**Capstone Project: The Battle of Toronto Neighbourhoods**

## **Author: Gnana S**

## **Date: Oct 2019**

# **Introduction**

This project is aimed at finding an optimal neighbourhood location for a Thai restaurant. Specifically, this report is to advice a Thai restaurant chain that is interested in opening an outlet in Toronto.

In Canada, especially in Toronto, being a food capital of Canada, Thai food is becoming very popular. Thai food has a multiethnic appeal. It is basically enjoyed by South East Asian, Chinese and Indian community which is a sizeable population in Toronto. The chefs of this chain can modify the basic curry local spices and understand the need for western tastes and modified the level of spices and hotness. This chain has several outlets in the USA running successfully and wants to start operations in Canada. In Toronto, there are several successful Thai restaurants already and the question to be answered is with regard to the choice of location to beat existing competition and yet be profitable. That is to be able to find new catchment areas with similar characteristics as that of locations corresponding to the existing competitors.

# **Data**

Key data sets that I used to proceed with solution are:

1. Neighbourhood details for Toronto. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>, will be scraped using the sing the BeautifulSoup package to get the basic neighbourhood data. This will be augmented by geographical coordinates using the data in the csv file: <http://cocl.us/Geospatial_data>
2. Qualitative web crawling to understand the Toronto neighbourhood preferences, demographics and its food industry.
3. Venue details (Name, geospatial location, category etc.) that includes Thai and non Thai restaurants. For this Foursqaure API Venues data will be used.

# **Methodology**

## Data loading and cleansing

**Toronto Neighbourhood data:**

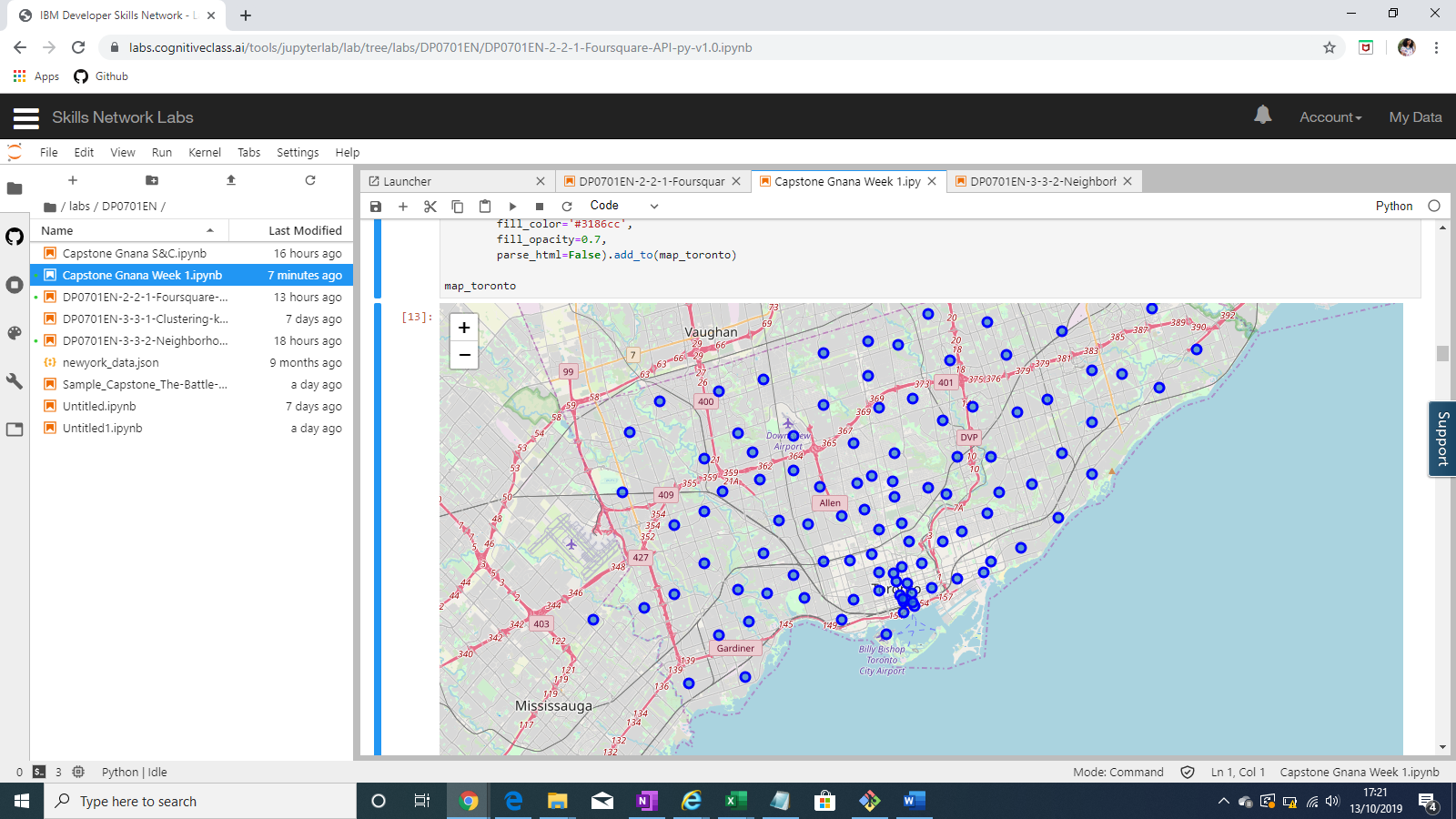
1. I loaded the data downloaded or scraped from multiple sources into pandas data frames.
2. Combined geospatial data of neighbourhoods with Canada postal code data.
3. Performed basic clean up of the neighbourhood dataset to make it suitable for running any data science analytics. Columns to which Borough was not assigned was dropped. Same Postcodes were combined into one column etc.
4. Next I filtered the Toronto neighbourhoods from the Canada data by using a keyword filter on the borough name. For further processing I only used this limited Toronto dataset.

**Venues data:**

1. I used my Foursquare credential to download the venues data for top 100 venues that are in each of Toronto neighbourhoods within a radius of 500 meters.
2. I also created a function to extract the venue category and load it into the key venues dataframe.

## Exploring data

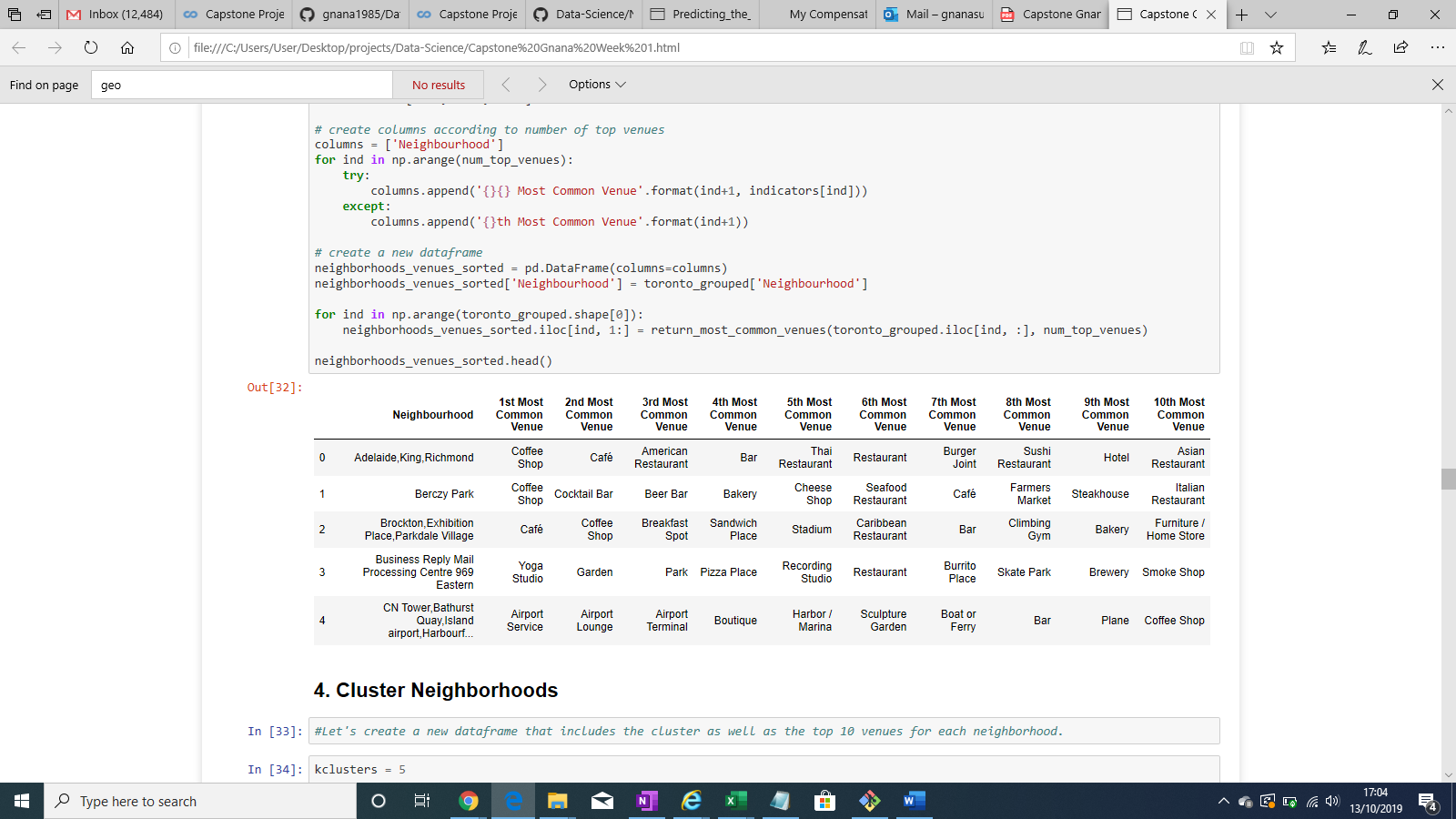
1. I used the geolocator function to find the geographical coordinate of Toronto City and used this to create a map to view all the neighbourhoods in the dataframe.



1. In the next step, I combined the venues data and neighbourhood data to create a combined view of all the Toronto venues.
2. The I did a basic group by on the above dataset to identify the busy neighbourhoods in terms of number of venues. The neighbourhood of Adelaide seems to be the one buzzing with activities.

|  |  |
| --- | --- |
| **Neighbourhood** | **Venue count** |
| Adelaide,King,Richmond | 100 |
| Berczy Park | 57 |
| Brockton,Exhibition Place,Parkdale Village | 25 |
| Business Reply Mail Processing Centre 969 Eastern | 17 |
| CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara | 17 |

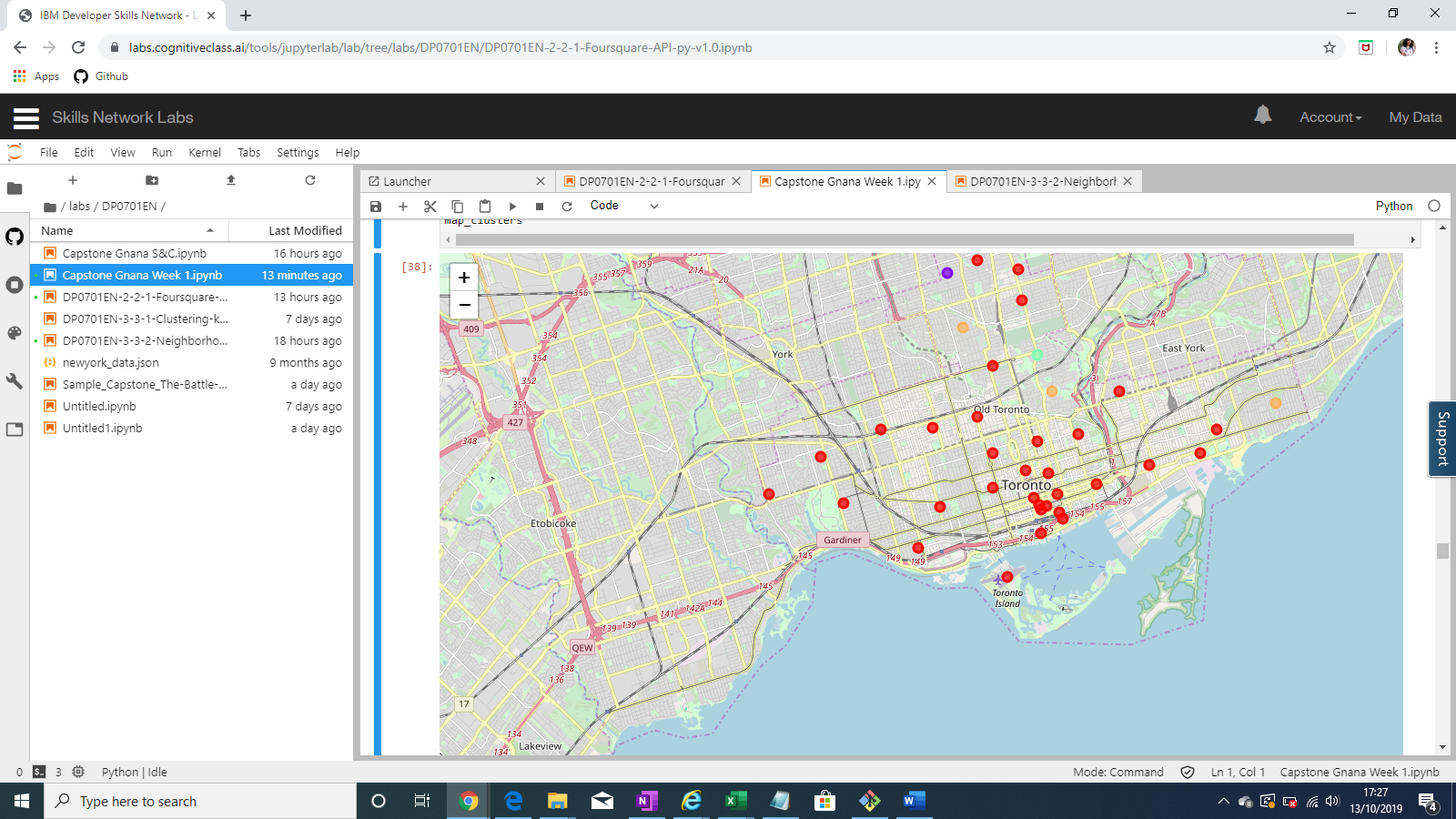
1. In the next part of exploration, I wanted to understand the nature of venues in the neighbourhoods, so I created a data frame with the top 10 common venue categories for each neighbourhood. This dataset for the basis for the clustering discussed in the next section.



## Clustering Neighbourhoods

I ran the k-means to cluster the neighbourhood into 5 clusters and created a new dataframe that includes the cluster as well as the top 10 venues for each neighbourhood. And then merged to add latitude/longitude for each neighbourhood.

Here is the map to visualize the clusters.



Next I examined each cluster in detail to infer cluster characteristics and explored the web to support my findings.

##### Cluster0 seems to have venues that are popular in the busy areas as is probably marks the business district of Toronto with lots of cafes, pubs and light weight entertainment centers to suit the busy lifestyle of the working class.

This cluster also encompasses the bulk of Toronto neighbourhoods indicating the lifestyle or key characteristic of the city itself.

##### Cluster1 seems like a residential zone with the basic and essential venues particularly useful for the vulnerable ie children and aged.[¶](https://e-ccf172908f.cognitiveclass.ai/lab#Cluster1-seems-like-a-residential-zone-with-the-basic-and-essential-venues-particularly-useful-for-the-vulnerable-ie-children-and-aged.)

There is also geographical proximity of these neighourhoods most stretching away from downtown to the north

##### Cluster2 seems to have venues characteristic of people who want to enjoy peaceful and leisure living.

Is a residential locality surrounded by posh neigbourhoods. Is geographically toward the north, away from the key business district.

##### Cluster3 seems like a zone inhabited with health concious poeple.

Wiki notes on Moore park - "Moore Park is one of Toronto's most affluent neighbourhoods."

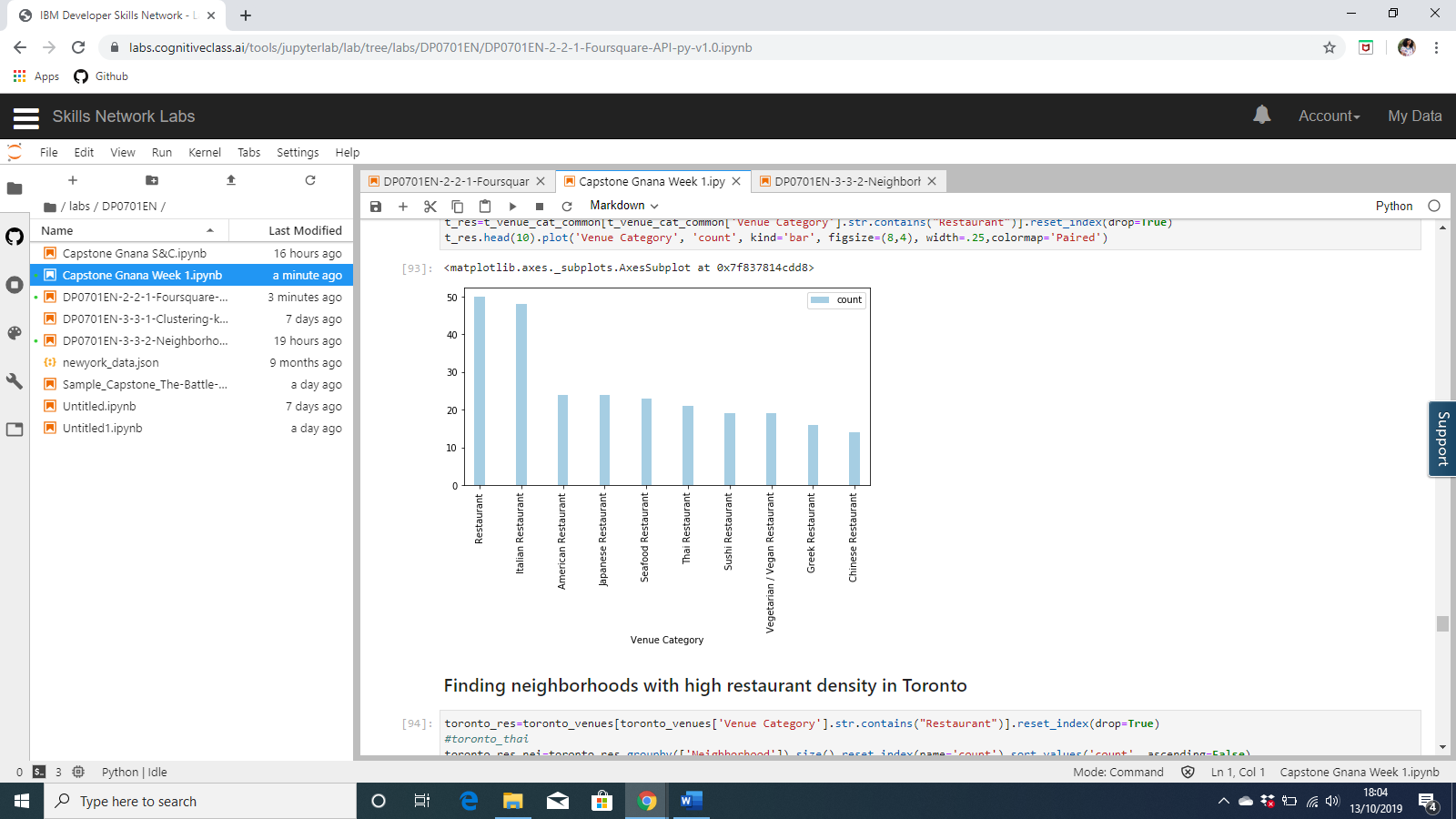
##### Cluster4 seems to have venues particularly suited for a peaceful community and large greenspace in central Toronto and seems prosperous.

It is aligned to wikipedia content on Rosedale: "It is located north of Downtown Toronto and is one of its oldest suburbs. It is also one of the wealthiest and most highly priced neighbourhoods in Canada.[2] Rosedale has been ranked the best neighbourhood in Toronto to live in by Toronto Life.[3] It is known as the area where the city's 'old money' lives,[4] and is home to some of Canada's richest and most famous citizens including Gerry Schwartz, founder of Onex Corporation, and Ken Thomson of Thomson Corporation, the latter of whom was the richest man in Canada at the time of his death in 2006"

## Venue Statistics

### I used the basic group y and Rank function to understand the following

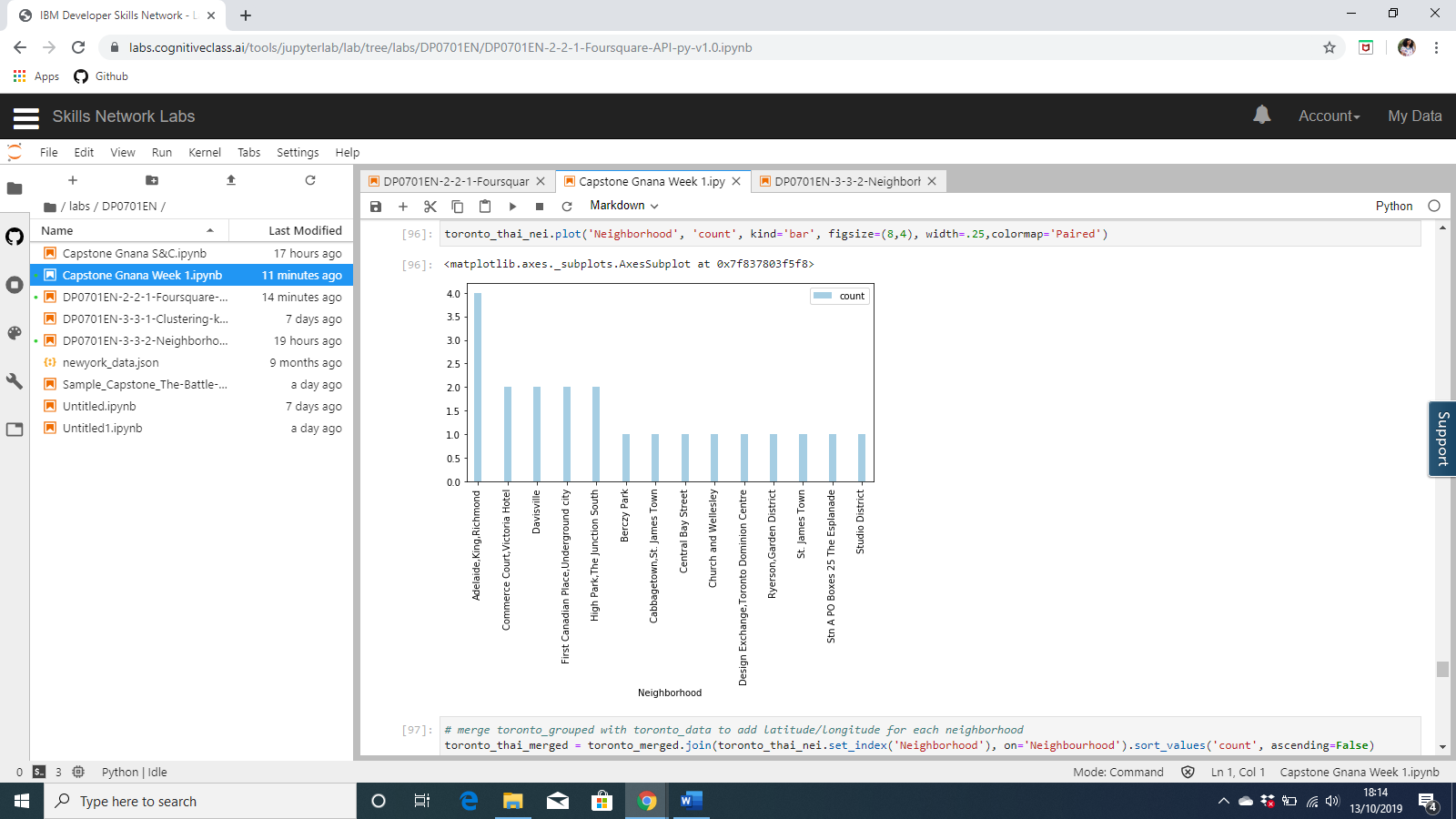
### Top 10 common restaurant categories in Toronto



1. Neighbourhood ranks based on the number of restaurants

|  |  |
| --- | --- |
| **Neighbourhood** | **count** |
| 6 | Chinatown,Grange Park,Kensington Market | 34 |
| 0 | Adelaide,King,Richmond | 30 |
| 8 | Church and Wellesley | 29 |
| 14 | First Canadian Place,Underground city | 26 |
| 5 | Central Bay Street | 25 |
| 9 | Commerce Court,Victoria Hotel | 25 |
| 25 | St. James Town | 23 |
| 24 | Ryerson,Garden District | 23 |
| 12 | Design Exchange,Toronto Dominion Centre | 22 |
| 26 | Stn A PO Boxes 25 The Esplanade | 21 |
| 20 | Little Portugal,Trinity | 20 |
| 30 | The Danforth West,Riverdale | 15 |
| 17 | Harbourfront East,Toronto Islands,Union Station | 13 |
| 10 | Davisville | 11 |
| 1 | Berczy Park | 11 |
| 4 | Cabbagetown,St. James Town | 11 |
| 27 | Studio District | 10 |
| 16 | Harbord,University of Toronto | 9 |
| 23 | Runnymede,Swansea | 9 |
| 18 | Harbourfront,Regent Park | 8 |
| 19 | High Park,The Junction South | 7 |
| 11 | Deer Park,Forest Hill SE,Rathnelly,South Hill,... | 4 |
| 28 | The Annex,North Midtown,Yorkville | 4 |
| 22 | Parkdale,Roncesvalles | 4 |
| 2 | Brockton,Exhibition Place,Parkdale Village | 3 |
| 29 | The Beaches West,India Bazaar | 3 |
| 21 | North Toronto West | 3 |
| 3 | Business Reply Mail Processing Centre 969 Eastern | 2 |
| 13 | Dovercourt Village,Dufferin | 2 |
| 7 | Christie | 2 |
| 15 | Forest Hill North,Forest Hill West | 1 |

1. All neighbourhoods with Thai restaurants in Toronto

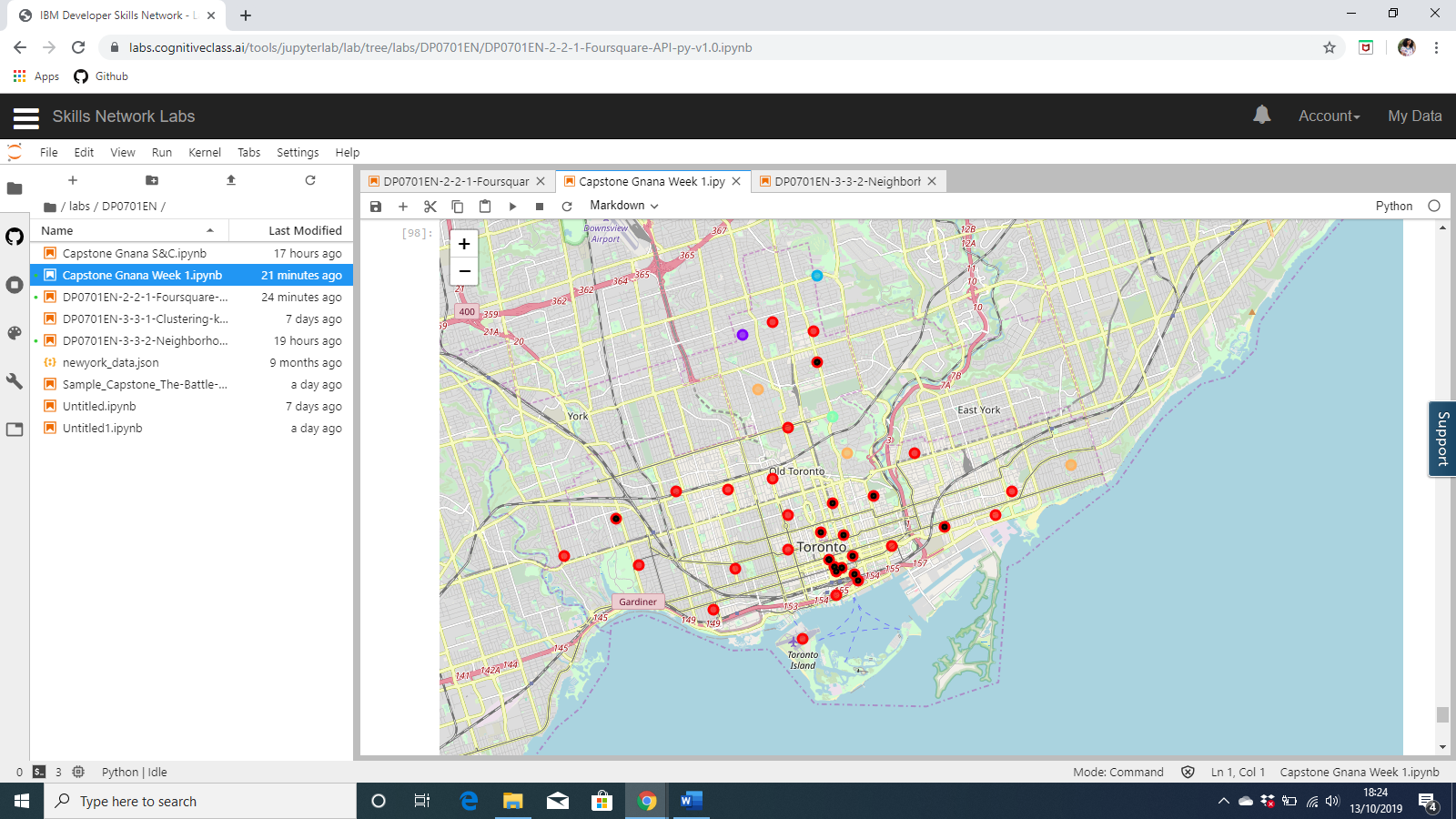


# **Results**

Post segmentation analysis, Cluster0 emerged as the most favourable cluster with neighbourhoods for opening any new restaurant. The

From other key statistics and ranking, it was clear to me that some neighbourhoods in this cluster were already highly populated with restaurants while others had only a few indicating that they could be high potential candidates for opening a restaurant.

There are already 21 Thai restaurants in Toronto in the same cluster. I tried to view the location of these existing Thai restaurants over the clustered neighbourhood map.



# **Discussion**

Toronto is a big city with a high immigrant population and is also the financial and business capital of Canada. Hence it is home to a variety of restaurants catering to various cusines. Restaurants density is high the key business district and adjoining areas which forms a large cluster of neighbourhoods making it difficult to critically narrow down the best neighbourhood for a new Thai restaurant.

There can be many factors that can be analysed, but in this project I have kept it simple just based on the existing all restaurant density, Thai restaurant density and geographical proximity to key business district. I used the Kmeans algorithm as part of this clustering study.

The black dots on the red neighbourhoods indicates the presence of a Thai restaurant already. From a visual inspection of the superimposed map, It was becoming clear that the south east parts of this cluster had neighbourhoods which could pose high competition to a new Thai restaurant.

Hence a neighbourhood away from these existing Thai restaurants where not only there were no Thai restaurants but also few other restaurants would be ideal location choices.

# **Conclusion**

On the basis of the visual inspection of the existing Thai restaurants in Cluster0, any new restaurant can be positioned in the western Cluster0 neighbourhoods i.e. in the area between High Park and Salad King which shares the characteristics of neighbourhoods in which the restaurant business is thriving.

Top picks are the neighbourhoods of Dover court Village and Christie given the low restaurant density in that, competition is less and hence can prove to be quite profitable.