

# **Capstone Final Project**

## ***Choosing the Best Store Category***

***By: Tony Kim***

### **1. Introduction/ Business Problem**

The city of Toronto is always growing, implementing changes, and finding new ways to attract new people. There are tons of different varieties of stores from big malls to diverse ethnic dining. However, considering how diverse the type of stores is already in Toronto, how should a person hoping to open a store in Toronto approach this? If they want to open a restaurant, will they have too many competitions if its an Italian restaurant? What about Chinese restaurant? This is where Foursquare API will extremely useful since I can obtain information about the trending stores within Toronto. Since Toronto is quite large, I can break up Toronto into small clusters to analyze within a cluster what types of stores are more popular against others. There are two ways of analyzing this problem. I can obtain more general analysis to see which type of stores are the most popular within each cluster and see if a type sticks out above others. Or I can examine a single cluster very closely to analyze the trend within the cluster and see what type of store can fit the trend better.

### **2. Methodology**

For the data, I used the Toronto locations data from

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) which is the same dataset provided in the previous week. I will also be using the Longitude and Latitude dataset([http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)) provided by the course Moderators. By combining these two datasets, I utilized the Foursquare API to find the most trending stores and divide the city of Toronto in different clusters.

### **3. Results/Discussion**

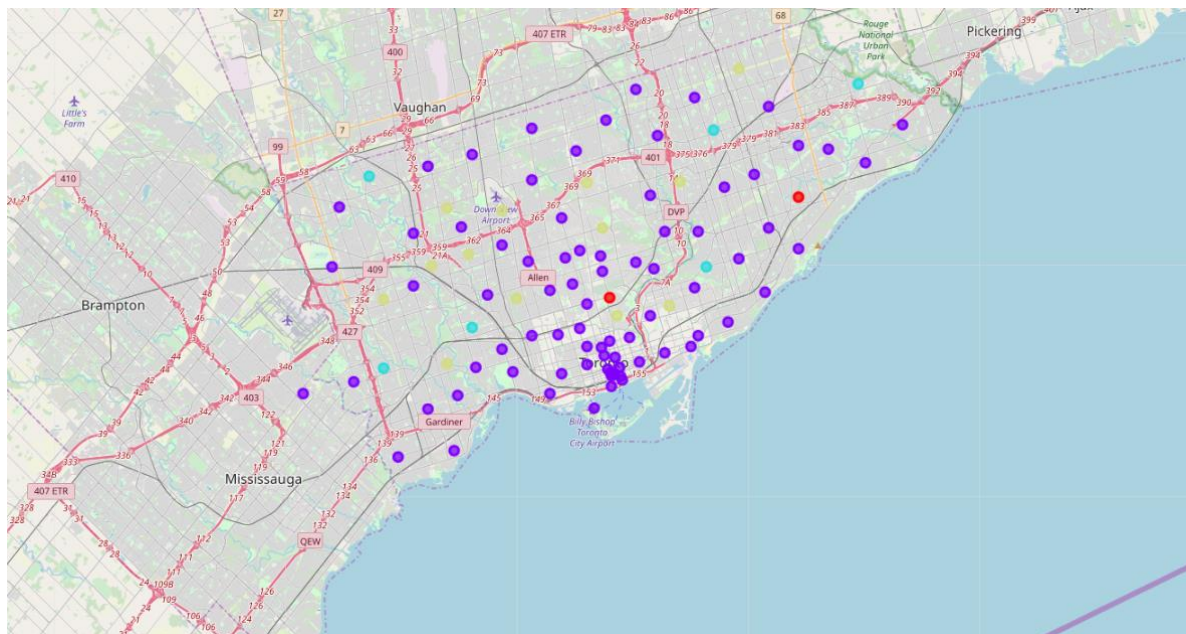
Coffee Shop	185
Café	98
Restaurant	63
Pizza Place	56
Italian Restaurant	55
Park	50
Bakery	49
Hotel	43
Bar	43
Sandwich Place	40

By counting the frequency of each Category venues, I found that Coffee Shop and Café are the most dominant. To do a further analysis on these findings, I used a Wordcloud to visualize the data.



From this figure, I can safely say the most common venue is related to Coffee where overall, the most common venues seem to be related to food in some way.

To do further analysis, I used Kmeans cluster method to divide neighbourhoods with similar characteristics into 4 different clustered and plotted them on a map using Folium.



The current findings seem to be quite vague to derive a conclusion out of it so I decided to confine my research to the top most common venue instead of top 5.

I followed the same formula to see how it differs from my initial findings.

Coffee Shop	19
Pizza Place	12
Café	5
Breakfast Spot	4
Discount Store	3
Park	3
Indian Restaurant	2
Furniture / Home Store	2
Japanese Restaurant	2
Fast Food Restaurant	2

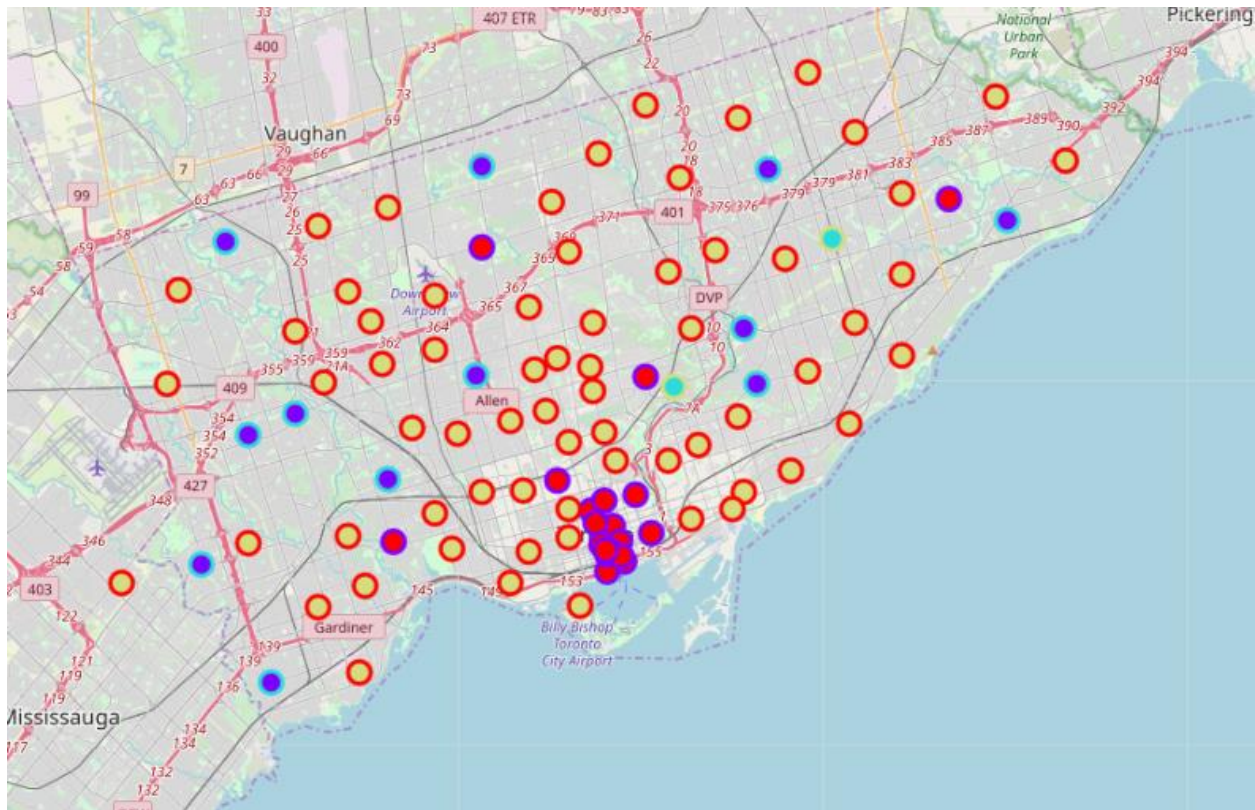
Just like before, Coffee Shop maintained the top position. However, Pizza Place took second position this time. Unlike the first dataset, coffee is not the most dominant where the frequency of Pizza Place is not far off from Coffee Shop. From this, I did further research and visualized the data once again using Wordcloud.



Coffee and Pizza are prevalent, but unlike the first Wordcloud, food categories do not seem to be as prevalent here.



Lets do a final analysis by plotting the Kmeans cluster of our new dataset on the Toronto map using Folium.



From the map, the colors can be defined as follows.

**Yellow = Cluster 0**

**Red = Cluster 1**

**Purple = Cluster 2**

**Teal = Cluster 3**

Considering I only used the top most common venue, it is not surprising that all Coffee Shops and Pizza Shops got sent to each respective cluster.

It is important to note that this figure is based on the *Most Common Venue*. This means that even if all the Coffee Shops are centered in one area, it does not mean Coffee Shops do not exist in other areas. It simply means that Coffee Shops are most common within that specific cluster. However, that made the observation a lot more interesting. Lets look at the cluster containing Coffee Shops as its value (red). All the Coffee Shops being clustered in the heart of Toronto gives us an understanding that Coffee Shops in that area are popular in

its availability, but there may be an abundance of Coffee Shops confined in a specific area. This means that if I were to open a store in that specific area, opening another Coffee Shops may not be the most appropriate since there will be too much competition.

Now, lets examine the cluster containing the value Pizza Place (purple). It seems to be fairly scattered around Toronto, but non-existent in the heart of Toronto. This means that, while there may be still Pizza Stores in that area, it is not the most common venue. Since the Pizza Places are quite popular, it may be beneficial to open a Pizza store in the center of Toronto.

Finally, let's examine the final cluster (Teal) with the value Indian Restaurant. There are two points on the map that indicate Indian Restaurant as the most common venue. Considering the venue Indian Restaurant is quite specific, we could open stores that exhibits Indian culture to compliment the Indian restaurants within those two neighborhoods.

## **Conclusion**

From the results, there may be an abundance of Coffee Stores at the heart of Toronto meaning opening another Coffee Store has a higher chance of failure. Further, some areas are specific to a certain ethnic culture and may benefit from opening stores that compliment that specific culture, which in our case was the Indian culture. There are many limitations to these findings. One is the use of free account for Foursquare API where I could not do an in-depth research due to the limited calls I could make. Further, my results are based on the most common venues around Toronto postal codes. I do not have data of how well each venue does in terms of revenue and profit.