SCHOOL OF INFORMATION SYSTEMS AND COMMUNICATION TECHNOLOGY

DA4310

DATA SCIENCE

ASSIGNMENT 01

DASHBOARD APPLICATION

MODULE LECTURER

NORFARRAH MOHD MASDI

SEMESTER 04

ACADEMIC SESSION 2023/2024

SUBMITTED BY

AWANGKU MUHAMMAD YAMIN BIN PENGIRAN IBNU [22FTT1344]

PROGRAM

DIPLOMA IN DATA ANALYTICS (DDAS02)

SUBMISSION DATE

26TH FEBRUARY 2024

TABLE OF CONTENT

Introduction

One of the most widely used services for streaming movies and videos is Netflix. By mid-2021, they had over 200 million subscribers worldwide and over 8000 films and TV series available on their website. This dataset includes directories, ratings, release year, runtime, and other information for every movie and television programme that is accessible on Netflix.(ARIYO, n.d.).

By using Python, streamlit, and additional libraries can be used to build a shared visualization application. In this instance, the customer requested assistance in visualizing their business, looking for hidden patterns, and gaining a deeper understanding of the dataset, such as the top director across nations.

Problem Statement

To have the ability to create a dashboard application. Interactive and insightful visualizations should be included in the application. Before utilizing the data for the visualizations, data scientists must preprocess it. Data scientists also have to think about how the application is designed visually. Finally, writing an application user manual is another task assigned to data scientists.

Aims and Objectives

The aim of this project is to be able to create a streamlit application for visualizing the netflix dataset, also able to provide users with interactive insights into content available on netflix.

However the objectives of this project:

1. Analyzing Trends among Attributes.
2. Identify Popular on each Attribute.
3. Geospatial Analysis of Content Availability.

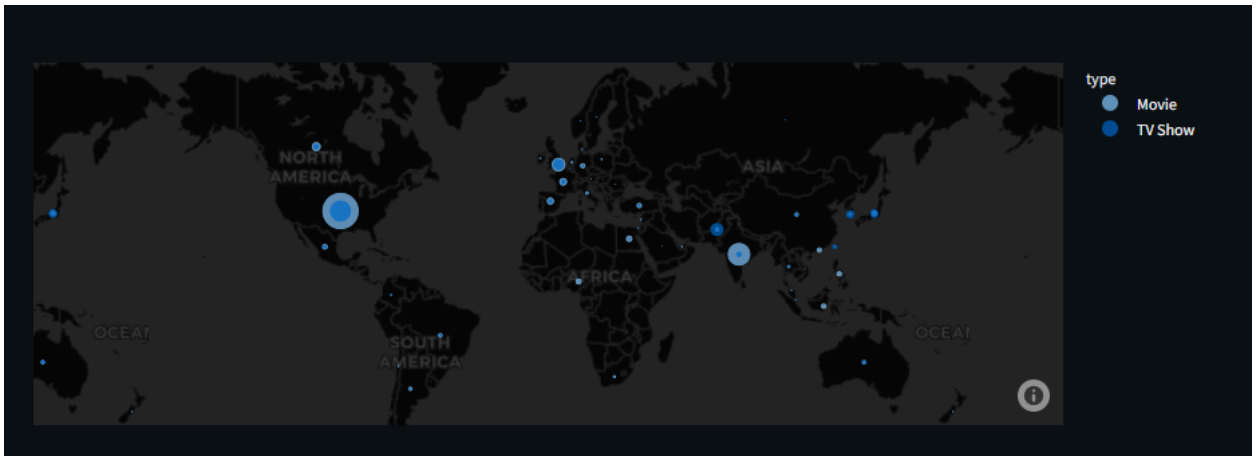
Application Overview

Streamlit is a simple, easy-to-use online application designed for interactive Netflix dataset exploration. This dataset allows users to conduct geospatial studies of popular director identification and analysis, as well as content availability and distribution by country. With the use of simple filters and interactive visualizations, users may quickly identify patterns and insights in the collection. They can now use the available data to inform their decisions and conclusions. All things considered, the app provides a quick and easy way to peruse Netflix's collection while picking up useful knowledge with only a few clicks.

Main Dashboard

Within this Main Dashboard, there a 7 visualization that fit in this dashboard, each offering unique perspectives, the first one is gonna be main dashboard and the second dashboard gonna be second dashboard:-

Occurrences of Types in Different Countries



(Fig 1.0:Scatter Plot Map)

This visualization shows a map with bubbles that represent various countries, scaled based on the prevalence of various content kinds in each nation. Every bubble in the dataset has a color code that corresponds to a certain sort of content. For example in this visualization, User can see that America has the largest bubble instead of among other countries.

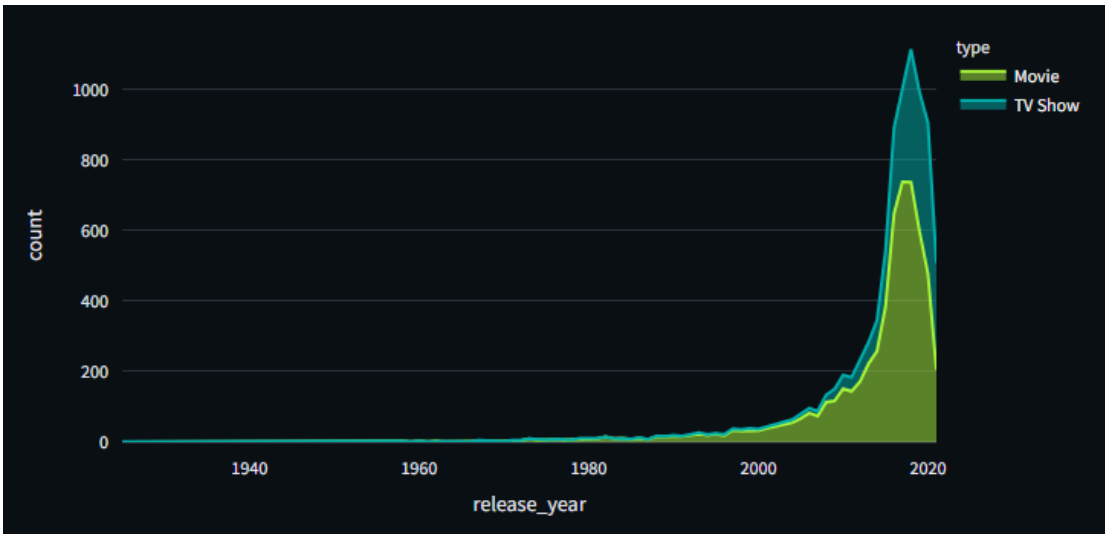
Importance in Achieving Objectives:

1. Analyzing Trends among Attributes: The ability to visualize attribute occurrences internationally aids in identifying regional trends in popular or diminishing content.

2. Identifying Popular Attributes: It assists with targeted marketing and product localization by displaying which features are popular in other nations.

3. Geospatial Analysis of Content Availability: It helps with targeted content distribution and market targeting by displaying where material is essentially available through the mapping of attribute occurrences.

Trend of Release Count by Year and Type



(Fig 1.1: Area Line)

This graphic illustrates the pattern of Netflix content releases throughout time, broken down into two categories: TV shows and movies. The graph's lines each indicate the number of releases of a certain category (TV show or movie) across a range of years. The number of releases is shown by the y-axis, while the years are represented by the x-axis. In this visualization, users can clearly see Netflix's spending on streaming content has increased significantly since 2015, registering an annual growth rate of 28.93% from 2016 to 2020. This may have allowed Netflix to produce and release more content to meet the growing demand of its subscribers, (Dean, 2023).

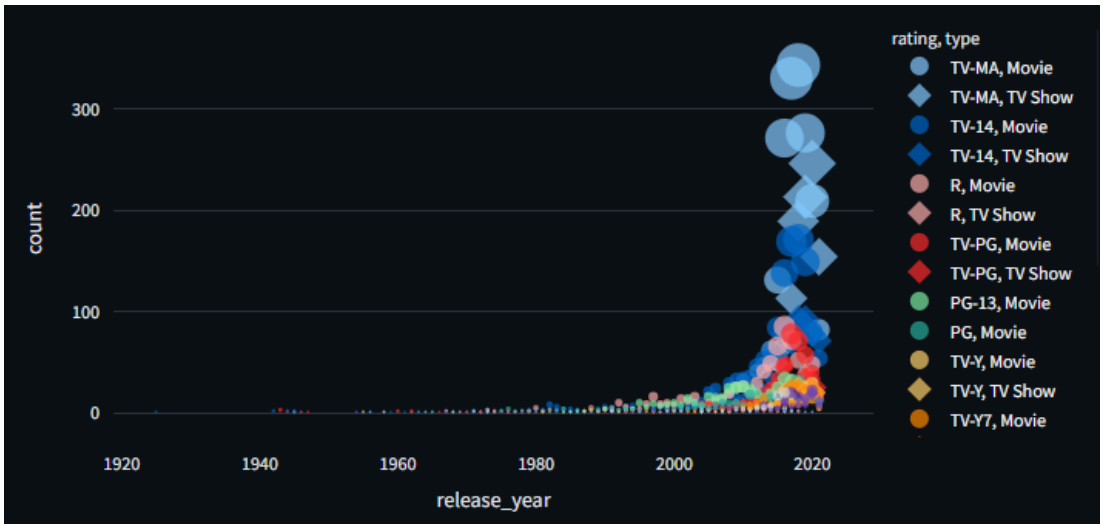
Importance in Achieving Objectives:

1. Understanding Release Trends: The Visualizations of the evolution of Netflix's movie and TV show releases throughout time. This provides insights on the evolution of Netflix's content strategy by highlighting periods where content production has increased, decreased, or stayed constant.

2. Identifying Popular Types: A comparison of Netflix's movie and TV programme release statistics over time illustrates which genre has dominated the service. Making judgements about content acquisition and production planning is influenced by the popularity of each genre.

3. Spotting Seasonal Patterns: The visualization reveals cyclical or seasonal release patterns, such highs or lows during particular quarters or months. By identifying these patterns, release schedules may be optimally adjusted to correspond with audience preferences and peak watching periods.

Scatter Plot of Data Counts by Release Year, Type and Rating



(Fig 1.2: Scatter Plot)

The distribution of Netflix content counts according to release year, genre (movies and TV shows), and rating is shown in this visualization as a scatter plot. With the x-axis representing the year of release, the y-axis representing the rating, and the size or color of the dots reflecting the number of content entries, each point on the scatter plot reflects a particular combination of these factors. In this Scatter plot shows that TV-MA movies and tv shows are increased from 2000 to 2020 due to netflix originally are rated TV-MA, that indicating a significant focus on this rating rather than family friendly program(Movieguide | The Family Guide to Movies & Entertainment, 2018).

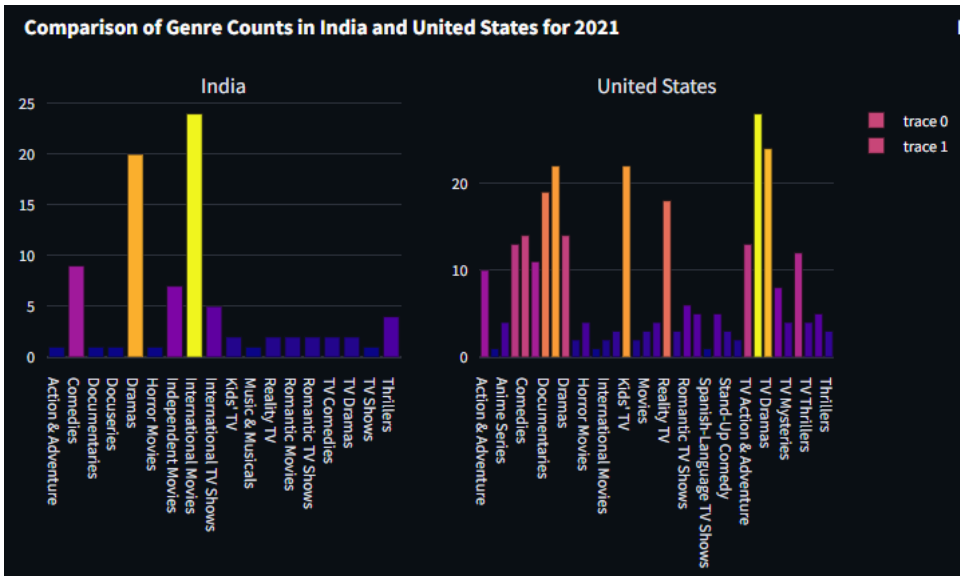
Importance in Achieving Objectives:

1. Track Content Evolution: It offers a moment in time of how Netflix has distributed its content, sorted by genre and rating. This makes it easier to monitor how content offerings change over time and comprehend how the platform's focus changes.

2. Analyze Content Performance: The correlation between release year, kind, and rating facilitates the examination of trends in content performance. Finding effective content strategies and areas for development is made easier with the use of this understanding.

3. Inform Decision-Making: Determining the distribution of content counts according to these factors helps in making strategic choices about content creation, audience targeting, and acquisition. In order to improve content offerings and increase audience engagement, it offers actionable insights.

Comparison of Genre Counts Between Country in each Year



(Fig 1.3:Subplot Bar)

A subplot bar graph comparing the numbers of various genres in several nations for each year is displayed in this visualization. Every subplot represents a particular year, and the number of each genre within each country is shown by grouped bars. The comparison makes it possible to track changes in country-specific preferences and genre preferences over time, offering insights into how Netflix's content consumption patterns are changing. In this case India in 2021 has International Movies as highest due to movie that release in that year often reflect cultural nuances and themes that resonate strongly with Indian audiences, driving high levels of engagement, However in United states TV dramas often have longer seasons and episodes compared to other genres, leading to a higher count of content being produced and aired shows like General Hospital and Days of Our Lives (Rochotte, 2023).

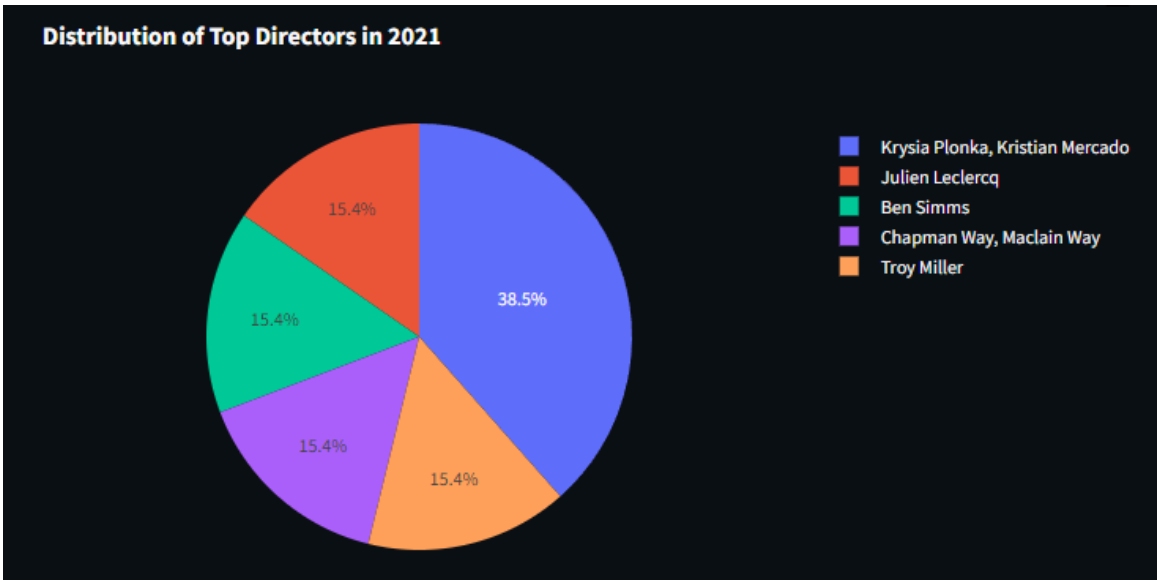
Importance in Achieving Objectives:

1. Identify Genre Preferences: It makes it possible to track changes in genre preferences over time in several different nations. Decisions on the creation and purchase of content are influenced by this awareness of audience preferences and tastes.

2. Track Content Consumption Trends: Netflix's shifting patterns in content consumption may be tracked by comparing the genre counts for each country each year. Being able to adapt content strategy in response to shifting audience interests requires having access to this information.

3. Inform Global Content Strategy: Decisions on global content strategy are informed by knowledge of regional differences in genre popularity. It facilitates the customisation of content offerings to particular regional tastes, hence optimizing viewer happiness and engagement in a variety of marketplaces.

Distribution of Top Directors in Specific Year



(Fig 1.4:Pie Graph)

This pie chart-style visualization shows the distribution of material directed by leading filmmakers for each year. Each slice in the pie chart indicates a different filmmaker who contributed to Netflix material in that particular year. The size of each slice shows how much of the material is credited to a certain filmmaker, emphasizing how important directors are to Netflix every year.

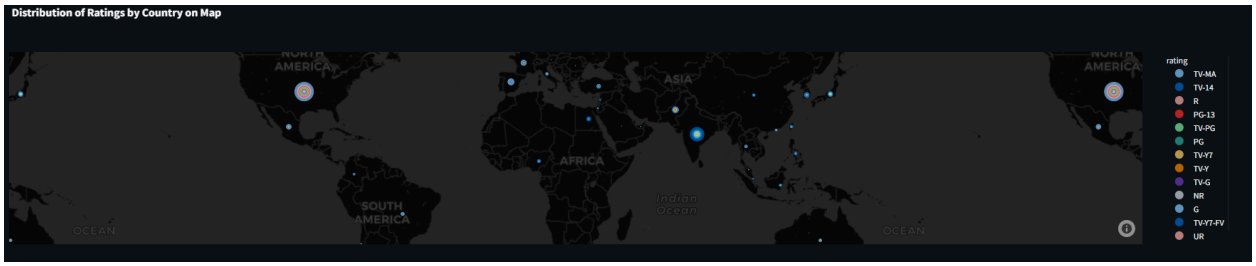
Importance in Achieving Objectives:

1. Understanding Director Impact: The amount of content that famous directors provide to Netflix annually is depicted in this visualization. Each year, it assists in determining which filmmakers have had the most influence on Netflix's content catalog, offering a glimpse into their evolution over time.

2. Tracking Director Trends :The annual pie charts can be used to track trends in director contributions. This indicates shifts in the popularity of directors, Netflix's tastes, and trends in talent. Decisions about alliances, people management, and content acquisition are informed by an understanding of these changes.

3. Evaluating Content Quality:Top directors are a sign of prospective content quality and viewer appeal in Netflix's collection. Directors with a proven track record recommend better material. Top directors' analyses of content distribution aid in evaluating the diversity and quality of Netflix's annual output.

Distribution of Ratings by Country on Map



(Fig 1.5:Scatter Plot Map)

The distribution of Netflix content ratings across various nations is shown on a scatter plot map in this visualization. Every point on the map represents a distinct geographic area (country); the color or size of the dots indicates the average rating of material that is available in each nation; the x- and y-axes show the longitude and latitude, respectively. This scatter plot map offers a clear illustration of the regional variations in content ratings, shedding light on local tastes and Netflix content reception. By using a scatter map, it is possible to quickly spot best and worst performing areas, trends, and outliers, and gain insights into the distribution of ratings by country and how to tackle these issues, for example bringing more content from that specific region .

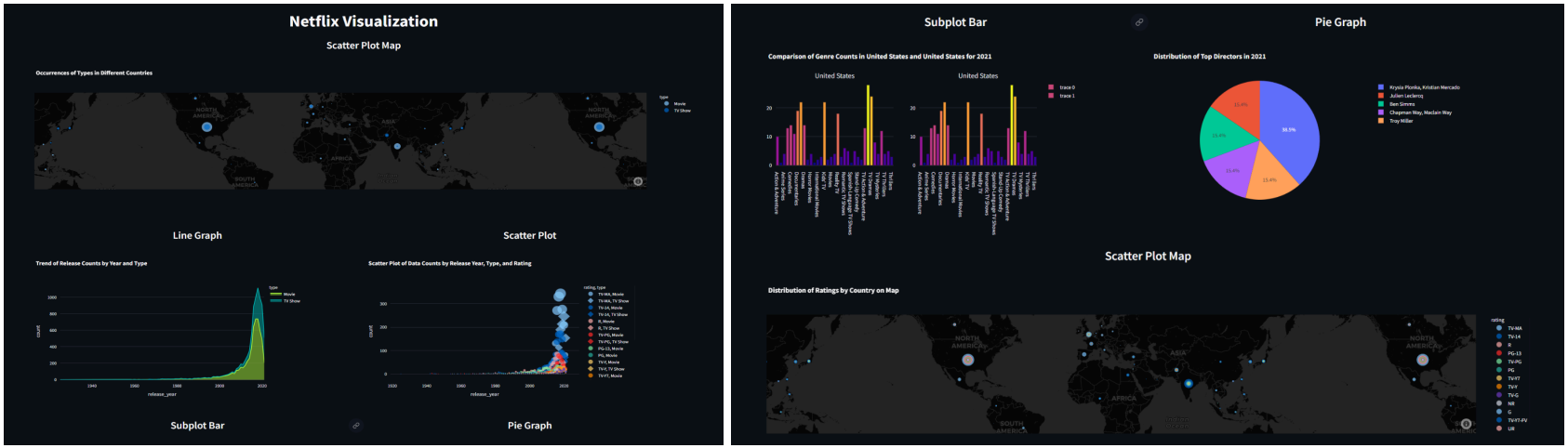
Importance in Achieving Objectives:

1. Understanding Regional Preferences:This graphic illustrates the differences in Netflix content ratings between nations. It assists in determining which geographical areas tend to obtain better or lower ratings for content, which helps with judgements about localization, audience targeting, and content selection.

2. Assessing Reception: The scatter plot map illustrates how different nations view Netflix content. Average ratings are shown by colors or point sizes, which enable cross-region comparisons. Examining these variances directs efforts to increase the relevancy and attractiveness of material and helps measure audience engagement.

3. Gaining Market Insights: The visualization provides insights that guide plans for growing internationally and entering new markets. Gaining insight into how material is assessed in various locations may help uncover potential markets and areas where local audience catering needs can be strengthened. Decisions about content localization, audience engagement tactics, and market prioritization are guided by this information.

Main Dashboard



(Fig 1.6 Main Dashboard)

Netflix's content landscape may be analyzed with insight thanks to the extensive set of visualizations provided by this dashboard. With the use of interactive maps and graphs, users can delve into a variety of aspects of Netflix's enormous collection of content, such as how it is distributed throughout the globe, how release counts have changed over time, and how different content categories, genres, ratings, and director contributions interact? Users may use these visualizations to learn more about Netflix's content ecosystem and to derive practical insights that can help guide strategic choices about the creation, distribution, and purchase of content.

With the help of the Netflix Dashboard, consumers may examine Netflix's content portfolio and make well-informed choices. Users may identify regional preferences and customize distribution techniques by visualizing different content kinds across national borders. Scatter plots provide information about the characteristics of the content and the audience's reaction, while trend analysis of release counts helps predict demand and optimize production schedules. In-depth insights are also offered by genre and director comparisons, and scores are mapped using geospatial analysis to comprehend regional variances. All things considered, the Netflix Dashboard provides users with useful information to improve audience engagement worldwide.

User Manual

Features

This application has a number of features implemented, including built-in Plotly Express tools, several pages, a slider, multiple selections, an expander from Streamlit, and shape on Plotly Express libraries. This feature makes it easier for users to distinguish between values and see greater insights in visualizations.

Plotly Express Tools

(Fig 2.0 Ploty Express Tools)

How it work:-

1. Zoom In/Zoom Out: With the use of these tools, users can enlarge or reduce the size of particular plot points for more detailed study or context. Generally, users may accomplish this by using the plot interface's buttons or by clicking and dragging to specify the region they want to zoom in on.

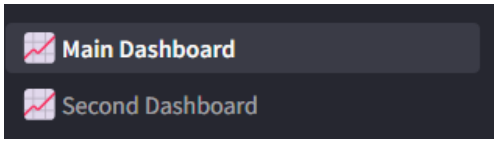
2. Download: With the use of this function, users may download the plot as a static file or picture for offline reading or sharing. Typically, the plot interface's menu or button allows users to access this option.

3. Full Screen: With the full-screen option, users may view the plot over the whole screen, giving them more space for in-depth research. By pressing a button or choosing an option from the plot interface, users may enable full-screen mode.

4. Resize: Users can change the plot's dimensions to better fit their screen or suit their viewing preferences by using the resize capability. The plot may normally be resized by clicking and dragging its corners or edges, or by utilizing the resizing handles that are included in the plot interface.

5. Reset View: This function undoes any zooming or panning that may have been done by returning the plot to its original perspective. Through the plot interface, users can choose an option or click a button to return to the original view.

Multi-Pages



(Fig 2.1 Eg Multipage)

How it works:

Handle Page Navigation: A page's content is rendered when a user selects it from the navigation sidebar because a corresponding Python script or module is run at that time. This might incorporate any other Streamlit element, data tables, input forms, or interactive visualizations.

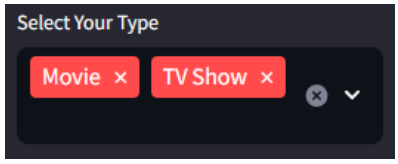
Slider

(Fig 2.2 Eg Slider)

How it works:

Updating Values: Real-time updates are made to the specified value when the user moves the slider. This implies that the updated selection will be automatically reflected in any downstream elements or visualizations that rely on the slider value.

Multiselect



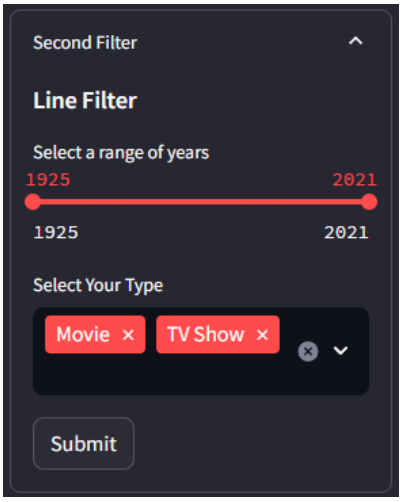
(Fig 2.3 Eg Multiselect)

How it works:

Updating Selected Values: The selected values are updated in real-time when the user makes selections or modifies their choices within the multiselect widget. This implies that any visualizations or downstream components that rely on the values chosen will likewise instantly change to reflect the updated choices.

Expander

(Fig 2.4 Eg Before Click Expander)



(Fig 2.4 Eg After Click Expander)

How it works:
Expanding and Collapsing Content: The Expander widget's content expands or collapses to switch between visible and hidden states when a user clicks on the widget's header. This minimizes distractions from other areas of the UI and enables users to concentrate on particular content portions.

Shape

(Fig 2.1 Example Shape)

How it works:
Using Shapes for Annotations: Shapes can be utilized not just for visual components but also for callouts and comments. To offer more background or plot details, users can annotate or label shapes with words.

Recommendation

Strengths And Weakness of Application

Strengths	Weakness
Easy to use: Use the terminal to type the command "pip install streamlit." When importing the Streamlit library, it is customary to use the following command: "import streamlit as st."	Limited Styling Options: The lack of advanced UI customization options in Streamlit may lead some users to seek alternative frameworks that offer more extensive styling capabilities
Open Source: For those who don't know HTML but know how to code in python, this will benefit the user a lot.	Production Scalability: For high-traffic or intensive workloads, consider performance optimization techniques or alternative deployment approaches.
Minimal code requirement: Streamlit requires less code than typical web development frameworks to produce visually appealing apps..	Limited Testing Tools: Built-in testing tools are still evolving, so consider using external libraries or frameworks for comprehensive testing.

Suggestion on improvement on Application

When it comes to improvement, some of the weaknesses can be improved. For example, for Limited Styling Options, The developer may learn how to implement python into html or other languages such as C/C++, Java, Javascript,CSS and other languages such as SQL. This may take the development of the application to take so long but the benefit in the future may be good for the developer himself or the company(ryanlampkin, 2022). For product Scalability,Caching: Utilize Streamlit's caching mechanisms to store and reuse computations, reducing processing load and improving app performance Efficient Data Loading: Optimize data loading processes by pre-processing data and using efficient data structures to enhance overall performance the improvement ,(*How to *Really* Scale a Streamlit App?*,2023).

By leveraging this improvement in limited styling options and product scalability, this will make huge improvement toward the application in terms of performance and UI/UX. Not only that will benefit the developer and the company.

Conclusion

In conclusion, the creation of a Streamlit application that visualizes Netflix's dataset provides insightful information about the content landscape by making it simple for users to examine patterns, pinpoint popular qualities, and do geographic studies. The programme makes use of Python modules like Streamlit, Plotly Express, and others to offer interactive visualizations that help users decide on distribution, audience targeting, and content acquisition. Although Streamlit is open-source and user-friendly, it could be more flexible in terms of style possibilities and scalable for high-traffic situations.

The user experience and scalability of the programme may be greatly improved by integrating the latest developments in UI customisation and performance optimisation strategies. The programme can meet the expectations of an increasing user base and achieve improved performance efficiency by adopting caching technologies and addressing style option constraints. The Streamlit application is well-positioned to provide useful insights and facilitate strategic decision-making processes for efficiently navigating Netflix's ever-changing content ecosystem through ongoing innovation and refinement.

References

Netflix Data: cleaning, analysis and visualization. (2022, August 26). Kaggle. https://www.kaggle.com/datasets/ariyoomotade/netflix-data-cleaning-analysis-and-visualization

GeeksforGeeks. (2018, December 26). *Split a text column into two columns in Pandas DataFrame*. https://www.geeksforgeeks.org/split-a-text-column-into-two-columns-in-pandas-dataframe/

st.sidebar - Streamlit Docs. (n.d.). https://docs.streamlit.io/library/api-reference/layout/st.sidebar

Pie. (n.d.). https://plotly.com/python/pie-charts/

Line. (n.d.). https://plotly.com/python/line-charts/

Dean, B. (2023, March 27). *Netflix Usage and Growth Statistics: How Many People Watch Netflix in 2021?* Backlinko. https://backlinko.com/netflix-users

Movieguide | The Family Guide to Movies & Entertainment. (2018, September 20). *61% of Netflix Originals Have Mature Content – Few Are Family Friendly*. Movieguide | the Family Guide to Movies & Entertainment. https://www.movieguide.org/news-articles/61-of-netflix-originals-have-mature-content-few-are-family-friendly.html

Rochotte, E. (2023, September 14). 14 U.S. TV Shows With the Greatest Number of Episodes of All Time. Redbook. https://www.redbookmag.com/life/g44749412/tv-shows-most-episodes/

Bar Charts. (n.d.). Plotly.com. https://plotly.com/python/bar-charts/

Scatter. (n.d.). https://plotly.com/python/scattermapbox/

Create a multi-page app - Streamlit Docs. (n.d.). https://docs.streamlit.io/get-started/tutorials/create-a-multipage-app

st.expander - Streamlit Docs. (n.d.). Docs.streamlit.io. Retrieved February 19, 2024, from https://docs.streamlit.io/library/api-reference/layout/st.expander

Kilcommins, S. (2021, April 28). *Streamlit — Everything You Need To Know*. Medium. https://medium.datadriveninvestor.com/streamlit-everything-you-need-to-know-665eb90fcf4a#11ed

Wawrzyński, A. (n.d.). Pros and cons of using Streamlit for simple demo apps. SoftwareMill. https://softwaremill.com/pros-and-cons-of-using-streamlit-for-simple-demo-apps/

ryanlampkin. (2022, August 22). *Will Streamlit layout and UI become more customizable?* Streamlit. https://discuss.streamlit.io/t/will-streamlit-layout-and-ui-become-more-customizable/29595

*How to *really* scale a Streamlit App?* (2023, June 6). Streamlit. https://discuss.streamlit.io/t/how-to-really-scale-a-streamlit-app/44662