Brandilyn Hall
Yvonne Martinez
Lyndah Mupfunya
Kwadwo Asante

**Project 1 Proposal _ Group 3**

We will be focusing on Netflix TV shows and movies using the below data set. We have seen through Kaggle various analysts use this data set, which gives us confidence we can create our own unique analysis.

Data Set: https://www.kaggle.com/shivamb/netflix-shows

Inspiration: Netflix is a popular streaming platform for both movies and tv shows. We believe that through Covid the platform has become more and more used and are interested in analysis relating to years, countries, ratings, etc. Below we have added two inspiration links we found.

- https://www.kaggle.com/rsesha/autoviz-on-netflix-dataset?scriptVersionId=83019093
- https://medium.com/geekculture/data-analysis-of-netflix-movies-imdb-rating-using-jupyter-notebook-d923186da6c7
- https://www.datacamp.com/workspace/templates/dataset-python-netflix-movie-data

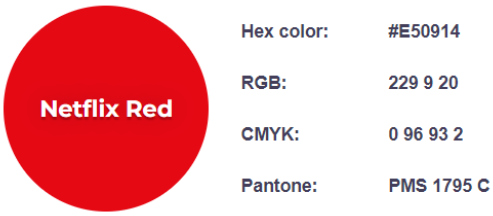
Some of the initial questions we wanted to ask of the data are:

- Which countries have contributed most movies in recent years?
- What country has the longest movie duration time?
- What genre has the most movies created?

Some of the initial thoughts on regression or correlations are:

- What is the correlation between the date of release and duration of the movie/show? Are current movies longer than older releases? We predict that movies currently are longer in minutes.
- What is the correlation between maturity ratings and the duration of the movie? Our assumption is based off children's attention span length.

Color theme: Because we are referencing Netflix's data, we would like to have a theme that includes their current logo colors of red (#E50914), black (#000000), and white (#FFFFFF). Some of the chart types we would like to include are bar charts, scatter plot, histogram, violin, line chart, and boxplots.



Brandilyn Hall
Yvonne Martinez
Lyndah Mupfunya
Kwadwo Asante