

# **IBM Data Science Capstone Project by Yasmine Mahmoud**

## **The Best Neighborhood in Toronto to open a Gluten Free Bakery**

### **1. Introduction**

For the last few months, I have been working on earning the IBM Data Science professional certificate. IBM Data Science professional certificate is a set of specialized courses with hands on Labs designed to train you to work as a Data Scientist.

The courses are very engaging with real datasets and real world problems to work on.

The Specialization consists of 9 courses: Data science Methodology, Tools for Data Science, Python for Data science and AI, Databases and SQL with Python, Data analysis with Python, Data visualization with Python, Machine Learning with Python and lastly, the Capstone project where you define a problem, describe and collect data from online resources to create a solution for an exciting problem.

In this blog, I explain my process for determining the perfect spot to open a Gluten Free Bakery in Toronto, Canada. Along the way, I will explain all the processing steps starting with problem formulation, data preparation, methodology, the data analysis and visualization and lastly, the machine learning algorithm to predict the perfect spot for the restaurant.

### **2. Business Problem**

We're more than a decade into the Gluten Free market growth. Thanks to increased awareness of Celiac Disease and gluten sensitivity -- and improving rates of diagnosis, there is a variety of Gluten Free products in almost every grocery store.

While demand for gluten free meals in restaurants eventually gives those affected with Celiac disease and health conscious individuals more choice, the situation does come with pitfalls. The most important one, is the need to ensure that any restaurant with a gluten free menu knows not just what ingredients are used in their recipes...but how to safely prepare the meal without putting you at risk from gluten cross contamination.

The next important pitfall, is that it's not unusual to go into a bakery that advertises gluten-free baked goods and be told that they don't have it or simply they can't ensure it's not gluten contaminated. So, you basically feel like a second-class citizen just because you have a disease or you are health conscious.

### **3. Target Audience**

Since people around the world are becoming more and more conscious of what they put into their bodies and how that impacts their health, opening a Gluten Free Bakery in Toronto will attract a large audience, not only those affected with Celiac disease.

This project demonstrates to investors that opening a Gluten Free Bakery is a lucrative business idea.

## 4. The Data

The following Datasets contain various unique information about Toronto and its neighborhoods to discover the best spot to open a Gluten free Bakery:

**4.1 Foursquare Location Data:** to get the location and venue information (<https://developer.foursquare.com/docs>)

**4.2 Neighborhood Profiles Data:** To get detailed information about each neighborhood in Toronto. (<https://open.toronto.ca/dataset/neighbourhood-profiles/>)

**4.3 Neighborhoods Dataset:** To get information about Toronto neighborhoods such as location, and the name of all the neighborhoods present in Toronto. (<https://open.toronto.ca/dataset/neighbourhoods/>)

**4.4 List of Postal Codes of Canada:** [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

## 5. Methodology

After cleaning and exploring the data, we will apply the K-means machine learning algorithm to create clusters of Toronto neighborhoods and choose the best spot(s) to open a bakery.

First I scraped the web to get the postal codes of Toronto:

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned

I cleaned and preprocessed the data:

Postal Code	Borough	Neighbourhood
0	M3A	Parkwoods
1	M4A	Victoria Village
2	M5A	Regent Park, Harbourfront
3	M6A	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park, Ontario Provincial Government
...	...	...
98	M8X	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Church and Wellesley
100	M7Y	Business reply mail Processing Centre, South C...
101	M8Y	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Mimico NW, The Queensway West, South of Bloor,...

Then I imported a csv file to extract the latitudes and longitudes for various neighborhoods in Toronto:

	<b>Postal Code</b>	<b>Latitude</b>	<b>Longitude</b>
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Then merged the two previous tables on Postal Codes:

	<b>Postal Code</b>	<b>Borough</b>	<b>Neighbourhood</b>	<b>Latitude</b>	<b>Longitude</b>
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Next, I extracted the data needed for the analysis:

	<b>Neighbourhood</b>	<b>Latitude</b>	<b>Longitude</b>
2	Regent Park, Harbourfront	43.654260	-79.360636
4	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
9	Garden District, Ryerson	43.657162	-79.378937
15	St. James Town	43.651494	-79.375418
19	The Beaches	43.676357	-79.293031

I then visualized all the Neighborhoods of the above data frame using Folium



I then used geocode to find the coordinates of the city Toronto:

The geographical coordinate of Toronto City are  
43.65238435, -79.38356765.

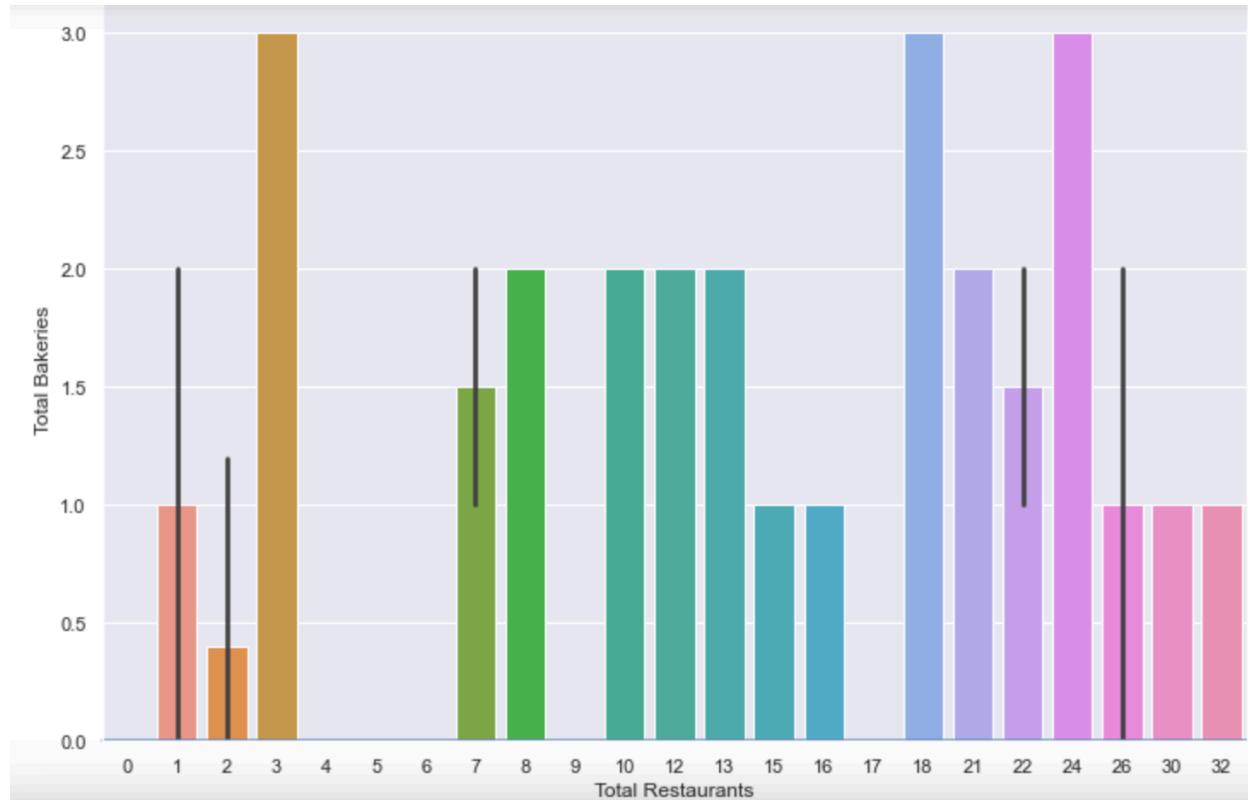
After That, I used the Foursquare developer app and the one hot encoder to extract the different venue types in Toronto:

	Neighbourhood	Adult Boutique	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Tibetan Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store
0	Berczy Park	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0
1	Brockton, Parkdale Village, Exhibition Place	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	Business reply mail Processing Centre, South C...	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	CN Tower, King and Spadina, Railway Lands, Har...	0	1	1	1	2	2	2	0	0	...	0	0	0	0	0	0
4	Central Bay Street	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0

Then I compared the number of Restaurants to the number of Bakeries in Toronto per Neighborhood:

	Neighbourhood	Total Bakeries	Total Restaurants	Latitude	Longitude
0	Berczy Park	2	12	43.644771	-79.373306
1	Brockton, Parkdale Village, Exhibition Place	2	2	43.636847	-79.428191
2	Business reply mail Processing Centre, South C...	0	2	43.662744	-79.321558
3	CN Tower, King and Spadina, Railway Lands, Har...	0	0	43.628947	-79.394420
4	Central Bay Street	0	17	43.657952	-79.387383

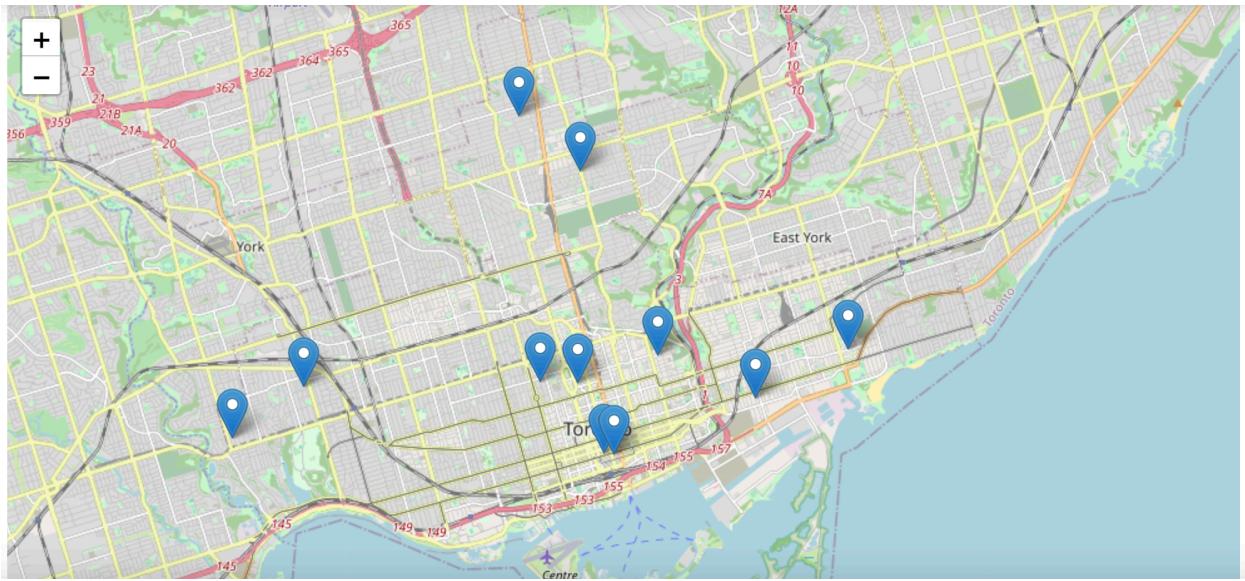
And Visualized it:



Since there was no gluten free venues datasets, I clustered the data using the number of bakeries compared to restaurants per Neighborhood using the Kmeans Clustering machine learning Algorithm:

Cluster Labels	Neighbourhood	Total Bakeries	Total Restaurants	Latitude	Longitude
0	4	Berczy Park	2	43.644771	-79.373306
1	1	Brockton, Parkdale Village, Exhibition Place	2	43.636847	-79.428191
2	1	Business reply mail Processing Centre, South C...	0	43.662744	-79.321558
3	1	CN Tower, King and Spadina, Railway Lands, Har...	0	43.628947	-79.394420
4	4	Central Bay Street	0	43.657952	-79.387383

Cluster Labels	Neighbourhood	Total Bakeries	Total Restaurants	Latitude	Longitude
7	3	Commerce Court, Victoria Hotel	1	43.648198	-79.379817
8	2	Davisville	0	43.704324	-79.388790
11	3	First Canadian Place, Underground city	1	43.648429	-79.382280
15	2	High Park, The Junction South	1	43.661608	-79.464763
16	2	India Bazaar, The Beaches West	0	43.668999	-79.315572



I found that these two neighborhoods would be the two best spots to open a bakery:

] :

	Neighbourhood	Total Bakeries	Total Restaurants	Latitude	Longitude
7	Commerce Court, Victoria Hotel	1	32	43.648198	-79.379817
11	First Canadian Place, Underground city	1	30	43.648429	-79.382280

In the future, I will scrape the web to find data on gluten free venues and analyze the data.