# Data Science Project Report

**Finding the best location for new business**

By Sam Zhang

2019-07-31

## Introduction

Toronto is an interesting city with diversity. I’ve been to it twice and enjoyed both times. The Chinese food is fascinating. As I almost chose to live there, I wonder if I were given the chance to start a Chinese restaurant, how do I choose the best location?

This actually is a very practical question, and potentially useful to anyone who are looking to invest.

## Data

My solution will be based on the following data (for now):

1. Defined location areas in Toronto. Will use neighborhood list scraped off a web Site (https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M).
2. Geo location data of the neighborhoods. For this I will use geolocator to find center of each neighborhood
3. Chinese restaurant data, with location data. For this I will use Foursquare Venue service to list Chinese restaurant in that area, then use Foursquare Venue detail service to get details of it. For version one will just get the createdAt property and use that as the business start date.

Combine the above data, we get a list of Chinese restaurants, each with its neighborhood, location, and date business started.

## Methodology

Overall I will take the experimental methodology. Start with something simple, and try to improve it with better models; version and document each model; compare and evaluate.

With version one, my idea is to find the hottest area for Chinese restaurants, and recommend it. For version one I will use density as the measure of hotness. In plain words, simply count the number of Chinese restaurants in each neighborhood.

The advantage of this approach is its simplicity but is likely practically effective.

The disadvantage is that it may be too simple and not being logically 100% sound. Imagine, if everyone choose next location based on this, then all new future business will be opened in the same area. We all know that cannot be right.

But it is a starting point. Let’s see how it goes.

## Results section

where you discuss the results.

## Discussion section

where you discuss any observations you noted and any recommendations you can make based on the results.

## Conclusion section

where you conclude the report.