## Final project: Guidelines for the initial write-up

### Due date: Sunday, Apr 2

The initial write-up for the final team project should give a brief description of the type of data mining task you are interested in working on and the particular dataset you will be working on. It can be organized by the following sections:

1. Description of the data and relevant background information (please attach portion of the data; for example, the first 20 records)
2. Identification of the particular question(s) of interest
3. Specification of the data mining task that you have in mind that will help answer these questions (for example: classification, prediction, clustering, association, etc.)
4. Identification of the variables included in the data – how many and what type (numerical, nominal/categorical, etc.)
5. Any initial pre-processing/data exploration methods that you may think could be appropriate for the data at hand – these may include:
   1. Detecting missing values and outliers
   2. Specification of the “relevant” variables (do you need all of them?)
   3. Need for sampling – random or “selective”?
   4. Use of visualization methods and summary statistics
   5. Need for data reduction, aggregation, zooming, filtering, variable transformation
   6. Detecting any “obvious” patterns/clusters or structures

For Step 5, you are not expected to state a lot of details – just include whatever initial thoughts or insights you may have gathered and which technique(s) you may want to apply.