

Finding Target Areas in Toronto for Hot New Restaurants

Using Foursquare Data to Find General Areas of Toronto Missing Restaurant Types

David Gillman

April 9, 2019

Associated Jupyter Notebook with Python Code

<https://github.com/davidgillman/IBMCapstone/blob/master/CapstoneWeek5.ipynb>

Table of Contents

Introduction	3
Data.....	4
Methodology.....	5
Center/Centroid Selection	5
Map of Toronto with Centers	6
Algorithms Used.....	6
Results.....	7
Dumplings	8
Dumplings Closest.....	8
Dumpling Average.....	8
Halal	9
Halal Closest.....	9
Halal Average	9
Ethiopian	10
Ethiopian Closest.....	10
Ethiopian Average.....	10
Discussion.....	11
Recommendation.....	11
Conclusion.....	12

Introduction

As an entrepreneurial restaurant chain, Chef Creations is always looking for new concepts in different markets. As part of an expansion into the Canadian casual dining market Chef Creations (Chef) needs to position new concept restaurants in the best locations within Toronto.

This project identifies areas of Toronto that are best to open hot new concept restaurants.

Chef sees the cultural and financial similarities between New York and Toronto. As eatery trends start in one city they are likely to be picked up in the other within a year. The delay is usually due to other restaurant operators needing to observe the changes. This takes time.

Chef wants to be ahead of that crowd and open a trendy restaurant several months earlier. Using location and trending data from Foursquare for New York, we quickly spot hot trends in restaurants.

Looking to expand in Toronto, Chef needs to know the hot, trendy restaurant types in

New York and then see areas of Toronto do not have any of those types of restaurants or are underserved.

I will determine the 3 types of restaurants in New York that have the most trending indicated on Foursquare.

The solution to the problem is less exact than specific locations since factors such as amount of traffic in the area, pedestrian access, parking, frontage and specific retail location must be weighed. However, using the postal code data is too fine grained. I will use the clusters generated in Week 3 assignment to determine the best region of Toronto to investigate putting a new restaurant.

My project will guide the Chef (or other restaurant entrepreneurs) on opening new restaurants more quickly. Reducing the time to open hot new locations in targeted areas puts Chef (or other restaurant entrepreneurs) ahead of the competition.

The methodology could be useful for other types of retail establishments. Using trending data from other sources, smaller retailers going into new markets could use general location methods to focus their search efforts to different neighborhoods. Large retailers almost certainly do this now but the methodology could be useful for smaller retailers that do not yet have this experience as part of the standard operations.

Data

The project happens in two phases both using Foursquare data. The client wants to focus on restaurant trends centered on the borough of Manhattan in New York City.

The first data set will be to identify points in Manhattan that are of most interest and then pull the hot trend data from Foursquare. The hottest 3 types of restaurants are identified.

This is the first data set - the Trending API return data from New York. This trending API is not very flexible. It is not able to be fed parameters for distance from a location nor can it filter on just restaurants. The Trending API result data contains does not extend out from its specified location more than 10 blocks.

Another limitation is the Trending API returns only current data and does not have any type of historical view. Only those locations that are hot and trendy at the moment of the API request are identified.

As part of this project, I ran the API repeatedly and at different times of the day and different days of the week. I tallied the hot restaurants being returns manually. I determined there were several hot and trend restaurants in the Manhattan area but that time of the day and day of the week affected their placement in the trending API.

Using my manual tally over repeated API request, I determined 3 interesting restaurant types. Those are the one used.

From there all of those restaurants of the 3 identified types are pulled throughout Toronto. This is the second data set using the Foursquare Places API. This is a straight forward and relatively non-volatile data source.

There was a 3rd data source used from a previous project for this class. The Week 3 data that located centroids of postal codes in Toronto was used. This provided 5 set locations from which the distances of all pulled restaurants in the 2nd data source could be calculated.

Methodology

The first task was to establish locations within New York City to target. The smaller sections were used to exclude areas of the city that Chef does not feel represents the forward edge of restaurant trends. Two locations in lower Manhattan were used.

An initial issue observed was the trending data is volatile. It changes throughout the day so no definitive list could be assembled from any single pull of data from the Foursquare API. As a result, the Trending data was pulled approximately 10 times during various parts of the day. The most common types of restaurants were recorded.

The restaurant types were edited manually to exclude very common restaurant types. Specifically, Delis and Food Trucks were removed from the list.

The resulting 3 types of restaurants that are hot and trendy in New York used were Dumplings, Halal, and Ethiopian.

Those 3 restaurant types were then used to pull data for all matching restaurant types in Toronto. Those results were kept in 3 separate data frames. The total count of each type is below.

Restaurant Type	Count
Dumplings	38
Halal	51
Ethiopian	38

Each of the restaurants summarized in the table above came from Foursquare with latitude and longitude. Using that the closest and average distance to each centroid is calculated

Center/Centroid Selection

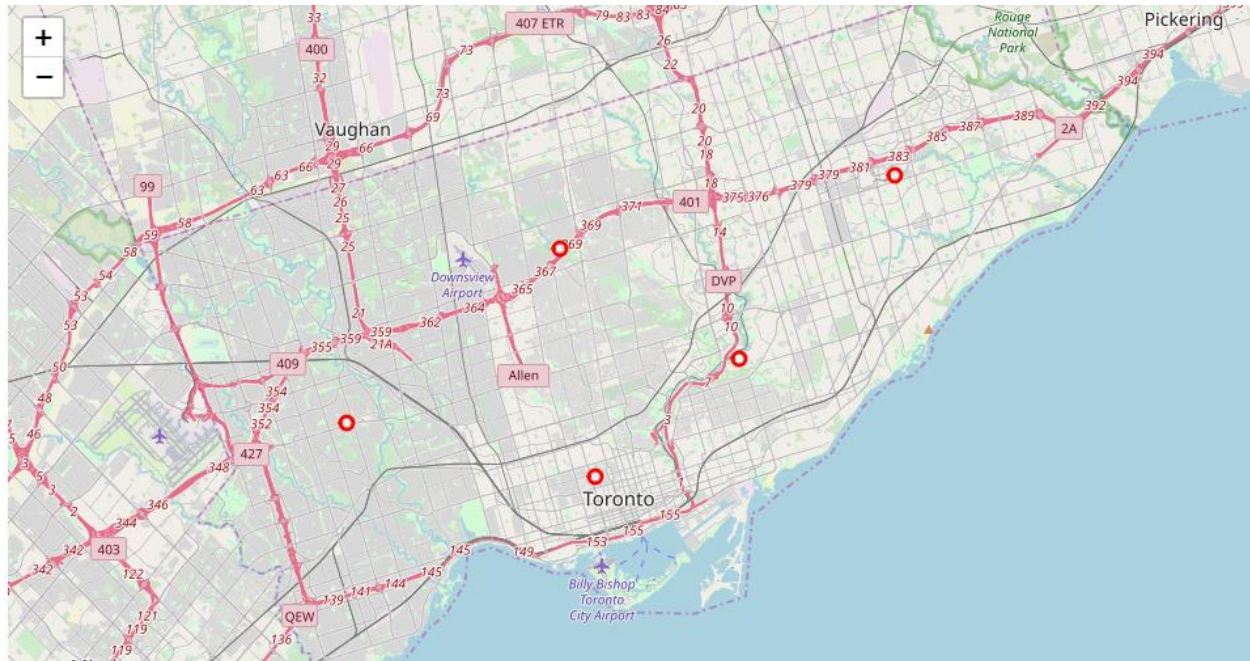
For reference, these are the latitude and longitude locations for each centroid used as a location center.

	Latitude	Longitude
0	43.653860	-79.398558
1	43.665725	-79.351178
2	43.689694	-79.394773
3	43.664518	-79.414128
4	43.653519	-79.395844

After this table, the locations are referred to in this report as Centroid1 through Centroid5 in the order listed above.

Below is a map with the Centroids as markers.

[Map of Toronto with Centers](#)



[Algorithms Used](#)

This base project used K-NearestNeighbor to determine the centroid locations. Five centroids were used as that is a number which lends itself to meaningful separation (meaning not too few) without burdening the human reader with too many to consider.

Likewise, other factors will need to be considered by the human interpreter. Those include very critical placement factors such retail space condition, price and availability.

After centroid determination use K-NearestNeighbor, the Euclidean distance from each centroid was calculated in a custom algorithm using the location data for every matching restaurant.

The results of this process of analytics and the progression of algorithms employed, yields simple results that a person looking to place new restaurants can interpret.

For easy interpretation, the closest matching restaurant's distance and the average of all restaurants distance is display in charts. There are easy to interpret and guide people planning locations to concentrate efforts on specific centroids. These results are only a guide to aid people. They are not definitive locations that are optimally located.

Results

The algorithms employed yields some interesting results.

The two selected metrics for inclusion in the report are closest to the centroid location and the average distance. Knowing the closest location lets the location planner understand the closest competition. Knowing the average distance is a proxy for the density of matching results so that a location planner finds a portion of the city that has the least access to matching restaurant types.

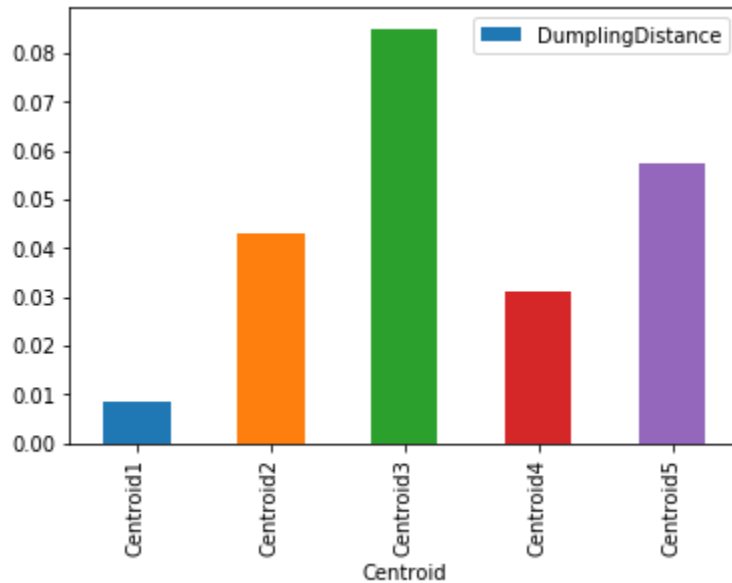
Each restaurant type is independent of the others, so each is graphed relative to the centroid locations individually.

The interesting finding is two of the restaurant types, Dumplings and Ethiopian, both show that Centroid3 is a good area for both types. On the other hand, analysis shows Halal type restaurants are already distributed throughout the Toronto area.

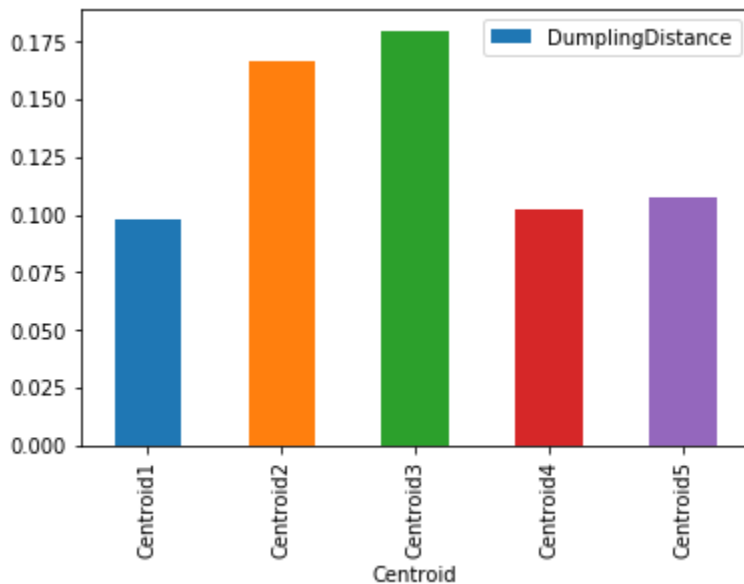
Dumplings

The Dumpling restaurant type had 38 matching locations in Toronto. As can be seen from the charts below, Centroid3 has the largest distance to the closest Dumpling restaurant and the largest average distance. This indicates that among the centroid locations, the Centroid3 area is the best candidate for a Dumpling restaurant concept.

Dumplings Closest



Dumpling Average

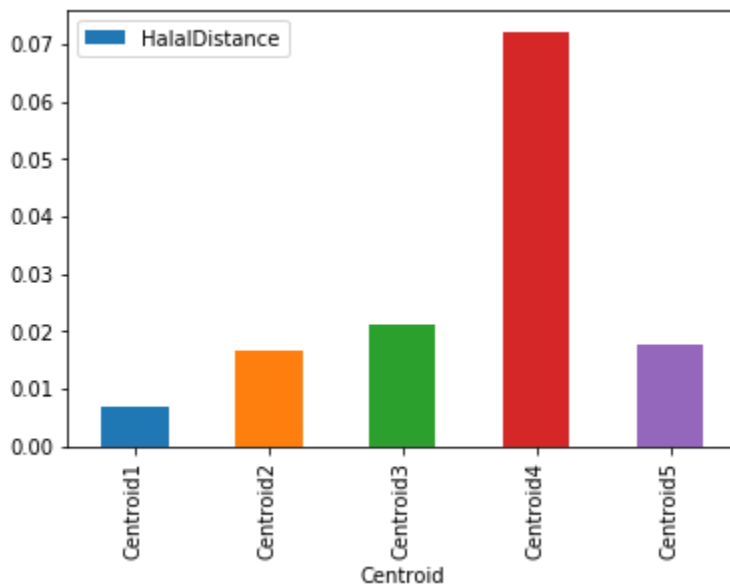


Halal

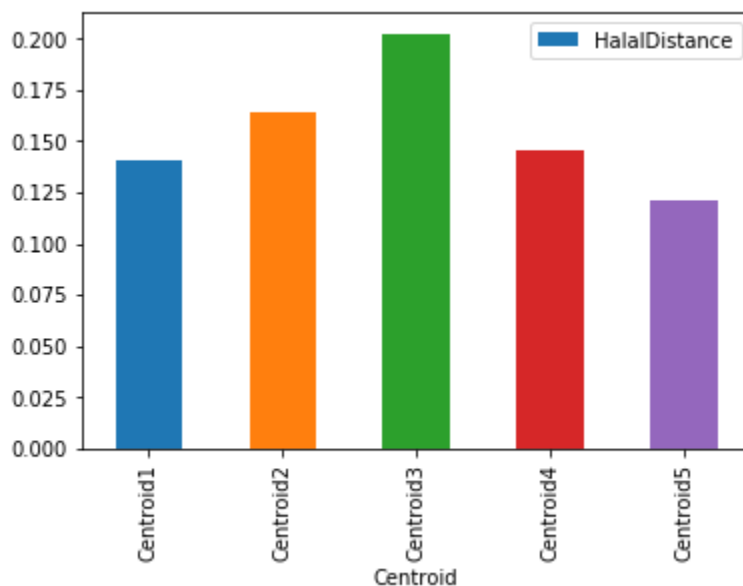
Halal restaurant location recommendations are less clear cut. Centroid4 has, by far, the furthest distance to the closest Halal restaurant. However, Centroid 3 has the largest average distance.

Given that the average distance extremes for all centroid were with a factor of 2, it may be that Halal restaurants are fairly evenly distributed around Toronto. While Centroid4 is a good location to examine, Toronto may already have a substantial number of distributed Halal restaurants with no only one specific area being particularly good for a new restaurant.

Halal Closest



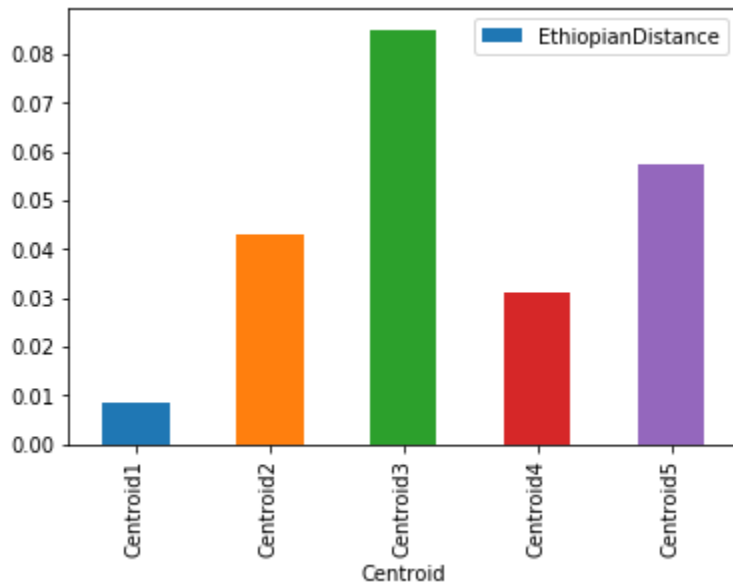
Halal Average



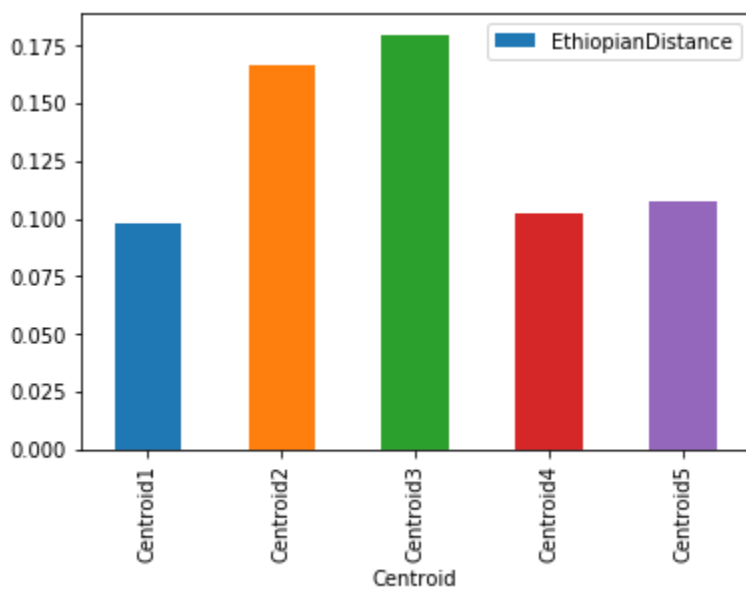
Ethiopian

Ethiopian restaurants look to have a similar pattern of distance as the Dumpling restaurants. Centroid3 is the location with the furthers distance with the nearest Ethiopian restaurant and also the centroid with the greatest average distance.

Ethiopian Closest



Ethiopian Average



Discussion

Each restaurant type needs to be considered individually since the concepts, kitchen and marketing would all need to be independent.

Dumpling restaurants in the area of Centroid3 would be better positioned to be different or unique among local dining experiences.

Likewise, Ethiopian restaurants would be similarly position new Centroid3.

Halal restaurants are likely a less new or different option in the Toronto area. They appear already well represented in Toronto and not a good option to establish a new location.

The recommendation from this report is for human location specialists to concentrate first efforts on Centroid3 to locate either Dumpling or Ethiopian new restaurants.

Recommendation

Chef Creations should focus their new restaurant opening efforts on Dumpling or Ethiopian restaurants near Centroid3.

Since the area around Centroid3 is good for both, Chef Creations maybe be able to open both types more easily.

Halal type restaurants already appear well distributed around Toronto so that type, despite being trendy in New York, are not the better option to open a new location.

Conclusion

The methodology used in the report is basic and easily understood. Given a more specific task for a specific client a better predictive model is possible. The client would be able to give better parameters that would guide both data acquisition and the analysis.

For a general guide to locations in Toronto, this report shows the theoretical client an area to look towards to locate two new concept restaurants. This was the goal of the project.

The analytical concept is generalizable to other types of retail as well. Using one or a few metro areas as the basis, companies could use the same methodology to target general areas of smaller metro areas to find matching characteristics. For that level of detail, the client would need to give more guidance which would be used as parameters in data selection.

