Netflix Data Analysis Project Plan

## Fe

## **Step 1: Data Collection and Cleaning:**

Link to dataset: <https://www.kaggle.com/datasets/shivamb/netflix-shows/data>

About the Data: Netflix is a renowned streaming service that offers a wide variety of award-winning TV shows and Movies that individuals can binge if they purchase a membership with the company. The streaming service was first founded in 1997 and launched in April of 1998. This dataset consists the listing of all the movies and TV shows available on Netflix, along with details such as cast members, directors, ratings, release year, and duration.

Methodology: The dataset was originally downloaded to Microsoft Excel for brief data cleaning and then later uploaded to Microsoft SQL Server for further exploratory analysis. After discovering patterns and KPIs, I will move the data to Tableau for a detailed visualization.

Questions that Need to be Answered:

1. Does Netflix have more focus on TV Shows or Movies?
2. What content is available in different countries?
3. What is the network analysis of Actors/Directors?

Data Cleaning Process:

* Excel
  + Noticed that duration has shifted to rating column for show\_id s5542, s5795, and s5814 so I shifted them to the correct column.

## **Step 2: Data Exploratory in SQL:**

Notebook:

1. Early part of data exploratory I wanted to know the amount of data that I am working with and then showing then breaking it down further by showing the total number of Movies and TV shows that are contained in the dataset.

A screen shot of a computer

Description automatically generated

A screen shot of a movie

Description automatically generated

1. To show what contents are available in different countries, I wanted to first see what companies the biggest producers are.

A screenshot of a data

Description automatically generated

2a.) I further then showed which TV/Movie Ratings were most common throughout the dataset.

A screenshot of a table

Description automatically generated

2b.) I then took the Top 3 countries (US, India, and United Kingdom) and compared what TV/Movie ratings were available in each country.

A screenshot of a table

Description automatically generated A screenshot of a table

Description automatically generated A screenshot of a table

Description automatically generated

1. To look at the network relationship between directors and cast members, I first looked at which director has produced the most films in the dataset.

A screenshot of a computer

Description automatically generated

3a.) I then looked at which cast member(s) were most common.

A screenshot of a computer

Description automatically generated

3b.) For the next step I looked at the top 5 producing directors and see which cast were associated with their movies/TV shows

A screenshot of a computer

Description automatically generated

3c.) For further analysis I took the highest connection between director and cast member and looked at what type of film they produced the most.

## **Step 2: Visualization in Tableau:**

Contents:

* Movie/Show Distribution
* Heatmap of countries where films were produced.
* Tree map of most common Genre
* Count of films released throughout the years.
* Top Directors
* Top Cast members

Link to Dashboard: <https://public.tableau.com/app/profile/cedrick.cavan/viz/NetflixDashboard_17083612145830/Dashboard1>

## **Step 3: Analysis/Key findings:**

1. Netflix has more of a focus on movies
2. The top 3 countries that produce the most amount of TV shows/Movies are the United States, India, and United Kingdom in that order.
   1. United States and India had more movies filmed in their country. United Kingdom almost had equal amount of TV shows and movies.
3. Rajiv Chilaka directed the greatest number of films in this case the most TV shows.
4. Most films were released to Netflix in the late 2000’s
5. Highest number of Genre produced were Dramas, International Movies

## **Step 4: Conclusion:**

Revisiting Questions that need to be answered:

1. Does Netflix have more focus on TV Shows or Movies?
   * According to this dataset, Netflix has more of an emphasis on Movies compared to TV shows.
2. What content is available in different countries?
   * Based on the top 3 producing countries, the United States allows all types of contents while India and United Kingdom only allow a few ratings in their country
3. What is the network analysis of Actors/Directors?
   * Rajiv Chilaka and cast members Vatsal Dubey, Julie Tejwani, Rupa Bhimani, etc. have the most connection between director and cast members.
     1. This is due to majority of films produced were an animated Indian movies