

Q. Which is an interesting data set you discovered recently? Why is it your favorite? No datasets on Kaggle, please.

Ans: Recently I worked on Netflix movies and TV shows dataset. This contains movie/show name, category, director, cast, release year, genres and many other columns. This data set is kind of semi-structured database. Apply some feature engineering techniques, some NLP on it and predict the number of clusters with which can belong. As well as recommended top 5 movies/shows of the particular movie/show.

I feel this data is more interesting cause there is first challenge in front of me is to convert unstructured columns to structured. Then visualize this data in different ways to get important information from this whole dataset.

This dataset consist of almost 80% data is in form of categorical so, another challenge is to convert that into numerical by feature engineering and NLP.

You can show my project “Netflix movies and tv shows clustering and recommendation system” on GitHub. Link is given below....

Link: [Netflix movies and tv shows clustering and recommendation system](#)