**Toronto Neighborhoods and its Latin American Areas**

**Alexandra Butke**

**August 5th, 2020**

1. **Introduction**
   1. **Background**

The Latin culture and people are a fast and growing population that brings beauty, knowledge and tradition to every country they are in. While the United States is the mecca for this community, the Canada Latin population is also rapidly growing.

According to 2016 Statistics the highest concentration of Latin American’s resides in Ontario Canada, with the majority of them located in Toronto. Toronto sits at the edge of the United States, which is ideal for immigrants looking to have a natural transition. There are approximately 73K Spanish speaking people living in Toronto.

This project is catered to Latinos who are seeking to immigrate to Canada. Leveraging data from Foursquare, Geopy and Wikipedia, we will apply data science and python to find Latin Venues and explore the various neighborhoods in Toronto.

* 1. **Problem**

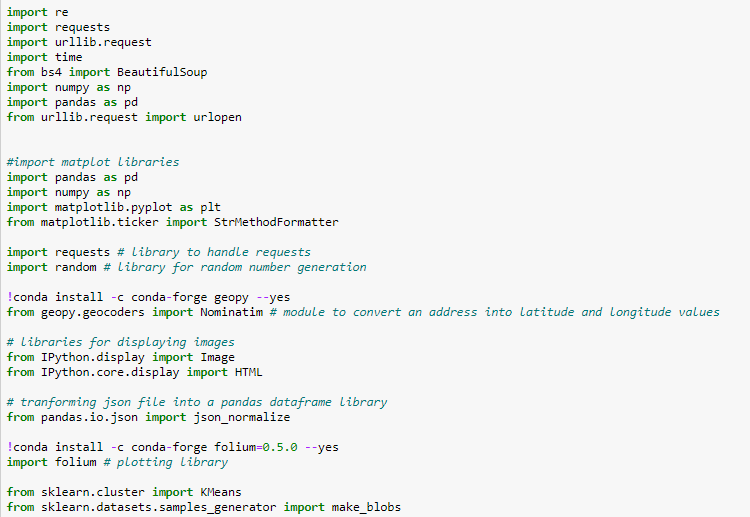
While, the Latin American Population in Canada is growing, it is still at the lower end of the overall population. Without data and research, it might be difficult for a Latin Person, to target areas in which they might be able to easily matriculate into. Using the Applied Data Science tools and skills we learned, this research aims to provide insight into where the largest Latin Population in Canada is, where to find venues that cater to Latin’s and a general understanding of the overall areas’ venues/attractions.

* 1. **Interest**

Canada’s proximity to the United States makes an ideal country for those looking to visit or immigrate to Canada. In addition, Canadian leaders looking to attract Latin Tourist, could garner insight for potential investment areas.

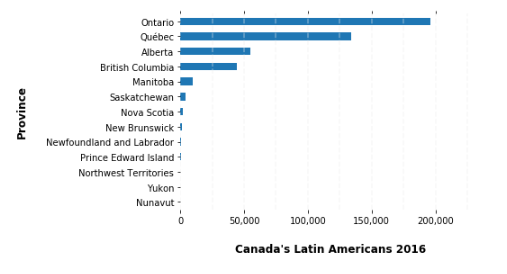
1. **Data and Resources**
   1. **Links**

* Latin Demographic Data Link: <https://en.wikipedia.org/wiki/Latin_American_Canadians>
* Location Data: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
  1. **Resources**
* **Foursquare.com** database to explore and search various venues and to compare neighborhoods in Toronto
* **BeautifulSoup** Python package for parsing HTML documents
* **Geopy** to locate coordinates of Toronto Postal Codes
* **Folium** for mapping visualization
* **matplotlib** to create a barchart for visualization
* from **sklearn.cluster** import **KMEANS** to explore and analyze the Neighborhoods
  1. **Python Libraries**



1. **Methodology**
   1. **Methodology Section 1:**

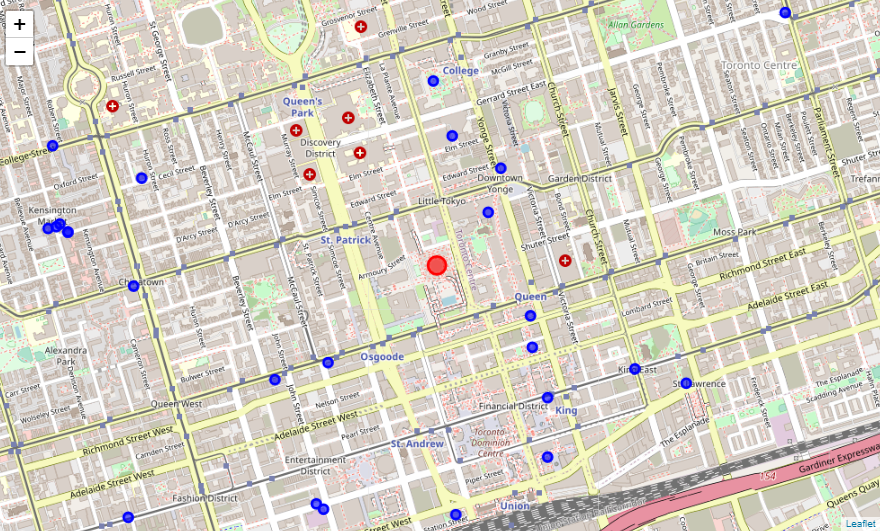
To understand where in Canada the Latin Population resides, data was downloaded from Wikipedia <https://en.wikipedia.org/wiki/Latin_American_Canadians>. **BeautifulSoup** Python Package was then used to scrape and parse the data. The parsed data was then cleaned to create a pandas dataframe. **Matplotlib** barchart was then used to illustrate the 2016 Latin American Canadian Population in descending order (largest to smallest). Using this illustration, the analysis clearly shows that most Latin American's reside in four Canadian Provinces: Ontario, Quebec, Alberta, and British Columbia.



* 1. **Methodology Section 2:**

According to Canada's 2016 census Ontario (see above bar chart) has the largest population at 196K. This would make Ontario an ideal location for Latinos seeking to immigrate to Canada. Toronto is the provincial capital of Ontario and boast one of the largest Latin population at ~133K. While, Latin American’s are the minority of the population, Toronto has many venues, restaurants and localities that cater to the Latin Demographic.

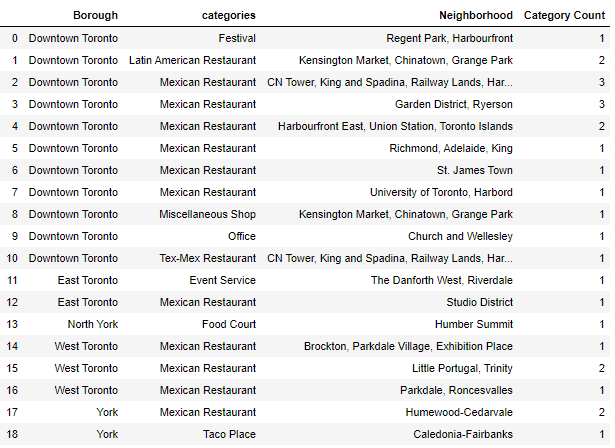
**Foursquare API** search method is applied to find the Latin Venues. I limited my search to two key words “Latin” and “Mexican”, which allowed me to be more targeted in the venues I wanted to pull. I searched a 5000 radius to capture as many neighborhoods as possible and limited my search to 150 venues. **Geolocator** was used to get the locations latitude and longitude. I conducted two searches for venues the first search was for any venue with the word “Latin”, and create a Latin\_DF dataframe. I then conducted a second search for the word “Mexican” and created a Latin\_DF1 dataframe. I combined both dataframes (Latin\_DF2) using python append functionality. Use a Python **def** function to filter through the data, clean and wrangle it. **Folium map** was used to show the Center of Toronto (red filled circle) and all the venues that cater to Latin’s (blue filled circle)

**Where to find Latin Venues**

* 1. **Methodology Section 3:**

Downtown Toronto is a hotspot for those seeking to find a variety of Latin venues. The Latin Sparks Festival which started in 2012 and considered Canada’s largest block parties is held in Downtown Toronto. If you are looking for a Mexican Style Restaurant the Garden District and Kensington Market Neighborhoods, both in Downtown Toronto, are excellent and have a variety of choices for you.

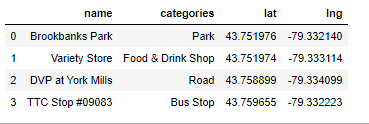
**BeautifulSoup** was used to parse and scrape the Wikipedia <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> , to get the postal codes, boroughs and neighborhoods and create a dataframe. **Geopy** was used to create the latitude and longitude dataframe. Then the two dataframes where merge to capture: Postal Code, Boroughs, Neighborhood, Latitude and Longitude. Pythons **Groupby** functionality was used to get the count by venue category. The resulting dataframe you can deduce that Downtown Toronto is a hot spot.



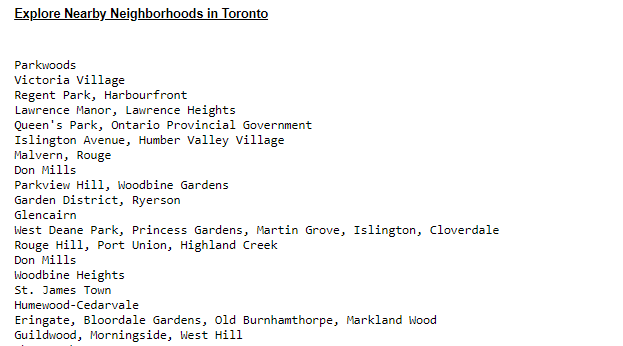
* 1. **Methodology Section 4:**

Toronto has many neighborhoods and has over 7K venues of which a total 264 are unique categories. There are many selections for eating, shopping or entertainment. Using **Foursquare, K-MEANs** and cluster we will explore these neighborhoods and the variety of venues.

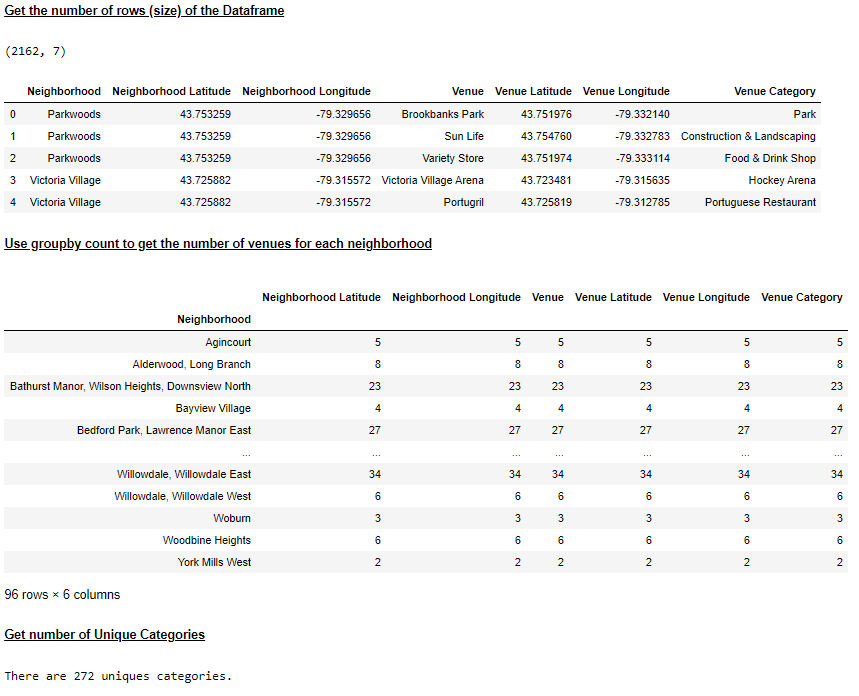
**Foursquare** explore methodology and Toronto Postal Code Longitude and Latitude data was used to explore the neighborhoods. Exploring the first neighborhood you can see the Parkwoods is the first neighborhood, then use python def function to extract the categories of the venues and create the dataframe.



Another python **def** function was used to iterate through the neighborhoods to compile and explore nearby venues.

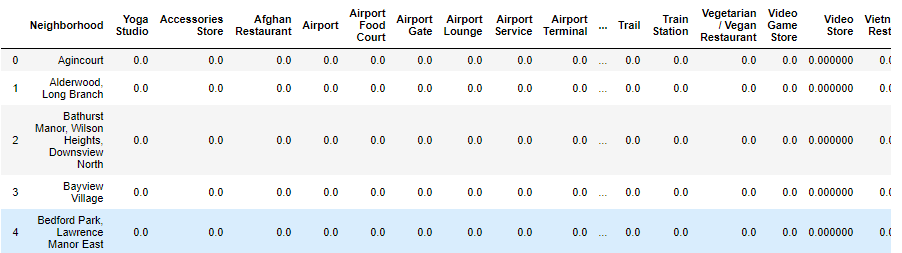


You can get the number of venues in the dataframe by using the **.shape** python function. You can see that there are 2,162 venues. Applying **groupby count()** function you can create a dataframe to get the category count. Using the **unique** python function, you can see that there are 275 unique categories of venues.

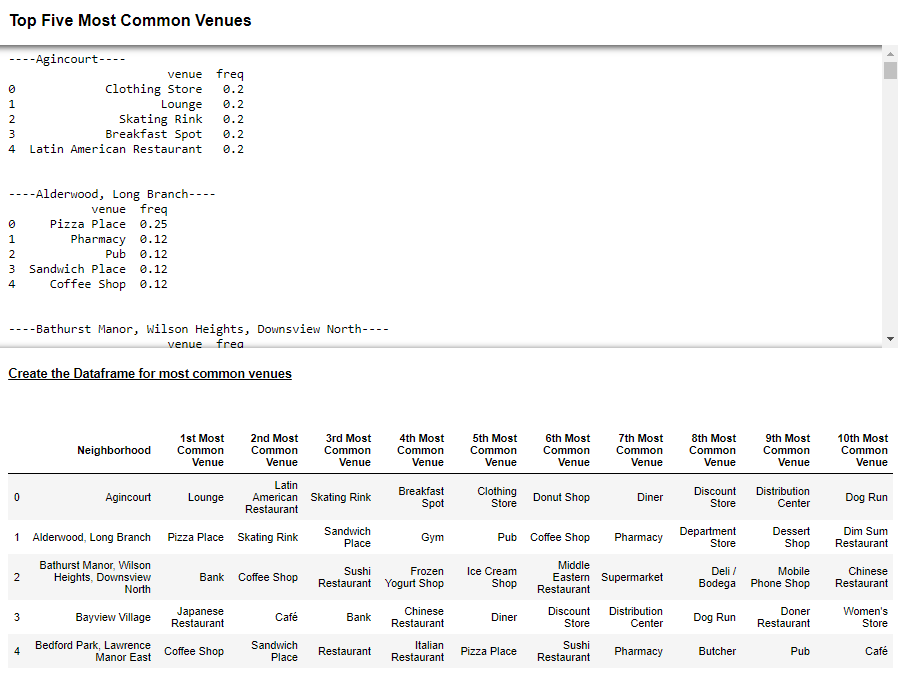


* 1. **Methodology Section 5:**

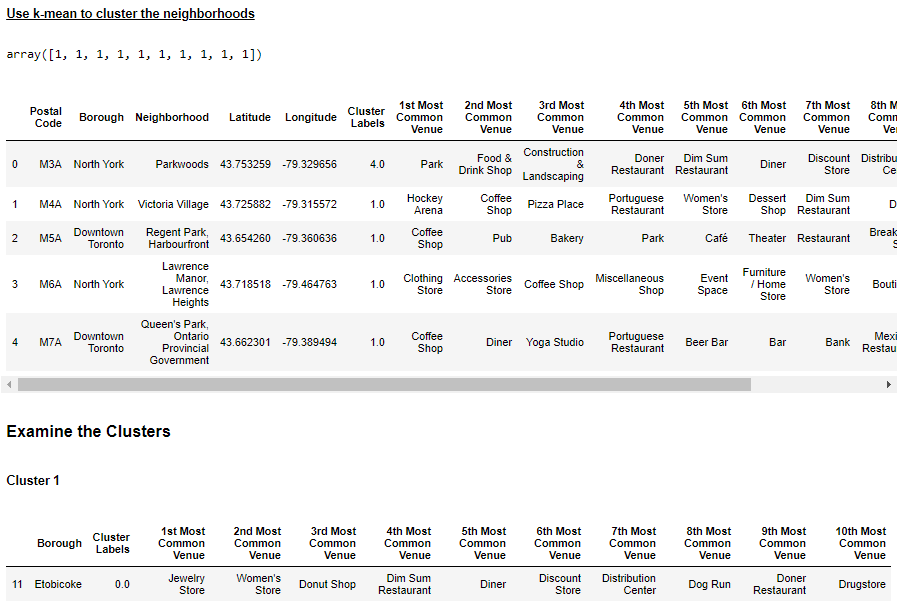
To Analyze the neighborhoods, I leveraged **K-mean** and **sklearn.cluster** methodology. **Onehot** learning was applied to get dummy venue categories for each neighborhood.

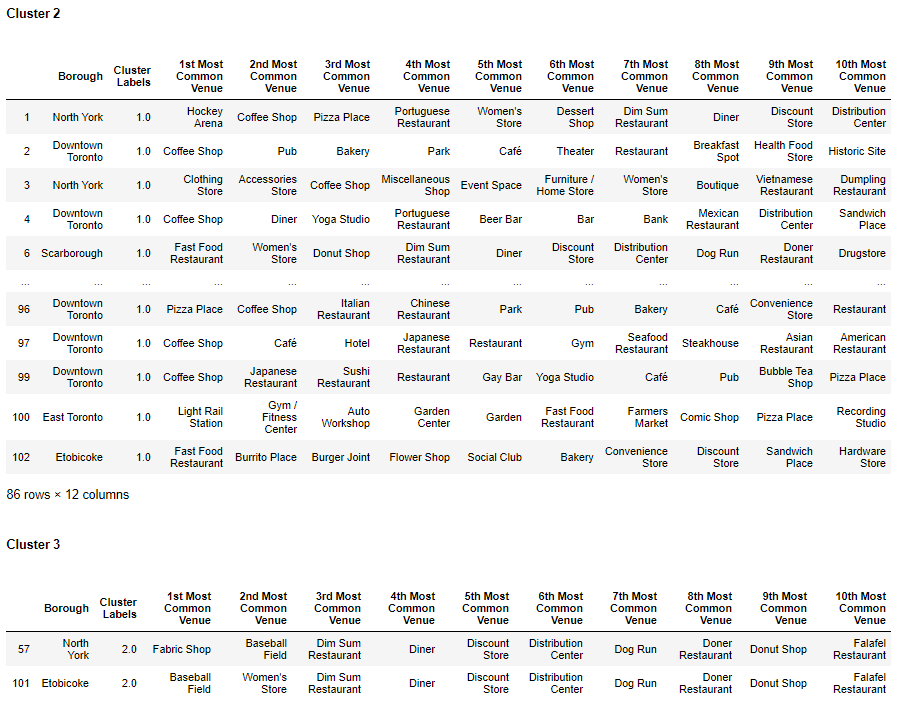


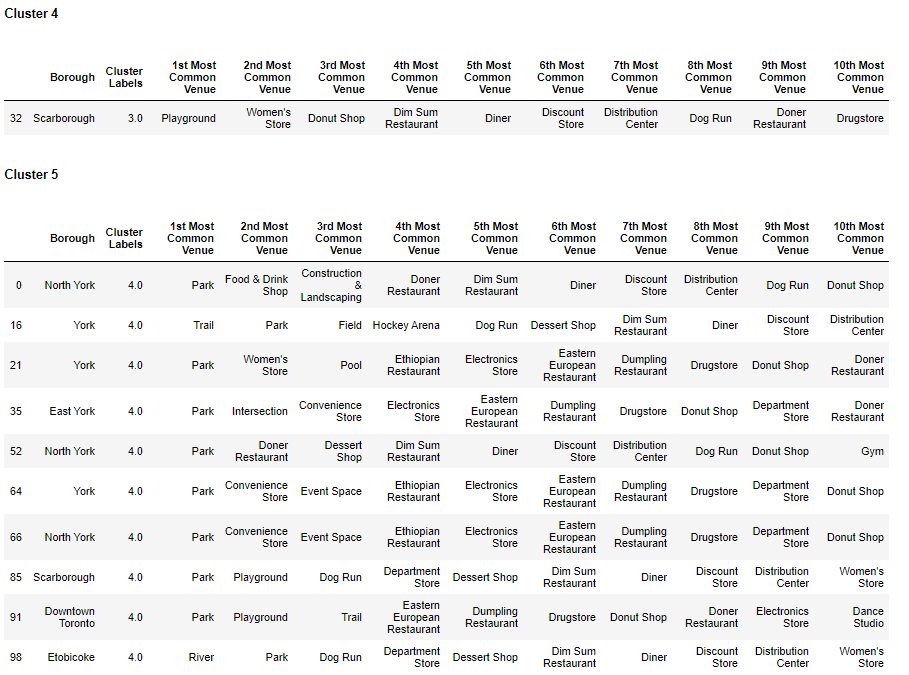
You can **groupby** top 5 venues by filtering through the dataframe above using a python **forloop** and getting the venue frequency for the various neighborhoods. Then clean it by creating a **def** function to loop through the common venues and list the 10 Most Common Venues,



**K-Means** to **Cluster** the neighborhoods is used to examine the neighborhoods further.







# ****Results and Discussion****

Toronto, the provincial capital of Ontario is multi-culture. The Latin American Community is rapidly growing. Using Foursquare, BeautifulSoup, Geopy, matplotlib k-mean, clustering and Folium you can wrangle and visualize the Latin Demographics, Latin Venues and overall Toronto Neighborhoods.

Methodology section 1 results: The BeautifulSoup package is used to create a dataframe of the wikipage “Latin American Canadians” population. The program extracts the backend table class from the wikipage for population, enabling data wrangling and clean-up to create a dataframe. Then visualize the data using matplotlib. Though through the dataframe alone you can quickly ascertain that Ontario has the largest Latin American Population, matplotlib bar chart does a better job with illustrating the difference for the consumer of this analysis. Methodology section 2 results: Using foursquare, geopy and folium together to conduct a search of venues within a 5000 radius Toronto then to represent these locations on a folium map was amazing. The consumers of this article will learn that Kensington Neighborhood has cluster of venues that cater to Latin Americans they can also quickly view where in Toronto to find the various locations.

Methodology section 3 results: This section foursquare and geopy is used to categorize the Latin venues and provide a quick dataframe containing the number of Latin Venues for each Neighborhood and Borough. While, the grid is limited by the key search words applied, I could have created a for-loop to provide a list of key words and broadened my search. In this example I chose to keep it simple, and search for key words Latin and Mexican.

Methodology section 4 and 5 results: This was the most complex section in my learning. In this section I applied Beautiful Soup, k-mean, clustering, and geopy to conduct an analysis about Toronto. This analysis was done to a broader analysis of what Toronto has to offer. In evaluating the most common venues, I that Latin American Restaurants ranked as the second most common venues in the Agincourt Neighborhood, located in the Borough of Scarborough which was just outside my radius search.

Using the explore method in foursquare allowed me to find other areas outside the radius of my original search in Methodology section 2. The explore foursquare method enable me to go through each neighborhood in Toronto and exploring a 750 radius through each. This allowed me to discover more areas that might cater to the Latin Community. In addition, provides views of other areas of interest.

K-means algorithm to partition the venue categories, you can examine the specified 5 clusters. Using the cluster approach, you can confirm that Downtown Toronto does have many Mexican Restaurants, as Mexican Restaurant appeared as the 8th most common venue in cluster 2.

# ****Conclusion****

As a Latina who has never visited Canada, I found the results of this research very interesting. The findings will enable other Latin American’s to target areas in Toronto that might cater to their demographic and will also provide other areas of interest to them.

From a Decision Scientist perspective, I see the value in leveraging the many tools available in developing your analysis and putting your story together. I would say that k-mean, clustering, beautifulsoup and folium are tools I now have in my skills toolkit that will be valuable in my work.