

Battle of the Neighborhoods: Toronto

Market Analysis: Sandwich Shop

Applied Data Science Capstone

Jonathan Donnelly

April 30, 2021



Introduction:

The capital city of Ontario, Canada is arguably the most important city to Canada's culture and economy. Toronto the single most populous city in the entire country also acts as the center of Canada's finance and commerce. This is a multicultural city resting on the border with the United States making it an easy center of trade between the two countries with the benefit of also having access to shipping across the Atlantic. Toronto over the last several decades has seen its population explode from just over 1 million residents in 1951 to over 5 million by 2006. With this expansion in population there came an increase in the number of businesses located within the city as well. All told this makes for a market that can be extremely profitable as well as brutal to break in to.

The purpose of this project is to utilize multiple data sources to identify areas within Toronto where a sandwich shop would have the greatest opportunity for success. Choosing the right location to open a business is a critical step. The wrong location could lead to having too much competition, not enough customer base, or potentially pricing yourself out of the region. This project aims to eliminate these risks and be able to make recommendations on which neighborhood the business owner should target for their opening.

Data:

Multiple data sources will be needed throughout this project:

- Data regarding the existing businesses throughout Toronto including classification, names, and coordinates.
- Data for the names of boroughs, neighborhoods, and coordinates pertaining to Toronto.
- Data on the population and income of residents of Toronto based on the most recent census performed in 2016.

Methodology:

For this project we will need to rely on multiple methods learned throughout the courses within this certification. We start by webscraping data from Wikipedia for a list of boroughs, neighborhoods, and postal codes for Toronto. We will clean the data to bring it into a usable state and create a dataframe.

We then will read in a csv obtained from Cognitive Class containing the postal codes and geographic coordinates into a new dataframe and then merge the two dataframes into a single dataframe with all data matched by postal code. We will be able to utilize this dataframe with Folium to create a visualization showing a map of Toronto with each of the neighborhoods flagged.

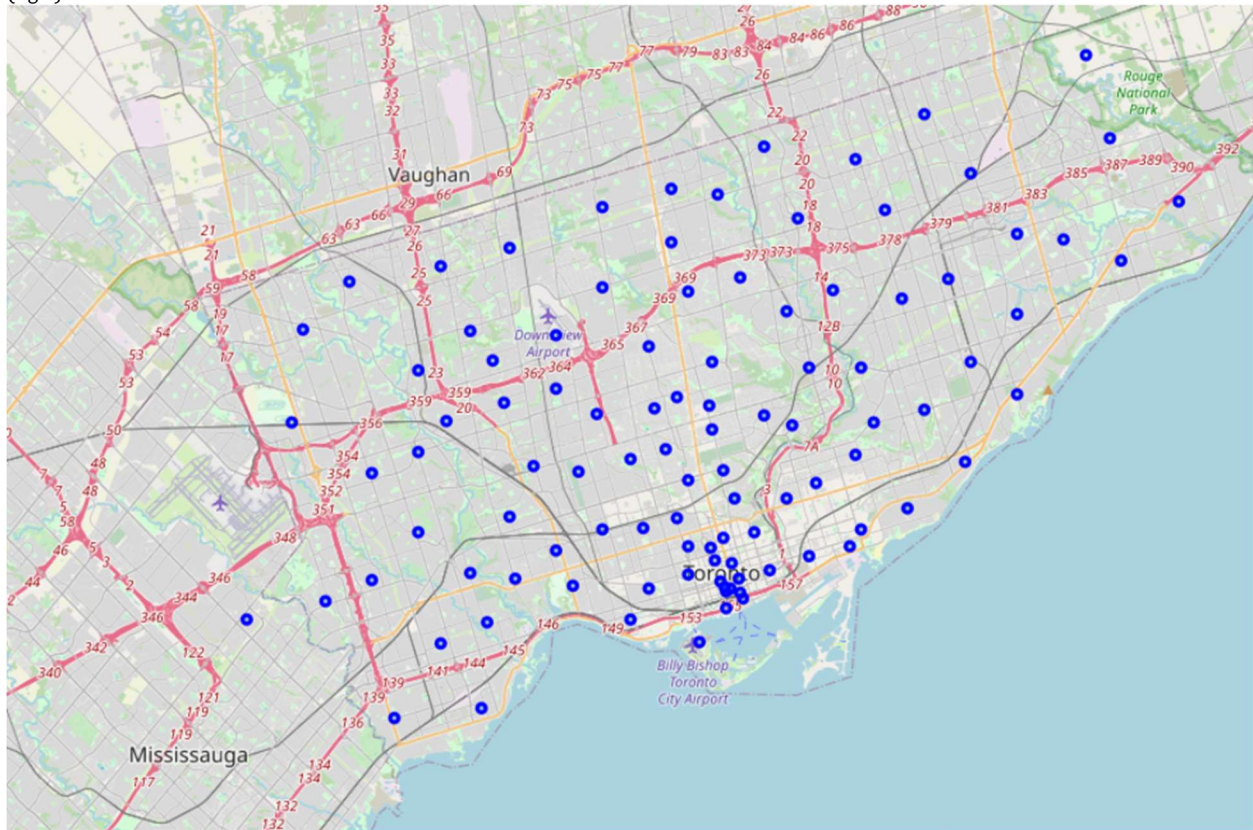
By utilizing the fetch capability with the FourSquare API, we can generate a list of all existing venues within Toronto, their names, locations, and classifications and match those with our list of neighborhoods to determine what the market in each neighborhood includes. We can filter this down to look specifically for venues which specialize in sandwiches (Sandwich Place and Deli/Bodega per the FourSquare classification system).

The data obtained will be analyzed and used to create visualizations to easily communicate the findings.

Results:

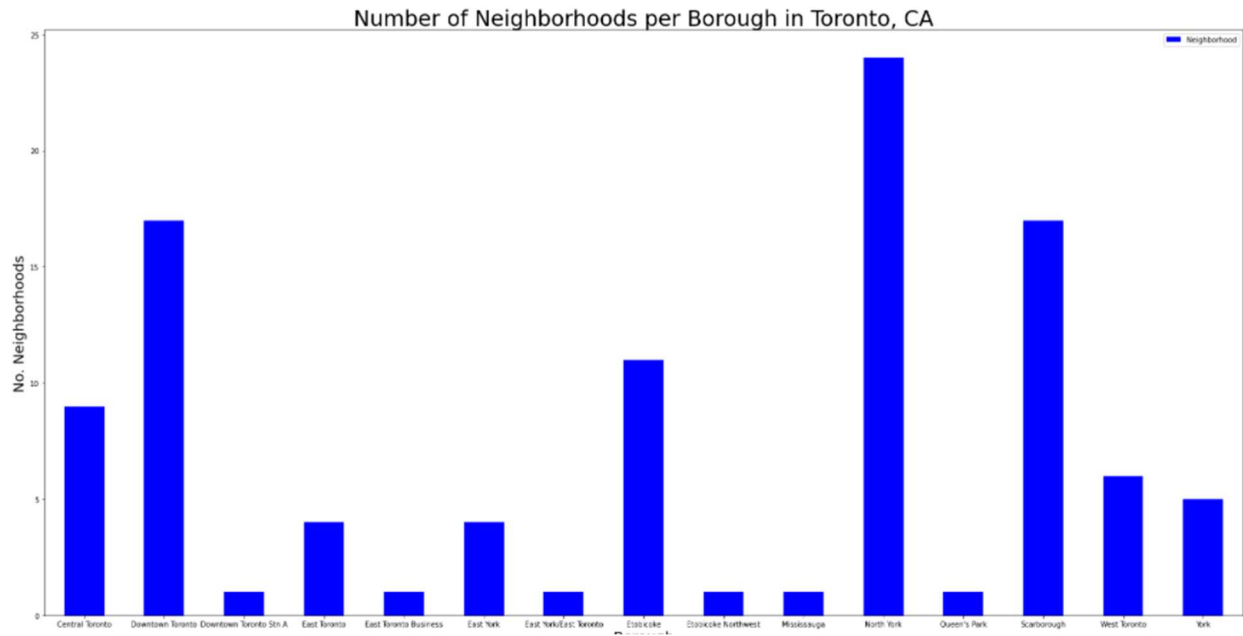
By merging the coordinates for each postal code found through Cognitive Class with the Wikipedia data listing the neighborhood names we are able to map the locations of each neighborhood (fig.1). As we can see Toronto has a large number of individual neighborhoods. A small business owner would be hard pressed to understand the market of all of these neighborhoods enough to make an informed decision on where to open a new sandwich shop.

(fig.1)



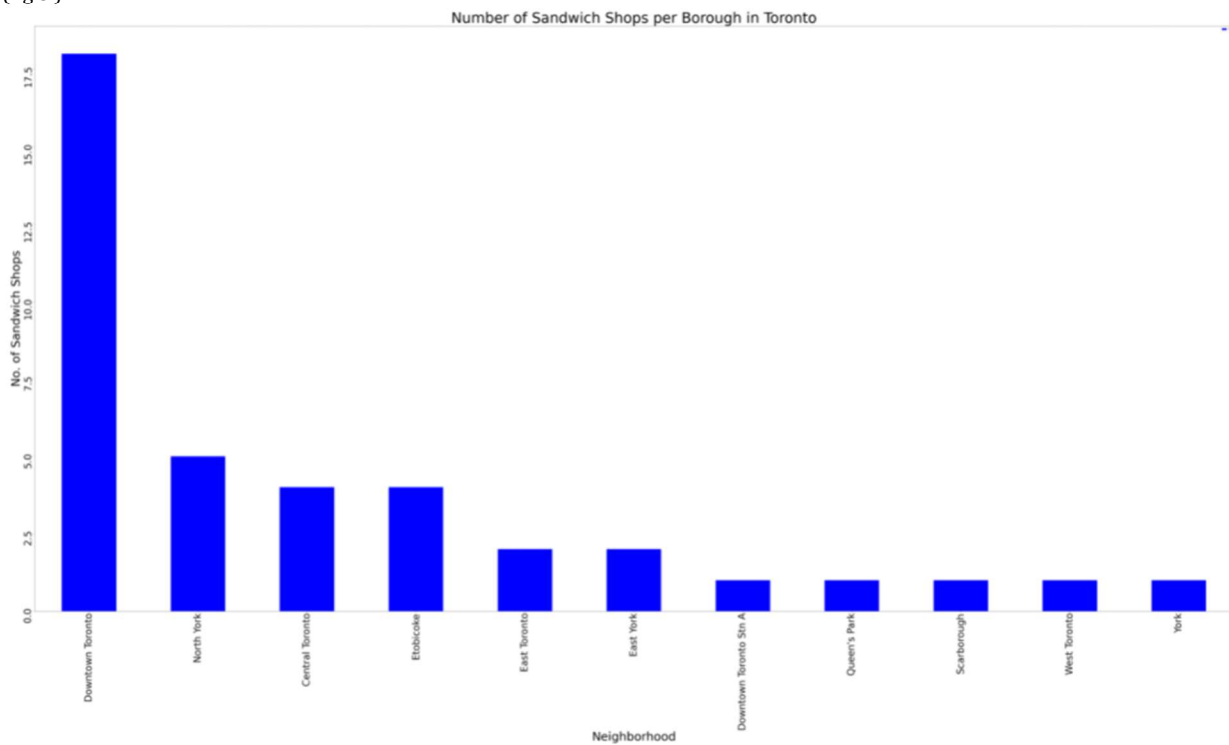
Like many other cities (New York City for example), Toronto is not broken into mere neighborhoods. Instead, it has both boroughs and neighborhoods. Each borough contains multiple neighborhoods as subdivisions. In fig. 2 below we can see that the size and number of neighborhoods within each borough varies significantly making boroughs inappropriate to utilize for our analysis.

(fig 2.)



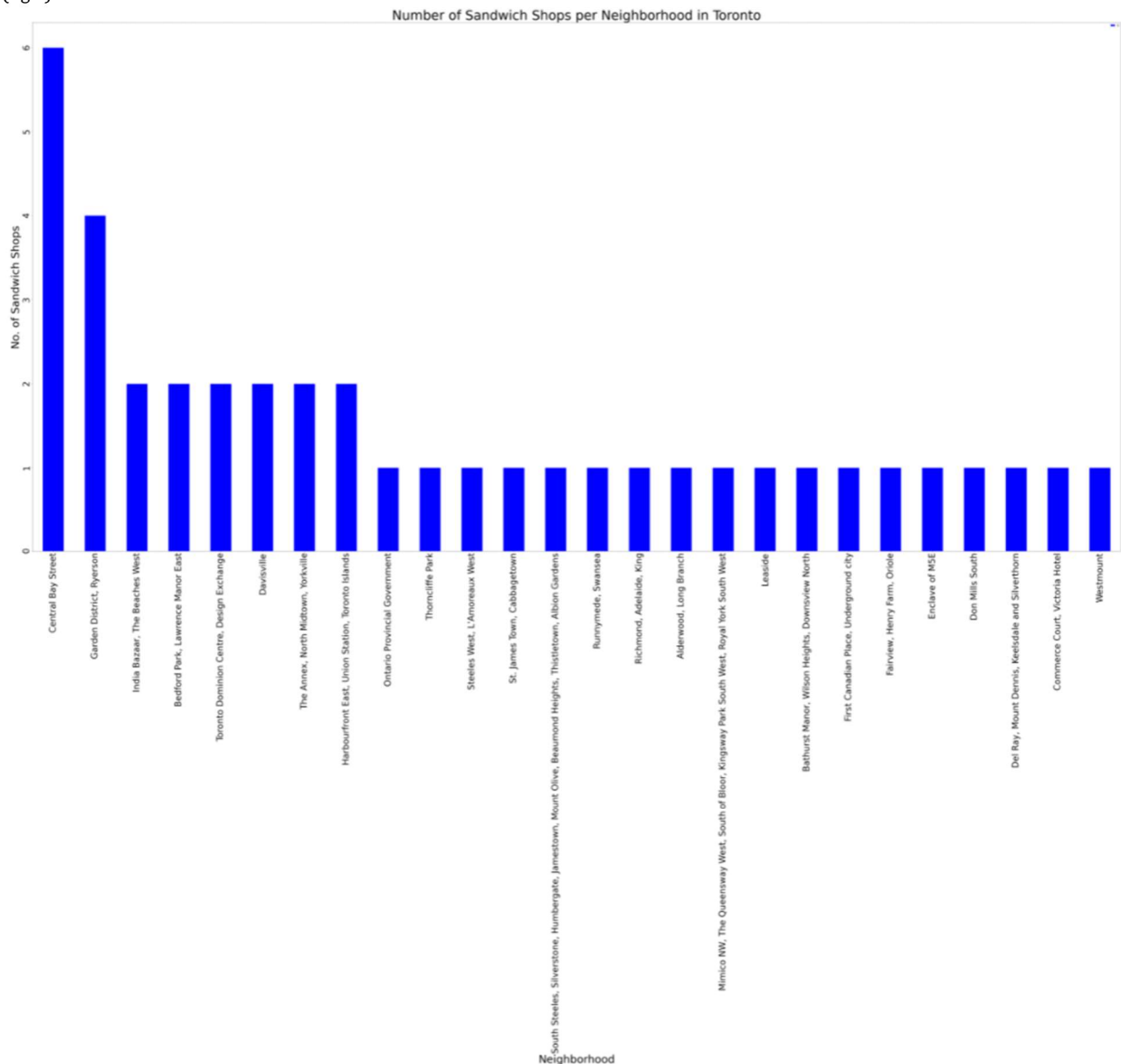
By bringing in the FourSquare venue data we are able to provide a count of venues specializing in sandwiches per Borough as seen in fig 3.

(fig 3.)



At this point we know we are going to be looking at Neighborhoods for our decision making and recommendations instead of Boroughs. In fig. 4 we have further narrowed our focus to the neighborhood level. To ensure we had an accurate count we reviewed the FourSquare classifications and determined that there were two classifications that fit our needs: sandwich place and deli/bodega. Through this analysis we were able to confirm that Toronto did not have any venues that were classified as a deli/bodega. However, we left that definition in the analysis to ensure accuracy in the future. In fig. 4 we display the counts per neighborhood.

(fig.4)



Based on the data visualized in fig 4. we identify 11 neighborhoods for consideration. The remaining analysis is narrowed in focus to just these potentials. We look at population within each and income. The dataframe in fig 5. displays the number of households within each neighborhood that fall within income brackets. This will help us to identify neighborhoods that will fit our target market.

(fig 5.)

	Under \$5,000	5,001 9,999	10,000 14,999	20,000 24,999	25,000 29,999	30,000 34,999	35,000 39,999	40,000 44,999	45,000 49,999	50,000 59,999	60,000 69,999	70,000 79,999	80,000 89,999	90,000 99,999	100,000 199,999	200,000+
Alderwood	55	45	80	145	150	155	170	160	165	335	300	320	275	250	1915	360
Bathurst Manor	130	85	155	245	230	220	255	240	245	400	430	390	310	300	2025	515
Cabbagetown- South St. James Town	175	160	330	330	265	245	240	235	255	425	420	350	295	270	1925	700
Henry Farm	460	170	210	280	290	265	280	280	285	545	465	470	415	340	1320	230
Leaside- Bennington	120	65	80	150	160	150	155	140	165	310	265	265	270	230	3775	2260
Mimico (includes Humber Bay Shores)	570	370	635	750	665	660	645	705	675	1270	1255	1120	1105	920	5455	1245
Mount Dennis	110	125	255	285	290	310	295	240	280	490	425	315	320	275	860	50
Runnymede-Bloor West Village	45	30	45	95	85	95	110	95	100	195	190	195	170	170	2110	875
Steeles	195	115	150	315	405	410	395	300	315	565	605	515	475	430	2245	380
Thistletown- Beaumont Heights	40	35	65	150	140	160	165	170	180	290	270	185	210	185	940	120
Thorncliffe Park	250	160	255	375	490	490	465	470	415	730	560	425	360	270	925	100

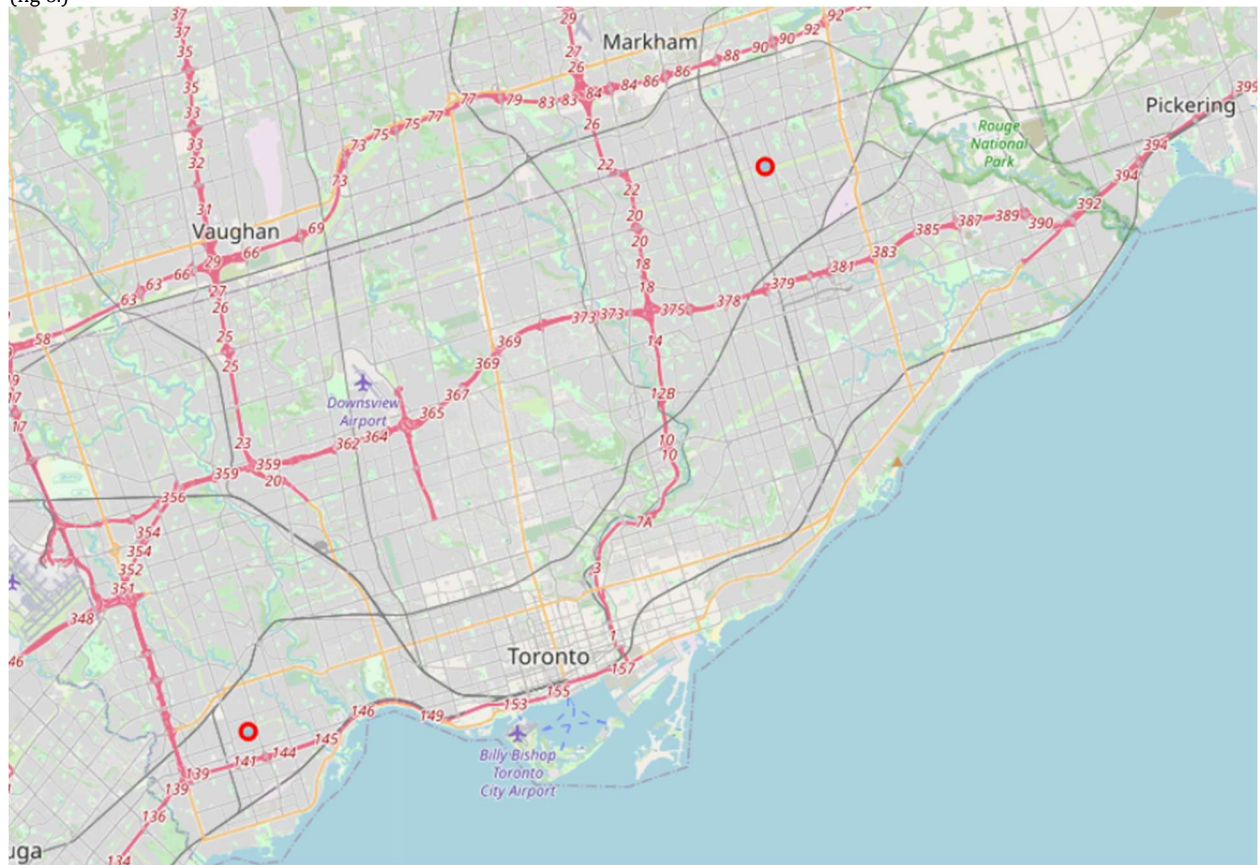
Discussion:

The analysis we have performed has looked to identify neighborhoods within Toronto that had room for a new sandwich shop to gain a foothold. To this effort we were able to identify that there were 11 neighborhoods that had 1 sandwich shop indicating that there was interest in this type of venue but not being an oversaturated market. Out of these neighborhoods we then looked at population and income to further narrow our results. Looking first to population there were three that stood out as having a larger population than the rest. For our purposes, a large population is preferable as it means more customers for us to target. The three largest populations were in Mimico, Steeles, and Thorncliffe Park. Mimico alone has approximately double the population of many of the other neighborhoods. The final part of our analysis looked at household income and counted the number of households present in each neighborhood that fell within specific brackets. Again, our attention is drawn to Mimico and Steeles. Both neighborhoods display a significant number of households that have an income of 100,000 or more. This would

indicate that they potentially have more disposable income to spare towards activities such as dining.

After reviewing available data our recommendations would be to consider the Mimico or Steeles neighborhoods as viable locations to open a new sandwich shop. The potential customer base combined with income in these neighborhoods make them ideal candidates while neither is seen as an oversaturated or unwelcoming market for a sandwich shop. Should the business owner be able to open multiple shops the recommendation would be to open one in each. As seen below in fig 6. The neighborhoods are separated from each other such that they would not be in competition with each other.

(fig 6.)



Conclusion:

Through the course of this project, we have worked to identify the problem and data needed to solve the problem, as well as gathered and cleansed the data available for our purpose. We took that data and worked our way

through a detailed analysis to so that we may provide the client our recommendations. Specifically, we narrowed possible locations down from 140 neighborhoods to just 2 which would suit their needs. All of this was possible by utilizing all the skills learned throughout the course of this certification.