MOVIE RECOMMENDATION SYSTEM

WONUOLA ABIMBOLA

Business Case

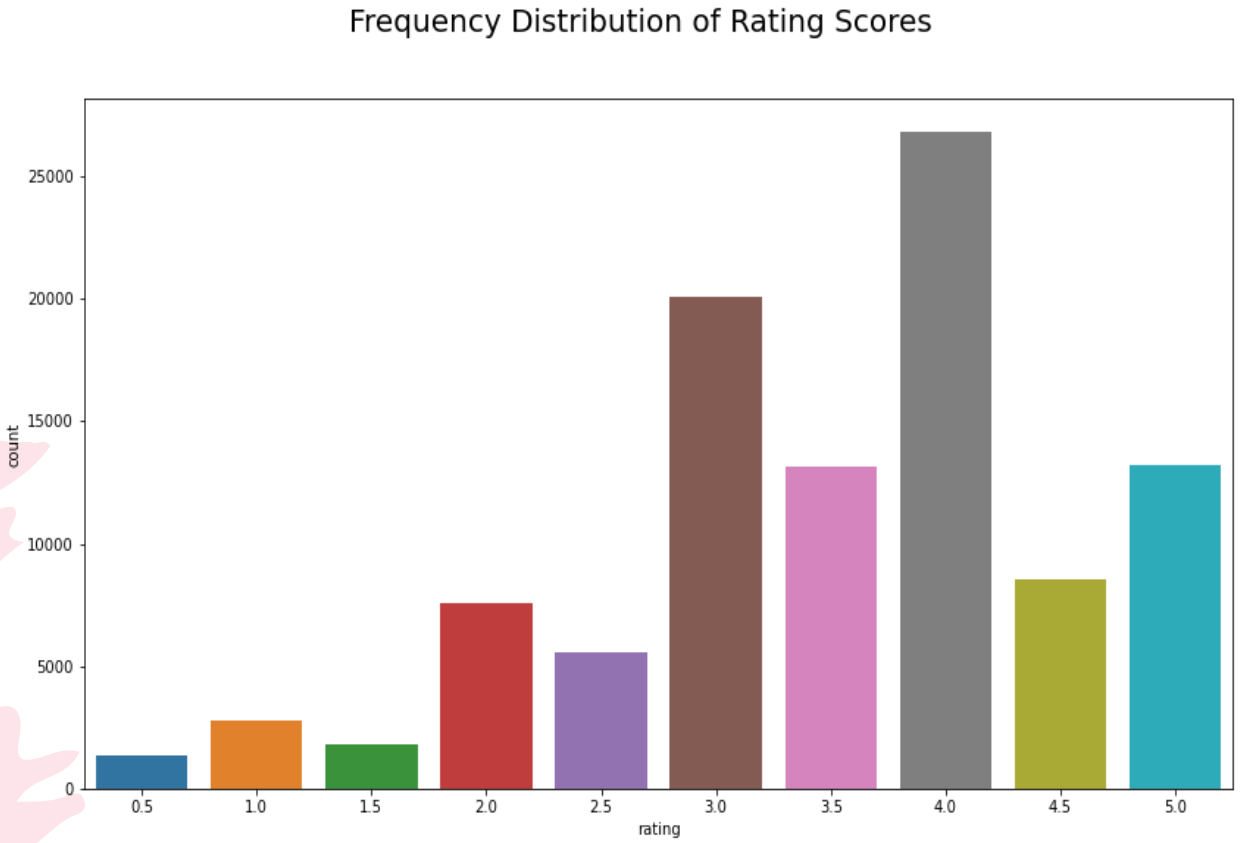
- Build a movie recommendation system for a streaming service that offers a wide variety of award-winning movies on their platform.
- Data used in this project was obtained from the latest MovieLens Dataset.

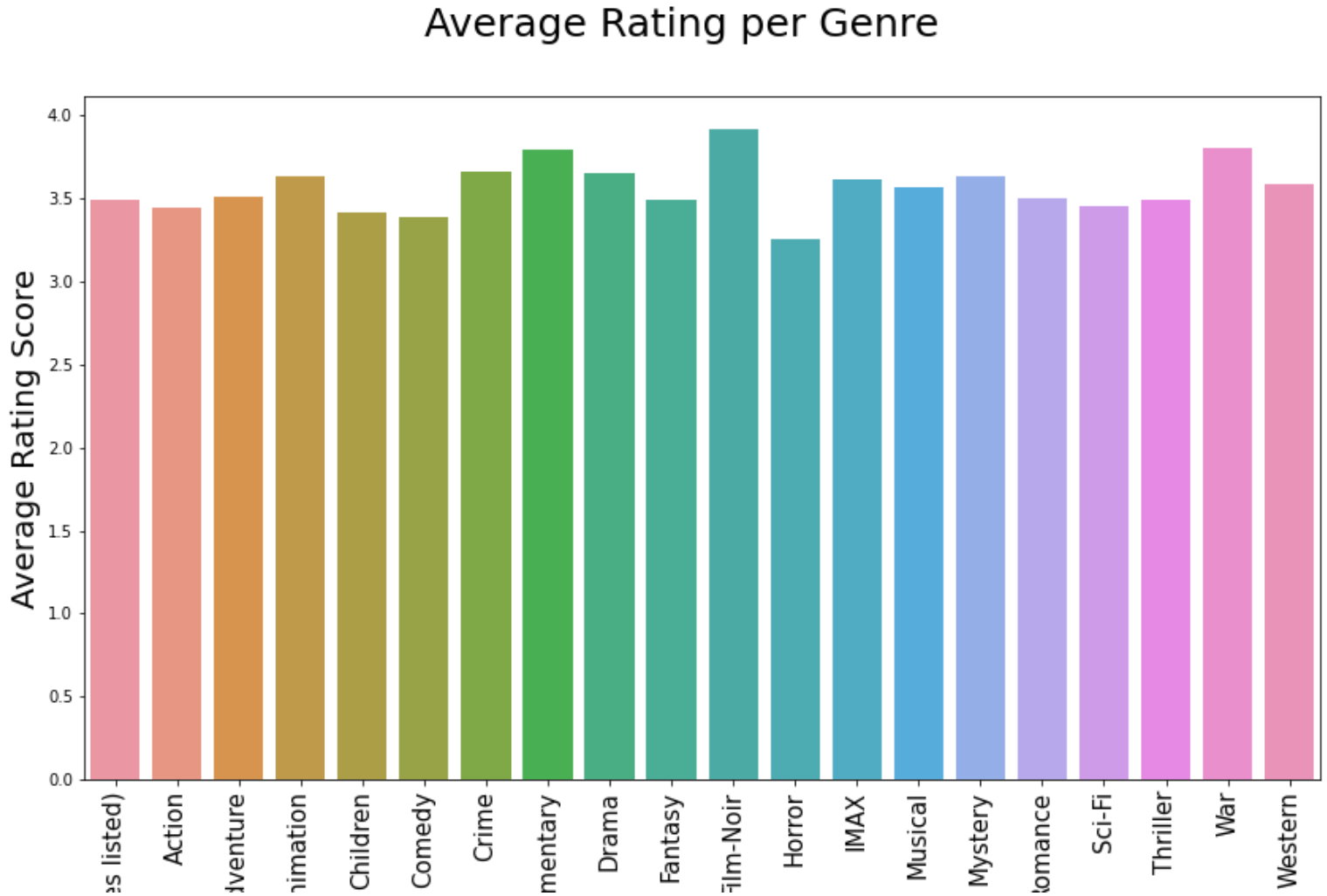
Descriptive Statistics

- Looking at the statistics of the rating scores, ratings are made on a 5-star scale, with half-star increments. The overall mean score was ~3.5 with a standard deviation of ~1.04

Exploratory Data Analysis

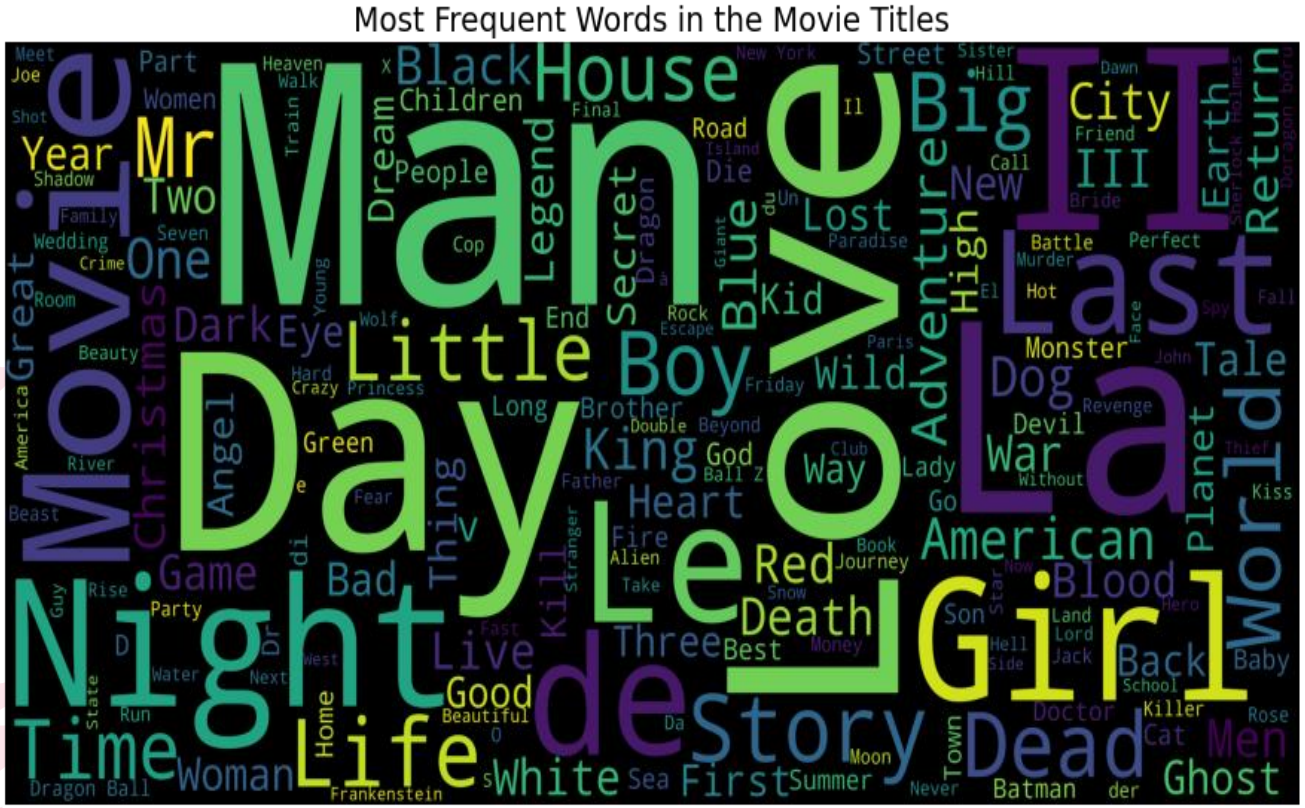
- Most movies rated 3.0 and above
- Most common rating was 4.0




Average Rating per Genre
Average Rating Score
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
es listed)
Action
dventure
imation
Children
Comedy
Crime
mentary
Drama
Fantasy
ilm-Noir
Horror
IMAX
Musical
Mystery
omance
Sci-Fi
Thriller
War
Western

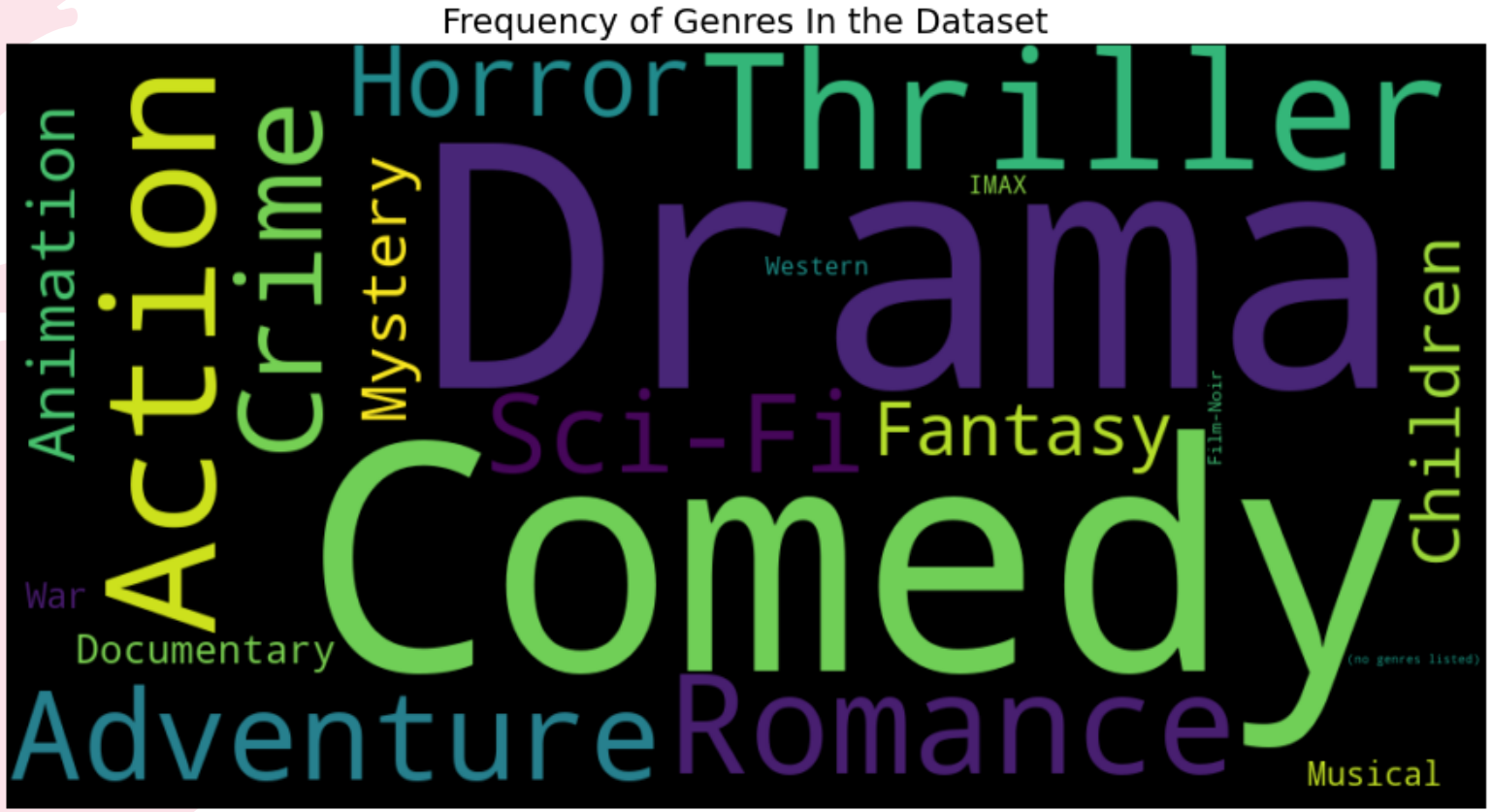
Exploratory Data Analysis

- Most prevalent words in the movie titles



Exploratory Data Analysis

- Most popularly mentioned genres in the dataset are displayed.



Content Based Model

- The cosine similarities using the movies 'genres' and 'tags' columns was calculated.
- A user provides the title of a movie they have watched.
- The model then recommends movies with the cosine highest similarity with the movie

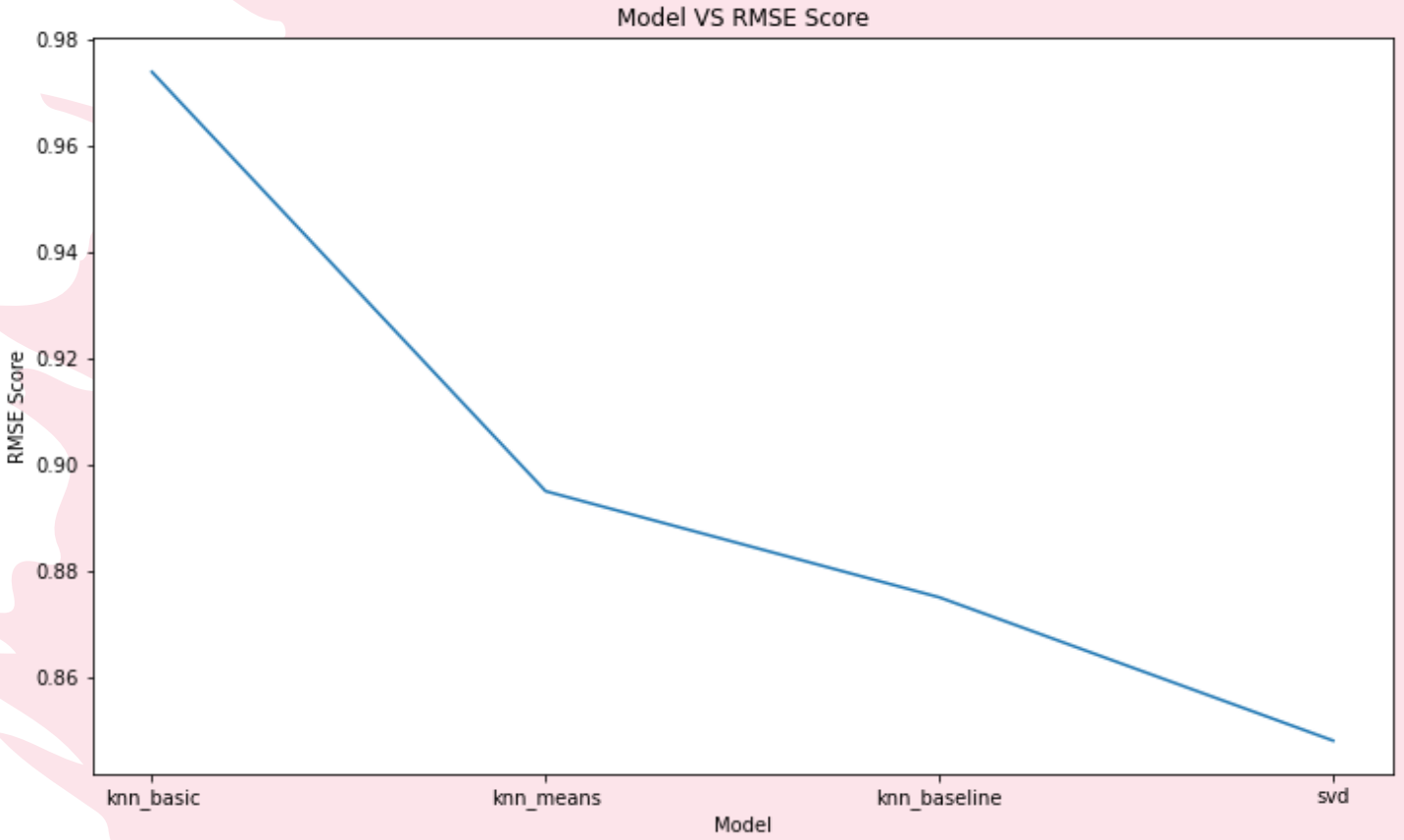
```
movies_like_hulk  = ContentBasedRec('Hulk (2003)', mt_ratings)
```

```
movies_like_hulk.movie_list()
```

```
Here are a few recommendations for you
['Hulk (2003)',
 'Star Wars: Episode IV - A New Hope (1977)',
 'Star Wars: Episode V - The Empire Strikes Back (1980)',
 'X-Men (2000)',
 'Star Trek (2009)',
 'Terminator Salvation (2009)',
 'Avatar (2009)',
 'Star Wars: Episode I - The Phantom Menace (1999)',
 'Superman (1978)',
 'Superman III (1983)',
 'Star Wars: Episode III - Revenge of the Sith (2005)',
 'Fantastic Four (2005)',
 'Serenity (2005)',
 'Fantastic Four: Rise of the Silver Surfer (2007)',
 'Green Lantern (2011)']
```

Collaborative Filtering Model

- For this, an SVD model was used
- Many iterations were done with various hyperparameter values
- The model we ended up using gave resulted in an rmse score of ~0.85



Conclusion

- Both models are useful and work hand-in-hand.
- Content based recommendations are straightforward and help with the 'cold start' problem but do not expose the user to new content.
- Collaborative Filtering helps to expose the user to content they may not have seen before based on users that share similar tastes