# Proposal for Project: Cleaning, analysis and Visualization of Netflix Data

By: Saurav Purkayastha

1. Introduction:

The proposed project aims the delivery of an easy and convenient picturization of a large dataset with the help of graphs and charts. Netflix is a popular streaming service that offers a vast catalog of movies, TV shows, and original contents. Hence, this project focuses on the delivery of significant data visualization for the trends observed in the ratings and consumption of content in this large online platform. By taking care of all the rules and guidelines, and the documented instructions delivered, this project has been successfully carried out for the production of valuable insight.

2. Targeted beneficiaries

* Marketing teams of Netflix: The marketing team can receive large benefits from the data visualization especially while performing audience segmentation for the company
* Team focusing on Content Strategy: These teams can use the data visualization done through this data to identify the most popular genres of TV shows and movies while making vital production decisions.
* The Product team of Netflix: This team will benefit from the information to create better user interfaces. Better user interfaces will allow increased precision in recommending content catering to audience choices.
* Production and filmmaking houses: These companies can benefit from the data visualization because they will be able to analyze the changes that have taken place in audience preferences.
* Directors and writers in the film industry: These stakeholders in the industry will be also benefited in terms of analyzing the changes that have taken place in audience preferences.
* Students pursuing media study courses: this data can be helpful for students pursuing media study courses for getting a glimpse into the effects of streaming platforms on the consumption of media.
* Culture analysis experts: This project can prove beneficial for such analysts to understand the shifts in cultures in correlation with viewing trends.

3. Resources used

Data source: The primary data source that will be used in this case is the “netflix1.csv” dataset. This dataset contains large amount of data focusing on TV Shows and movies released in Netflix. This dataset contains categories such as ratings of movies and TV shows, the directors, the respective countries, durations and genres.

Other sources: Cleaning of data and removal of duplicate data will take place through the use of Google Colab software and further data visualization will be carried using Tableau software.

4. Intended goals

* To produce an idea about the large range of content that is produced in the filmmaking industry.
* To develop an understanding of the frequency of viewership in different countries by focusing the amount of TV shows and movies produced.
* To understand the performance of several content types. This analysis will allow us to develop knowledge regarding the most popular genres or formats of movies and TV shows over a period of time.
* To understand trends such as the top directors who make impactful content.
* To deliver information to several teams and production companies etc about the types of visual narratives that can be developed for viewer experience.

5. Conclusion

In conclusion, the examination of Netflix content data has yielded important information about user involvement trends, content trends, and viewer preferences. Through the analysis of this extensive dataset, we have discovered important genres that appeal to the audience, allowing us to make better judgments about the production and acquisition of content.  In addition to guiding strategic content choices, these data will help create a more customized and interesting entertainment platform for an audience that is spread globally.