**Capstone: The Battle of Neighborhoods**

**Data Description**

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.

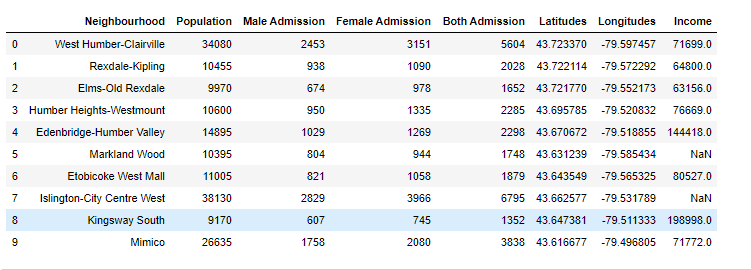
1. **Collections**

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

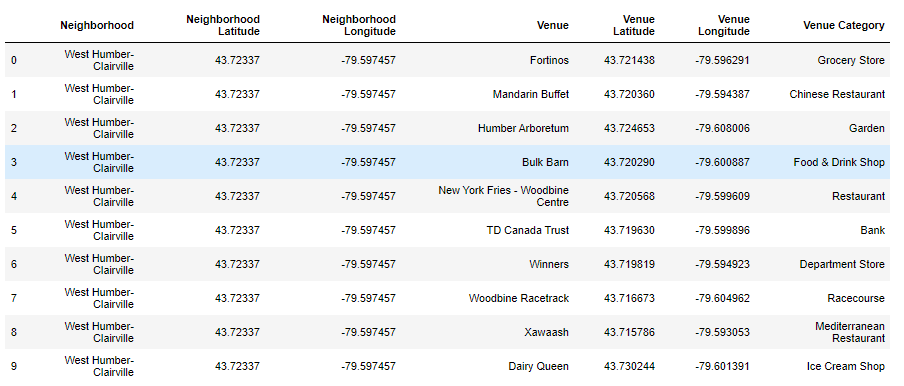
1. **Datasets**
2. Neighborhoods Data

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



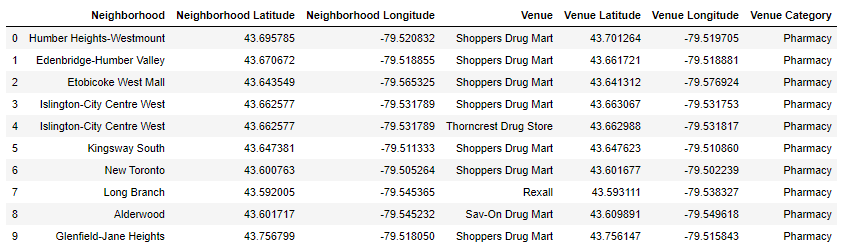
1. Toronto Venues Data

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



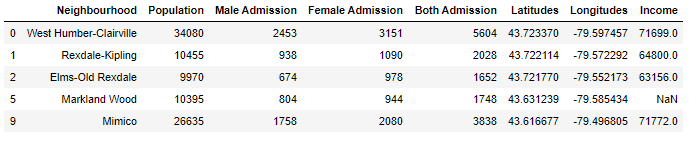
1. Toronto Pharmacy Data

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



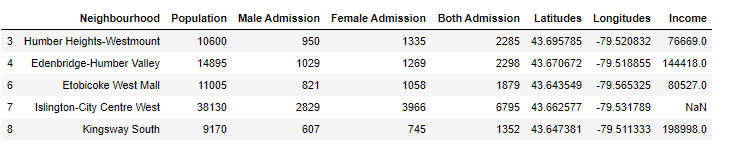
1. Neighborhoods without Pharmacies Data

It stores the information of neighborhoods without any retail Pharmacies.



1. Neighborhoods with Pharmacies Data

It stores the information of neighborhoods with retail Pharmacies.



1. **Conclusion**

We will analyze these data frames visually and try to identify the location for setting up the Pharmacy Store for Orange Drugs.

With all these features, techniques and data, we will then be able to come up with a best recommendation to the management of Orange Drugs to their problem which is where is the best neighborhood for them to first start off to offer their services. For an example, we will not want to enter a neighborhood whereby there is already a high concentration of Pharmacy available or there is a high trending of such stores upcoming in the neighborhood. We will like to recommend a neighborhood where we know that there will be a higher demand of such delivery service due to the lack of supply in that area.