

Final project (25%)

Group Project

For the final project you will work in a team of 3 or 4 people. You need to form your team by the start of the second synchronous session. A discussion board will be created to help people form teams.

Selecting a Dataset

The team will have to pick a dataset to be analyzed. The dataset needs to have a minimum of 10,000 “values” (ex. a dataset with 10 attributes and 1,000 observations). The dataset does not have to be publically available. For example, it could be a dataset from a student’s company (assuming any confidential information is masked).

Questions to be addressed in the report

There are no required questions to be addressed, or techniques to be used. However, the team is expected to use advanced data science concepts, and not just traditional descriptive statistics.

Project Updates:

Each team will use a Agile Kanban Data Science project methodology (see AK document). The project updates should include your:

- Kanban board (which is explained in the AK document)
- Project summary. The project summary is one page and contains four quadrants. The quadrants focus on:
 - What was accomplished since the last update (or since the project started) – *these should be highlighted on the Kanban board*
 - What is working well for the team
 - Plans for the next update
 - Issues / what is not working well.

Project Timeline

There are four project updates noted in the syllabus. Each of these updates count 1% of your grade. The focus of these updates is to provide feedback and help move your project forward. These updates will be five minutes long, and will take place during our synchronous session. These updates will be just with the team members and the instructor, with the other students working on the lab during this time.

Project Deliverables

The final report will be worth 21% of your grade.

Each team will prepare a presentation, which will report on the findings of the analysis. The PowerPoint presentation will be such that managers without data science expertise could understand and use the insight generated - this will be 7% of your grade.

In addition to the presentation, a word document should be submitted that describes the work done on the project (ex. type of analysis, the R code). One can think of this as being a report to the data science group team leader. This detailed document will count 14% of your grade.