

Juan David Jaramillo

---

## **New Restaurant in Colombia**

### ***Introduction***

While opening a restaurant can be a very lucrative business, a lack of demand causes many restaurants to close within the first year of opening. There are many different factors that can account for a restaurant's success such as location, competition and quality of the food. This is an important question that every business owner must face when choosing whether to open a restaurant or not, as well as location of the business.

Colombia has a growing economy that attracts a lot of international investor given its high potential to create large returns on capital. One of the most profitable business in the last years is the American restaurants that offers American style burgers. Thus, a lot of new burger restaurants have opened its doors in the past few years. However, a Chef believes that he has the best recipe for the perfect burger. He doesn't want to compete against a lot of restaurants, so he wants to open his restaurant in a low-density zone or in a zone with bad rated restaurant so he can take all the unhappy clients. He is willing to invest all his savings in the restaurant to offer the best service and product to the costumer. For this reason, he needs to be sure that the place chosen will be the best in terms of surrounding restaurants.

Here we will try finding if someone wants to open a new restaurant in Colombia which location is best suited for it keeping in mind the competitors and which income group of people will be attracted most. We will consider variables as population of the city, density of American restaurants around and rating given by the people.

Since there are lots of restaurants in Colombia we will try to detect locations that aren't crowded with restaurants. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

### ***Data***

Data on Colombia cities and its population

([https://en.wikipedia.org/wiki/List\\_of\\_cities\\_in\\_Colombia\\_by\\_population](https://en.wikipedia.org/wiki/List_of_cities_in_Colombia_by_population)). We will import this database in order to explore it and choose the top 5 cities with the most population.

We will access data through FourSquare API interface and arrange them as a dataframe for visualization of American Restaurants in 5 the major Colombian cities which will be determined based on the database mentioned above. FourSquare will provide us with information about surrounding restaurants and ratings. By merging data on Colombian restaurants and rating data, we will be able to recommend profitable investments.

## Methodology

The main target here is to assess which city would have the lowest American Restaurant density. I used the Four Square API through the venues channel. I used the near query to get venues in the cities. Also, I use the CategoryID to set it to show only American Restaurant type Places. An Example of my requests:

[https://api.foursquare.com/v2/venues/explore?&client\\_id=&client\\_secret=&v=20180605&Bogota,&limit=100&categoryId=4bf58dd8d48988d14e941735](https://api.foursquare.com/v2/venues/explore?&client_id=&client_secret=&v=20180605&Bogota,&limit=100&categoryId=4bf58dd8d48988d14e941735)

That 4bf58dd8d48988d14e941735 is the Id of the American Restaurants. Also, Foursquare limits us to maximum of 100 venues per query.

Moreover, I repeated this request for the 5 studied cities and got their top 100 venues. I saved the name and coordinate data only from the result and plotted them on the map for visual inspection.

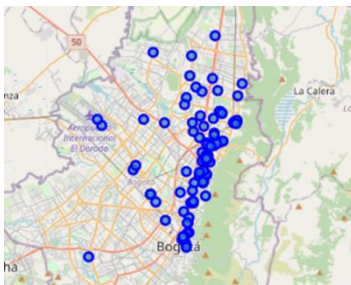
Next, to get an indicator of the density of The Restaurant Places, I calculated a center coordinate of the venues to get the mean longitude and latitude values. Then I calculated the mean of the Euclidean distance from each venue to the mean coordinates. That was my indicator; mean distance to the mean coordinate.

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:

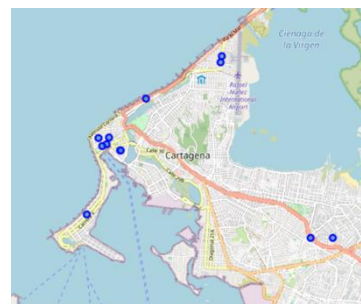
1. Collect Inspection Data
2. Explore and Understand Data
3. Modeling

## Results and Discussion

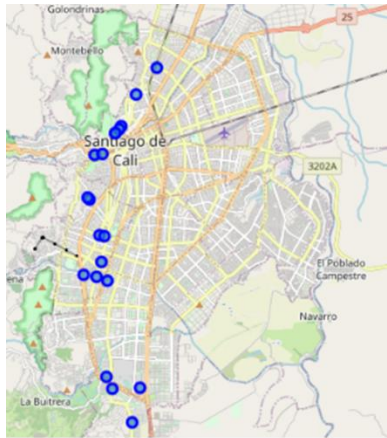
The next step was to visualize the number of restaurants in different cities of Colombian to obtain a general understanding of the distribution and density of restaurants. As seen from the map, all the cities are densely clustered near downtown and spread out as the distance from downtown increases. This phenomenon can specially be watched in cities such as Medellin and Bogotá. This is important because while some cities might not have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant.



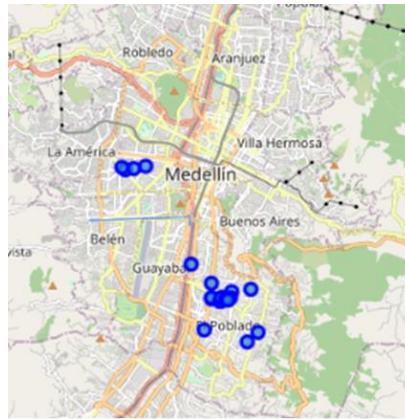
Bogotá



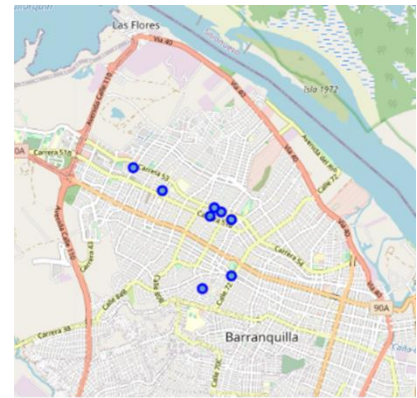
Cartagena



Cali



Medellín



Barranquilla

As we can see, Bogota es the biggest city in terms of population however by far it has 100 or more restaurants all over de city, thus the market shar will be really small and would have to invest a lot of money to outstand the other good restaurants. On the other hand, the other Cities such as as Medellín, Cali, Barranquilla or Cartagena, only have about 20 or less American restaurants so there is a big market to explore a lot of clients to attract. Moreover, cities as Medellín and Cali also has a big amount of population so this two are the ideal city to choose. As I said before I would consider each city as a cluster, thus me cost function will be base on the city with the least Euclidian distance from the center to every restaurant. That means that for a client would be easier to reach my restaurant. Base on all of the above criteria, Medellin was chosen as the best city to open an American Restaurant.

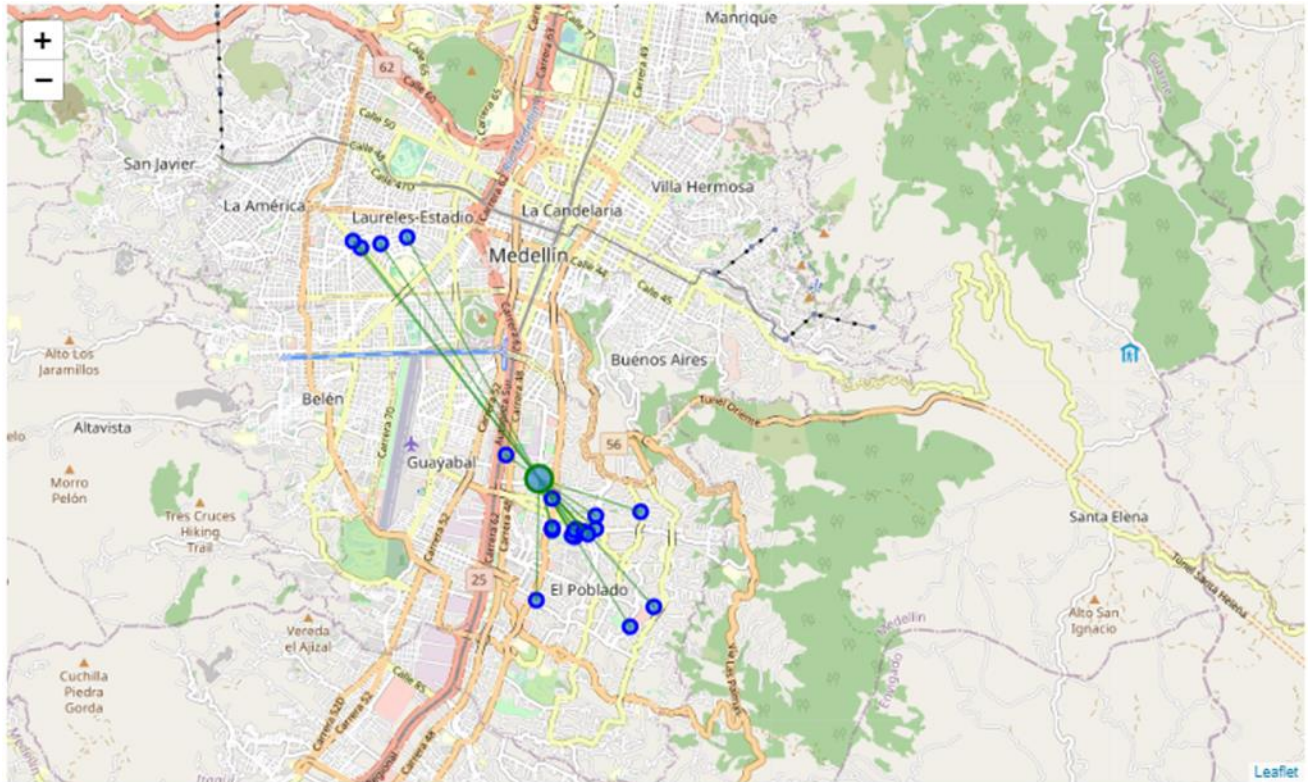
Cluster	Characteristics
Cluster 0	Positive Spending Power (0.3 – 1.8)
Cluster 1	Negative Spending Power (-1.2 -- -0.8)
Cluster 2	Near Zero Spending Power (-0.5 – 0.5)
Cluster 3	High Positive Spending Power (1.7+)
Cluster 4	Negative Spending Power (-0.8 – 0) With Large Number of Restaurants

**Table 1:** Characteristics of the clusters resulting from k-Means clustering algorithm

From the results of the clustering algorithm, it was determined that neighborhoods corresponding to cluster 3 (Medellin) were the best choice for opening an American restaurant in Colombia based on the normalized spending power and population.

From the three remaining regions, I would recommend that the client open his/her restaurant in either Cali, Barranquilla or Cartagena.

Especially Barranquilla or Cartagena Both regions have very few restaurants, so the offer of American products is small.



## **Conclusion**

Opening a restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, extensive research about the area would greatly increase the likelihood of the restaurant succeeding. From the project above, I demonstrated the workflow necessary for a client to determine what area the restaurant should open. For specifically, I determined that the optimal location to open an American Restaurant in Colombia, would be in Medellín.