The Battle of The Neighborhoods: Providing food to neighbourhood

Joel Tan

1. Introduction

Due to COVID-19, restriction of patrons to dine in restaurants has resulted in a dramatic loss of income for restaurant owners. One of the ways of relieving this negative impact was to engage in food delivery service. However, the cost of food delivery and duration of delivery affecting the quality of the food drove the revenue down once again. A potential solution is to move the business closer to homes, where they may scale down their business and focus on takeaway orders. It could be a preorder takeaway drive-through business model. The limitation to this is to find a neighborhood with little competition therefore increasing the chance of a comeback. In this report, I will be using Toronto as a city model where I analyzed food venues in each neighborhood and identify potential neighborhood to move the business to.

2. Data

The data used is split into 2 parts, the first is the coordinate of each neighborhood in Toronto. We web scrap Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) to get the postal code of each borough and neighborhood. Next, from (https://cocl.us/Geospatial_data) we get the geocoordinates of the postal code of Toronto. The second part is using foursquare API to obtain venues around each neighbourhood. By using geocoordinates obtained in part 1, we will make call to have a list of venues in each neighbourhood.

2.1 From Data to solution

By using the data, we will be able to plot Toronto based on the frequency of food venues and give suggestion of areas lacking of food establishments.

3. Methodology section

This section describe how I process the data and the tool and model utilized.

3.1 Grouping neighbourhood based on postal code.

More than one neighbourhood can exist in one postal code area. As the geocoordinate from geospatial is tagged to postal code instead of neighbourhood names, we have to merge rows under “Neighbourhood” column having the same name using code:

```
DF1= DF2.groupby(['A'],as_index=False)['B','C’].agg(lambda x: list(x))
```

This will allow us to join rows from multiple columns using values of 1 column.

Getting nearby venues from foursquare API

3.2 Getting nearby geocoordinates of venues around neighbourhood

By making calls to foursquare using getNearbyVenues:

Toronto_venues = getNearbyVenues(names=df['Neighborhood'],latitudes=df['Latitude'],longitudes=df['Longitude']

We set a radius 1000m around the neighbourhood coordinates to call for venues registered with foursquare.

3.3 Getting coordinates of food establishment

We defined the following list as venues referring to food establishment:

terms = ['Bakery', 'Restaurant', 'Burger', 'Pizza', 'Cafe', 'Coffee', 'Bar', 'Wine', 'BBQ', 'Brewery', 'Burrito']

with the code:

str.contains('|'.join(terms))]

Next, we group the neighbourhood according to the number of food establishment returned from foursquare API in each neighbourhood with the code:

groupby([])[].count().reset_index()

We can sort the dataframe by their counts via code:

sort_values(by='')

3.4 Illustrating the distribution of food establishment as heatmap

To have a relative view of number of food venues in each neighbourhood, we used heatmap from folium with the code:

HeatMap(FoodHeat, name=HeatMap, min_opacity=0.5, max_zoom=18, radius=10, blur=10, gradient=None, overlay=True, control=True, show=True).add_to(map_toronto)

3.5 Viewing the distribution of number of food establishment in each neighbourhood

We can plot a distribution boxplot to get the 25%, 50%, 75% range of the number of food establishment in each neighbourhood using code:

boxplot(column =['Venue Count'], grid = False)

describe()

3.6 We can compare the distribution of food establishment with unsupervised clustering model based on venues count

Using sklean.cluster.k_means, we size the cluster into 4, to compare it with boxplot. By getting the cluster data, we can split view their counts() and compare the data of clustering model to statistical boxplot. We compare grouping of neighbourhood by illustration it the map using folium.

3.7 To view an overall frequency of venues in each neighbourhood

Using onehot, we are able to tabulate the frequency of venues returned from foursquare calls. This will give a huge overview of the venues in each neighbourhood and if business owner can find complimentary food establishment.

4. Results

The following section present analyzed data in chart and map illustration aid in discussion later on.

4.1 Food Venues

As described in methodology, we filtered returned list of venues from foursquare by our "terms" and the list was cut from 3370 entries to 1457 entries.

In [18]:
```
FoodDF.head()
```

Out[18]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Malvern, Rouge	43.806686	-79.194353	Harvey's	43.800020	-79.198307	Restaurant
2	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.802008	-79.198080	Fast Food Restaurant
4	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
5	Malvern, Rouge	43.806686	-79.194353	Caribbean Wave	43.798558	-79.195777	Caribbean Restaurant
7	Malvern, Rouge	43.806686	-79.194353	Tim Hortons	43.802000	-79.198169	Coffee Shop

In [19]:
```
FoodDF.shape
```

Out[19]: (1457, 7)

In [20]:
```
FoodDFClean = FoodDF.groupby(['Neighborhood','Neighborhood Latitude','Neighborhood Longitude'])['Venue Catego
FoodDFClean2 = FoodDFClean.rename(columns={'Venue Category': 'Venue Count'})
FoodDFClean2.sort_values(by='Venue Count').head()
```

Out[20]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Count
44	Humberlea, Emery	43.724766	-79.532242	1
47	Islington Avenue	43.667856	-79.532242	1
52	Lawrence Park	43.728020	-79.388790	1
14	Clairville, Humberwood, Woodbine Downs, West H...	43.706748	-79.594054	1
72	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	2

Figure 1. Showing the number of food venues and the first 5 rows of dataframe showing neighbourhood with their respective venue count (refer to methodology 3.3).

4.2 Heatmap showing the spread of food venues

Using the data from Figure 4.1, we illustrate it using heatmap imposed on a folium map of Toronto.

In [23]:
```python
from folium import plugins
from folium.plugins import HeatMap

# Filter the DF for rows, then columns, then remove NaNs
FoodDF2 = FoodDF[['Venue Latitude', 'Venue Longitude']]

# List comprehension to make out list of lists
FoodHeat = [[row['Venue Latitude'],row['Venue Longitude']] for index, row in FoodDF2.iterrows()]

# Plot it on the map
HeatMap(FoodHeat, name=HeatMap, min_opacity=0.5, max_zoom=18, radius=10, blur=10, gradient=None, overlay=True

# Display the map
map_toronto
```

Out[23]:

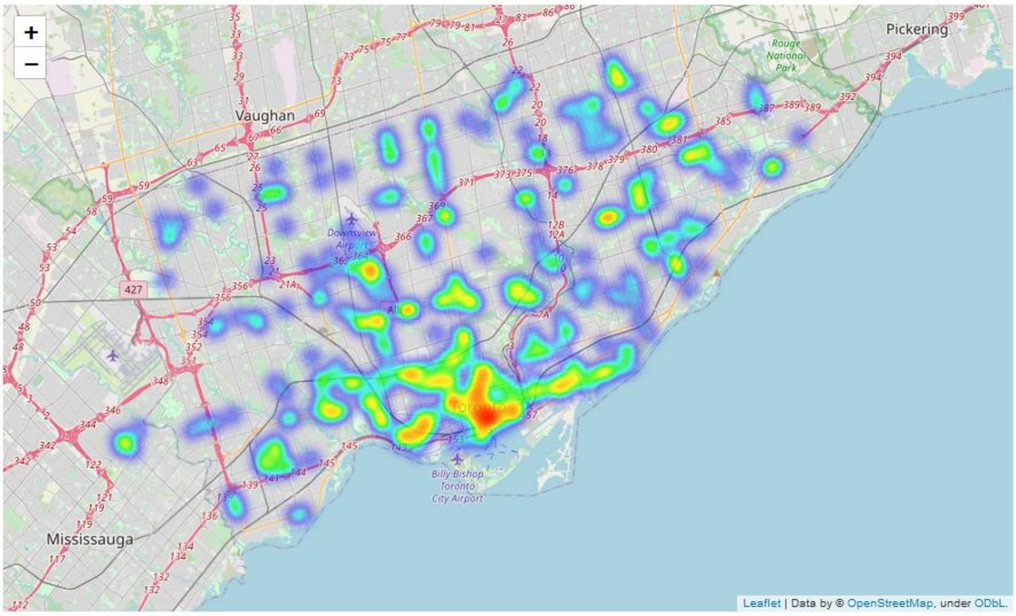
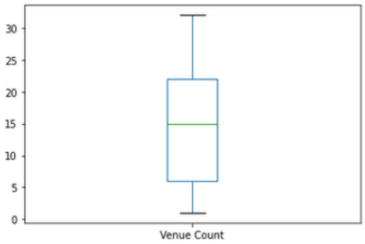


Figure 2. Heat map imposed on a folium map of Toronto

4.3 Boxplot analysis of the venue count in each neighbourhood

In [24]:
```python
VenueCountHist = FoodDFClean2.boxplot(column =['Venue Count'], grid = False)
```



In [25]:
```python
print('The neighbourhood will be splitted into 4 segment, <=25%, 25%><50%, 50><75% and >=75%')
FoodDFClean2.describe()
```

```
The neighbourhood will be splitted into 4 segment, <=25%, 25%><50%, 50><75% and >=75%
```

Out[25]:

	Neighborhood Latitude	Neighborhood Longitude	Venue Count
count	101.000000	101.000000	101.000000
mean	43.702782	-79.399271	14.425743
std	0.051021	0.096176	8.838944
min	43.602414	-79.615819	1.000000
25%	43.659526	-79.464763	6.000000
50%	43.696319	-79.389494	15.000000
75%	43.739416	-79.340923	22.000000
max	43.815252	-79.160497	32.000000

Figure 3. Boxplot of venue count per neighbourhood and their statistical describing. Note: LatLong in this table is meaningness.

4.4 KClustering model to cluster the neighbourhood

We purposely set the cluster number to 4 so as to compare it to the analyze made by boxplot.describe() function.

In [31]:
```
FoodDFGroup
```
Out[31]:

	Cluster Labels	Neighborhood	Venue Count
0	0	[Bathurst Manor, Wilson Heights, Downsview Nor...	[8, 8, 10, 11, 14, 9, 9, 13, 8, 10, 14, 8, 8, ...
1	1	[Agincourt, Brockton, Parkdale Village, Exhibi...	[23, 25, 26, 27, 25, 24, 28, 23, 25, 28, 32, 2...
2	2	[Bedford Park, Lawrence Manor East, Berczy Par...	[18, 19, 15, 20, 21, 18, 20, 18, 21, 22, 21, 2...
3	3	[Alderwood, Long Branch, Bayview Village, Birc...	[4, 4, 2, 2, 1, 5, 2, 2, 3, 3, 2, 1, 1, 4, 1, ...

In [32]:
```
FoodDFClean3['Cluster Labels'].value_counts()
```
Out[32]:
```
2    29
3    27
1    24
0    21
Name: Cluster Labels, dtype: int64
```

Figure 4. Table showing the list of neighbourhood and the venue count in each of the cluster

In [39]:
```
# create map
map_clusters = folium.Map(location=[43.651070,-79.347015],zoom_start=10)

# set color scheme for the clusters
x = np.arange(k)
ys = [i + x + (i*x)**2 for i in range(k)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, neighborhood, cluster in zip(FoodDFClean2['Neighborhood Latitude'], FoodDFClean2['Neighborhood Lon
    label = folium.Popup(str(neighborhood) + str(cluster), parse_html=True)
    popup=label,
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```
Out[39]:

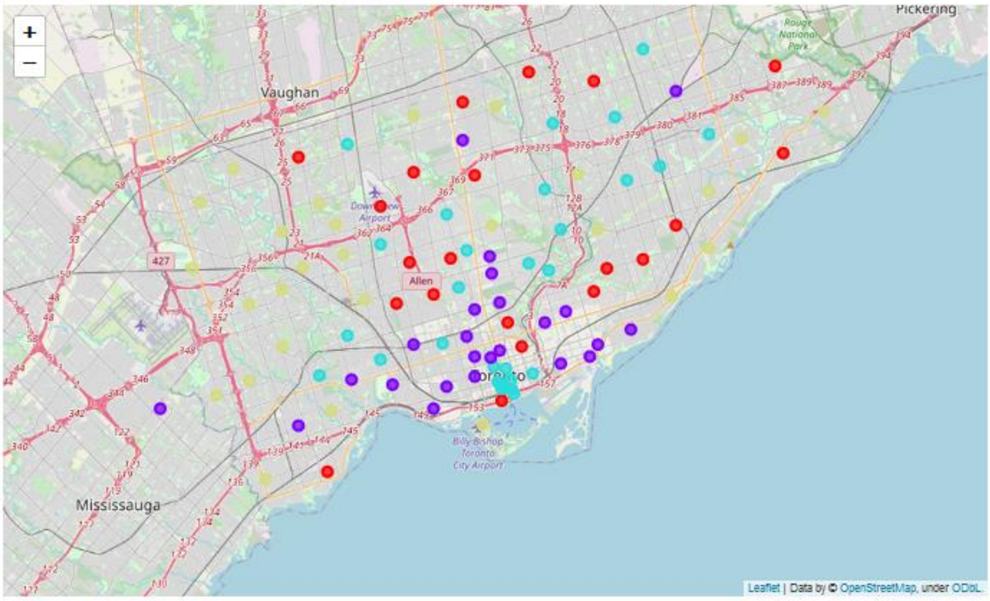


Figure 5. Imposing neighbourhood cluster on Toronto map

4.5 Neighbourhood cluster with the least number of food venues

In [64]:

```
# create map
map_clusters2 = folium.Map(location=[43.651070,-79.347015],zoom_start=10)

# set color scheme for the clusters
x = np.arange(k)
ys = [i + x + (i*x)**2 for i in range(k)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, neighborhood, cluster in zip(LowestCountN['Neighborhood Latitude'], LowestCountN['Neighborhood I
    label = folium.Popup(str(neighborhood) + str(cluster), parse_html=True)
    popup=label,
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#87cefa',
        fill_opacity=0.5).add_to(map_clusters2)

map_clusters2
```

Out[64]:

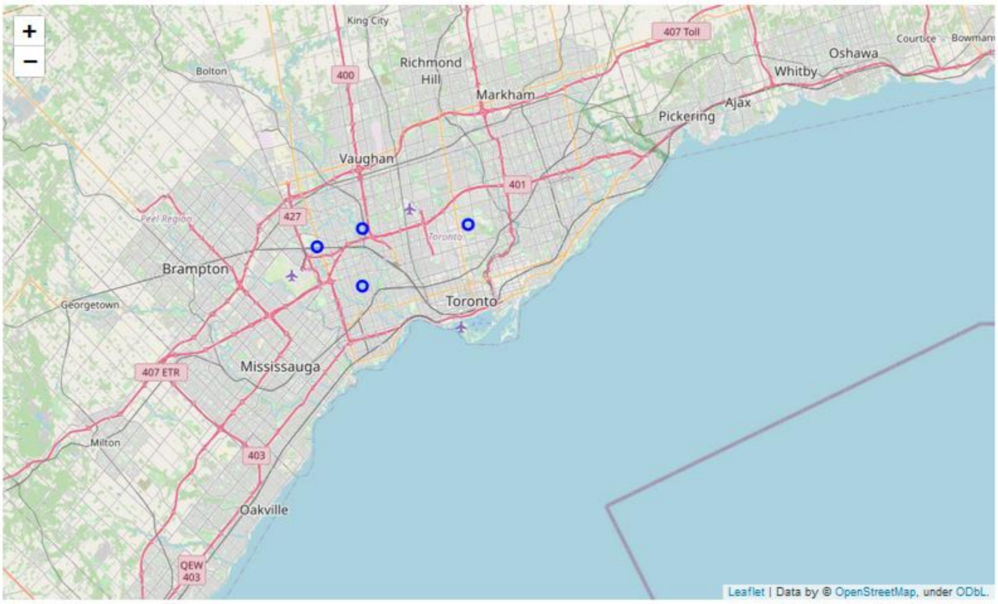


Figure 6. Imposing neighbourhood with least food venue count on Toronto map

4.6 An overall view of the venues in Toronto neighbourhood

This will give a huge overview of the frequency of venues in each neighbourhood to allow business owner to find complimentary food establishment or avoid similar food services.

Out[38]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Shopping Mall	Caribbean Restaurant	Bakery	Bank	Seafood Restaurant	Sushi Restaurant	Discount Store	Supermarket	Intersection
1	Alderwood, Long Branch	Discount Store	Park	Pizza Place	Pharmacy	Gas Station	Shopping Mall	Liquor Store	Donut Shop	Moroccan Restaurant	Sandwich Place
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Convenience Store	Park	Coffee Shop	Supermarket	Sushi Restaurant	Bridal Shop	Sandwich Place	Fried Chicken Joint	Deli / Bodega
3	Bayview Village	Bank	Japanese Restaurant	Grocery Store	Gas Station	Playground	Park	Skating Rink	Café	Intersection	Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Sandwich Place	Bank	Fast Food Restaurant	Pub	Bagel Shop	Bakery	Indian Restaurant	Intersection

5. Discussion

After cleaning the results returned from foursquare call, the percentage of food establishment in Toronto is ~43.2% (Figure 1). Using heatmap to illustrate the location of these food venues, it is observed that majority are located in downtown Toronto (Figure 2) and along Yonge Street and while resident area has the least number of food venue count. Using statistical distribution analyzed (Figure 3), there are neighborhood with only 1 food venue while the average is 15 food venues. List of neighborhood that has only 1 food venues are:

{'Clairville, Humberwood, Woodbine Downs, West Humber, Kipling Heights, Rexdale, Elms, Tandridge, Old Rexdale', 'Humberlea, Emery', 'Islington Avenue', 'Lawrence Park'}

From KClustering (Figure 5), it is also observed that the cluster around downtown Toronto and along Yonge Street with higher number of food venues are clustered together, while neighborhood that are further from downtown Toronto with lower food venues count are clustered together. Hence it is possible to look into the location these neighborhoods (Figure 6) and explore the possibility of opening a food establishment there.

6. Conclusion

Based on statistical analysis, the further the region is from downtown Toronto, the lesser the number of food venues. The number of food venues in the neighborhood can be as low as 1. Based on the venues returned from a single venues API service provider (foursquare), the list can still serve as a good indicator for future business planning to open food ventures closer to residence neighborhood.