

NATIONAL UNIVERSITY SCIENCE AND TECHNOLOGY (NUST)

(High Impact Skills Development Program for Gilgit Baltistan)

Name: Hasnain Asad

RollNo: GIL-DSAI-137

SECTION : 03

Email Address: Hasnainasad2016@gmail.com

Project Title: Interactive Dashboard Development
Using Tableau for Exploratory Data Analysis (EDA)

Tableau Public Account link:

https://public.tableau.com/views/DataVisulization_17260953362700/Netflix?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Objective:

In this project, you will perform Exploratory Data Analysis (EDA) using Tableau, which is a crucial step before any AI or machine learning task. EDA helps you understand the underlying patterns, distributions, and relationships within your data, setting the stage for more sophisticated analysis. By engaging deeply with the data through Tableau, you will practice making informed decisions that not only improve the effectiveness of your visualizations but also lay a solid groundwork for future predictive modeling or machine learning applications. You will select a substantial dataset of your choice and develop an interactive dashboard that uncovers and presents meaningful insights.

The dataset should include around 1000 items (rows) and 10 attributes (columns). Your final submission will include a Tableau Public link to your dashboard and a concise report detailing your EDA process, design decisions, and any challenges encountered.

Introduction:

The Netflix Titles dataset sourced from Kaggle provides a comprehensive collection of information on movies and TV shows available on Netflix. This dataset includes attributes such as title, director, cast, country, release year, rating, and duration, making it an excellent resource for exploratory data analysis (EDA). The primary objectives of this EDA are to explore the distribution and trends of Netflix content, identify patterns related to genres, countries, and release years, and create a data visualization dashboard that effectively communicates these insights. The insights gained from this analysis will support future AI/ML tasks, such as building recommendation systems and predicting content popularity.

Visualization Process:

To create the visualizations, the following steps were taken:

1. Content Distribution by Genre:

Bar Chart: A bar chart was chosen to display the number of titles per genre as it clearly shows the comparison between different genres.

- Pie Chart: A pie chart was used to show the percentage distribution of genres, providing a quick visual representation of genre proportions.

2. Content Distribution by Country:

- World Map: A world map was used to highlight the number of titles produced by each country, offering a geographical perspective.
- Bar Chart: A bar chart was used to compare the top 10 countries with the most content, making it easy to see which countries contribute the most.

3. Release Year Trends

- Line Chart: A line chart was chosen to show the number of titles released each year, as it effectively displays trends over time.
- Histogram: A histogram was used to display the distribution of release years, providing insight into the frequency of releases over different periods.

4. Ratings Distribution:

- **Bar Chart:** A bar chart was used to show the count of titles for each rating category, making it easy to compare the popularity of different ratings.
- **Box Plot:** A box plot was used to analyze the distribution of ratings across different genres, highlighting any outliers or variations.

5. Duration Analysis:

- **Histogram:** A histogram was used to display the distribution of movie durations, showing the range and frequency of different lengths.
- **Bar Chart :** A bar chart was used to show the number of seasons for TV shows, clearly comparing TV show lengths.

Decision-Making Justification:

The design choices for the visualizations were made to enhance clarity, aesthetics, and interactivity:

- **Clarity:** Each chart type was chosen based on its ability to clearly convey the intended information. For example, bar charts are excellent for comparing categories, while line charts are ideal for showing trends over time.
- **Aesthetics:** Color schemes and chart designs were selected to be visually appealing and easy to interpret. Consistent use of colors helps in distinguishing different categories and maintaining a cohesive look.
- **Interactivity:** Interactive elements such as tooltips and filters were incorporated to allow users to explore the data in more detail. This enhances user engagement and provides a more dynamic experience.

Challenges and Solutions:

Several challenges were encountered during the project

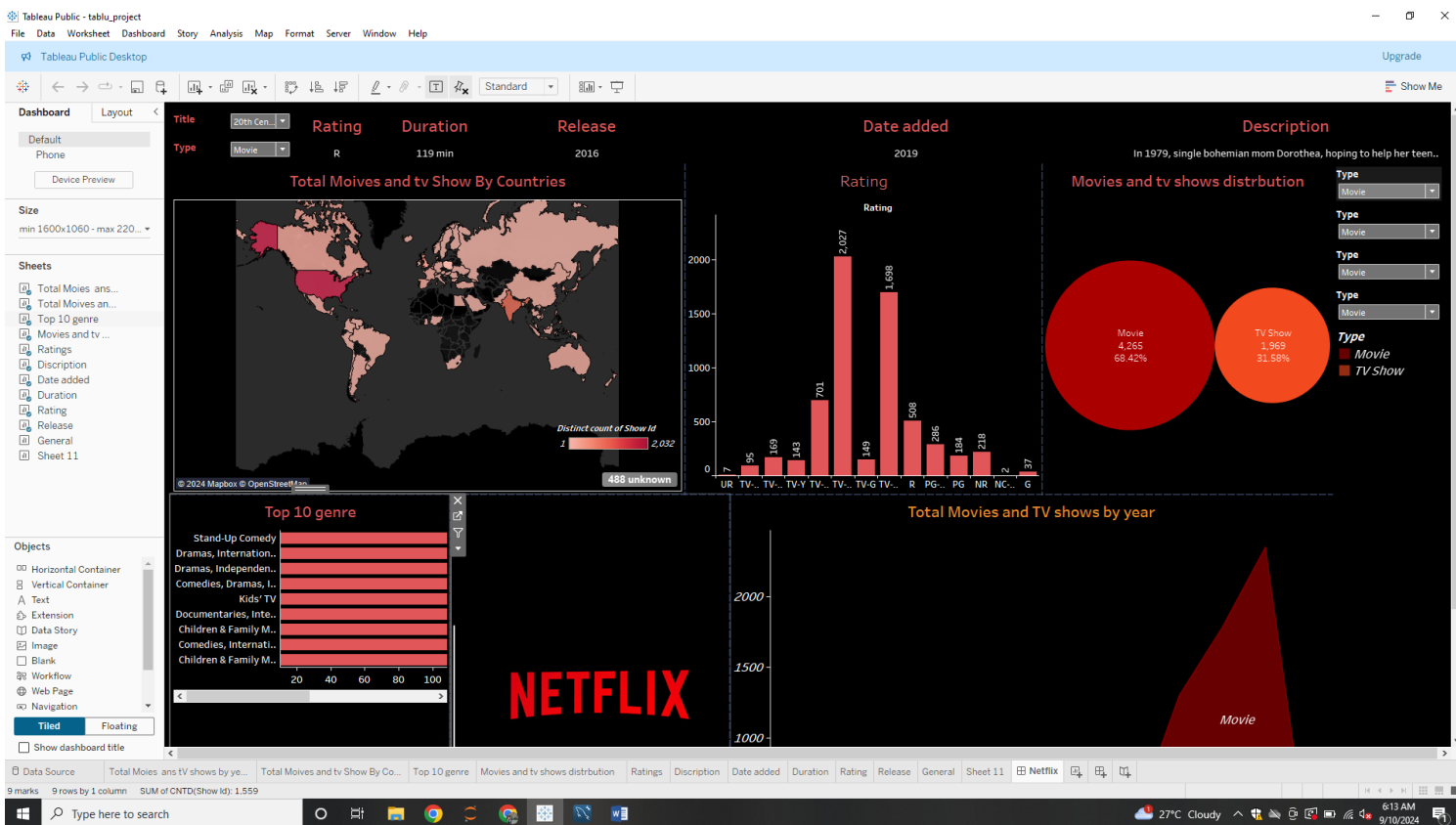
- **Standardizing Data:** Ensuring consistency in categorical data, such as country names and genres, was challenging. This was resolved by standardizing formats and using consistent naming conventions.
- **Choosing Appropriate Visualizations:** Selecting the right chart types to effectively communicate the data was crucial. This was achieved by considering the nature of the data and the insights we aimed to convey.

Conclusion:

The data visualization dashboard for the Netflix Titles dataset provides a comprehensive overview of the content available on Netflix. Key insights derived from the EDA include:

- **Genre Popularity:** The most and least popular genres were identified, providing insights into content preferences.
- **Country Contributions:** The analysis revealed which countries produce the most content, highlighting global content distribution.
- **Trends Over Time:** Observing the number of releases over the years helped identify trends in content production.
- **Rating Patterns:** The distribution of content ratings was analyzed, offering insights into the types of content available.
- **Duration Trends:** The typical lengths of movies and TV shows were understood, providing a better understanding of content formats.

The dashboard serves as a powerful tool for understanding the distribution and characteristics of Netflix's content library, supporting data-driven decision-making in content strategy and development.



Github Link: <https://github.com/hasnainasad1/Data-Visualization-Project>