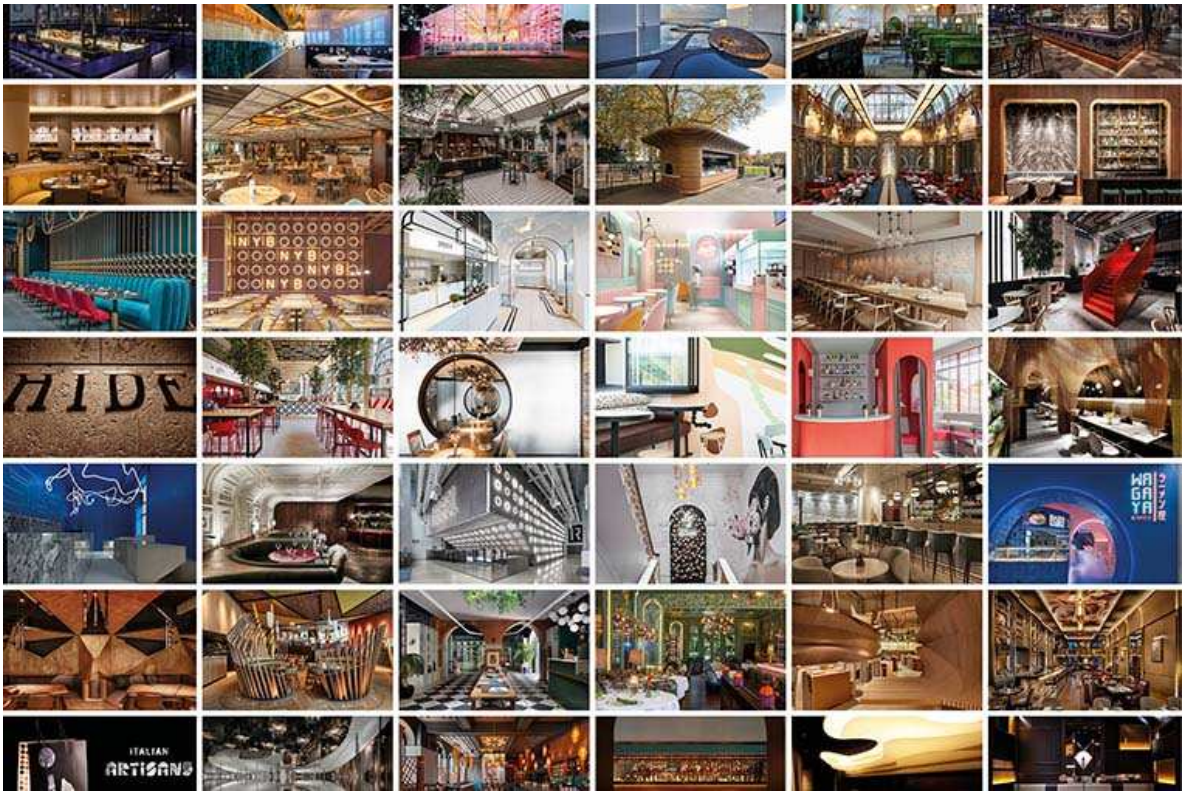
RESTAURANTS IN TORONTO
By David Rua



As a part of the Applied Data Science Capstone Project, I will work with data about restaurants in Toronto, to give some insights about what we can find in this city for **tourists** or what options are missing for **new businesses**, depending on the location and type of restaurant.
Now that we have been equipped with the skills and the tools to use location data to explore a geographical location, I came up with this idea and expect it can be useful for anyone who wants to visit Toronto or to start a new business idea.

In this project, I'll go through the process step by step from problem designing, data preparation to analysis and a conclusion that can be considered by tourists or business investors concerning restaurants categories in Toronto.

1. Introduction – Business Problem

Toronto is the provincial capital of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The Greater Toronto Area (GTA) as a whole had a 2016 population of 6,417,516. The city covers an area of 630.20 square kilometres (243.32 sq mi) and comprises six districts – East York, Etobicoke, North York, Old Toronto, Scarborough and York – which were amalgamated to form Toronto's present boundaries in 1998. Toronto is an international centre of business, finance, arts, and culture, and is recognized

as one of the most multicultural and cosmopolitan cities in the world (From Wikipedia, free encyclopedia).

The **cuisine of Toronto** reflects Toronto's size and multicultural diversity. Different ethnic neighbourhoods throughout the city focus on specific cuisines, such as authentic Chinese and Vietnamese found in the city's Chinatowns, Korean in Koreatown, Greek on The Danforth, Italian cuisine in Little Italy and Corso Italia, and Indian in Little India. Numerous other world cuisines are available throughout the city, including Portuguese, Hungarian, Japanese, and Caribbean. Toronto's large Jewish population has also ensured a variety of Jewish restaurants and delis, with varying adherence to kosher rules. In addition to ethnic cuisines, Toronto is also home to many fine dining establishments and chain restaurants ranging from fast food to casual or upscale dining.

Several culinary festivals take place in Toronto each year. In addition, food tours in Toronto, like those offered by Tasty Tours Toronto (tastytourstoronto.com) or Culinary Adventure Co (Abel, Ann. March 13, 2017. "Eat the World: 9 Best Food Tours") are an increasingly popular way for locals and tourists to explore the food culture of the city.

Toronto-based culinary events and festivals include:

- Charlie's Burgers is a Toronto-based series of private dining events established in 2009.
- Summerlicious and Winterlicious.
- Taste of the Danforth.
- Toronto Wing Festival.
- Vegetarian Food Festival (vegfoodfest.com).

Having all this information about Toronto's cuisine, I would like to know more about restaurants in the city. In addition, this battle of neighborhoods is a perfect option to get deep in this subject and discover a little more about the city.

What I would like to solve is where should I or any other tourist go to find a specific type of restaurant in Toronto, according to the location.

2. Data description

The first step was import and install all libraries I needed in the project, like pandas, numpy, folium, matplotlib and sklearn, among others.

In the second step, I continued with the data collection. I created a Toronto dataframe from a csv file (downloaded from Wikipedia, free encyclopedia) and import the geographical coordinates of Toronto (from url 'https://cocl.us/Geospatial_data'). Then, I merged Toronto and Coordinates dataframes into a new dataframe and created a map of Toronto's neighborhoods using folium (folium.Map). This was an important key because here I saw the number of boroughs and neighborhoods to start the search of restaurants.

After this, I got get the location data of Toronto using **Foursquare Developers** (ID & SECRET), based on the geographical coordinates previously loaded. I explored Toronto's venues in a 2km radius, got a new dataframe and explored the nearby venues (80 per neighborhood). Then, I grouped these venues by neighborhood and did one hot encoding for getting dummies of venue category. So that

I could calculate count of all venue group by neighborhoods. Finally, I printed each neighborhood along with the top 5 most common venues and displayed the top 10 venues for each neighborhood.

The next was running a k-means to cluster the neighborhoods into 5 clusters, created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood. After that, I visualized the resulting clusters in a map and examined the 5 clusters common venues.

Next, I explored the frequency of categories in the neighborhoods with a word cloud and created a new dataframe with the venues category that contain the word "Restaurant", that was my target venue. Following, I displayed a map of Toronto's restaurants.

In the exploratory data analysis, I got some detail about neighborhoods and restaurants in the city, found the number and frequency of categories, saw the venue's names and locations, used a word cloud for restaurants, and then mapped them.

Finally, I explored the restaurants categories using bar plots of number of restaurants vs boroughs and categories. With this information, I saw the most common category and venues location. At last, I showed the names and coordinates of the restaurants and mapped them in a specific borough.

3. Methodology

The information about Toronto's neighborhoods from Wikipedia had just three fields (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M): Postal Code, Borough and Neighborhood. So, I had to merged it with the coordinates from the city in a new dataframe. The result was 10 boroughs and 103 neighborhoods.

Setting the geographical coordinates of Toronto with the geolocator, it resulted in a latitude of 43.6534817 and a longitude of -79.3839347. Using folium.Map, the neighborhoods looked like this:

After this, I used Foursquare to get the location data of Toronto venues. The resulted dataframe was this one:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Lawrence Park	43.728020	-79.388790	J.N. Munkongolo & Associates - Family Lawyer	43.727437	-79.390264	Lawyer
2	Lawrence Park	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
3	Lawrence Park	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
4	Davisville North	43.712751	-79.390197	Sherwood Park	43.716551	-79.387776	Park
5	Davisville North	43.712751	-79.390197	Summerhill Market North	43.715499	-79.392881	Food & Drink Shop
6	Davisville North	43.712751	-79.390197	Homeway Restaurant & Brunch	43.712641	-79.391557	Breakfast Spot
7	Davisville North	43.712751	-79.390197	Winners	43.713236	-79.393873	Department Store
8	Davisville North	43.712751	-79.390197	Best Western Roehampton Hotel & Suites	43.708878	-79.390880	Hotel
9	Davisville North	43.712751	-79.390197	Subway	43.708474	-79.390674	Sandwich Place

In this dataframe there were 266 unique categories. Using get_dummies I grouped rows by neighborhood, took the mean of the frequency of occurrence of each category, and printed each neighborhood along with the top 5 most common venues. Here are some examples:

```
----Cedarbrae----
                venue  freq
0     Hakka Restaurant  0.12
1  Fried Chicken Joint  0.12
2                Bank  0.12
3   Athletics & Sports  0.12
4      Thai Restaurant  0.12
```

```
----Central Bay Street----
                venue  freq
0         Coffee Shop  0.17
1      Sandwich Place  0.06
2  Italian Restaurant  0.06
3               Café  0.05
4  Japanese Restaurant  0.05
```

```
----Church and Wellesley----
                venue  freq
0     Sushi Restaurant  0.07
1         Coffee Shop  0.07
2  Japanese Restaurant  0.05
3          Restaurant  0.04
4             Gay Bar  0.04
```

Then, I ran a k-means to cluster the neighborhood into 5 clusters (kclusters = 5). Created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood. The result for Cluster 1 was:

	Postal Code	Cluster Labels	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	M4N	0	Bus Line	Park	Swim School	Lawyer	College Gym	College Rec Center	Drugstore	Donut Shop	Doner Restaurant	Dog Run
1	M4P	0	Hotel	Park	Gym	Department Store	Sandwich Place	Pizza Place	Food & Drink Shop	Breakfast Spot	Convenience Store	Dessert Shop
2	M4R	0	Clothing Store	Coffee Shop	Yoga Studio	Sporting Goods Shop	Gift Shop	Ice Cream Shop	Fast Food Restaurant	Diner	Mexican Restaurant	Park
3	M4S	0	Pizza Place	Sandwich Place	Dessert Shop	Coffee Shop	Gym	Café	Italian Restaurant	Sushi Restaurant	Indian Restaurant	Seafood Restaurant
5	M4V	0	Pub	Coffee Shop	American Restaurant	Liquor Store	Supermarket	Sports Bar	Sushi Restaurant	Bank	Pizza Place	Fried Chicken Joint
...	...	...	...	...	...	...	...	...	...	...	...	...
96	M6R	0	Gift Shop	Breakfast Spot	Cuban Restaurant	Bookstore	Dog Run	Italian Restaurant	Bar	Bank	Restaurant	Dessert Shop
97	M6S	0	Café	Coffee Shop	Restaurant	Pub	Pizza Place	Sushi Restaurant	Diner	Italian Restaurant	Gourmet Shop	Gym
98	M6C	0	Trail	Playground	Hockey Arena	Field	Curling Ice	Dance Studio	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
100	M6M	0	Sandwich Place	Bar	Convenience Store	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Women's Store	Dance Studio
101	M6N	0	Caribbean Restaurant	Grocery Store	Convenience Store	Breakfast Spot	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Women's Store

92 rows × 12 columns

To find the restaurant category, I looked for the frequency of venues category in the neighborhoods and graphed the next word cloud, were “Coffee Shop” showed up as one of the most common.

So, I created a new dataframe with the venues category that contain the word "Restaurant", that was my target, resulting in a 445 venues dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
15	North Toronto West, Lawrence Park	43.715383	-79.405678	C'est Bon	43.716785	-79.400406	Chinese Restaurant
19	North Toronto West, Lawrence Park	43.715383	-79.405678	Sushi Shop	43.713861	-79.400093	Restaurant
21	North Toronto West, Lawrence Park	43.715383	-79.405678	Tio's Urban Mexican	43.714630	-79.400000	Mexican Restaurant
22	North Toronto West, Lawrence Park	43.715383	-79.405678	A&W	43.715149	-79.399944	Fast Food Restaurant
35	Davisville	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant

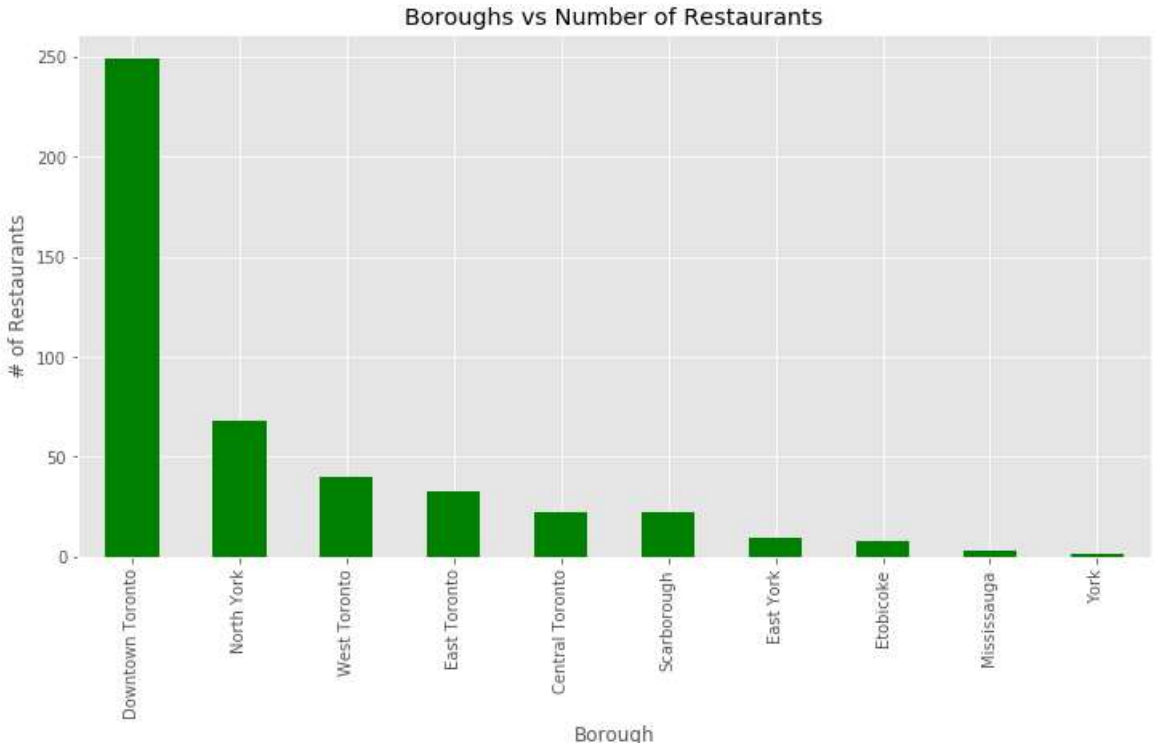
<u>Exploratory data analysis</u>

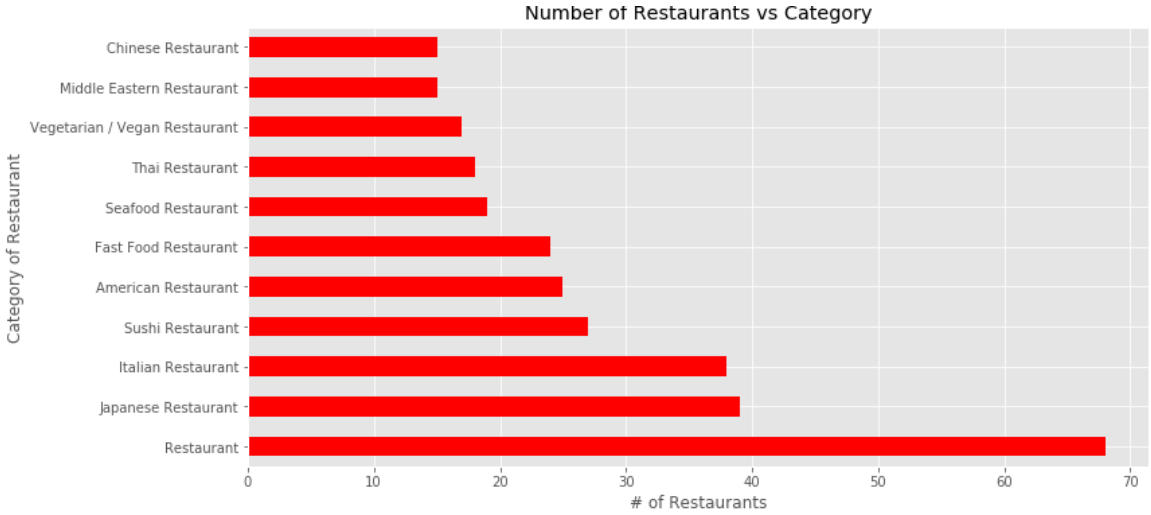
I began this exploring by getting some details about neighborhoods and restaurants category. There were 47 categories that include “Restaurant” label. This was the resulted description and word cloud:

```
Description of restaurants category is
count     47.000000
mean       9.468085
std       12.664313
min        1.000000
25%        1.000000
50%        3.000000
75%       13.000000
max       66.000000
Name: Venue Category, dtype: float64
The number of restaurants in this query is
(445,)
```

Within this 47 groups, the 10 most frequent were: Restaurant, Japanese, Italian, Sushi, American, Fast Food, Seafood, Thai, Vegetarian / Vegan, and Middle Eastern.

With bar plots I explored the number of restaurants by borough and category.





4. Results

After running this code and analyzing the outputs, here are the main results.

The main restaurant category is Restaurant (68) without a specific cuisine, followed by Japanese Restaurant (39) and Italian Restaurant (38). Almost 56% of restaurants are in Downtown Toronto.

Venue Category	Borough
Restaurant	68
Japanese Restaurant	39
Italian Restaurant	38
Sushi Restaurant	27
American Restaurant	25
Fast Food Restaurant	24
Seafood Restaurant	19
Thai Restaurant	18
Vegetarian / Vegan Restaurant	17
Middle Eastern Restaurant	15

Borough	Venue Category
Downtown Toronto	249
North York	68
West Toronto	40
East Toronto	33
Central Toronto	22
Scarborough	22
East York	9
Etobicoke	8
Mississauga	3
York	1

If you want to know a specific restaurant category location, I left a code for search it and map it, for example Thai Restaurant:

	Neighborhood	Thai Restaurant
0	Agincourt	0.000000
1	Bathurst Manor, Wilson Heights, Downsview North	0.000000
2	Bayview Village	0.000000
3	Bedford Park, Lawrence Manor East	0.100000
4	Berczy Park	0.083333
5	Brockton, Parkdale Village, Exhibition Place	0.000000
6	Business reply mail Processing Centre, South C...	0.000000
7	Canada Post Gateway Processing Centre	0.000000
8	Cedarbrae	0.333333
9	Central Bay Street	0.111111

If a tourist is interested in knowing a great variety of restaurants, I would suggest go Downtown Toronto, look for a specific category and find it in the map.

	Borough	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
22	Downtown Toronto	St. James Town, Cabbagetown	Kingyo Toronto	43.665895	-79.368415	Japanese Restaurant
23	Downtown Toronto	St. James Town, Cabbagetown	F'Amelia	43.667536	-79.368613	Italian Restaurant
24	Downtown Toronto	St. James Town, Cabbagetown	Murgatroid	43.667381	-79.369311	Restaurant
25	Downtown Toronto	St. James Town, Cabbagetown	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
26	Downtown Toronto	St. James Town, Cabbagetown	Mr. Jerk	43.667328	-79.373389	Caribbean Restaurant
...	...	...	...	...	...	...
266	Downtown Toronto	Christie	Actinolite	43.667858	-79.428054	Restaurant
267	Downtown Toronto	Queen's Park, Ontario Provincial Government	Mercatto	43.660391	-79.387664	Italian Restaurant
268	Downtown Toronto	Queen's Park, Ontario Provincial Government	Sushi Box	43.662960	-79.386580	Sushi Restaurant
269	Downtown Toronto	Queen's Park, Ontario Provincial Government	Tokyo Sushi	43.665885	-79.386977	Sushi Restaurant
270	Downtown Toronto	Queen's Park, Ontario Provincial Government	Como En Casa	43.665160	-79.384796	Mexican Restaurant

5. Discussion

For a business investor I would recommend start exploring North York and West Toronto, that were the second and third boroughs with more restaurants. Then take a look of categories in these Boroughs and compare with Downtown, to have an initial idea for further analysis.

To the owners, I would suggest adding some other adjective to the Restaurant category, due to it is easy for a tourist if there is an extra label for decide of visit a venue.

After doing this project, I suggest importing and installing all necessary libraries at the beginning of the notebook, so you can run all the code or sequences you need without concerning of which ones are missing.

As I am a travel enthusiast, with this project I also wanted to create a code that can be used or reply in many cities around the world, with a few changes in the targeted city.

6. Conclusion

With this “Battle of Neighborhoods” I found what categories of restaurants are more common and where are they located in the city, so we can explore a particular neighborhood. For example, now I can find Thai restaurants easily with the notebook.

A tourist can find this project useful if is planning to visit a specific borough of Toronto and want to explore a special kind of food.

A businessperson would have an initial idea of what options are missing in some boroughs, according to the categories with few frequency and potential customers.