Battle of the Neighborhoods | Opening a new Coffee Shop in Toronto

Efraín López

August, 2019

1. Introduction

1.1 Background

Deciding to become an Entrepreneur is a very daunting experience. There's a lot of unknowns to fight against. Take, for example, the case of a restaurant. What kind of food would my potential customers would like? Where are they located? Is there going to be competition? How fierce is it? All of this uncertainty can be diminished by using Data Science.

1.2 The Problem

My client owns several Coffee Shops in New York City. She has been told about the huge opportunities that Toronto poses, so she's seriously considering the idea of opening one of her Coffee Shops there but, where?

To determine the best location to open up the business, we'll look for a mix of distance from a point of interest, such as the center of a very well known neighborhood, density of potential competitors (restaurants) per area and density of direct competitors (other coffee shops).

2. Data acquisition and cleaning

2.1 Data Sources

For our analysis we'll need two kinds of data: geo data and market data. For geo data, we'll be using Google Map's API and for Toronto's restaurant data, we'll use Foursquare's. Because we will make a relatively small number of calls and the features needed will be very basic, both free tiers should suffice.

2.2 Data Cleaning

Because Google and Foursquare are very big brands, the way they store and manage data is very clean. No further data cleaning was needed, besides cleaning up redundant data, such as

deleting “Canada” from the addresses, for example, as we already know that all addresses will be located in Toronto, Canada.

2.3 Feature Selection

The advantage of extracting data from APIs such as Foursquare’s and Google Maps’ is that you can directly select the features you want- there’s no need to drop features.

From Google Maps data, we extracted the address of the potential locations, their gps coordinates, cartesian coordinates and distance from the point of interest. On the other hand, on Foursquare we used the Venue Names, Categories, Location and Distance.

3. Exploratory Data Analysis

3.1 Defining the area for analysis

Since Toronto is a very big city, we focused our efforts in one of the most populated areas within the city: Downtown Toronto. We used the area’s center point as a starting location for our analysis. Within a radius of 6km, we created a grid of 364 candidate neighborhoods with a radius of 600m each. We then assigned an address to each neighborhood candidate center. (Figure 1)

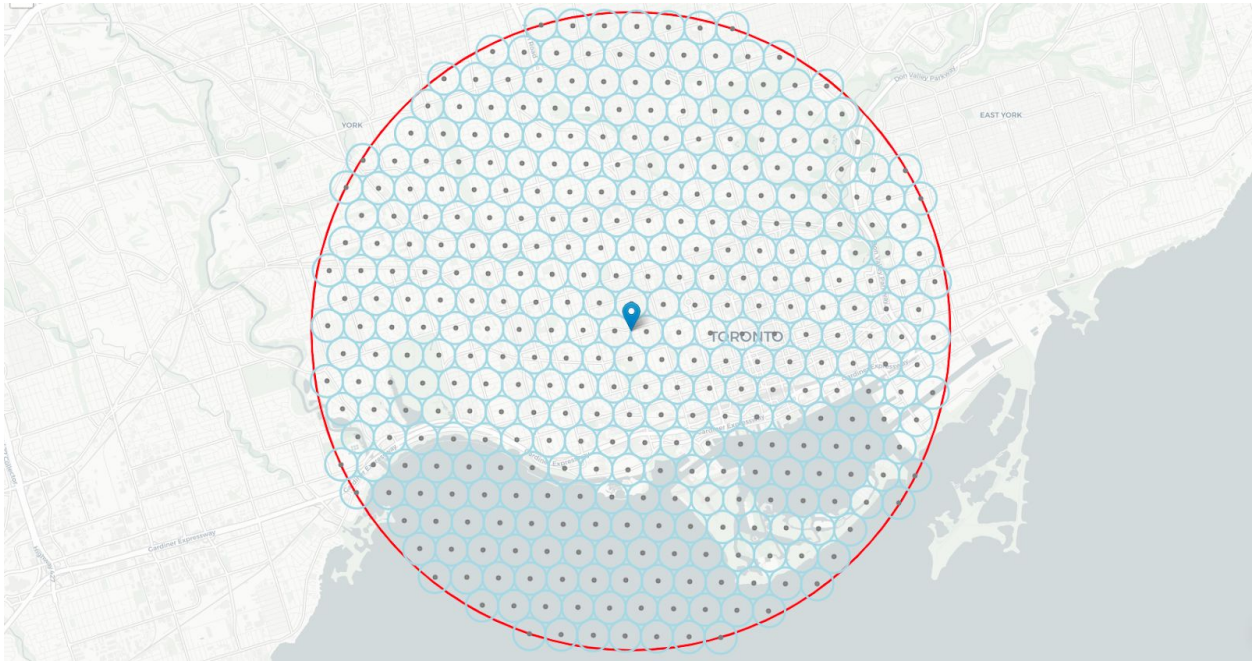

Figure 1. Visualizing the grid.

3.2 Analyzing the area

The grid helped us to focus our research. We used it to channel Foursquare's data extracting. After we had our data extracted, we found 1768 restaurants within that area. Out of those, 365 of them were coffee shops. That is, 20.64%.

The most important finding of our analysis was finding the hottest area of the city. As you can see in Figure 2, most of the action happens in the Financial District. This can give us a hint: it's probably better to focus on where the population density is higher. In this case, it makes a lot of sense: not only people commute everyday to where the businesses are, but it probably has the biggest amount of affluent people per area in Toronto. It's just a matter of finding the best spot with the right mix of factors. That's why we focused the next part of our research into exploring that area further.

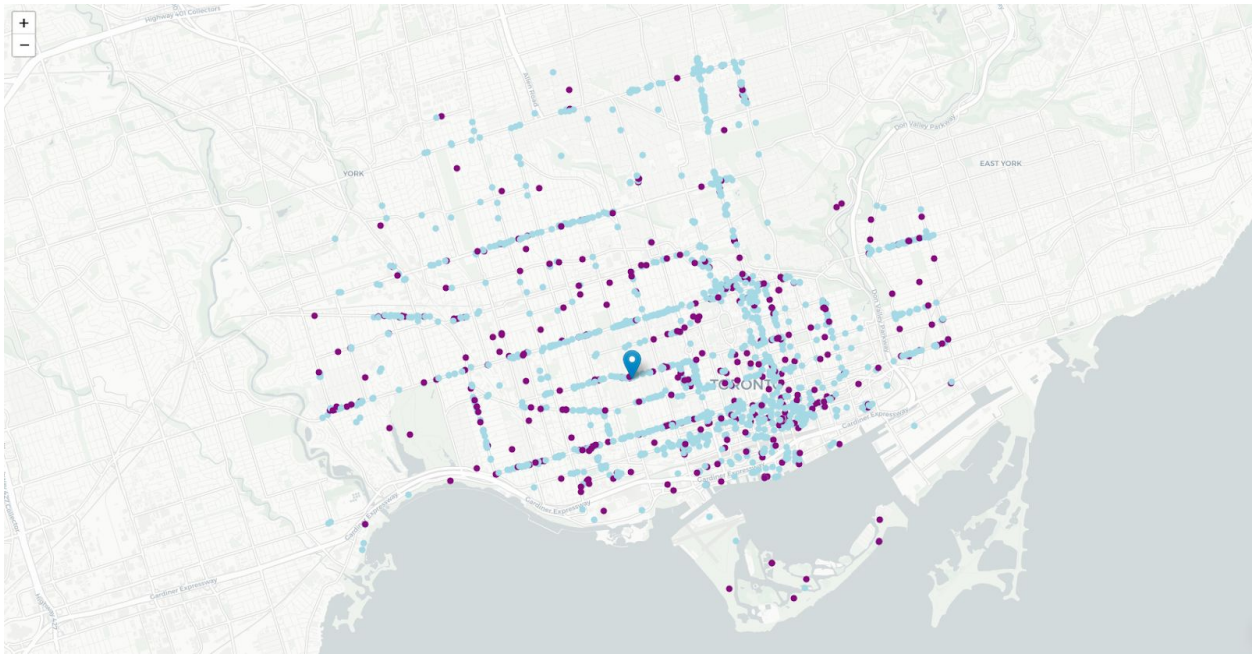

Figure 2. Map of Downtown Toronto with restaurants visualized. Notice how most of them are located in the Financial District.

3.3 Moving our area of analysis - narrowing our scope

In order to dig further, we needed to narrow our scope. (Figure3) We used the same methodology of creating a grid and then looking for potential competition within that area, but this time we narrowed it down to a smaller total area (1.5km radius from the Financial District center) with a tighter grid as well (100m radius).

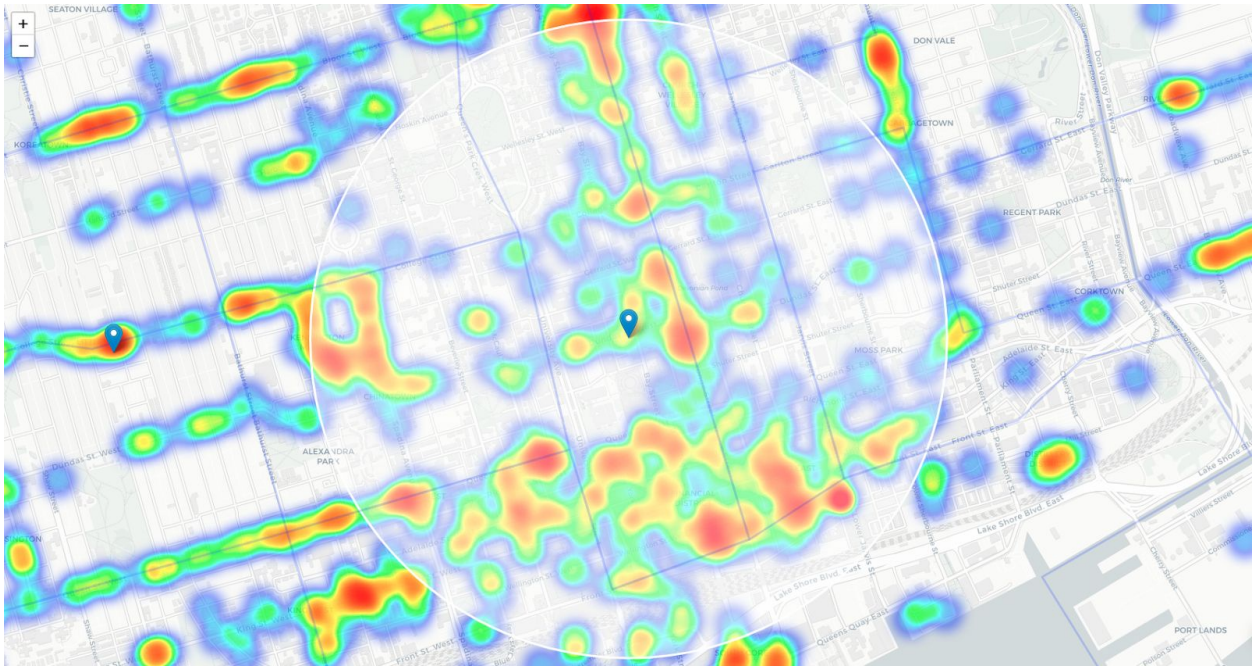

Figure 3. Heatmap of restaurants within the Financial District Area.

After our analysis, we found 109 locations with no more than two restaurants nearby, 124 areas with no cafeterias within 250m and 80 locations that met both conditions. (Figure 4)

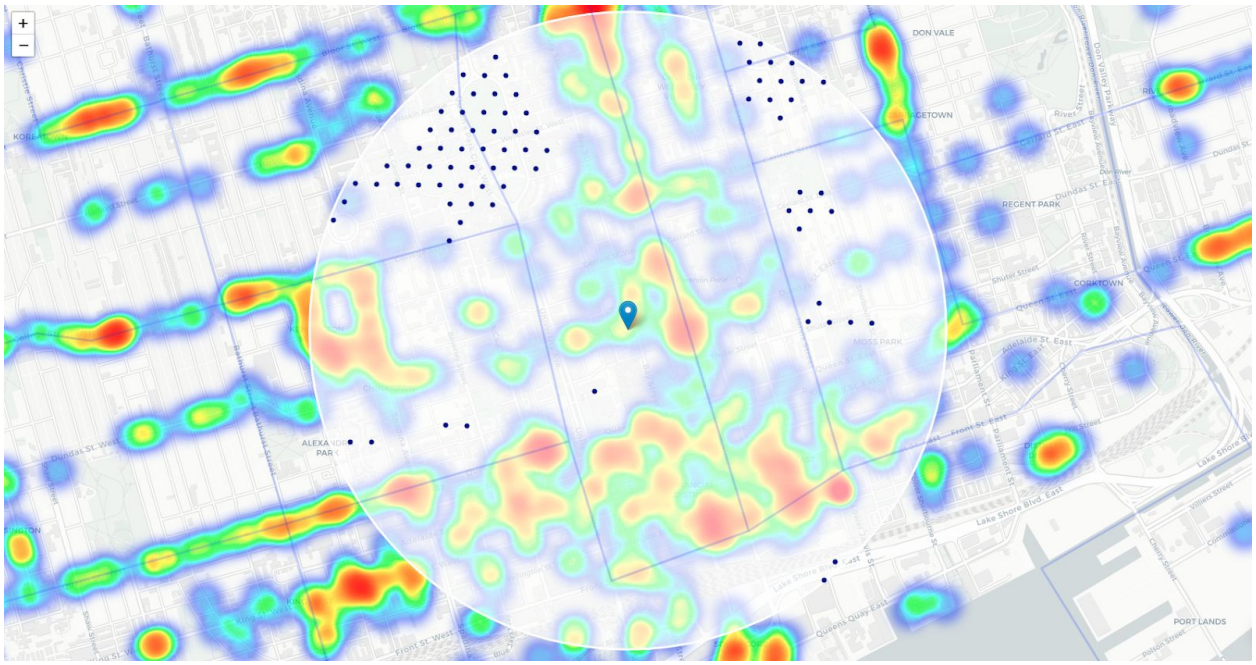

Figure 4. Heatmap of the Financial District with locations that met both conditions highlighted.

We then used Kmeans to cluster those locations in order to extract the addresses and get a list of potential locations. (Figure 5)

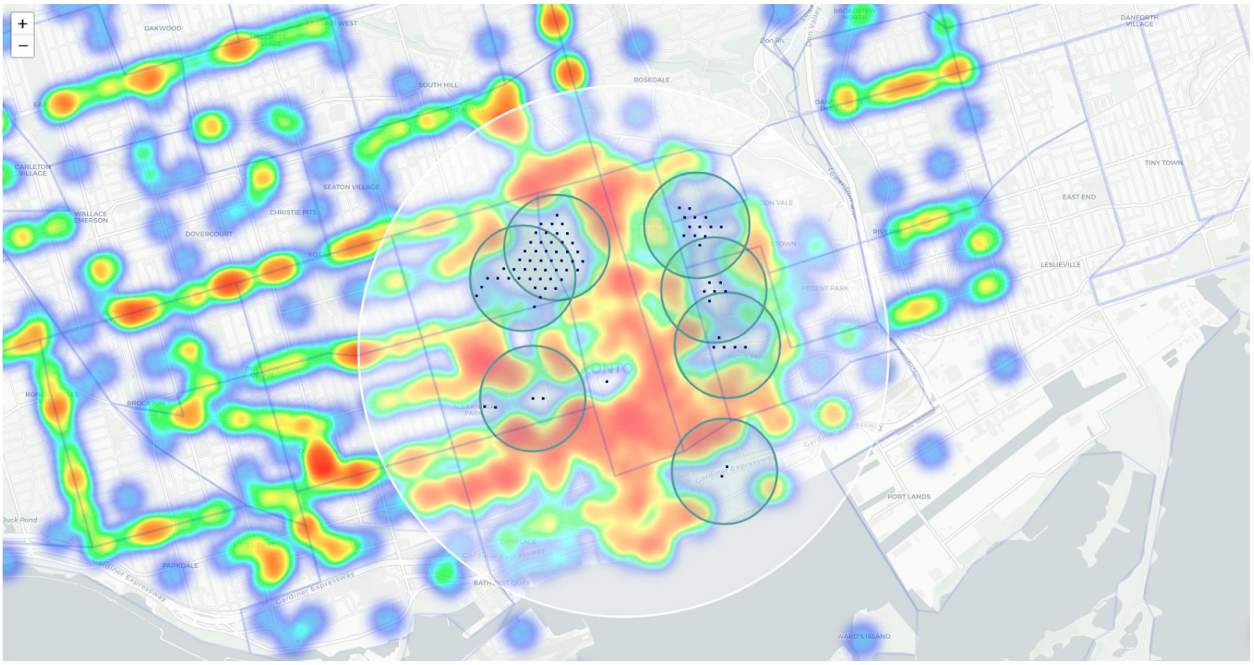

Figure 5. Heatmap with clusters, visualized.

Finally, our analysis came to an end when we found 7 areas that met the specific criteria we initially set up, as seen in Table 1.

Addresses of centers of areas recommended for further analysis	
25 King's College Cir, Toronto, ON M5S 3K1	1.2km from Financial Center
155 Wellesley St E, Toronto, ON M4Y 1J4	1.4km from Financial Center
120 Pembroke St, Toronto, ON M5A 2N8	1.0km from Financial Center
241 Lake Shore Blvd E, Toronto, ON M5E	1.5km from Financial Center
37 Sullivan St, Toronto, ON M5T 1B8	1.0km from Financial Center
Queen's Park, 111 Wellesley St W, Toronto, ON M5S	1.2km from Financial Center
128 Shuter St, Toronto, ON M5A 1V8	1.0km from Financial Center

Table 1. Addresses of centers of areas recommended for further analysis

4. Results/Discussion

We have found a way to narrow down, from an almost shapeless, non related data, a list of locations that will definitely prove useful for any entrepreneur. This analysis started by selecting a city, then selecting a point where to start our analysis (in this case, Downtown Toronto). Then, we used Google Map's API to fetch location data to get a list of possible addresses where our business could be located.

Foursquare data proved to be invaluable. Combining both data sources was vital to understanding the distribution of restaurant businesses in Downtown Toronto. This led us to the conclusion that it was probably a good idea to focus were the action was: the Financial District.

By readjusting the area and location to the Financial District, we found some pockets of opportunities that our stakeholders can take advantage of.

It's important to notice that, since we had very limited access to Foursquare's API, we couldn't find more qualitative data for our analysis. Of special interest would've been not only to have access to the locations and categories of the restaurants, but their scores as well. This could've helped us determine if there was a correlation between certain aspects (such as location, distance to specific points of interest, etc) to the perceived value of the restaurants in order to give better recommendations. Hence, further research is needed to improve these recommendations.

5. Conclusion

We have successfully created an analysis based on freely available data. This kind of business research should become standard practice in the near future. It's important to note, however, that the quality of the results will be greatly improved by having better quality data sources, such as having an improved access to Foursquare's data.

This analysis should be taken as a very solid first step towards deeper research. In the end, the stakeholders should take into account other aspects when making a decision.