

Where would it be the best place for Brazilians to live abroad?

Emanuel Monção Lima

April 17, 2020

1. Introduction

1.1 Background

A 24-years Brazilian military government has come to an end followed by an enormous financial crisis which has forced the country to change its currency for the eighth time within 42 years, and so its economic model in the 80s and 90s. The move to the new currency (BRL) now attached to the American dollar (USD) appears to be one of the most successful currency changes of recent times and has made the Brazilian economy be in spotlights again.

Nevertheless, with more freedom to come and go and better financial conditions Brazilians still have got some issues related to public administration which seems that it is possible to lead the country to some kind of financial, security, or political crisis in any time. This leads to a sort of tension and uncertainty mostly due to a trauma caused by several mistakes made by governments since just after World War II time but having critically collapsed back in the '90s. All of these combined have made some people try on new lives and experiences in another country.

1.2 Problem

The project has the idea of tracking and mapping best places in the world that might have familiar attributes to those Brazilians who wonder to move to another country someday in a near future based on choices of people that have already moved out and on Brazilian-related places and communities in the neighbourhoods so newcomers may feel more comfortable at first.

1.3 Interest

This might be individually interesting for those who might want to move to another country and have not decided whereto or want to check some pieces of information out first before making such a big move. Also might be interesting to some exchange agencies which aim to customize their packages depending on their customers' wishes since this report takes some parameters into consideration but are totally up to customizing though.

2. Data Management

2.1 Data Sources

Firstly data has been scraped from a *Wikipedia* page ([Brazilian diaspora](#)) containing countries in the world which have got a significant number of born Brazilians living it. Secondly, it was scraped from another webpage that provides a bunch of data about US cities and their citizens ([Top 101 cities with the most residents born in Brazil \(population 500+\)](#)) but here it indicates the top 101 cities in the US with the most residents born in Brazil. Finally, the *Foursquare* API was used to get some data about places related to Brazil in the chosen cities through a query call in a *Jupyter Notebook* that has been generated.

2.2 Data Cleaning

The pages were downloaded using *urlopen* method from *urllib.request* module and parsed from HTML to a *BeautifulSoup* instance so it was possible to better look into tags within the page and have easier access to data therein.

On the first page, the data was within a table tag with a specified class called "*infobox vcard*", so it was a nice try to use the *find()* method since there was only one table with that class name. Next step was to remove some links that were on our way to get a clean data from the table and it was observed that all those links had a "*sup*" tag in common which has made it easier to remove with the *find_all()* and *decompose()* methods in such order. Then all table rows have been taken from it then selected which rows had useful data and put into variables lists. With good rows in hand, it was parsed to a data frame from *pandas* module to be better manipulated such as taking some remaining brackets out and converting necessary rows to the right data types to be manipulated correctly afterwards.

The second webpage had the same scraping strategy and methods and the table which had the desired data was tagged as *table class="tabBlue tblsort tblsticky"*. Like the data treating of the first page, the same methods for removing not useful tags in the table, taking the good rows out of the table and parsing them to a *pandas.DataFrame()* object have been used. Both scraped data were treated to their better layout so graphs were to be taken out of them more easily and so data were better visualized.

Finally, the third dataset was the *Foursquare API* data which has been requested (*GET*) using a search query and a radius of interest alongside with the developer credentials such as client ID, client secret as well as the *Foursquare* version set to "20180604" and a specific limit number set to 50. Once all of these were set up, they have been formatted such as

"https://api.foursquare.com/v2/venues/search?client_id={} & client_secret={} & ll={},{} & v={} & query={} & radius={} & limit={}".

Its result was normalized to .json and had their rows selected as so convenient. These contain Brazil-related places such as restaurants and communities that may be useful as support for newcomers to a new city with such different culture and habits. So, if clustered, this may indicate potential good areas to live or stay in and yet feel as comfortable as possible at least at first.

3. Methodology (Exploratory Data Analysis)

The entire process was divided into stages so will this Exploratory Data Analysis. The first thing needed was to know which countries have Brazilians gone to the most in order to live so it is possible to filter the search to only one or two countries. After having a country to focus on, the next step is to fetch for those cities from that country which have a population with the most people born in Brazil. Having that city or cities, it is better to look for areas and neighbourhoods which contain the most Brazil-related places such as restaurants, pubs, cultural centres, steakhouses and so on. Once these venues have been collected, they will be put into a map and clustered and the best areas at first sight to a Brazilian newcomer to move in are shown are then shown.

3.1 Choosing Country

The first dataset which has been scrapped from the web (*Wikipedia*) will then be analysed. The page contains the approximate number of Brazilians living abroad and the countries which they have moved to. This data has been scrapped as mentioned in Section 2 and after cleaned and wrangled was put into a data frame. Since data is well organized, it's possible to see it clearly through a bar plot. It's been thought the horizontal bar plot would be a better way to visualize this specific data.

So, in order not to overfill or pollute the plot, only the first 15 countries with significant Brazilian population living in will be plotted using *barh()* method from *matplotlib.pyplot* module and the population data was scaled (divided by a thousand due to too big numbers) and so the label of the x-axis turned to be cleaner

The plot has taken place as follows:

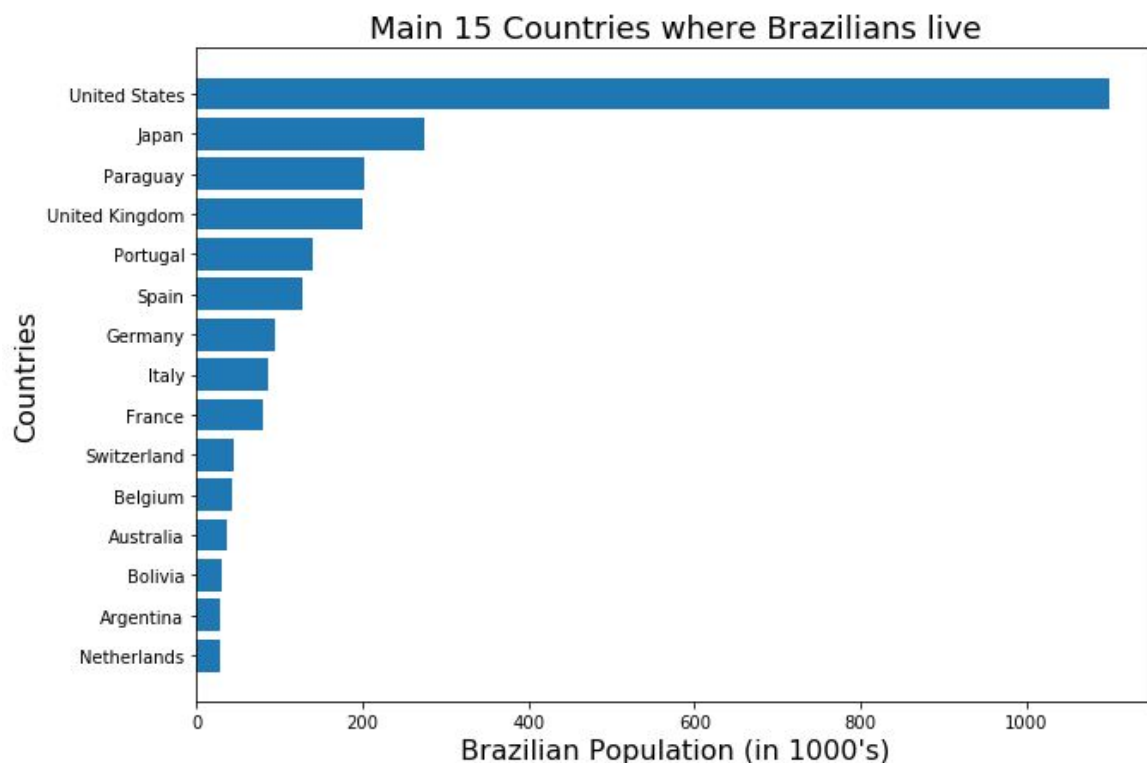


Figure 1: Horizontal bar graph plot representing countries with the most Brazilians

Since there are over 1 million Brazilian-born people living in the US and the difference to the second country (Japan) is huge, the United States of America was the chosen one to focus this project on because it is understood that a newcomer would feel more

comfortable in a place where it could find and contact with its compatriots (at least at first contact with a new country) easier and hence it takes us to the next stage of this analysis.

3.2 Choosing City

Now that the US is the chosen country due to its numerous Brazilian population, it is easier to seek for a city. The same features as above were taken to make this decision but taking the second dataset (from *city-data.com*) in consideration which contains the top 101 cities with the most Brazilians and also the percentage of people born in Brazil in comparison to their population. It was then added to the dataset a column with the calculated (estimated) number of Brazilians living in each city from it. The dataset then became as below:

	City Name	City Population	Percentage	#Brazilians
0	Loch Lomond, FL	3786	14.9	564
1	Avalon, FL	679	12.1	82
2	Everett, MA	42935	11.0	4723
3	Framingham, MA	68318	9.0	6149
4	Long Branch, NJ	30390	8.4	2553

Figure 2: Table of Brazilians in the cities of the USA.

Looking quickly at the table representation of the data frame above in Figure 2 it's worth to check that the first city *Loch Lomond, FL* has a percentage of 14.9% of its population tagged as Brazilian and the fourth city which is *Framingham, MA* and has a percentage of Brazilians living therein equals to 9% of its population. Though the city in the state of Florida has a higher percentage of its inhabitants represented by Brazilians the city in Massachusetts has more Brazilians living in: approximately 6150 in Framington whilst only 564 in Loch Lomond, almost 11 times bigger. This is because the population in Framington is also bigger and hence this percentage might not be the best feature to base on the decision.

Therefore, this data has been observed in two ways: firstly, taking into consideration that if it's a so-called big city, it has a bigger chance of being more immigrant-friendly and of being more supportive to Brazilian newcomers either through communities or places which are related to Brazil. Secondly, the number of Brazilians living in a city is a very important feature to have in mind for the decision-taking moment.

So, having the dataset mentioned above ordered by the population of the cities within it using the `sort_values()` method from `pandas.DataFrame`, a bar chart was plotted with the top 5 cities in such order.

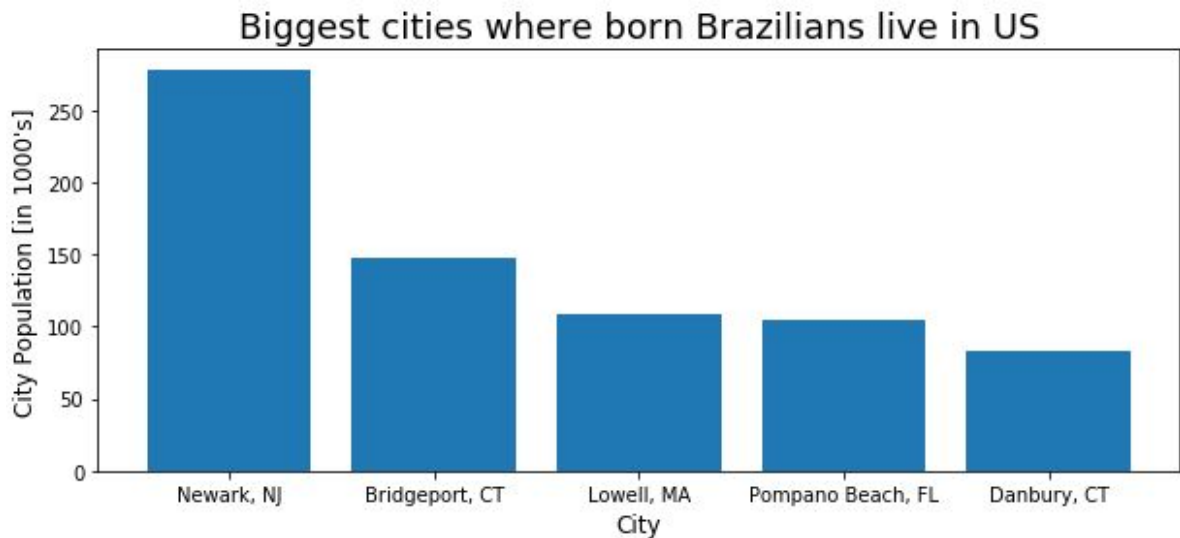


Figure 3: Bar chart showing cities ordered by its population size

Also, using the same methods as above the dataset was reordered by the number of Brazilians living in them another bar chart was plotted as follows.

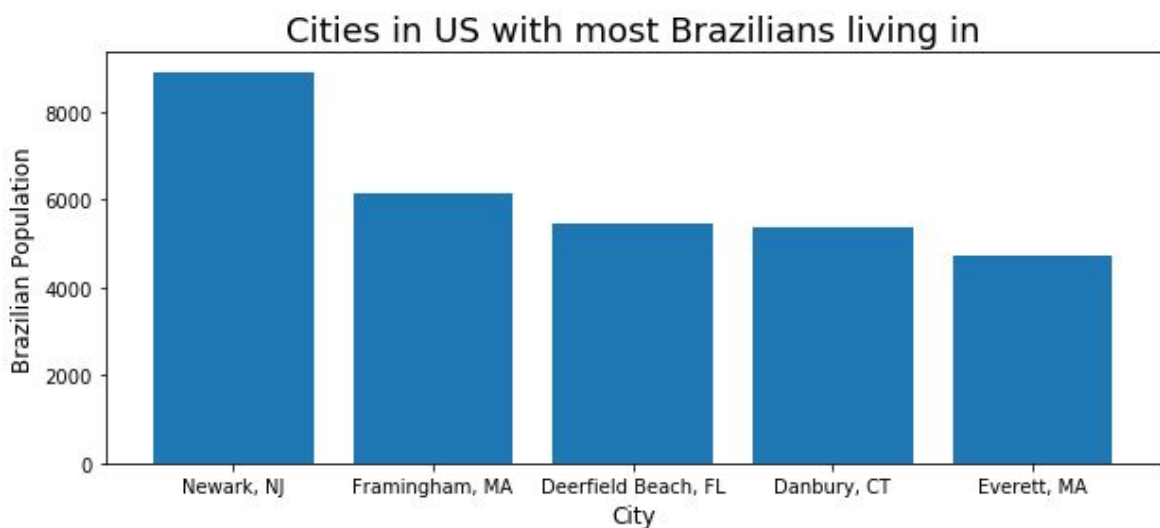


Figure 4: Bar chart showing cities ordered by the number of Brazilians living in them

As a result of the analysis of both graphs above, one is able to observe that *Newark, NJ* and *Danbury, CT* are contained within both of them, i.e., they are included in the top 5 of both features. Thus, a particular analysis on these two cities will take place then.

3.3 Clustering Best Areas in the City

In order to cluster the best areas in chosen cities, their locations are firstly needed and they were obtained using the *Nominatim* method from *geopy.geocoders* module. A call has been made and returned an object containing the requested city's latitude and longitude. A *Map* object from *folium* library was then created and its centre would become the coordinates just found. This process was executed for both cities.

Next step should then be the requesting of the venues related to Brazil culture within each city. A *Foursquare* call through an API was used with a search query radius set to 7 km centred on the same city's coordinates used on Map. Those venues would be what the clustering algorithm is working on soon. As an example, the figure below shows the venues of one of the cities mentioned.

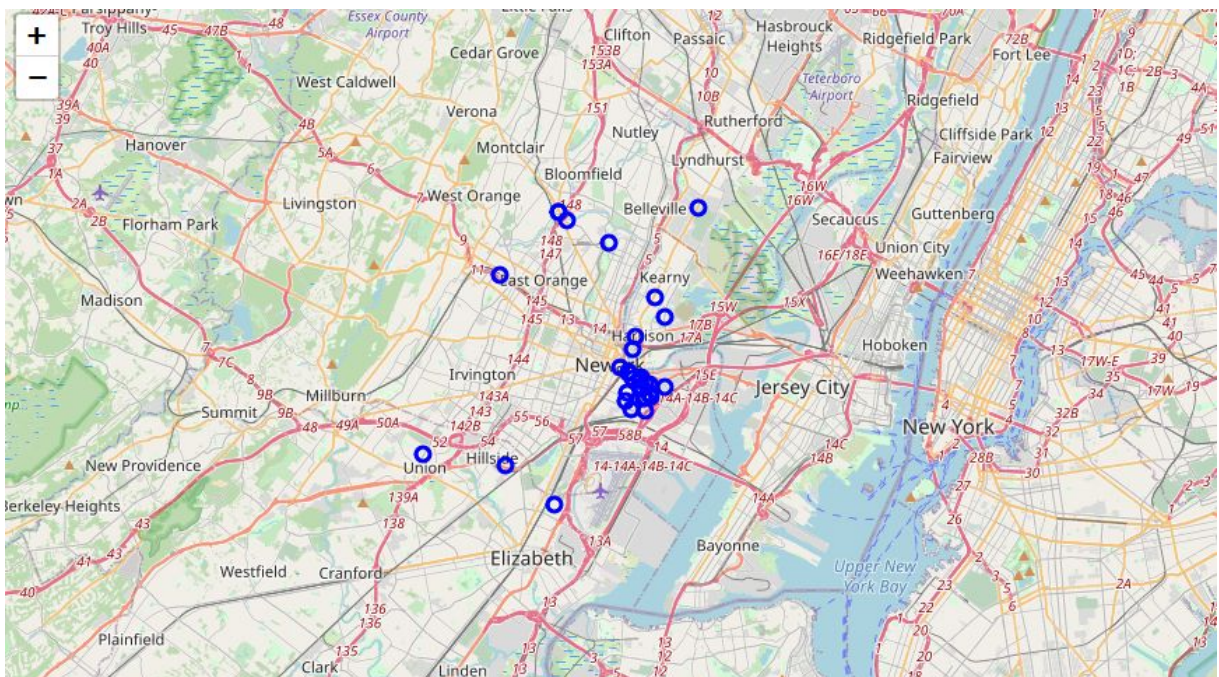


Figure 5: Screenshot of the map showing Brazil-related venues nearby the city of *Newark, NJ*.

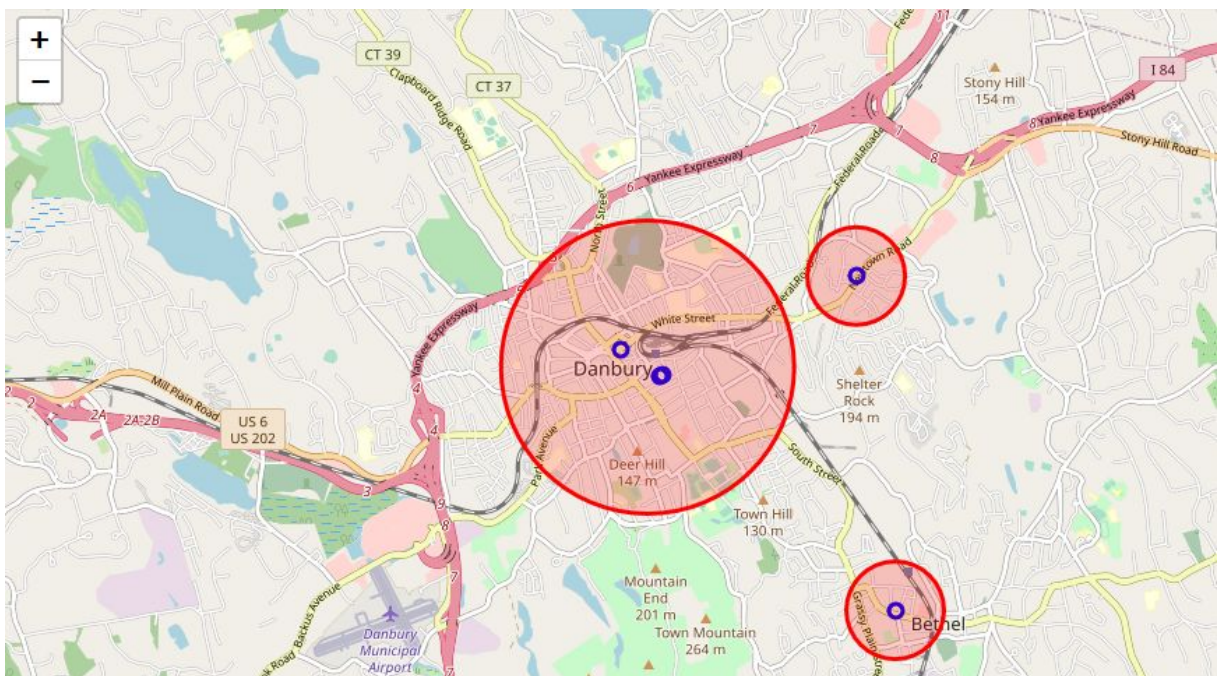
The circle markers on that map have been added using *CircleMarker* from *folium.features* module previously imported and the latitude and longitude of each venue which are part of the *Foursquare* request result. Based on an observation upon the maps, it was possible to decide a decent number of clusters needed so the clustering algorithm may produce a decent and acceptable result.

Once the number of clusters has been decided, it is applied to the clustering algorithm which for this project the *KMeans* algorithm from *sklearn.cluster* module was the chosen one due to its versatility and simplicity to use, moreover, it is largely used for clustering tasks with maps involved. For the call, the only parameter passed into the function was the number of clusters. And, since there is no need for predicting any future events, only the training process was required (with *fit()* method). The dataset for training was a data frame in which columns contained the latitude and longitude of each venue therein the correspondent city.

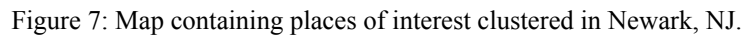
After the algorithm has finished the training process, the clusters' centres were taken and had a circle for each of them added to the city map using *folium.Circle()* so the circle radius doesn't change as we zoom in and out since the radius is measured in meters in that method instead of being measured in pixels as in *folium.features.CircleMarker()* method. At last, the radius was based on the number of venues labelled or the area covered by them.

4. Results and Discussion

The clustering process has been applied both in Danbury and Newark cities in order to group Brazil-related places in and draw what it would be an optimal area for Brazilian newcomers. So, Danbury, CT has its map turned into the next figure.



In the same way, Newark, NJ had its places of interest for Brazilian support clustered as follows:



On the other hand, Newark appears to be a lot denser than Danbury in terms of Brazilian places and so one may wonder that other city sectors are also dense, i.e. more places condensed in the same or smaller area since the Brazilian percentage within the city is around 3.2% and can be seen as the clusters are much closer to each other than the Danbury's ones (and yet these have had their radius modified) and even have them intercepting each other like a chain. Though, Newark has plenty of options to any Brazilian newcomer feel more comfortable when arriving in a completely unknown land.

At the moment that the Newark map was being analysed, has come to mind that another important feature to be considered is the city's distance from the biggest cities and metropolises around it.

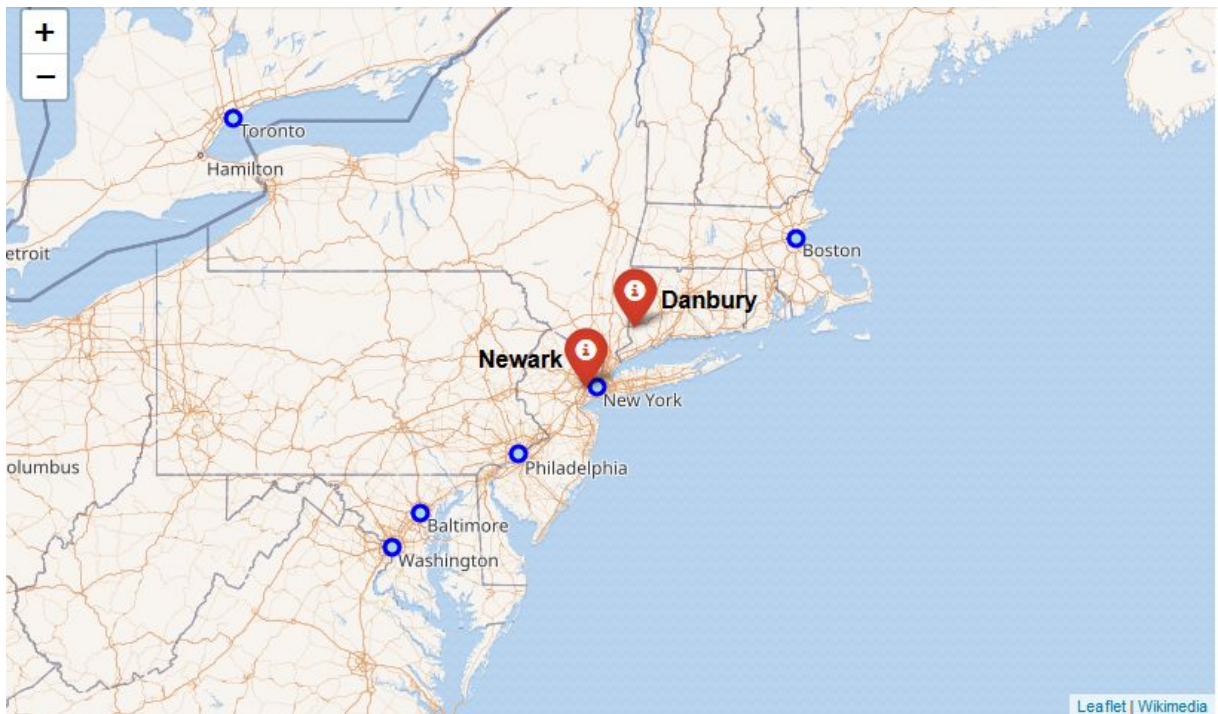


Figure 8: Map showing the location of some important cities around both cities of analysis.

The map above shows some of the most important cities around Danbury, CT and Newark, NJ and it is possible to be somewhat aware of how distant are they from those big and important cities. This might perhaps be helpful to make the decision.

5. Conclusion

It is indeed a personal choice of where the person or family is going to try on new lives soon and a different country. Different people, probably a different language spoken, different sights and weather. A whole different culture, to summarize, even if there is only a border setting those countries apart. Basically, everyone knows what they don't like about their home or previous cities and there stands a good chance that those are unique opinions. So, if people thoughts about the city they have lived in for years or their entire lives change from person to person, so it will for choosing where to settle in for the years to come.

Thus, the ideas hereby presents a set of two options of cities having some parameters set. However, these parameters might be changed accordingly to customers demand. for best attendance. In this case, the two cities hereby presented were *Danbury, CT* and *Newark, NJ* and each had some of their characteristics hereby presented as well - at least when related to Brazil or Brazilian people.

In a nutshell, it is up to the individual to choose between a bigger or a smaller city to live in (or whether its size even matters at all), a city nearer to a big city or another a bit further though still, it is a not so long distance and plenty of other parameters depending on the individual's necessities or wishes. Also, the highlighted areas in the cities' maps have empirical radius based on Brazil-related venues density and might be either larger or smaller.

Note that the results herein presented do not take into consideration the current pandemics which the world is experiencing or its effects either on Brazil or the USA politics, economics nor the relationships between both nations.

References

- https://en.wikipedia.org/wiki/Brazilian_diaspora. Access on April 17th, 2020.
- <http://www.city-data.com/top2/h153.html>. Access on April 17th, 2020.
- <https://pt.foursquare.com/city-guide>. Access from April 18th to 26th, 2020.