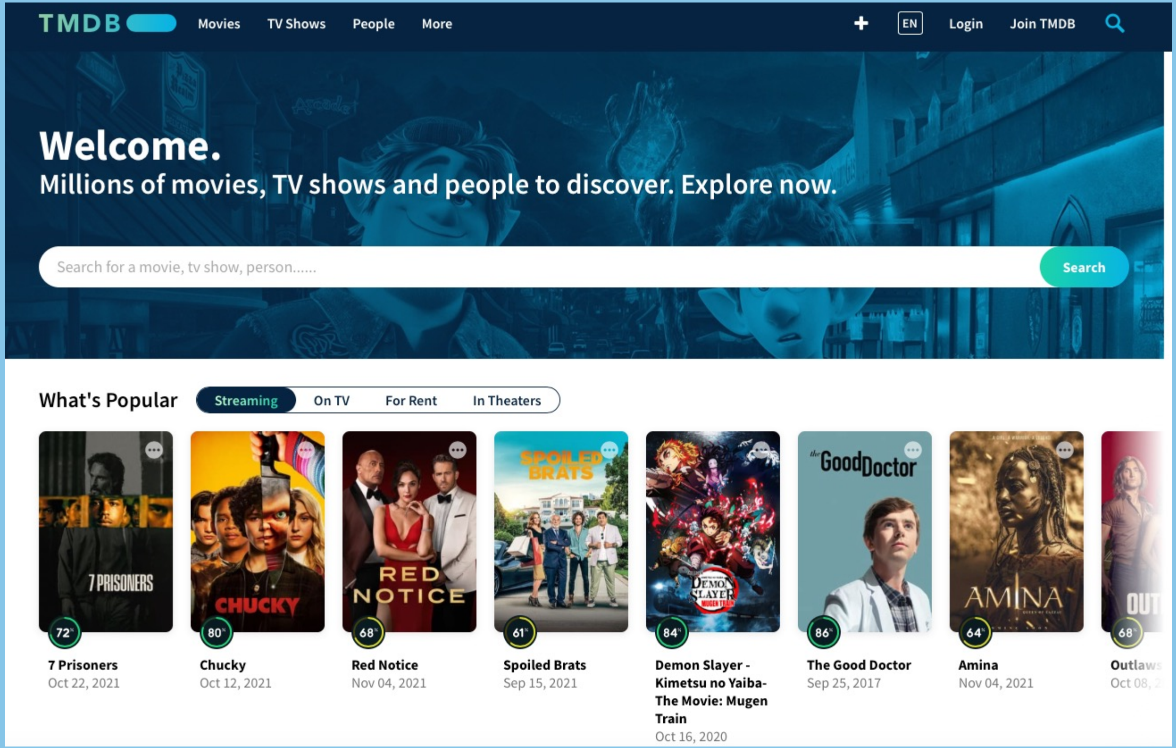
NLP/Unsupervised Learning Presentation Slides

A Hybrid Recommender System for Titles in the TMDB Movie Database

George Pappy - 15 December 2021

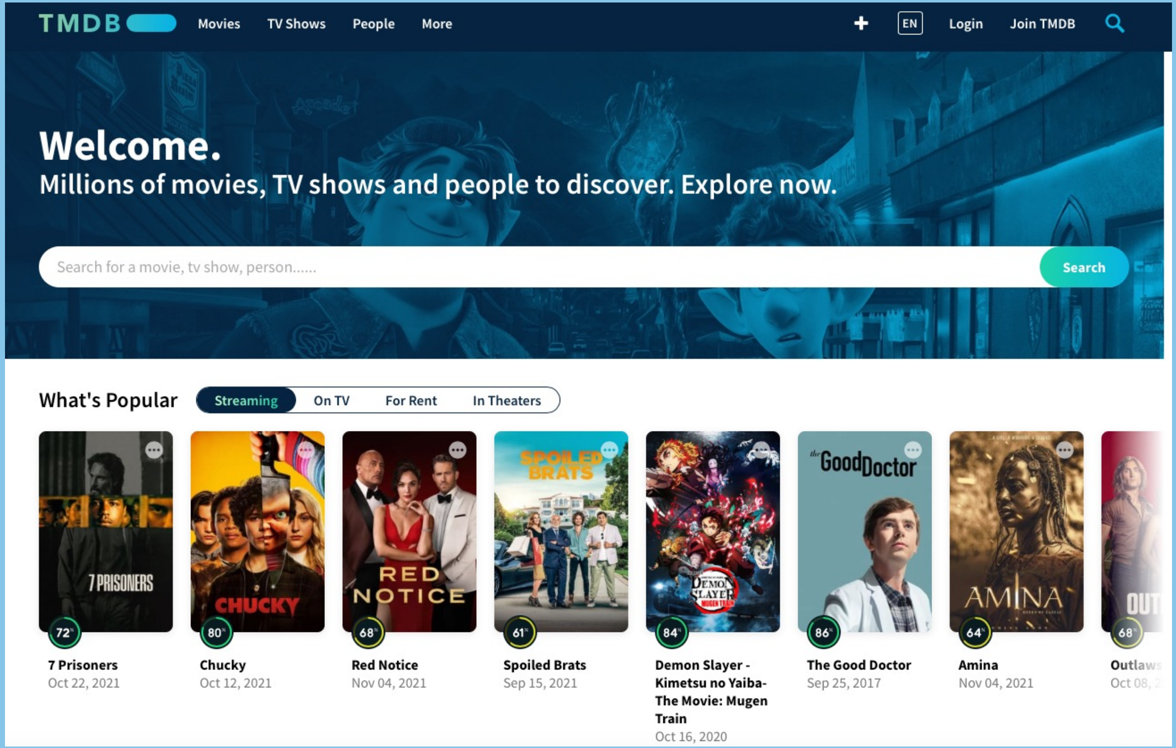
Introduction

- The Movie Database (TMDB.org) offers a service rivaling IMDB.com



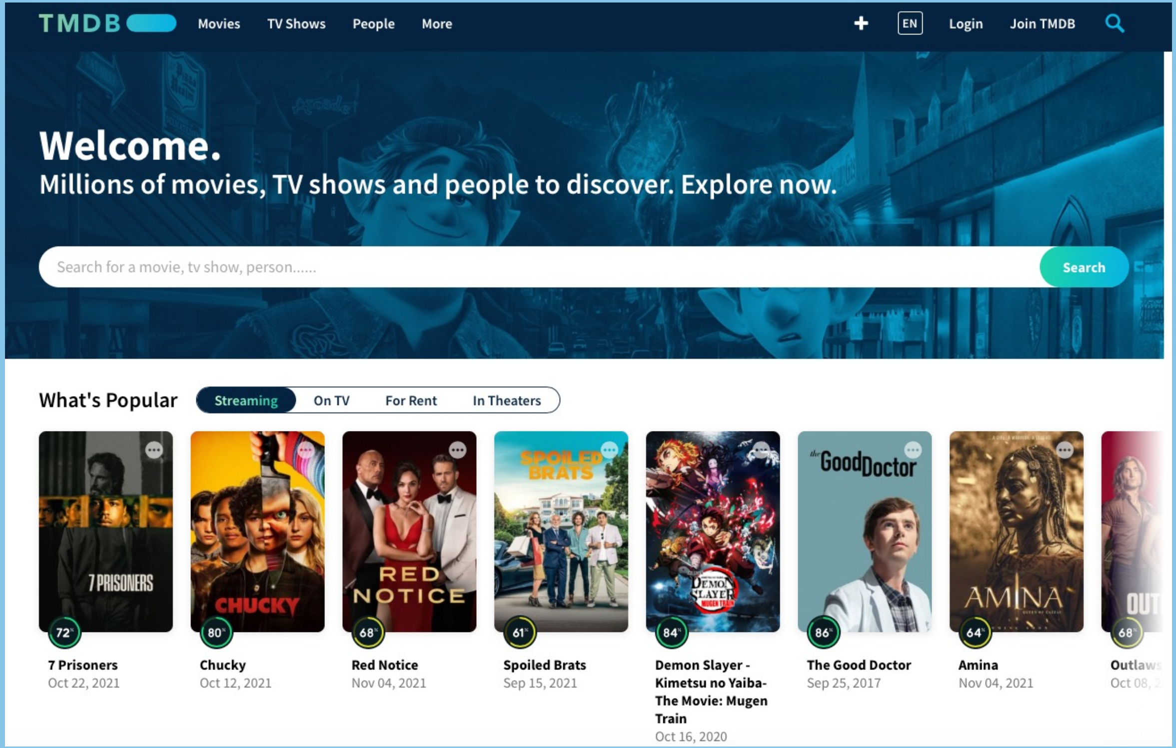
Introduction

- The Movie Database (TMDB.org) offers a service rivaling IMDB.com
  - Has just a fraction of IMDB's daily visits and registered users



Introduction

- The Movie Database (TMDB.org) offers a service rivaling IMDB.com
  - Has just a fraction of IMDB's daily visits and registered users
  - Lives in IMDB's "shadow"



Introduction (con’t.)

Goal: TBDB wants to offer a Movie Recommender System

Introduction (con't.)

Goal: TBDB wants to offer a Movie Recommender System:

- Lure users away from IMDB

Introduction (con't.)

Goal: TBDB wants to offer a Movie Recommender System:

- Lure users away from IMDB
- Attract more registered users based on a superior user experience

Introduction (con’t.)

Goal: TBDB wants to offer a Movie Recommender System:

- Lure users away from IMDB
- Attract more registered users based on a superior user experience
- Encourage registered users to rate more movies

Methodology

- Primary Data Set: Titles from TMDB.org

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:
    - Genre(s)

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:
    - Genre(s)
    - Director Name

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:
    - Genre(s)
    - Director Name
    - Top-4-Billed Actors' Names

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:
    - Genre(s)
    - Director Name
    - Top-4-Billed Actors' Names
    - Text-based Plot Summary

Methodology

- Primary Data Set: Titles from TMDB.org
  - Each title augmented using the TMDB query API:
    - Genre(s)
    - Director Name
    - Top-4-Billed Actors’ Names
    - Text-based Plot Summary
  - After data cleaning 47,723 titles remain

Methodology (con't.)

- Additional Data Set: TMDB Users' Movie Ratings

Methodology (con't.)

- Additional Data Set: TMDB Users' Movie Ratings
  - 5,004,591 Ratings

Methodology (con't.)

- Additional Data Set: TMDB Users' Movie Ratings
  - 5,004,591 Ratings
    - 50,000 Distinct TMDB Users

Methodology (con't.)

- Additional Data Set: TMDB Users' Movie Ratings
  - 5,004,591 Ratings
    - 50,000 Distinct TMDB Users
    - 28,044 Distinct Titles

Methodology (con’t.)

- Additional Data Set: TMDB Users’ Movie Ratings
  - 5,004,591 Ratings
    - 50,000 Distinct TMDB Users
    - 28,044 Distinct Titles
    - Only possible ratings: {0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 , 4.5, 5.0}

Results: Content-Based Recommenders

Results: Content-Based Recommenders

1. **Baseline Recommender:**

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries

Results: Content-Based Recommenders

1. **Baseline Recommender:**
    - NLTK wordtokenized Plot Summaries
    - CountVectorized (NLTK English stopwords)

Results: Content-Based Recommenders

1. **Baseline Recommender:**
    - NLTK wordtokenized Plot Summaries
    - CountVectorized (NLTK English stopwords)
    - TruncatedSVD

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries
   - CountVectorized (NLTK English stopwords)
   - TruncatedSVD
2. **Various Alternatives with Similar or Inferior Performance:**

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries
   - CountVectorized (NLTK English stopwords)
   - TruncatedSVD

2. **Various Alternatives with Similar or Inferior Performance:**
   - CountVectorizer hyperparameter tuning, TF-IDF instead of CountVectorizer

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries
   - CountVectorized (NLTK English stopwords)
   - TruncatedSVD
2. **Various Alternatives with Similar or Inferior Performance:**
   - CountVectorizer hyperparameter tuning, TF-IDF instead of CountVectorizer
   - Keyword Extraction of Plot Summaries (tried Rake-NLTK & SpaCy)

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries
   - CountVectorized (NLTK English stopwords)
   - TruncatedSVD
2. **Various Alternatives with Similar or Inferior Performance:**
   - CountVectorizer hyperparameter tuning, TF-IDF instead of CountVectorizer
   - Keyword Extraction of Plot Summaries (tried Rake-NLTK & SpaCy)
3. **Gensim (Best Performance):**

Results: Content-Based Recommenders

1. **Baseline Recommender:**
    - NLTK wordtokenized Plot Summaries
    - CountVectorized (NLTK English stopwords)
    - TruncatedSVD
2. **Various Alternatives with Similar or Inferior Performance:**
    - CountVectorizer hyperparameter tuning, TF-IDF instead of CountVectorizer
    - Keyword Extraction of Plot Summaries (tried Rake-NLTK & SpaCy)
3. **Gensim (Best Performance):**
    - SpaCy data cleaning & SpaCy English stopwords

Results: Content-Based Recommenders

1. **Baseline Recommender:**
   - NLTK wordtokenized Plot Summaries
   - CountVectorized (NLTK English stopwords)
   - TruncatedSVD
2. **Various Alternatives with Similar or Inferior Performance:**
   - CountVectorizer hyperparameter tuning, TF-IDF instead of CountVectorizer
   - Keyword Extraction of Plot Summaries (tried Rake-NLTK & SpaCy)
3. **Gensim (Best Performance):**
   - SpaCy data cleaning & SpaCy English stopwords
   - Gensim TF-IDF and LSI Topic Modeling (used to perform SVD)

Results: Content-Based Recommenders (con't.)

Baseline Model's Recommendations for "Toy Story"

```
Recommendations based on your interest in Toy Story:
```

title	genres	director	actors
Turbo	[adventure, animation, children, comedy, fantasy]	[david_soren]	[ryan_reynolds, paul_giamatti, michael_peña, s...
Asterix and the Vikings	[adventure, animation, children, comedy, fantasy]	[stefan_fjeldmark]	[roger_carel, lorànt_deutsch, sara_forestier, ...
Hawaiian Vacation	[adventure, animation, children, comedy, fantasy]	[gary_rydstrom]	[tom_hanks, tim_allen, joan_cusack, don_rickles]
The Adventures of Rocky & Bullwinkle	[adventure, animation, children, comedy, fantasy]	[des_mcanuff]	[rene_russo, jason_alexander, piper_perabo, ra...
A Connecticut Yankee in King Arthur's Court	[adventure, children, comedy, fantasy]	[mel_damski]	[keshia_knight_pulliam, michael_gross, jean_ma...
Jack-Jack Attack	[adventure, animation, children, comedy]	[brad_bird]	[bret_'brook'_parker, bud_luckey, eli_fucile, ...
Help! I'm A Fish	[adventure, animation, comedy]	[stefan_fjeldmark]	[sebastian_jessen, pil_neja, morten_kerrn_niel...
Paddington 2	[adventure, animation, children, comedy]	[paul_king]	[ben_whishaw, michael_gambon, imelda_staunton,...
A Bug's Life	[adventure, animation, children, comedy]	[john_lasseter]	[dave_foley, kevin_spacey, julia_louis-dreyfus...
Ribbit	[adventure, animation, children, comedy]	[mamat_khalid]	[johan, nurul_elfira_loy, awie, aznil_hj_nawawi]

Results: Content-Based Recommenders (con't.)

Baseline Model's Recommendations for "Toy Story"

Recommendations based on your interest in Toy Story:

title	genres	director	actors
Turbo	[adventure, animation, children, comedy, fantasy]	[david_soren]	[ryan_reynolds, paul_giamatti, michael_peña, s...
Asterix and the Vikings	[adventure, animation, children, comedy, fantasy]	[stefan_fjeldmark]	[roger_carel, lorànt_deutsch, sara_forestier, ...
Hawaiian Vacation	[adventure, animation, children, comedy, fantasy]	[gary_rydstrom]	[tom_hanks, tim_allen, joan_cusack, don_rickles]
The Adventures of Rocky & Bullwinkle	[adventure, animation, children, comedy, fantasy]	[des_mcanuff]	[rene_russo, jason_alexander, piper_perabo, ra...
A Connecticut Yankee in King Arthur's Court	[adventure, children, comedy, fantasy]	[mel_damski]	[keshia_knight_pulliam, michael_gross, jean_ma...
Jack-Jack Attack	[adventure, animation, children, comedy]	[brad_bird]	[bret_'brook'_parker, bud_luckey, eli_fucile, ...
Help! I'm A Fish	[adventure, animation, comedy]	[stefan_fjeldmark]	[sebastian_jessen, pil_neja, morten_kerrn_niel...
Paddington 2	[adventure, animation, children, comedy]	[paul_king]	[ben_whishaw, michael_gambon, imelda_staunton,...
A Bug's Life	[adventure, animation, children, comedy]	[john_lasseter]	[dave_foley, kevin_spacey, julia_louis-dreyfus...
Ribbit	[adventure, animation, children, comedy]	[mamat_khalid]	[johan, nurul_elfira_loy, awie, aznil_hj_nawawi]

→ These recommendations are essentially based on Genre alone

Results: Content-Based Recommenders (con’t.)

Best (Gensim) Model’s Recommendations for “Toy Story”:

title	genres	director	actors
Hawaiian Vacation	[adventure, animation, children, comedy, fantasy]	[gary_rydstrom]	[tom_hanks, tim_allen, joan_cusack, don_rickles]
Toy Story 2	[adventure, animation, children, comedy, fantasy]	[john_lasseter]	[tom_hanks, tim_allen, joan_cusack, kelsey_gra...
Small Fry	[adventure, animation, children, comedy, fantasy]	[angus_maclane]	[tom_hanks, tim_allen, joan_cusack, estelle_ha...
Toy Story 3	[adventure, animation, children, comedy, fanta...	[lee_unkrich]	[tom_hanks, tim_allen, joan_cusack, don_rickles]
Buzz Lightyear of Star Command: The Adventure ...	[adventure, animation, children, comedy, sci-fi]	[tad_stones]	[tim_allen, nicole_sullivan, stephen_furst, la...
Halloweentown	[adventure, children, comedy, fantasy]	[duwayne_dunham]	[debbie_reynolds, kimberly_j._brown, judith_ho...
Olaf's Frozen Adventure	[adventure, animation, children, comedy, fantasy]	[kevin_deters]	[josh_gad, kristen_bell, idina_menzel, jonatha...
Turbo	[adventure, animation, children, comedy, fantasy]	[david_soren]	[ryan_reynolds, paul_giamatti, michael_peña, s...
The Pasta Detectives	[adventure, children, comedy]	[neele_vollmar]	[anton_petzold, juri_winkler, karoline_herfurt...
Mr. Bug Goes to Town	[animation, children, comedy, fantasy, musical]	[dave_fleischer]	[kenny_gardner, gwen_williams, jack_mercer, te...

Results: Collaborative Recommender

Results: Collaborative Recommender

- Design Assumptions:

Results: Collaborative Recommender

- Design Assumptions:

    1. User must have at least 5 rated movies

Results: Collaborative Recommender

- Design Assumptions:

  1. User must have at least 5 rated movies

  2. Ratings overlap with “similar” users must span at least 5 titles:

Results: Collaborative Recommender

- Design Assumptions:

  1. User must have at least 5 rated movies

  2. Ratings overlap with “similar” users must span at least 5 titles:

- Rationale: determining “Similarity” is difficult using too few titles

Results: Collaborative Recommender

- Design Assumptions:
    1. User must have at least 5 rated movies
    2. Ratings overlap with “similar” users must span at least 5 titles:
- Rationale: determining “Similarity” is difficult using too few titles
- Similarity Measure: **Dot Product** of each movie’s ratings

Results: Collaborative Recommender (con't.)

Recommender Functionality:

1. Specify Minimum Acceptable Rating

```
## Get recommendations for User #5 (who has rated 71 movies)

recommendations(user_id=5, min_rating=4)
```

```
Recommendations Rated 4.0 or Better Based on Your Rating History
(Movies Liked by Similar User # 5403):
```

	movieId	title	projected_rating
0	32	Twelve Monkeys	4.5
1	111	Taxi Driver	4.5
2	356	Forrest Gump	4.5
3	608	Fargo	4.5
4	750	Dr. Strangelove or: How I Learned to Stop Worr...	4.5
5	908	North by Northwest	5.0
6	1057	Everyone Says I Love You	4.5
7	1206	A Clockwork Orange	4.0
8	1230	Annie Hall	5.0
9	1416	Evita	4.0

Results: Collaborative Recommender (con't.)

Recommender Functionality:

1. Specify Minimum Acceptable Rating

2. All overlapping users found

```
## Get recommendations for User #5 (who has rated 71 movies)

recommendations(user_id=5, min_rating=4)
```

```
Recommendations Rated 4.0 or Better Based on Your Rating History
(Movies Liked by Similar User # 5403):
```

	movieId	title	projected_rating
0	32	Twelve Monkeys	4.5
1	111	Taxi Driver	4.5
2	356	Forrest Gump	4.5
3	608	Fargo	4.5
4	750	Dr. Strangelove or: How I Learned to Stop Worr...	4.5
5	908	North by Northwest	5.0
6	1057	Everyone Says I Love You	4.5
7	1206	A Clockwork Orange	4.0
8	1230	Annie Hall	5.0
9	1416	Evita	4.0

Results: Collaborative Recommender (con't.)

Recommender Functionality:

1. Specify Minimum Acceptable Rating
2. All overlapping users found
3. Similarities computed

```
## Get recommendations for User #5 (who has rated 71 movies)

recommendations(user_id=5, min_rating=4)
```

```
Recommendations Rated 4.0 or Better Based on Your Rating History
(Movies Liked by Similar User # 5403):
```

	movieId	title	projected_rating
0	32	Twelve Monkeys	4.5
1	111	Taxi Driver	4.5
2	356	Forrest Gump	4.5
3	608	Fargo	4.5
4	750	Dr. Strangelove or: How I Learned to Stop Worr...	4.5
5	908	North by Northwest	5.0
6	1057	Everyone Says I Love You	4.5
7	1206	A Clockwork Orange	4.0
8	1230	Annie Hall	5.0
9	1416	Evita	4.0

Results: Collaborative Recommender (con't.)

<u>Recommender Functionality</u>:

1. Specify Minimum Acceptable Rating
2. All overlapping users found
3. Similarities computed
4. Most similar user's ratings returned

```
## Get recommendations for User #5 (who has rated 71 movies)

recommendations(user_id=5, min_rating=4)
```

```
Recommendations Rated 4.0 or Better Based on Your Rating History
(Movies Liked by Similar User # 5403):
```

	movieId	title	projected_rating
0	32	Twelve Monkeys	4.5
1	111	Taxi Driver	4.5
2	356	Forrest Gump	4.5
3	608	Fargo	4.5
4	750	Dr. Strangelove or: How I Learned to Stop Worr...	4.5
5	908	North by Northwest	5.0
6	1057	Everyone Says I Love You	4.5
7	1206	A Clockwork Orange	4.0
8	1230	Annie Hall	5.0
9	1416	Evita	4.0

Results: Collaborative Recommender (con't.)

Sanity Check:

```
two_users_overlap(user_id=5, similar_user_id=5403)
```

```
Common Movie Ratings Between User #5 and Similar User #5403:
```

	title	user_rating	similar_user_rating	rating_difference
0	Pulp Fiction	5.0	5.0	0.0
1	The Shawshank Redemption	5.0	4.5	0.5
2	Schindler's List	4.5	4.5	0.0
3	Trainspotting	5.0	5.0	0.0
4	One Flew Over the Cuckoo's Nest	4.0	4.5	-0.5
...	...	...	...	...
28	Gone Baby Gone	4.0	4.0	0.0
29	American Gangster	4.0	4.5	-0.5
30	No Country for Old Men	4.5	4.5	0.0
31	Juno	5.0	5.0	0.0
32	There Will Be Blood	4.5	5.0	-0.5

33 rows × 4 columns

```
Mean Absolute Deviation of rating_difference = 0.4
```

Results: Collaborative Recommender (con't.)

Sanity Check:

1. Overlapping movies identified

```
two_users_overlap(user_id=5, similar_user_id=5403)
```

```
Common Movie Ratings Between User #5 and Similar User #5403:
```

	title	user_rating	similar_user_rating	rating_difference
0	Pulp Fiction	5.0	5.0	0.0
1	The Shawshank Redemption	5.0	4.5	0.5
2	Schindler's List	4.5	4.5	0.0
3	Trainspotting	5.0	5.0	0.0
4	One Flew Over the Cuckoo's Nest	4.0	4.5	-0.5
...	...	...	...	...
28	Gone Baby Gone	4.0	4.0	0.0
29	American Gangster	4.0	4.5	-0.5
30	No Country for Old Men	4.5	4.5	0.0
31	Juno	5.0	5.0	0.0
32	There Will Be Blood	4.5	5.0	-0.5

33 rows × 4 columns

```
Mean Absolute Deviation of rating_difference = 0.4
```

Results: Collaborative Recommender (con't.)

<u>Sanity Check</u>:

1. Overlapping movies identified
2. Rating differences computed

```
two_users_overlap(user_id=5, similar_user_id=5403)
```

```
Common Movie Ratings Between User #5 and Similar User #5403:
```

	title	user_rating	similar_user_rating	rating_difference
0	Pulp Fiction	5.0	5.0	0.0
1	The Shawshank Redemption	5.0	4.5	0.5
2	Schindler's List	4.5	4.5	0.0
3	Trainspotting	5.0	5.0	0.0
4	One Flew Over the Cuckoo's Nest	4.0	4.5	-0.5
...	...	...	...	...
28	Gone Baby Gone	4.0	4.0	0.0
29	American Gangster	4.0	4.5	-0.5
30	No Country for Old Men	4.5	4.5	0.0
31	Juno	5.0	5.0	0.0
32	There Will Be Blood	4.5	5.0	-0.5

33 rows × 4 columns

```
Mean Absolute Deviation of rating_difference = 0.4
```

Results: Collaborative Recommender (con't.)

Sanity Check:

1. Overlapping movies identified
2. Rating differences computed
3. Mean Absolute Deviation returned

   → **Smaller is better**

```
two_users_overlap(user_id=5, similar_user_id=5403)
```

```
Common Movie Ratings Between User #5 and Similar User #5403:
```

	title	user_rating	similar_user_rating	rating_difference
0	Pulp Fiction	5.0	5.0	0.0
1	The Shawshank Redemption	5.0	4.5	0.5
2	Schindler's List	4.5	4.5	0.0
3	Trainspotting	5.0	5.0	0.0
4	One Flew Over the Cuckoo's Nest	4.0	4.5	-0.5
...	...	...	...	...
28	Gone Baby Gone	4.0	4.0	0.0
29	American Gangster	4.0	4.5	-0.5
30	No Country for Old Men	4.5	4.5	0.0
31	Juno	5.0	5.0	0.0
32	There Will Be Blood	4.5	5.0	-0.5

33 rows × 4 columns

```
Mean Absolute Deviation of rating_difference = 0.4
```

Conclusions/Recommendations

1) Content-Based Recommender for “new” users (0-4 rated movies)

Conclusions/Recommendations

1) Content-Based Recommender for “new” users (0-4 rated movies)

2) Hybrid Recommender for users with at least 5 rated movies

Conclusions/Recommendations

1) Content-Based Recommender for “new” users (0-4 rated movies)

2) Hybrid Recommender for users with at least 5 rated movies:
   - Collaborative: Based on prior list of rated movies

Conclusions/Recommendations

1) Content-Based Recommender for "new" users (0-4 rated movies)

2) Hybrid Recommender for users with at least 5 rated movies:
   - Collaborative: Based on prior list of rated movies

   AND

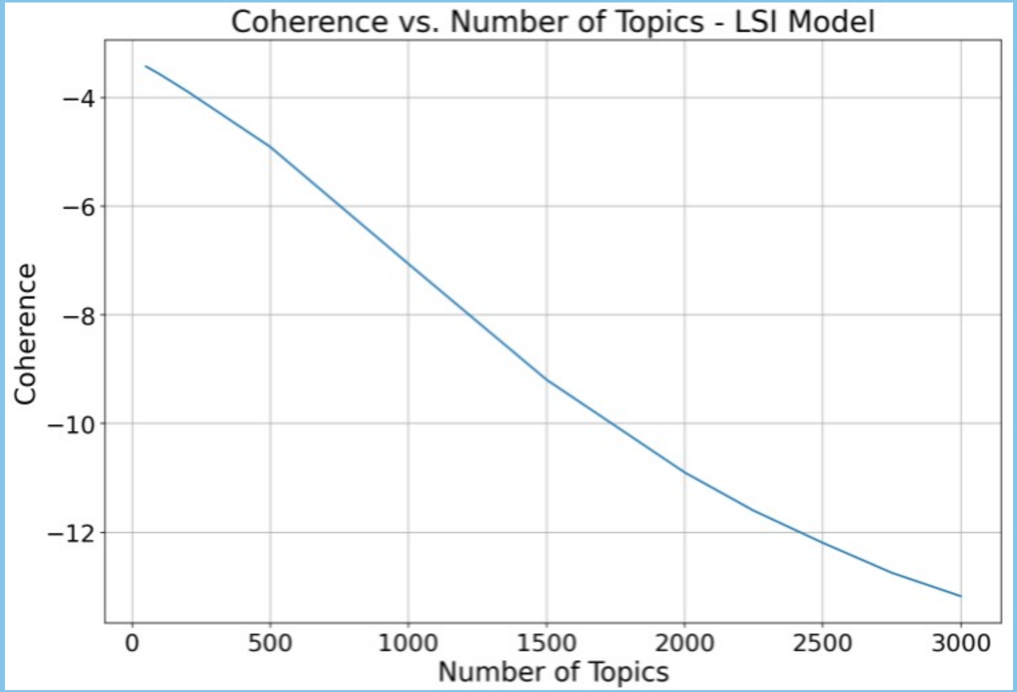
   - Content-Based: Additional list generated as described above

Appendix

Gensim LSI Topic Modelling

<u>Used to Dimension-Reduce Corpus</u>:

- Each movie represented by a linear combination of a subset of the topics
  - → Combinations of topics function as the components of a dimension-reduced space
- <u>Topic Coherence</u>:
  - Measures degree of semantic similarity between high-scoring words in a topic
  - A smaller coherence is better (indicates more cohesion amongst the movies associated with a given topic)
- <u>3000 topics is justified</u>:
  - Represents elbow of coherence values
  - Also, a practical computational limit



Gensim LSI Topic Modelling (con't.)

Some samples from the 3000 Topics generated:

Topic # 1662: -0.063*"socialite" + 0.055*"john_hurt" + 0.054*"alfred_hitchcock" + -0.053*"donald_sutherland" + -0.050*"intent" + 0.048*"rose" + 0.047*"interested" + 0.047*"collect" + -0.047*"headed" + 0.046*"edward_burns"

Topic # 1663: 0.059*"murdering" + -0.052*"influence" + 0.052*"media" + 0.051*"enigmatic" + 0.051*"deals" + 0.051*"reputation" + 0.050*"socialite" + 0.049*"clan" + -0.047*"mining" + -0.047*"immediately"

Topic # 1664: 0.069*"talented" + 0.060*"musicians" + -0.054*"massive" + -0.049*"crush" + 0.049*"sun" + 0.049*"discovering" + 0.049*"headed" + 0.048*"festival" + 0.048*"roman" + -0.048*"deserted"

Topic # 1665: -0.064*"drifter" + -0.060*"colleagues" + -0.058*"exactly" + -0.053*"steps" + 0.052*"thirty" + -0.052*"cowboy" + 0.051*"allows" + -0.051*"peaceful" + -0.048*"trained" + 0.047*"rebellious"

Topic # 1666: 0.063*"faced" + -0.060*"watch" + 0.060*"places" + -0.053*"widower" + -0.049*"gregory_peck" + -0.049*"phone" + -0.049*"completely" + 0.049*"citizens" + 0.045*"fulfill" + -0.043*"weeks"

Gensim Content-Based Recommender

Another Example:

```
Recommendations based on your interest in Star Wars:
```

title	genres	director	actors
The Empire Strikes Back	[action, adventure, sci-fi]	[irvin_kershner]	[mark_hamill, harrison_ford, carrie_fisher, bi...
Return of the Jedi	[action, adventure, sci-fi]	[richard_marquand]	[mark_hamill, harrison_ford, carrie_fisher, bi...
Star Wars: The Force Awakens	[action, adventure, fantasy, sci-fi, imax]	[j.j._abrams]	[harrison_ford, mark_hamill, carrie_fisher, ad...
The Star Wars Holiday Special	[adventure, children, comedy, sci-fi]	[steve_binder]	[harrison_ford, mark_hamill, anthony_daniels, ...
Star Wars: Episode III - Revenge of the Sith	[action, adventure, sci-fi]	[george_lucas]	[hayden_christensen, ewan_mcgregor, natalie_po...
Star Wars: The Last Jedi	[action, adventure, fantasy, sci-fi]	[rian_johnson]	[mark_hamill, carrie_fisher, adam_driver, dais...
Captain America	[action, adventure, sci-fi]	[elmer_clifton]	[dick_purcell, lorna_gray, lionel_atwill, char...
Star Wars: Episode I - The Phantom Menace	[action, adventure, sci-fi]	[george_lucas]	[liam_neeson, ewan_mcgregor, natalie_portman, ...
The Thief of Bagdad	[action, adventure, fantasy]	[raoul_walsh]	[douglas_fairbanks, snitz_edwards, charles_bel...
Captain Video, Master of the Stratosphere	[adventure, sci-fi]	[spencer_gordon_bennet]	[judd_holdren, george_eldredge, gene_roth, lar...

Gensim Content-Based Recommender (con't.)

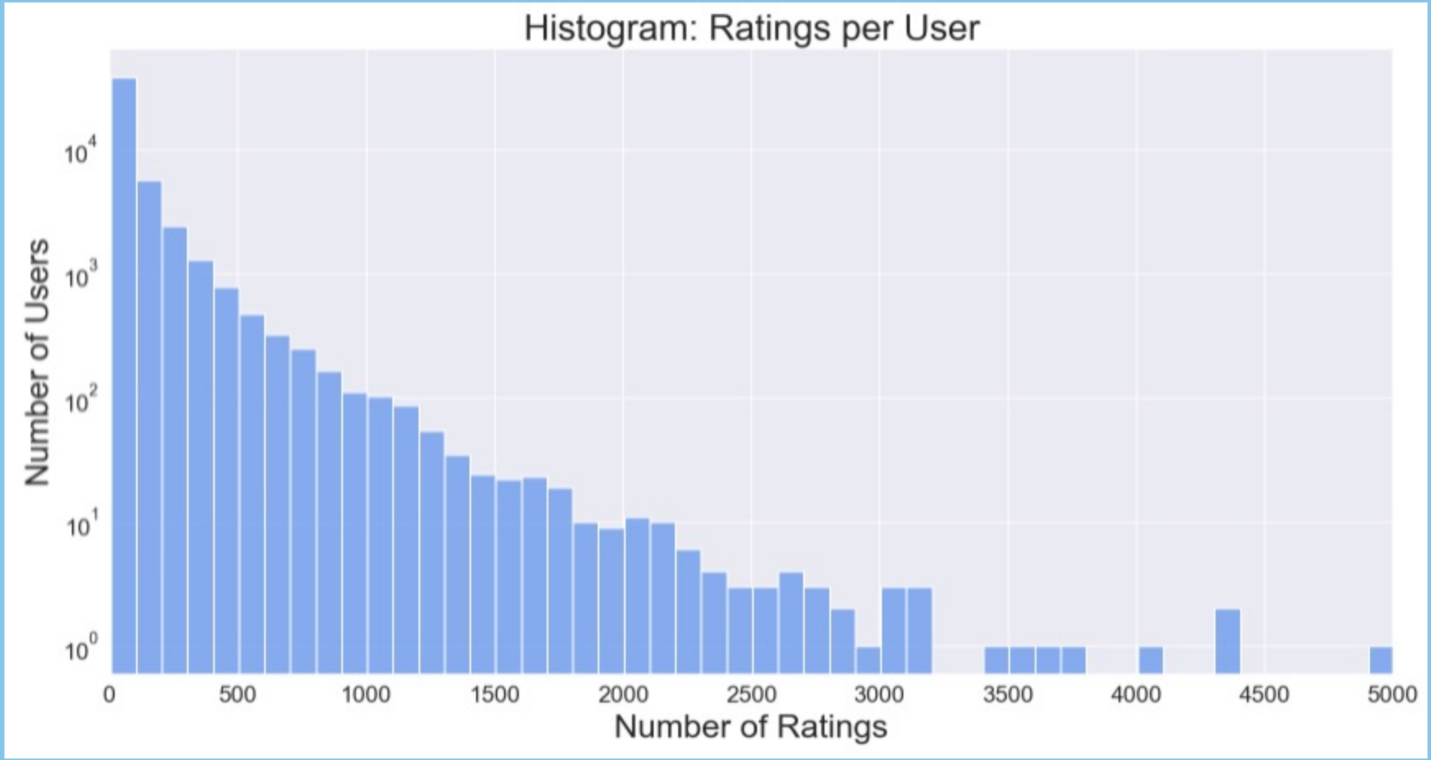
Another Example:

```
Recommendations based on your interest in Rocky:
```

title	genres	director	actors
Rocky III	[action, drama]	[sylvester_stallone]	[sylvester_stallone, talia_shire, burt_young, ...
Creed	[drama]	[ryan_coogler]	[michael_b._jordan, sylvester_stallone, tessa_...
Rocky IV	[action, drama]	[sylvester_stallone]	[sylvester_stallone, talia_shire, carl_weather...
Rocky Balboa	[action, drama]	[sylvester_stallone]	[sylvester_stallone, burt_young, antonio_tarve...
Rocky V	[action, drama]	[john_g._avildsen]	[sylvester_stallone, talia_shire, burt_young, ...
Rocky II	[action, drama]	[sylvester_stallone]	[sylvester_stallone, talia_shire, burt_young, ...
Black Night	[drama]	[olivier_smolders]	[fabrice_rodriguez, yves-marie_gnahoua, marie_...
Knockout	[action, drama]	[lorenzo_doumani]	[sophia_adella_luke, eduardo_yáñez, tony_plana...
The Bronx Bull	[drama]	[martin_guigui]	[william_forsythe, joe_mantegna, paul_sorvino,...
Final Impact	[action]	[joseph_merhi]	[lorenzo_lamas, kathleen_kinmont, michael_wort...

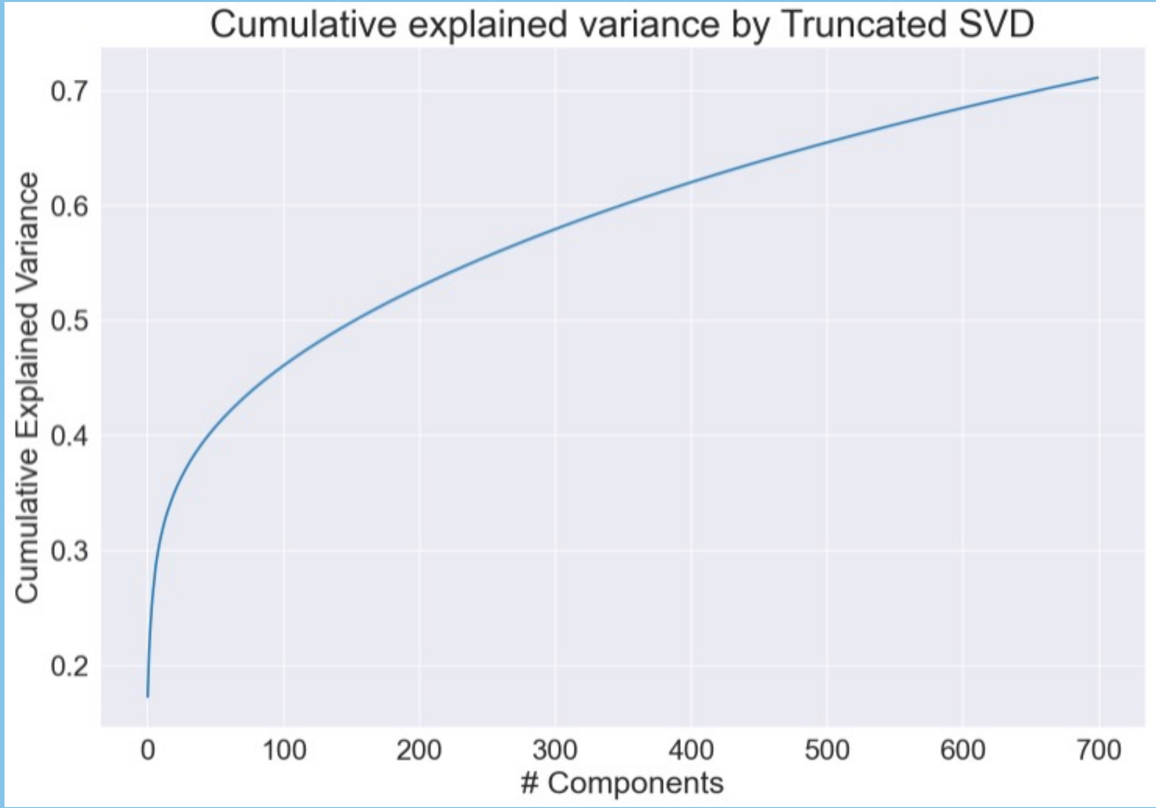
Collaborative Recommender Details

Most users have rated < 100 movies (note that y-axis is log-scale):



Collaborative Recommender Details (con't.)

- User-Ratings Matrix (X):
  - 50,000 rows (1 per user)
  - 28,044 columns (movies)
- Apply Truncated SVD:
  - $X = U \Sigma V^T$
  - 700 Components
  - 71.1% Explained Variance
- Resulting User Matrix (U):
  - 50,000 rows (1 per user)
  - 700 columns (components)



Collaborative Recommender Details (con't.)

Another Example:

```
## Get recommendations for User #2 (who has rated 15 movies)

recommendations(user_id=2, min_rating=5)
```

```
Recommendations Rated 5.0 or Better Based on Your Rating History
(Movies Liked by Similar User # 95643):
```

	movieId	title	projected_rating
0	50	The Usual Suspects	5.0
1	296	Pulp Fiction	5.0
2	318	The Shawshank Redemption	5.0
3	778	Trainspotting	5.0
4	858	The Godfather	5.0
5	1193	One Flew Over the Cuckoo's Nest	5.0
6	1617	L.A. Confidential	5.0
7	2858	American Beauty	5.0
8	2959	Fight Club	5.0
9	4011	Snatch	5.0

```
two_users_overlap(user_id=2, similar_user_id=95643)
```

```
Common Movie Ratings Between User #2 and Similar User #95643:
```

	title	user_rating	similar_user_rating	rating_difference
0	Hackers	3.5	3.5	0.0
1	Sex, Lies, and Videotape	3.5	4.0	-0.5
2	Harold and Maude	3.0	5.0	-2.0
3	Manhattan	3.0	4.5	-1.5
4	A Room with a View	4.5	5.0	-0.5
5	Stripes	3.0	4.5	-1.5
6	Driving Miss Daisy	4.0	4.5	-0.5
7	L.A. Story	3.5	5.0	-1.5
8	The Big Chill	4.0	4.5	-0.5
9	Little Shop of Horrors	4.0	4.5	-0.5
10	Risky Business	3.5	4.0	-0.5
11	American Graffiti	4.0	4.5	-0.5

```
Mean Absolute Deviation of rating_difference = 0.5
```

Collaborative Recommender Details (con’t.)

Edge Cases:

```
## Get recommendations for User #41 (who has rated fewer than 5 movies)

recommendations(41, user_ids, U, X, movie_ids, movies, ratings_count_df, min_rating=5)
```

```
Sorry, the user must have at least 5 previously-rated movies to find similar users.
```

```
## Get recommendations for User #94843 (who has rated 4965 movies - more than any other user in the final users_ratings_matrix)

recommendations(user_id=94843, min_rating=3)
```

```
Sorry, User #94843 does not have any similar users.
```

```
two_users_overlap(user_id=44276, similar_user_id=105056)
```

```
Sorry, User #44276 and User #105056 have not rated enough of the same movies and therefore cannot be compared.
```