Coursera Capstone

Christopher Kalanderopoulos

Opening a new Japanese Restaurant in Toronto

Introduction

We would like to determine the best location for a new Japanese Restaurant in Toronto.

The 'best location' will be defined as a location that:

- has other similar restaurants nearby
  (the assumption is made that this indicates demand)

- has a high population density
  (the assumption is made that this indicates potential customers)

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

Data

List of Metropolitan Toronto postal codes

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Population data sourced from Statistics Canada (StatsCan)

https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A

Cleaned this data to only include Toronto postal codes as the original data set includes all of Canada

FourSquare

Utilized FourSquare Developer API in order to pull neighborhood population data and cross-reference with postal codes and population data (above)

Methodology

K Means Clustering

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.

K Means Clustering

Two Clusters actually 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.

```
# FIND GEOGRAPHIC CENTRE OF EACH CLUSTER

cluster1coords = cluster1[['Latitude', 'Longitude']]
cluster1coords = list(cluster1coords.values)
lat = []
long = []

for l in cluster1coords:
  lat.append(l[0])
  long.append(l[1])

Blatitude = sum(lat)/len(lat)
Blongitude = sum(long)/len(long)
print(Blatitude)
print(Blongitude)
```

```
43.69177028571429
-79.41449641904762
```

Cluster 1

```
cluster3coords = cluster3[['Latitude', 'Longitude']]
cluster3coords = list(cluster3coords.values)
lat = []
long = []

for l in cluster3coords:
  lat.append(l[0])
  long.append(l[1])

blatitude = sum(lat)/len(lat)
blongitude = sum(long)/len(long)
print(Blatitude)
print(Blongitude)
```

```
43.695958680000004
-79.39204819333334
```

Cluster 3

Merging the Data

We must first merge the data to create a single data frame with all of our data - Neighborhood, Most Common Venues, Cluster, Population, and their Latitude and Longitude. This also confirms that the shape of our data is the same before performing further analysis.

	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.

Only the first few rows shown for reference.

Silhouette Analysis

Utilizing Silhouette Analysis, we can determine the best number of Clusters to use, 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.

From Scikit-learn: “Silhouette analysis can be used to study the separation distance between the resulting clusters.”

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

```
# CLUSTER MODELLING
# USE SILHOUETTE TO FIND BEST CLUSTER GROUPS

groupedclusters = grouped.drop('Neighborhood', 1)

kclusters = np.arange(2,10)
results = {}
for size in kclusters:
    model = KMeans(n_clusters = size).fit(groupedclusters)
    predictions = model.predict(groupedclusters)
    results[size] = silhouette_score(groupedclusters, predictions)

best_size = max(results, key=results.get)
best_size
```

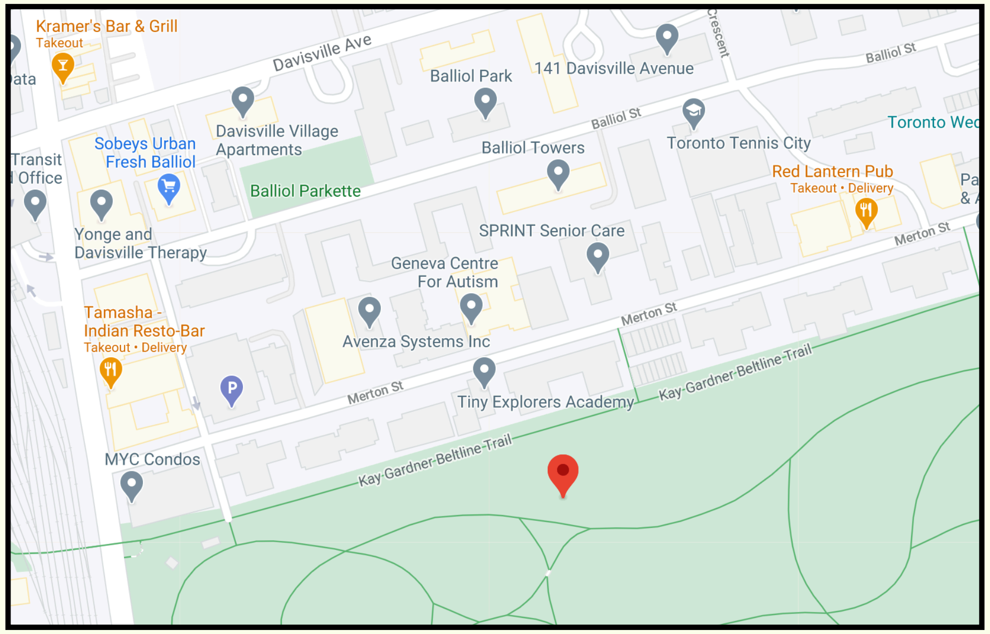
```
9
```

Source: https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

Plotting Coordinates

Plotting the coordinates through a mapping service such as Google Maps, we see that 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
Mostly Residential
LESS SUITABLE



Cluster 3
Shops, Grocery, Restaurants
MOST SUITABLE

Results

We can use the tool OpenCage 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**

Results

We can use the tool Folium to display a map with a circle to 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 others (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 you 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 are already 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.