

# Movie Recommendation System

Presented by: Jiarong Chen



# Background

## Business problem

Using the movie lens 100k data to analyze the movies' rating and make a movie recommendation system.

## Outline

1. **Best Rating Movie**
  - Count the total movie rating and rating times
  - Sort the movies by mean rating
  - Determine the best movie base on mean rating and rating times
2. **Best Movie Genre**
  - Build a genre data frame for all movies
  - Fill the genre data frame with total rating and rating times
  - Sort the genre data frame by mean rating and determine the best one
3. **Recommendation System**
  - Train a suitable algorithm
  - Tune hyperparameter for this model
  - Build a recommendation system





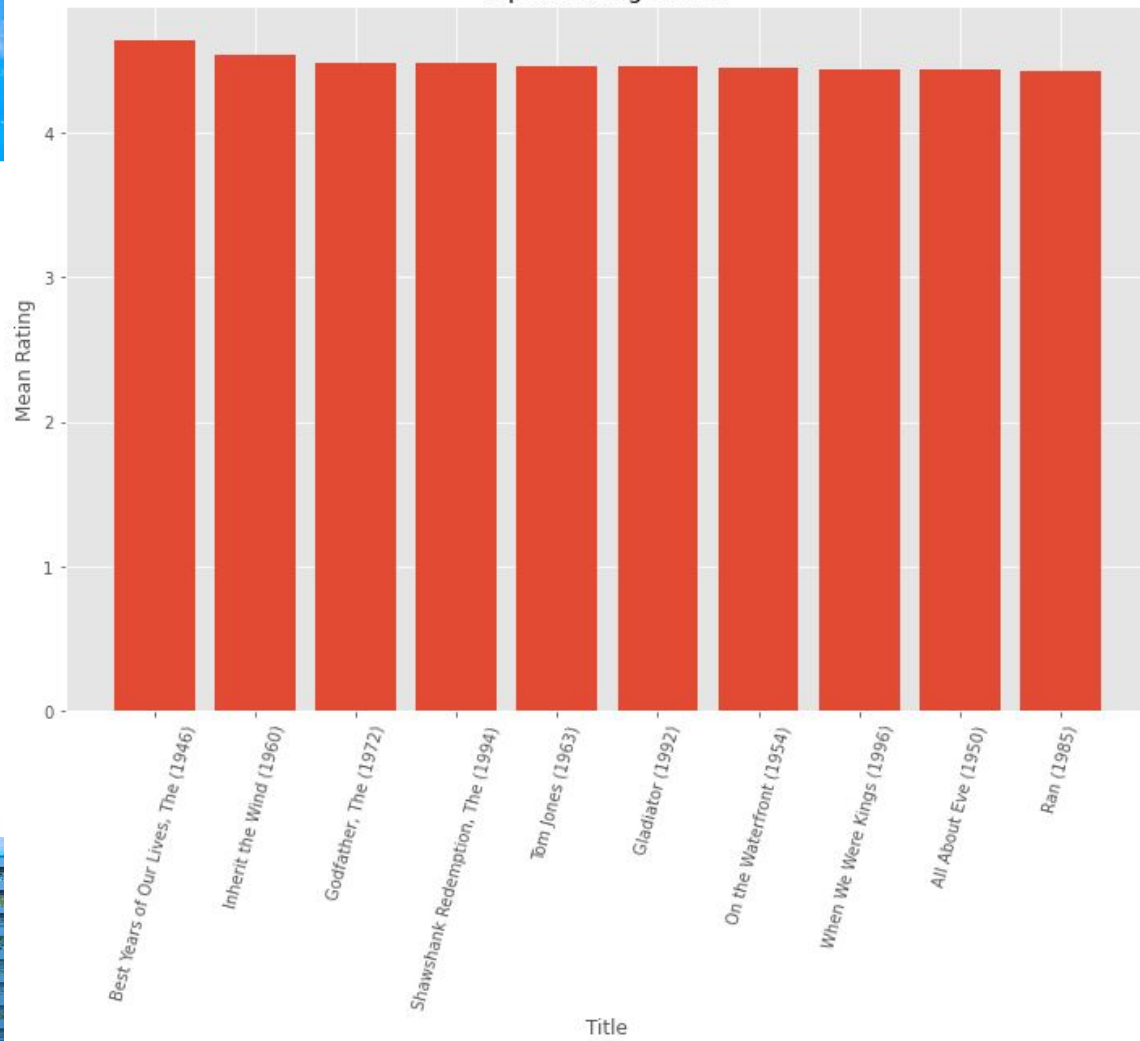
# 1. Best Movie

- a. Movie Rating Table
- b. Movie Rating Bar Graph
- c. Summary

## Top 10 best rating movies

	rating	rating_times	title	genres
movielf				
1939	4.636364	11	Best Years of Our Lives, The (1946)	Drama War
3469	4.541667	12	Inherit the Wind (1960)	Drama
858	4.487500	200	Godfather, The (1972)	Crime Drama
318	4.487138	311	Shawshank Redemption, The (1994)	Crime Drama
1948	4.458333	12	Tom Jones (1963)	Adventure Comedy Romance
8132	4.454545	11	Gladiator (1992)	Action Drama
1945	4.448276	29	On the Waterfront (1954)	Crime Drama
1147	4.437500	16	When We Were Kings (1996)	Documentary
926	4.434211	38	All About Eve (1950)	Drama
1217	4.423077	26	Ran (1985)	Drama War

Top 10 Rating Movies





# Summary

**Best Movie: Godfather, The(1972) and Shawshank Redemption, The(1994)**

**Godfather, The(1972) :**  
**Mean Rating: 4.487500**  
**Rating Times: 200**  
**Genre: Crime, Drama**

**Shawshank Redemption, The(1994) :**  
**Mean Rating: 4.487138**  
**Rating Times: 311**  
**Genre: Crime, Drama**



## 2. Best Genre

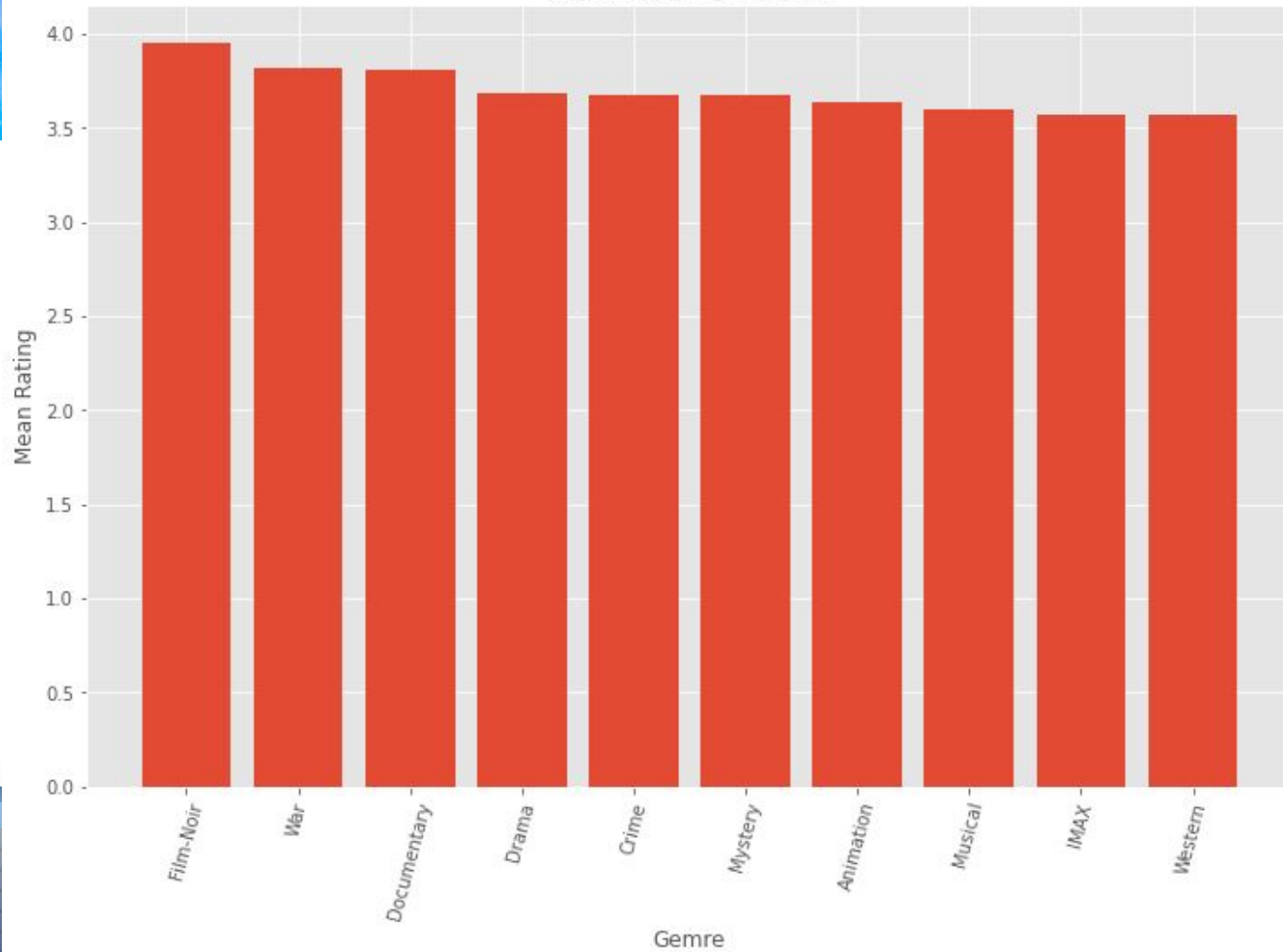
- a. Genre Rating Table
- b. Genre Rating Bar Graph
- c. Summary

All genres ranking by  
mean rating:

	totalRating	meanRating	count
<b>Film-Noir</b>	4509.5	3.955702	1140
<b>War</b>	19181.5	3.817214	5025
<b>Documentary</b>	5964.0	3.813299	1564
<b>(no genres listed)</b>	68.0	3.777778	18
<b>Drama</b>	164767.0	3.681780	44752
<b>Crime</b>	59853.0	3.679639	16266
<b>Mystery</b>	28056.5	3.679541	7625
<b>Animation</b>	22434.5	3.636062	6170
<b>Musical</b>	16993.5	3.598793	4722
<b>IMAX</b>	11270.5	3.571134	3156
<b>Western</b>	6819.0	3.566423	1912
<b>Romance</b>	68762.0	3.556165	19336
<b>Adventure</b>	77508.5	3.520393	22017
<b>Thriller</b>	88807.0	3.518502	25240
<b>Fantasy</b>	37492.0	3.518063	10657
<b>Children</b>	30086.5	3.466187	8680
<b>Sci-Fi</b>	53169.5	3.460430	15365
<b>Comedy</b>	131039.0	3.446037	38026
<b>Action</b>	93224.5	3.445613	27056
<b>Horror</b>	22510.5	3.315243	6790



Top 10 Rating Genres



# Summary

**Best Genre: Film-Noir**

**Mean Rating: 3.955702**

**Rating Times: 1140**

# 3. Recommendation System

- a. Algorithm Training(appendix)
- b. Hyperparameter Tuning(appendix)
- c. Recommendation System



```
1 # Obtain user ratings
2 user_rating = movie_rater(movies, 4, 'Comedy')
```

	movieId	title	genres
7682	82150	Bunny and the Bull (2009)	Adventure Comedy Drama

How do you rate this movie on a scale of 1-5, press n if you have not seen :

4

	movieId	title	genres
8463	106782	Wolf of Wall Street, The (2013)	Comedy Crime Drama

How do you rate this movie on a scale of 1-5, press n if you have not seen :

5

	movieId	title	genres
4055	5299	My Big Fat Greek Wedding (2002)	Comedy Romance

How do you rate this movie on a scale of 1-5, press n if you have not seen :

4

	movieId	title	genres
7719	83506	Ricky Gervais Live 4: Science (2010)	Comedy

How do you rate this movie on a scale of 1-5, press n if you have not seen :

4

## Recommendation List:

Recommendation # 1 : Shawshank Redemption, The (1994)

Recommendation # 2 : Wallace & Gromit: A Close Shave (1995)

Recommendation # 3 : Godfather, The (1972)

Recommendation # 4 : Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)

Recommendation # 5 : Modern Times (1936)

# Conclusion

- **Godfather, The(1972) is the best movie by rating**
- **Film-Noir is the best movie genre by rating**
- SVDpp(Singular Vector Decomposition) is a best algorithm in this case with RMSE around 0.89 and MAE around 0.69



# Recommendation

For Users:

- \* Some best movies are in crime and drama genres. And these two genres both having around 3.68 mean rating which are the top five genres in our movies data.
- \* If you don't have any idea what genre movies to watch, film-noir will be your best choice because it has the highest mean rating(3.955702) with 1140 rating times.
- \* If you need more accuracy recommendation, you can use the recommendation system.

For analysts:

- \* SVDpp(Singular Vector Decomposition) is an accuracy algorithm in this case but the shortcoming is costing a long run time.
- KNN(K-Nearest Neighbors) also an other option with lower accuracy but cost less time.
- \* GridsearchCV is a handy tool to help us tune model parameters, but the shortcoming also is costing a long run time.

When watching funny movie alone



When watching funny movie with family



# Future Work

Algorithm: Try more algorithms from other packages

Hyper Parameter Tuning: Tune more parameters of the models and use random search before grid search

Exploratory Data Analysis: Exploratory more characteristics from the dataset such as timestamp column





# Thank you!

**Any questions?**

You can find me at:

<https://www.linkedin.com/in/jiarong-jr-chen-ba87b214a/>

<https://github.com/JRRRRRRR>



# Appendix

Algorithm	RMSE	MAE	Fit Time
KNN Basic	0.9765	0.7518	0.1263
KNN Base Line	0.9048	0.6931	0.2384
KNN Means	0.9271	0.7098	0.1493
KNN ZScore	0.9275	0.7058	0.2704
SVD	0.9039	0.697	7.5432
SVDpp	0.892	0.6845	601.5069
NMF	0.9584	0.7366	7.6169
SlopeOne	0.938	0.7188	7.9913
CoClustering	0.9726	0.7539	2.6456

```
{'rmse': 0.8914524747832285, 'mae': 0.6850484764220486}
```

```
{'rmse': {'n_epochs': 10, 'lr_all': 0.009}, 'mae': {'n_epochs': 20, 'lr_all': 0.007}}
```