Capstone Project Submission

Team Member's Name, Email, and Contribution:

Team Member's Role: -

- **Kanika Raj**

  Email-rajkanika214@gmail.com
    - o Data Understanding
    - o Data Cleaning
    - o Data Visualization
    - o Multivariate analysis
    - o Natural language processing
    - o Hyperparameter tuning
    - o Different clustering model Building
    - o Research and technical documentation

Please paste the GitHub Repo link.

GitHub Links
Kanika -> https://github.com/kanika214/kanika214-NETFLIX-MOVIES-AND-TV-SHOWS-CLUSTERING.git

Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Netflix, Inc. is an American subscription streaming service and production company. Launched on August 29, 1997, it offers a film and television series library through distribution deals as well as its own productions, known as Netflix Originals.

As of March 31, 2022, Netflix had over 221.6 million subscribers worldwide, including 74.6 million in the United States and Canada, 74.0 million in Europe, the Middle East and Africa, 39.9 million in Latin America and 32.7 million in Asia-Pacific.[11] It is available worldwide aside from Mainland China (due to local restrictions), Syria, North Korea, Kosovo, Russia (due to the 2022 Russian invasion of Ukraine) and Crimea (due to US sanctions). Netflix has played a prominent role in independent film distribution, and is a member of the Motion Picture Association (MPA).

Netflix can be accessed via internet browser on computers, or via application software installed on smart TVs, set-top boxes connected to televisions, tablet computers, smartphones, digital media players, Blu-ray Disc players, video game consoles and virtual reality headsets on the list of Netflix-compatible devices. It is available in 4K resolution.[18] In the United States, the company provides DVD and Blu-ray rentals delivered individually via the United States Postal Service from regional warehouses.

This dataset consists of tv shows and movies available on Netflix as of 2019. The dataset is collected from Fixable which is a third-party Netflix search engine.
In 2018, they released an interesting report which shows that the number of TV shows on Netflix has nearly tripled since 2010. The streaming service's number of movies has decreased by more

than 2,000 titles since 2010, while its number of TV shows has nearly tripled. It will be interesting to explore what all other insights can be obtained from the same dataset.

Conclusion -:

1. Director and cast contain a large number of null values so we will drop these 2 columns.

2. In this dataset there are two types of contents where 30.86% includes TV shows and the remaining 69.14% carries Movies.

3. We have reached a conclusion from our analysis from the content added over years that Netflix is focusing movies and TV shows (Form 2016 data we get to know that Movies is increased by 80% and TV shows is increased by 73% compare)

4. From the dataset insights we can conclude that the greatest number of TV Shows released in 2017 and for Movies it is 2020

5. On Netflix USA has the largest number of contents. And most of the countries preferred to produce movies more than TV shows.