

# MSDA 607 Final Project

## Deliverables Schedule

Deliverable	Date	Points
One Paragraph Proposal	Tuesday May 12 <sup>th</sup>	25
Final Project	Sunday May 24 <sup>th</sup>	125
Final Project Presentation	Before or during Final Office Hours on May 26 <sup>th</sup>	50

## Policy on Collaboration

You may work in a team of up to three people. Each project team member is responsible for understanding and being able to explain *all* of the submitted project code. Remember that you can take work that you find elsewhere as a base to build on, but you need to acknowledge the source, so that I base your grade on what you contributed, not on what you started with!

## Approval Meeting

Once you've turned in your one paragraph proposal, I want to schedule a 15 minute phone meeting with each team, where you'll describe the reason (benefit) for doing this work and/or question you're seeking to answer, where you'll source the data, and the overall flow. For team projects, I also want you to articulate the roles and responsibilities of each team member.

## Final Project Checklist

To receive full credit, you'll need to deliver on all of the items in the checklist below. Please read carefully through this checklist before you make your project proposal. You are (within these checklist constraints) strongly urged to limit scope and make the necessary simplifying assumptions so that you can deliver your work on time!

- ☐ Proposal describes your motivation for performing this analysis.
- ☐ Your project has a recognizable "data science workflow," such as the OSEMN workflow or Hadley Wickham's Grammar of Data Science. [Example: First the data is acquired, then necessary transformations and clean-up are performed, then the analysis and presentation work is performed]
- ☐ Project includes data from at least two different types of data sources (e.g., two or more of these: relational or CSV, Neo4J, web page, MongoDB, etc.)
- ☐ Project includes at least one data transformation operation. [Examples: transforming from wide to long; converting columns to date format]
- ☐ Project includes at least one statistical analysis and at least one graphics that describes or validates your data.
- ☐ Project includes at least one graphic that supports your conclusion(s).
- ☐ Project includes at least one statistical analysis that supports your conclusion(s).
- ☐ Project includes at least one feature that we did not cover in class! There are many examples: "I used ggmap; I created a decision tree; I ranked the results; I created my presentation slides directly from R; I figured out to use OAuth 2.0..."
- ☐ Presentation. Was the presentation delivered in the allotted time (3 to 5 minutes)?
- ☐ Presentation. Did you show (at least) one challenge you encountered in code and/or data, and what you did when you encountered that challenge? If you didn't encounter any challenges, your assignment was clearly too easy for you!
- ☐ Presentation. Did the audience come away with a clear understanding of your motivation for undertaking the project?
- ☐ Presentation. Did the audience come away with a clear understanding of at least one insight you gained or conclusion you reached or hypothesis you "confirmed" (rejected or failed to reject...)?
- ☐ Code and data. Have you delivered the submitted code and data where it is self-contained—preferably in rpubs.com and github? Am I able to fully reproduce your results with what you've delivered? You won't receive full credit if your code references data on your local machine!
- ☐ Code and data. Does all of the delivered code run without errors?
- ☐ Code and data. Have you delivered your code and conclusions using a "reproducible research" tool such as RMarkdown?
- ☐ Deadline management. Were your draft project proposal, project, and presentation delivered on time? *Any part of the project that is turned in late will receive a maximum grade of 80%.* Please turn in your work on time! You are of course welcome to deliver ahead of schedule!