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

:

Student Name:

Prachi Singhroha

Student No.: *21355131*

Module ID: *CS7DS4/CSU44056*

Assignment 3

Declaration:

*I understand that this is an **individual** assessment and that collaboration is not permitted. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I understand that by returning this declaration with my work, I am agreeing with the above statement.*

General Description of the visualisation

For this assignment, I have created a visualisation for Netflix data available on Kaggle. The visualisation is in form of a dashboard.

Netflix is a very popular streaming platform with thousands of movies and TV shows. The dataset included titles of movies and TV shows till 18th October 2021 along with their respective details like their ID, name of the director, rating, the country it was produced in, release year, duration, date it was uploaded on Netflix, and genre(s).

The visualisation has following attributes:

- Quantitative
    - Release year (of movie or TV show available on Netflix)
    - Duration (of movie or TV show available on Netflix)
- Derived quantitative
    - Percent distribution of movies and TV shows on Netflix
    - Trends in movies and TV shows count on Netflix through the years
    - Unique count of ID w.r.t country, years, genres and ratings
- Categorical
    - Genres (of movies or TV shows available on Netflix)
    - Rating (of movies or TV shows available on Netflix)
    - Type (Movie or TV show)
    - Director (of movie or TV show available on Netflix)
    - ID (of movie or TV show available on Netflix)
    - Country (of movie or TV show where it was produced)
- Derived categorical
    - The top 10 genres on Netflix
    - The top 5 ratings in movies on Netflix
    - The top 3 ratings in TV shows on Netflix
    - Details of a particular movie or TV show available on Netflix

Tasks supported by the visualisation

This visualisation answers the following questions.

1. What are the movies and TV shows percentages distributed on Netflix?
2. What are the top 10 genres on Netflix?
3. What are the top 5 ratings in movies on Netflix?
4. What are the top 3 ratings for TV shows on Netflix?
5. What is the trend of movies and TV shows counting on Netflix through the years?
6. What is the count of movies and TV shows on Netflix by the country where it was produced?
7. What are the details (ID, name of director, rating, release year, genre(s) and duration) of a particular movie or TV show available on Netflix?

Tools used for the visualisation

The dashboard was created on Tableau Desktop and before creating the visualisation the data was cleaned using Excel. I deleted a few columns that were not required for the visualisation and empty cells were filled.

Encoding channels and idioms used for creating the visualisation.

1. Area chart
    a. Data:
        i. Table: categorical variable (type i.e. movie or TV show) and quantitative variable (Years, 2008 – 2021)
    b. Encoding channel: Colour to differentiate between the two types because it makes the visualisation attractive. The count value of each type in the coloured area.
    c. Task: to understand the trend of movies and TV shows count on Netflix through the years.
2. World regional map
    a. Data:
        i. Table: categorical attribute (country where the movie/ TV show was produced) and unique count of Show ID.
    b. Encoding channel: colour (sequential segmented colormap) to represent the amount of movies/TV shows produced in each country making the map more engaging.
    c. Task: count movies and TV shows on Netflix by the country where it was produced
3. Bar chart
    a. Data:
        i. Table: Categorical attribute (Genres) and quantitative attribute (count of ID w.r.t genres)
    b. Encoding channel: express value with aligned position/length in the primary direction, separate key attribute with position in second direction
    c. Task: lookup top 10 genres on Netflix
4. Pie chart
    a. Data:
        i. Table: Categorical attribute (Type) and quantitative attribute (count of ID w.r.t type)
    b. Encoding channel: area marks (wedges) with angle channel to increase efficacy, colour to make it visually appealing
    c. Task: percent distribution of movies and TV shows on Netflix
5. Packed bubbles:
    a. Data:
        i. Table: Categorical attribute (Ratings) and quantitative attribute (count of ID w.r.t ratings of movies)
        ii. Table: Categorical attribute (Ratings) and quantitative attribute (count of ID w.r.t ratings of TV shows)
    b. Encoding channel: colour and number (count of show IDs for each rating) to make the dashboard visually appealing and easy to read
    c. Task:
        i. Lookup top 5 ratings in movies on Netflix
        ii. Lookup top 3 ratings in TV shows on Netflix

Discussion about the novelty of visualisation

Although the data has been visualized before, this visualization is novel because the dashboard can be used to find details about the particular movie/ TV show from their respective lists and also gives insights about Netflix which can help understand the trends and distribution of available movies/ TV shows. Multiple dashboards have been created for Netflix in the past but none allowed the user to filter details and understand popular trends at the same time. This distribution can be used by people who are looking for information about a particular movie, want to understand trends of movies/ TV shows on Netflix, directors/ producers looking to choose a popular genre for

their movie/ show, or find movies/TV shows to watch if they are confused. Also, the visualisation fulfils the following two requirements.

1. The overall visualization uses multiple idioms
2. Multiple facets (dashboard of different idioms)

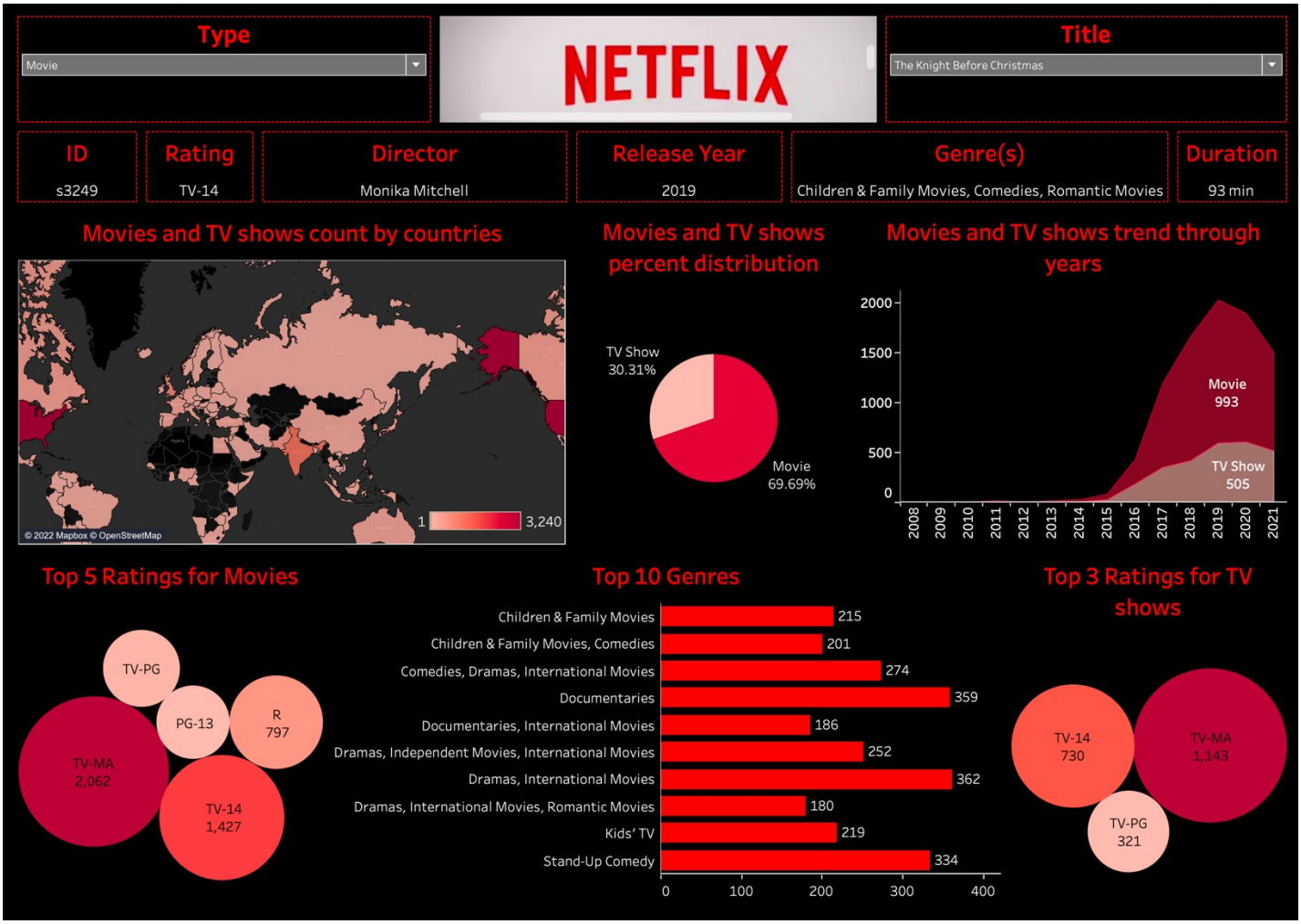
Critical analysis of the visualisation

The visualization is very useful for people looking to understand trends and distribution of movies and TV shows on Netflix. It can be used by producers, writers, and directors to find popular genres and top ratings of shows and movies. Although I tried to keep the visualisation minimal and to the point, the dashboard still might be a little distracting due to the use of red colour. The sole reason to use this sole was to match the colour palette of Netflix and make it more visually appealing.

References

- https://www.kaggle.com/datasets/shivamb/netflix-shows
- https://www.kaggle.com/code/manuelcabrerag/netflix-data-cleaning-analysis-and-visualisation
- Lecture notes available on Blackboard

Screenshot of dashboard



Link for the video and tableau file:

- https://drive.google.com/file/d/1uuLeIeyvuVSMNYIW2k6g6wsyxXNB9sCF/view?usp=sharing
- https://github.com/singhrohaP/Data-Visualisation/tree/main/Netflix%20dashboard