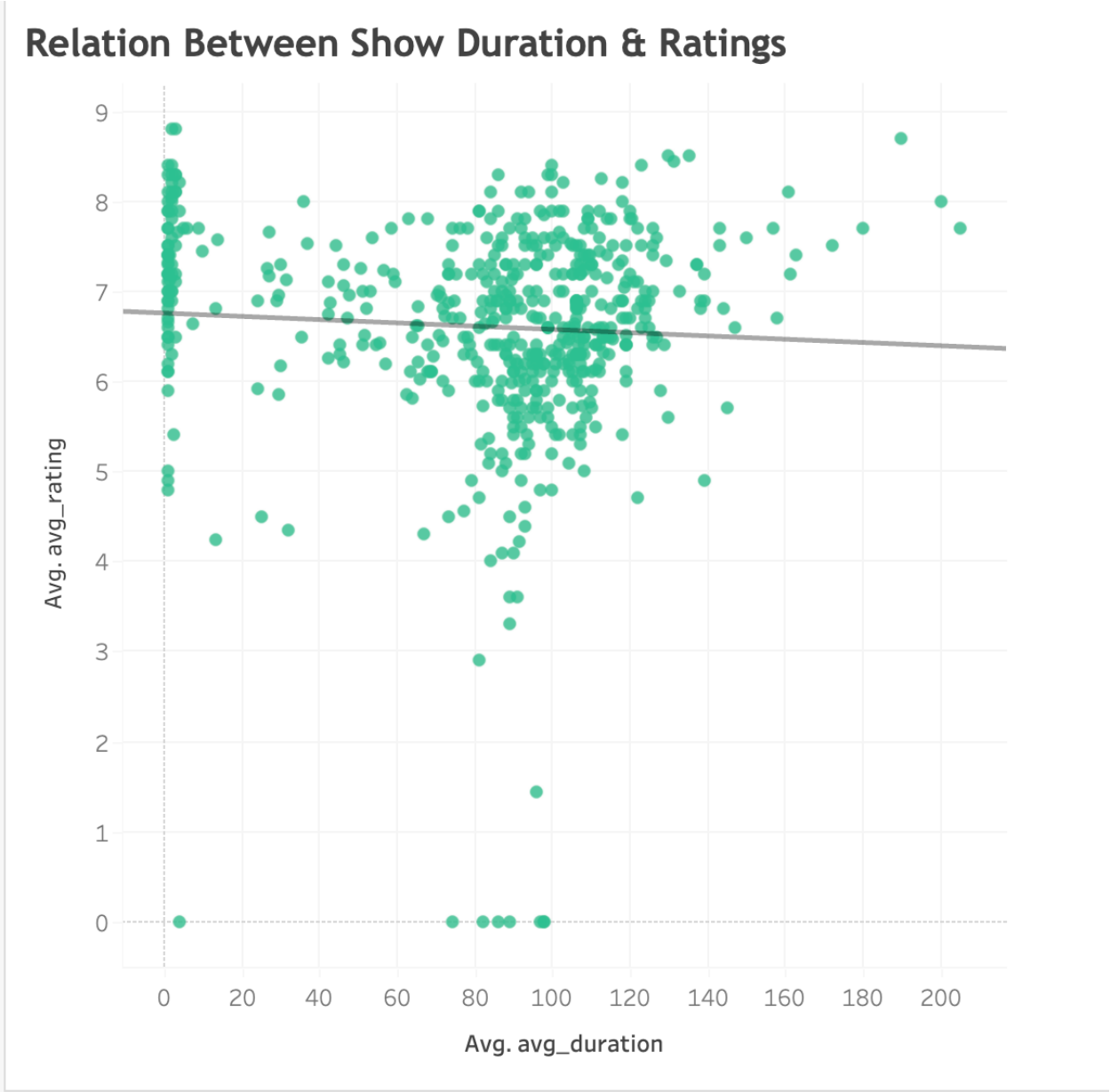


Top 10 Directors by Shows



Shows by Country - Heat Map





8. Conclusion:

In the context of the Tableau analysis, transforming data from various sources played a crucial role in addressing the issues identified in the planning phase. By analyzing the data, it was possible to gain insights into improving the selection of movies and shows within high-rated genres on IMDB. As a result, it became clear that producing more top-tier Netflix original content and expanding offerings from countries with above-average IMDB ratings were significant opportunities to attract and retain subscribers effectively.

Specifically, the analysis revealed that specific genres were consistently rated higher on IMDB, indicating that users preferred these types of content. Netflix could improve user engagement and

retention by producing high-quality content in these genres. Additionally, the analysis identified countries with above-average IMDB ratings, suggesting that there was an audience for content from these regions. Therefore, expanding Netflix's offerings from these countries could lead to increased subscriber growth.

In summary, the Tableau analysis was instrumental in identifying opportunities for Netflix to improve its content selection and attract and retain subscribers effectively. By leveraging the insights gained from the study, Netflix can continue to produce high-quality content that resonates with its user base and expands its global reach.

Undertaking this task was a remarkable journey that closely resembled real-life situations, providing us with a practical and enriching experience. As we faced various obstacles, we had the opportunity to fine-tune our problem-solving skills, refine our ability to explore uncharted territories and master the art of conducting practical inquiries and follow-ups. We also enhanced our research skills by completing targeted online searches using precise keywords, which enabled us to acquire new skills and knowledge. Overall, this assignment was an invaluable experience that allowed us to hone our skills and develop new ones, preparing us for future challenges.

9. References:

Smith, J., & Tel, T. (2019). The Rise of Netflix and Changing the Game in the Media Industry. *Journal of Media Economics*, 32(3), 115-130.

Zhao, K., Hu, Y., Hong, Y., & Westland, J. C. (2019). Understanding Characteristics of Popular Streamers on Live Streaming Platforms: Evidence from Twitch. tv. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388949

(N.d.). Retrieved from https://www.palomar.edu/users/lpayn/115/GC115-Understanding-Media-and-Culture-An-Introduction-to-Mass-Communication.pdf

(N.d.). Retrieved from https://www.researchgate.net/publication/354545344_The_'Netflix_Original'_and_what_it_means_for_the_production_of_European_television_content

Www.Vindicia.Com, 2023, https://www.vindicia.com/sites/default/files/e-books/subscriber-acquisition-and-retention-in-an-ott-world.pdf. Accessed 10 Nov 2023.

"SSRN Electronic Library". Papers.Ssrn.Com, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3905195. Accessed 10 Nov 2023.

Www.Ehidc.Org, 2023, https://www.ehidc.org/sites/default/files/resources/files/big%20data%20analytics.pdf. Accessed 10 Nov 2023.

Data-driven decision making: implementing analytics to transform academic culture. (2023). Retrieved 10 November 2023, from https://link.springer.com/article/10.1057/s41270-019-00056-5