A NEW PERUVIAN RESTAURANT IN MANHATTAN, NY

Applied Data Science Capstone

ABSTRACT

A new business case of opening a new Peruvian restaurant in Manhattan, NY, is evaluated based on the best optimal location to open it.

cesar galvez nuñez

Capstone final project

Table of Contents

1) Introduction to Business Case.

A Peruvian entrepreneur that counts with more than 50 restaurants in Peru hired my data science consulting company to investigate the possible location to open a new restaurant in New York City, specifically in Manhattan city. New York City has been the most populous city in the United States. It has a long history of multicultural immigration, which makes it very attractive to open a new Peruvian restaurant. Furthermore, most people associate Manhattan, NY, as the city never sleeps, which makes it an excellent feature to initiate a new culinary business. Peruvian food has experimented with an exponential change in the last ten years, and it has been considered one of the world's best culinary destinations. The biodiversity and multicultural heritage are the main features that make Peruvian food unique and the best place in the world to dine. Therefore, the gastronomic entrepreneur wants to take advantage of the Peruvian food excellent reputation and start a new business opportunity of opening a new restaurant in the borough's most popular restaurants, Manhattan city, NY.

2) Business Case Questions

The next questions should be answered after the evaluation of business case.

Q1) How many Peruvian restaurants are in Manhattan?

Q2) What is the best location in Manhattan City to open a Peruvian restaurant?

Q3) What are other potential neighborhoods in Manhattan for Peruvian Cuisine?

Q4) Which neighborhoods lack Peruvian restaurants?

3) Data Section.

The following data is needed to get the solution for this case.

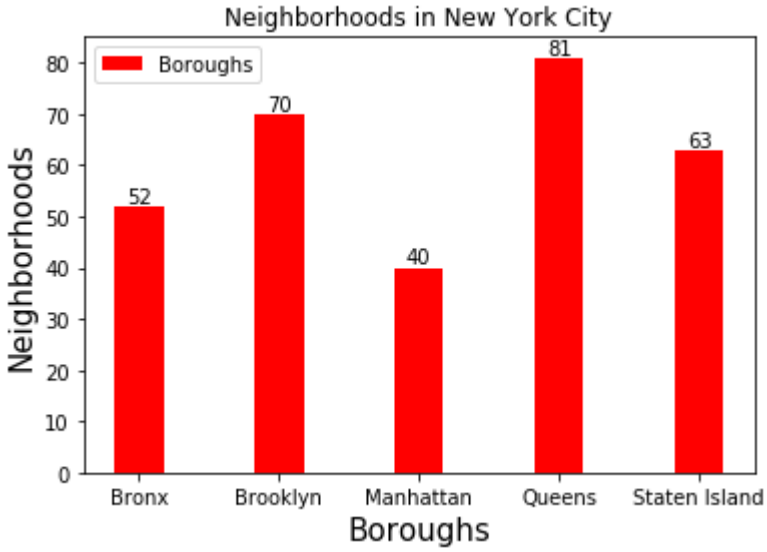
- ✓ List of Boroughs, Neighborhoods, as well as the latitudes and longitudes from each neighborhoods in New York: https://cocl.us/new_york_dataset
  The list of Boroughs, Neighborhoods, latitudes, and longitudes are needed to build the business case. After loading and exploring the data, it is then transformed into a data frame using Pandas. As the evaluation is focused on Manhattan, it is filtered to obtain only Manhattan's neighborhoods.
- ✓ List of 100 venues around 1000 meters of radius in Manhattan: **Foursquare API**
  This list will provide information on businesses around 1000 meters from Manhattan city.
- ✓ Venue data of Peruvian restaurants in Manhattan city: **Foursquare API**

The previous data is then filtered to obtain the existing business of Peruvian restaurants already in operation around Manhattan city. This data will be merged with the data of neighborhoods in Manhattan.

4) Methodology

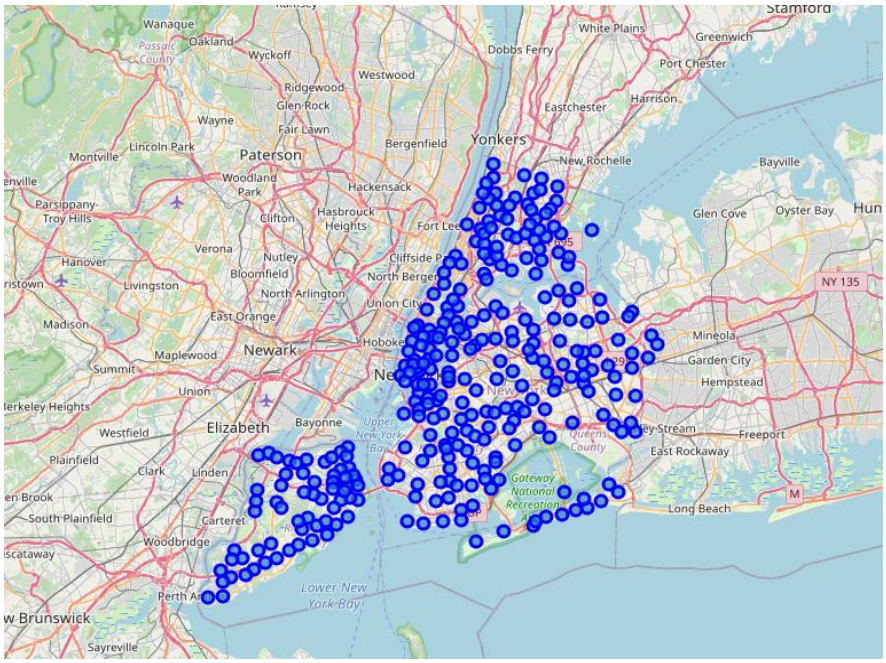
a) The evaluation starts by determining the number of boroughs and neighborhoods in New York. The data is extracted using the first URL link from the data section. The data displays that New York has 5 boroughs and 306 neighborhoods. Likewise, figure 1 is built utilizing the data and it displays Manhattan is the borough with the lower number of neighborhoods. Even though Manhattan has the lower number of neighborhoods, it has the most popular restaurants in New York.

Figure 1



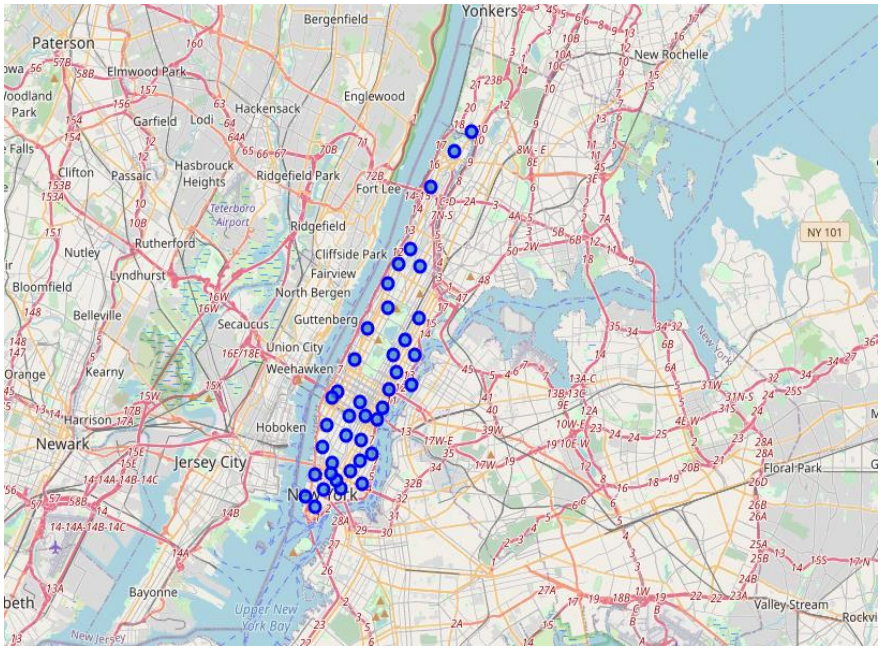
The details of each neighborhood in New York is displayed in figure 2. It shows the 305 neighborhoods in reference zone of the state of New York.

Figure 2



b) As mentioned in the introduction, the Peruvian Entrepreneur wants to open a new Peruvian restaurant in Manhattan city, hence the analysis will be focused in Manhattan's neighborhoods as depicted in the figure 3, Manhattan city counts with 40 neighborhoods.

Figure 3



c) Foursquare API with 100 venues of limit and 1000 meters radius is used to obtain the venues from Manhattan, which yields 3178 venues around the city. As shown in table 1, it displays the first five venues in Marble Hill neighborhood. For a space issue in the report only the first five venues are displayed in table 1.

Table 1

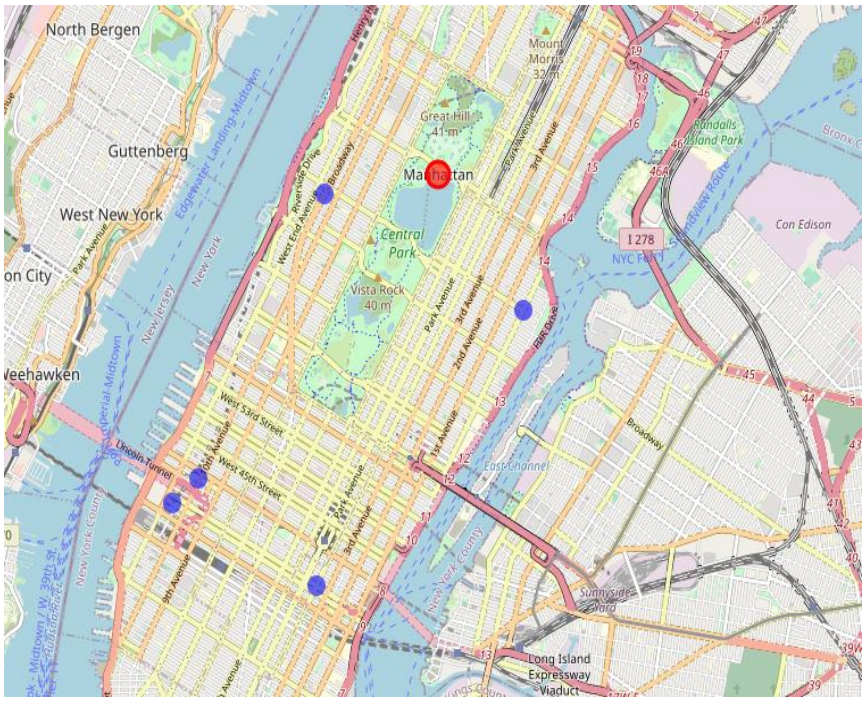
```
print(manhattan_venues.shape)
manhattan_venues.head()
```

```
(3178, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

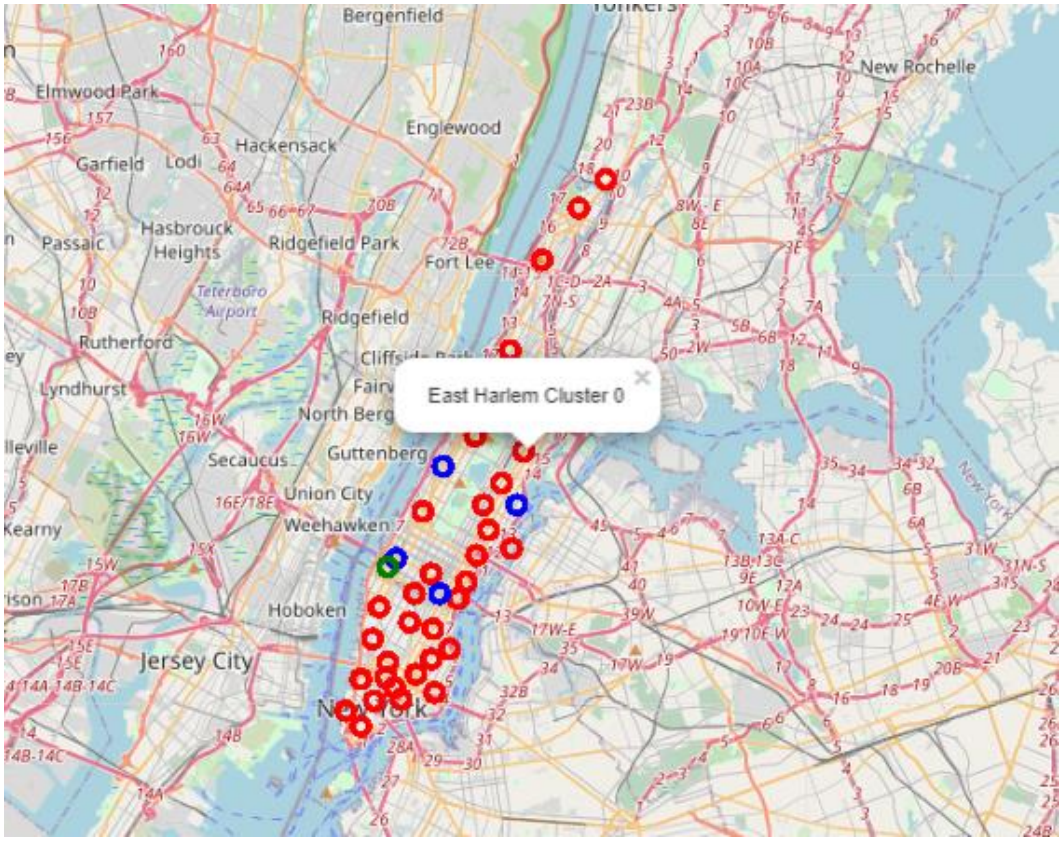
d) Once the venues from Manhattan are converted to data frame, it can be filtered to determine the number of Peruvian restaurants in operation in the city. As shown in the figure 4, there are five Peruvian restaurants around Manhattan city. The information from figure 4 emphasis the great opportunity to initiate a new gastronomic business.

Figure 4



e) The next evaluation is to know in which neighborhoods around Manhattan are the Peruvian restaurants. Therefore, the K-means clustering method is used to cluster the existing Peruvian restaurants by neighborhood. As depicted in figure 5, the Manhattan neighborhoods are grouped into three clusters that display the information needed to make decisions about the best place to open a new Peruvian restaurant. The blue cluster contains four restaurants in Clinton, Upper West Side, Murray Hill, and Hudson Yards. Likewise, there is another Peruvian restaurant in the green cluster located at Hudson Yards. On the other hand, the red cluster doesn't contain any Peruvian restaurant providing relevant evidence that the gastronomic business can be opened in any of the neighborhoods located in the red cluster.

Figure 5



The same information displayed in figure 5 is depicted in the table 2, 3 and 4. As shown in table 2 the red cluster is assigned to the zero-cluster label and it doesn't have any Peruvian restaurant inside of its neighborhoods. Otherwise, table 3 which is labeled with blue color and cluster label 1 displays four Peruvian restaurants which is by far the cluster

that contains more restaurants inside of Manhattan's neighborhoods. Finally, the cluster label 2 that shown in green color has only one restaurant in Hudson Yards neighborhood.

Table 2

```
#Cluster 0
to_merged.loc[(to_merged['Cluster Labels'] ==0) & (to_merged['Venue Category'] == 'Peruvian Restaurant') ]
```

Neighborhood	Peruvian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category

Table 3

```
#Cluster 1
to_merged.loc[(to_merged['Cluster Labels'] ==1) & (to_merged['Venue Category'] == 'Peruvian Restaurant') ]
```

	Neighborhood	Peruvian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
26	Murray Hill	0.010204	1	40.748303	-73.978332	Pio Pio	40.745535	-73.977626	Peruvian Restaurant
39	Yorkville	0.010000	1	40.775930	-73.947118	Pio Pio	40.779887	-73.947202	Peruvian Restaurant
6	Clinton	0.010000	1	40.759101	-73.996119	Pio Pio	40.760636	-73.994714	Peruvian Restaurant
36	Upper West Side	0.011494	1	40.787658	-73.977059	Flor de Mayo	40.785966	-73.976312	Peruvian Restaurant

Table 4

```
#Cluster 2
to_merged.loc[(to_merged['Cluster Labels'] ==2) & (to_merged['Venue Category'] == 'Peruvian Restaurant') ]
```

	Neighborhood	Peruvian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Hudson Yards	0.017857	2	40.756658	-74.000111	Chirp	40.753377	-73.996116	Peruvian Restaurant

5) Results

In this section, the answers released in the business case will be answered, having as basis the analyzed data.

a. How many Peruvian restaurants are in Manhattan?

Based on the evaluated data, there are five Peruvian restaurants around Manhattan.

b. What is the best location in Manhattan City to open a Peruvian restaurant?

Based on the k-means clustering algorithm used in Manhattan neighborhoods' data, the best place for opening a new Peruvian restaurant is in the red clustered neighborhoods. The optimal site can be in an area in the red cluster, not far from an

existing restaurant on the blue cluster. For example, a possible optimal location could be the East Harlem neighborhood. As shown in figure 5, the East Harlem area is not far from the Yorkville neighborhood that is clustered with blue color.

c. What are other potential neighborhoods in Manhattan for Peruvian Cuisine?

The previous answer states that the best option is to open a restaurant in a clustered neighborhood in red that is not far from an existing Peruvian restaurant located in the blue cluster. Another option is to open a Peruvian restaurant in a red cluster neighborhood but far away from a popular Peruvian restaurant because it can be an excellent opportunity to obtain new customers without any similar competition around the neighborhood. Another recommendation is to initiate the gastronomic business in the Hudson Yards neighborhood. There is only one Peruvian restaurant with an excellent reputation and opening another one can be a great alternative.

d. What neighborhoods lack Peruvian restaurants?

The Manhattan neighborhoods in the red cluster don't have a Peruvian restaurant. This information is relevant because it shows a great business opportunity to open a new Peruvian restaurant in one of these neighborhoods.

6) Recommendations

Based on the evaluation made, the Manhattan city provides a great business opportunity to initiate a new gastronomic business of Peruvian food. The document shows that there are only five Peruvian restaurants around Manhattan city where each restaurant is in a different neighborhood. Thirty-five neighborhoods in Manhattan don't have Peruvian restaurants. Therefore, a new Peruvian restaurant can be opened in one of these neighborhoods, not far from an existing Peruvian restaurant located in the blue cluster, for instance in the East Harlem neighborhood.