Capstone Project 2 - Proposal

Objective and Goal:

Recommender systems have been around since at least 1992. Today we see different flavours of recommenders, deployed across different verticals:

- Amazon
- Netflix
- Facebook
- Last.fm.

What recommender systems do is they predict whether or not a user will like an item (s)he has not seen. The system does this either using the user's history of rating(content based filtering) or the user's similarity to other users (collaborative filtering). The goal of this project is to build a recommender system.

Dataset:

Last.fm provides a dataset for music recommendations. For each user in the dataset it contains a list of their top most listened to artists including the number of times those artists were played. There are a total of **92,834** user-artist ratings in the dataset. It also includes user applied tags which could be used to build a content vector.

Last.fm's data is aggregated, so some of the information (about specific songs, or the time at which someone is listening to music) is lost. However, it is the only dataset in our sample that has information about the social network of the people in it.

http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-readme.txt

Methodology:

The proposed approach is to build the recommendation system using a collaborative filtering algorithm by using pearson correlation coefficient to define similarity between users. User rating her is defined as the number of times a user listened to an artist (i.e. the 'weight' variable). RMSE can be used to evaluate system's performance.

Final Deliverable:

The final deliverable will be the code, a slide deck and a 10-12 page report.