Project Management Plan

# Division of Work

The project involves three principal components. We should plan these components to be relatively independent of each other, so different people can work on different pieces in parallel.

1. Recommendation Engine
2. Simulation of recommendation feedback
3. Analysis + Visualization of simulated data

The goal here is to divide up code into the logical blocks, but allow us to work as independently as possible. To divide these successfully, we should:

1. Divide into sub-teams. 3 people focusing on the engine, 1 on simulation, and 1-2 on analysis. The final presentation creation should be a joint effort.
2. Define outputs ahead of time: Analysis + Simulation teams need to lay out what the simulation team’s deliverable will be (what data, in what format, will the sim produce?). This is so the analysis team will know how to code their analysis before receiving data, and how the sim team needs to code the simulation to produce that data. Likewise, the Engine and Simulation team need to decide what the engine will look like when passed to the simulation.
3. Early deliverable of dummy outputs: To get team 2 going, team 1 should create a simple, incomplete, imperfect, but functioning recommendation engine. They can then take the rest of the semester to finish the engine while the simulation team has something to work with. Then near the end of the project, the engine team can pass the final engine to time sim team, and the sim team should be set to simply re-run with the updated engine. Repeat that process for the sim team.

# Deliverables

1. Engine: A .py file containing a function that runs the recommendation engine. It should accept a pandas data frame (such as the base data provided for this lab), and output predicted scores for all movies/people based on that recommendation engine. (As an alternative, it might simply output the X# of movies with high predicted scores, for every person.) While the engine team needs to code, generate, iterate, refine, etc to make the model, the delivered function should merely be that final model as a function. That is, the simulation team can’t re-generate a new ML model in every iteration of the simulation.
2. Simulation: A data file containing, for every person, the top X movies recommended to them, for every generation of the simulation. Likely a CSV, or other format as requested by analysis team.

Example data might look like:

| Name | Generation | Rank | Movie |
| --- | --- | --- | --- |
| Mike | 1 | 1 | Batman |
| Mike | 1 | 2 | Superman |
| Mike | 1 | 3 | Spiderman |
| Mike | 2 | 1 | Superman |
| Mike | 2 | 2 | Iron Man |
| Mike | 2 | 3 | Batman 2 |

This output *must* be adjusted to the analysis team’s needs. Analysis team here has authority to request the types of output they want for their work.

1. Analysis: Results of statistical tests and data visualizations, in whatever format is most easy for presentation. This deliverable will be incorporated into the presentation, likely a PPT or similar. Pulling together the final presentation will probably be a joint effort with some revisions from the analysis deliverables, so team should be prepared to make minor modifications.
2. Each team must additionally create a set of slides explaining their portion of the project, to be compiled and used as the group presentation.

# Proposed Teams

-Sathvik: Engine. Has already produced a doc laying out some design. Can provide support with developing the system engine and also in simulation.

-Alex: Simulation. Should be simple enough for 1 person and it’s the part I wanted to execute on. Analysis team support. I’ve used python professionally so I can help work through any coding issues too.

-Ross: Engine. I can provide support with ML (or collaborative filtering) models. I have also looked at some other examples of recommendation engines written in python, so we can use those for support if needed.

-Ankit: Will help with developing the Recommendation System engine. Will help with simulation and documentation.

-Edwin: I can provide assistance on the analysis. But I think we would have to define our objectives more clearly so that we know how to drive the analysis of the project.

Fill in with the part you’re most interested in working on. The recommendation engine will need the most support, but we need someone confident in stats to take the lead on the data analysis portion.

# Proposed Due Dates

Modeling Team

* Dummy Model: 11/9
* Final Model: 11/21
* Slides: 11/28

Simulation Team

* Dummy Simulation Data: 11/16
* Final Simulation Data: 11/23
* Slides: 11/28

Analysis Team

* Proposed stat. tests/viz: 11/23
* Final Analysis Data/viz as slides: 11/28

All teams

* Review other team’s slides, with specific attention on the final analysis results: 11/30

# Team Resources

* Google Drive for sharing all files
* I’ve got an office in Thompson 824 that we can use for any in-person collab. work.