

CS7DS3 Main Assignment

This assignment will consist of a report, with a maximum length of 20 pages. The assignment deadline is 5p.m. on Friday April 6th, 2018. I would like you to analyse the Yelp academic dataset: <https://www.yelp.com/dataset>.

Specifically, you should focus your analysis on restaurant businesses that are currently open in the city of Toronto. In the report I would like you to address the following questions:

1. Compare the ratings of different neighbourhoods. Are any neighbourhoods clearly superior to others? If so, by how much?
2. What variables are most influential at predicting restaurant rating? How accurate are these predictions?
3. Is there any association between neighborhoods and restaurant categories? Can you identify neighbourhoods that are more likely to contain certain types of restaurant category than others?

The assignment is worth 70% of the module mark, and will be evaluated under the following criteria:

- Data management (10%)
- Clarity of writing and exposition (10%)
- Modelling (60%)
- Creativity (20%)

You may use any software or programming language you wish, as long as you can clearly describe what model the software is implementing, and how. Similarly, the data is available in two formats, either of which is acceptable to analyse.

If you can identify an alternative dataset that you would prefer to analyse, then contact me and we can discuss if it will be suitable for the project.