**NETFLIX MOVIES AND TV SHOWS**

**Introduction**

With over 8000 movies and TV shows available on its platform and a staggering 250M global subscribers as of mid-2023, Netflix stands as one of the most popular media streaming giants. This digital repository has reshaped the global entertainment landscape, exemplifying the transformative influence of streaming platforms on audience consumption habits. In this report, we will be visualizing and analyzing the dataset ‘Netflix Movies and TV Shows’ sourced from Kaggle. The tabular dataset consists of listings of all the movies and TV shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc. We will be performing an interactive visual analysis of the dataset, encompassing data preparation in Python, and leveraging various types of visualization methods through Tableau. Our primary objective is to extract meaningful insights from this dataset, unraveling trends, preferences, and patterns prevalent among Netflix users. These revelations could serve as invaluable guidance for content creators and industry decision-makers, potentially enhancing content quality and aligning offerings with evolving viewer preferences. The significance of these findings extends not only to Netflix but also across diverse digital platforms.

**Methodology**

We followed a structured approach to visualize and analyze the ‘Netflix Movies and TV Shows’ dataset, consisting of the following steps:

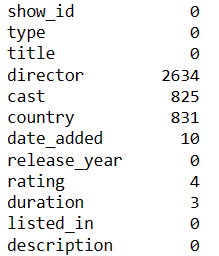
1. Raw data collection
2. Data Readiness
3. Interaction through Visualization Discussion and Interactive Dashboards
4. Data Visualization and Analysis
5. **Raw data collection**

[*https://www.kaggle.com/datasets/shivamb/netflix-shows/data*](https://www.kaggle.com/datasets/shivamb/netflix-shows/data)

1. **Data readiness**

***2.1. Data Overview***

The dataset consists of 12 columns and 8807 unique values. Within the dataset, there are 6 columns that have missing values. Notably, the 'Director' field accounts for approximately one-third of these missing values.

A screenshot of a computer code

Description automatically generated

*Figure 1: Missing Values in Columns (Quantity and Percentage)*

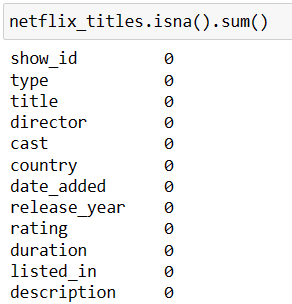
***2.2. Dealing with missing value***

To ensure the accuracy and reliability of our analysis, we performed the following data cleaning and processing steps:

* **Handling Missing Country Values**: Null values in the 'Country' column were substituted with the mode country value to ensure minimal impact on the dataset's representativeness.
* **Preserving Director Information**: Despite a considerable number of missing values in the 'Director' column, it was retained due to its potential relevance in exploring specific directors' works.
* **Replacing Null Entries**: Null values in 'Director' and 'Cast' columns were replaced with 'No Data' to signify the absence of available information.
* **Dropping Remaining Missing Values**: Any remaining missing values across other columns were dropped from the dataset.

***2.3. Post-Cleaning Dataset Summary***

After cleaning the data, there are 8790 rows remaining with 12 columns.

A screenshot of a computer

Description automatically generated

*Figure 2: Post-Cleaning - Absence of Null Values and Dataset Info*

Where possible, I will use Netflix brand colors that enhances visual consistency and reinforces brand identity, aiding in a more engaging and recognizable presentation.

A red rectangular object with black text

Description automatically generated

1. **Interaction through Visualization Discussion and Interactive Dashboards**

The discussion and presentation below highlight the interactive features and insights derived from two prominent dashboards.

A screenshot of a computer screen

Description automatically generated

*Dashboard 1*

A screenshot of a computer screen

Description automatically generated

*Dashboard 2*

There are three outstanding interactions that I will want to discuss and use throughout this data visualization report, which apply to all worksheets using the data source.

**Filter Actions:**

I set up filter actions in these dashboards that allow to click on elements within one chart and view corresponding changes in other related charts, showcasing the interconnectedness of data visualizations. Deselecting a chosen part of a chart reverts all charts back to their original state, allowing for a comprehensive view of the data.

*Example Interaction:*

A screenshot of a computer

Description automatically generated

*Dashboard 1: Filter Action: Select ‘United States’*

Upon selecting 'United States' in the map, the remaining charts dynamically respond to this selection, providing specific insights:

* Movies dominate the content distribution in the US, accounting for 68% of Netflix content.
* Netflix appears to cater largely to 'Adult' audiences in the United States, with nearly half (46%) of its content targeting this demographic.
* Notably, TV Shows exhibit a shorter lag time compared to Movies, possibly due to annual season releases on Netflix.

**Filter based on different types or categories:**

Filter by 'Type’ or Filter by ‘Maturity Groups’ allows for the isolation and analysis of specific segments of information within a dataset, aiding in the examination and comparison of distinct categories, enabling clearer insights and targeted analysis. Specific illustrations are in *Chart 6* and *Chart 7* in the following section – part 4.

**Parameter ‘Top N’:**

I created the "Top N" parameter to enable dynamic control over the displayed items. Users can easily switch between different subsets of data without manual adjustments. Top N with limit from 1 to 15 with current value of 10 and can be changed the value by typing in the number, we want.

*Example Interaction:*

A screenshot of a computer screen

Description automatically generated

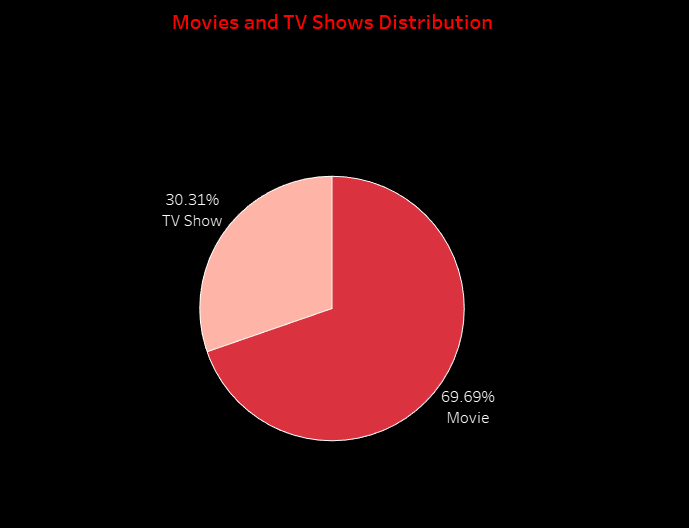
*Dashboard 2: Top N: Change Top 10 to Top 1*

By changing top N to top 1, we can easily see that the most popular genre, 'Dramas,' reveals a rapid rise post-2010, peaking in 2018 before a subsequent decline in the line chart or Director ‘Rajiv Chilaka’ stands out as the director with the most movies on Netflix.

Next, let’s visualize and evaluate each chart to glean insights.

1. **Data Visualization and Analysis**

**What is the distribution of TV shows versus movies available on Netflix?**

****

*Chart 1: Movies and TV Shows Distribution*

**Results and Findings:**

From the above pie chart (*chart 1*), we can see:

* 'Movies' constitute approximately 69.69% and 'TV Shows' approximately 30.31% of Netflix's content.
* The pie chart underscores the dominance of movies over TV shows in Netflix's content inventory, showcasing a higher quantity of movies compared to TV show offerings.

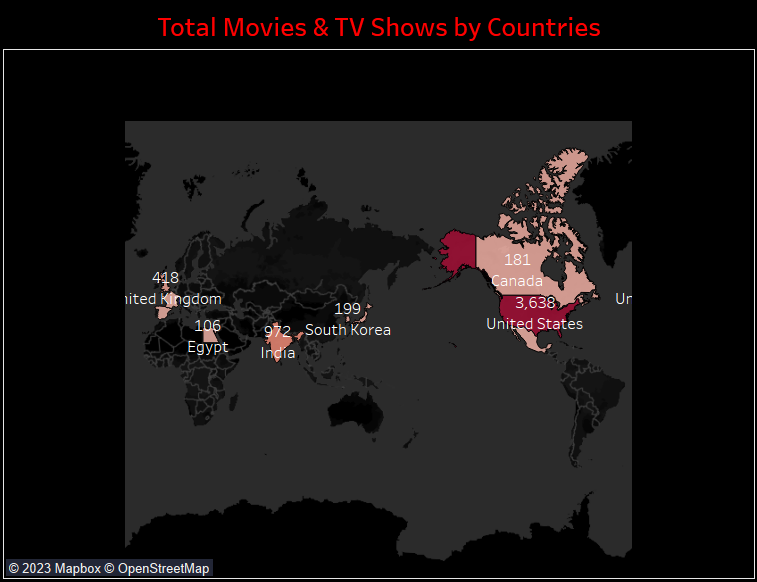
**Technical details of the used visualization method(s):**

* Visualization Technique: Pie Chart
* Variables Utilized: 'Type' column distinguishing between 'TV Show' and 'Movie'
* Calculation Method: Employed Table Calculation to derive percentages.
* Percentage Calculation: Calculated as % of Total Distinct Count of ShowID along the table.
* Visual Enhancement: Added labels indicating type and respective percentages within pie chart segments.

**Discussion on the justification of the visualization method(s) used:**

The selection of a pie chart for illustrating the distribution of TV shows versus movies on Netflix was deliberate due to its suitability for showcasing proportional relationships. The nature of the data—categorical with distinct categories ('TV Show' and 'Movie')—aligns well with the pie chart format. The utilization of percentages within the segments enhances the clarity of understanding, providing a direct comparison between the two categories.

**How does the quantity of movies/TV shows vary across different countries?**

****

*Chart 2: Total Movies & TV Shows by Countries*

**Results and Findings:**

As we can see from the map (*chart 2*):

* The USA stands out as the top creator of content for Netflix, followed by India and the UK, although they produce notably less content in comparison.
* This trend is understandable considering Netflix's American origins, which naturally lead to the USA being the leading contributor to its content library.

**Technical details:**

* Visualization Technique: Map Chart
* Variables Utilized: 'CNTD(Show Id)' for color representation, 'CNTD(Show Id)' and 'Country' for labeling
* Filter Applied: Country and Type

**Discussion on the justification:**

The utilization of a map chart to display movie and TV show quantities across various countries strategically allows for a clear comparison. Color codes indicating the count of shows ('CNTD(Show Id)') and labels ('CNTD(Show Id)' and 'Country') make it easy to identify specific counts per country. This approach offers an intuitive overview of Netflix content distribution across regions.

**Are there noticeable differences in the distribution of movies and TV shows across these countries?**

**A graph with red and white bars

Description automatically generated with medium confidence**

*Chart 3: Top 10 Countries – Movie and TV Show Split*

**Results and Findings:**

Based on the percent bar chart (chart 3), we can see that:

* Interestingly, Netflix in India is made up nearly entirely of Movies. Bollywood is big business, and perhaps the focus of this industry is Movies and not TV Shows.
* Conversely, South Korean Netflix is almost entirely TV Shows.
* The underlying reasons for the difference in content must be due to market research conducted by Netflix.

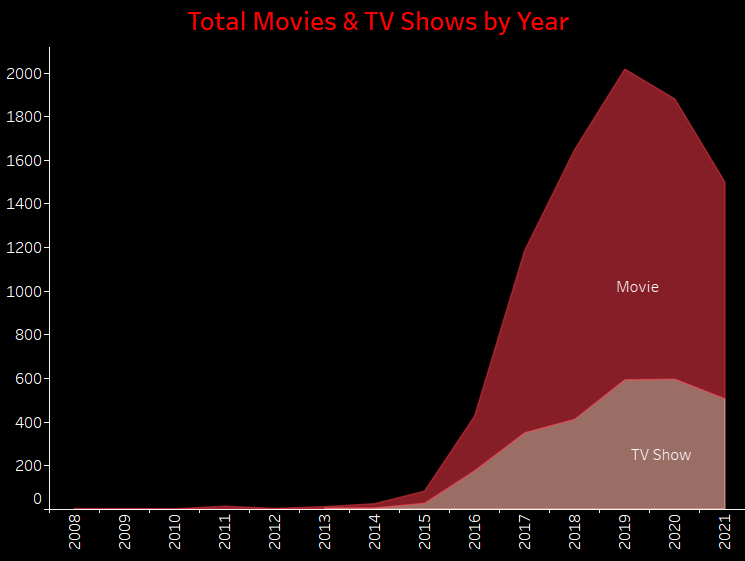
**Technical details:**

* Visualization Technique: Percent Stacked Bar Chart
* Variables Utilized: Column: CNTD(ShowID), Row: Country
* Filter: Country (Top N countries)
* Color: Type (Movies and TV Shows)
* Calculation Method: Created a calculated field 'MoviePercentage' to sort countries by movie percentage.

**Discussion on the justification:**

I opted for a percent stacked bar chart to visualize the distribution of movies and TV shows across various countries due to its ability to display these differences comprehensively. I utilized 'CNTD(ShowID)' to quantify the content, 'Country' for comparative analysis among nations, and 'Type' to differentiate between movies and TV shows. Sorting the countries by 'MoviePercentage' allowed for a clearer comparison of the prevalence of movies versus TV shows within the top 10 countries.

**How has the addition of content on Netflix changed over time?**

****

*Chart 4: Total Movies & TV Shows by Year*

**Results and Findings:**

From the above line chart (*chart 4*):

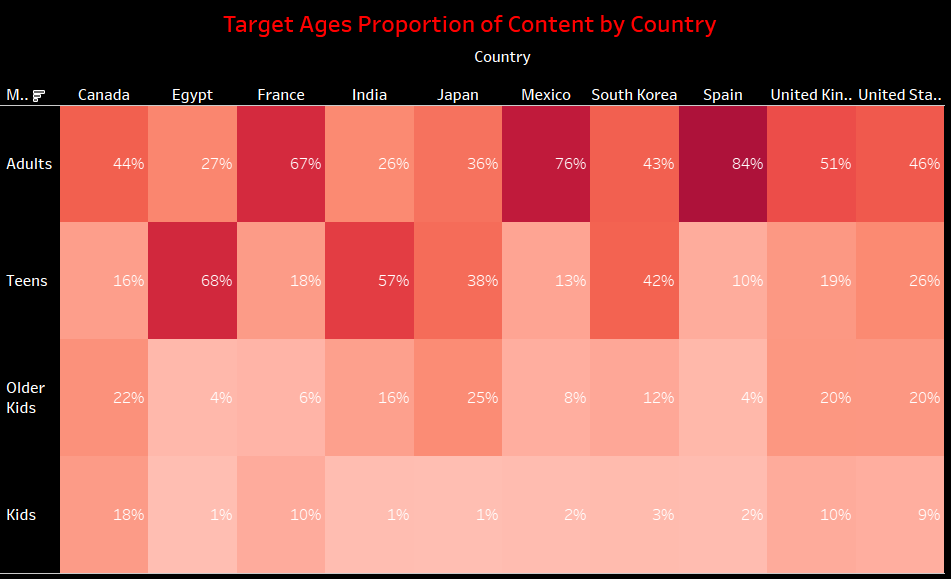
* Netflix's initial years showed a gradual beginning. The upward trend started in 2015, followed by a sharp rise beginning from 2016 onwards.
* The peak in Netflix's global content addition occurred in 2019, followed by a visible decline in 2020. This drop is likely linked to the COVID-19 pandemic, causing disruptions in production schedules and content creation.
* It appears that Netflix has focused more attention on increasing Movie content than TV Shows. Movies have increased much more dramatically than TV shows.

**Discussion on the justification:**

The line chart was chosen as it is good for showing how things change over time, making it ideal for displaying Netflix's content additions year by year. I used the 'Year(Date Added)' to show each year's content and 'Count(Show ID)' to count how much was added each year. Using different colors for movies and TV shows helped in seeing the difference between their additions easily.

**Target Ages Proportion of Content by Country**

How does the distribution of target age groups for content vary across different countries on Netflix? Does Netflix constantly focus on specific demographics universally, or does this strategy differ depending on the country?



*Chart 5: Target Ages Proportion of Content by Country*

**Results and Findings:**

The above heat map (*Chart 5*) revealed interesting differences among countries:

* Netflix appears to focus on offering shows for ‘Adult’ audiences across most countries.
* Shows in countries like India and Egypt seem predominantly targeted at teenagers.
* It is also interesting to note similarities between culturally similar countries - the US & UK are closely aligned with their Netflix target ages. However, these trends notably differ from countries like India or Japan.

**Technical Details:**

* Visualization Technique: Heat Map
* Variables Utilized: Column: Country, Row: Maturity Rating (group)
* Filter: Country (Top N countries); Maturity rating (Excluding 'Others' - UR: Unrated, NR: Not Rated, and keeping 4 other groups: Kids, Older Kids, Teens, Adults)
* Color and Label: CNTD(Show Id)
* Table Calculation: Percentage calculation of Maturity Rating (group) for each country using CNTD(Show Id)

**Discussion on the justification:**

I opted for a heat map to illustrate the distribution of target ages for content across different countries within the top 10. This visualization method efficiently showcases the proportional representation of age groups. Utilizing 'Country' and 'Maturity Rating (group)' provides a comprehensive understanding of how Netflix targets specific demographics across various regions. The exclusion of 'Others' and emphasis on ‘Kids’, ‘Older Kids’, ‘Teens’, and ‘Adults’ groups allows for a focused analysis.

**How old are the Movies (TV Shows)?**

Let’s take a look at the lag between the release of content and its availability on Netflix.A graph of different colored bars

Description automatically generated

*Chart 6: The lag time between the release of content and its availability on Netflix*

A graph of a number of movies

Description automatically generated with medium confidenceA graph of a television show

Description automatically generated

*Chart 6.1 & Chart 6.2: Filter by ‘Movie’ and ‘TV Show’ respectively*

**Results and Findings:**

Based on these bar charts presented above (*chart 6*, *6.1*, *6.2*), it is observable that:

* The time lapse between the release of content and its arrival on Netflix differs across countries.
* The time gap for TV Shows seems to be shorter and more consistent compared to Movies. This pattern likely results from the seasons of TV shows being released annually on Netflix, creating a more regular timeline for their appearance on the platform.
* In Spain, newer movies seem to dominate Netflix, while Egypt and India tend to feature older movies on average.
* Overall, Spain stands out for having the most recent content available on Netflix.

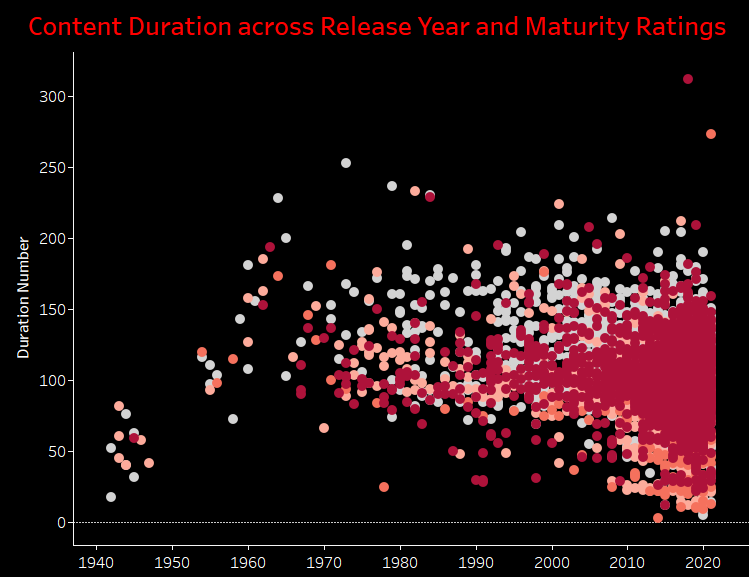
**Technical Details:**

* Visualization Technique: Bar Chart
* Variables Utilized: Columns: Country, Rows: AVG(Lag Time)
* Calculated Field: Lag Time = DATEDIFF('year', [Release Year], [Year Added])
* Filter: 'Type' → Allows for a side-by-side bar chart comparing the time gap between Movies and TV Shows in different countries. Alternatively, examining each type separately in bar charts highlights countries with recent content updates.

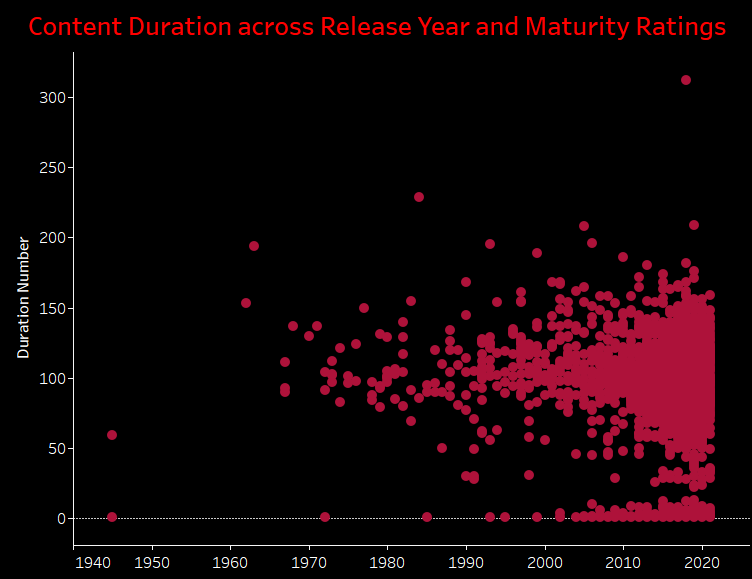
**Discussion on the justification:**

The choice of bar charts facilitates the comparison of average delays between content release and appearance on Netflix across different countries and content types. It helps identify countries with recent content updates and allows for a direct comparison between Movies and TV Shows.

**How does the duration of content vary concerning the release year across different maturity ratings?**

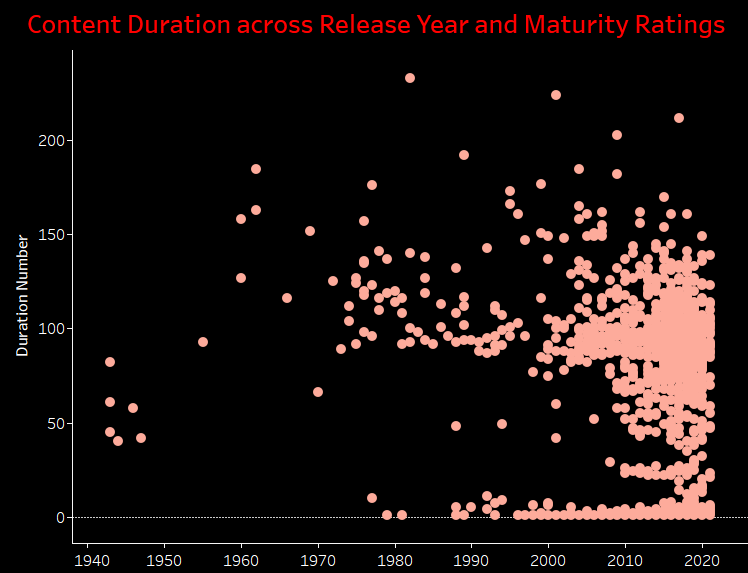
****

*Chart 7: Content Duration across Release Year and Maturity Ratings*

**A screen shot of a graph

Description automatically generated**

*Chart 7.1 & 7.2: Filter by Maturity Ratings, ‘Adults’ and ‘Teens’ Group respectively.*

**A graph of red dots

Description automatically generated**

*Chart 7.3 & 7.4: Filter by Maturity Ratings, ‘Older Kids’ and ‘Kids’ Group respectively.*

**Results and Findings:**

We can see from the above scatter plot (*chart 7*):

**For Movies:**

* Each of the four maturity rating groups displays an increasing diversity in movie durations by the time. Prior to 2000, movie durations were typically between 100 to 150 minutes. After 2000, the range expanded significantly, encompassing shorter films (around ten minutes) to longer ones (up to more than 200 minutes).
* Despite the increasing diversity, movies predominantly center around the 100 to 150-minute duration mark across maturity groups.
* Adult and teenage-oriented movies exhibit similar duration patterns, centered around 100 to 150 minutes. In contrast, movies intended for children tend to have shorter durations, typically ranging from 50 to 100 minutes.

**For TV Shows:**

* All four maturity rating groups showcase a similar trend. Starting from 2010, there's a noticeable increase in the number of TV Shows, offering more diverse durations in terms of season count. However, the predominant focus remains on TV Shows with fewer seasons.

**Technical Details:**

* Visualization Technique: Scatter Plot
* Variables Utilized: Column: Release Year, Row: Duration
* Color: Maturity Rating
* Filter: Rating, Maturity Rating, Type
* Details: Titles (using ATTR function)

**Discussion on the justification:**

The scatter plot visualization method effectively illustrates how the duration of content varies concerning the release year across different maturity ratings. By plotting Release Year against ‘Duration’ and utilizing color and filter by ‘Maturity Ratings’, this chart facilitates to compare content duration distributions for different maturity ratings, particularly in relation to the release year.

**Genre Evolution Over Years**

How has the diversity and distribution of genres available on Netflix changed over the years? Which genres have shown the most significant increase or decrease in representation?

A graph of a number of years

Description automatically generated

*Chart 8: Genre Evolution Over Years*

**Results and Insights:**

The above line chart (*chart 8*) shows that:

* Netflix began with a limited number of genres, each with a small quantity. Subsequently, from approximately 2010, there was a rapid rise in the total quantity of genres available.
* Netflix showcased the highest total quantity of genres around 2017-2019. However, there was a subsequent sharp decline in genre quantity post-2019.
* The marked decrease in genre quantity might be linked to the covid pandemic, impacting the introduction and availability of new genres.

**Technical Details:**

* Visualization Technique: Line Chart
* Variables Utilized: Column: Release Year, Row: CNTD(Show Id)
* Calculated Field:  'Genre\_Individual' field is derived from the 'Listed In' column to address movies with multiple genres. It isolates the primary genre from combined listings, typically representing the main genre of the movie.
* Color: Genre\_Individual
* Filter: Genre\_Individual & Type (Top N genres most popular)

**Discussion on the Justification:**

The chosen line chart effectively illustrates the changing landscape of genre diversity and distribution on Netflix over different years. By employing 'Release Year' on the x-axis and the count of unique show IDs ('CNTD(Show ID)') on the y-axis, it vividly portrays how the quantity and variety of genres have evolved over time. Utilizing 'Genre\_Individual' as a color-coded representation enhances the analysis by focusing on specific genres derived from a calculated field based on 'Listed In.'

**Director Content Distribution**

How does the distribution of movies and TV shows vary across different directors? Are there specific directors who contribute more content to Netflix?

A screen shot of a movie

Description automatically generated

*Chart 9: Director Content Distribution*

**Results and Insights:**

While most directors contribute a modest number of movies and TV shows, a select few directors stand out for their prolific contributions to Netflix's content library. Therefore, I filter the top 10 contributing directors displayed in the Tree Map. Notably, Rajiv Chilaka holds the highest content counts among this elite group.

**Technical Details:**

Visualization Technique: Tree Map

Variables Used: Director\_Individual, CNTD(ShowID), Color: CNTD(ShowID)

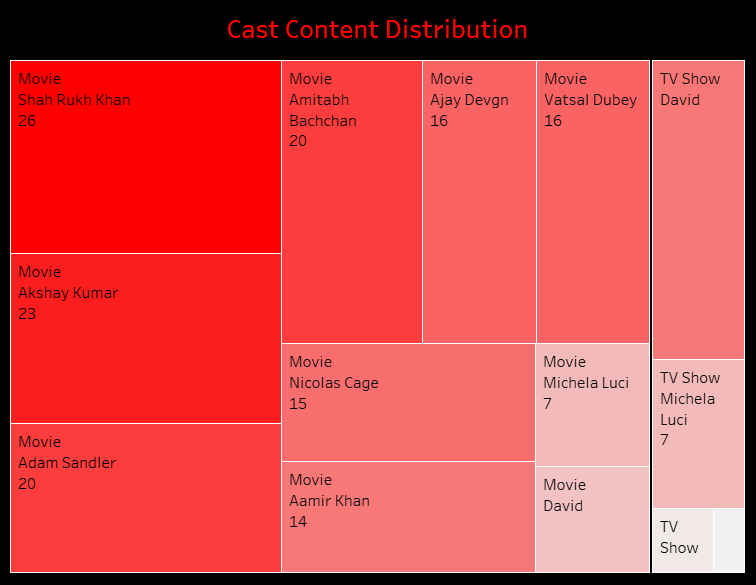
Filtered By: Type (Movies/TV Shows), Director\_Individual (Top N directors)

**Discussion on the Justification:**

The Tree Map visualizes content distribution across directors, showcasing their impact on Netflix content. By isolating top directors and specific content types, it highlights which directors significantly contribute, offering insights into who shapes Netflix's library.

Do the same steps as chart 9, we can get the ‘Cast Content Distribution’ as below (chart 10):

**Cast Content Distribution**



*Chart 10: Cast Content Distribution*

**Conclusion**

In conclusion, the comprehensive analysis of Netflix content distribution across different countries, genres, maturity ratings, and timeframes has revealed interesting patterns and insights. The visual representations and interactive dashboards employed in this study have offered valuable perspectives on the diversity and evolution of content on the platform. Through various charts and interactivities, we have understanded and explained correlations, trends, and focal points within Netflix's content landscape more clearly. The strategic utilization of visualization techniques has provided a dynamic exploration of data, facilitating a deeper understanding of content distribution trends, audience preferences, and the platform's strategic focus. As the streaming landscape continues to evolve, leveraging such data-driven insights remains crucial for content strategists, marketers, and stakeholders to adapt, innovate, and cater effectively to diverse viewer preferences.