

# **Coursera IBM**

## **Applied Data Science Capstone**

The Coursera logo, featuring the word "coursera" in a white, lowercase, sans-serif font, centered within a solid blue rectangular background.

**coursera**

**Report on**  
**["The Battle of Neighborhoods"]**  
**Capstone Project**

## Table of Contents

Chapter 1: Introduction .....	3
1.1 Problem Background .....	3
1.2 Problem Description .....	4
1.2 Target Audience .....	5
1.3 Success Criteria .....	5
Chapter 2: Data Collection .....	6
2.1 Requirements .....	6
2.2 Collections .....	6
2.3 Datasets .....	7
2.3.1 Neighborhoods Data .....	7
2.3.2 Toronto Venues Data .....	7
2.3.3 Toronto Pharmacy Data .....	8
2.3.4 Neighborhoods without Pharmacies Data .....	8
2.3.5 Neighborhoods with Pharmacies Data .....	9
Chapter 3. Methodology .....	10
Chapter 4: Result .....	11
Chapter 5: Conclusion .....	15

However, that also means the market is highly competitive and as a well-developed city, the cost of doing business is also one of the highest in the country. And thus, any new business venture or expansion in the country needs to be reviewed carefully and strategically targeted so that the return on investment will be sustainably reasonable and more importantly the investment can be considerably less riskier.

## 1.2 Problem Description

There are approximately 88,000 pharmacies in the United States. Over half (about 48,000) are located within drug stores, grocery stores, hospitals, department stores, medical clinics, surgery clinics, universities, nursing homes, prisons, and other facilities. What about Community Pharmacies? Community Pharmacies offer basic preliminary health assessments in addition to dispensing prescriptions. The community pharmacist is considered to be the most accessible healthcare professional to the general public, as they are available to anyone on a walk-in basis. Approximately 9 out of 10 Americans live within 5 miles of a community pharmacy



Community pharmacists are considered to be the most accessible health professional to the public, as they are available to provide personalized advice about health and medicine on a walk-in basis, without the need for an appointment. 89% of the population in the United Kingdom can access a community pharmacy within a 20-minute walk. The community pharmacies are found to be scattered in city of Toronto. This is clearly a problem that people need to walk/drive more to access

dedicated pharmacies. If they are out of time, they are compelled to make an appointment for hospitals.

## 1.2 Target Audience



To solve this problem, Data Scientist team led by myself has been engaged by Orange Drugs Inc. The objective is to locate and recommend to the management which neighborhood will be the best choice to start off their first retail Pharmacy. The management also expects to understand the rationale of the recommendations in the report.

## 1.3 Success Criteria

The success criteria of this project will be a good recommendation of the neighborhood choice in Toronto to the management of Orange Drugs Inc based on following factors:

- i. Lack of Pharmacy stores available (less competition).
- ii. Number of Hospital around.

- iii. Higher number of Hospital Admissions (higher demand).
- iv. Population Size.

## Chapter 2: Data Collection

### 2.1 Requirements

As we need to explore, segment, and analyze the neighborhoods in the city of Toronto, the Toronto neighborhoods data is key for this project. We need the names of neighborhood, their populations, hospital admission scenario. Here, we also need the exact geolocations of each neighborhoods.

We need the data how hospitals in scattered all over Toronto. For this, their geolocation is must. We need the data how pharmacy stores are distributed all over the city. The data should contain the coordinates for each of the Pharmacies in Toronto that will help us to further obtaining more information critical for this project. We will also like to obtain the key information; such as average income for each neighborhood which might be one of the key factors for the neighborhood of choice in the final recommendation report.

### 2.2 Collections

From '<http://www.torontohealthprofiles.ca/>', we obtain hospital admission rate in different neighborhoods in Toronto, Canada. It includes: neighborhoods name, their population, admission scenario. We use geopy for integrating these observations with geolocation. It doesn't have average income of person in each neighborhood. We scrape and clean the income data of each neighborhood from '<https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research>' and left join with above neighborhood data.

We use Foursquare API for getting Venue's data using the geolocation of Neighborhoods. We Filtered this data for getting the Pharmacies Data in Toronto, Ontario

## 2.3 Datasets

### 2.3.1 Neighborhoods Data

It has the all the neighborhoods name, their respective populations, hospital admission scenario. Along with these, it also comprises geolocation, and income.

	Neighbourhood	Population	Male Admission	Female Admission	Both Admission	Latitudes	Longitudes	Income
0	West Humber-Clairville	34080	2453	3151	5604	43.723370	-79.597457	71699.0
1	Rexdale-Kipling	10455	938	1090	2028	43.722114	-79.572292	64800.0
2	Elms-Old Rexdale	9970	674	978	1652	43.721770	-79.552173	63156.0
3	Humber Heights-Westmount	10600	950	1335	2285	43.695785	-79.520832	76669.0
4	Edenbridge-Humber Valley	14895	1029	1269	2298	43.670672	-79.518855	144418.0
5	Markland Wood	10395	804	944	1748	43.631239	-79.585434	NaN
6	Etobicoke West Mall	11005	821	1058	1879	43.643549	-79.565325	80527.0
7	Islington-City Centre West	38130	2829	3966	6795	43.662577	-79.531789	NaN

### 2.3.2 Toronto Venues Data

It has Venues names with their respective geolocation, its category, and the neighborhood where it lies.



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	West Humber-Clairville	43.72337	-79.597457	Fortinos	43.721438	-79.596291	Grocery Store
1	West Humber-Clairville	43.72337	-79.597457	Mandarin Buffet	43.720360	-79.594387	Chinese Restaurant
2	West Humber-Clairville	43.72337	-79.597457	Humber Arboretum	43.724653	-79.608006	Garden
3	West Humber-Clairville	43.72337	-79.597457	Bulk Barn	43.720290	-79.600887	Food & Drink Shop
4	West Humber-Clairville	43.72337	-79.597457	New York Fries - Woodbine Centre	43.720568	-79.599609	Restaurant
5	West Humber-Clairville	43.72337	-79.597457	TD Canada Trust	43.719630	-79.599896	Bank
6	West Humber-Clairville	43.72337	-79.597457	Winners	43.719819	-79.594923	Department Store
7	West Humber-Clairville	43.72337	-79.597457	Woodbine Racetrack	43.716673	-79.604962	Racecourse

### 2.3.3 Toronto Pharmacy Data

It stores the Pharmacies, their respective names, and the neighborhood where it lies.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Humber Heights-Westmount	43.695785	-79.520832	Shoppers Drug Mart	43.701264	-79.519705	Pharmacy
1	Edenbridge-Humber Valley	43.670672	-79.518855	Shoppers Drug Mart	43.661721	-79.518881	Pharmacy
2	Etobicoke West Mall	43.643549	-79.565325	Shoppers Drug Mart	43.641312	-79.576924	Pharmacy
3	Islington-City Centre West	43.662577	-79.531789	Shoppers Drug Mart	43.663067	-79.531753	Pharmacy
4	Islington-City Centre West	43.662577	-79.531789	Thorncrest Drug Store	43.662988	-79.531817	Pharmacy
5	Kingsway South	43.647381	-79.511333	Shoppers Drug Mart	43.647623	-79.510860	Pharmacy
6	New Toronto	43.600763	-79.505264	Shoppers Drug Mart	43.601677	-79.502239	Pharmacy
7	Long Branch	43.592005	-79.545365	Rexall	43.593111	-79.538327	Pharmacy
8	Alderwood	43.601717	-79.545232	Sav-On Drug Mart	43.609891	-79.549618	Pharmacy
9	Glenfield-Jane Heights	43.756799	-79.518050	Shoppers Drug Mart	43.756147	-79.515843	Pharmacy

### 2.3.4 Neighborhoods without Pharmacies Data

It stores the information of neighborhoods without any retail Pharmacies.



	Neighbourhood	Population	Male Admission	Female Admission	Both Admission	Latitudes	Longitudes	Income
0	West Humber-Clairville	34080	2453	3151	5604	43.723370	-79.597457	71699.0
1	Rexdale-Kipling	10455	938	1090	2028	43.722114	-79.572292	64800.0
2	Elms-Old Rexdale	9970	674	978	1652	43.721770	-79.552173	63156.0
5	Markland Wood	10395	804	944	1748	43.631239	-79.585434	NaN
9	Mimico	26635	1758	2080	3838	43.616677	-79.496805	71772.0

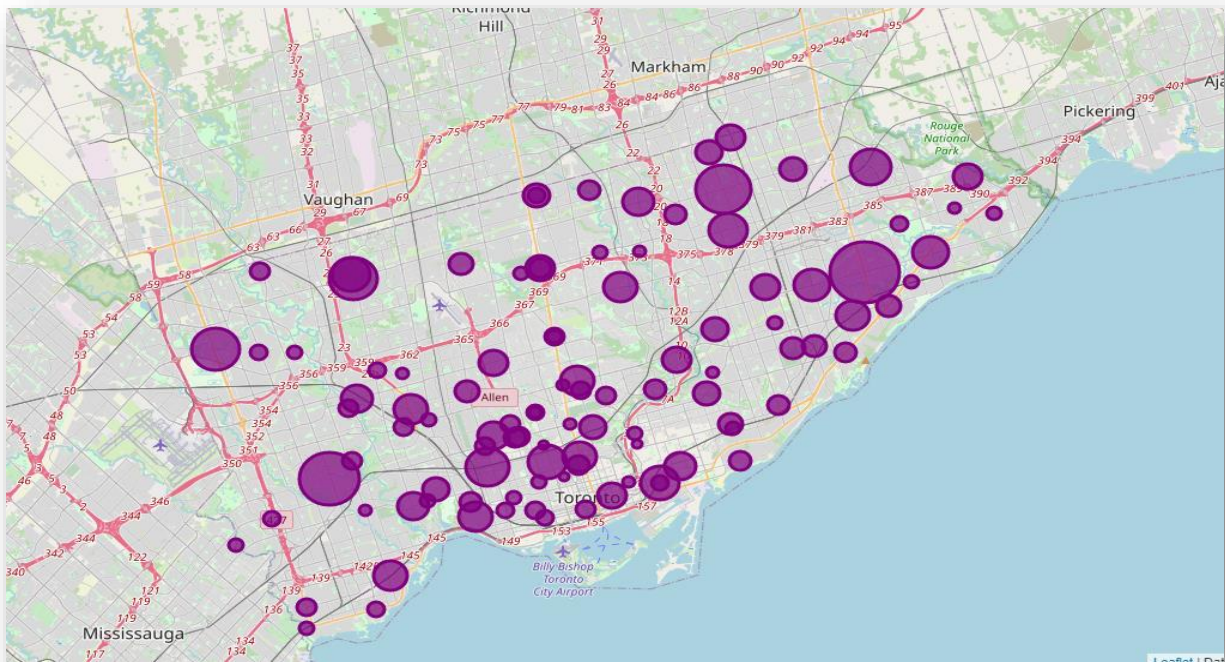
### 2.3.5 Neighborhoods with Pharmacies Data

It stores the information of neighborhoods with retail Pharmacies.

	Neighbourhood	Population	Male Admission	Female Admission	Both Admission	Latitudes	Longitudes	Income
3	Humber Heights-Westmount	10600	950	1335	2285	43.695785	-79.520832	76669.0
4	Edenbridge-Humber Valley	14895	1029	1269	2298	43.670672	-79.518855	144418.0
6	Etobicoke West Mall	11005	821	1058	1879	43.643549	-79.565325	80527.0
7	Islington-City Centre West	38130	2829	3966	6795	43.662577	-79.531789	NaN
8	Kingsway South	9170	607	745	1352	43.647381	-79.511333	198998.0

## Chapter 3. Methodology

The project completes following different methodologies and techniques from using the publicly available data in Canadian data resource portal to scraping information from Wikipedia and other sites. The data from Toronto Health Profiles website contains observation of all neighborhoods of Canada. It includes vital information like name of neighborhood, their respective population, hospital admission scenario, etc. This is critical to understand the population of each Toronto neighborhood, their hospital admission scenario which are one of the key elements in the neighborhood of choice in this project. Furthermore, we need to know the coordinates and locations of this neighborhoods, and therefore the geocoder API has been used for achieving this objective.



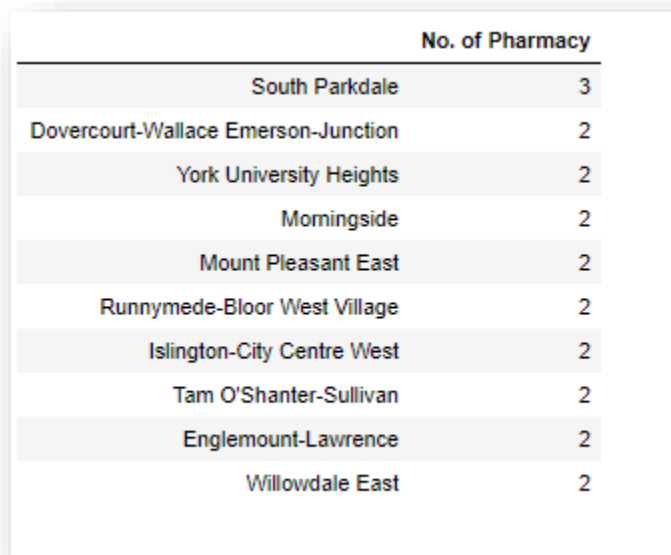
This geolocation is important because we input this information into the location information provider such as Foursquare.com to obtain venue information in these neighborhoods, and this is precisely what we have done for it in this project. When we obtain all the venues, their geolocation, and category we filtered the data based upon the category Pharmacy. So, this is how we get all the pharmacies in Toronto

City. This is critical as we need to recommend to the management the best neighborhood and other suitable neighborhoods so that Orange Drugs can start their first Pharmacy Store in the best one and plans their future extensions in following suitable neighborhoods. According to Pharmacy data, we'll visualize which regions are least competent and suitable for starting the Pharmacy Store.

Finally, with all these methodologies, we will then be able to come up with a best recommendation to Orange Drugs based on store competition, population, number of hospitals, and hospital admission scenario. In other words, we will not want to recommend to the management to enter a neighborhood whereby there is already a high concentration of Pharmacy stores available, no hospitals, and least admission scenario.

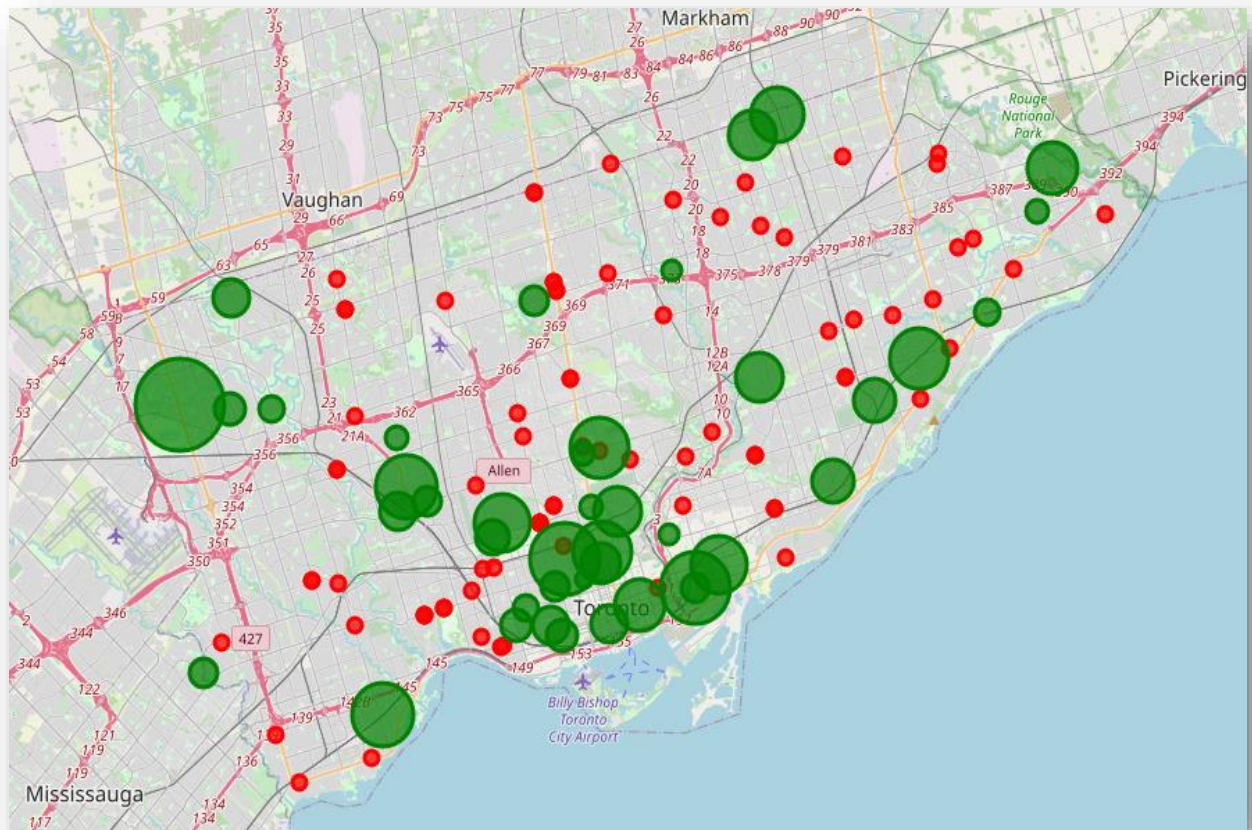
## Chapter 4: Result

Firstly, we separate neighborhoods with and without Community Pharmacies. The top ten neighborhoods with most pharmacies are as shown in figure below:



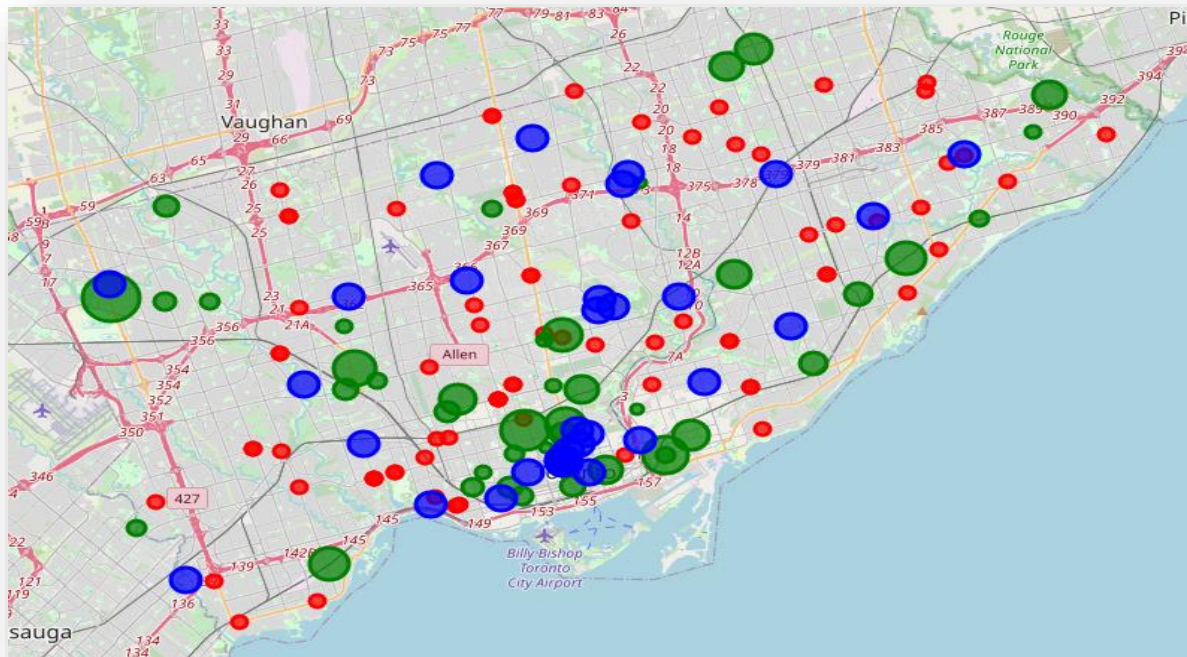
No. of Pharmacy	
South Parkdale	3
Dovercourt-Wallace Emerson-Junction	2
York University Heights	2
Morningside	2
Mount Pleasant East	2
Runnymede-Bloor West Village	2
Islington-City Centre West	2
Tam O'Shanter-Sullivan	2
Englemount-Lawrence	2
Willowdale East	2

It shows that South Parkdale has most no. of Pharmacies as three followed by Dovercourt-Wallace Emerson-Junction, York University Heights, Morningside, and so on. There are around 42 neighborhoods without any Community Pharmacies. The different size green circles figure below shows that neighborhood admission ratio per average population of the neighborhood.

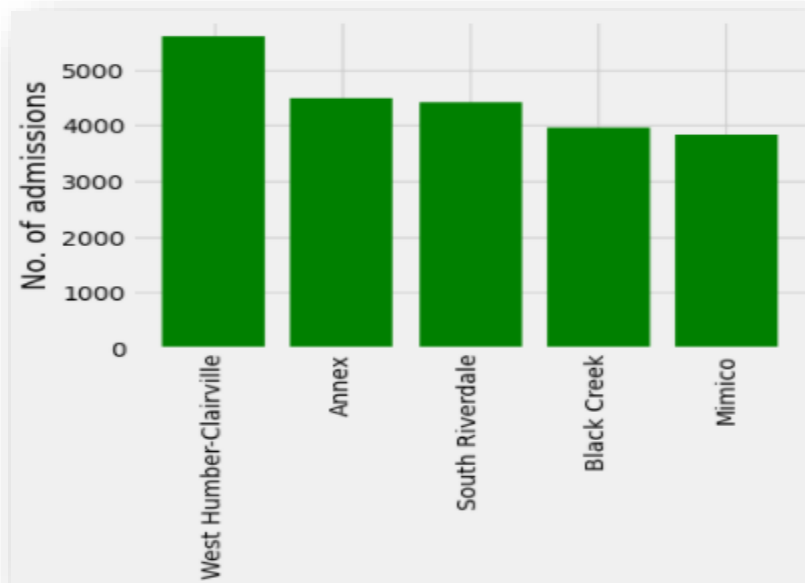


From the above visualization map, we came to know that there is lack of adequate number of Pharmacies near University of Toronto. The major neighborhoods include Annex, South-River Dale, Church-Yonge Corridor, etc.

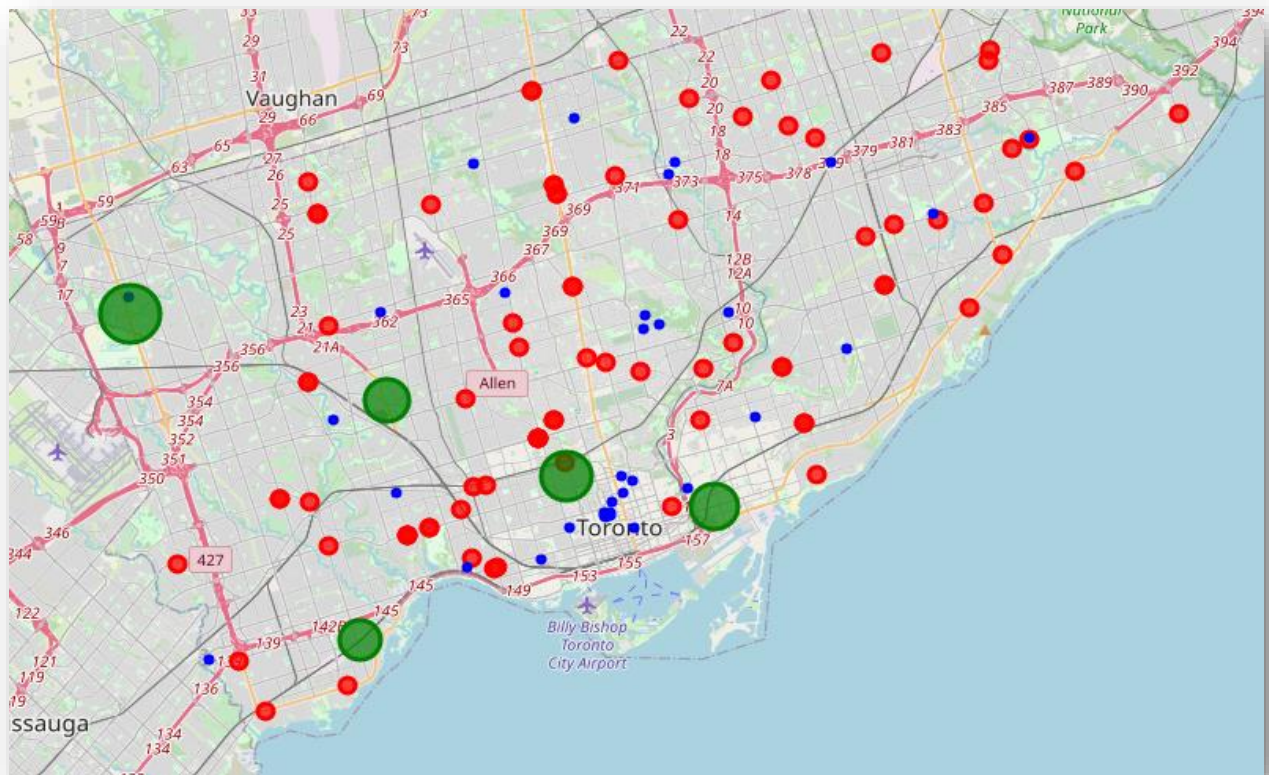




Again, when we inserted the hospital locations in the map, the major top hospitals lie in these areas. It has adequate hospitals, the rate of admission of patients seems maximum in these areas but it lacks adequate number of Pharmacies. Let's narrow down our results, the top neighborhood without any pharmacies but number of admission rate is shown by the following bar graph.



Let's visualize above neighborhoods in Map along with Pharmacies distribution and Hospitals



According to above graphs, with referencing our evaluation criteria, the top neighborhood to start to Pharmacy is in the following order:

- i. West Humber-Clairville
- ii. South River Dale
- iii. Black Creek

## Chapter 5: Conclusion

We have set following evaluation criteria

- i. No Pharmacy stores available.  
Less the number of pharmacies, less the number of competitions.
- ii. Number of Hospital around.  
If there are hospitals around, store is more likely to run at higher pace
- iii. Number of Hospital Admissions.  
In neighborhood where more people are admitted, the pharmacies is more likely to operate more because of higher demands of medications.
- iv. Population Size.  
Higher the population, larger the market size is.

Based upon our results and evaluation criteria, it seems there is high demand of Community Pharmacy in West Humber-Clairville. This neighborhood has the maximum admission rate, it has large number of populations, and within several miles there are no any Community Pharmacy around. So, there is less competition, high demand of drugs, and the market is large. We'll like to recommend South River Dale and Black Creek as honorable mentions which can be considered for future extensions.