**The battle of Toronto neighborhoods**

**April 1st, 2019**

1. **Introduction:** 
   1. **Background: Traveling this days can be an expensive hobbit during this times, but there is always a way to hack the expenses for the duration of the trip. One of the hacks is to stay in affordable accommodations like Hostels, it might be not for everyone but maybe you are one of those who love to travel and explore new experiences.** But, starting a Hostel can be an exciting and profitable venture for an entrepreneur with a passion and love for traveling and customer service.Toronto is one of the most popular and busiest Canada cities. A business man decides it would be a good business move to open a new Hostel in Toronto.
   2. **Problem:** Before opening a hostel, it would be obvious to analyze the possible factors that could aid or harm the business. In the scope of this project my goal is to analyze the best neighborhood in Toronto to open a new Hostel. The possible factors to be considered are: Which is the liveliest neighborhood amongst all neighborhoods? How many hostels and hotel are there in each neighborhood?
   3. **Interest:** The stakeholders involved in opening a new hostel would be interested in the analysis of the neighborhoods and the hostel competition in each neighborhoods.**The target audience is the young people who travel with low bugged and look for affordable locations to sleep over. The importance of the solution for this problem is because the market for low bugged travelers is increasing and Canada do not offer several options for this kind of customers.**

**2. Data Acquisition and Cleaning:**

**2.1. Data Sources: The data set relevant to this project were retrieve from several places like Wikipedia, Foursquare Api, and Web Scraping.**

**2.2.** Data **Cleaning:** The data scraped from the web was converted into a data frame. There were some rows with information not need it for our project that were dropped. Some data frame were merged to get the final data frame.

**3. Methodology:**

To get a soluction for the problem we are going to use a Unsupervised Classification Task, the machine learning Algorithm will be K-means Clustering.

**4. Exploratory data analysis:**

**Exploring the most popular neighborhood in Toronto City Downtown:**  It would be ideal to analyze the number of total venues and the number of total hostel’s and hotel’s in the neighborhoods located in Downtown area.

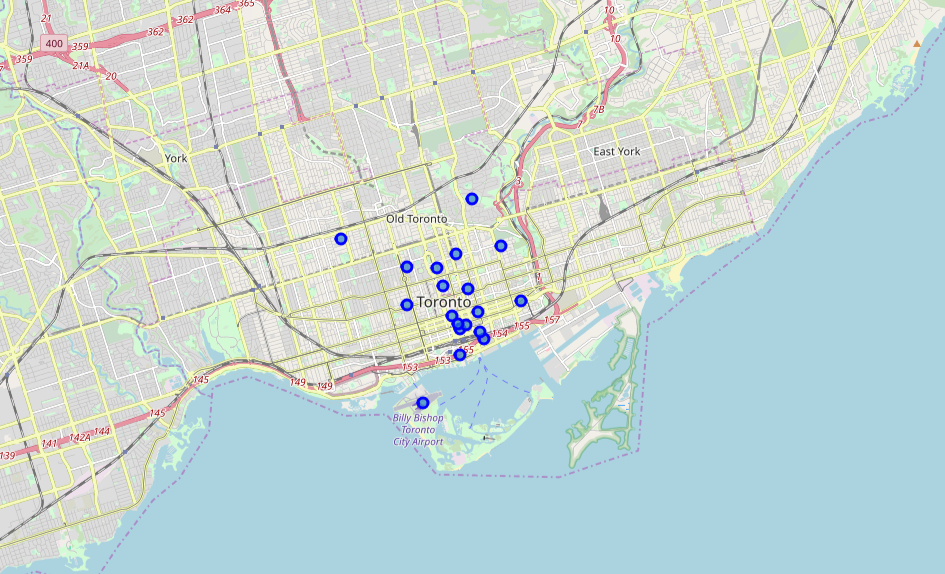


Fig 1: Map indicating the location of Toronto downtown neighborhoods

For the Downtown area I get 19 neighborhoods to be analyst as we can see in the next table.

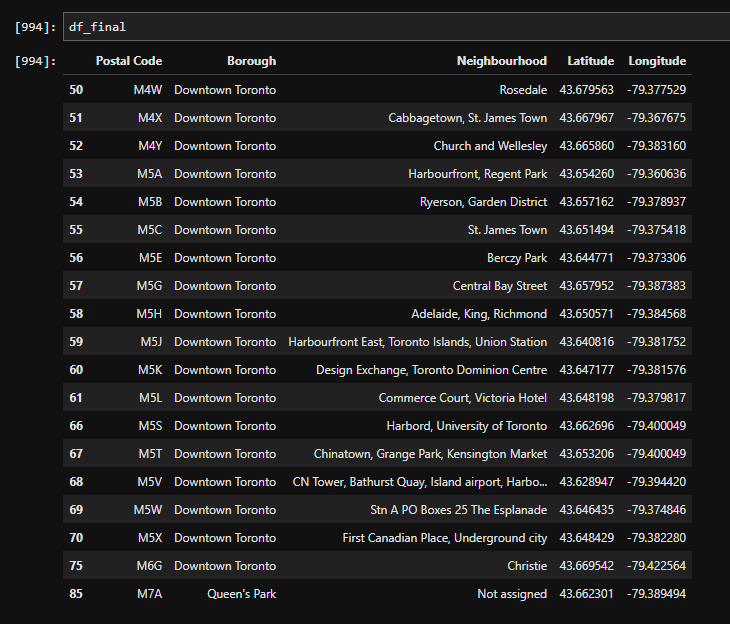


Fig 2: Table with Toronto downtown neighborhoods.

We got 211 uniques categories, but we are not going to use all of them for our Project.

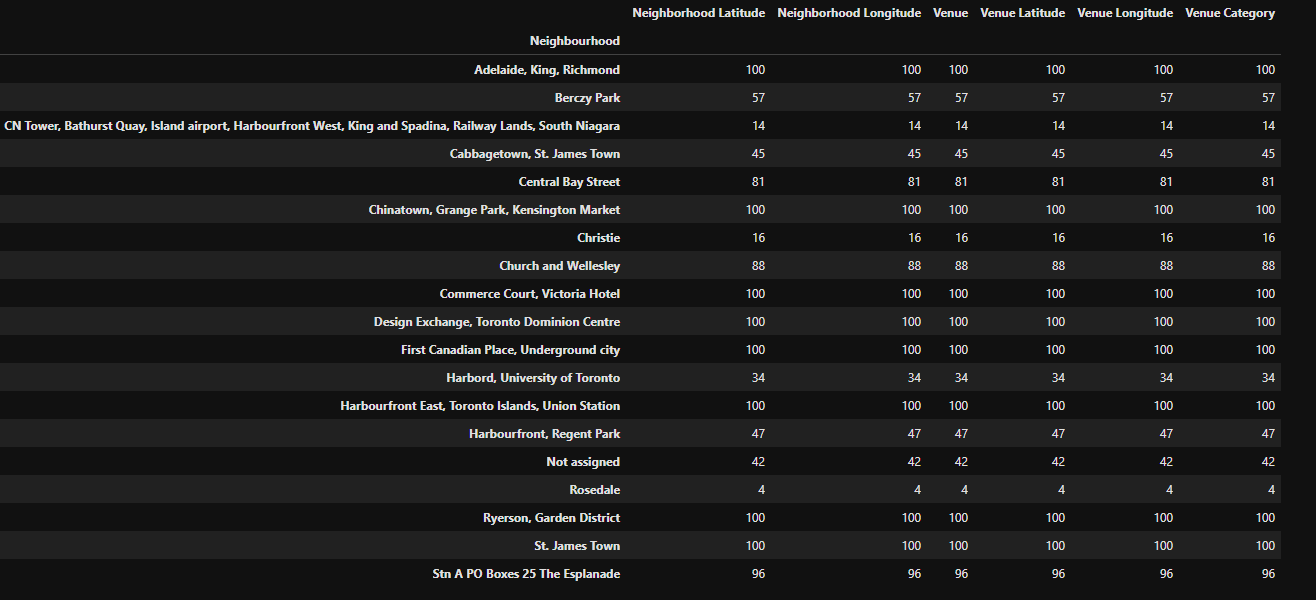


Fig 3: Quantity of Venue Category on each neighborhood

The retrieve data from Foursquare Api show us that there is only two hostels in Downtown Area on the neighborhoods of St James Town and The Esplanade, it looks like a great opportunity to open hostels in the area because there is not many competition.

Quantity of Hostels per Neighborhood

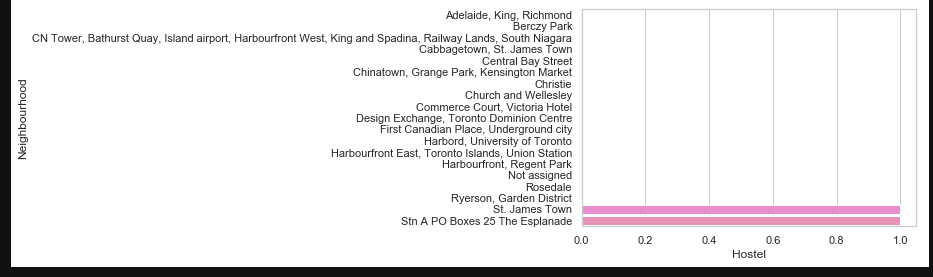


Fig 4: Bar Chart of Hostels count per neighborhood vs Quantity.

Hotels are more populated in the area as we can see in the next bar chart, it could be a good option to open hostels in this neighborhoods.

Quantity of Hotels per Neighborhood

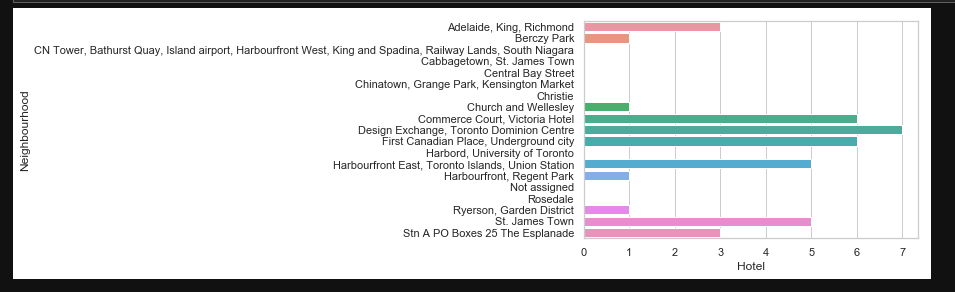


Fig 5: Bar Chart of Hotels count per neighborhood vs Quantity.

5. Results

After applying the Machine Learning Algorithm we got five Cluster in Downtown area as we can see in the next image

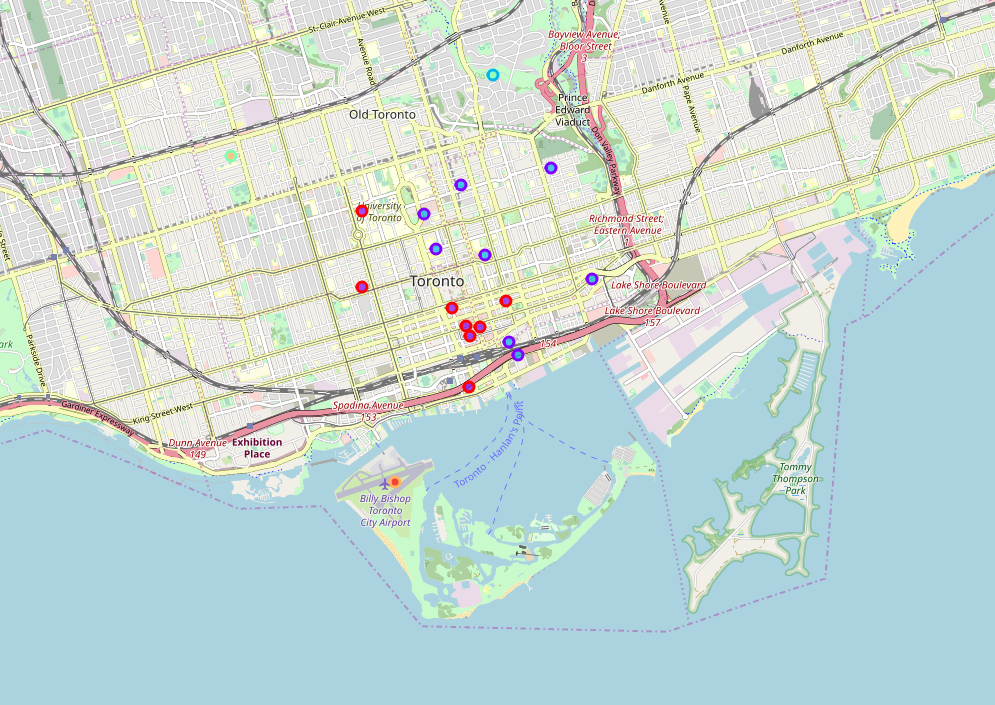


Fig 6: Clusters in Downtown Area.