Applied Data Science Capstone

Week 5 Presentation

Kyle Thomsen

22 Dec 2019

- For my project, I have decided to tackle the question of determining which Toronto neighborhood would be most suitable for the opening of a new franchise for a restaurant based on the current density of similar places.
- Neighborhoods with lower restaurant densities would provide less competition for the new franchise. The target audience of such a study would be the owners of a company seeking to expand its operations into the Toronto area.
- They could use the results of this analysis to narrow down the potential sites for the new location. Furthermore, doing so would allow them to focus their resources on investigating just the list of locations obtained here instead of the whole of Toronto.

To assess which Toronto neighborhood would most benefit from the addition of a new restaurant, it would be necessary to determine which neighborhood has the lowest density of current restaurants. This requires a couple of parameters:

- the total number of restaurants in the neighborhood
- the total area of the neighborhood

- By finding the ratio of these two numbers, it is possible to determine which neighborhood has the lowest restaurant density.
- Therefore, it is possible to determine the neighborhood in which the company will have the least competition for consumers.
- While the area of each neighborhood is a matter of trivial geometry, the number of venues within its borders is more difficult to ascertain. By leveraging the Foursquare API, it is possible to determine the number of venues with a type similar to "restaurant" for each neighborhood and thereby determine the baseline competition the new franchise will encounter in each neighborhood.
- An example of the sort of data that will be collected is a list of the venues for each neighborhood. A feature which can be extracted from these data is the category of each venue, which will then allow me to focus on only those which are considered restaurants. A more concrete example of this is included in the following table.

Neighborhood	Venue	Venue Lat	Venue Long	Venue Cat
Malvern, Rouge	Wendy's	43.807448	-79.199056	Fast Food Restaurant
Port Union, etc.	Royal Canadian Legion	43.782533	-79.163085	Bar
West Hill, etc.	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
West Hill, etc.	G & G Electronics	43.765309	-79.191537	Electronics Store
West Hill, etc.	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant

- To answer this question, I began by associating the neighborhoods near each postal code in the Toronto area with the latitude and longitude of the center of the postal code's designated area.
- I accomplished this by first scraping the list of Toronto postal codes on Wikipedia using BeautifulSoup.
- This involved identifying the HTML tags associated with the table, namely <tbody>, associating the elements of the *.contents attribute with the table's rows, and appending each row to an array of strings.
- In order to transmute this array into a DataFrame, I split the string corresponding to each row into its three component cells, corresponding to a triple consisting of a postcode, borough, and neighborhood.
- In order to facilitate integration with the Foursquare API, I opted to group together the neighborhoods which shared a postcode before associating each group with the latitude and longitude of its center. These mappings were obtained by reusing the Geospatial_Coordinates.csv file from the week 3 assignment.

- After initializing my connection to the Foursquare API, I configured the parameters defining the maximum number of venues to return for each query and the radius around each latitude/longitude pair in which venues should be located.
- I eventually settled on a venue maximum of 1000 and a radius of 1000 meters. After obtaining the results from Foursquare as a JSON object, I then extracted the category for each venue.
- After obtaining the venues for each neighborhood group, I then used one-hot encoding to analyze the categories present in each neighborhood.
- I filtered the categories to only retain those indicating a type of restaurant before summing the total number of restaurants associated with each group of neighborhoods.

- Finally, to obtain the restaurant density for each group, I simply divided the total number of restaurants by the area over which the Foursquare API searched for venues.
- In this case, that area is 1000 m * 1000 m * $\pi = \pi$ * 10^6 $m^2 = \pi$ km^2.
- While the densities produced are quite small, as they are given in units of restaurants per square meter, the sorting of the neighborhoods by density is independent of the units used.

The results of my analysis revealed that the following neighborhoods all have the minimum restaurant density:

- Port Union, Rouge Hill, Highland Creek
- Malvern, Rouge
- West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
- Weston
- Scarborough Village
- Humberlea, Emery
- York Mills, Silver Hills

while these neighborhoods all have the same maximum value for theirs:

- Toronto Dominion Centre, Design Exchange
- St. James Town
- Victoria Hotel, Commerce Court
- Garden District, Ryerson
- Union Station, Toronto Islands, Harbourfront East
- Underground city, First Canadian Place
- Richmond, King, Adelaide

- To have such a large number of neighborhoods tied for both density extrema is an artifact of the ultimately discrete nature of the Foursquare API.
- Specifically, the minimum value of 3.183099 * 10^{-7} restaurants per m^2 corresponds to only one restaurant in the search area, while the maximum of 3.183099 * 10^{-5} restaurants per m^2 corresponds to 100 restaurants in the search area.
- The minimum seems sensible, as there is almost always a restaurant within a kilometer of any given location in a major city.
- However, the repetition of the maximum value is curious; there doesn't appear to be a reason behind the 100-restaurant maximum inherent in one's intuition about urban environments.
- In this respect, I suspect that this is some sort of internal limitation of my free Foursquare developer account.

The lists of neighborhoods give the best and the worst places for the new restaurant to be opened. With the minimum restaurant density, the new restaurant will have less competition; correspondingly, those neighborhoods with the maximal density are already saturated with restaurants. This means that any of these neighborhoods would be a suitable location for the new restaurant:

- Port Union
- Rouge Hill
- Highland Creek
- Malvern
- Rouge
- West Deane Park
- Princess Gardens
- Martin Grove
- Islington
- Cloverdale
- Weston
- Scarborough Village
- Humberlea
- Emery
- York Mills
- Silver Hills

- Since this doesn't narrow down the potential optimal locations to a satisfactory degree, my recommendation to the restaurant owners would be to consider a follow-up analysis on this subset of neighborhoods in order to further pare it down.
- This could be based on, for example, the accessibility and local tastes of each neighborhood.
- In this way, this list of sixteen neighborhoods may be filtered down to just a few.
- Upon selecting a neighborhood, it would be beneficial to give further consideration to the new restaurant's location within the neighborhood in order to maximize its business opportunities.

- This analysis has produced a list of sixteen neighborhoods which are each associated with only one restaurant within a kilometer of their centers.
- Thus, it would be of benefit for the restaurant owners to select one of them for the new location.
- This would minimize the competition encountered at the new location, which in turn would offer it the greatest chances to thrive.
- While more analysis can be conducted to refine these results, the reduction from the full list of all 140 Toronto neighborhoods to 16 represents a narrowing of the options by 88.6%. This would yield, if not a specific location recommendation for the future restaurant, something far closer to one than existed before this analysis was conducted.