Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
1. Mohammad Jibran jibranmd9@gmail.com 1. Data Cleaning & Data Visualization. 2. EDA 3. Create Cluster for our data. 4. NLP Data. 5. Data Preprocessing. 6. Silhouette score. 7. Elbow Method. 8. DBSCAN. 9. Get Recommendations. **2. Siddhi H Thakur** siddhi.thakur04@gmail.com 1. Data Cleaning & Data Visualization. 2. EDA 3. Create Cluster for our data. 4. NLP Data. 5. Data Preprocessing. 6. Silhouette score. 7. Elbow Method. 8. DBSCAN. 9. Get Recommendations.
Please paste the GitHub Repo link.
Github Link:- https://github.com/DSJibran/NETFLIX-MOVIE-AND-TV-SHOWS
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

SUMMARY NETFLIX-MOVIE-AND-TV-SHOWS

Nowadays we all are connected with social media and watch lots of movies and tv shows via those subscription platforms. Netflix is one of those platforms which puts a lot of effort into entertaining us with its contents. Netflix is an online streaming service that allows subscribers to watch television series and movies on their personal computers, mobile devices, and gaming consoles. This is mainly because watching movies at home is cheaper and more convenient than going to the theater.

The dataset has two types of content: movies and TV shows. It has records from 2008 to 2021, there are 7770 entries and 14 columns.

We cleaned the data to check NULL values, duplicate values and also extracted date columns to find insight from the data. After data cleaning, we performed EDA to make accurate analysis on our data.

In this project we do lots of Exploratory data analysis to get the understanding of the data and how the content is distributed in the dataset. We got to Know about top ranked directors and actors, profoundly say
Indian Actor Anupam Kher is ranked first with 38 overall appearances in TV shows and films.

Countries producing most no of contents and check more popular genres of different countries.

Unique title and description word we used maximum no. of times.

We perform pre-processing for removing the punctuation and removing stop words, also used stemming to reduce words to their basic form, and standardizing data before starting the development of the model.

We use Hypothesis Testing to check if there is any relation between year and type.

We used different clustering algorithms and then checked the model performance using Silhouette's Score, Elbow Method to find the number of clusters, to identify the best fit model, DBSCAN and Recommendation on Movie and tv shows.

Netflix has increasingly focused on TV rather than movies in recent years.

The number of movies on Netflix is growing significantly faster than the number of TV shows. Because of covid-19, there is a significant drop in the number of movies and television episodes produced after 2019. But it still seems like a bright future for Netflix with ever increasing interest in the audience.