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 business 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

The methodology established for this business case follows the process shown in figure 1. It starts by acquiring the data and cleaning it. In this process, the data is collected from the link https://cocl.us/new_york_dataset and converted to a data frame format using Pandas. After the data is ready in a data frame format, the second process of exploratory data analysis starts by evaluating it to obtain relevant information that serves as input to the predictive model. Finally, one predictive model is chosen to observe the best location for opening a new Peruvian Restaurant in Manhattan. How we are exploring neighborhoods in Manhattan, the k-means clustering will be used for this proposal. K-means is an unsupervised machine learning algorithm that groups similar data points by trying to discover fundamental patterns. Hence it must fix numbers of K to make the clusters of the dataset.

Figure 1

a) Data acquisition and cleaning.

The evaluation starts by acquiring the data of New York neighborhoods from the following URL link https://cocl.us/new_york_dataset ,then the raw data in json extension is conversated to data frame format using pandas, a shown in table 1, it is filtered to show only the first five rows of data for a space issue.

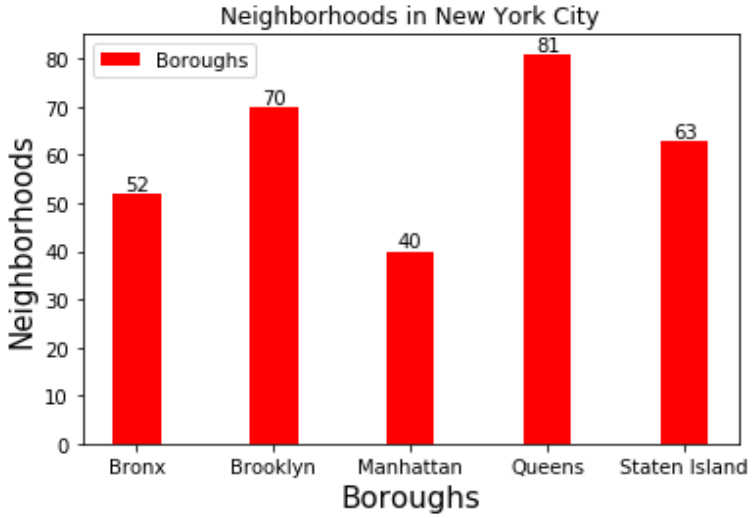
Table 1

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

b) Exploratory Data Analysis.

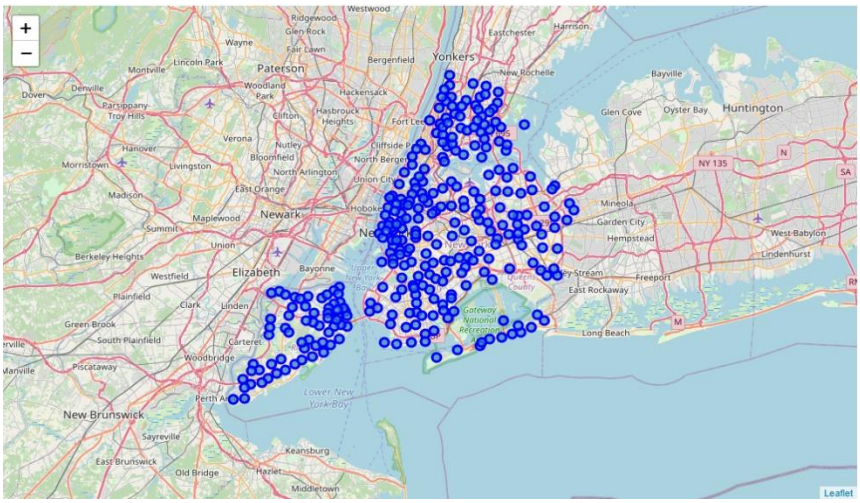
The evaluation in the second process begins by determining the number of boroughs and neighborhoods in New York. The data shows that New York has 5 boroughs and 306 neighborhoods as depicted in figure 2. Likewise, the graph displays that Manhattan is the borough with the lower number of neighborhoods. Even though Manhattan has the lower number of neighborhoods, it has the most popular restaurants in New York. Thus, the new gastronomic businesses will initiate in one of the neighborhoods from Manhattan.

Figure 2



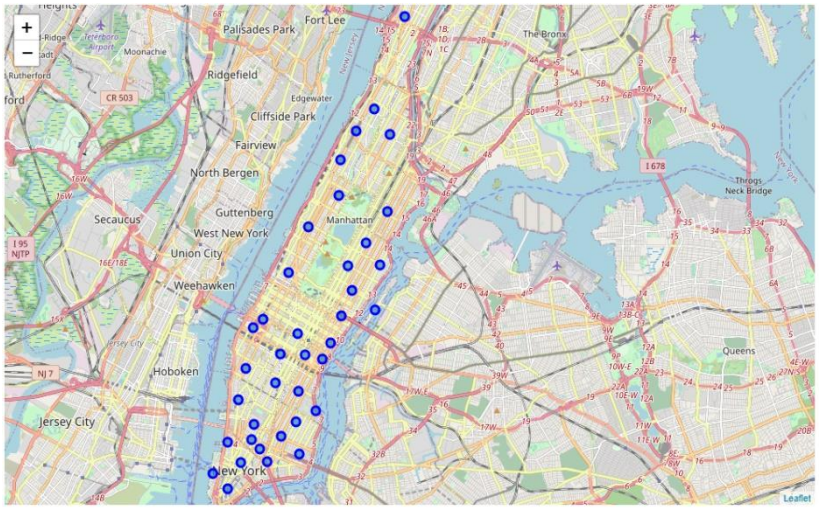
The details of each neighborhood in New York can also be displayed in the map of figure 3 utilizing Folium visualization library. It shows the 306 neighborhoods in reference zone of the state of New York.

Figure 3



As mentioned in the introduction, the Peruvian Entrepreneur wants to open the new Peruvian restaurant in Manhattan city, hence the evaluation will be focused in Manhattan's neighborhoods as depicted in the figure 4. The previous map of New York neighborhoods is segmented and clustered only for the Manhattan Neighborhoods.

Figure 4



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 2, 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 2

```
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

Once the venues from Manhattan are converted to data frame, we use dummies function to convert it from categorical to numerical values. Then the data frame can be grouped by

neighborhoods by taking the mean of the frequency of occurrence of each category, as shown in table 3, it shows the first ten neighborhoods grouped by each category.

Table 3

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	Battery Park City	0.000000	0.00	0.00	0.000000	0.014925	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Carnegie Hill	0.000000	0.00	0.00	0.000000	0.010989	0.00	0.000000	0.000000	0.010989	0.000000	0.010989	0.000000
2	Central Harlem	0.000000	0.00	0.00	0.066667	0.044444	0.00	0.000000	0.000000	0.000000	0.022222	0.000000	0.000000
3	Chelsea	0.000000	0.00	0.00	0.000000	0.040000	0.00	0.000000	0.010000	0.000000	0.050000	0.000000	0.000000
4	Chinatown	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Civic Center	0.000000	0.00	0.00	0.000000	0.030000	0.01	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000
6	Clinton	0.000000	0.00	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000
7	East Harlem	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	East Village	0.000000	0.00	0.00	0.000000	0.020000	0.00	0.000000	0.010000	0.010000	0.010000	0.000000	0.010000
9	Financial District	0.000000	0.00	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Flatiron	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000

To display the actual Peruvian restaurants, a justification is made to grouped values to look for Peruvian restaurants around Manhattan. It yields Peruvian restaurants operating in different neighborhoods as sown in table 4.

Table 4

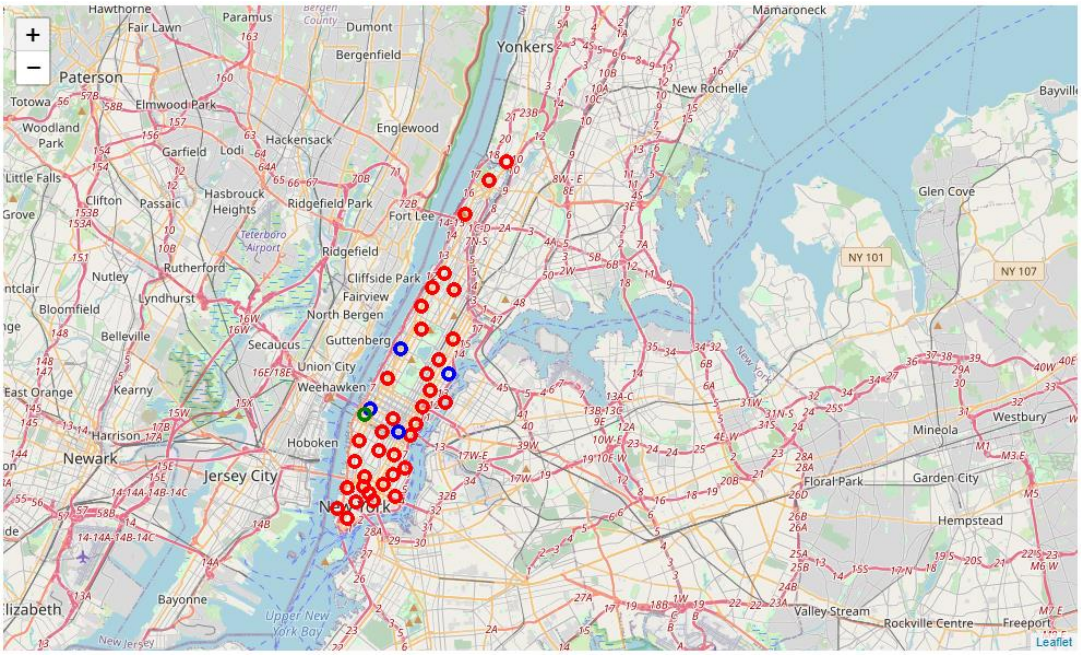
	Neighborhoods	Peruvian Restaurant
0	Battery Park City	0.0
1	Carnegie Hill	0.0
2	Central Harlem	0.0
3	Chelsea	0.0
4	Chinatown	0.0

c) Predictive Modeling.

The next step is to cluster the current Peruvian restaurants inside the Manhattan neighborhoods. K-means is an unsupervised machine learning algorithm that groups similar data points by trying to discover fundamental patterns. Hence it must fix numbers of K to make the clusters of the dataset. It is very simple tool for clustering and clearly suited for modeling the data in this project. The Manhattan neighborhoods are grouped into 3 clusters based in the frequency of occurrence of Peruvian restaurants. 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 Neighborhoods with more Peruvian Restaurants. The green cluster has less Peruvian Restaurants. 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



- ✓ Cluster 0: Neighborhoods with cero Peruvian restaurants
- ✓ Cluster 1: Neighborhoods with more Peruvian Restaurants
- ✓ Cluster 2: Neighborhoods with less Peruvian Restaurants

The same information displayed in figure 5 is depicted in the table 5, 6 and 7. 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 5

```
#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 6

```
#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 7

```
#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 Peruvian 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. The reason to open the new Peruvian restaurant not far from an actual Peruvian restaurant clustered in blue color is that actual restaurants already have a gained reputation, thus it can be favorable to open the new restaurant not far from a restaurant located in the blue cluster.

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

Another option is to open a Peruvian restaurant in a red cluster neighborhood but far away from an existing Peruvian restaurant because it can be an excellent opportunity to obtain new customers without any similar competition around the neighborhood.

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. This evaluation shows that most of the Peruvian restaurants are in the blue cluster neighborhoods (cluster 1) which is around Murray Hill, Yorkville, Clinton, and Upper West Side. The green cluster neighborhoods (cluster 2) has the less Peruvian restaurants around Manhattan. Finally, the red clusters (cluster cero) doesn't have Peruvian restaurants in its neighborhoods. Therefore, a new Peruvian restaurant can be opened in one of these neighborhoods clustered in red, not far from an existing Peruvian restaurant located in the blue cluster, for instance in the East Harlem neighborhood that is clustered in red.