**Project Proposal**

**Dataset selected**: MovieLens 25M Dataset

**Analysis to be performed:**

1. **Number of movies released per year (Descending order)**
2. **Number of movies per genre**
3. **Number of times a movie was tagged (Descending order)**
4. **Number of times a user tagged some movie**
5. **Number of movies per rating**
6. **Perform join on movies.csv and links.csv and extract links for each movie**
7. **Perform join on ratings.csv and tags.csv and extract tags and ratings for each movie**
8. **Perform join on ratings.csv and tags.csv and extract ratings and tags by each user**
9. **Find the respective genome tag for each movie(join genome-scores.csv and genome.tags.csv)**

**1.Number of movies released per year (Descending order)**

**2.Number of movies per genre**

**3.Number of times a movie was tagged(Join movies, tags) (Descending order)**

**4.Number of times a user tagged some movie**

**5. Probable Spam/non-spam users**

**6. Number of movies per rating**

**7. Perform join on movies.csv and links.csv and extract links for each movie (Pig)**

**8. Perform join on ratings.csv and tags.csv and extract tags and ratings for each movie (Pig)**

**9. Find the movies released in 2011 in “Action” Genre**

**10. Movies based on top relevance. (Pig)**

**11. Top 5 movies of all time based on average rating per Genre (Pig)**