

Title: Data Analysis of Netflix Dataset

Introduction:

Netflix is the world's largest online TV show and Movie streaming platform. Netflix service is widely offered in a number of countries, including the United States, United Kingdom, India, Japan, South Korea, Japan etc. Netflix headquarters is in the United States. Netflix service is available only to those individual who pay a monthly subscription fee. Initially, Netflix only served as a DVD retailer and DVD rental business. After 2010, Netflix launched its streaming service for the first time and due to continuous improvements and growth this platform became popular among the people. Currently, Netflix has become a highly popular online streaming platform and is constantly releasing latest TV shows and movies for the audience. The Netflix data was available on "kaggle" which is an open-source platform. The primary source of this data is "Flixable", which is a third-party platform for Netflix.

In this assignment, I have analyzed the Netflix data with the help of different python libraries. Python is a popular programming language due to its flexibility, adaptability and large extension libraries. Data Analysis in python is relatively easy as compared to other languages. Data analysis is mainly done to get key insights from the data. In this analysis I have applied different skills and techniques that I have learned so far in this course. I have done following tasks:

1. Downloaded the Netflix data and imported it in python.
2. Data Pre-processing and cleaning.
3. Data visualization using different libraries and parameters.

Dataset description:

The Netflix dataset features include:

- ❖ "cast" – cast members of TV shows or movies.
- ❖ "country" – released location of TV show or a movie was released.
- ❖ "data_added" – date of release of a TV show or movie on Netflix.
- ❖ "release_year" – year of release of TV show or movie.
- ❖ "rating" – parental guidance ratings for movies and TV shows.

- ❖ "duration" – duration of a TV show or a movie.
- ❖ "listed_in" - genre of a TV show or movie.
- ❖ "description" – a short plot of the movie or TV show.

Python libraries:

The major Python libraries that I have used in this analysis are pandas, seaborn and matplotlib.

Data Pre-processing and Cleaning:

Data pre-processing and cleaning is very important step in data analysis. Accurate results are possible only if the data is cleaned. The raw Netflix data contained many missing values so it was preprocessed and cleaned using different python libraries prior to Data visualization.

Data Visualization

The following visualizations were made after analyzing the Netflix data in python

1. Pie plot:

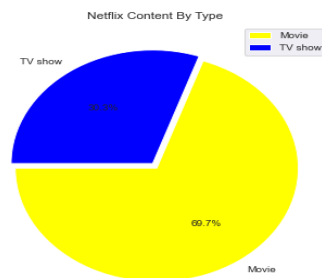


Fig. 1. Netflix content by type

2. Time-series plot:

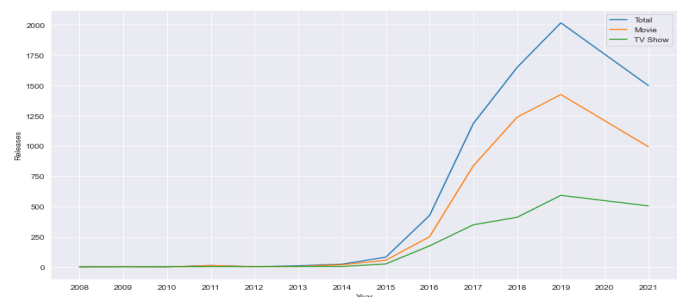


Fig. 2. Type of content produced on Netflix from 2008 to mid of 2021

3. Bar plot:

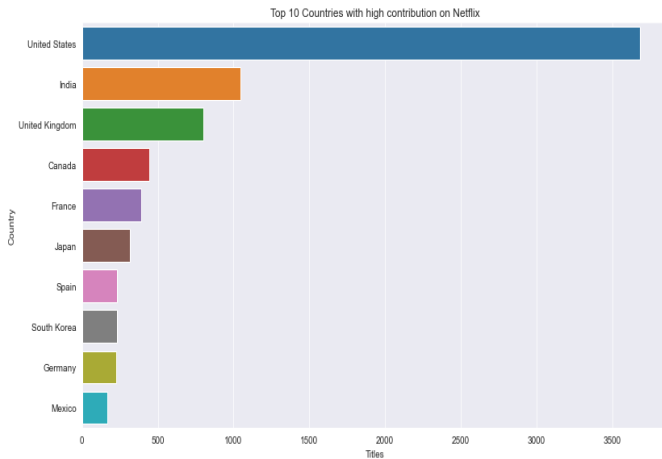


Fig. 3. Top 10 countries with highest contribution on Netflix

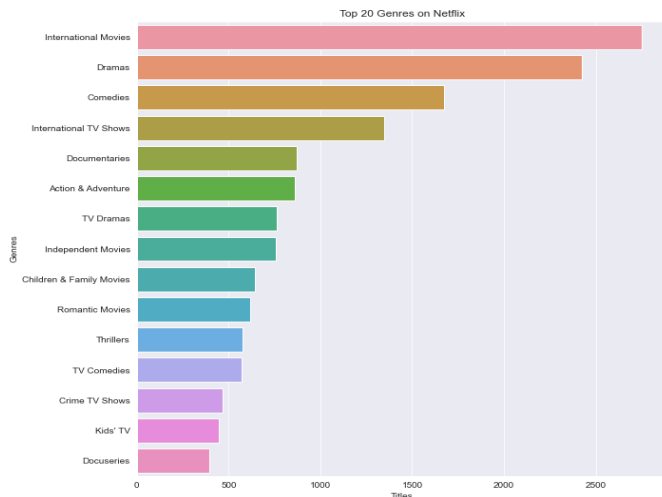


Fig. 4. Top 20 genres available on Netflix

Conclusions:

The following conclusions were made by analyzing Netflix dataset.

1. The major content that is available online on Netflix is movies (i.e. 70%).
2. After going online in 2010, Netflix started gaining interest from the people gradually. From 2014, Netflix started producing more content amid the interest of the audience. The time line graph also shows that Netflix produced the

highest content during the COVID-19 era (i.e. 2019-20).

3. The bar plots show that United States leads the charts of content contribution on Netflix followed by India and United Kingdom.
4. The bar plot also shows that the top 2 genres available on Netflix are International Movies and Dramas.