

The Battle of Neighborhoods

Padmalaya Sahu

June 6, 2019

1. Introduction

1.1 Background

Tourism industry is one of the most important business industry in the world with millions of money worldwide. Which city of the world would interest a particular person is an important factor that determines the tourism business of a city or a particular website. For example, the more a website provides details about a city and its neighborhood's details, a person will be interested to use that website more, thus enhancing the tourism for that city. However, the interest of tourism location changes from person to person. Each year there are a number of tourists who aspire to travel to a new location. Thus providing the information about that location in clear and correct way is most important. These information are a profit for both the tourists and the tourism industry. It brings a lot of value, both competitively and economically, to the website the tourist's uses for the location information. Their importance is widely recognized by the tourism Industry.

1.2 Problem

Data that might contribute to determining the neighborhood details of a place might include its location code, nearby cafes, nearby restaurants, nearby library, nearby museums, and trends that are famous a particular time determine the tourism rate of that place.. This project aims to compare details of two cities namely, New York and Toronto to help a person with respect to his/her demands about the neighbourhood of the city.

1.3 Interest

Obviously, Tourism industry would be very interested in getting the details of the interest of a particular tourist to enhance their business . Tourists who are interested in visiting New York or Toronto may also be interested. It may also interest certain websites to use these data to enhance their site by providing the details.

2. Data Needed :

New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

Data source : https://cocl.us/new_york_dataset

Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.

Toronto city data that contains list PostalCode,Boroughs,Neighborhoods .

Data_source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, De

scription : This site is a wikipedia page containing the postal codes and its corresponding boroughs and neighborhoods. The actual table from this page was scraped and stored in a dataframe for further use.

Various places in each Neighborhood of new york city and Toronto.

Data source : Foursquare API

Description : By using this api we will get all the venues in each neighborhood. We can filter these venues to get only indian restaurants.

GeoSpace data for New York and Toronto

Data source : <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

Data source : http://cocl.us/Geospatial_data

Description : By using this geo space data we will get the New york Borough boundaries that will help us visualize different neighborhoods of particular city.

3. Methodology

In this section we discuss the analysis of our project. I have given the methods I used to explore the data and various results I got. The machine learning techniques, the foursquare api usage everything is explained in this section.

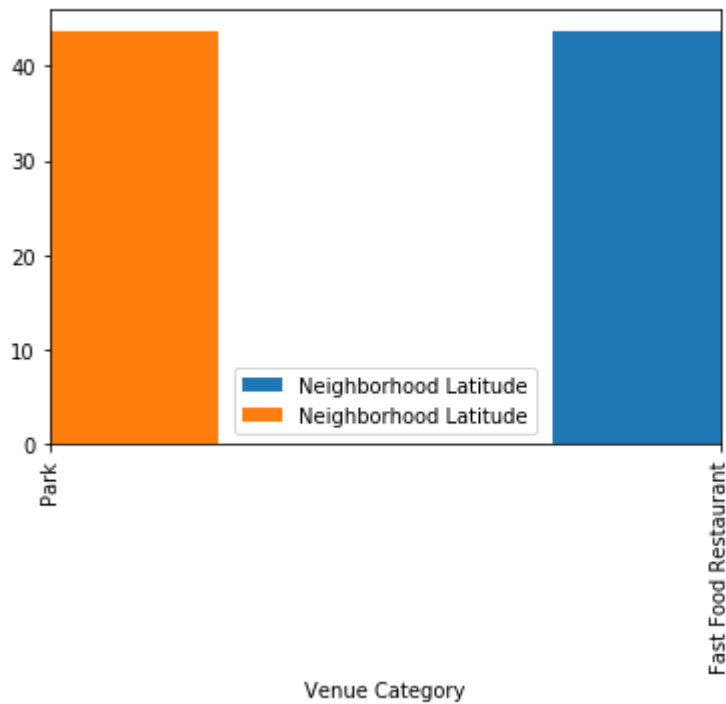
For example, like in the following figure I have represented the data about if there is any park at a particular latitude and longitude at the Toronto City. The bar graph indicates the presence of a park at the (latitude=42.5 and longitude= -72) position. I have the matplotlib "plot" method for plotting the information about the particular latitude and longitude value.

We can plot different longitude and latitude values to get data about them.

```

toronto_venues.plot(x=venue_category, y=[neighborhood_latitude])
plt.xlim("Park", "43.75", "-79.32")
plt.show()

```



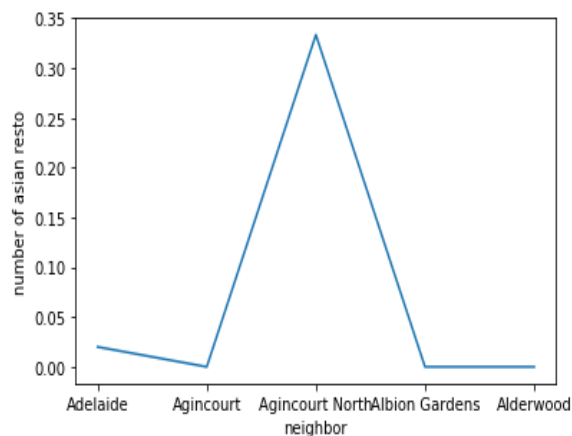
Now in the below picture I have plotted the frequency of the asian restaurants in Toronto city between different neighborhoods. It can be seen that Agincourt neighborhood has the highest frequency of neighborhood restaurants.

```

plt.xlabel(neighbor)
plt.ylabel('number of asian resto')

```

Out[119]: Text(0, 0.5, 'number of asian resto')



In this picture the most visited venues of New York city is given. It can be seen that the venues are listed with their ranking i.e. the top visited venue, the second most visited etc. I have used the Foursquare api method to get the data about the venues in the nearby neighbourhood. The explore function of the Foursquare api helps us to get all the data we want.

[46]:

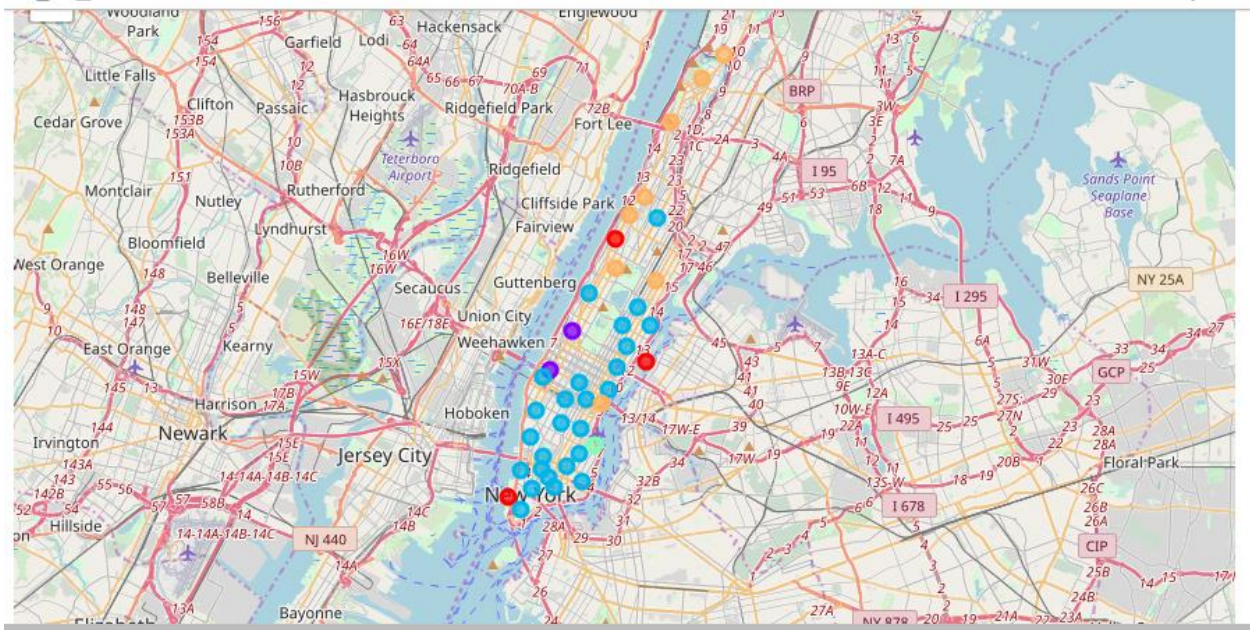
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Wine Shop	Clothing Store	Italian Restaurant	Department Store	Women's Store
1	Carnegie Hill	Coffee Shop	Pizza Place	Café	Yoga Studio	Bookstore	Cosmetics Shop	French Restaurant	Bar	Japanese Restaurant	Spa
2	Central Harlem	African Restaurant	Art Gallery	Seafood Restaurant	American Restaurant	Gym / Fitness Center	French Restaurant	Cosmetics Shop	Chinese Restaurant	Public Art	Grocery Store
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Nightclub	Bakery	Seafood Restaurant	American Restaurant	Theater	Art Gallery	Hotel
4	Chinatown	Chinese Restaurant	Cocktail Bar	American Restaurant	Salon / Barbershop	Spa	Dumpling Restaurant	Bubble Tea Shop	Ice Cream Shop	Vietnamese Restaurant	Bakery

The top 3 venues of Battery Park City neighborhood are “Park”, “Coffee Shop”, “Hotel”. Similarly a dataframe for Toronto City is also constructed to get the trends of the city.

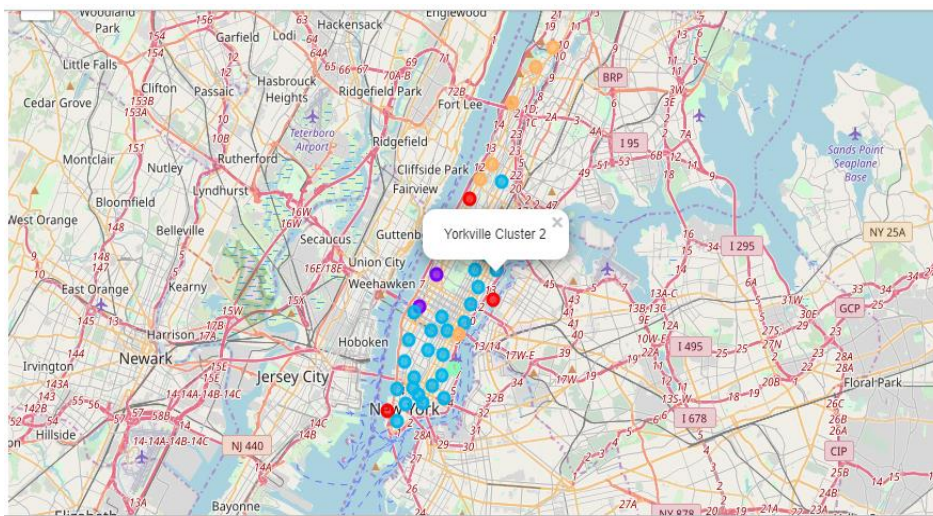
In the below picture, a scatter plot for different neighborhood of Toronto city is constructed over the map of the city. I have used the Folium library for this.



In the picture below ,different clusters of the neighborhood information of Manhattan in New York City is shown.It can be seen that the different clusters are of different colors.



When we click over a particular circle its neighborhood name and cluster label is shown in the popup.All the clustering is done using the sci-kit learn library.It contains different methods to perform various machine learning techniques.Clustering is an important machine learning technique to group similar data together in a cluster.This classifies our data according to our Interest.In this project our point of interests are different neighborhoods.



4. Results

The results of this project is not very particular. The results depends upon the neighborhood we want to explore and its attributes. This can be useful for any tourism industry ,websites,tourists etc to get the results about any particular neighborhood.

5. Conclusion

The conclusion for this project varies with the interests of the users .For example , if a person like Indian Food he/she is mostly likely to visit an Indian Restaurant. Thus if he/she searches that we can predict the tourism for that restaurant. This project is actually helpful for all the businessmen and the customers.

Some points can be -

- Manhattan have potential Indian Restaurant Market
- Staten Island ranks last in average rating of Indian Restaurants.
- Manhattan is the best place to stay if you prefer Indian Cuisine.
- The most common venue in the Alderwood neighborhood of Toronto city is Pizza place followed by Gym.