



# MOVIE RECOMMENDATION SYSTEM



BY Brian Kipruto Ngeny



# Introduction.

---

The movie industry has grown to be a multi-billion industry within the last century and with the advent of technology, It is expected to accelerate in growth.

This project is aimed at developing a movie recommendation system that will provide recommendations of movies specific to users based on their rating of other movies. This should help in eradicating the hurdle of choosing and deciding on their own what to watch from the numerous amounts of movies on offer today.

# Business Understanding.

---

The dataset was provided by GroupLens which is a research group in the Department of Computer Science and Engineering at the University of Minnesota.

This project's target audiences are, streaming sites and movie enthusiasts who will no longer have to worry about what to watch next because the resulting system will be able to recommend movies they might be interested in watching based on their rating of other movies.

# Problem Statement.

---

With so many movies being released, the challenge now is to select what to watch. These has prompted me to come up with a solution in the form of a recommendation system for movies. By so doing, I hope to solve the problem faced by movie enthusiasts and online streaming companies by developing a system that recommends movies based on the rating given to other movies by users.

# Objectives.

---

The primary objective of these project was to come up with a system that would be able to recommend movies to a user based on the ratings they have given to other movies.

The secondary objective was to build a model that uses content based recommendation to recommend movies to a user that are similar to a particular movie within the dataset.

# Data Understanding.

---

The dataset contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018. This dataset was generated on September 26, 2018. The data folder contains `ratings.csv`, `tags.csv`, `movies.csv`, and `links.csv`.

# Data Cleaning.

---

The steps to clean the data acquired were;

- I) Drop the columns we don't need.
- II) Identify any missing values and drop the rows.
- III) Engineer features or new columns where necessary.
- IV) Merge datasets if necessary.

# Exploratory Data Analysis.

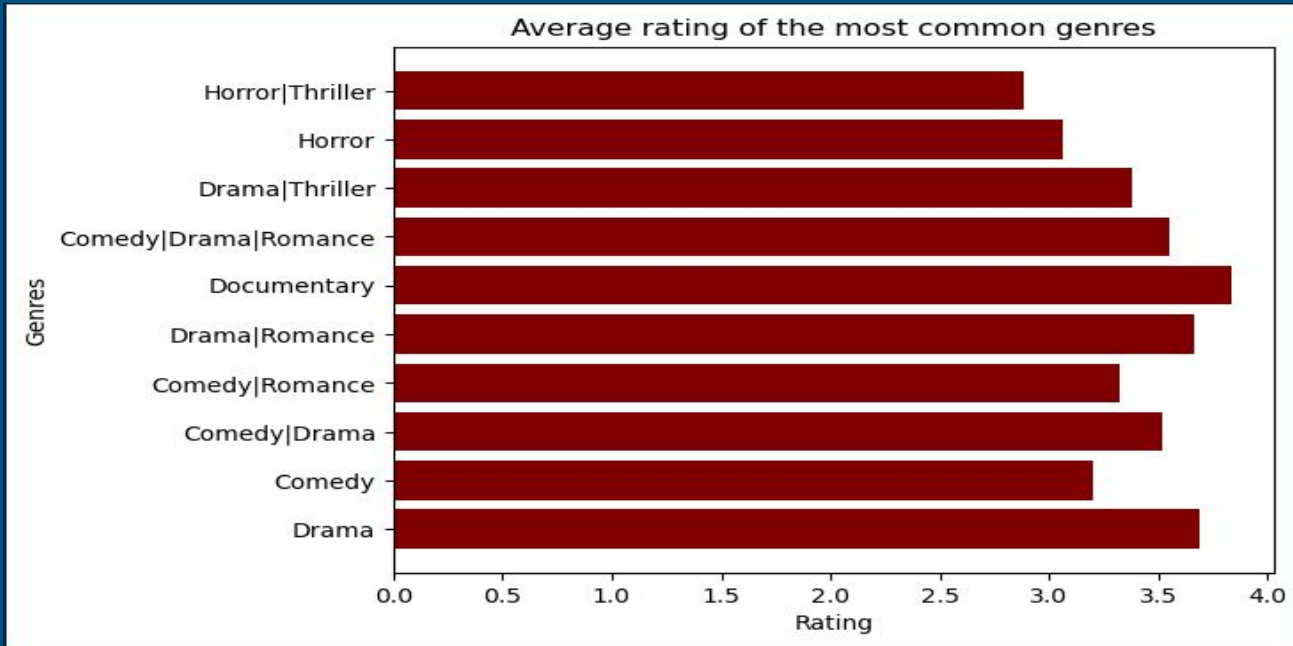
---

First, I analyzed the `___movie_data___` dataset to discover the top ten genres of movies that are most prevalent within the dataset and the average rating of these genres of movies.

Here is the visualization;



# Visualization.



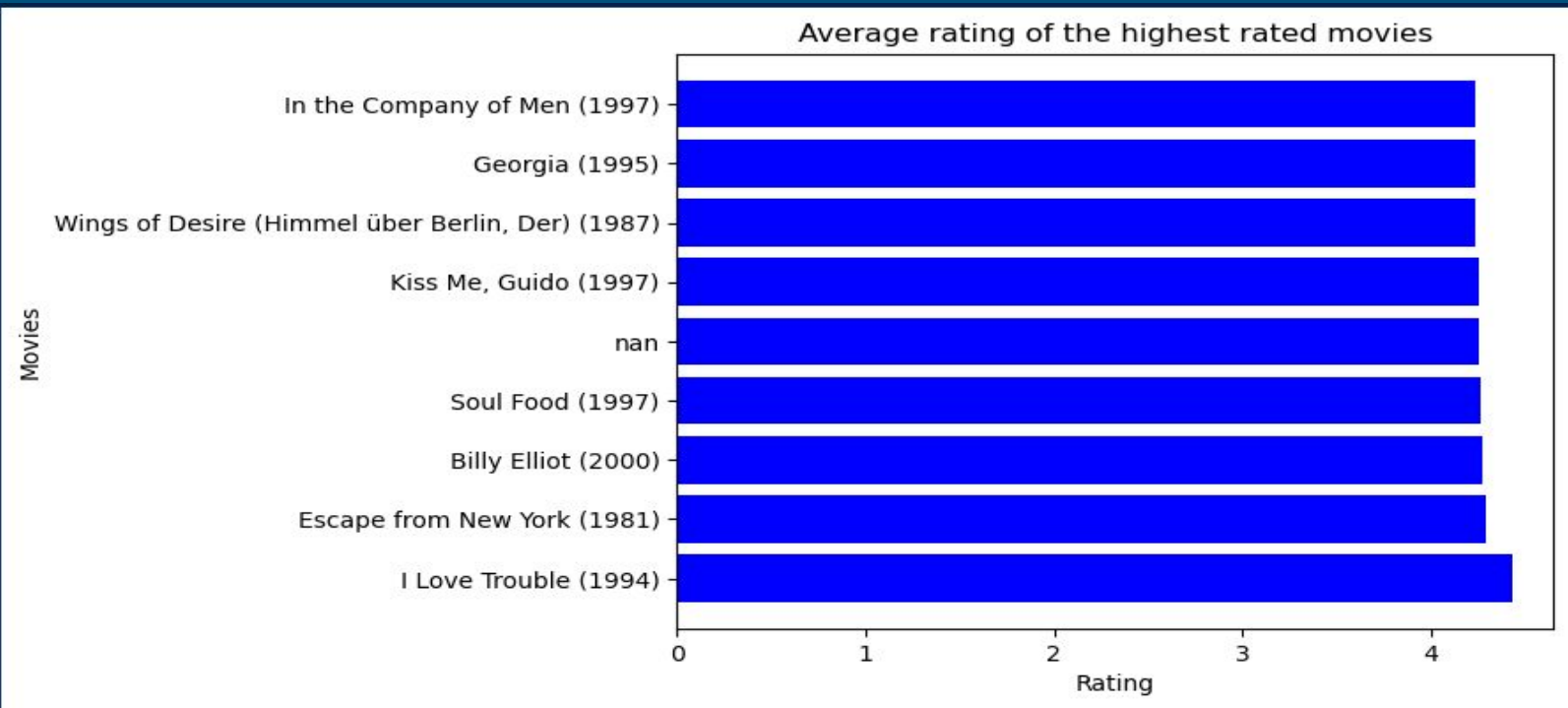
## E.D.A.

---

Finally, I analyse the `___rating_data___` data frame to determine the top ten highest rated movies that have been rated by at least a hundred users.

Here is the visualization;

# Visualization.



# Model 1.

---

The first model aimed to give recommendations of movies similar to a given movie. The movies which were more similar to the given movie will had a higher weight in the rating computation for the movie.

Overall the model appeared to perform well by recommending movies that were closely related to the given movie.

# Model 1 results.

---

```
... Movie Id: 100 Simillar Movies are:

452          Rising Sun (1993)
390      Fear of a Black Hat (1994)
391          With Honors (1994)
456      Romeo Is Bleeding (1993)
44          Pocahontas (1995)
590      Cold Comfort Farm (1995)
341          Bad Company (1995)
534  All Dogs Go to Heaven 2 (1996)
408          House Party 3 (1994)
156          Net, The (1995)
Name: title, dtype: object

Movie Id: 100000 Simillar Movies are:

Movie Id Not Found...
```

## Model 2.

---

The second model was going to give movie recommendations to specific users based on the movies viewed and rated by that user. For these particular task, I used the pearson's correlation coefficients to determine the most suitable movies for the user. The model performed well and had the following scores;

RMSE: 3.20.

MAE: 2.96.

# Model 2 results.

User Id: 10 Recommended Movies are:

	movieId	title	genres
43	47	Seven (a.k.a. Se7en) (1995)	Mystery Thriller
130	157	Canadian Bacon (1995)	Comedy War
136	163	Desperado (1995)	Action Romance Western
184	216	Billy Madison (1995)	Comedy
720	940	Adventures of Robin Hood, The (1938)	Action Adventure Romance

User Id: 100000 Recommended Movies are:

UserId Not Found..

# Conclusion.

---

I) The first model is useful in recommending movies to users who are yet to make their preferences known but are interested in specific movies and would like recommendations of similar movies.

II) I was able to come up with another model that recommends five movies to a user based on the ratings that he/she has given other movies. These was the main objective of these project and it has been achieved with \_\_RMSE\_\_ and \_\_MAE\_\_ as the evaluation metrics.



# Recommendations.

---

Based on the \_\_\_RMSE\_\_\_ and \_\_\_MAE\_\_\_ values obtained, the next step I recommended with regards to building a better recommendation system, was to broaden the scope of analysis. A deeper dive beyond the rating of the movies is necessary in order to better understand the preferences of individual users.