**Predicting the Best Location for Chinese Restaurant in North York Toronto**

1. **Introduction**
   1. **Background**

Toronto is one of the most popular cities in the world for immigrates. The diversity of culture brings a diverse food culture as well. North York is the outer suburbs of Toronto, growing into the secondary business districts outside Downtown Toronto. High-rise development in the area has given the former municipalities distinguishable skylines of its own, with high-density transit corridors serving.

* 1. **Problem**

Chinese restaurants are always popular across the world. However, they might not be the mainstream restaurant type in the city, especially in North York. This project aims to find the most popular location in North York to establish a Chinese restaurant.

* 1. **Interest**

This project may potentially benefit entities who wants to start a new Chinese restaurant in North York Toronto. As well as investor who are looking for the area in North York where Chinese restaurants are the most popular choice.

1. **Data acquisition and cleaning**
   1. **Data sources**

The Toronto post code, borough and neighbourhood data are scrapped from [here](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), and the venue details are provided by Foursquare.

* 1. **Data cleaning**

Geospatial co-ordinate is one of the key elements for API requests from Foursquare. Therefore, after scraping the data frame of the neighbourhoods in Toronto, geospatial co-ordinates are imported from csv file and joined with existing data frame, sharing the primary key as neighbourhood.

When the geospatial co-ordinate of North York is acquired, API requests were made for the detailed venues of North York.

Missing data were dropped, venues are returned with name, location and categories, categorised by neighbourhoods in North York.

* 1. **Feature selection**

Only the popularity feature of venues is selected during the feature engineering process, given the purpose of the project is to find the popularity of different venues. 99 unique venue categories were filtered out and one hot encoded for model fitting. The target is the name of neighbourhood, where Chinese restaurant is the first most common venue.