Phase 4 Project Non-Technical Presentation

MovieLens Recommendation System

BRIAN O'DONELL JANUARY 2023

Overview & Business Understanding

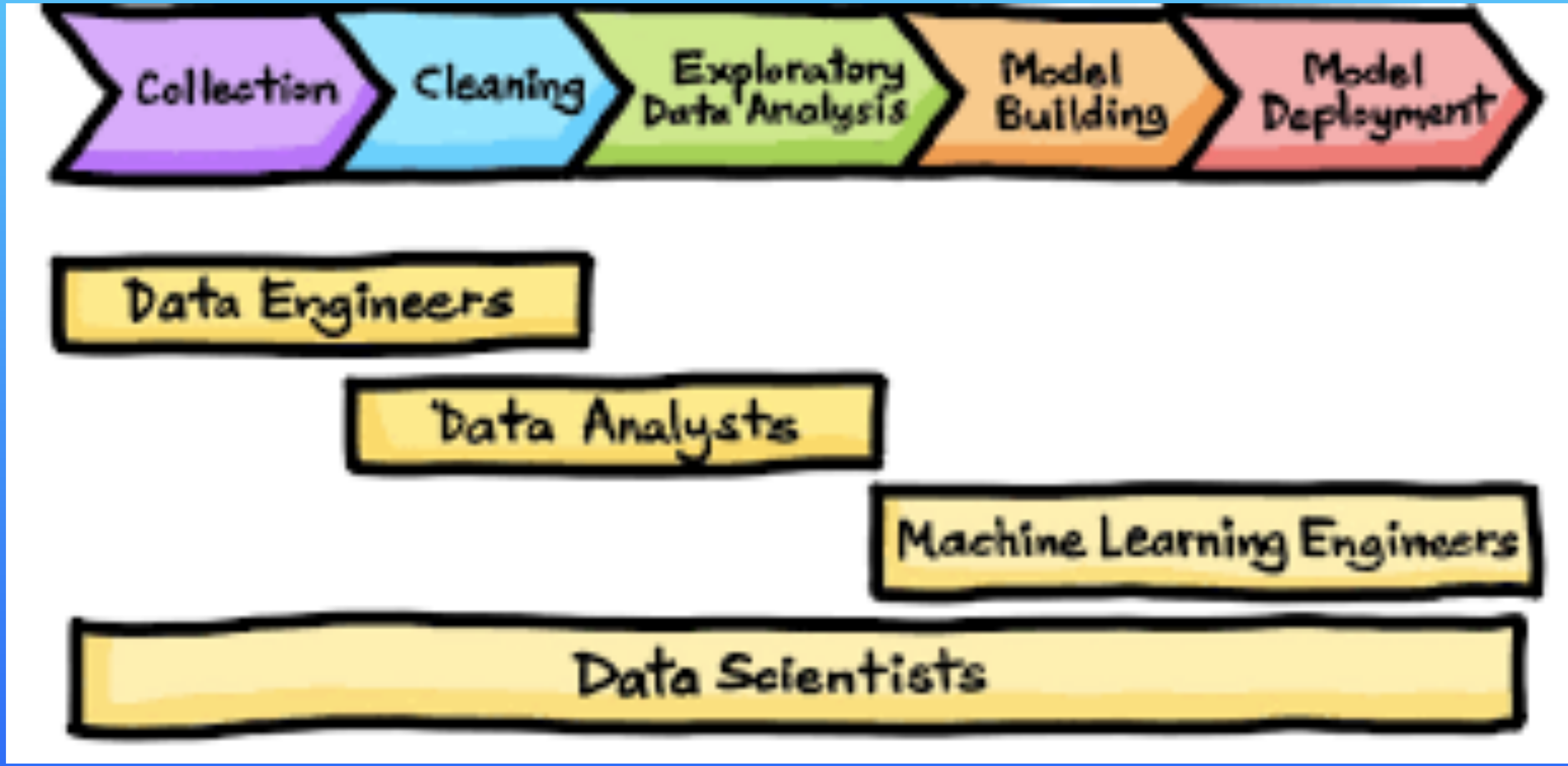
- Given the 'MovieLens' dataset I was tasked with creating a new recommendation system based off the history of users and their movie ratings
- This recommendation system will bring unique value to the customers and give our stakeholders a strong start with entering the competitive movie-streaming industry

Data Understanding

- The MovieLens Dataset contains over:
  - 9,000 Movies
  - 600 Users
  - 100,000 Ratings

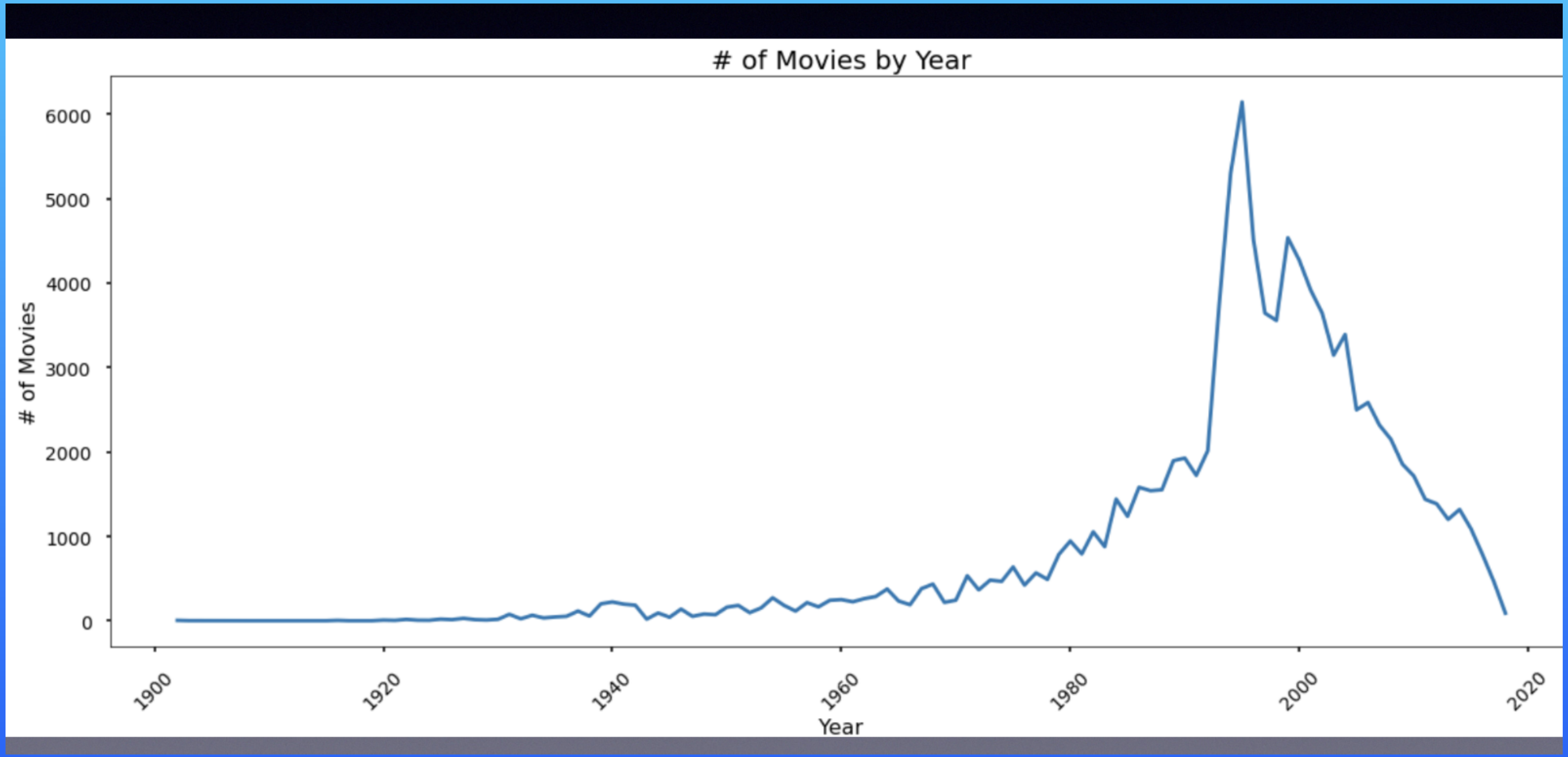
Data Preparation

- I Followed the basic Data Science Methodology to ensure the best results for our model.
- The data was thoroughly:
    - Cleaned
    - Inspected
    - Fit
    - Modeled



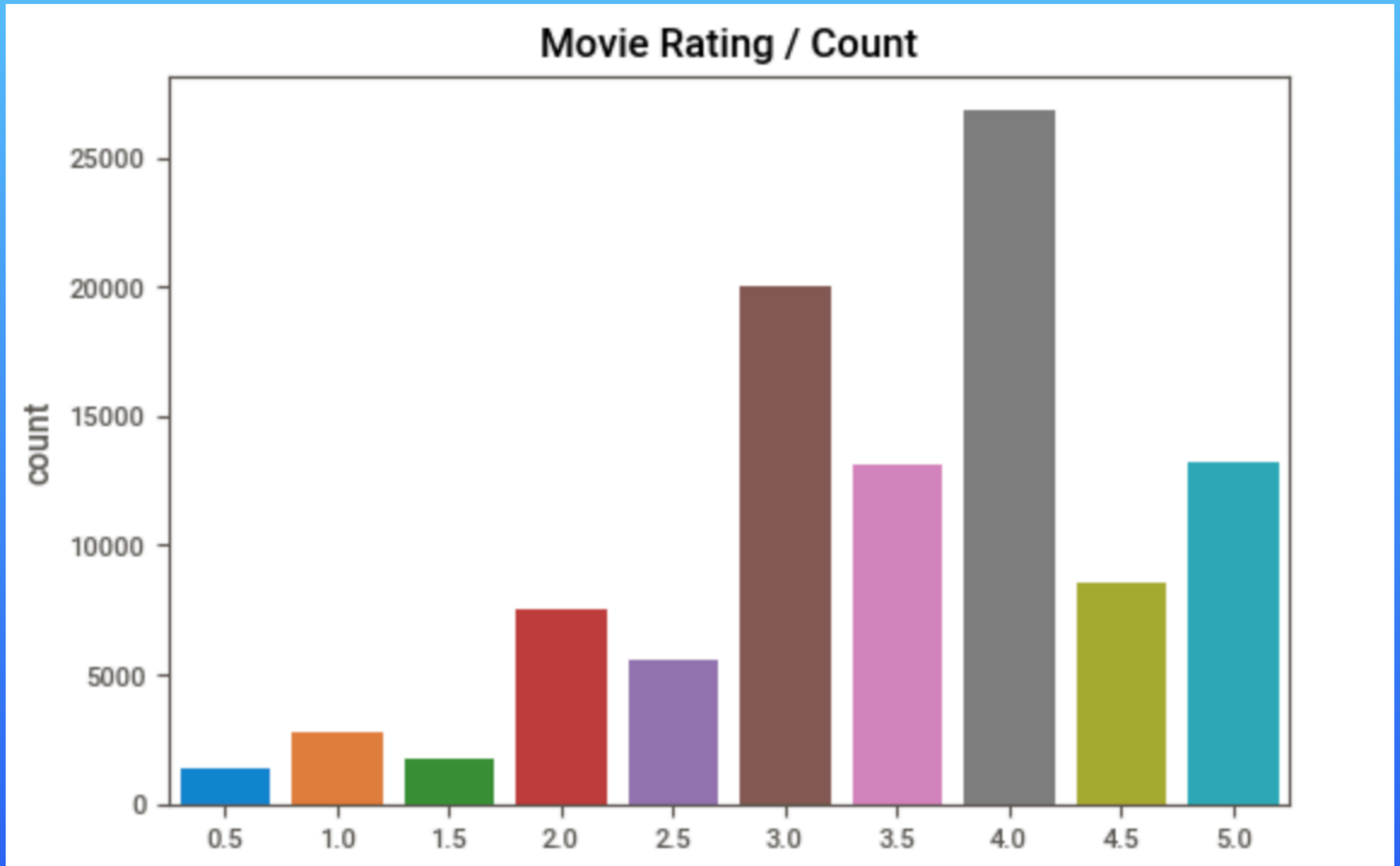
Data Understanding

- Some info about the data:
  - Movies Primarily from 1990s through 2018



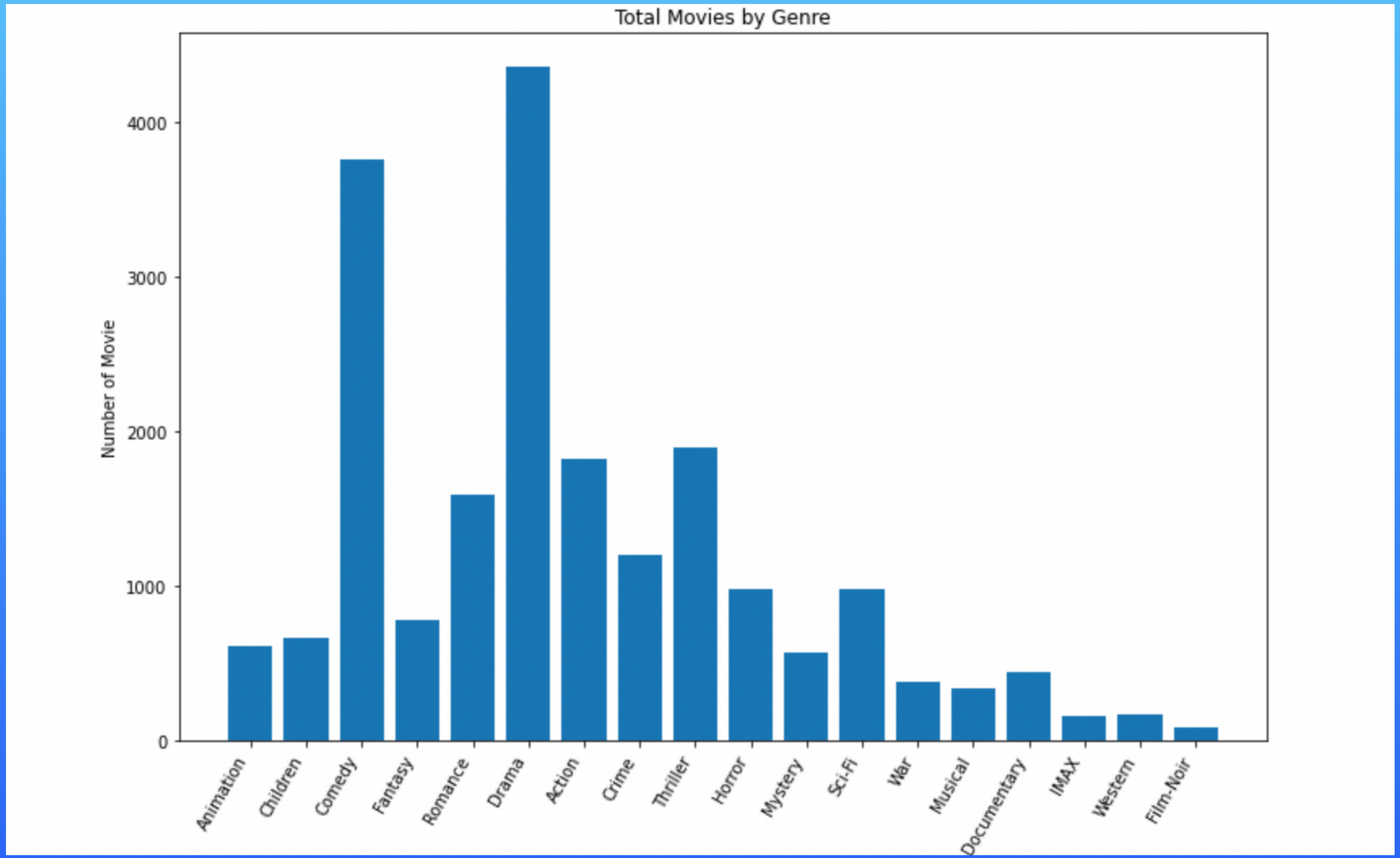
Data Understanding

- Some info about the data:
  - People tend to give higher reviews
  - Possible that people only watch movies they know they will like...



Data Understanding

- Some info about the data:
  - Drama was the most popular genre
  - Followed by:
    - Comedy
    - Thriller
    - Action



Models

- 2 Different Machine Learning algorithms were used:
  - Singular Value Decomposition(SVD)
    - Metrics after several iterations:
      - Root Mean Squared Error: 0.82
      - Mean Absolute Error: 0.63

  - This metrics tells us how far our predictions are from the actual rating
    - Example: If a user rated a movie 3.5 stars our model would predict within .82 stars.

Models

- The other algorithm :
  - K-Nearest-Neighbors(KNN)
    - Metrics after several iterations:
      - RMSE : 0.83
      - MAE: 0.64

Evaluating Results & Recommendations

- Both models performed well, but the KNN model seems to return more consistent results in terms of ratings and genre
- Many movies being recommended were 'Classics' so to speak, consider weighting the year the movie was released and putting more emphasis on genre to appeal to appropriate target market.
- Continue to add to the dataset, append movies released since 2018 and continue to tune and iterate models to keep constant results

Final Recommendation

- The KNN model results shown below reinforce why this models should be used to launch
- As mentioned continue to add to the dataset to keep up with trends
- Take advantage of the 'Cold-Start' Engine to get best results for New Users

```python
# Get the top recommendations for specific user....
user_id = 608
recommendations = get_top_n_k(predictions_k,minimum_num_ratings=100, n=5)[user_id]

# Print the recommended movie IDs and predicted ratings
print([(iid, rating) for (iid, rating) in recommendations])
```

```
[('Dark Knight, The (2008)', 4.215843395524766), ('Prisoners (2013)', 4.212079225026457), ('Spotlight (2015)', 4.193783809287072), ('Departed, The (2006)', 4.1654399244949305), ('There Will Be Blood (2007)', 4.147380483286304)]
```

KNN Recommendation System Results:

Thank YOU!!!

- Questions??