# **Movie Recommendation System Project Report**

**Introduction**

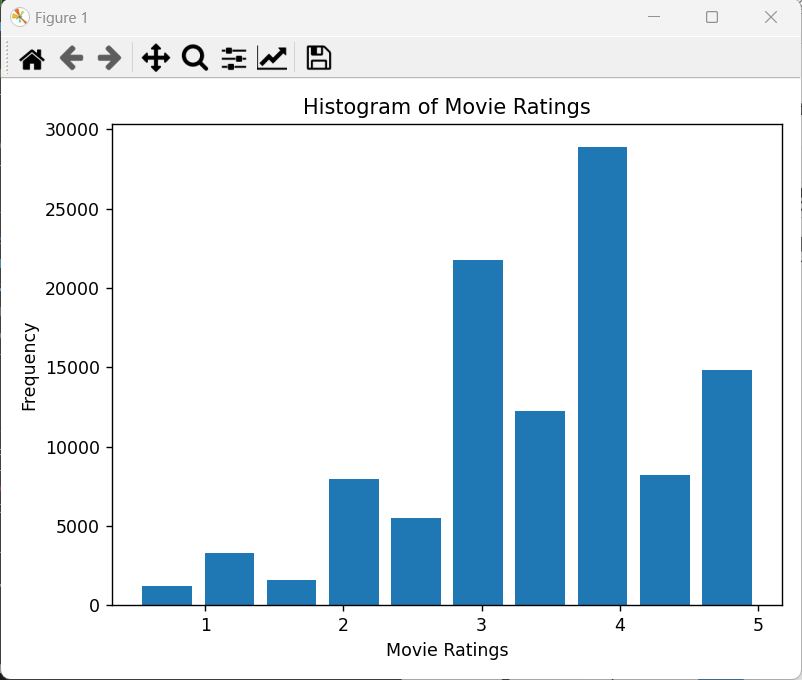
The purpose of this project is to develop a movie recommendation system using Python and Pandas. The system will use both content-based filtering and collaborative filtering techniques to recommend movies to users based on their preferences. The IMDB dataset and IMDB dataset2 are used for data analysis and building the recommendation system.

**Data Exploration**

Exploratory data analysis was performed to understand the distribution of the data. The following histograms were plotted:

**1-a) Histogram of movie ratings**

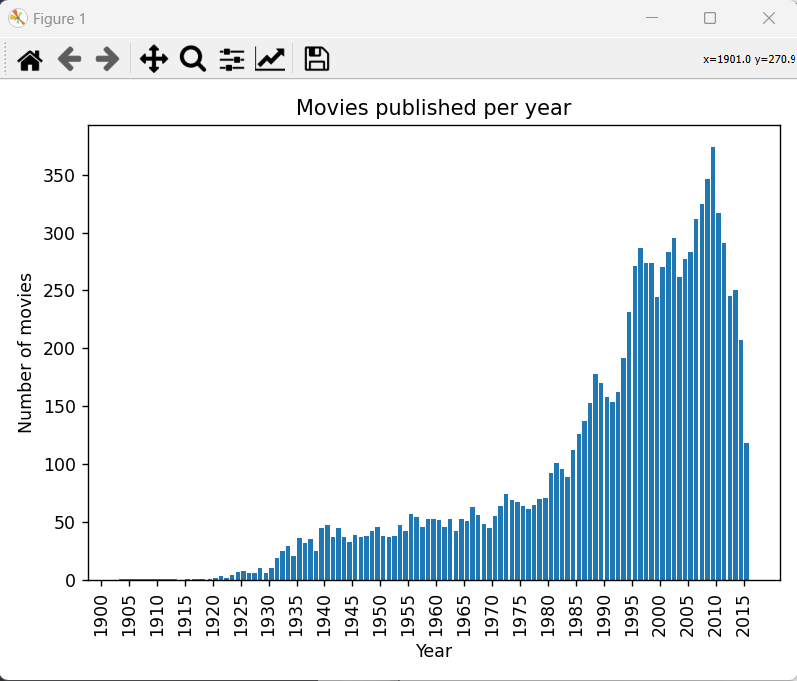
**Output:**



The histogram shows that most of the movies have ratings between 3 and 5.

**1-b) Histogram of movies published per year**

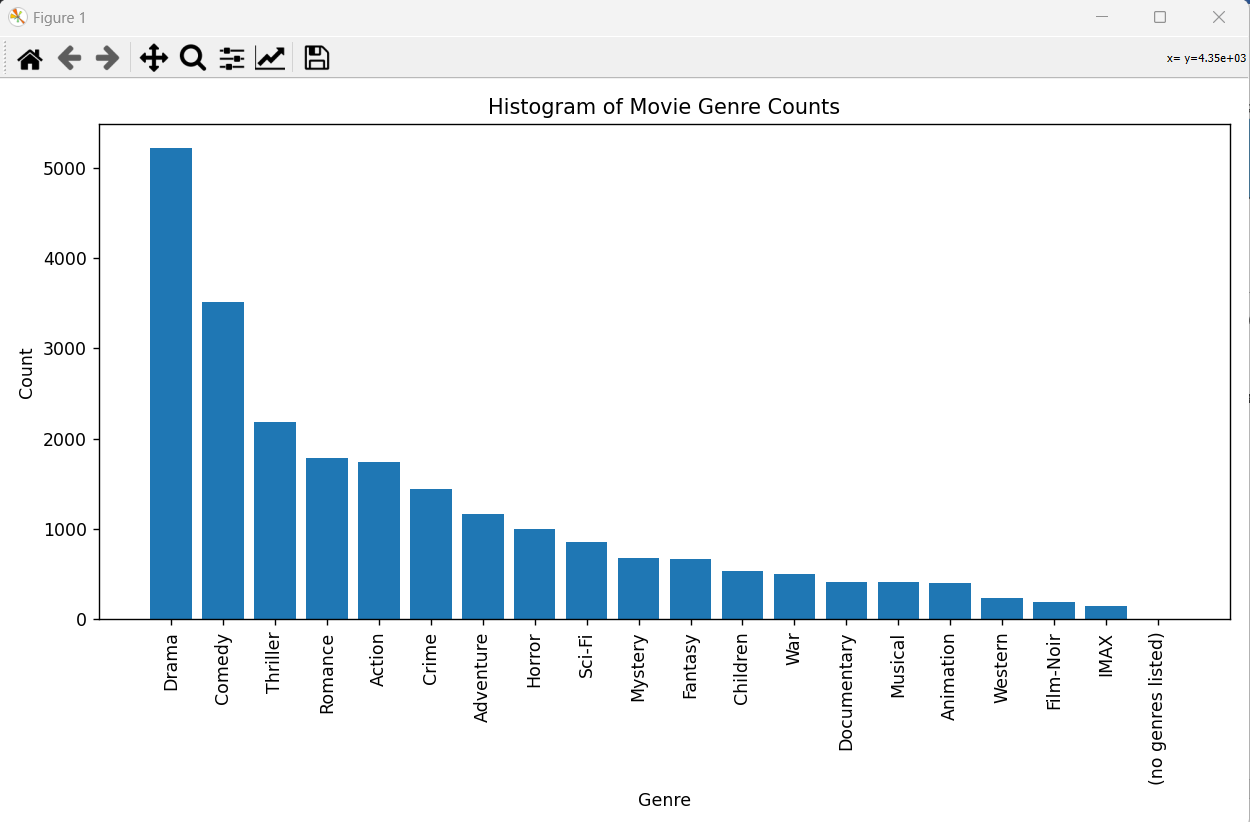
**Output:**



The histogram shows that most of the movies were published in the 2000s.

**1-c) Histogram of count of genres showing the most popular movie genres**

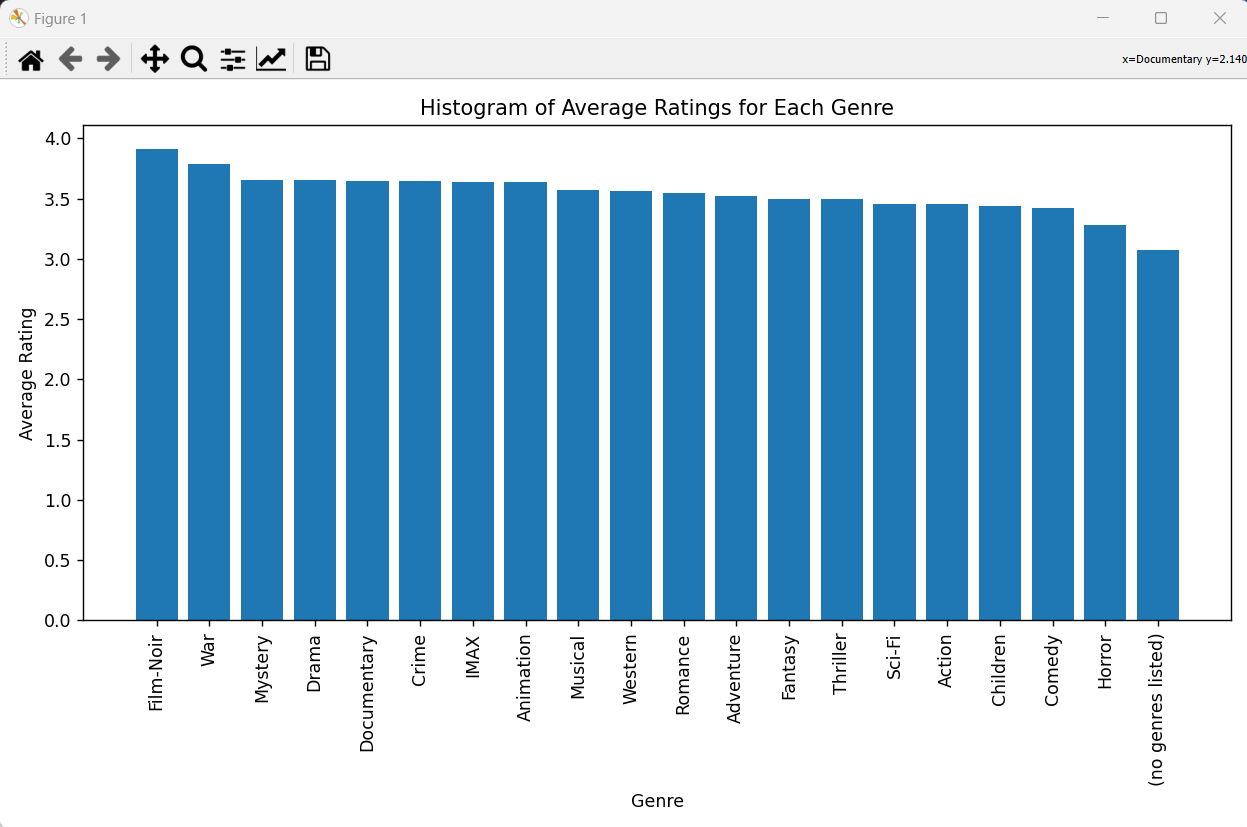
**Output:**

****

From the histogram, we can see that the most popular genre is Drama, followed by Comedy and Thriller.

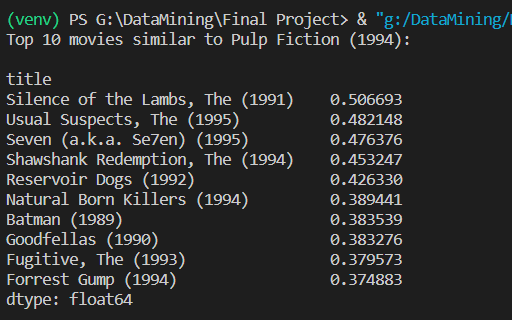
**1-d) Histogram of the average ratings for each genre**

**Output:**



**1-e) Using pearson's R correlation on the ratings, recommend top 10 movies similar to Pulp Fiction (1994).**

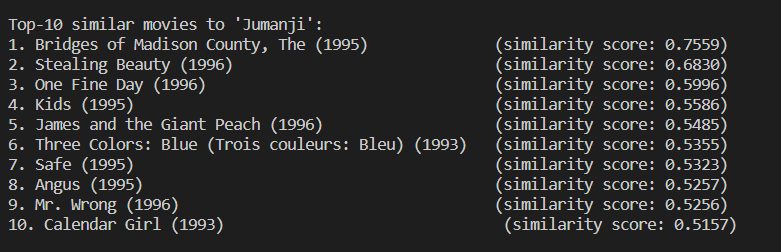
**Output:**

****

These are the top 10 movies similar to "Pulp Fiction (1994)" based on Pearson's R correlation on the ratings.

**1-f) Using K-Nearest Neighbor algorithm and cosine distance similarity, recommend the top-10 movies similar to 'jumanji'.**

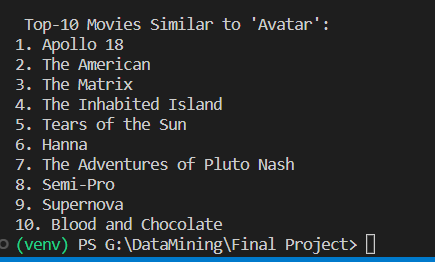
**Output:**



**Similarity Score = 1 – Distance vector of K-Nearest Neighbor**

**2-a) Using the TF-IDF on the 'overview' text, give recommendation of top-10 movies similar to 'Avatar'.**

**Output:**

****

**Conclusion**

In conclusion, I have implemented a basic movie recommendation system using Python and Pandas. I have demonstrated how to use the IMDB dataset to create histograms of movie ratings, movies published per year, the count of genres, and average ratings for each genre. We have also shown how to use Pearson's R correlation and K-Nearest Neighbor algorithms to recommend top-10 movies similar to a user's movie choice.

Furthermore, I have applied the TF-IDF algorithm on the movie overviews to provide a recommendation of top-10 movies similar to a specific movie.