Predicting a suitable location for a new Coffee Shop in Toronto

Jason P Keeler

May 29, 2019

**1. Introduction**

1.1 Background

There are 230 Coffee Shops in the Toronto area. Therefore, it is advantageous for an entrepreneur to select a location that is not saturated with coffee shops when opening a new business in that highly competitive industry.

1.2 Problem

Data that might contribute to determining the best location for a Coffee Shop might include existing venues data to include competing venues and supportive venues in the prospective locations. This project aims to narrow the list of prospective locations based on the competing and supportive venues in those locations.

1.3 Interest

There is an entrepreneur that is already interested in opening a Coffee Shop and will use this analysis.

**2. Data acquisition and cleaning**

2.1 Data sources

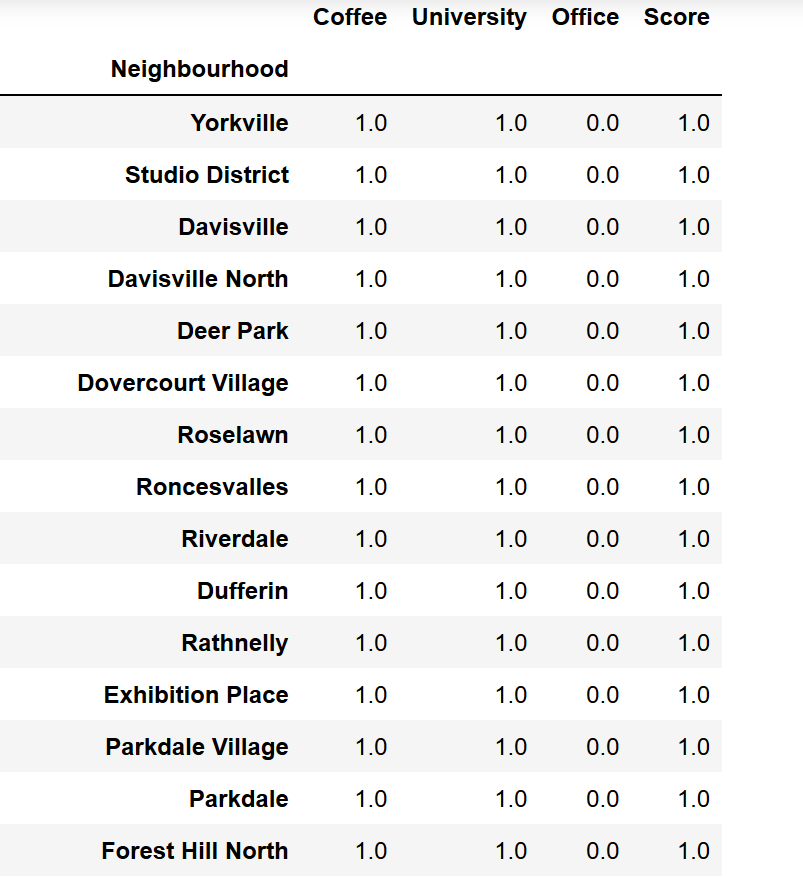
The location data comes from Wikipedia: https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M. Venue data comes from the Foursquare Developer API.

2.2 Data cleaning

Data downloaded or scraped from multiple sources and combined into one table. Panda Data frames were used to explore the data. Matplotlib, Folium, KMeans and BeautifulSoup are some of the packages used in the Project Notebook.

2.3 Feature selection

After data cleaning, there were 210 neighborhoods in the Toronto area. In those 210 neighborhoods there were 4191 venues, with 90 unique categories of venue. I selected 3 venue categories as having significance to our selection of Coffee Shop location. The venue categories selected were existing Coffee Shops, Universities and Offices. Upon examining those venue categories, a weight was given to each category with existing Coffee Shops having a negative weight and the other 2 categories having differing positive weights. The weighted score for each neighborhood was then determined based on the density (number) of significant venues in each neighborhood. Using this method, I was able to limit the potential locations for a Coffee Shop from the countless random options, to 210 neighborhoods, to just 37 neighborhoods with a high weighted score. The high weighted score is calculated as those neighborhoods with weighted score greater than equal to 1 based on the above criteria.

 (Table 1).

**3. Exploratory Data Analysis**

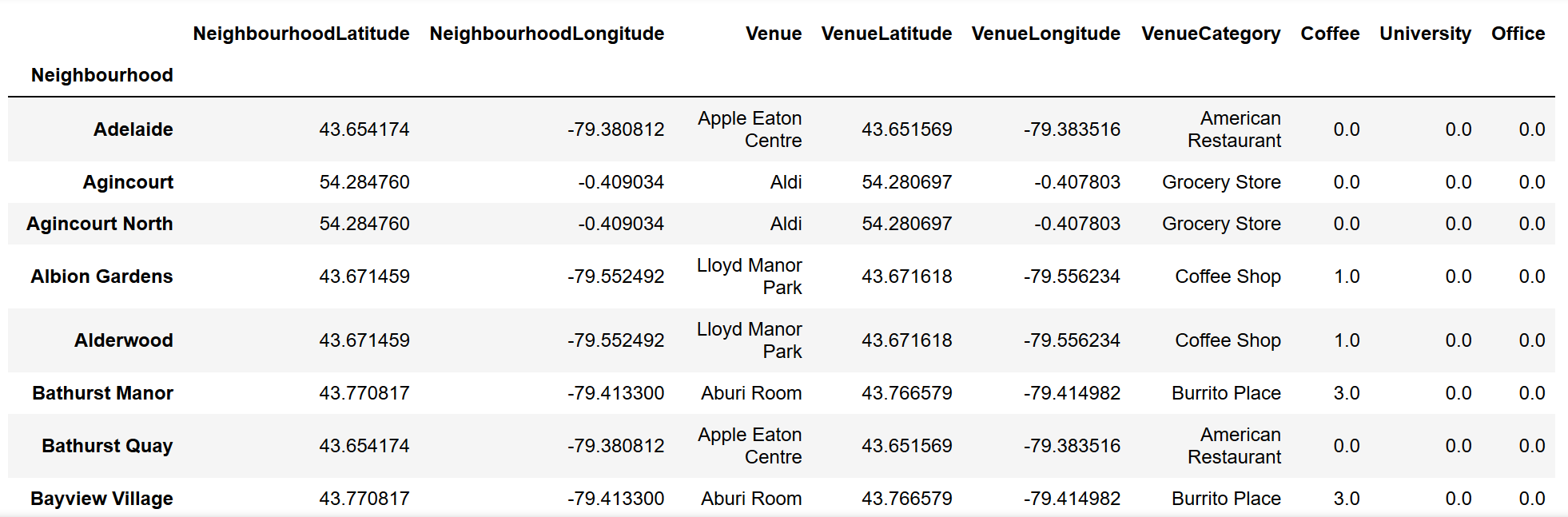
3.1 Calculation of target variable

(

(Table 2).



(Figure 1).

 (Table 3).

**4. Conclusions**

It was determined from this Data Science Project that the best neighborhoods for opening a new Coffee Shop in the Toronto area are: Yorkville, Studio District, Davisville, Davisville North, Deer Park, Dovercourt Village, Roselawn, Roncesvalles, Riverdale, Dufferin, Rathnelly, Exhibition Place, Parkdale Village, Parkdale, Forest Hill North, Forest Hill SE, Forest Hill West, North Toronto West, North Midtown, Moore Park, Little Portugal, High Park, Lawrence Park, South Hill, Runnymede, Brockton, The Annex, Trinity, Summerhill East, Business Reply Mail Processing Centre 969 Eastern, The Junction South, The Danforth West, The Beaches West, The Beaches, India Bazaar, Swansea and Summerhill West. This list of neighborhoods along with the tables and figures included in this report were provided to the entrepreneur looking to open a new Coffee Shop in the Toronto area.

**5. Future directions**

If additional support is needed in narrowing the list of locations for the entrepreneur, then future iterations of this project could include frequency of coffee shop vs other venues in calculations rather than count. Also, additional venue categories could be added to the features for predicting best location. Finally, population density and demographics data could be incorporated into the project for added specificity.