Movie Recommender

Dmitriy Pavlov - Data Scientist
GA - DSI 6
February 12,2019

Overview

- Data Science Problem
- Data Collection
- Exploratory Data Analysis
  - Rating Insights
  - Genre Insights
- Segmentation
- Recommenders
  - Feature Based vs Collaborative
  - Recommendations

Data Science Problem:

Build out a recommender engine for movies based on movie features and collaborative user rating. In addition gather greater insights on the movie cohorts.

Exploratory Data Analysis

Data Collection: MovieLens public 20m dataset was used for this project

Files	ratings.csv	tags.csv	movies.csv	links.csv	genome-score.csv	genome-tags.csv
Data	userId	userId	movieId	movieId	movieId	tagId
	movieId	movieId	title	imdbId	tagId	tag
	rating	tag	genres	tmdbId	relevance	
	timestamp	timestamp				

EDA: The last decade has the most movies, but they are not the best rated
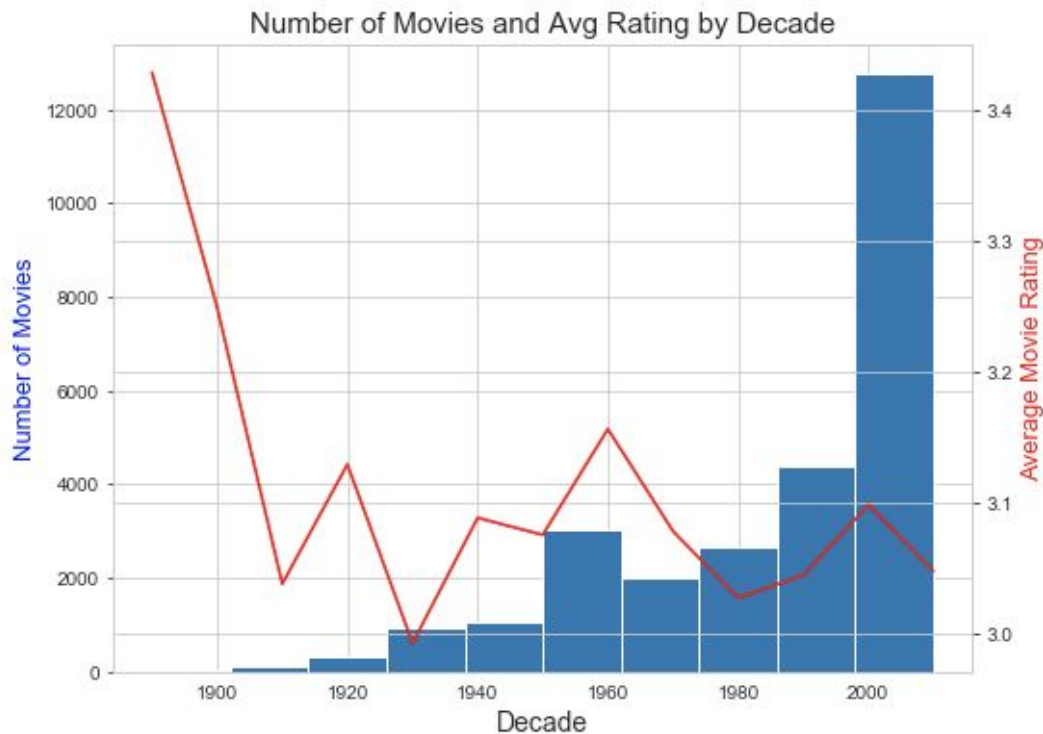
Number of Movies and Avg Rating by Decade

EDA: As complexity of movies increase the rating increases as well

Number of Genres and Avg Rating by Decade

EDA: As complexity of movies increase the rating increases as well

Number of Genres and Avg Rating by Decade
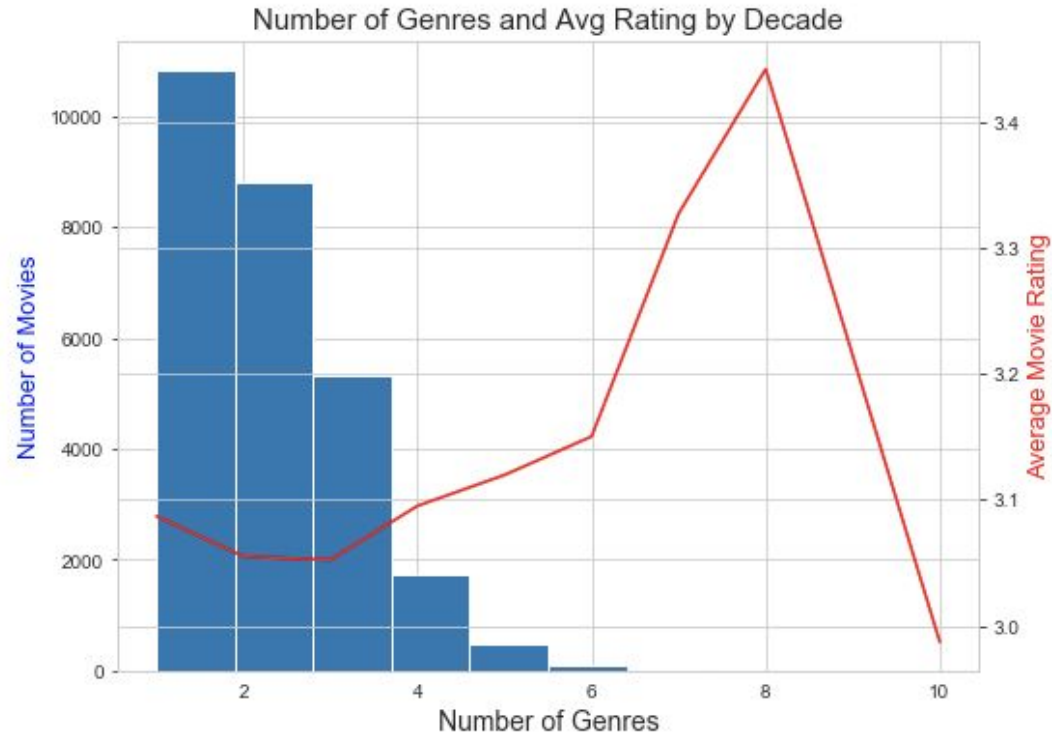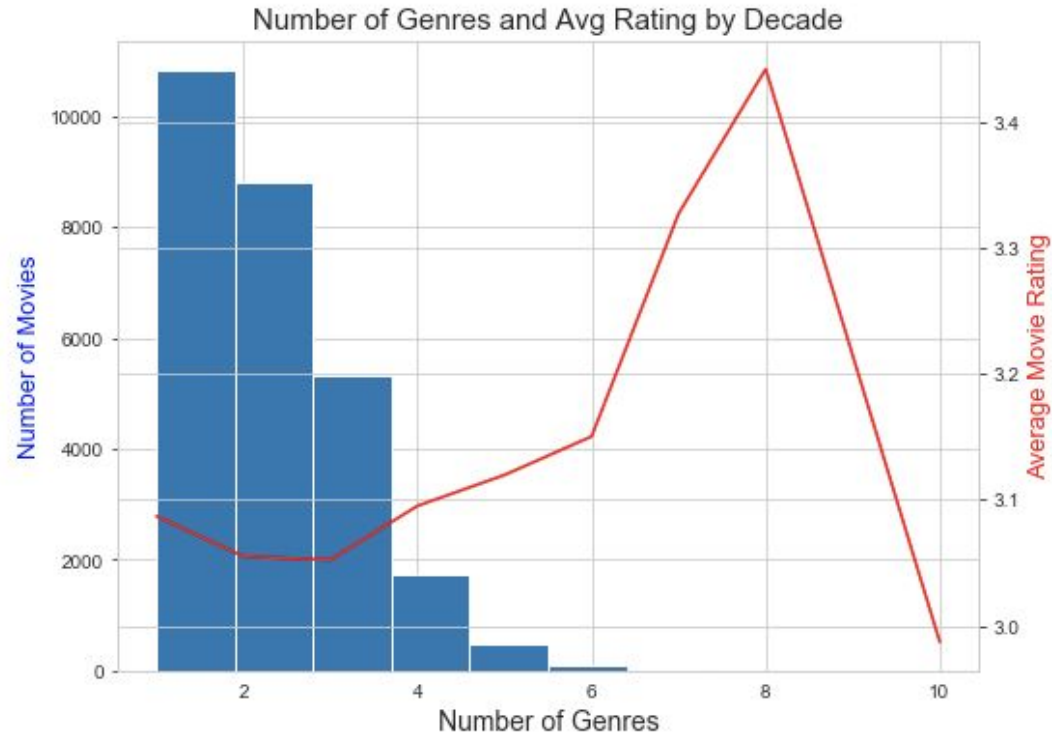
EDA: Noir movies tend to have the highest rating, while horror are rated the worst

Avg Movie Rating by Genre
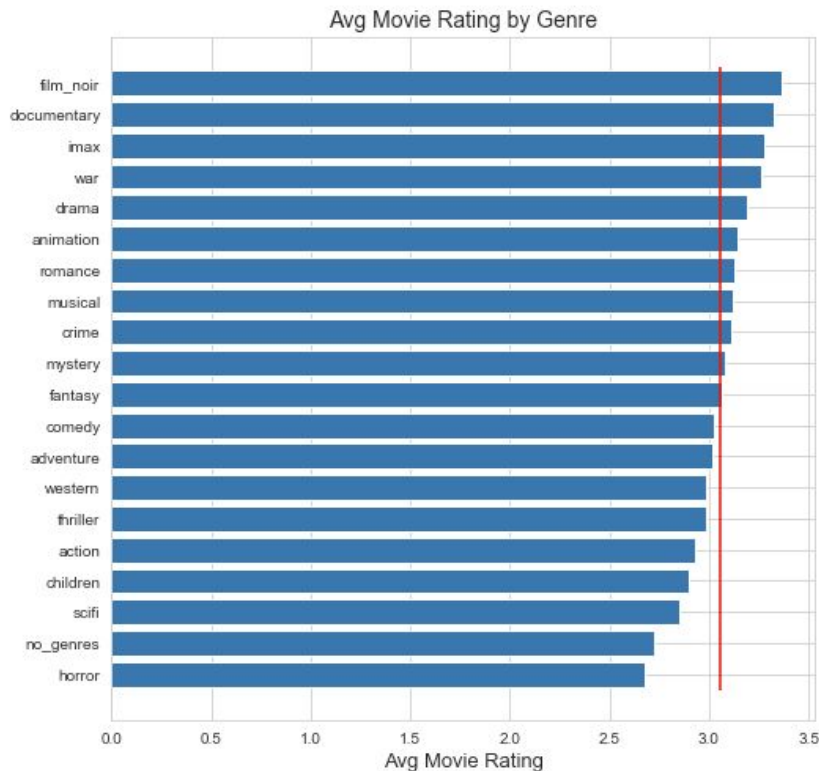
EDA: Movie genres change over time, with documentaries and thrillers increasing the most from 1980 to 2010 while comedy and horror declined
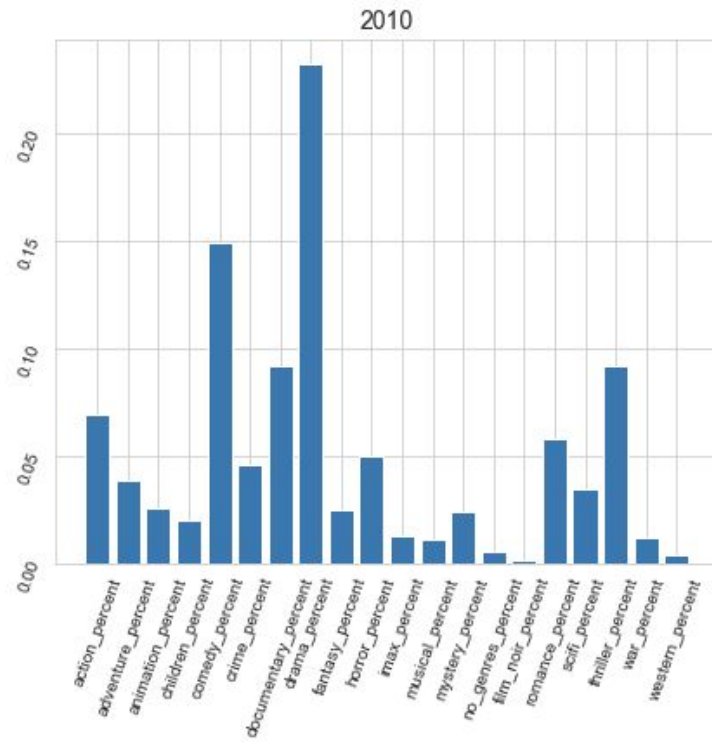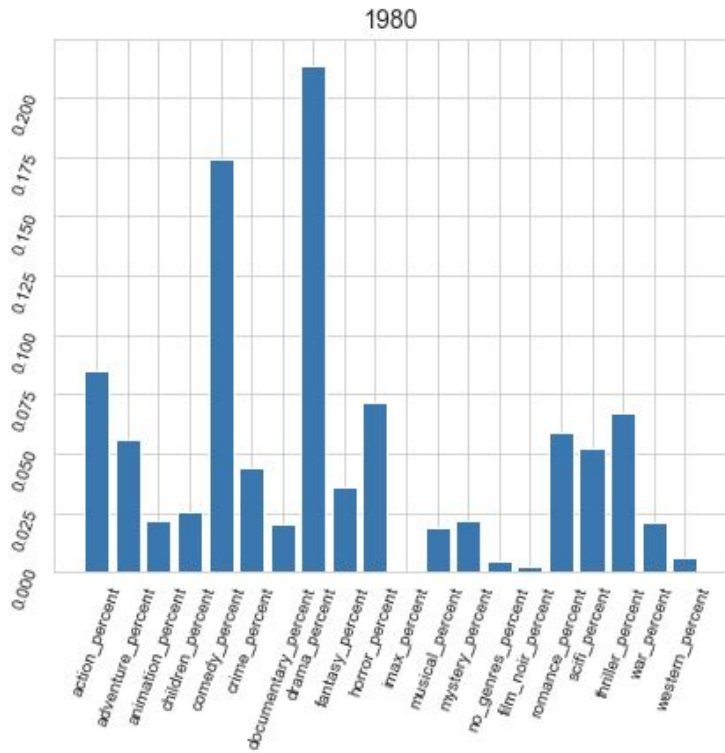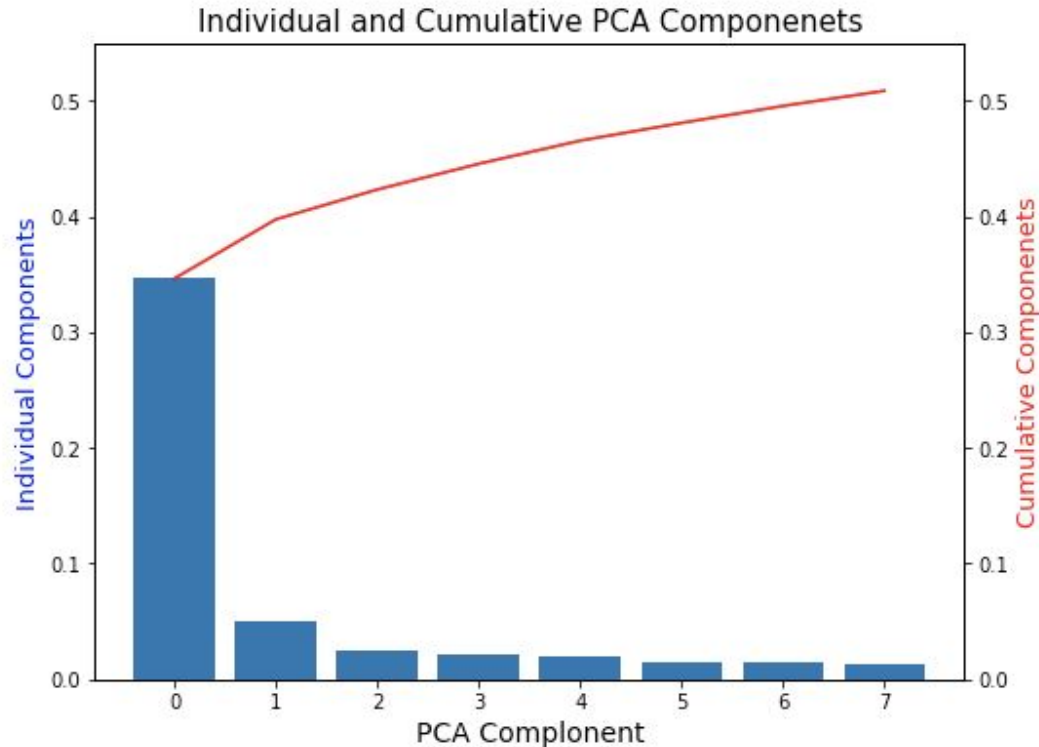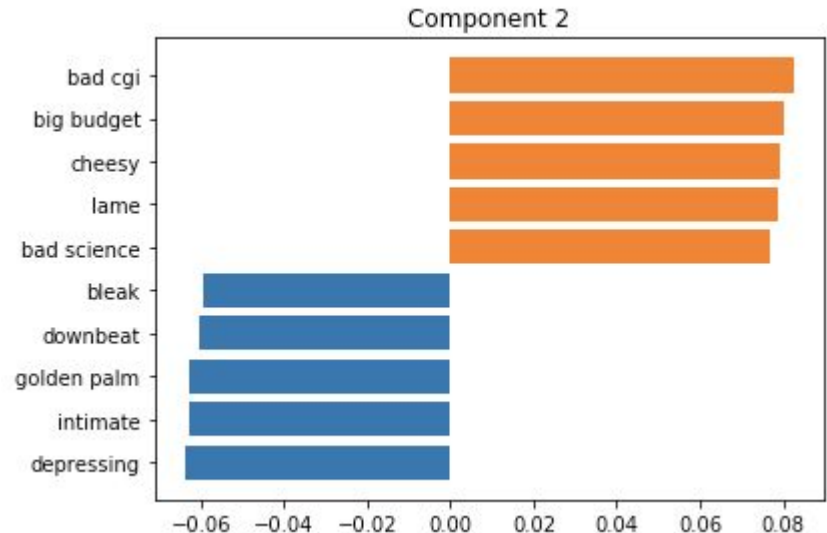
Principal Component Analysis: Segmentation

Segmentation: First 8 PCA components explain over 50% of the variability within films

Segmentation: PCA components can be analyzed further to grasp a better understanding of genre and genome variability

Full list of components can be found in Notebook 5
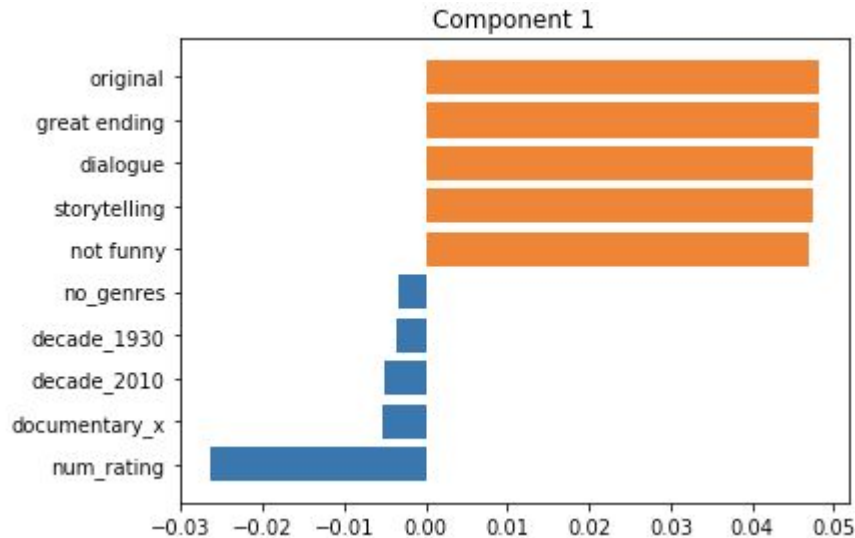
Recommenders

Recommenders: Features and collaborative item based recommenders

Feature Based Recommender - uses movie features such us genres, genomes, and decade in order to recommend similar movies

Collaborative Item Based Recommender - uses user rating to find movies that are rated in similar fashion, answers the question - "*People who enjoy this movie also enjoyed this {}*"

Movie: Monty Python and the Holy Grail

Feature Based Recommender

1. Monty Python's Life of Brian (1979)
2. Young Frankenstein (1974)
3. Fish Called Wanda, A (1988)
4. Sleeper (1973)
5. Bananas (1971)
6. Duck Soup (1933)
7. Man Who Would Be King, The (1975)
8. This Is Spinal Tap (1984)
9. Groundhog Day (1993)
10. Raising Arizona (1987)

Collaborative Item Based Recommender

1. Monty Python's Life of Brian (1979)
2. Indiana Jones and the Raiders of the Lost Ark (1981)
3. Star Wars: Episode V - The Empire Strikes Back (1980)
4. Princess Bride, The (1987)
5. Back to the Future (1985)
6. Groundhog Day (1993)
7. Indiana Jones and the Last Crusade (1989)
8. Star Wars: Episode IV - A New Hope (1977)
9. Fish Called Wanda, A (1988)
10. Blade Runner (1982)

Movie: Jurassic Park

Feature Based Recommender

1. Terminator 2: Judgment Day (1991)
2. Speed (1994)
3. Independence Day (a.k.a. ID4) (1996)
4. Timecop (1994)
5. Twister (1996)
6. Cliffhanger (1993)
7. Virtuosity (1995)
8. Executive Decision (1996)
9. Dragonheart (1996)
10. Star Trek: Generations (1994)

Collaborative Item Based Recommender

1. Terminator 2: Judgment Day (1991)
2. Forrest Gump (1994)
3. Fugitive, The (1993)
4. Speed (1994)
5. Batman (1989)
6. True Lies (1994)
7. Apollo 13 (1995)
8. Braveheart (1995)
9. Silence of the Lambs, The (1991)
10. Dances with Wolves (1990)

Movie: Cinema Paradiso

Feature Based Recommender

1. Manon of the Spring (Manon des sources) (1986)
2. Jean de Florette (1986)
3. Dead Poets Society (1989)
4. Stand by Me (1986)
5. On Golden Pond (1981)
6. My Life as a Dog (Mitt liv som hund) (1985)
7. Kolya (Kolja) (1996)
8. Mediterraneo (1991)
9. Sling Blade (1996)
10. Big Blue, The (Grand bleu, Le) (1988)

Collaborative Item Based Recommender

1. Room with a View, A (1986)
2. Graduate, The (1967)
3. My Left Foot (1989)
4. Amadeus (1984)
5. Annie Hall (1977)
6. Casablanca (1942)
7. When Harry Met Sally... (1989)
8. Manhattan (1979)
9. My Life as a Dog (Mitt liv som hund) (1985)
10. Crying Game, The (1992)

Movie: Pan's Labyrinth

Feature Based Recommender

1. Fall, The (2006)
2. Let the Right One In (2008)
3. Orphanage, The (Orfanato, El) (2007)
4. Babel (2006)
5. Perfume: The Story of a Murderer (2006)
6. Letters from Iwo Jima (2006)
7. Slumdog Millionaire (2008)
8. Atonement (2007)
9. Dumplings (Gaau ji) (2004)
10. Tsotsi (2005)

Collaborative Item Based Recommender

1. Little Miss Sunshine (2006)
2. Children of Men (2006)
3. Prestige, The (2006)
4. Departed, The (2006)
5. V for Vendetta (2006)
6. No Country for Old Men (2007)
7. 300 (2007)
8. Casino Royale (2006)
9. Eternal Sunshine of the Spotless Mind (2004)
10. Juno (2007)

Movie: Napoleon Dynamite

Feature Based Recommender

1. Anchorman: The Legend of Ron Burgundy (2004)
2. Harold and Kumar Go to White Castle (2004)
3. PCU (1994)
4. Envy (2004)
5. Stuck on You (2003)
6. Dodgeball: A True Underdog Story (2004)
7. Confessions of a Teenage Drama Queen (2004)
8. Lonely Guy, The (1984)
9. Friday After Next (2002)
10. I Heart Huckabees (2004)

Collaborative Item Based Recommender

1. Garden State (2004)
2. Life Aquatic with Steve Zissou, The (2004)
3. 40-Year-Old Virgin, The (2005)
4. Dodgeball: A True Underdog Story (2004)
5. School of Rock (2003)
6. Shaun of the Dead (2004)
7. Wedding Crashers (2005)
8. Anchorman: The Legend of Ron Burgundy (2004)
9. I Heart Huckabees (2004)
10. Super Size Me (2004)