DATA VISUALIZATION DSC-530

Phase 1A and Phase 1B Project Report

University of Massachusetts, Dartmouth

Professor: Amir Akhavan Masoumi

Submission Date: 11/07/2022

Group 21
Sriram Jagadeesan
Manish Yadav

Project Title: <u>Netflix - Data Exploration and Visualization</u>

What?

        Netflix is one of the largest providers of online streaming services with 223 million paid memberships in over 190 countries enjoying TV series, documentaries, feature films and reality shows across a wide variety of genres and languages. The platform has successfully established their presence in the entertainment sector and people's lives. Personalization and customization of products or services are significant features of Netflix.

Why?

        The aim of this project is to analyze the content released by Netflix and gain some insights into how the OTT platform is using the viewers content consumption (over a period) and changing business strategies of the company based on it. Traditionally the streaming service has focused mainly on movie content, but from a bird's eye overview of the data, TV shows and series content has grown rapidly. This project will showcase different insightful visualizations which will include three areas of data on Netflix: geographic, time based and popularity.

How?
1. **Viz1: Geographic Impact**: The first dashboard's data focused on Netflix's geographic impact. Using D3.js, a geospatial map will be produced. We will show which countries have what kind (Movies/TV shows) of content and how much of it they consume. To make it more interactive and informative, filters will be added.

2. **Viz2: Timeline:** The second dashboard is to create an interactive line graph with the purpose of showcasing the growth of Netflix over the years from the very beginning.

3. **Force Directed Graph:** It is used to visualize the connections between objects in a network. The graph shows a set of Genre as nodes, with links between a pair of genres indicating that audience who liked one of them are likely to enjoy the other.

4. **Viz3: Popularity:** The third one was based on showcasing the popularity of Netflix worldwide. We will also include a dataset with the costs of plans in each country. We will use bar chart for subscription rates comparisons for each country.

Design:

1. **Searching and Cleaning up the data:** The data on Netflix was also not clean, so we cleaned that data on open refine, then split countries into different cells so that D3 can easily detect those countries and give the output. We also added two other datasets with the main one to add more details to the final visualization.
2. **Brainstorming:** The second and important step after searching and cleaning the data was to think about how it will look and what all features it will have. We did some brainstorming before directly jumping to the software. Also, we chose colors from the Netflix logo as the primary color palette.
3. **Geographic Impact**: The first dashboard's data focused on Netflix's geographic impact. To make it more interactive and informative, filters were added. As we hover and click on each country, we get the Country name and Number of Netflix Titles. Using this we can get a Geospatial interface to visibly understand the country wise no. of titles.
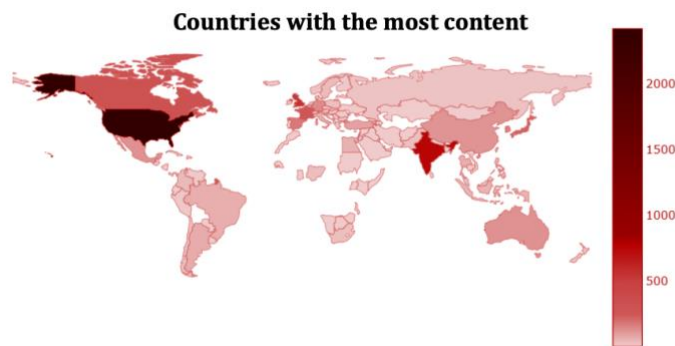

Fig 1:

4. **Time-Series Analysis:**

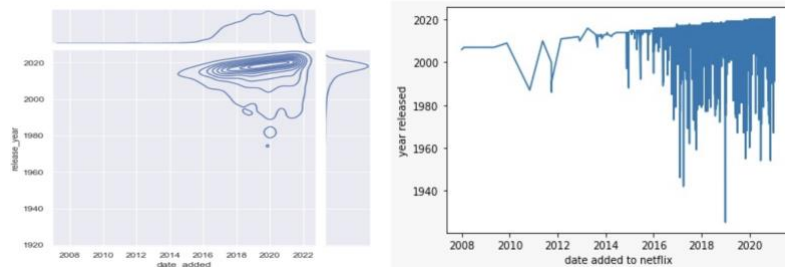Fig 2 & 3:
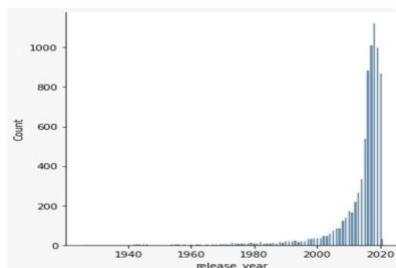

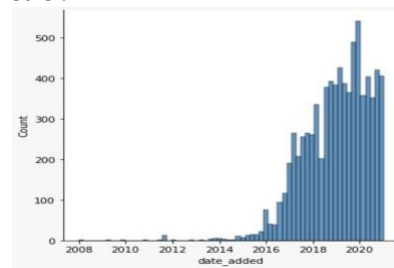Fig 4   Fig 5:

- **Fig 2 &3 : The visualization depicts the trend at which content added to Netflix has been diversified based on year of release.**
- **Fig 4: Depicts the year of release of the content that has been added to Netflix over the years.**
- **Fig 5: Visualization depicts the trend at which content added to Netflix has been diversified based on year of release.**

We created these visualizations using python and seaborn. First filtered the data to get rid of null values, duplicates. Then we used seaborn to create the above graphs. The above visualization shows the distributions of additions to Netflix and their release year separately. By this we see that content available in Netflix has an exponential growth initially indicating the tremendous growth of the company followed by a reduced growth with large variance year on year indicating the increase in competition from rival companies. Whereas in the graph depicting the release year we can see a straight upward trend with a peak in 2017 and then a downward trend. The above visualizations shows that initially Netflix added content that is fresh in the market with a trend of purchasing more and more old content relative to the new ones.

5. **Force Directed Graph:** It is used to visualize the connections between objects in a network. The graph shows a set of Genre as nodes, with links between a pair of genres indicating that audience who liked on one of them are likely to enjoy the other. We first created an svg container then created forced layout using D3, loaded the data and created nodes and finally linked the nodes and appended.
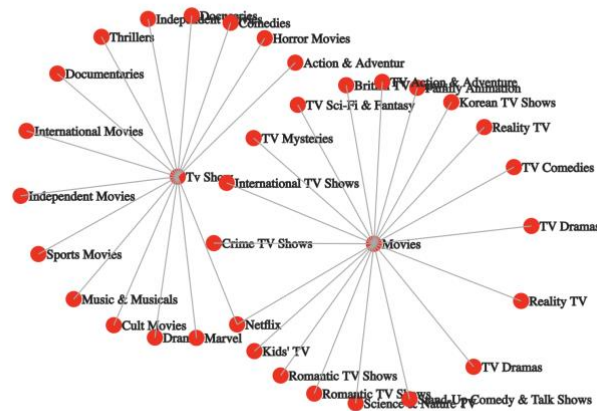

Fig 6

6. **Bar Graph:** The fourth one was based on showcasing the subscription rates of Netflix worldwide. This was not an advanced graph, but we believe it accurately represented the data, so I used it. Because there isn't a lot of information and it's easy to understand, we didn't add any filters. This graph helps in showcasing the Subscription Rates in Major countries. We can see that it's the Highest for USA where even the content production is more compared to other regions.
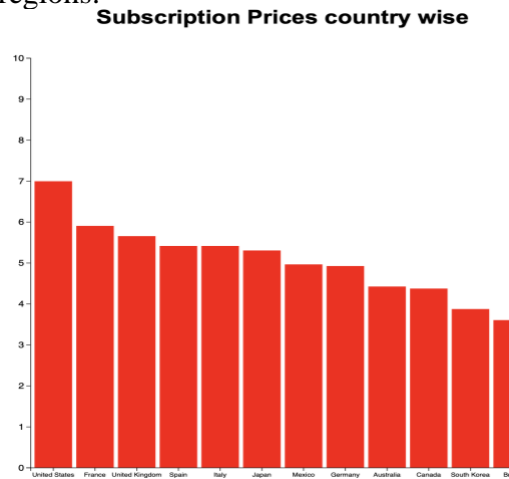

Fig 7

7. **Homepage:** We wanted to design a dashboard where users could easily find information and navigate. The goal was to make it visually appealing while also avoiding the need for additional navigation instructions.
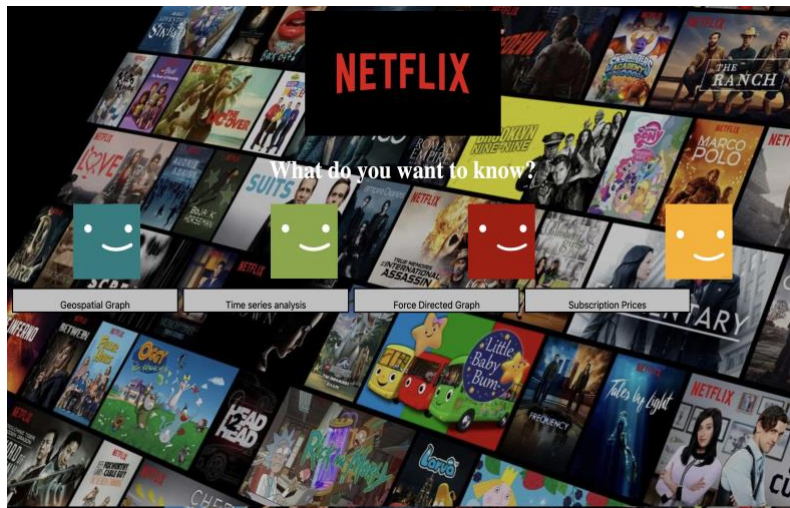

Fig 8

Relation between Viz 1 and 2 3 and 4: All the four visualizations are interconnected and seeing these visualizations, user should be able to understand how Netflix has grown in terms of business, content creation capacity, variations in type of content being consumed

and how Netflix uses all these statistical data to enhance there overall reach to customers and gain profit.

Conclusion and Future Work :

- Data Analysis is a fundamental step to address the various needs of a client in any professional spectrum. The varied range of insights that can be derived from a data is itself.
- Overall, the visualization was successful in conveying significant Netflix information with a story. The dashboards made that quite clear. We can infer from all the Graphs that USA is leading country in terms of content creation, subscription costs, number of subscribers and so on.By clicking on the homepage image below, you can navigate to the final result.
- We made the webpage easy to navigate in the throughout the visualization as the whole visualization is as the UI of the original application.
- We could add more information in the future, such as IMB ratings and data on how many people share a single subscription, add the link of the data resource in each dashboard.

Datasets:

https://www.kaggle.com/niharika41298/netflix-visualizations-recommendation-eda
https://www.kaggle.com/datasets/shivamb/netflix-shows
https://www.comparitech.com/blog/vpn-privacy/countries-netflix-cost/

Webpage URL: http://127.0.0.1:5500/homepage.html