# Personalization Project

Objective: Build content based recommendation systems using MovieLens 1M dataset. One of the main takeaways from our findings in Part 1 of our project was that we did not incorporate the inherent user generated genre tags of movies. Therefore, if a user doesn’t like horror movies, this information can be fed into a model to allow us to better predict a user’s top movie preferences. In Part 2 we aim to determine the correlation between these inherent movie genres to further increase the accuracy of our recommendations. It would also be interesting to look at how movie popularity might affect a user’s predicted movie rating and how it might correlate with genre.

## Deliverables

The main deliverable will be a GitHub repository with the following:

* A README file outlining the repository’s contents
* A requirements file with all software/package requirements to run our code
* A top level directory for Part II that has a Python notebook containing our approach and basic results

Project Outline

1. Develop a small dataset sample (10000 users / 100 items)
2. Build content based algorithm
3. Develop evaluation methods of the content based model
   1. Cross Validation Setup
   2. Accuracy on training and test data
   3. Coverage on training and test data
   4. Systematically test a range of hyperparameters in the models and plot the results
      1. Model size/Sample size: how does overall accuracy change? Run time?
4. Drawbacks and caveats of the developed content based model