**Final Report**

**Applied Data Science Capstone Project - The Battle of The Neighborhoods.**

**Growth in the Restaurant Business.**

**Alberta canada**

**Introduction**

Alberta is a province in western Canada. It's one of the 3 Prairie Provinces in Canada bordered by British Columbia to the west, Saskatchewan to the east, the Northwest Territories to the north and the U.S. state Montana to the south. The capital of Alberta is Edmonton, located near the center of the province and 180 miles north of Calgary, the largest city of Alberta.

Alberta is the 4th most populous province of Canada and the most populous of the prairie provinces. Alberta has an estimated population of 4.08 million, up from 3.645 million in 2011. The province has a population density of 6 people per square kilometer, or 15 people per square mile. To read more about Alberta demography go to [Alberta Population 2020](https://worldpopulationreview.com/canadian-provinces/alberta-population)

Alberta has a unique and vibrant culture that has been largely shaped by immigrants who have settled in the province. Throughout the province, but especially in the major cities of Calgary and Edmonton, one can find the culture, cuisine, music and arts of communities from all over the world.

For more on Alberta’s culture go to [https://www.canadavisa.com](https://www.canadavisa.com/about-alberta.html#gs.mzxesd)

Alberta has a rugged and very beautiful landscape. The population is ethnically diverse. This project aims to take a trip to Alberta Canada by means of Data Science.

**Business problem**

A business problem is always a business opportunity. Therefore the goal of

the project is to decide if there is an opportunity to open more restaurants in Alberta canada?

We know when people get together the best conversations take place around foods and drinks. Currently with the presence of Covid-19, life is about being alone or being in quarantine and that goes for nearly the whole wide world.

However we know at one point in the near future normalcy will return. With that in mind, economies will start growing again. People will want to get together to eat, to drink, to laugh and to have a good time at restaurants and other venues.

I aim to explore the province of Alberta in Canada to see if there is room to open a few more restaurants.

**Description of the Data**

**In term of projecting how the work will flow:**

Dataframes will be created based on the province Alberta Canada.

Geographical coordinates are essential in terms of Latitudes and Longitudes.

Exploring nearby places in different neighborhoods to compare venues, segmenting them to understand means of opportunities.

**Libraries that are likely to be used for the project.**

Pandas - to create and manipulate the dataframes.

Beautiful Soup and Requests - to scrap and library to handle http requests.

Geocoder - will be used to retrieve geographical coordinates.

Folium - is for data visualization with interactive mapping.

Matplotlib - for plotting modules.

JSON - to handle JSON files.

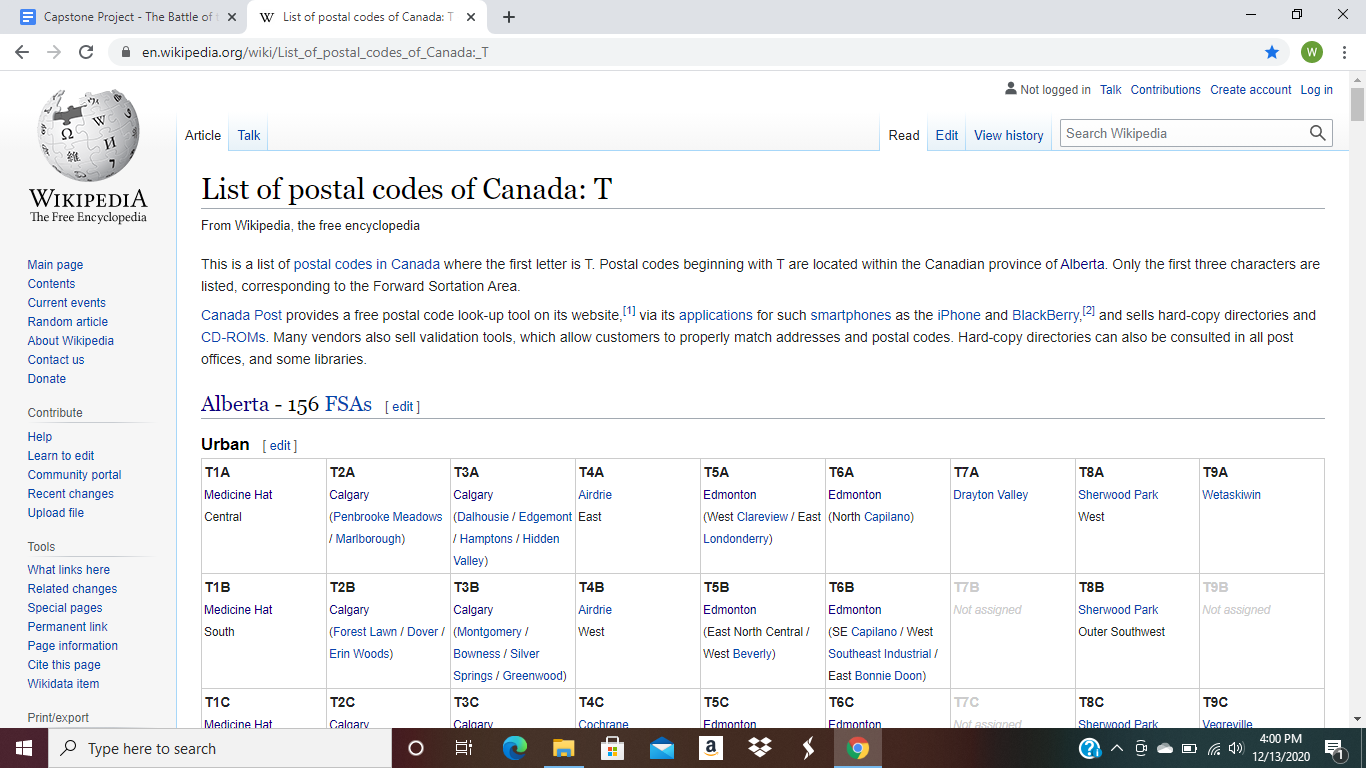
Scikit Learn - for importing k-means clustering.

Notes: this is only a projection. Other libraries such as Numpy, XML, Geopy will be installed and imported if necessary.

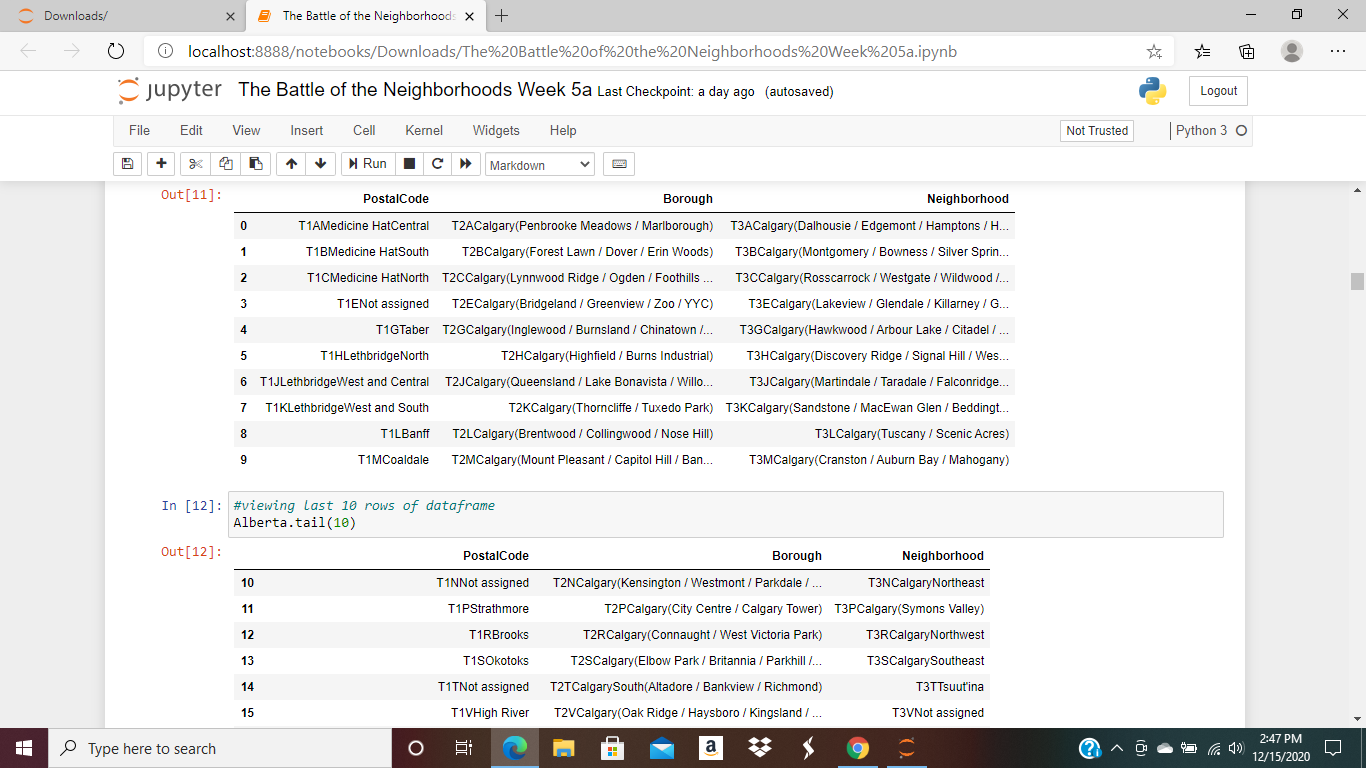
**Methodology**

The data is where it all begins. Without data we cannot perform data analysis. Finding the right data source can be a most difficult challenge. On week number 3 during this course we worked with data on New York City. As a New Yorker who loves the city I thought that was awesome. Next we worked with Toronto Canada. Now I would like to explore Canada further by working with the province of Alberta. It is not as populated as NYC but we shall see how it goes.

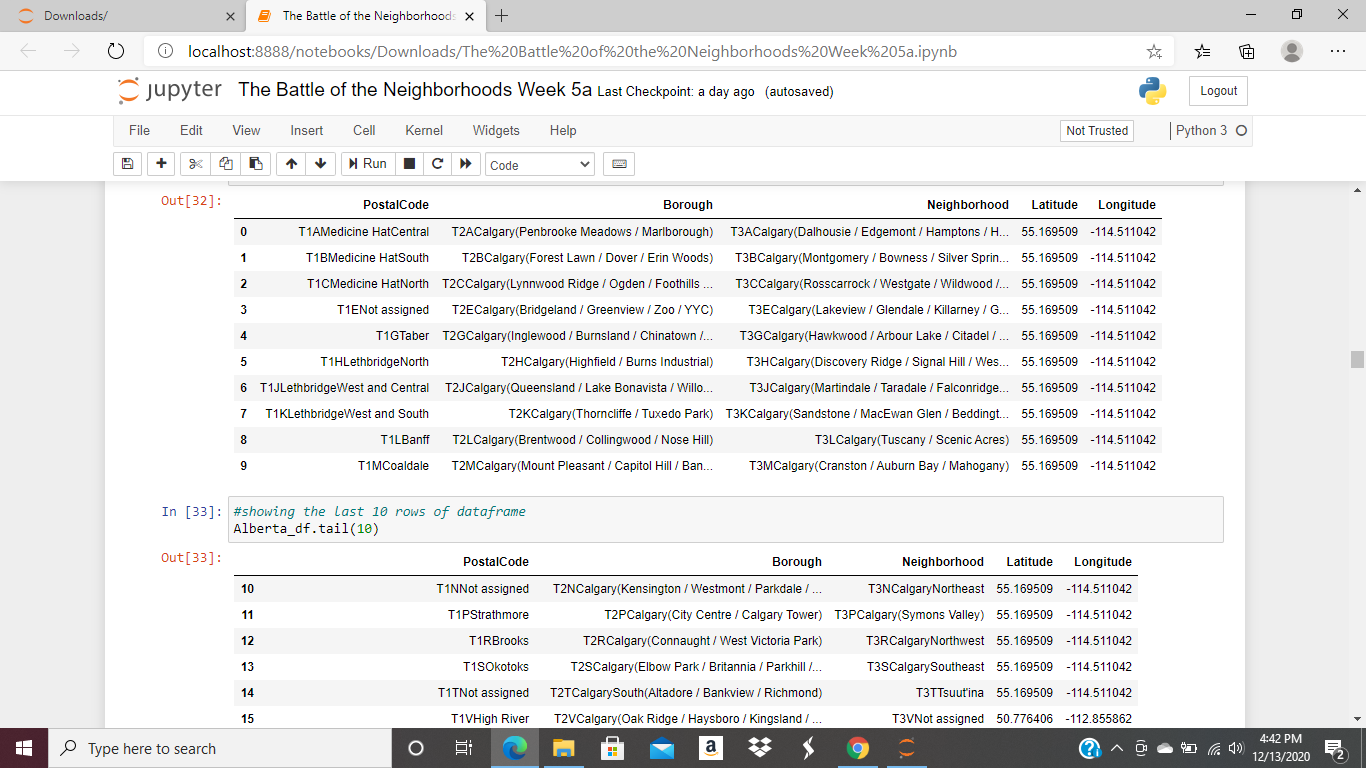
From Wikipedia [List of postal codes of Canada: T](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T) stands for Alberta, province of Canada which are categorized by postal codes.



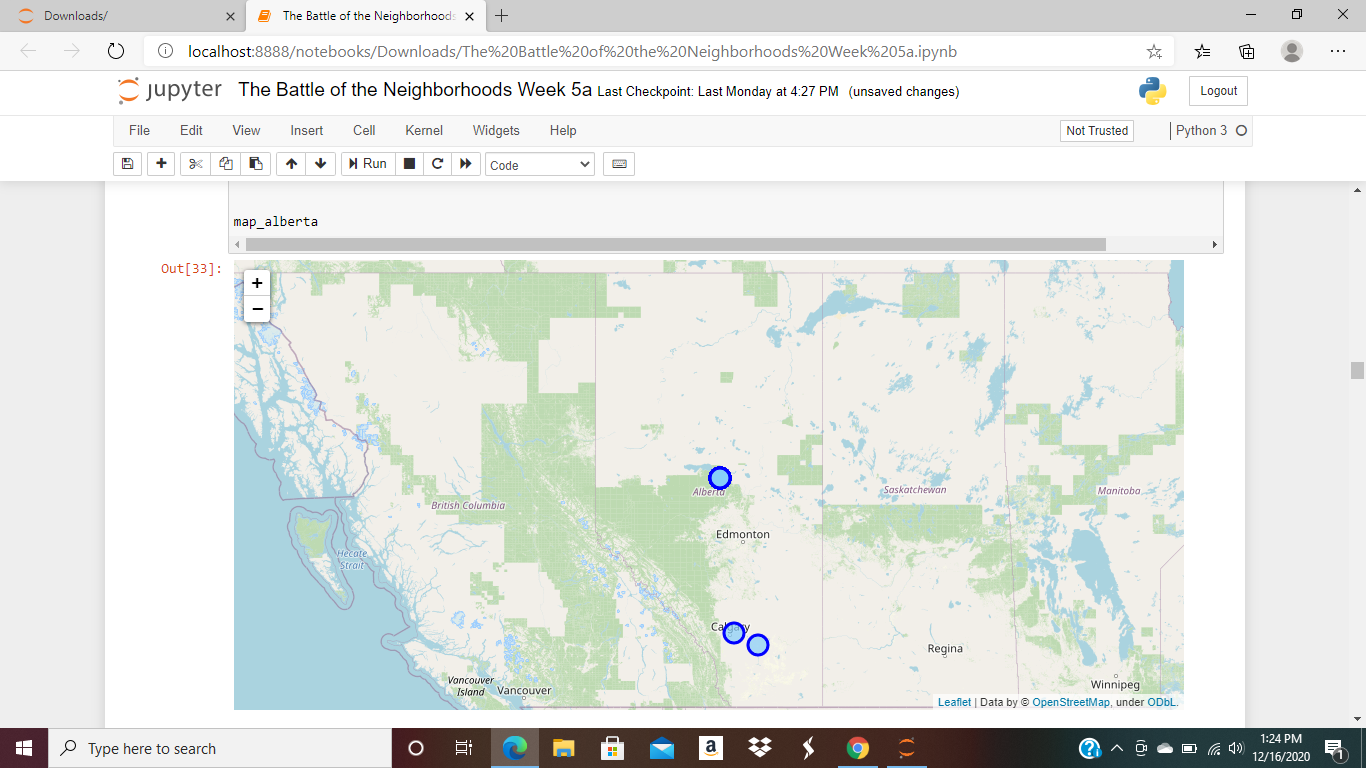
I scraped the data by means of Beautiful Soup and Requests to create a Pandas dataframe. After formatting and cleaning the data, the result is a dataframe with a focus on populated areas of Alberta, a province of Canada resulting in 20 boroughs and 20 neighborhoods.



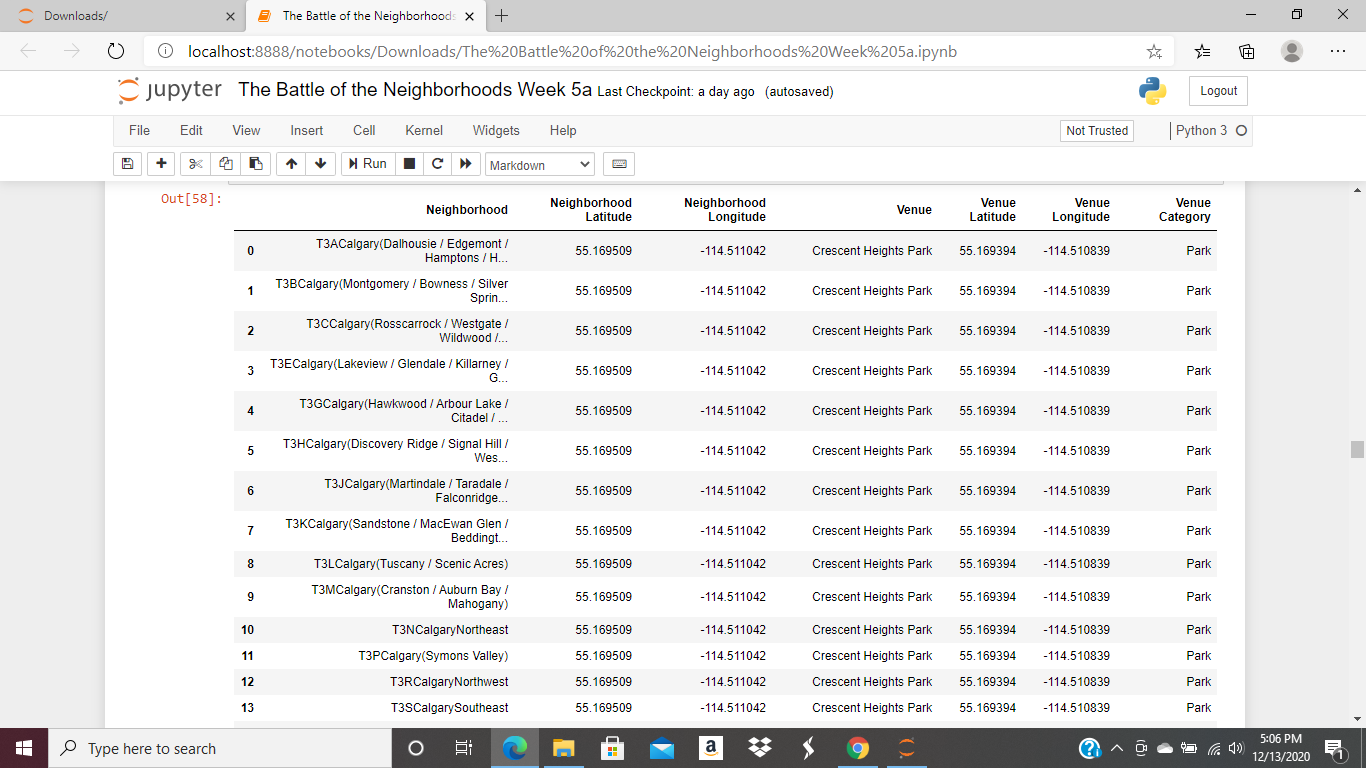
The nominatim function of Geopy added geospatial coordinates of latitude and longitude to the dataframe. The merger consists of PostalCode, Borough, Neighborhood, Latitude and Longitude. The shape is 20 rows and 5 columns. I can observe that many of the neighborhoods share common latitudes and longitudes.



Once the Geospatial coordinates were added, the opportunity to show a map of Alberta Canada presented itself. The folium package was used to show the 3 distinct sets of latitudes and longitudes that are present based on the data scraped and formatted from Wikipedia



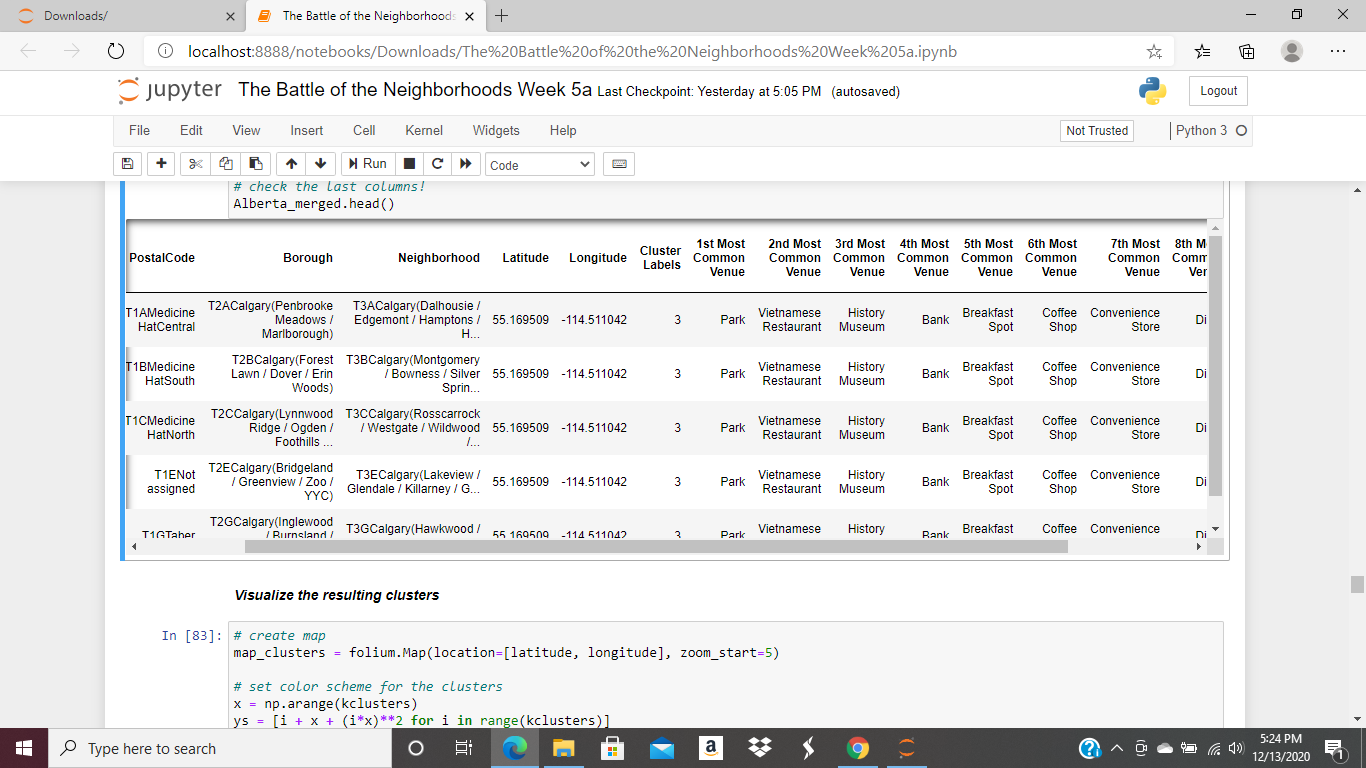
The Foursquare API locational platform gathered information about venues inside each and every neighborhood. The type of venues it rendered contained information such as the neighborhood, the neighborhood’s latitude and longitude, the venue, the venue’s category, the venue’s latitude and the venue’s longitude. A total of 66 venues were accounted for.



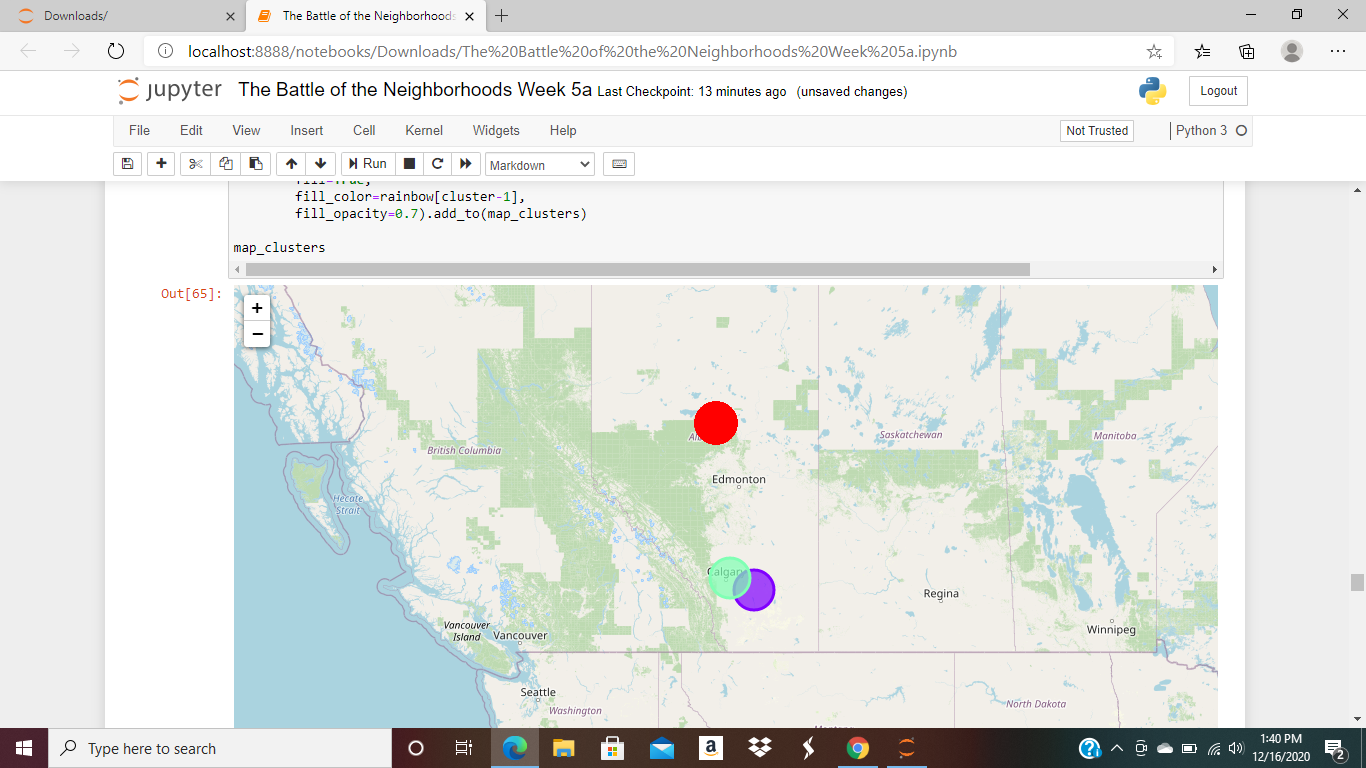
Unsupervised machine learning algorithm approach call K-means Clustering segmented and

grouped the neighborhoods into clusters. After a few testing I settled on K=3.

We have the most common venues from one to ten with 3 cluster labels.

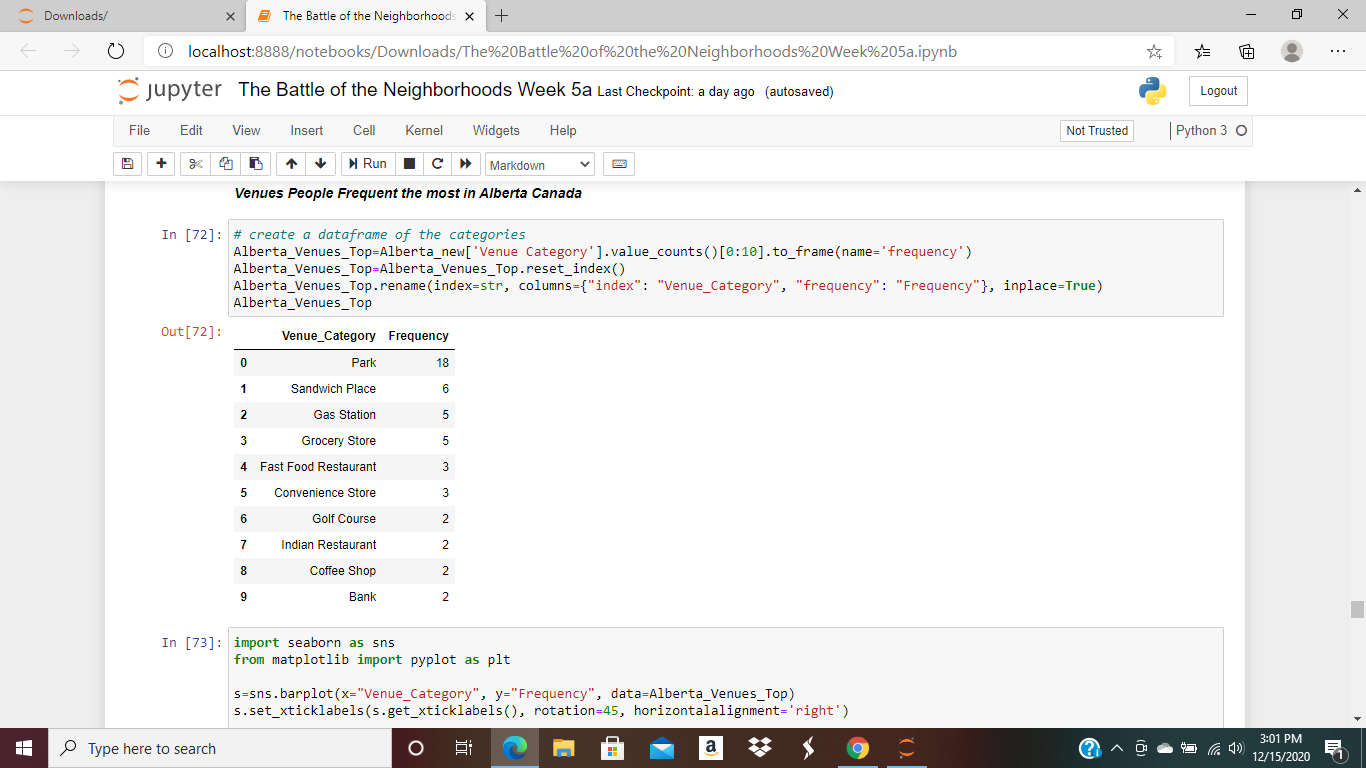


A map with the clusters is shown below:

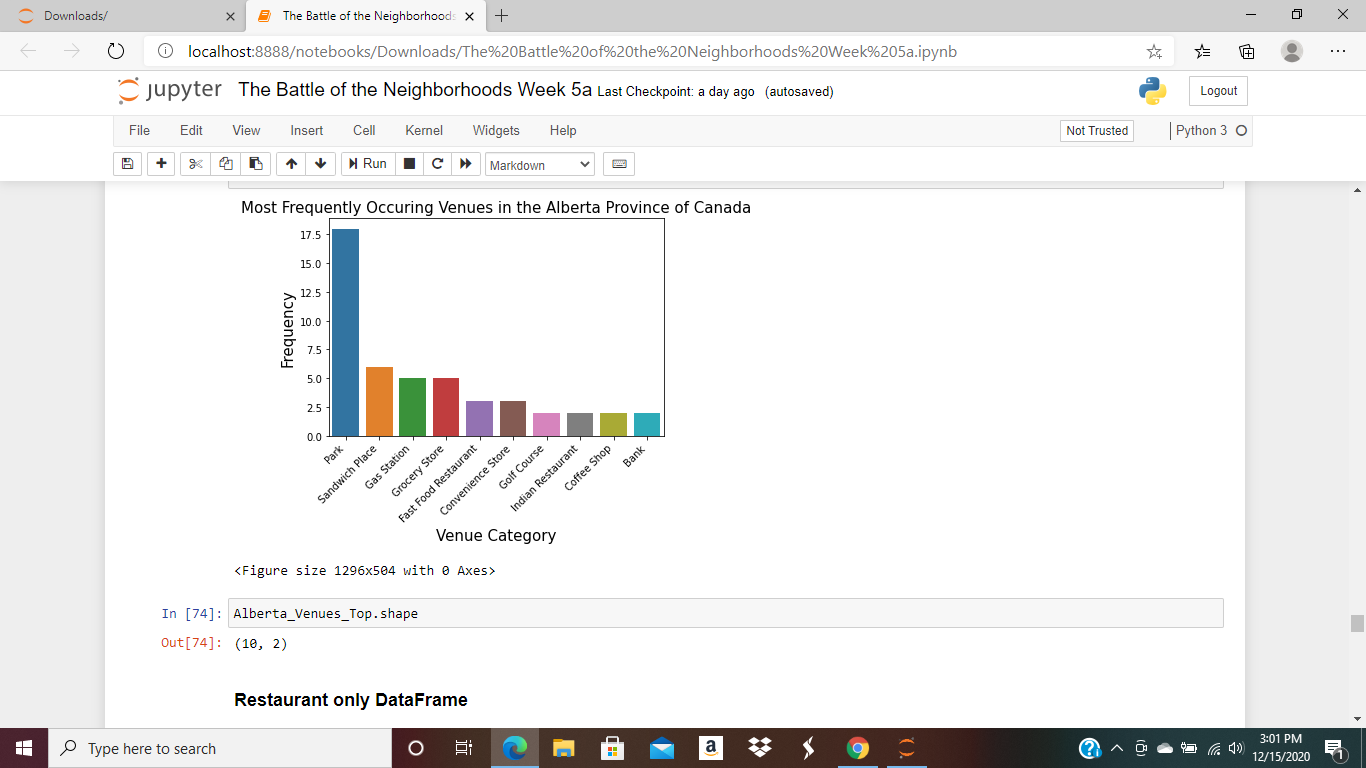


**Results**

To explore the province of Alberta Canada closer I created a dataframe of the venues people frequented the most. It helps to give an understanding of what is trending culturally since we are using data science to take a trip to Alberta.



A bar chart is plotted of the venues that are frequented the most and parks are the most frequently occurring venues as shown below

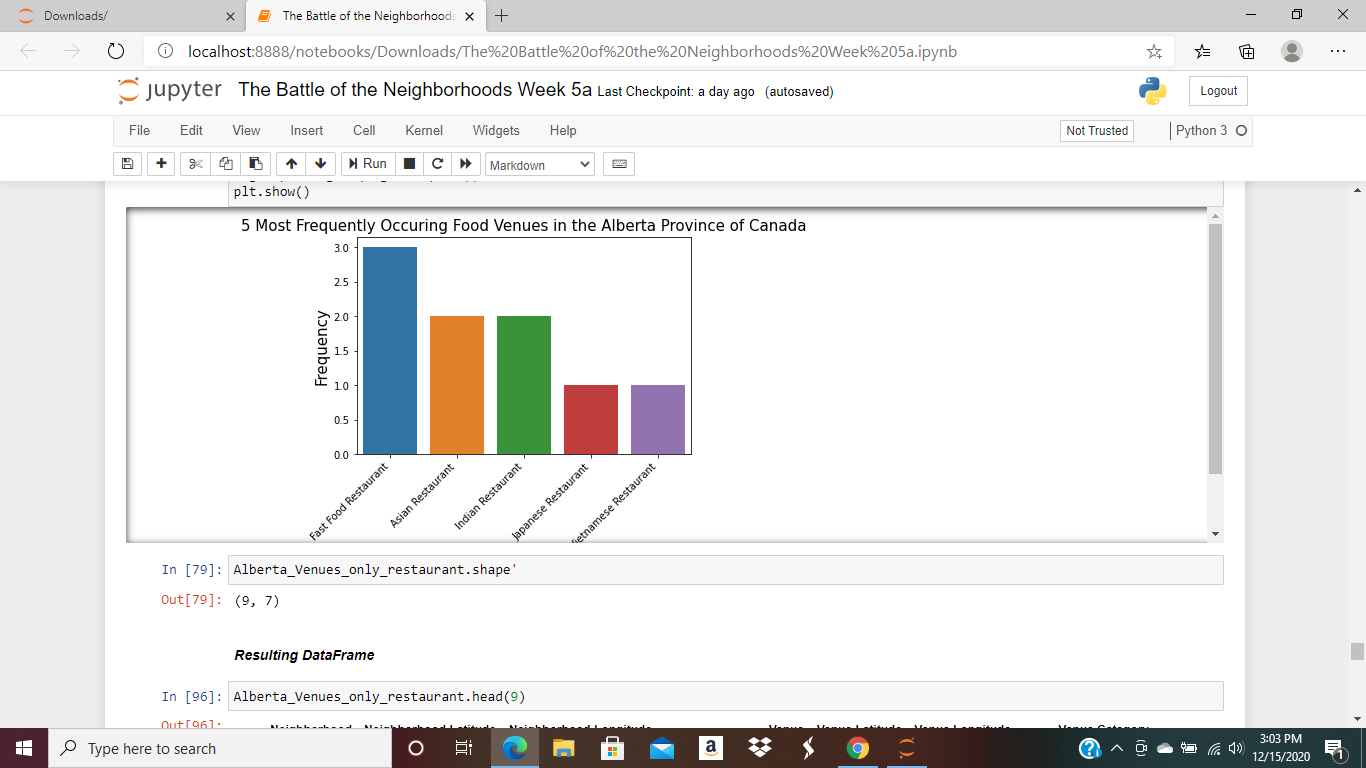


Why so many parks you might ask ? From the Foursquare Api the count is 18 parks. Calgary hosted a Winter olympic. The parks are great venues for both summer and winter outdoor activities. Although the province is vast, it’s population is less than half of that of NYC.

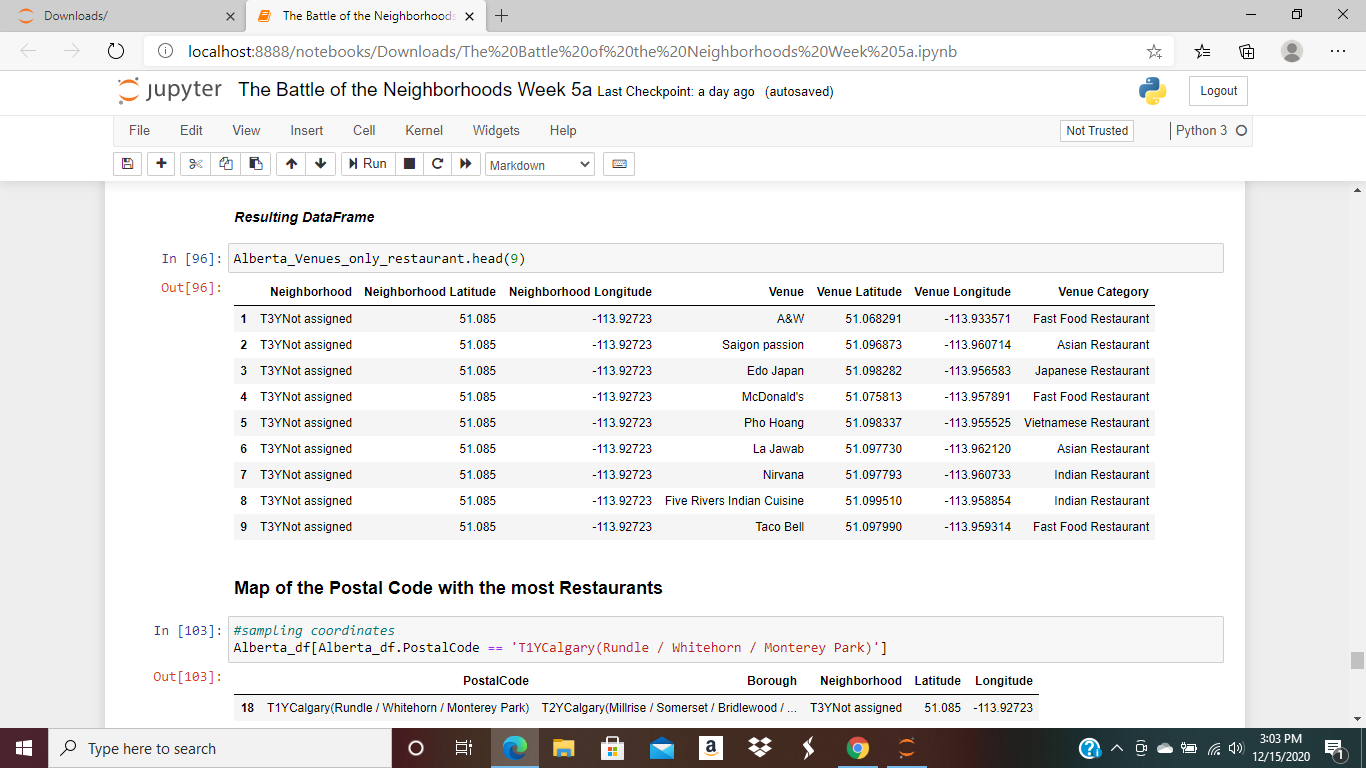
**Next,**

I wanted to answer the question “ is there opportunity for growth in the restaurants’ business? ”

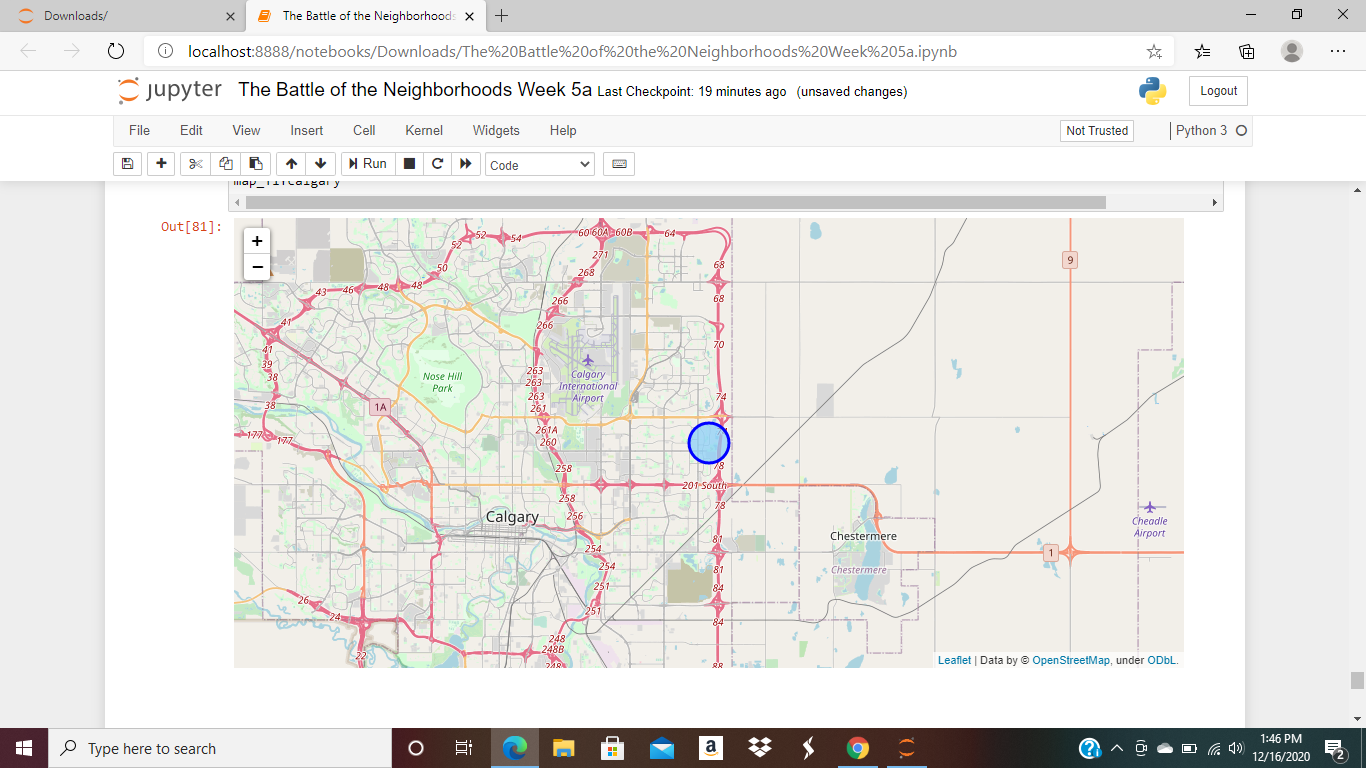
To help me answer that question I plotted a graph of the 5 most frequently occuring food venues.



The data revealed that although many casual food venues exist, only 9 restaurants were present. In addition 3 of them are fast food restaurants as shown in the restaurant only data frame below:



Furthermore these venues were concentrated in the region postal code TY1 (Rundle/Whitehorn /Monterey Park) an area close to the airport of Calgary as shown on the map below:



**Discussion**

My observation is the area has 18 parks but only 9 restaurants among them 3 are fast food restaurants. I have not yet visited Alberta Canada in person so I am happy to have taken this trip by means of data science. After doing the research I learned that the province of Alberta is a great place for all types of outdoors activities during the winter and the summer times.

The population is growing and is ethnically well diversed. Alberta Canada definitely has room for growth in the restaurant’s business. Some suggestions are an Italian restaurant, a French restaurant, a steakhouse, maybe a Greek restaurant or other.

**Conclusion**

The data presented above was gathered from sources where accuracy may not be pure. However based on it I concluded that Alberta Canada especially the Calgary region does present an opportunity to open a few more restaurants considering the population is growing and is ethnically diverse.

As I said before people love to eat, to drink and to have a good time. We are social creatures. Unfortunately for now because of the Covid-19 we are all isolated in some ways or other. Hopefully very soon all of that will change. Until such time “happy travel via Data Science and Bon appetite as well”. I thank you for taking this trip with me to Alberta Canada.