

*The goal for your final project is for you to build out a recommender system using a large dataset (ex: 1M+ ratings or 10k+ users, 10k+ items).*

There are three deliverables, with separate dates:

- [1] Planning Document** Find an interesting dataset and describe the system you plan to build out. If you would like to use one of the datasets you have already worked with, you should add a unique element or incorporate additional data. (i.e. explicit features you scrape from another source, like image analysis on movie posters). The overall goal, however, will be to produce quality recommendations by extracting insights from a large dataset. You may do so using Spark, or another distributed computing method, OR by effectively applying one of the more advanced mathematical techniques we have covered. There is no preference for one over the other, as long as your recommender works! The planning document should be written up and published as a notebook on GitHub or in RPubS. **Please submit the link in the Assignment 5 folder, due Sunday, July 9.**
- [2] Implementation.** In this final project deliverable, you'll build out the system that you describe in your planning document. This will be due on **Sunday, July 16** and must be turned in as an RMarkdown file or a Jupyter notebook, and posted to GitHub or RPubS.com.
- [3] Presentation.** Make a five-minute presentation of your system in our final meetup. If you're not able to attend the meetup, you're responsible for either recording your presentation, or scheduling one-on-one time to deliver your presentation prior to the meetup. You should be prepared to present on **Thursday, July 20.**

To help you structure your work, you may find it useful to review Aurélien Géron's project checklist, which can be found in your Week 1 folder, under "Guidance on Effective Projects and Presentations."

You should use this project to showcase some of the concepts that you have learned in this course, while delivering on the (probably) less familiar Spark platform.

You are welcome to submit a compelling alternative proposal (subject to approval), such as implementing a recommender system using in Microsoft Azure ML Studio or with Google TensorFlow, or building out an application of a certain complexity using another tool.

*You are strongly encouraged but not required to work in a small group on this assignment.*