Coursera Capstone

Opening a new Japanese Restaurant in Toronto

Christopher Kalanderopoulos
February 18, 2021

INTRODUCTION

Background

We are looking to open a new Japanese Restaurant in Toronto. In order to maximize our investment and draw in the largest volume of patrons, we need to determine the ideal location for the new restaurant.

We will present a visual analysis of the ideal location, based on multiple factors, including current Japanese restaurants and neighbourhood populations, to solve this problem.

Problem

What is the best neighbourhood to locate the new restaurant in?
What is the best actual latitude and longitude to locate the new restaurant?

Interest

Those who are looking to open a new Japanese Restaurant in Toronto would benefit from this data at time of publishing, February 18, 2021.

DATA

We will use the following sources for our data. We clean and extract the necessary data to lead us to the correct decision based on the Business Problem. We are fortunate with this data that not a lot of cleaning is required. As the Wikipedia data is minimal and sorted into an appropriate dataframe already, no changes are necessary. StatsCan being a government organization follows conventions for organizing data appropriately, and the source can be trusted as a primary source of information. Foursquare, as the provider of data for restaurants and other venues is also an excellent source of information, and using their API calls means that we do not need to clean their data at all, but rather just pull in the data we need.

Wikipedia

As we are working in Toronto, we will be using the same source of Postal Codes in Canada as the Week 3 assignment. This will give us the Postal Codes that begin with M, all of which are for Toronto.

> *URL:* https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Statistics Canada (StatsCan)

The following spreadsheet provides information of population count based on the postal code of communities across the country. We clean the data to only extract the postal codes that exist in Toronto.

> 'Statistics Canada. 2017. Population and dwelling counts, for Canada and forward sortation areas© as reported by the respondents, 2016 Census (table). Population and Dwelling Count Highlight Tables. 2016 Census.''
>
> *URL*: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A

This data will be used to show whether there is an adequate population in a neighborhood in order to support the restaurant. We would ideally like to find the highest density neighborhood of Toronto, however this must also be cross-referenced with data from FourSquare (see below) which will show where existing Japanese restaurants are.

The assumption we have here is that a neighborhood that already has a high-density of similar restaurants as this shows that there is likely demand for these restaurants. It is also said that this competition will help the entire market thrive, so we will use this business rule as a basis for the hypothesis as well.

We are also looking at neighbourhood populations to ensure that there is a good balance of higher populations that would presumably go to these restaurants. Thus, we are making the assumption that the people who travel to these restaurants are likely within the same neighbourhood, or likely in a neighbouring one in which the population won't vary widely.

FourSquare

We will use the FourSquare Developer API in order to pull a wide range of data. We will use the free developer account to pull requests for:

> Neighborhood Names
> Neighborhood Latitude
> Neighborhood Longitude
>
> Unique Venue Categories
> *This is a list of all the different types of venues that FourSquare has data for. We then narrow down the data to include only Japanese and Japanese-style restaurants—Ramen Restaurant, Japanese Restaurant, Sushi Restaurant, Noodle House, Sake Bar.*

This data is crucial in determining the number of restaurants in each neighborhood, which we can then cross-reference to the population data from the StatsCan information above. From here, we are able to determine the best location based on existing restaurants (using the same assumptions noted in the StatsCan source) and population density.

METHODOLOGY

K Means Clustering allows us to assign a cluster value based on feature similarity (in this case, restaurant similarity) which will help us on the path to determine the best geographic location for the restaurant.

The data was then merged, to create a single data frame with all of the information we have collected, cleaned, and interpreted so far.

Utilizing Silhouette Analysis, we determined the best number of Clusters to use,

	PostalCode	Population2016	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	...	Noodle House	Ramen Restaurant	S B
0	M1S	37769.0	Scarborough	Agincourt	43.794200	-79.262029	Sushi Restaurant	Noodle House	Japanese Restaurant	Sake Bar	...	0.333333	0.000000	0
1	M8W	20674.0	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484	Japanese Restaurant	Sushi Restaurant	Sake Bar	Ramen Restaurant	...	0.000000	0.000000	0
2	M3H	37011.0	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	Sushi Restaurant	Japanese Restaurant	Ramen Restaurant	Sake Bar	...	0.000000	0.166667	0
3	M2K	23852.0	North York	Bayview Village	43.786947	-79.385975	Japanese Restaurant	Sushi Restaurant	Ramen Restaurant	Sake Bar	...	0.000000	0.142857	0
4	M5M	25975.0	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	Sushi Restaurant	Japanese Restaurant	Ramen Restaurant	Sake Bar	...	0.000000	0.166667	0

otherwise we would be forced to run analyses on each individual neighborhood, regardless of whether those neighborhoods are actually the best way to break down the data.

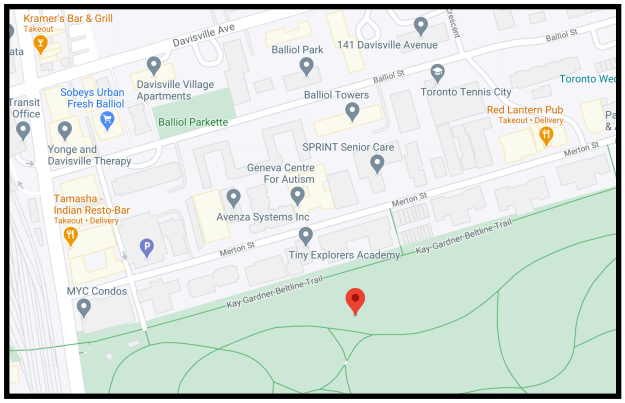
The ideal number of Clusters as determined by Silhouette Analysis in this case is 9.

From these 9 clusters, two Clusters have the same density, both with a shape of (21, 19) so they are both similarly appropriate for the restaurant. From here, we can find the actual geographic centre of each cluster, which provides more information.

Once we plotted these coordinates, it was easy to see that one location was much less ideal than another.

The geographic centre of Cluster 1 is a residential street surrounded mostly by other homes, whereas Cluster 3 is along the edge of a park surrounded by businesses and commercial spaces. Therefore, there are likely to be more available units for new restaurants, and more of the population coming here for dining and shopping. Therefore, Cluster 3's geographic centre is more suitable.

Cluster 1 - LESS IDEAL



Cluster 3 - MOST IDEAL

RESULTS

Using the tool OpenCage GeoCoder to convert the coordinates from Cluster 3 into an actual street address.

This provides us the following address:

117 Merton Street, Old Toronto, ON M4S 3G1, Canada

This address is confirmed to be directly across the street from the park from the original coordinates and is the closest physical address to that location, confirming the information is correct.

It also provides that the best neighborhood for the restaurant is:

Davisville, Toronto

Within the Notebook, we can use the Folium tool to render the map and indicate the ideal location:

DISCUSSION

Some assumptions are made in this report as noted, and the following potential factors are ignored as a result:

- availability of units that will support a restaurant
- rent and lease prices
- quality of competing restaurants
- other style of restaurants
- price point of other restaurants

The reason we have ignored these factors is because many of them (rental prices, availability, etc.) will change on a frequent enough basis to invalidate them for the purposes of this assignment. While we are presenting an address in results, we are not determining the best exact possible **actual** address, as rental/lease availability, price, and dozens of other factors mean that this changes on a too frequent basis to be valid or relevant for these purposes.

Other factors (price point of other restaurants) can be largely mitigated by the nature of many Japanese restaurants having a variety of quick meals (ramen, for example, is traditionally an under-20 minute lunch meal) and sushi, which may be a longer sit-down experience, or a grab-and-go lunch, especially in the pandemic world of 2020 and 2021.

While a full actual analysis would be required for an actual business, this presentation provides a solid foundation for which neighbourhood to drill down into and look further into for additional information.

CONCLUSION

As one might expect, the greatest concentration of restaurants was determined to be Central and Downtown Toronto. These are also, unsurprisingly, the most densely populated areas of the city, due to the abundance of high-rise buildings and walkable areas with plenty of restaurants, shopping, and nightlife.

By focusing on the Davisville area where there is a relatively high number of Japanese restaurants, the competition will continue to encourage others to keep their quality high, and therefore keep discerning customers coming back for another quality Japanese retaurant.

Opening the new restaurant in Davisville is sure to make the most sense based on the research shown above, backed up with proper data analysis.