

# **Coursera Applied Data Science Capstone Project**

**The Battle of Neighborhood: Where to start the  
business in Guangzhou?**

Author: Ziqing Xu

Date: Oct 28, 2020

# 1. Introduction

In our lives, people are diverse. Some people prefer a cozy lifestyle and choose to work for someone else. However, some people are passionate about challenges, and they start their own businesses to follow their dreams and fulfill their passion. Types of entrepreneurship are diverse, such as opening restaurants, coffee shops, IT companies and so on. Then, a significant problem arises, namely, in which specific location to recommend entrepreneurs to start their businesses. For example, if you want to open a restaurant in Toronto, it is recommended that you open a restaurant on Yonge St, as restaurants are clustered in that area.

The goal of the project is to segment and cluster the neighbourhoods by exploring and comparing the neighbourhoods. By analyzing the clusters, we can figure out the best-recommended location to start a specific type of business. As my hometown is Guangzhou, in this project, we will focus on the neighbourhoods in Guangzhou. Definitely, we can also apply this application to other cities.

## 2. Data

1. Neighbourhood data of Guangzhou extracted from:  
[https://en.wikipedia.org/wiki/List\\_of\\_township-level\\_divisions\\_of\\_Guangdong](https://en.wikipedia.org/wiki/List_of_township-level_divisions_of_Guangdong)
  - a. districts information
  - b. subdistricts information
2. Geospatial Coordinates of Guangzhou Neighbourhoods will be obtained by using the geocode package. Geospatial Coordinates data is important for getting venues data.
3. Venues data of each Neighbourhood will be retrieved by using Foursquare API

