Mihir Shekhar 21355037 **CS7DS4-202324 DATA VISUALISATION**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Introduction

This report presents an analysis of the Netflix dataset, a comprehensive collection of titles available on the streaming platform. Sourced from Kaggle, it offers a glimpse into the dynamic world of digital entertainment. Using Tableau, this study aims to unveil trends and patterns in Netflix's content distribution, exploring genre diversity, directorial influence, and geographical spread. The insights derived from this analysis seek to provide a deeper understanding of Netflix's strategy in adapting to the evolving tastes and preferences of a global audience.

Dataset and Background

The Netflix dataset from Kaggle contains a diverse array of movies and TV shows, offering a snapshot of the streaming service's vast library. It includes various elements like title, director, cast, country, release year, and genre, providing a rich foundation for analysis. This dataset reflects the global nature of Netflix's content, encompassing a wide range of genres and productions from numerous countries. It serves as a crucial resource for understanding the streaming giant's approach to content curation and distribution. By analyzing this data, we can gain insights into the trends that shape Netflix's offerings, its response to viewer preferences, and its positioning in the competitive streaming market.

Dataset Pre-processing

Before analysis, the Netflix dataset underwent essential pre-processing. This included cleaning missing values and standardizing entries for consistency, particularly in fields like country and genre. Categories were streamlined to simplify analysis, ensuring uniformity in how different types of content are represented. These steps were vital to ensure the accuracy and reliability of the data, setting a solid foundation for a meaningful and insightful exploration of Netflix's content trends.

Data Types

The Netflix dataset includes a mix of data types, enhancing the depth of analysis. Categorical data such as 'type' (movie or TV show), 'genre', and 'country' allows for segmenting content and understanding distribution patterns. Numerical data like 'release year' and 'duration' enable trend analysis over time and comparative study between distinct categories. This combination of

categorical and numerical data offers a comprehensive view, crucial for a detailed exploration of Netflix's content strategy, audience preferences, and market positioning.

Tools and Technologies Used

Tableau, a powerful data visualization tool, was the primary technology used for this analysis. Its user-friendly interface and advanced visual capabilities made it ideal for exploring the Netflix dataset. Tableau's ability to handle large datasets, coupled with its diverse range of visualization options, facilitated a detailed and interactive examination of content trends. The tool's features like dynamic filtering, custom visualizations, and interactive dashboards were crucial in presenting complex data in an accessible and engaging manner.

Tasks Supported by the Visualization

The Tableau dashboard was designed to facilitate various analytical tasks, enhancing the understanding of Netflix's content strategy. It allows for trend analysis, tracking how content types and genres have evolved over time. Comparative analysis is another key task, enabling a side-by-side comparison of different genres, directors, and countries, revealing patterns in content production and popularity. Geographical analysis through the map visualization highlights the global distribution of titles, showcasing Netflix's international reach. The dashboard also supports directorial impact analysis, illustrating the influence of different directors on the platform's content. Lastly, the rating distribution analysis sheds light on the target audience demographics and content suitability. These tasks, enabled by Tableau's interactive visualizations, provide a comprehensive view of the dataset, uncovering significant insights into the dynamics of Netflix's content.

Encoding Channels and Idioms

- **Donut Chart - Content Type Distribution**: This chart’s color contrast simplifies distinguishing between Movies and TV Shows, enabling immediate visual grasp of the content composition. The donut chart was chosen over a pie chart for its modern look and to better center the text that indicates the total number of titles.
- **World Map - Geographic Distribution**: Circle sizes reflect the number of titles to offer visual cues about Netflix's market penetration in different countries. We used a dark map background to ensure that the circles stand out, facilitating quick geographical comparisons.
- **Bar Chart - Yearly Content Addition**: Bar height illustrates the number of titles added each year, with the aim to demonstrate growth trends at a glance. Bars were selected for their familiar representation of time-series data, allowing intuitive year-over-year comparisons.
- **Bubble Chart - Genre Popularity**: Bubble size indicates genre frequency, chosen for its visual impact and ease of comparison. The placement and color variations add depth to the analysis, guiding the eye towards the most prevalent genres.
- **Horizontal Bar Chart - Directorial Influence**: The bar length for each director provides a clear, comparative measure of output. Horizontal bars were utilized to handle long director names and to facilitate easy reading from left to right, aligning with Western reading patterns.
- **Treemap - Content by Rating**: The use of area size to represent the number of titles per rating category allows for efficient space usage and the grouping of similar content, making it an effective way to communicate content distribution.

Discussion about the novelty of the visualization

The dynamic elements of the Tableau dashboard significantly enrich the data analysis experience. Interactive filters enable users to select specific genres, time periods, or countries, allowing for a customized view of the data. Hover-over tooltips provide additional details on specific data points, offering deeper insights. The ability to interact with the data in real-time enhances user engagement, making the dashboard not just a tool for displaying information but a platform for discovery and exploration. These interactive features transform the data into a more accessible and engaging narrative.

Handling of Complexity

The visualization techniques employed in the Tableau dashboard effectively managed the complexity of the Netflix dataset. By segmenting data into distinct visual components like maps, bar charts, and bubble charts, it made complex data patterns more understandable and accessible. The dashboard's design emphasized clarity, allowing users to easily navigate through various layers of information. This approach not only made the insights more digestible but also catered to users with various levels of data literacy, ensuring a broad appeal and comprehension.

Analysis of Strengths and Weaknesses

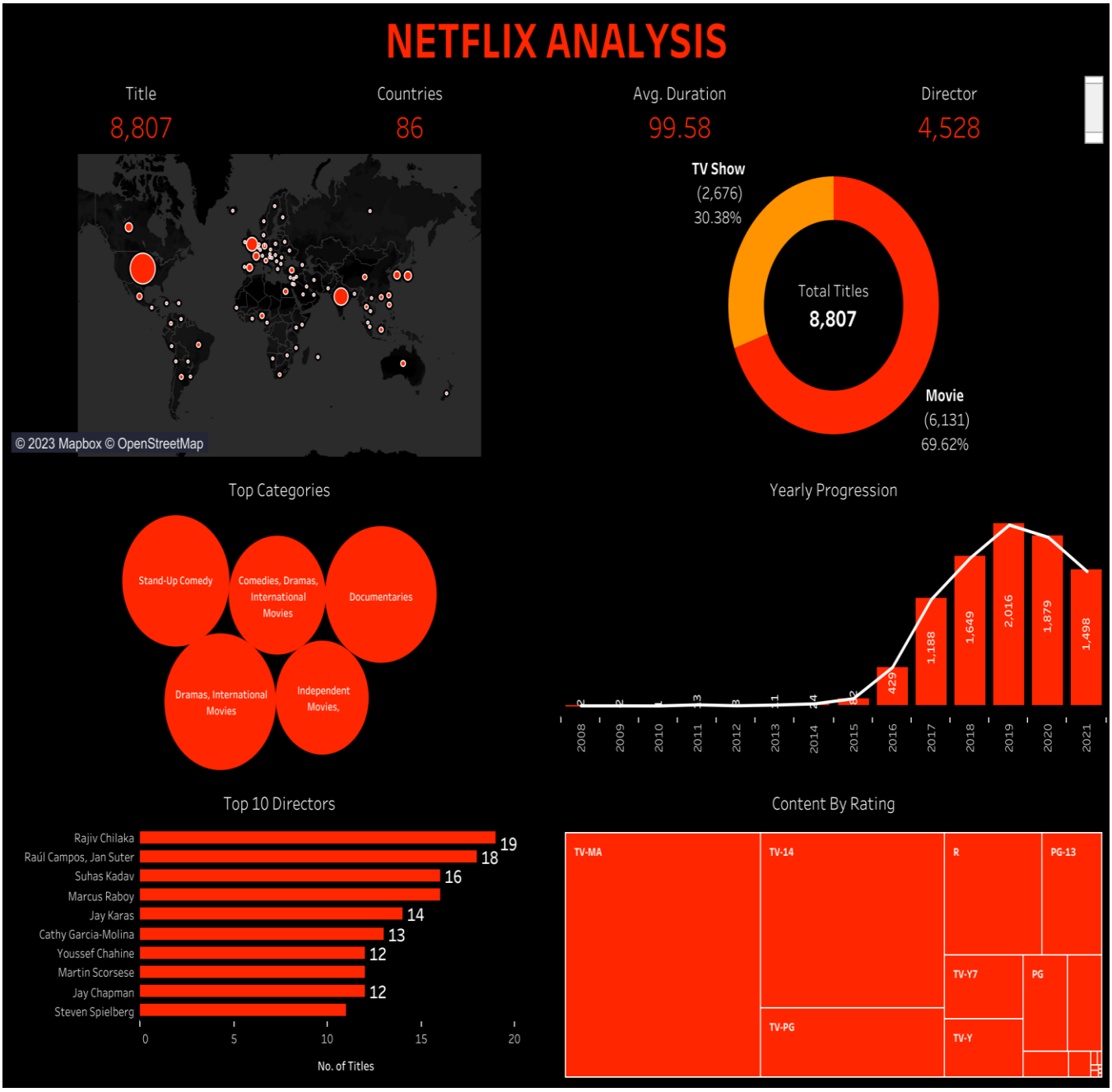
The Tableau dashboard excels in presenting Netflix's dataset with clarity and interactivity, making complex data easily accessible. Its strengths lie in the intuitive design and interactive elements, which engage users and facilitate a deeper understanding of the data. The diverse visualization techniques effectively highlight distinct aspects of the dataset, such as content distribution and genre popularity.

However, there are areas for improvement. The dashboard primarily focuses on descriptive analytics, offering limited predictive insights or deeper analysis into potential correlations and trends. Additionally, while it successfully presents surface-level data, it could benefit from incorporating more advanced analytical features, like statistical analysis or trend prediction, to provide a more comprehensive view of the data.

Conclusion

This analysis of Netflix's content using Tableau has illuminated the intricate patterns and trends within the streaming giant's expansive library. It underscores the diverse, global nature of Netflix's offerings and its adaptive content strategy. While the visualizations effectively demystify complex data sets, offering clear and engaging insights, there is room for deeper analytical exploration. Future enhancements could include predictive modelling and advanced correlation analysis. Overall, this project highlights the power of visual analytics in extracting meaningful stories from data, providing valuable insights into the preferences and behaviours of the streaming audience.

Dashboard Image:



References:

- "Netflix Titles Dataset." Kaggle. This dataset provides a comprehensive list of movies and TV shows available on Netflix, which was the primary data source for this analysis. [https://www.kaggle.com/shivamb/netflix-shows]
- Tableau Software. The main tool used for data visualization in this project. The official website offers detailed information about the software's capabilities and features. [https://www.tableau.com/]
- Relevant course materials from CS7DS4, including lectures and readings on data visualization techniques and best practices.
- "Tableau Help Documentation." This resource provided guidance on utilizing various features and functionalities of Tableau for effective data visualization. [https://help.tableau.com/current/pro/desktop/en-us/default.htm]
- OpenAI's ChatGPT.[https://openai.com/chatgpt]

Link for video presentation:

https://drive.google.com/drive/folders/1asYT4mp3GFq-jfz06rYB3-ii5NIT3WPa?usp=drive_link

Link for the Tableau Visualization Dashboard:

https://public.tableau.com/views/NetflixAnalysisDashboard_17029321088180/NetflixDashboard?:language=en-US&:display_count=n&:origin=viz_share_link

Dataset Link:

https://www.kaggle.com/shivamb/netflix-shows