# THE BATTLE OF NEIGHBORHOODS
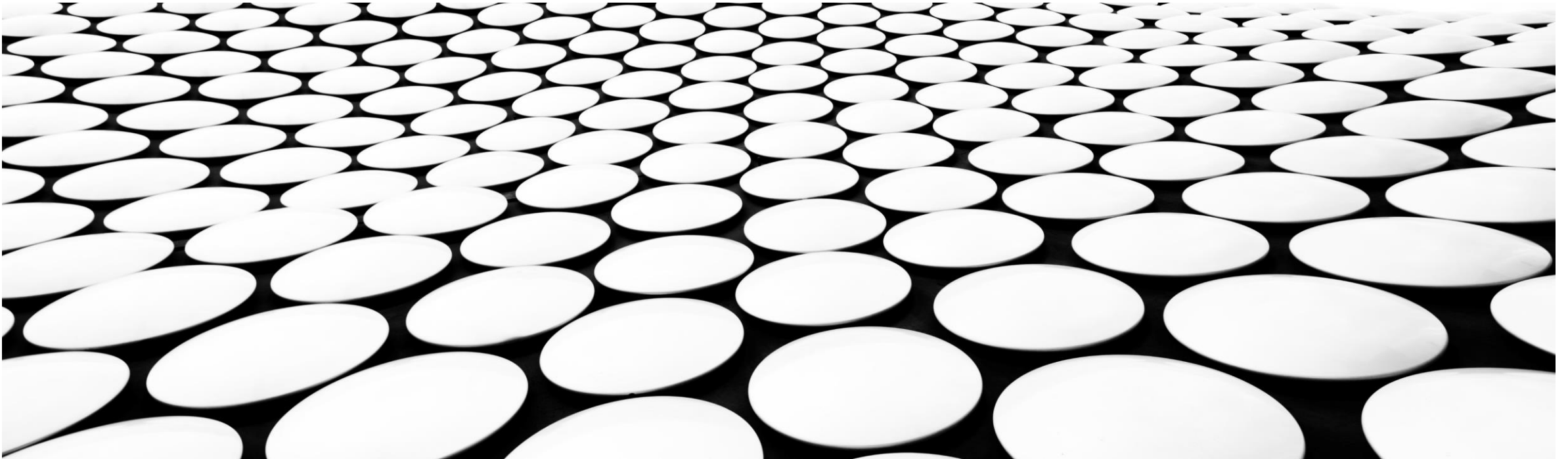
## IDENTIFY THE PERFECT PLACE IN THE CITY OF TORONTO TO OPEN A CAFÉ

Alexander. V. Larionov

April 27, 2021

# INTRODUCTION



**Background**

- According to Wikipedia, Toronto is Canada's largest city, the administrative center of Ontario.

- It has a population of 2,731,571 (2016). Toronto is part of the Golden Fory, a densely populated region around the western part of Lake Ontario with a population of about 7 million people.

- About one third of Canada's total population lives within a 500 km radius of Toronto. About a sixth of all Canadian jobs are within the city limits.

- The city of Toronto is also known as the "economic engine" of Canada, considered one of the leading metropolises in the world and has a lot of weight both in the region and at the state and international level.

- In The Economist'sannual global quality of liferanking, which measures overall quality of life, Toronto ranks fourth in the world among 140 participating cities.

# INTRODUCTION



**Problem**

- But where is the best place to open your own café?

- Using Foursquare location data and regional clustering, location information to determine which area in Toronto might be the "best" to open

- The cafe aims to predict the most appropriate place to open a new cafe in Toronto, Canada.

# INTRODUCTION



**Interest**

- Obviously, new entrepreneurs are very interested in accurately predicting the optimal location when deciding to buy, or rent a property when deciding to open a new café, to gain competitive advantages and business values.

# DATA ACQUISITION AND CLEANING



- The data that will be required will be a combination of CSV files that have been produced for analysis from multiple sources that will provide a list of areas in Toronto (via Wikipedia), the geographic location of the districts (via the Geocoder Package) and location data related to the cafe (via Foursquare).

# DATA ACQUISITION AND CLEANING



First, we need to extract data from data sources:

*Toronto Areas via Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)*

# DATA ACQUISITION AND CLEANING

| | Postal Code | Borough | Neighborhood |
|---|---|---|---|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 4 | M7A | Queen's Park | Ontario Provincial Government |
| 5 | M9A | Etobicoke | Islington Avenue |
| 6 | M1B | Scarborough | Malvern, Rouge |
| 7 | M3B | North York | Don Mills North |
| 8 | M4B | East York | Parkview Hill, Woodbine Gardens |
| 9 | M5B | Downtown Toronto | Garden District, Ryerson |
| 10 | M6B | North York | Glencairn |
| 11 | M9B | Etobicoke | West Deane Park, Princess Gardens, Martin Grov... |
| 12 | M1C | Scarborough | Rouge Hill, Port Union, Highland Creek |
| 13 | M3C | North York | Don Mills South |
| 14 | M4C | East York | Woodbine Heights |

- The Wikipedia website, as shown above, provides almost all information about neighborhoods. Including the postcode, the district and the name of the districts present in Toronto. Because the data is not in a format suitable for analysis, the data was cleaned from this site

# DATA ACQUISITION AND CLEANING

| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Malvern, Rouge | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Rouge Hill, Port Union, Highland Creek | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |
| 5 | M1J | Scarborough | Scarborough Village | 43.744734 | -79.239476 |
| 6 | M1K | Scarborough | Kennedy Park, Ionview, East Birchmount Park | 43.727929 | -79.262029 |
| 7 | M1L | Scarborough | Golden Mile, Clairlea, Oakridge | 43.711112 | -79.284577 |
| 8 | M1M | Scarborough | Cliffside, Cliffcrest, Scarborough Village West | 43.716316 | -79.239476 |
| 9 | M1N | Scarborough | Birch Cliff, Cliffside West | 43.692657 | -79.264848 |
| 10 | M1P | Scarborough | Dorset Park, Wexford Heights, Scarborough Town... | 43.757410 | -79.273304 |
| 11 | M1R | Scarborough | Wexford, Maryvale | 43.750072 | -79.295849 |

**Geographical Location data.**

Second source of data provided (https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv)

us with geographical coordinates of areas with appropriate postcodes. Because the CSV file, attaching it to the Pandas data frame was easy

## GEOGRAPHICAL LOCATION DATA.



| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Malvern, Rouge | 43.806686 | -79.194353 | Wendy's | 43.807448 | -79.199056 | Fast Food Restaurant |
| 1 | Rouge Hill, Port Union, Highland Creek | 43.784535 | -79.160497 | Royal Canadian Legion | 43.782533 | -79.163085 | Bar |
| 2 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | RBC Royal Bank | 43.766790 | -79.191151 | Bank |
| 3 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | G & G Electronics | 43.765309 | -79.191537 | Electronics Store |
| 4 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Sail Sushi | 43.765951 | -79.191275 | Restaurant |
| 5 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Big Bite Burrito | 43.766299 | -79.190720 | Mexican Restaurant |
| 6 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Enterprise Rent-A-Car | 43.764076 | -79.193406 | Rental Car Location |
| 7 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Krispy Kreme Doughnuts | 43.767169 | -79.189660 | Donut Shop |
| 8 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Woburn Medical Centre | 43.766631 | -79.192286 | Medical Center |
| 9 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Lawrence Ave E & Kingston Rd | 43.767704 | -79.189490 | Intersection |
| 10 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Eggsmart | 43.767800 | -79.190466 | Breakfast Spot |
| 11 | Woburn | 43.770992 | -79.216917 | Starbucks | 43.770037 | -79.221156 | Coffee Shop |

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | Dairy Queen | Ice Cream Shop | 43.710378 | -79.290701 |
| 1 | Warden Ave & St. Clair Ave E | Intersection | 43.712057 | -79.281005 |
| 2 | TTC Bus #68 Warden | Bus Line | 43.711778 | -79.279714 |
| 3 | Warden Station Bus Loop | Bus Station | 43.711241 | -79.279576 |
| 4 | TTC Bus 102 Markham Road | Bus Line | 43.711381 | -79.279588 |

Location data using Foursquare

The location, name and category of various sites in Toronto was collected using the Foursquare explore API.

To get the data, you needed to create an account where it provided a "Secret Key" as well as a "Customer ID" that would allow me to retrieve any data.

You can see that the neighborhoods are grouped around the neighborhood, so clustering data is further simplified.

Once all the data has been collected and placed in the data frame, the analysis process must be cleaned and merged.

After, the venue data pulled from the Foursquare API was merged with the table above providing us with the local venue within a 500-meter radius shown below.

# GEOGRAPHICAL LOCATION DATA.
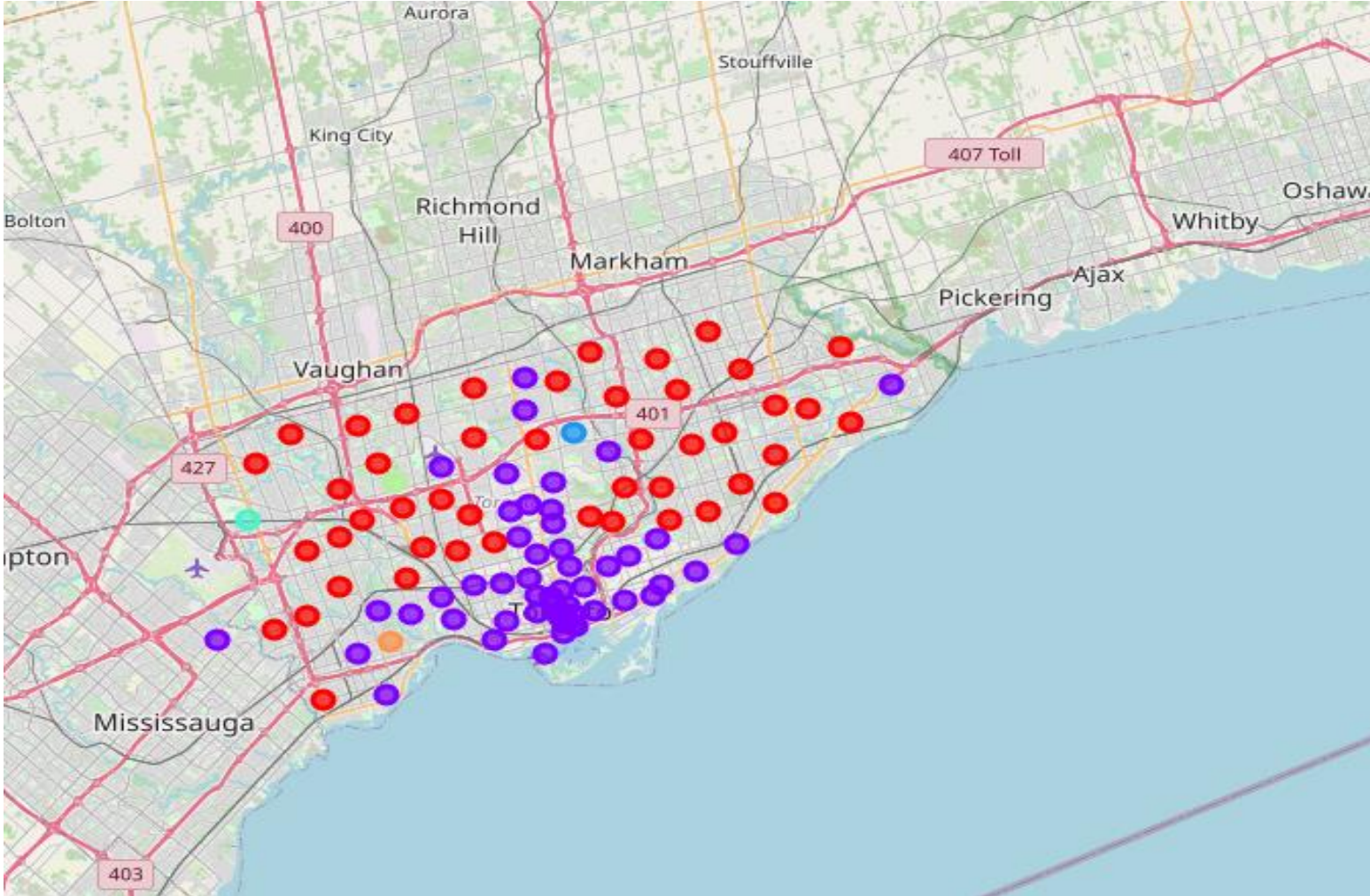


Now after cleansing the data, the next step was to analyze it.

We then created a map using folium and color coded each Neighborhood depending on what Borough it was located in.

# EXPLORATORY DATA ANALYSIS



```
to_merged['Venue Category'].value_counts()['Café']
5]: 93
```

Next, we used the Foursquare API to get a list of all the Venues in Toronto which included Parks, Schools, Café Shops, Asian Restaurants etc. Getting this data was crucial to analyzing the number of Italian Restaurants all over Toronto.

There was a total of 93 cafe in Toronto.

# EXPLORATORY DATA ANALYSIS

We then merged the Foursquare Venue data with the Neighborhood data which then gave us the nearest Venue for each of the Neighborhoods.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Malvern, Rouge | 43.806686 | -79.194353 | Wendy's | 43.807448 | -79.199056 | Fast Food Restaurant |
| 1 | Rouge Hill, Port Union, Highland Creek | 43.784535 | -79.160497 | Royal Canadian Legion | 43.782533 | -79.163085 | Bar |
| 2 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | RBC Royal Bank | 43.766790 | -79.191151 | Bank |
| 3 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | G & G Electronics | 43.765309 | -79.191537 | Electronics Store |
| 4 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Sail Sushi | 43.765951 | -79.191275 | Restaurant |
| 5 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Big Bite Burrito | 43.766299 | -79.190720 | Mexican Restaurant |
| 6 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Enterprise Rent-A-Car | 43.764076 | -79.193406 | Rental Car Location |
| 7 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Krispy Kreme Doughnuts | 43.767169 | -79.189660 | Donut Shop |
| 8 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Woburn Medical Centre | 43.766631 | -79.192286 | Medical Center |
| 9 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Lawrence Ave E & Kingston Rd | 43.767704 | -79.189490 | Intersection |
| 10 | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 | Eggsmart | 43.767800 | -79.190466 | Breakfast Spot |

Then to analyze the data we performed a technique in which Categorical Data is transformed into Numerical Data for Machine Learning algorithms. This technique is called One hot encoding. For each of the neighborhoods, individual venues were turned into the frequency at how many of those Venues were located in each neighborhood.

| | Neighborhood | Accessories Store | Adult Boutique | Airport | Airport Food Court | Airport Gate | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Antique Shop | Aquarium | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Ath & S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Malvern, Rouge | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | Rouge Hill, Port Union, Highland Creek | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 2 | Guildwood, Morningside, West Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | Guildwood, Morningside, West Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | Guildwood, Morningside, West Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

# EXPLORATORY DATA ANALYSIS

| | Neighborhood | Café |
|---|---|---|
| 0 | Agincourt | 0.000000 |
| 1 | Alderwood, Long Branch | 0.000000 |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | 0.000000 |
| 3 | Bayview Village | 0.250000 |
| 4 | Bedford Park, Lawrence Manor East | 0.045455 |
| 5 | Berczy Park | 0.016949 |
| 6 | Birch Cliff, Cliffside West | 0.250000 |

After, we created a new data frame which only stored the Neighborhood names as well as the mean frequency of Cafe in that Neighborhood. This allowed the data to be summarized based on each individual Neighborhood and made the data much simpler to analyze.

# EXPLORATORY DATA ANALYSIS

K vs Error
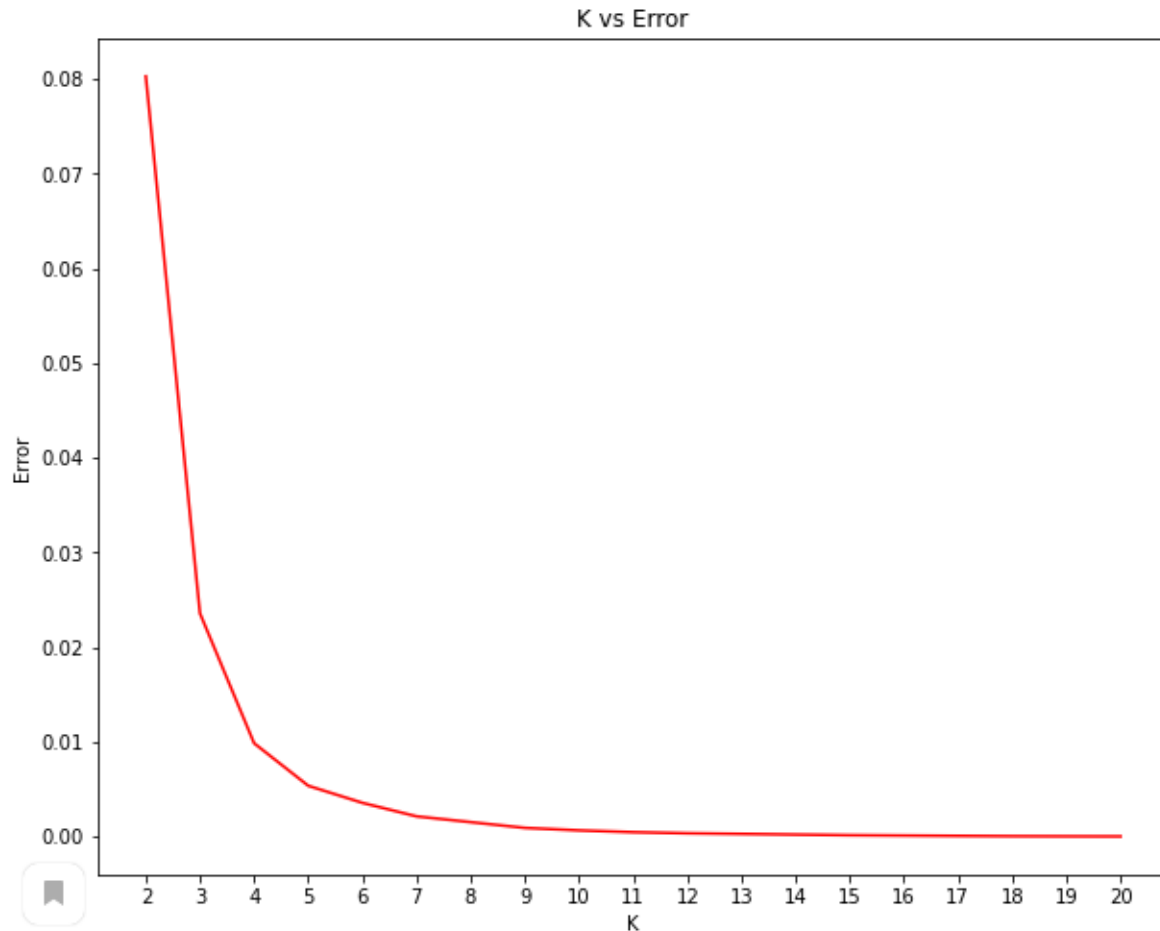


To make the analysis more interesting, we wanted to cluster the neighborhoods based on the neighborhoods that had similar averages of Café in that Neighborhood. To do this we used K-Means clustering.

To get our optimum K value that was neither overfitting or underfitting the model, we used the Elbow Point Technique. In this technique we ran a test with different number of K values and measured the accuracy and then chose the best K value. The best K value is chosen at the point in which the line has a sharpest turn.

In our case we had the Elbow Point at K = 4.

That means we will have a total of 4 clusters.

# EXPLORATORY DATA ANALYSIS

| | Neighborhood | Café | Cluster Labels |
|---|---|---|---|
| 0 | Agincourt | 0.000000 | 0 |
| 1 | Alderwood, Long Branch | 0.000000 | 0 |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | 0.000000 | 0 |
| 3 | Bayview Village | 0.250000 | 3 |
| 4 | Bedford Park, Lawrence Manor East | 0.045455 | 2 |
| 5 | Berczy Park | 0.016949 | 3 |
| 6 | Birch Cliff, Cliffside West | 0.250000 | 3 |
| 7 | Brockton, Parkdale Village, Exhibition Place | 0.136364 | 3 |
| 8 | CN Tower, King and Spadina, Railway Lands, Har... | 0.000000 | 0 |
| 9 | Caledonia-Fairbanks | 0.000000 | 0 |
| 10 | Cedarbrae | 0.000000 | 0 |
| 11 | Central Bay Street | 0.049180 | 2 |

| | Neighborhood | Café | Cluster Labels | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | 0.0 | 0 | 43.794200 | -79.262029 | Panagio's Breakfast & Lunch | 43.792370 | -79.260203 | Breakfast Spot |
| 35 | Forest Hill North & West | 0.0 | 0 | 43.696948 | -79.411307 | Kay Gardner Beltline Trail | 43.698446 | -79.406873 | Trail |
| 35 | Forest Hill North & West | 0.0 | 0 | 43.696948 | -79.411307 | Forest Hill Road Park | 43.697945 | -79.406605 | Park |
| 35 | Forest Hill North & West | 0.0 | 0 | 43.696948 | -79.411307 | Nikko Sushi Japenese Restaurant | 43.700443 | -79.407957 | Sushi Restaurant |
| 35 | Forest Hill North & West | 0.0 | 0 | 43.696948 | -79.411307 | Oliver jewelry | 43.700374 | -79.407644 | Jewelry Store |

In K-Means clustering, objects that are similar based on a certain variable are put into the same cluster. Neighborhoods that had similar mean frequency of Italian Restaurants were divided into 4 clusters. Each of these clusters were labelled from 0 to 3 as the indexing of labels begin with 0 instead of 1.
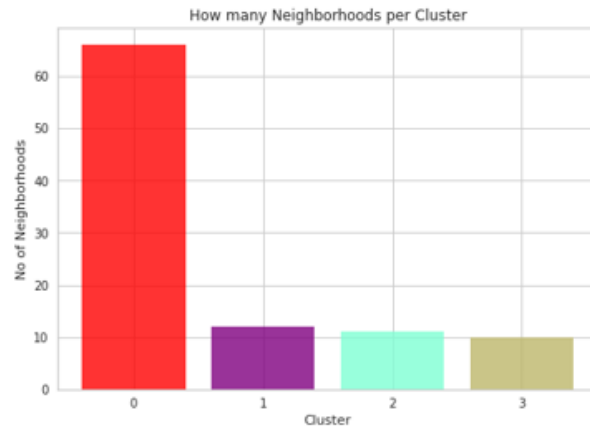
## EXPLORATORY DATA ANALYSIS
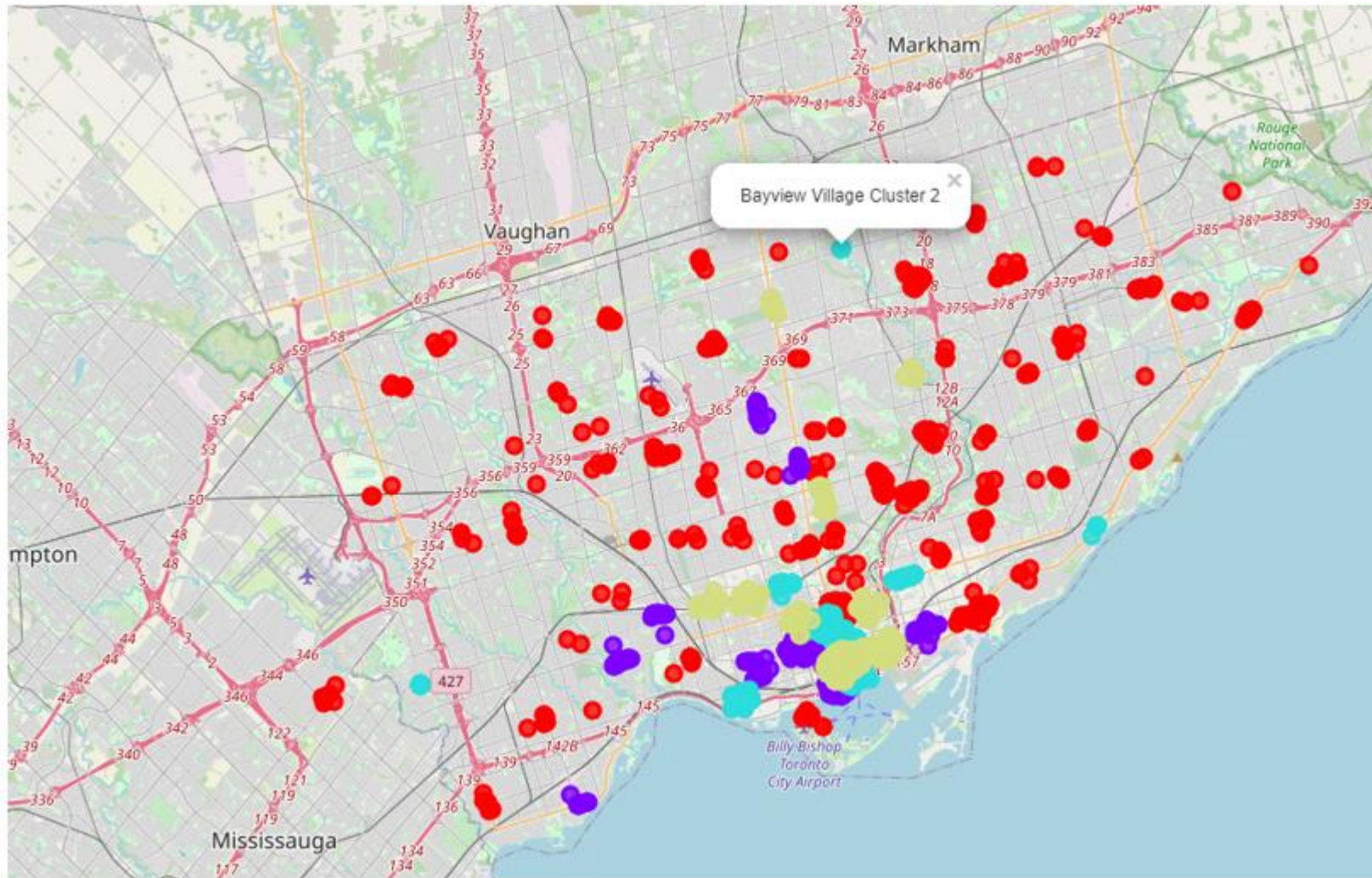


Then we created a map using the Folium package in Python and each neighborhood was colored based on the cluster label. For example, cluster 1 was purple and cluster 2 was aquamarine.

The map above shows the different clusters that had similar mean frequency of Café.

## ANALYSIS



How many Neighborhoods per Cluster

```
# How many neighborhoods in each cluster
cafe['Cluster Labels'].value_counts()

1]:  0    66
     3    12
     1    11
     2    10
```



Average number of Café per Cluster

We have a total of 4 clusters (0,1,2,3). Before we analyze them one by one lets check the total amount of neighborhoods in each cluster and the average café in that cluster.

From the bar graph that was made using Matplotlib, we can compare the number of Neighborhoods per Cluster. We see that Cluster 1 has the least neighborhoods (1)

while cluster 0 has the most (66). Cluster 3 has 12 neighborhoods, cluster 1 - 11  and cluster 2 has only 10. Then we compared the average Café per cluster.

Therefore, the ordering of the average Café in each cluster goes as follows:

1. Cluster 0 (≈0.012987)

2. Cluster 2 (≈0.031250)

3. Cluster 1 (≈0.049180)

4. Cluster 3 (≈0.063830)

## ANALYSIS

| | Borough | Neighborhood | Café | Cluster Labels |
|---|---|---|---|---|
| 500 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 512 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 514 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 515 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 516 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 517 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 518 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 519 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 520 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 521 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 522 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |
| 523 | Downtown Toronto | Church and Wellesley | 0.012987 | 0 |

| | Borough | Neighborhood | Café | Cluster Labels |
|---|---|---|---|---|
| 0 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 1 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 2 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 3 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 4 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 5 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 6 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 7 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 8 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 9 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 10 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |
| 11 | Queen's Park | Ontario Provincial Government | 0.031250 | 2 |

```
df_cluster2['Venue Category'].value_counts(ascending=False)['Café']
```
`]: 18`

| | Borough | Neighborhood | Café | Cluster Labels |
|---|---|---|---|---|
| 0 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 1 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 2 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 3 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 4 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 5 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 6 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 7 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 8 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 9 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 10 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |
| 11 | Downtown Toronto | Central Bay Street | 0.049180 | 1 |

```
df_cluster1['Venue Category'].value_counts(ascending=False)['Café']
```
`: 29`

| | Borough | Neighborhood | Café | Cluster Labels |
|---|---|---|---|---|
| 0 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 1 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 2 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 3 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 4 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 5 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 6 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 7 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 8 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 9 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 10 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |
| 11 | Downtown Toronto | Regent Park, Harbourfront | 0.063830 | 3 |

```
_cluster3['Venue Category'].value_counts(ascending=False)['Café']
```
`45`

# DISCUSSION



Most of the Café are in cluster 3, represented by "darkkhaki" clusters.

Areas located in the Downtown Toronto area, which have the highest average level of Café Regent Park, Harbourfront and St. James Town. In spite of what is in the cluster of 0 huge number of districts, café are almost non-existent. Looking To nearby establishments, the optimal place to host a new café is in the city centre Toronto, as there are many areas in the area, but virtually no café , that excludes any competition.

Having 66 districts in the area without Italian restaurants provides a good opportunity for opening a new café . Some of the drawbacks of this analysis are that clustering based entirely on data from the Foursquare API. And only the geographical  location of similar establishments  is considered.

This sums  up  the initial conclusions for this project      and  recommends  that       the entrepreneur think about  opening cafes  in these  places with little or no  competition.

# CONCLUSION



Finally, to complete this project, we had the opportunity to solve a business problem, and it was solved in a way that a real data scientist would.

We've used multiple Python libraries to extract information, manage content, and break and visualize these datasets.

We used the Foursquare API to study the settings in the Toronto area, and we got a lot of data from Wikipedia that we collected using the Beautifulsoup Web Scraper Library.

We have also visualized the use of different graphs present in seaborn and matplotlib libraries.

Similarly, we used the AI strategy to anticipate the error, taking into account the information, and used Folium to map it.

Places that are there to improve or certain shortcomings give us the opportunity to further improve this project with the help of additional information and the distinctive machine of the Learning Strategy.

In addition, we can use this venture to investigate any situation, such as opening an alternative kitchen or opening a movie theater and so on.

# CONCLUSION
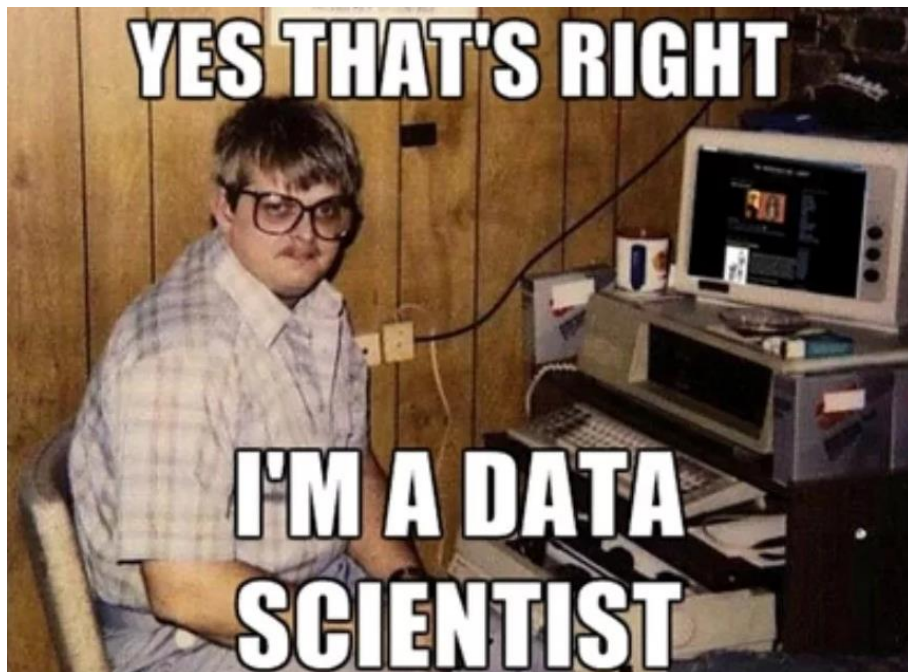


In a more serious analysis, with sufficient resources, computing power and funding, additional analysis is needed in the final decision, taking into account, among other things:

Popularity of the supposed place of opening of cafes among locals and tourists (the number of daily flow of people in this place)

Transport logistics

Cost of rent or cost of property purchase

✓ Local law (opportunity to sell alcoholic beverages)

✓ Income level of residents living in the area

✓ crime rate in the immediate vicinity,

✓ And many other factors.