Coursera Capstone Project: Applied Data Science
The Battle of Neighborhoods

Haoyuan (Miley) Gao

Introduction

Vehicle becomes a daily essential for most people in Canada. According to the data provided by the Government of Canada, there are more than 35 million vehicle registrations in 2019, which means every 1.05 people owns a vehicle. These 35 million cars need a place to park when they are not moving. Parking lots that are mainly locating on streets, next to business areas have limited number of spaces, but the traffic is huge. Especially, for those business on streets that do not have their own parking space, the only option is street parking.

The center of several business areas is the ideal location for a parking lot, because people could easily reach the places they want within a certain distance. To succeed with a parking lot business, one must choose the accessible locations that is convenient for customers to park their vehicles near their destinations.

Business Problem

Citizens and employees need to commute to their places on a regular basis, as well as random city travelers. Although some places provide on-site parking or street parking, it is usually not enough for busy cities. This project will aim on finding ideal locations for neighborhoods in Toronto to set up parking lots for the business on the main streets where is hard to find a parking place.

Data

The data for this project is retrieved from multiple sources with the considerations to the accuracy.

The data of the neighborhoods in Toronto is extracted out by using Python web scraping from the Wikipedia page, which including postal code, borough names, neighborhood names.

The latitude and longitude are retrieved using the Geocoder Python package by sending a given postal code to get its coordinates; and an acquired csv file could also be downloaded with the link provided in the appendix section.

The venue data is retrieved by passing required parameters to the FourSquare API. The data contains the contains the information regarding venue name, geographical coordinates of the

given venues and venue category. A combined DataFrame that contains all the venue details and respective neighborhoods information will be created for further analysis.