**Netflix EDA & Recommendation System**

Saad Alkhathlan

**Abstract**

Netflix is a streaming service that allows their users to watch TV shows & Movies on internet-connected devices, Netflix have been growing over the last few years with its popularity and content. The goal of this project is to build an EDA or a storytelling through Netflix data alongside a recommendation system.

**Design**

The purpose is to helps Netflix users find similar content to what they like by building content-based recommendation system using two models TF-IDF and Count Vectorizers then filtering it on multiple metrics, project also helps producers and managers with different graphs and visuals to make better decisions on what kind of content the customers want, at what time it should be added, and which countries has more interest in a specific moving rating such as kids, teens or adults movies.

**Data**

The dataset contains 8807 data points with 12 features for exploratory analysis, most of the features I’m using are categorical such as, type (Either a Movie/TV Show), title (Name of the show), director (Name of the director of the show), cast (Name of actors and other cast of the show), country (Name of countries the show is available to watch on Netflix), listed\_in (show genre), with only one integer feature which is release\_year (Release year of the show)

**Algorithms**

*Feature Engineering*

Creating more features to make the visuals more appealing such as target\_ages of the movie/TV show, genre, and a first\_country feature that only shows the first country mentioned in the production of a movie/TV show, since many productions have several countries listed, this feature will avoid results being unbalanced.

*Models*

Content-based recommendation system by using Count Vectorizer and TF-IDF models then filtering it on multiple metrics: title, cast, director, listed\_in and plot.

**Tools**

* Programming Language: Python.
* Environment: Jupyter Notebook.
* Libraries: NumPy, Pandas, Matplotib, Plotly, and Seabon.
* Models: Scikit-learn; Count Vectorizer & TF-IDF.

**Communication**

Presentation that includes visuals for storytelling the data

Graphical user interface, text, application

Description automatically generated

Chart

Description automatically generated