Coursera Capstone Project
The Battle of Neighborhoods

Shajid Hossain Hemal

December 30, 2020

Introduction/Business Problem

Suppose you are an entrepreneur and you are looking to open a Pub in Toronto, Ontario. But you cannot decide the location of the pub because the location of the pub has a significant impact on the expected returns.

- You want to open the Pub in a location where the business would be profitable, where there are many customers. So a populated spot would be perfect to open this pub.
- Also you would want a place where there few to none competition i.e. you don't want to place the pub in the immediate proximity of existing ones.

In order to answer this question, we would have to build a model to get recommendations on where to start your business.

Data

A description of the data: The data used to solve this problem is "List of postal codes of Canada" data collected from [Wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Data is a single dataframe, containing the postal codes, boroughs and the neighborhoods in Toronto, Ontario.

Postal Code	Borough	Neighborhood
M1B	Scarborough	Malvern, Rouge.
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek.
M3A	North York	Parkwoods.
M4A	North York	Victoria Village.
M5A	Downtown Toronto	Regent Park, Harbourfront.
M1H	Scarborough	Cedarbrae

The locations of the neighborhoods are collected by using python's geocoder package to get location information. Now the data contains **Latitude** and **Longitude** along with **Postal Code**, **Borough**, **Neighborhood**. Latitude and Longitude is absolutely necessary to get the venues from the Foursquare api. Example of the data after using gecoder:

Postal Code	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Methodology

Heatmap-based kernel density estimation was used. Heatmap was already implemented as plugin for Folium, which was used to visualize data to map. Visualization is shown in figure 1.
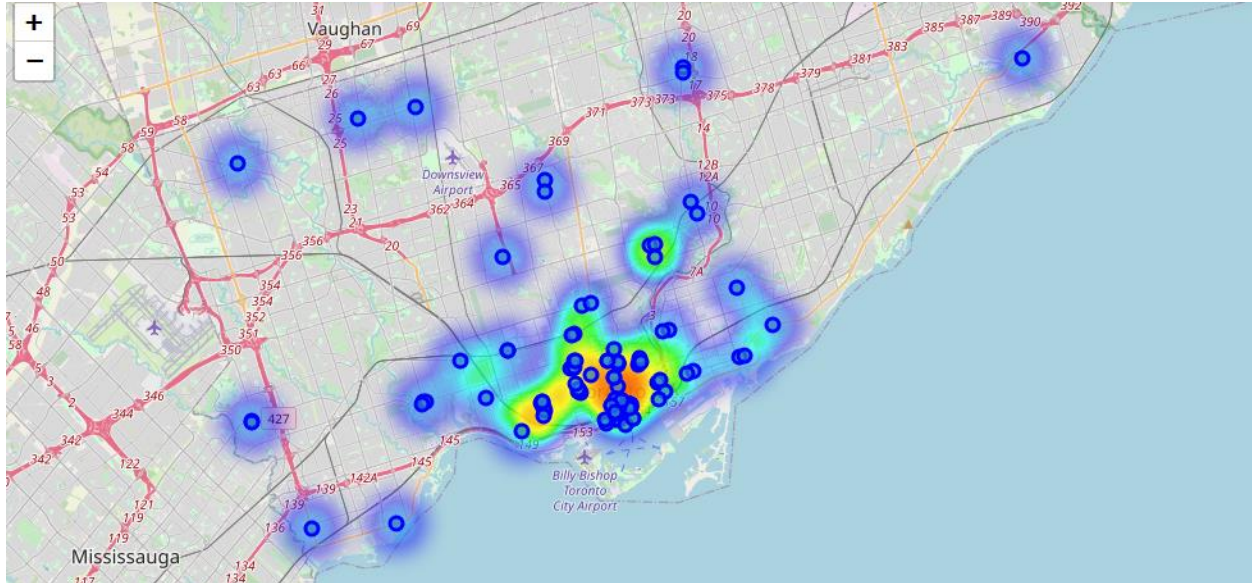

Figure 1: Data visualized to the map of Toronto, including heatmap based kernel density estimation.

Results

Based on these results, possibly good locations for new Pub would be in crossroad of Queens west, Fashion district and Entertainment district, shown in figure 2.
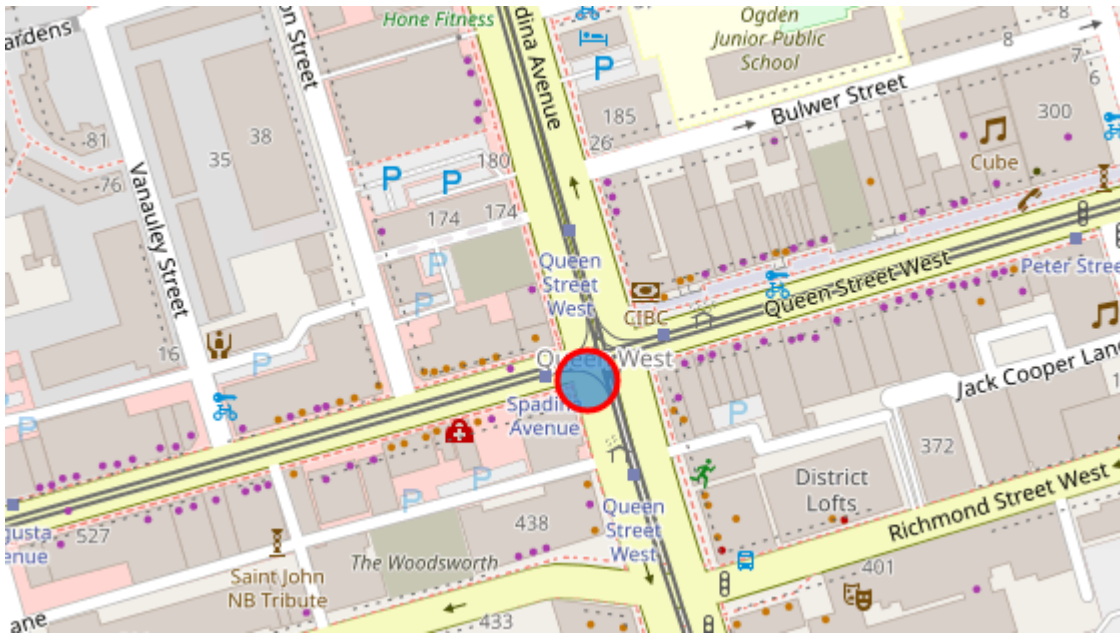

Figure 2: Proposed location for a new Pub.

Conclusions

Optimal location for a new Pub in center of Toronto was estimated based on data from FourSquare.