Opening a Japanese Restaurant in Toronto

Miguel Silva – Quality Engineer

Coursera Capstone Project

Introduction

This Project aims to determine the best location to open a Japanese restaurant in the city of Toronto.

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

- Similar restaurants nearby (assuming that this indicates high demand)
- High population density (assuming that has lots of potential customers)

<u>Audience</u>

Investors looking to open a restaurant in Toronto, as it shows relevant insights from data that might be crucial to decide the best neighbourhood to establish a new restaurant.

Data

This project uses data from three data sources: Wikipedia, Statistics Canada (StatsCan) and Foursquare.

1. Toronto's neighbourhoods and boroughs along with their postal codes: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

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

3. Geographical Location Data using Foursquare's API

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 to determine the best geographic location to open a Japanese restaurant.

<u>Determining the K Cluster Number using Silhouette Analysis</u>

It was used the Silhouette Analysis to determine the best number of Clusters to use and the ideal number obtained was 9.

```
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
```

K-Means Clustering

Clustering Results:

It was observed that the clusters 4 and 6 had the maximum density, having both a shape dimension of (15, 16), being theoretically both similarly appropriate to open a restaurant. The following code was performed to show the centre locations of this clusters.

```
cluster4coords = cluster4[['Latitude', 'Longitude']]
cluster4coords = list(cluster4coords.values)
lat = []
long = []

for l in cluster4coords:
  lat.append(l[0])
  long.append(l[1])

blatitude4 = sum(lat)/len(lat)
blongitude4 = sum(long)/len(long)
print(blatitude4)
print(blongitude4)
```

```
43.710583493333345
-79.39145293333334
```

Cluster 4

```
cluster6coords = cluster6[['Latitude', 'Longitude']]
cluster6coords = list(cluster6coords.values)
lat = []
long = []

for l in cluster6coords:
  lat.append(l[0])
  long.append(l[1])

blatitude6 = sum(lat)/len(lat)
blongitude6 = sum(long)/len(long)
print(blatitude6)
print(blongitude6)
```

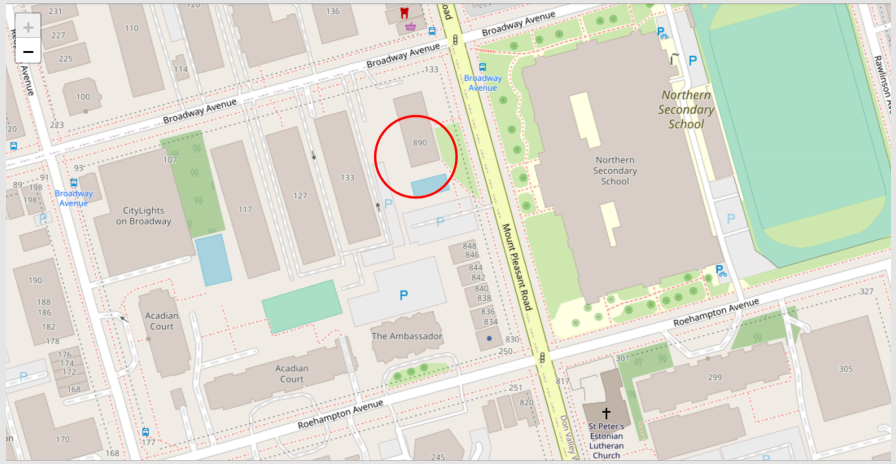
```
43.70664769333334
-79.38809083333332
```

Cluster 6

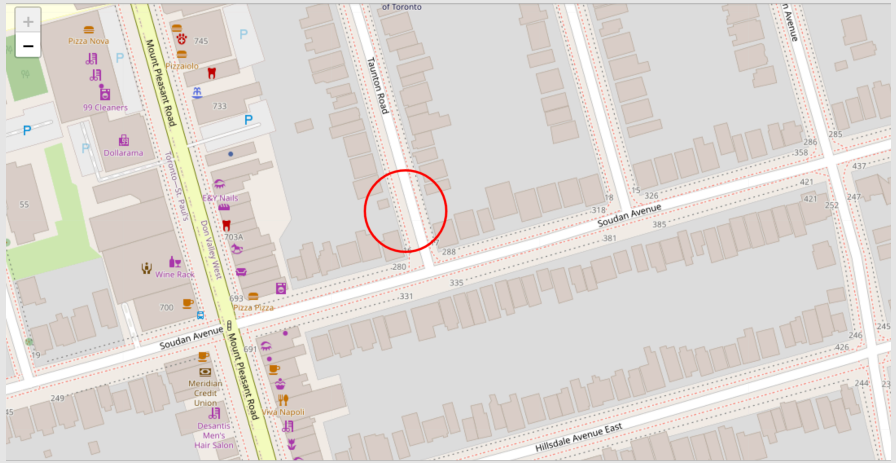
K-Means Clustering

Plotting Results:

In order to understand the exact locations of both clusters, it was used folium in order to plot the maps.



Location of Cluster 4



Location of Cluster 6

K-Means Clustering

Discussion of Results:

By analysing the plotted maps, it is possible to see that the cluster 4 is located in a position a bit far away from other restaurants. On the other hand, the cluster 6 is closer to other restaurants. So we assume that the ideal location for opening a Japanese restaurant is in the location of cluster 6.

Finding the name of this location using Geocode API

```
# Finding the best location on cluster 6 using opencage (input coordinates of cluster 6)
key = 'a3c042ed8fe641e4abe8a2ae595e08a7'
geocoder = OpenCageGeocode(key)
results6 = geocoder.reverse_geocode(43.70664769333334, -79.38809083333332)
pprint(results6)
```

```
  'formatted': '22 Taunton Road, Old Toronto, ON M4S 2P2, Canada',
  'geometry': {'lat': 43.7066588, 'lng': -79.388155}}]
```

```
# Best location name on cluster 6
popstr6 = df3[df3['Postal Code'].str.contains('M4S')]

def str_join(*args):
    return ''.join(map(str, args))

popstr6 = str_join('Best Location: ', popstr6['Neighborhood'].values, ' in ', popstr6['Borough'].values)

print(popstr6)
```

```
Best Location: ['Davisville, Davisville'] in ['Central Toronto']
```

Best Location

Conclusion

As it was expected, the ideal location to open a Japanese restaurant is in the city centre of Toronto. Based on this model and on the assumptions taken, we assume that the ideal location for opening a Japanese restaurant is in Davisville in Central Toronto.