**Business understanding and requirements**

My client is young and perpsective enterpreneur from Korea who wants to start a Korean restaurant business in Toronto City. He hired me to find out the best neighborhood location for opening his first korean restaurant. Criterieas of the choice are:

* Neighborhood with highest average annual household income
* Neighborhood with least amount of competitive restauraunt (both serving Korean and other cuisines)
* Neighborhood that is under Business Imporevement Areas (BIA) program

Question Problem is “What is the best neighborhood to open Korean Restaurant Business?”

**Data Collection**

Data that is required for solving the question problem will be gathered and collected using different sources and different methods for data gathering. Data about average annual household income will be gathered by webscrapping method of the website which contains tabular data of 2016 average annual household income grouped by neighborhood in Toronto. I couldn’t find more relevant and fresh data of average household income that is why let’s assume that this data is of 2020. Data about venues in particular household will be gathered using Foursquare API. Data about neighborhoods that are under Business Improvement Areas program will be gathered from Toronto Open Data Source portal via CKAN Toronto OpenData API.

**Data preparation**

Data will be stored in different forms. Tabular data of neighborhoods under Business Imporevement Areas (BIA) program will be accessed via OpenData Toronto API and transformed or assigned to pandas dataframe. Data about average annual household income by neighborhood will be webscrapped from website using BeautifulSoup library and converted into dataframe using pandas. Data about venues in particular neighborhood will be accessed using Foursquare API and stored into pandas dataframe.

**Modelling**

Most suitable model for this kind of question will be clustering model. We can cluster the neighborhoods based on several features as “great”, “good” or “bad” neighborhood/location for opening Korean Restaurant in. Those features will be average annual household income, number of other restaurants (especially Korean restaurants). Different clustering models will be used, such as: K Means, Hierarchical and DBSCN.

**Conclusion**

After completing the project work I will make conclusion regarding the best neighborhood where to open Korean Restaurant