Recommender Movies System

1 Problem Statement

Recommender systems are among the most popular applications of data science today. They are used to predict the "rating" or "preference" that a user would give to an item. Almost every major tech company has applied them in some form. Amazon uses it to suggest products to customers, YouTube uses it to decide which video to play next on autoplay, and Facebook uses it to recommend pages to like and people to follow. There are also popular recommender systems for domains like restaurants, movies, and online dating. Recommender systems have also been developed to explore research articles and experts, collaborators, and financial services. YouTube uses the recommendation system at a large scale to suggest you videos based on your history. For example, if you watch a lot of educational videos, it would suggest those types of videos.

2 Project Objective

The main objective for this project we will build a movies recommender system that we have to consider various metrics like keywords, cast, most-watched genre, director, and inject them into a machine learning model which then produces what the user might like to watch next. We will check the cosine similarity and the predict the movie that user might like to watch next.

3 Dataset

All experiments are performing on the following dataset provided by MovieLens Dataset:
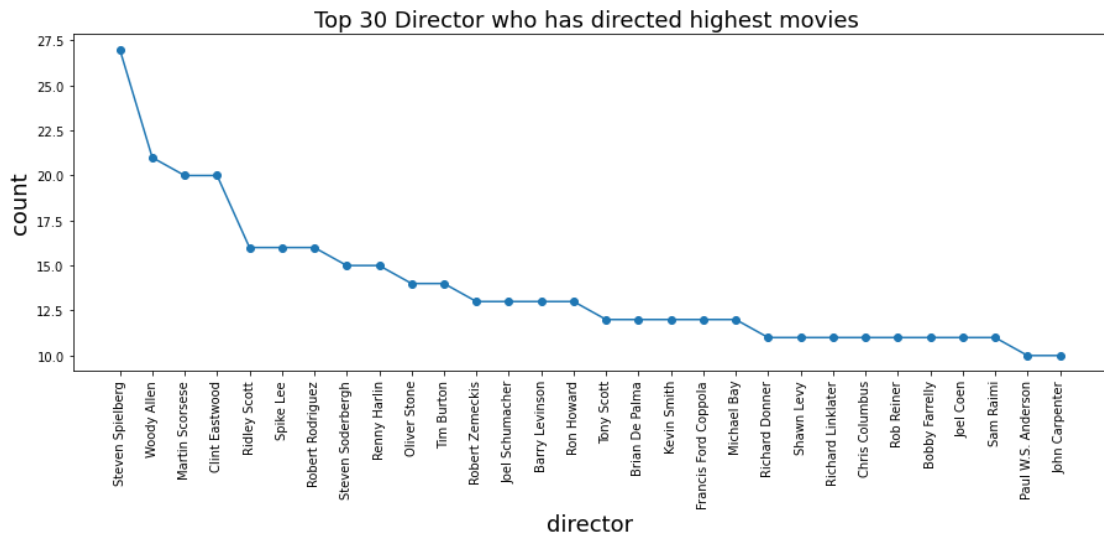
1. MovieLens Data Set [1]: The dataset consists of movies released on or before July 2017. This dataset captures feature points like cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts, and vote averages. These feature points could be potentially used to train your machine learning models for content and collaborative filtering.

4 Exploratory Data Analysis

Exploratory data analysis is performed to gain different useful information and hidden insights from dataset. In this section different statistical techniques have been used to gain insights and then being visualized into appropriate charts and plots.
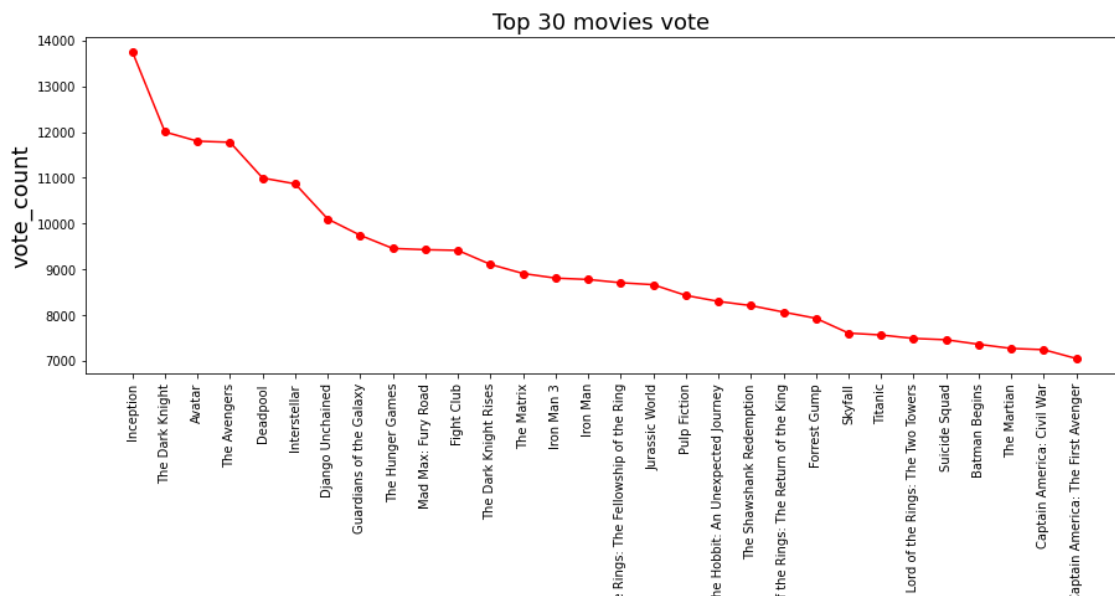
4.1 Director Distribution

This analysis is about the distribution of director in the movies dataset. The line chart representing this information is shown below.

Top 30 Director who has directed highest movies

From above chart it is clear that in most of the movies directed by Stephen Spielberg and then Woody Allen.

4.2 Movies Vote Distribution

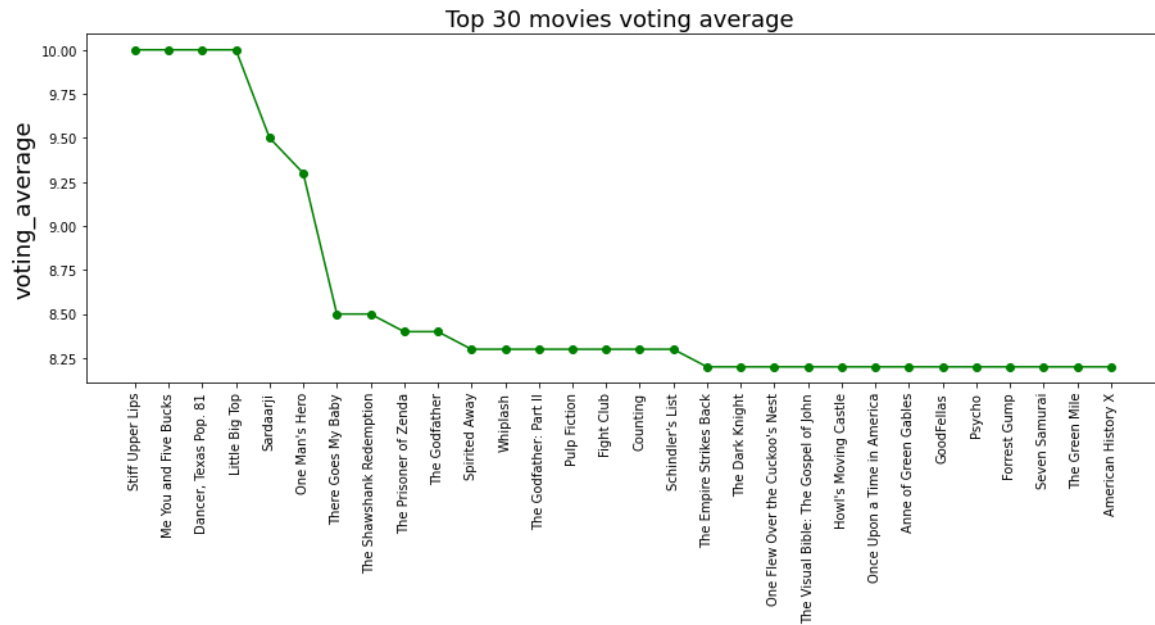This analysis is about the distribution of movies vote. The line chart representing this information is shown below.

Top 30 movies vote

From above chart it is clear that in most voted movies are Inception and The Dark Knight.

4.3 Movies Voting Average Distribution

This analysis is about the distribution of movies voting average. The line chart representing this information is shown below.

Top 30 movies voting average

From above chart it is clear that most voting average movies are Stuff Upper Lips, Me You and Five Bucks, Dancer Taxes pop and Little big top.

5 Data Preprocessing

Firstly, we will combined the four features, features are cast, genre, keywords and director and then handle the missing values with fill the empty string. We will use Counter Vectorizer for feature selection and then check the cosine similarity between the combined features.

The Movies which will be more cosine similarity to others movies so, we will recommend to the user for watch. You can check the code for this recommender system on my Github[2].

6 Results

```
# lets try to get the similar movies against Avatar movie
recommended_movies('Avatar',10)
```

Top 10 similar movies to Avatar are:

Guardians of the Galaxy
Aliens
Star Wars: Clone Wars: Volume 1
Star Trek Into Darkness
Star Trek Beyond
Alien
Lockout
Jason X
The Helix... Loaded
Moonraker
Planet of the Apes

```
# lets try to get the similar movies against The Dark Knight Rise movie
recommended_movies('The Dark Knight Rises',10)
```

Top 10 similar movies to The Dark Knight Rises are:

Batman Begins
The Dark Knight
Amidst the Devil's Wings
The Killer Inside Me
The Prestige
Batman Returns
Batman
Batman & Robin
Kick-Ass
RockNRolla
Kick-Ass 2

References

[1]   https://www.kaggle.com/grouplens/movielens-20m-dataset

[2]   https://github.com/Talha1818/Pakistan-Freelancing-Training-Program--Data-Sceince-Project-Recommender-Movies-System-