

Identifying Ideal Neighborhoods for a New Mexican Restaurant in New York City

Connor O'Dell

1 Introduction

One of the most important decisions any new brick-and-mortar business must make is the physical location of the business. For a restaurant, there are many competing interests in this choice like: the cost of the location; proximity to other venues like offices, entertainment venues, other restaurants, and bars; the amount of foot traffic; and the trendiness of the neighborhood.

For our hypothetical problem, an enthusiastic young chef plans to open a new high-end Mexican restaurant in New York City. The status of Mexican food as a serious cuisine is rising with the elevation of restaurants like [Pujol](#) in Mexico City to world-wide fame, and the dabbling of [Noma](#), widely considered the best restaurant in the world, in Mexican cuisine, he thinks there is a prime opportunity for such a venue in the restaurant capital of the US. But in order for the business to be a success, the chef knows he needs to find the perfect location.

In order to find the best for his new restaurant, the chef would like us, the data scientists, to find answers to the following questions.

1. Are there any similar restaurants already open in New York, and if so, where are they?
2. What kinds of neighborhoods are the existing Mexican restaurants in, and what neighborhoods are most similar?
3. Are there any neighborhoods that are similar to those with existing Mexican restaurants that don't currently have a Mexican restaurant?

2 Data

In order to answer the three questions posed by our client, the source of venue data will be Foursquare venue data via the Foursquare API. In order to use the Foursquare data, we will also need location data for each neighborhood in New York. For this purpose, we will use [New York City neighborhood data](#) from New York University. We built a dataset of venue information using the Foursquare

API of venues within 500 meters of the location of neighborhood. We collected the name, longitude, latitude, and category of each venue.

An important subset of our venue data is the set of restaurants that fall under the Mexican Restaurant category. Our chef-client is interested in high-end dining, so we will cull national chains of casual restaurants from the set of Mexican restaurants in order to make our set of restaurants relevant to his business. In total, we removed 18 Chipotle locations, 3 Qdoba locations, and 2 food trucks. This left us with 152 Mexican restaurants in New York City.

In order to generate useful graphics for the client, we used the Folium library in Python to plot locations of neighborhoods and venues. Finally, we accessed the location data for New York City using the Geopy library in Python.

3 Methodology

For the first question, we will provide a visualization of the data collected on the locations of Mexican restaurants.

3.1 Question #2

In order to answer the second question for the client, we need to begin by clustering all of the neighborhoods of New York based on the venues in each neighborhood. In order to apply machine learning algorithms to the data, we built find the relative frequency of each venue categories for each of the neighborhoods. Then we cluster the neighborhoods using this data to find neighborhoods that contain the same kinds of venues.

For the clustering, we used k-means clustering to group the neighborhoods together. On the assumption that New York is a diverse city and will have many different kinds of neighborhoods, we chose a large number of clusters (40) to use in the algorithm. A color coded map of the resulting clusters can be seen in Figure 1.

Finally, we labeled each Mexican restaurant with the cluster that its neighborhood belongs to.

3.2 Question #3

Once each restaurant was associated with a cluster, we could find the number of Mexican restaurants in each of the clusters. The results can be read from the bar chart in Figure 2. Thus we can answer the final question by finding the neighborhoods that do not currently have a Mexican restaurant in those clusters that already have Mexican restaurants.

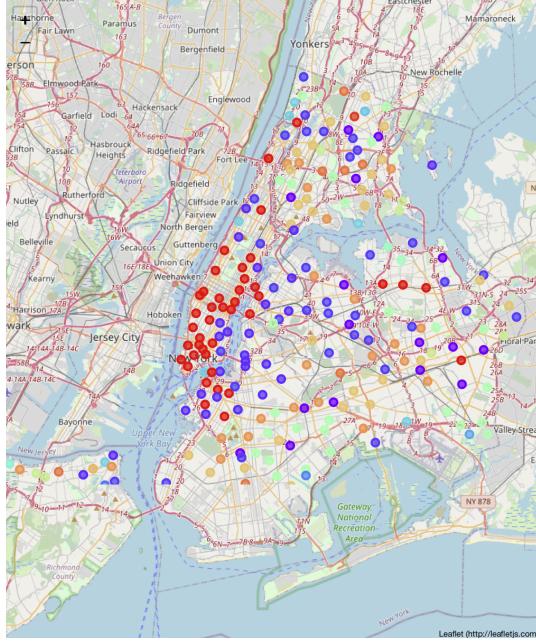


Figure 1: Color coded clusters of neighborhoods in New York

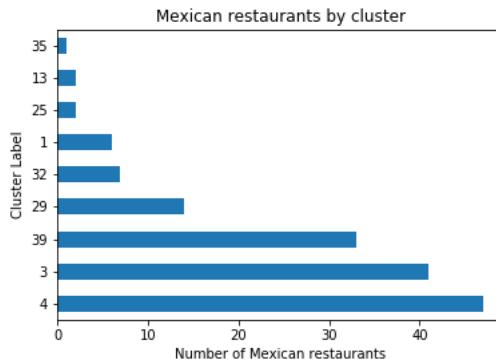


Figure 2: The number of Mexican restaurants in each cluster of neighborhoods

4 Results

4.1 Question #1

To answer this question, we provide our client with a map of all of the Mexican restaurants in New York. This can be seen in Figure 3. We can also provide the client with the raw data in table form.

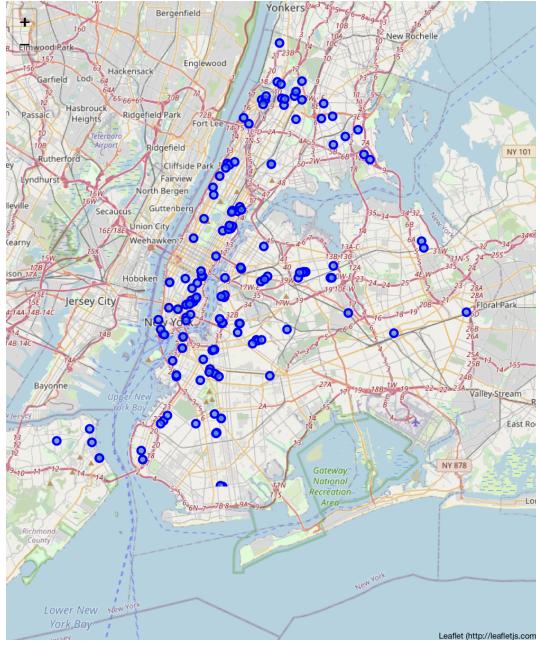


Figure 3: The Mexican restaurants of New York

4.2 Question #2

To give a clear picture of the distribution of restaurants among the clusters, we provide the client with a map of the Mexican restaurants color coded by cluster. The client can observe the geographic distribution of Mexican restaurants in the top three clusters in the bar chart in Figure 5.

4.3 Question #3

From Figure 2, it is clear that only three clusters contain the vast majority of Mexican restaurants. Thus we make the decision to focus on clusters 4, 3, and 39. A naive answer to this question would be to list of all of the neighborhoods that don't have Mexican restaurants in the three clusters that have the most Mexican restaurants. But this is a large list that could be unmanagable for the client to sort through, so we need to narrow the focus. To do this, we look at the top three clusters to see which boroughs to which the neighborhoods belong. This can be seen in the bar chart in Figure 5.

Considering the data displayed in Figure 5 and knowing that the client wishes to open a high-end restaurant, we suggest that the client focus on the most cosmopolitan cluster, number 39. Cluster 39 includes many neighborhoods of Manhattan whereas the other two include large parts of Brooklyn and Queens which are more residential.

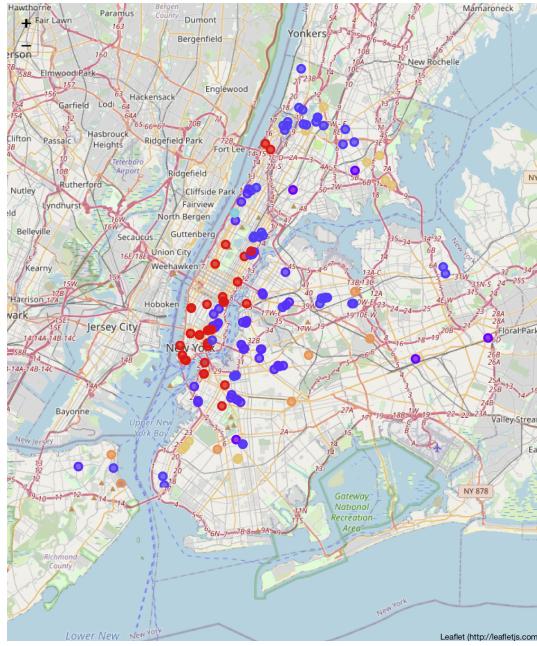


Figure 4: Map of the Mexican restaurants of New York color coded by cluster

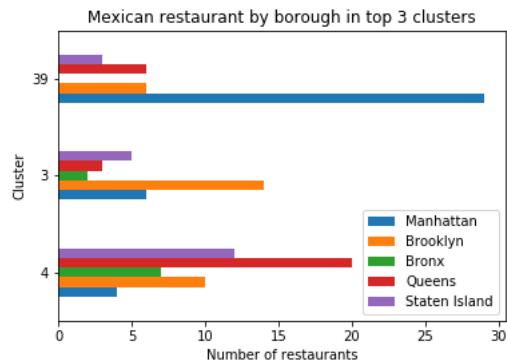


Figure 5: The number of neighborhoods in each borough for the top three clusters

We provide the client with a map of the neighborhoods in cluster 39 that don't currently have a Mexican restaurant. This can be seen if Figure 6.

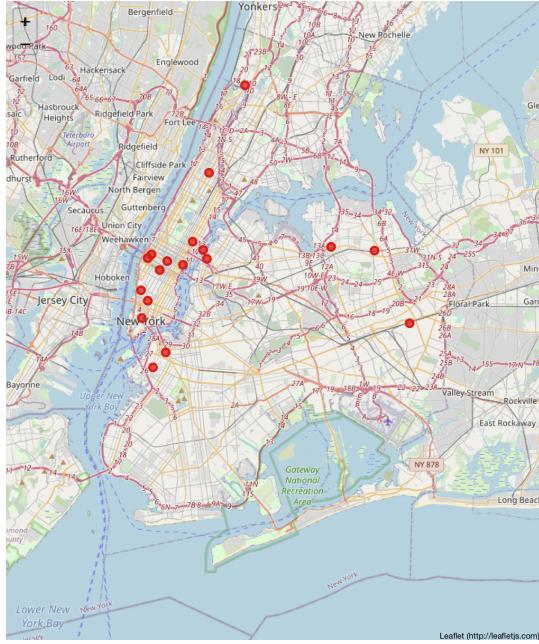


Figure 6: Recommended neighborhoods of New York for the location of the new restaurant

5 Discussion

Our primary recommendations to the client for the location of his new restaurant are contained in the map above. We believe that a new Mexican restaurant would be well-suited to any of these neighborhoods, and there would be no direct competitors. The neighborhoods include well-known parts of New York like Downtown Brooklyn, Harlem, and Hudson Yards along with several more surprising results further away from the heart of Manhattan. With this data in hand, we expect the client can take a focused approach when he begins searching for the best property to locate his new restaurant.

In order to refine our results and provide more focused recommendations, there are several steps that could be carried out. First, we would like to collect data about the price point of each of the Mexican restaurants in our data set. The client wants to open a high-end restaurant, so it would be helpful to know which, if any, of the neighborhoods support an existing high-end Mexican restaurant. Second, it could be helpful to include restaurants from cuisines adjacent to Mexican like Argentinian, Columbian, and Brazilian in our data set on the assumption that these restaurants would attract the same kinds of diners as the planned new restaurant.

6 Conclusion

The goal of this project was to locate neighborhoods in New York that an upcoming chef should consider as locations for his new, upscale Mexican restaurant. In order to complete this project, we used venue data from Foursquare. This data provided the locations of all of the Mexican restaurants in New York City and was the basis for our clustering of neighborhoods. We used k-means clustering to group similar neighborhoods together based on the kinds of venues they contain. In the end, we located neighborhoods that do not currently have a Mexican restaurant but are similar to those that do support one. We think that one of these locations could be the ideal place for a new restaurant to flourish.