# SOEN6611 Project – summer term 2022 (30%, team work – 2 to 4 students per team)

**Summary:** The aim of the soen6611 project is modeling Big Data Quality and applying measurements of the Big Data V's to real world data. Students are asked to group themselves in teams of 2 to 4 students. Each team will be given a large dataset (an excel file with the sources will be posted by the tutor of this course).

The first 4 steps in the project will be to identify their goals for each of the Big Data V's, the hierarchical measurement models and establish thresholds for each indicator.

Teams will be told that their data will need to be properly prepared for use in a machine learning algorithm. Due to time constraints, they will not need to apply a machine learning algorithm.

Instead, students will need to apply the measures of each the selected V's on their respective datasets and analyse the results. Then they must make take steps at cleaning the data to have it ready for a machine learning pipeline. After each change, students will need to re-apply their measurements and record the new changes.

After all data cleansing is done, soen6611 teams must decide if their data would be usable for a machine learning algorithm. Does it meet the minimum thresholds set by them at the beginning of the project?

Teams aren't graded on whether the dataset at the end of the project can be used. Instead, they will be evaluated on their ability to apply measurements on a real-world application and obtain a proper analysis of these measurements at each phase in their project.

**Project Goals**:

**G1. Software Measurement Plan (SMP)**. SMP identifies measurement objectives, and presents the plan to assess the objectives. The steps 1-4 associated with this case study are structured methods for identifying software measures that directly support, and are traceable to, your business goals.

**G2. Software Measurement Feedback Loop** The teams will practice in Step 5 the measurement feedback loop where the teams will collect, analyze and report decisions according to their SMP.

Outline of Project's Steps:

| Steps | Description | Dates |
|---|---|---|
| Step 1 | Identify SMART (Specific, Measurable, Achievable, Realistic, and Timely) measurement goals for given V's of Big Data (1.5 points). | Posted on July 1<br>Due on July 11th<br>Submit before midnight on July 11th, one submission per team |
| Step 2 | Operationalize the Measurement Goals (1.5 points) | |
| Step 3 | Derive Success Criteria and Indicators. Document the derived measures and base measures (10 points) | Posted on July 7th,<br>Submit before 6pm on July 21st, one submission per team |
| Step 4 | Software Measurement Plan (5 points) | Posted on July 14th<br>Submit before 6pm on July 28th, one submission per team |
| Step 5 | Measurement Feedback Loop (7 points) | Posted on July 21st<br>Due on August 4th by 6pm, one submission per team |
| Post Mortem Analysis Presentations | In class (5 points) | |