DATA EXPLORATION

1)

users in the social network = 10

movies = 64

directors (non-unique) = 100

actors (non-unique) = 1850

actresses (non-unique) = 895

The <u>movies that no-one liked</u> are 'The Outsiders' and 'If Lucy Fell'

The <u>most liked</u> movie is **'Mission Impossible'.**

Other well liked movies (liked by at least 4 users) are:

- 'Spies Like Us'

- 'Star Wars Episode 1: The Phantom Menace'

- 'The Usual Suspects'

- 'The Virgin Suicides'

- 'American Beauty'

- 'Eight Legged Freaks'

- 'Fargo'

- 'The Firm'

- 'Ghost World'

- 'The Godfather'

- 'The Ladykillers'

The least liked films (not counting the ones not liked at all) are:

- an_american_rhapsody

- anna

- the_black_dahlia

- cq

- fall

- frankenweenie

- a_good_woman

- in_good_company

- the_island

- just_cause

- lick_the_star

- manny__lo

- marie_antoinette

- match_point

- my_brother_the_pig

- north

- paris_je_t_aime

- the_perfect_score

- torrance_rises

- untitled_woody_allen_fall_project_2006

The actors that appear in some of the most well liked movies but not much frequency are

- Tom Cruise: The Firm & Mission Impossible

- Frank Oz: Star Wars 1 & spies like us

- Kevin Spacey: American Beauty and & The Usual Suspects

- Thora Birch: Ghost World & American Beauty

Actors/Actresses/Directors with more than 5 appearances in movies:

- Sofia Coppola: popular works are Virgin Suicides & Star Wars 1 & the Godfather

- Scarlett Johansson: popular works are Eight Legged Freaks & Ghost World

- Steve Buscemi: popular works are Fargo & Ghost World

- John Goodman

- Roman Coppola

- Francis Ford Coppola

- Ethan Coen

- Joel Coen

- Frances McDormand

- Holly Hunter

- Bruce Campbell

- Jon Polito

IMPLEMENTATION JUSTIFICATION

2)

        I gathered the information on the ratings of all of the movies that were in the movies.pl and added them to a separate file named 'ratings.pl'. The ratings for each of the movies were gathered from the Internet Movie Database (IMDb). The website provides them as values out of 10 so I multiplied the ratings by 10 to have the ratings of the movies be out of 100.

        I used this because I believed that to help the user be more social on the social network they should be recommended to watch movies that their friends have liked and that they have yet to see/like. However, the user's time should not be wasted if they aren't particularly fond of watching movies in their free time. The ratings are in place so that the user is only recommended movies that meet the previous criteria but that are also generally well received. A lot of more casual users would not really have much of an idea of who certain actors and directors are so they may not even care. Even though they may have liked a movie with a certain actress they may not like any other movie where that actress shows up. Ratings could be a more universal tool for decision making for all types of users as most people will not be interested in a movie that the general public regards as uninteresting.

        The goal would be that even if the movie was not to the user's taste, having watched a well-received movie would give them more to talk about and ultimately it would allow them to be more social by giving them a popular thing to talk about and for them to have formed some opinion on. The user's nephew may have liked 'Frankenweenie' and maybe that movie wouldn't have stuck out to the user initially, but since it was so well received by the whole public it's likely worth checking out. The recommendation algorithm ultimately encourages deepening connections with current friends by having more movies to talk about but also encourages the user to potentially build more connections through movie knowledge.

        As it stands there is a cutoff for the quality expectations of the movies that the user will be recommended. Any movies that receive below a 70 will never be recommended to them, and if the user has no friends, no movies below an 80 will ever be recommended. In another possible implementation the user may have also been able to adjust this parameter of their tolerance to the quality of movies if they chose to define it. This would potentially broaden or narrow the amount of movies recommended to them.