**Abstract:**

Producing a movie can represent a significant investment of time and resources by a movie company so due diligence is required to ensure that a new movie will be popular with potential audiences. Our team (Team 14) proposes that useful information can be extracted from a data set of over one million movie reviews and used to guide the development of a new movie. We first performed exploratory data analysis (EDA) to better understand the primary characteristics of the movie review data set. Guided by the information collected through EDA, we made various conjectures on the interactions and relationships between movie reviewers and their assessments of movie quality in order to identify areas of risk and opportunity.

Next, we identified popular movies based on number of reviews and analyzed them in search of common attributes. We observed that the action genre is associated with 7 out of 10 popular movies. While the standard deviation for ratings wasn’t significantly different between action and non-action genre movies, we did find that the correlation coefficient for gender actually *improved* when considering only movies with the Action genre indicator. So, our analysis indicated that action genre movies are similarly liked between Men and Women, the distribution of the ratings is similar to non-action movies and seven out of ten movies identified as popular are associated with the action genre.

Finally, we recommend that the movie company create a minimally viable movie product, in the form of a movie trailer, for an action genre movie. Doing so will make the movie available to potential consumers as soon as possible and provide the opportunity to collect as they watch, comment on and rate the trailer in various channels such as YouTube or IMDB.com.

**Motivation:**

Our team believes there is actionable information hidden in the data set that can help a movie production company experience success at the box office. It is our desire to unlock the value through the application of statistical rigor to identify the types of movies that are popular and experience fiscal success at the box office. Furthermore, we hope to share our findings with a movie production company anxious to improve their performance at the box office and improve their ranking within the industry.

**Introduction:**

As a starting point our team will convey some basic details regarding the MovieLens 1 million movie reviews data set. This data set contains data about users, movies and ratings provided by thousands of users for thousands of movies. Specifically, we analyzed three files which, when combined, collectively contained “1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users who joined MovieLens in 2000” [1.1].

User information comes from the users.dat file and contains demographic information regarding age range, gender, ZIP code and occupation categories.

Movie information comes from the movies.dat file and contains information such as Movie ID, title and genres. “Genre is the term for any category of literature or other forms of art or entertainment, e.g. music, whether written or spoken, audio or visual, based on some set of stylistic criteria. Genres are formed by conventions that change over time as new genres are invented and the use of old ones are discontinued. Often, works fit into multiple genres by way of borrowing and recombining these conventions”[1.2]. Genres are pipe-separated and are selected from the following genres: Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western[1.3].

Ratings information comes from the ratings.dat file and contains information for linking ratings to users and movies. The file of course also contains ratings (which are made on a 5-star scale and are whole-star ratings only) and a timestamp “represented in seconds since the epoch as returned by time(2)” [1.4]. Each user has at least 20 ratings in the ratings.dat file.

This data set contains significant insights into the relationships and interactions between user traits and movie traits and ultimately their potential effects on users’ ratings for movies. Team 14 will explore these areas and share our findings in this paper. As a first step, we will do some exploratory data analysis and present some basic details of the data our team collected and analyzed.

**Part 1 – Importing the MovieLens 1M data set and performing Exploratory Data Analysis (EDA)**

The perfunctory step was to import and merge the three files that make up the 1 million movie reviews data set. As mentioned earlier, three data files were imported then merged into a single Pandas dataframe. With the single dataframe established the results was then save to an external file in order to preserve this initial, combined data set. The file was saved in the HDF5 format which is “a data model, library, and file format for storing and managing data. [1.5]”

Some basic details for the MoveLens users are shown below with help from the Matplotlib library:

|  |  |  |
| --- | --- | --- |
| User Age, Histogram | User Gender, Bar Chart | Age and Gender Table |
|  |  | F M  age  1 78 144  18 298 805  25 558 1,538  35 338 855  45 189 361  50 146 350  56 102 278  1,709 4,331 |

Next, we took a look at the ratings data. Using pandas data frames and pivot tables we examined how many movies had an average rating greater than 4.5 star (on a 5-star scale) for the entire user population, then again by gender. With this first analysis, we already see some variation in ratings depending on the gender of the user. The findings are below and details are available in Appendix 1-A:

|  |  |  |
| --- | --- | --- |
| **All Users** | **Male Users** | **Female Users** |
| Question: How many movies have an average rating over 4.5 overall?  Answer: 21 | Question: How many movies have an average rating over 4.5 among men?  Answer: **23** | Question: How many movies have an average rating over 4.5 among women?  Answer: **51** |

We then considered gender and age when assessing movies with a median rating greater than 4.5. The first step was to restrict the ratings to Men and Women over the age of 30. Then we applied the aggregate function ‘median’ which effectively returned all movies with a rating of 5 stars due to the nature of the rating system where users could provide only whole star ratings.

It is important to note that the filter for age > 30 returns age ranges slightly out of scope. The age field is actually a code for an age range. So, when we restricted the data frame to “> 30” in reality we are analyzing a collection of age ranges; “35-44”, “45-49”, “50-55” and “56+”. So, the median ratings are actually for Men and Women *over 34.* The findings are below and details are available in Appendix 1-B:

|  |  |
| --- | --- |
| **Men over 30 with Median Movie Rating > 4.5** | **Women over 30 with Median Movie Rating > 4.5** |
| Question: How many movies have a median rating over 4.5 among men over age 30?  Answer: **86** | Question: How many movies have a median rating over 4.5 among women over age 30?  Answer: **149** |

A final basic analysis was to determine what are the Top 10 Popular Movies. The definition of “popular” can be subject to debate but our method was to simply identify the number of ratings per movie using pandas and pivot tables, sort the list in descending order by number of ratings and slice off the top 10 rows. The justification for this method being that "popular" movies are those which have the greatest number of people sharing their opinion about the movie. If a person shares their opinion about a movie we can safely assume that means the person saw the movie. So we can then consider each movie review to indicate a movie was watched, and each viewing of a movie is in essence a "vote" in favor of the movie being popular.

A movie can be good or bad by any number of movie critic standards but what likely matters to a movie company is revenue. In fact, some movies can be so bad by most movie critic standards that they actually develop a cult following. So, regardless of the ratings, what ultimately matters is if people were willing to pay for a ticket to see a given movie.

With that justification in mind, we applied the method described above which yielded the following results for Top 10 Popular Movies. A pivot table was created to calculate the mean rating and count of ratings by movie title as well as show the genres associated with each movie. The results were sorted in descending order by number of reviews and the top ten rows were sliced from the data frame which yielded the following:

|  |
| --- |
| **Ten Most Popular Movies**  (descending order by Number of Reviews) |
| | Movie Title | Genres | Rating Mean | Number of Reviews | | --- | --- | --- | --- | | American Beauty (1999) | Comedy|Drama | 4.317386 | 3428 | | Star Wars: Episode IV - A New Hope (1977) | Action|Adventure|Fantasy|Sci-Fi | 4.453694 | 2991 | | Star Wars: Episode V - The Empire Strikes Back (1980) | Action|Adventure|Drama|Sci-Fi|War | 4.292977 | 2990 | | Star Wars: Episode VI - Return of the Jedi (1983) | Action|Adventure|Romance|Sci-Fi|War | 4.022893 | 2883 | | Jurassic Park (1993) | Action|Adventure|Sci-Fi | 3.763847 | 2672 | | Saving Private Ryan (1998) | Action|Drama|War | 4.337354 | 2653 | | Terminator 2: Judgment Day (1991) | Action|Sci-Fi|Thriller | 4.058513 | 2649 | | Matrix, The (1999) | Action|Sci-Fi|Thriller | 4.315830 | 2590 | | Back to the Future (1985) | Comedy|Sci-Fi | 3.990321 | 2583 | | Silence of the Lambs, The (1991) | Drama|Thriller | 4.351823 | 2578 | |

The veracity of our list of popular movies is supported by the following data from Box Office Mojo (http: //www. boxofficemojo.com). This supplemental data from Box Office Mojo shows domestic, wide-opening weekend gross dollars earned for each movie title in our Ten Most Popular Movies list. Nine out of ten movie titles in our Ten Most Popular Movies list were the top grossing movies for their wide-opening weekend.

|  |  |
| --- | --- |
| **Movie Title** | **Wide-Opening Weekend: Rank & Gross Dollars** |
| American Beauty: | No.3 rank, 706 theaters, $11,598 average |
| Star Wars: Episode IV - A New Hope: | No.1 rank, 757 theaters, $8,992 average |
| Star Wars: Episode V - The Empire Strikes Back 1980): | No.1 rank, 823 theaters, $13,171 average |
| Star Wars: Episode VI - Return of the Jedi (1983): | No. 1 rank, 1,002 theaters, $22,973 |
| Jurassic Park (1993): | No. 1 rank, 2,404 theaters, $19,561 average |
| Saving Private Ryan (1998): | No. 1 rank, 2,463 theaters, $12,414 average |
| Terminator 2: Judgment Day (1991): | (Non-3D) No. 1 rank, 2,463 theaters, $12,414 average  (3D) No. 25 rank, 371 theaters, $1,490 average |
| Matrix, The (1999): | No. 1 rank, 2,849 theaters, $9,753 average |
| Back to the Future (1985): | No. 1 rank, 1,420 theaters, $7,853 average |
| Silence of the Lambs, The (1991): | No. 1 rank, 1,497 theaters, $9,196 average |

**Appendix 1-A: Average rating greater than 4.5 star**

|  |  |  |
| --- | --- | --- |
| All users | Male Users | Female Users |
| | title | rating | | --- | --- | | One Little Indian (1973) | 5.000000 | | Schlafes Bruder (Brother of Sleep) (1995) | 5.000000 | | Bittersweet Motel (2000) | 5.000000 | | Ulysses (Ulisse) (1954) | 5.000000 | | Follow the Bitch (1998) | 5.000000 | | Gate of Heavenly Peace, The (1995) | 5.000000 | | Song of Freedom (1936) | 5.000000 | | Lured (1947) | 5.000000 | | Baby, The (1973) | 5.000000 | | Smashing Time (1967) | 5.000000 | | I Am Cuba (Soy Cuba/Ya Kuba) (1964) | 4.800000 | | Lamerica (1994) | 4.750000 | | Apple, The (Sib) (1998) | 4.666667 | | Sanjuro (1962) | 4.608696 | | Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954) | 4.560510 | | Shawshank Redemption, The (1994) | 4.554558 | | Godfather, The (1972) | 4.524966 | | Close Shave, A (1995) | 4.520548 | | Usual Suspects, The (1995) | 4.517106 | | Schindler's List (1993) | 4.510417 | | Wrong Trousers, The (1993) | 4.507937 | | | title | gender | rating | | --- | --- | --- | | Angela (1995) | M | 5.000000 | | Gate of Heavenly Peace, The (1995) | M | 5.000000 | | Ulysses (Ulisse) (1954) | M | 5.000000 | | Smashing Time (1967) | M | 5.000000 | | Small Wonders (1996) | M | 5.000000 | | Schlafes Bruder (Brother of Sleep) (1995) | M | 5.000000 | | Lured (1947) | M | 5.000000 | | Follow the Bitch (1998) | M | 5.000000 | | Dangerous Game (1993) | M | 5.000000 | | Bells, The (1926) | M | 5.000000 | | Baby, The (1973) | M | 5.000000 | | Time of the Gypsies (Dom za vesanje) (1989) | M | 4.833333 | | I Am Cuba (Soy Cuba/Ya Kuba) (1964) | M | 4.750000 | | Lamerica (1994) | M | 4.666667 | | Window to Paris (1994) | M | 4.666667 | | Sanjuro (1962) | M | 4.639344 | | Apple, The (Sib) (1998) | M | 4.600000 | | For All Mankind (1989) | M | 4.583333 | | Godfather, The (1972) | M | 4.583333 | | Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954) | M | 4.576628 | | Shawshank Redemption, The (1994) | M | 4.560625 | | Raiders of the Lost Ark (1981) | M | 4.520597 | | Usual Suspects, The (1995) | M | 4.518248 | | | title | gender | rating | | --- | --- | --- | | 24 7: Twenty Four Seven (1997) | F | 5.000000 | | Dancemaker (1998) | F | 5.000000 | | Raw Deal (1948) | F | 5.000000 | | Prisoner of the Mountains (Kavkazsky Plennik) (1996) | F | 5.000000 | | Song of Freedom (1936) | F | 5.000000 | | Other Side of Sunday, The (Søndagsengler) (1996) | F | 5.000000 | | One Little Indian (1973) | F | 5.000000 | | Message to Love: The Isle of Wight Festival (1996) | F | 5.000000 | | Lamerica (1994) | F | 5.000000 | | I Am Cuba (Soy Cuba/Ya Kuba) (1964) | F | 5.000000 | | Twice Upon a Yesterday (1998) | F | 5.000000 | | Woman of Paris, A (1923) | F | 5.000000 | | Gate of Heavenly Peace, The (1995) | F | 5.000000 | | Gambler, The (A Játékos) (1997) | F | 5.000000 | | For the Moment (1994) | F | 5.000000 | | Skipped Parts (2000) | F | 5.000000 | | Saltmen of Tibet, The (1997) | F | 5.000000 | | Belly (1998) | F | 5.000000 | | Big Combo, The (1955) | F | 5.000000 | | Country Life (1994) | F | 5.000000 | | Coldblooded (1995) | F | 5.000000 | | Ayn Rand: A Sense of Life (1997) | F | 5.000000 | | Clean Slate (Coup de Torchon) (1981) | F | 5.000000 | | Brother, Can You Spare a Dime? (1975) | F | 5.000000 | | Bittersweet Motel (2000) | F | 5.000000 | | Ballad of Narayama, The (Narayama Bushiko) (1958) | F | 5.000000 | | Battling Butler (1926) | F | 5.000000 | | World of Apu, The (Apur Sansar) (1959) | F | 4.842105 | | Hearts and Minds (1996) | F | 4.833333 | | Apple, The (Sib) (1998) | F | 4.750000 | | Rain (1932) | F | 4.750000 | | Marcello Mastroianni: I Remember Yes, I Remember (1997) | F | 4.666667 | | Among Giants (1998) | F | 4.666667 | | Panther (1995) | F | 4.666667 | | Aparajito (1956) | F | 4.666667 | | Close Shave, A (1995) | F | 4.644444 | | Firelight (1997) | F | 4.600000 | | Before the Rain (Pred dozhdot) (1994) | F | 4.600000 | | Flower of My Secret, The (La Flor de Mi Secreto) (1995) | F | 4.600000 | | Wrong Trousers, The (1993) | F | 4.588235 | | General, The (1927) | F | 4.575758 | | Sunset Blvd. (a.k.a. Sunset Boulevard) (1950) | F | 4.572650 | | Pather Panchali (1955) | F | 4.571429 | | Wallace & Gromit: The Best of Aardman Animation (1996) | F | 4.563107 | | Schindler's List (1993) | F | 4.562602 | | Grand Illusion (Grande illusion, La) (1937) | F | 4.560976 | | Shawshank Redemption, The (1994) | F | 4.539075 | | Grand Day Out, A (1992) | F | 4.537879 | | To Kill a Mockingbird (1962) | F | 4.536667 | | Creature Comforts (1990) | F | 4.513889 | | Usual Suspects, The (1995) | F | 4.513317 | |

**Appendix 1-B: Women and Men over 30 with Median Movie Rating > 4.5**

|  |
| --- |
| Movies with a median rating over 4.5 among Men over age 30 |
| 42 Up (1998) M 5.0  All Quiet on the Western Front (1930) M 5.0  American Beauty (1999) M 5.0  Among Giants (1998) M 5.0  Angela (1995) M 5.0  Ayn Rand: A Sense of Life (1997) M 5.0  Bells, The (1926) M 5.0  Bicycle Thief, The (Ladri di biciclette) (1948) M 5.0  Boat, The (Das Boot) (1981) M 5.0  Bridge on the River Kwai, The (1957) M 5.0  Carmen (1984) M 5.0  Casablanca (1942) M 5.0  Chinatown (1974) M 5.0  Chushingura (1962) M 5.0  Citizen Kane (1941) M 5.0  City Lights (1931) M 5.0  Close Shave, A (1995) M 5.0  Conformist, The (Il Conformista) (1970) M 5.0  Creature Comforts (1990) M 5.0  Double Indemnity (1944) M 5.0  Dr. Strangelove or: How I Learned to Stop Worry... M 5.0  Ed's Next Move (1996) M 5.0  Fargo (1996) M 5.0  Firelight (1997) M 5.0  Follow the Bitch (1998) M 5.0  For All Mankind (1989) M 5.0  Gambler, The (A Játékos) (1997) M 5.0  General, The (1927) M 5.0  Godfather, The (1972) M 5.0  Godfather: Part II, The (1974) M 5.0  Grand Illusion (Grande illusion, La) (1937) M 5.0  Grapes of Wrath, The (1940) M 5.0  Grateful Dead (1995) M 5.0  Hearts and Minds (1996) M 5.0  I Am Cuba (Soy Cuba/Ya Kuba) (1964) M 5.0  Inherit the Wind (1960) M 5.0  It's a Wonderful Life (1946) M 5.0  Jean de Florette (1986) M 5.0  Jupiter's Wife (1994) M 5.0  Lamerica (1994) M 5.0  Lawrence of Arabia (1962) M 5.0  Lured (1947) M 5.0  Maltese Falcon, The (1941) M 5.0  Manon of the Spring (Manon des sources) (1986) M 5.0  Modern Times (1936) M 5.0  North by Northwest (1959) M 5.0  One Flew Over the Cuckoo's Nest (1975) M 5.0  Palm Beach Story, The (1942) M 5.0  Paralyzing Fear: The Story of Polio in America,... M 5.0  Pather Panchali (1955) M 5.0  Paths of Glory (1957) M 5.0  Patton (1970) M 5.0  Psycho (1960) M 5.0  Raiders of the Lost Ark (1981) M 5.0  Rear Window (1954) M 5.0  Return with Honor (1998) M 5.0  Sanjuro (1962) M 5.0  Saving Private Ryan (1998) M 5.0  Schindler's List (1993) M 5.0  Schlafes Bruder (Brother of Sleep) (1995) M 5.0  See the Sea (Regarde la mer) (1997) M 5.0  Seven Chances (1925) M 5.0  Seven Samurai (The Magnificent Seven) (Shichini... M 5.0  Seventh Seal, The (Sjunde inseglet, Det) (1957) M 5.0  Shawshank Redemption, The (1994) M 5.0  Silence of the Lambs, The (1991) M 5.0  Sixth Sense, The (1999) M 5.0  Small Wonders (1996) M 5.0  Smashing Time (1967) M 5.0  Star Wars: Episode IV - A New Hope (1977) M 5.0  Sunset Blvd. (a.k.a. Sunset Boulevard) (1950) M 5.0  Third Man, The (1949) M 5.0  Tigrero: A Film That Was Never Made (1994) M 5.0  Time of the Gypsies (Dom za vesanje) (1989) M 5.0  To Kill a Mockingbird (1962) M 5.0  To Live (Huozhe) (1994) M 5.0  Treasure of the Sierra Madre, The (1948) M 5.0  Two or Three Things I Know About Her (1966) M 5.0  Usual Suspects, The (1995) M 5.0  Vampyros Lesbos (Las Vampiras) (1970) M 5.0  Wallace & Gromit: The Best of Aardman Animation... M 5.0  Window to Paris (1994) M 5.0  Wizard of Oz, The (1939) M 5.0  Wrong Trousers, The (1993) M 5.0  Yojimbo (1961) M 5.0  Young Frankenstein (1974) M 5.0 |

|  |
| --- |
| Movies with a median rating over 4.5 among Women over age 30 |
| 24 7: Twenty Four Seven (1997) F 5.0  400 Blows, The (Les Quatre cents coups) (1959) F 5.0  Across the Sea of Time (1995) F 5.0  African Queen, The (1951) F 5.0  After Life (1998) F 5.0  Amadeus (1984) F 5.0  Among Giants (1998) F 5.0  Aparajito (1956) F 5.0  Apple, The (Sib) (1998) F 5.0  Ballad of Narayama, The (Narayama Bushiko) (1958) F 5.0  Before the Rain (Pred dozhdot) (1994) F 5.0  Belly (1998) F 5.0  Best Man, The (Il Testimone dello sposo) (1997) F 5.0  Bicycle Thief, The (Ladri di biciclette) (1948) F 5.0  Big Combo, The (1955) F 5.0  Blue in the Face (1995) F 5.0  Bogus (1996) F 5.0  Bridge on the River Kwai, The (1957) F 5.0  Brother, Can You Spare a Dime? (1975) F 5.0  Buena Vista Social Club (1999) F 5.0  Casablanca (1942) F 5.0  Christmas Story, A (1983) F 5.0  Cinema Paradiso (1988) F 5.0  Citizen Kane (1941) F 5.0  City Lights (1931) F 5.0  Clean Slate (Coup de Torchon) (1981) F 5.0  Cleo From 5 to 7 (Cléo de 5 à 7) (1962) F 5.0  Close Shave, A (1995) F 5.0  Cold Fever (Á köldum klaka) (1994) F 5.0  Conformist, The (Il Conformista) (1970) F 5.0  Country Life (1994) F 5.0  Creature Comforts (1990) F 5.0  Croupier (1998) F 5.0  Dancemaker (1998) F 5.0  Dark Command (1940) F 5.0  Dr. Strangelove or: How I Learned to Stop Worry... F 5.0  Dreaming of Joseph Lees (1998) F 5.0  Eighth Day, The (Le Huitième jour ) (1996) F 5.0  Enchanted April (1991) F 5.0  Everest (1998) F 5.0  Eyes Without a Face (1959) F 5.0  Faraway, So Close (In Weiter Ferne, So Nah!) (1... F 5.0  Fargo (1996) F 5.0  Firelight (1997) F 5.0  Flower of My Secret, The (La Flor de Mi Secreto... F 5.0  Force of Evil (1948) F 5.0  Funeral, The (1996) F 5.0  Gambler, The (A Játékos) (1997) F 5.0  Garden of Finzi-Contini, The (Giardino dei Finz... F 5.0  Gaslight (1944) F 5.0  General, The (1927) F 5.0  Godfather, The (1972) F 5.0  Gone with the Wind (1939) F 5.0  Grand Day Out, A (1992) F 5.0  Grand Illusion (Grande illusion, La) (1937) F 5.0  Grandfather, The (El Abuelo) (1998) F 5.0  Green Mile, The (1999) F 5.0  Haunted World of Edward D. Wood Jr., The (1995) F 5.0  Hear My Song (1991) F 5.0  Hearts and Minds (1996) F 5.0  Idiots, The (Idioterne) (1998) F 5.0  Innocents, The (1961) F 5.0  It Happened One Night (1934) F 5.0  It Takes Two (1995) F 5.0  It's a Wonderful Life (1946) F 5.0  Jupiter's Wife (1994) F 5.0  Kagemusha (1980) F 5.0  Killing Fields, The (1984) F 5.0  King of Masks, The (Bian Lian) (1996) F 5.0  Lamerica (1994) F 5.0  Lawn Dogs (1997) F 5.0  Lawrence of Arabia (1962) F 5.0  Legend of 1900, The (Leggenda del pianista sull... F 5.0  Local Hero (1983) F 5.0  Love Serenade (1996) F 5.0  Lucie Aubrac (1997) F 5.0  Maltese Falcon, The (1941) F 5.0  Man Facing Southeast (Hombre Mirando al Sudeste... F 5.0  Manon of the Spring (Manon des sources) (1986) F 5.0  Marcello Mastroianni: I Remember Yes, I Remembe... F 5.0  Microcosmos (Microcosmos: Le peuple de l'herbe)... F 5.0  My Fair Lady (1964) F 5.0  Nights of Cabiria (Le Notti di Cabiria) (1957) F 5.0  North by Northwest (1959) F 5.0  Notorious (1946) F 5.0  Official Story, The (La Historia Oficial) (1985) F 5.0  Only Angels Have Wings (1939) F 5.0  Other Side of Sunday, The (Søndagsengler) (1996) F 5.0  Palookaville (1996) F 5.0  Paradine Case, The (1947) F 5.0  Paradise Lost: The Child Murders at Robin Hood ... F 5.0  Paris, France (1993) F 5.0  Paths of Glory (1957) F 5.0  Perils of Pauline, The (1947) F 5.0  Philadelphia Story, The (1940) F 5.0  Prisoner of the Mountains (Kavkazsky Plennik) (... F 5.0  Producers, The (1968) F 5.0  Promise, The (La Promesse) (1996) F 5.0  Psycho (1960) F 5.0  Quiet Room, The (1996) F 5.0  Raiders of the Lost Ark (1981) F 5.0  Rain (1932) F 5.0  Ran (1985) F 5.0  Raw Deal (1948) F 5.0  Rear Window (1954) F 5.0  Rebecca (1940) F 5.0  Red Sorghum (Hong Gao Liang) (1987) F 5.0  Saltmen of Tibet, The (1997) F 5.0  Sanjuro (1962) F 5.0  Schindler's List (1993) F 5.0  Seven Samurai (The Magnificent Seven) (Shichini... F 5.0  Shadow of a Doubt (1943) F 5.0  Shawshank Redemption, The (1994) F 5.0  Singin' in the Rain (1952) F 5.0  Sixth Sense, The (1999) F 5.0  Skipped Parts (2000) F 5.0  Sling Blade (1996) F 5.0  Some Folks Call It a Sling Blade (1993) F 5.0  Some Like It Hot (1959) F 5.0  Song of Freedom (1936) F 5.0  Sound of Music, The (1965) F 5.0  Source, The (1999) F 5.0  Star Wars: Episode IV - A New Hope (1977) F 5.0  Still Breathing (1997) F 5.0  Substance of Fire, The (1996) F 5.0  Sum of Us, The (1994) F 5.0  Sunset Blvd. (a.k.a. Sunset Boulevard) (1950) F 5.0  Sunset Strip (2000) F 5.0  Taste of Cherry (1997) F 5.0  Terrorist, The (Malli) (1998) F 5.0  Third Man, The (1949) F 5.0  Three Days of the Condor (1975) F 5.0  To Kill a Mockingbird (1962) F 5.0  To Live (Huozhe) (1994) F 5.0  Trial, The (Le Procès) (1963) F 5.0  Trouble in Paradise (1932) F 5.0  Two Women (La Ciociara) (1961) F 5.0  Two or Three Things I Know About Her (1966) F 5.0  Usual Suspects, The (1995) F 5.0  Vertigo (1958) F 5.0  Wallace & Gromit: The Best of Aardman Animation... F 5.0  What Happened Was... (1994) F 5.0  Window to Paris (1994) F 5.0  Wizard of Oz, The (1939) F 5.0  Woman of Paris, A (1923) F 5.0  World of Apu, The (Apur Sansar) (1959) F 5.0  Wrong Trousers, The (1993) F 5.0  Yojimbo (1961) F 5.0  Young Frankenstein (1974) F 5.0 |

Citations:

* 1. MovieLens, <http://files.grouplens.org/datasets/movielens/ml-1m-README.txt>
  2. Wikipedia - genres, <https://en.wikipedia.org/wiki/List_of_genres#Film_and_television_formats_and_genres>
  3. MovieLens, <http://files.grouplens.org/datasets/movielens/ml-1m-README.txt>
  4. MovieLens, <http://files.grouplens.org/datasets/movielens/ml-1m-README.txt>
  5. The HDF Group, <https://support.hdfgroup.org/HDF5/>