Final assignment

Final Assignment for Capstone Project

In the following lines I will define some important aspects about the topic I am working to finish the final assignment.

Introduction

In order to open new venues or business and get a sense of which ones are the most popular ones, it is necessary to know at least some essential information about the place where it is planned to be open. This could give a strong indicator whether the venue will be successful or not. This information would be crucial for any investor and anyone who want to open a new business.

Knowing which venues are most common could give a hint of how much profitable those venues are. I am going to compare some important cities. If some specific type of venue is successful in a city and there's other city which is similar, it is very likely that the venue would be successful in the other city. To do it, I am going to retrieve all the venues from the cities and get the most common venues. This should give a sense that the cities that have similar venues are in some way very similar cities.

I am planning to plot a map with the most common venue to see which part of the city is not crowded with that venue. This information should let us know which venue put and where in order to get the most profit from it.

Retrieving Data

The most difficult part might not be to use the algorithms, because we can get the result we want with simple algorithms. Using simple algorithms would mean that it would be computational less expensive and easier to understand for everyone. In this case, the data must be found on the web and it might not be so easy to find complete information for all the neighborhoods. In this case, to be sure I am going to use data from California, Toronto and Manhattan. I will have to clean all the data and use dataframes to get the desired information. It will be necessary to learn to use a new algorithm to get the coordinates of the neighborhoods, because the past methods did not work at all in my computer. I am going to retrieve all the venues from Foursquare and show the maps using follium. I will use hot encoding to get all the venues in different categories and then sum the number of venues in each category. Then, I will plot the most common venues of the cities in bar charts. Then I will compare the cities and decide which are similar to the others. Finally, I am going to plot a map with

only the most profitable venue and it would be easy to determine where it would be best to put one.

Data from California:

https://www.geonames.org/postal-codes/US/CA/california.html

Data from Toronto:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Data from Manhattan:

https://cocl.us/new_york_dataset

```
import geocoder

Code_1=df_clean['Code']


for postcode in Code_1:
    g= geocoder.arcgis(postcode)#maybe we need to change equal to the format below
    current_g= g.json
    neighborhood_lat= current_g['lat']
    neighborhood_lon= current_g['lng']

    df1 = df1.append({'Latitude':neighborhood_lat,
                                'Longitude': neighborhood_lon}, ignore_index=True)
```

Methodology

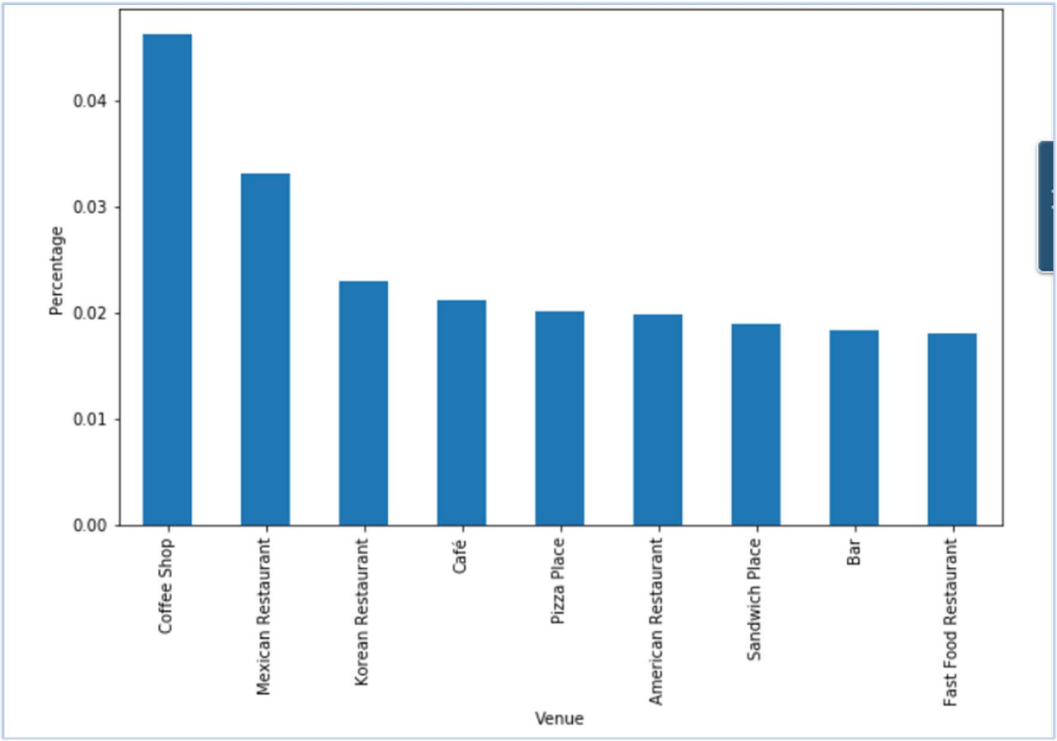
In the beginning I wanted to use the data from many cities and compare them in a big scale. However, it was impossible to find a dataset that contained every neighborhood with its coordinates in any city. For that reason, I tried to search for a dataset that at least contained all the neighborhoods and postal codes of any city. It was not easy, but I found 3 datasets: Washington, California and Abu Dhabi. So, In the beginning I tried to use at least these cities. I tried using the method suggested in week 3, but as the last time it was not able to get a single coordinate. I needed to search for a new method to get the coordinates. After learning how to use it, I tried with the 3 cities, but the only one that worked was California.

To be sure, I decided to use just data from California, Toronto and Manhattan. Once I got the coordinates, I needed to get the venues from each neighborhood and the coordinates of each venue. After that, it was necessary to use hot encoding to get all the unique categories of the venues. Once done I did some additional procedures.

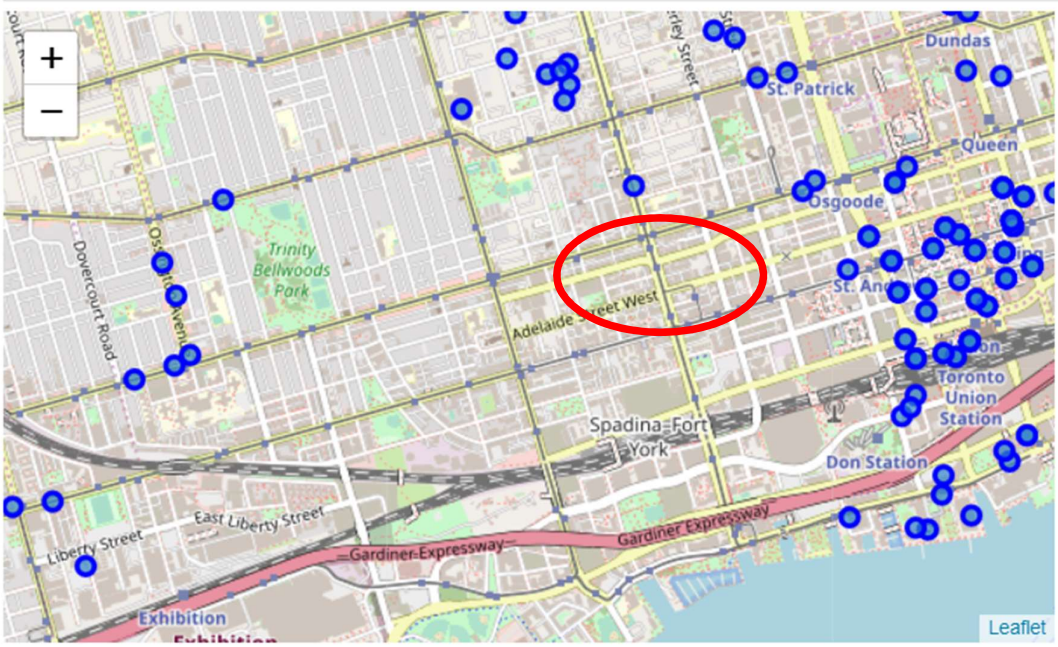
```
# one hot encoding
California_onehot = pd.get_dummies(California_venues[['Venue Category']], prefix="", pr

# add neighborhood column back to dataframe
California_onehot['Neighbourhood'] = California_venues['Neighbourhood']

# move neighborhood column to the first column
fixed_columns = [California_onehot.columns[-1]] + list(California_onehot.columns[:-1])
California_onehot = California_onehot[fixed_columns]

California_onehot.head()
```

	Neighbourhood	ATM	Accessories Store	Acupuncturist	Afghan Restaurant	African Restaurant	Airport	Airport Terminal	Am Resta
0	Beverly Hills	0	0	0	0	0	0	0	
1	Beverly Hills	0	0	0	0	0	0	0	
2	Beverly Hills	0	0	0	0	0	0	0	
3	Beverly Hills	0	0	0	0	0	0	0	
4	Los Angeles	0	0	0	0	0	0	0	

After doing the One hot encoding I summed every category to get the total number of each venue on the 3 cities. Then, I selected the 10 most common on each venue. After knowing the most popular venues I inferred which city was different from the others.



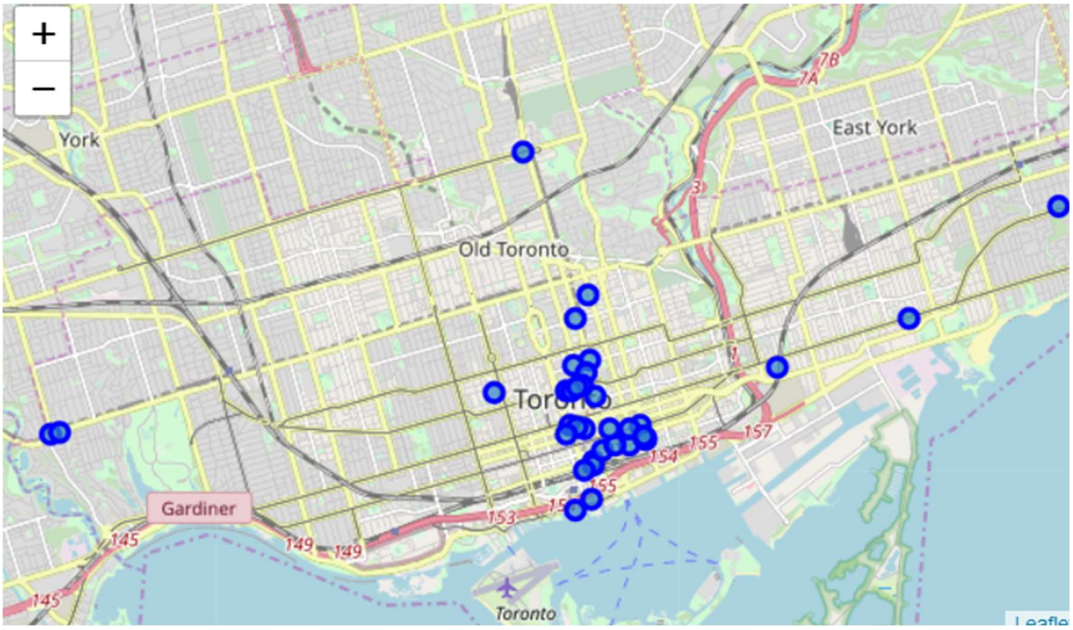
Finally, knowing which venue was the most popular one I plotted a map of the city with only that venue, in order to see which part of the city is not crowded with that specific venue, therefore it gives us an strong indicator where is the optimal place to build a venue .



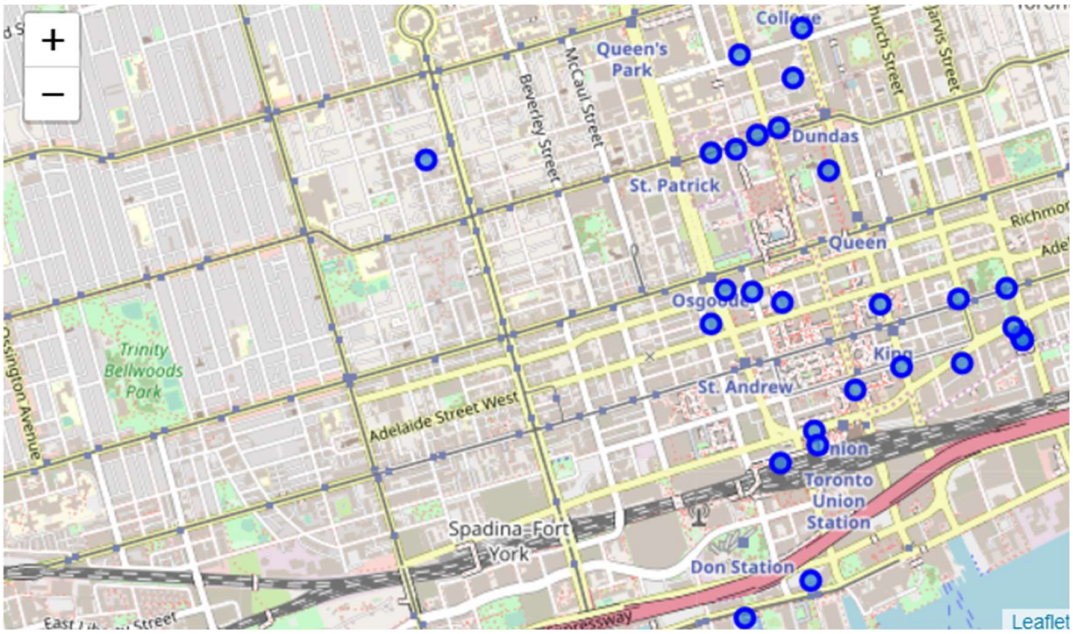
Near Spadina Fort York would be a good place to open a coffee shop

Results

The most common venue in all the cities were coffee shops or café, followed by restaurants. Nonetheless, the difference between the cities was in the difference of ratio of coffee shops in Toronto. Toronto had by far superior in ratio of coffee shops than California and Manhattan. Near Spadina Fort York was the best location to put a coffee shop in Toronto, because it was not crowded as the other areas. Also, another great idea would be to put more Pizza Places in Toronto, because Manhattan and California were not that different than Toronto, but they by far more Pizza places in ratio than Toronto. Pizza places in Toronto were on the 10^{th} place in the most popular venues, compared with Pizza places being in the 5^{th} in both Manhattan and California. Near Spadina Fort York suggests us again that it would be a good idea to open a venue there.



Pizza Places in Toronto



Near Spadina Fort York is again a convenient place to open a new venue

Discussion section

Although this assignment was originally made to use Google's API and Foursquare, not being able to use Google's API and being limited by the small quantity of algorithms to use taught me that we can do useful things with the simplest algorithms .The main idea of this course was to challenge us to create our own problem which will be useful on a particular situation, use Foursquare and apply what they taught us. I wish they had taught us how to use more complex algorithms such as those from Notebook in the outstanding submissions, but they did not. So, I had to think about what I should do and by doing it I realized that the simple algorithms are really powerful and easy to understand.

It was extremely difficult to get the coordinates of neighborhoods of the cities, because of the reasons I exposed before. However, I would suggest using the Arcgis method, as I did, because once found it was not difficult to use.

CONCLUSION

The most optimal venue and place I found was a Coffee shop in Toronto near the city center, for example near Spadina Fork York . Not necessarily it is assured that this is the exact optimum, because many factors influence in the success of any venue. The analysis I did was highly correlational, but it gives a strong indicator that It would be at least highly profitable. Also, another great option is to build a Pizza place in Toronto, because compared with the other cities that are in some way similar, Toronto has less Pizza places in ratio than the other cities. Pizza places are in the 10^{th} position in the most popular venues in Toronto, compared with Pizza places being in the 5^{th} position in both Manhattan and California. A good place to open the Pizza Place would be ,as in the coffee shops, Spadina Fork York.