**Netflix TV Shows / Movies Analysis**

**STEP ONE :**

**Reading and Exploration of our Data :** This dataset contains information on all of the movies and TV shows available on Netflix as of May 2022. In addition to basic information such as title, release year, andruntime, the dataset includes data on the cast and crew, IMDB score and number of votes, genres, production companies, and more.

**STEP TOW :**

**Data Modeling :** after exploring the data it’s time to model it **but** we found that our data is very simple to model which lead us to the third step.

**STEP THREE :**

**upload in power bi :** in this step i uploaded the data in power bi and then transform it on **power query.**

**STEP FOUR :**

**Data cleaning using power query:** in this step i started cleaning and makes the data looks better, **first** i have checked for columns data types in all files in the dataset, **second** i have checked for any blank rows to remove it , **third** i have removed a column called index from every table in the dataset because i found that column is a useless column , **fourth** i started checking for any **logical error** that may case any error when i visualize in power bi and here **some** of this errors 1- in file **raw\_titles** which contains an information of every movie/show in our dataset i have found that in column title there are 1 raw blank which is not acceptable (no movie/show has no name) so i had to remove it since it’s only 1 row that will not affect our data too much , 2- in the same file i have found many rows in age\_certified column is blank so in this case i can’t delete them all so i have replaced the blank with a new category i named it **Not Certified** , 3- in runtime column i have found some values with 0 !! fortunately they are few so we can delete them also , 4- in column genres i have found a genre with ‘[]’

in this case i have replaced it with unknown to indicate that we dont know this movie/show genre ,5- in column country we have the same problem in 4 so i replaced ‘[]’ with unknown

since our data is mostly categorical we don’t have many choices to do with it ,anyway our table now is good and fortunately the other tables is almost cleaned and nothing i can do for it.

**STEP FIVE:**

**check for any error in tables relation and validate it:** like i mentioned above our data is very simple to model i have already 6 files but i used 4 in modeling and 2 for visualizing ,our main files is raw\_titles file and raw\_credites file which has an informations of actors and directors and the id of the movies/shows they work on it,there are a **one-to-many** relation between them and when i uploaded the 2 files power bi automatically establish the relation between them ,there also 2 files i used in our model which is best movie by year and the second is best show by year, i used a merge query to merge this tow files with the raw\_title file since they have an information that we don’t know in raw\_title file.**after that** i have started visualizing and obtain insight from this data.