Learning Outcomes:

1. Regarding new analytic techniques: to be able to study them, understand them and use them appropriately.
2. To practice the following basic ideas of data analysis:
   1. Formulate questions and answer these questions from the data.
   2. Model, study results, model, study results, …. (iterate)
   3. To be able to interpret/translate the analytical findings to the context of a problem.
3. To be able to use Python to perform data analysis.

Grading Guides for Different Project Categories:

1. *Data Analysis Project Proposal:*
   1. Report quality
   2. Presentation quality
   3. Literature review of techniques for similar problems
   4. Proposed pipeline
      * Data collection tools, frequency, storage format
      * Analytic tools, models
      * Output (dashboard, report, etc.)
   5. Preliminary analysis of pilot project, or some proof of impact of proposed project.
   6. Example: You might have data on machine breakdowns (different types). You may want to model these as a simulation model. Think about the ingredients you would need:
      * Distribution of breakdown, perhaps broken down by machine type, age, etc.
      * Number of repairmen you have available.
      * Time of repair
        1. Labour,
        2. Time to purchase new parts
      * Cost of repair
      * Cost of downtime

With these, you might be interested in say,

* + - Optimal number of repairmen to keep
    - Total number of uptime hours per month
    - How much do I need to reduce repairtime by in order to increase my uptime hours?

1. *Methodology Review:*
   1. Report quality
   2. Presentation quality
   3. Theoretical value:
      * Depth
      * Correctness
      * Modernity, I.e. how modern are the techniques covered.
   4. Application to small examples
      * Code quality
      * Presentation
   5. Example: You may wish to expound on any one of the following topics (or others you may be interested in):
      * Association Rule Mining
      * Recommender systems
      * Random Forests
      * Regression models
      * Exponential Smoothing Time Series models
2. *Data Analysis:*
   1. Report quality
   2. Presentation quality
   3. Problem statement
   4. Quality of analysis
      * Exploration
      * Novelty of models
      * Code quality (Python)
   5. Interpretation and impact of results