Project Proposal:

1. The kind of data you'd like to work with/field you're interested in (e.g., geodata, weather data, etc.)

We will be working with movie data, specifically what factors cause a movie to have a higher user rating.

2. The kinds of questions you'll be asking of that data:

We will ask things like does a higher budget make a better movie?

Which production company makes the highest average rated movie?

Which genres have the highest rated movies?

And does the positivity of a movie affect its rating. This will be determined by running a sentiment analysis on each movies description.

3. Possible source for such data:

The data will be gather from 2 sources, the first source will be a csv from Kaggle having movie data up to 2016. The other set will be from api calls to tmdb to get movie data for 2017. We will also use Wikipedia to get a list of the movies released in 2017 to use in our calls.

http://wikitable2csv.ggor.de/