**Introduction/Business Problem**

My Capstone project is focused on helping Data Science students like myself understand how the Toronto and New York assignments in this class need to be redesigned, because they teach incredibly faulty applied machine learning. I hope the course is improved by spending a lot more time on:  
- the curse of dimensionality (too many columns, often correlated, and/or irrelevant, and/or relevant but missing, and/or mostly null, etc).  
- various optimization metrics (silhouette, distortion, Calinski-Harabasz, Davies Bouldin, etc) that could be relevant for different stakeholders/uses to determine the optimal (Elbow) number of clusters. The problem being that they do not provide the same optimal number of clusters!  
- lack of business analysis. (ie low restaurant density doesn’t imply opportunity, nor does high density imply lack of opportunity).

The list of lessons could go on. I hope the course is improved at least enough to not massively miseducate and create more critical and robust data scientists who understand how difficult and subjective it actually is to be an applied data science, especially for business applications.

**Data**  
I will use Manhattan, NY Foursquare API data and neighborhood locations precisely the way the instructors.