The Batlle of the Neighborhoods

African Restaurant at New York City

Capstone Project

Author:

Carlos Djedje

.

October 2020

1. Introduction

Uhuro is a well-known group of restaurants in Africa, and has a wide range of menus that are a faithful representation of African cuisine. With a franchise of more than 15 restaurants spread across different African countries, this group is recognized for its recognized high level of quality in its services.

This caused the Uhuro group to attract the attention of large investors who bet on expanding the group of results worldwide. Such expansion will occur in a phased manner, seeking to cover the world's major financial cities such as New York, Dubai, Shanghai, London etc.

In this process, the first city to receive the expansion is New York City. With this in mind, the Uhuro group decided to hire data science services to assess in which neighborhoods it would be best to install the restaurant.

1.2. Problem

New York City is known for its diversity in international cuisine, making it possible to find the best of each continent or country around every corner of the city. This factor presents a great risk for new investments in the food industry.

My client wants to open an African restaurant in New York City. The quality of service and field of cultures of African dishes the Uhuro group already has, the problem faced is the deficient knowledge that stakeholders have about the place where they intend to open the establishment.

This lack of knowledge constitutes a disadvantage and risks, as the choice of the wrong place will affect the success of the undertaking and may alienate investors from the expansion project, so it is in the group's interest to obtain enough data to make decisions.

2. Description of the data

In order to be able to offer the Uhuro group data with which they can safely make a decision, it is important to first establish that the question to answer must be What are the neighborhoods with the least restaurants to offer African cuisine?

To answer this question, we need to obtain geographic information from the neighborhoods of New York City and then obtain information on the types of restaurants in those areas. To obtain the geographic data of New York, the following website was used: https://geo.nyu.edu/catalog/nyu_2451_34572 where it was possible to obtain information about all the neighborhood data of the city of New York.

This dataset has data such as name of neighborhoods, name of Borough and geographic coordinates, with this information it is possible through a powerful platform to obtain information about the establishments around it, this platform is Foursquare.

With Foursquare we will make requests to the API to explore the venues of each neighborhoods in New York City, this API returns the information in JSON format, and there are all venues in the respective neighborhoods and more importantly, for each venue found its category and for the case of our study we are interested in the category of African restaurant.

With the use of these two datasets it will be possible to make an analysis and assess which of the neighborhoods are most recommended for the Uhuro group to open its restaurant.

3. Data acquisition and cleaning

3.1. Data acquisition

As stated above, to carry out this activity will be necessary to obtain information on the neighborhoods of New York City and for each one of the neighborhoods have information about their venues.

In order to obtain information about New York City neighborhoods, a survey was carried out with the aim of finding such information in the GeoJson format, this format already presents the desired coordinates, such a search resulted in the following website: https://geo.nyu.edu/catalog/nyu_2451_34572 that made it possible to obtain the desired information.

After finding a website for information, it was necessary to choose a tool that would make it possible to have information about the venues of New York City neighborhoods, with which the foursquare platform was chosen.

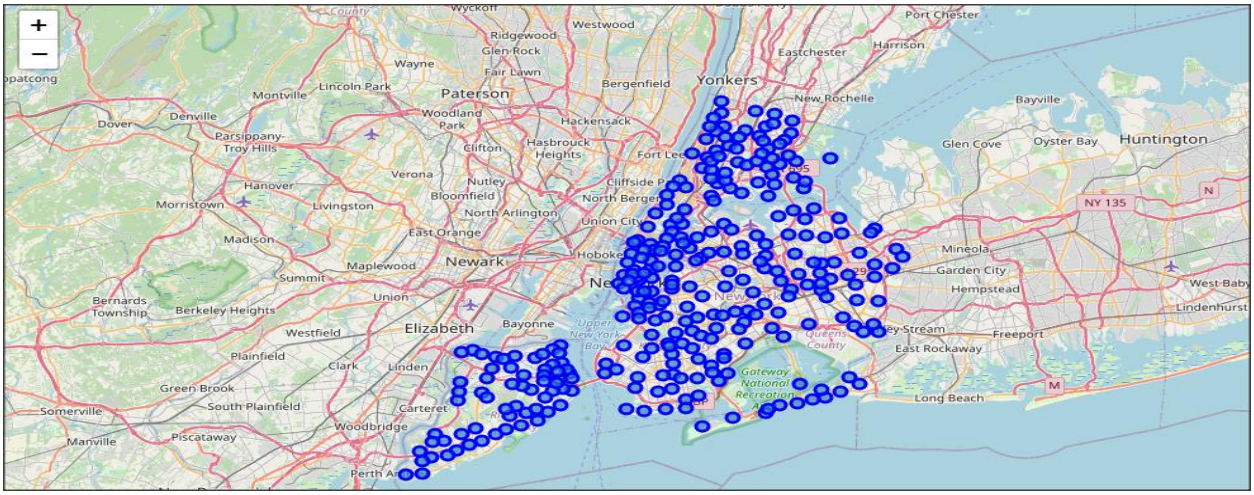
3.2. Data cleaning

As a first step in this process, we read the geojson with New York City's neighborhood data, after reading this information we can see that our dataset is composed of the following information:

Out[10]:

	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

Note that the figure above only shows 5 lines of our dataset. This dataset presents information about New York City's Borough, Neighborhood, Latitude and Longitude. In order to confirm that the coordinates are all within the city of New York, we can make a visualization of the city map.



Obtaining the data from the neighborhoods, the next step is to obtain the data from the venues, to obtain this data we use the Foursquare platform, where we make requests for its API to obtain the venues, in this request we intend to obtain the 100 venues near the neighborhoods in one 500m radius,

and at the end for each of the neighborhoods we list their venues. With this junction we obtain a new dataset where we have the list of each venue for each neighborhood, as shown in the figure below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

In the image above we can see that we have the name of the neighborhood, the latitude and longitude of the neighborhood, the name of the venue, the latitude and longitude of the venue, and to which category this venue belongs.

4. Analysis

In order to carry out the analysis, it is important to remember the question we must answer, which are the neighborhoods with the least restaurants with African cuisine, as these neighborhoods are more likely to receive a new establishment without competition. In order to facilitate the analysis we will make some changes to our dataset.

We will transport the information about the venues categories to the dataset columns and convert the information into numbers, we use the get_dummies method of the pandas library. After such a change we obtain the dataset in the following format.

	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Al Cra Stc
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0

As we can see, we have as one of the categories of our dataset the African Restaurant, as the type of establishment that the Uhuro group intends to open belongs to this category, we can filter our dataset and focus our analysis only on this category, thus guaranteeing improvement performance in the process to be carried out. After performing this filter we obtain a dataset with the following format:

	Neighborhood	African Restaurant
0	Wakefield	0
1	Wakefield	0
2	Wakefield	0
3	Wakefield	0
4	Wakefield	0

As we can see due to the junction of the neighborhoods dataset and the venues of each of them, our dataset made a cardinality between them, so we have the neighborhoods repeated for the number of times that the venue exists, this formatting does not help in the analysis. For the analysis to be concise we will group the neighborhoods and calculate the means, then obtaining a dataset with the information shown below.

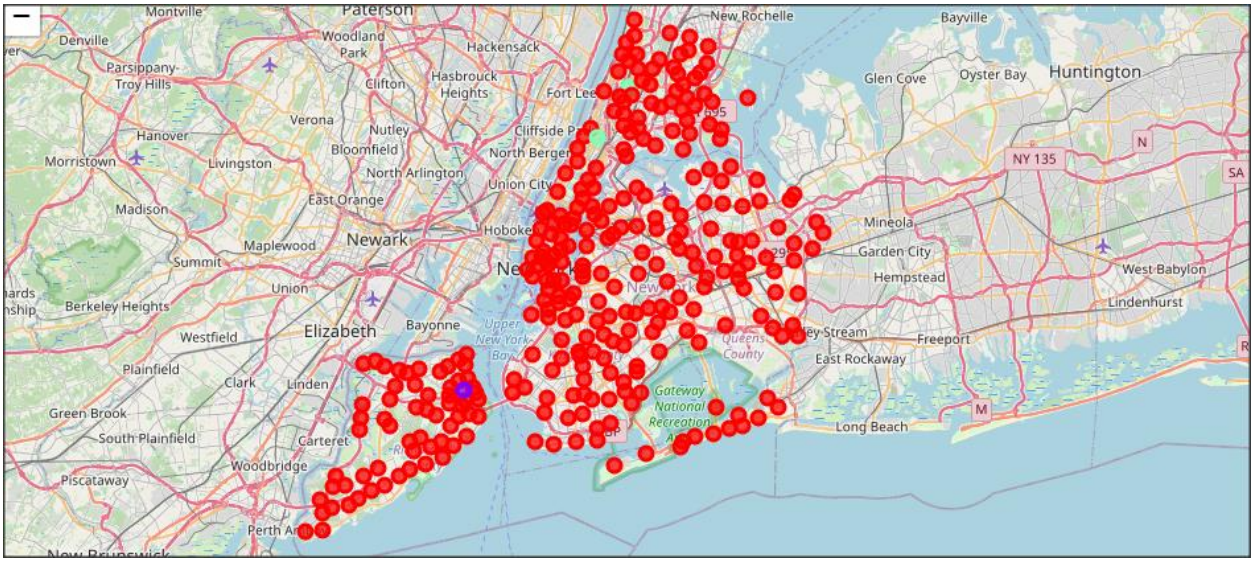
	Neighborhood	African Restaurant
0	Allerton	0.000000
1	Annadale	0.000000
2	Arden Heights	0.000000
3	Arlington	0.000000
4	Arrochar	0.000000
5	Arverne	0.000000
6	Astoria	0.000000
7	Astoria Heights	0.000000

Clustering

We will now segment neighborhoods in order to have information about which neighborhoods have the most African restaurants in relation to the others. To carry out this activity, we will use the kmeans algorithm. We set up 3 clusters, as we want to classify neighborhoods as low index, medium index and high index of African gastronomy restaurants. After applying the algorithm to the prepared dataset we get the following result:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	African Restaurant
0	Bronx	Wakefield	40.894705	-73.847201	0	0.0
1	Bronx	Co-op City	40.874294	-73.829939	0	0.0
2	Bronx	Eastchester	40.887556	-73.827806	0	0.0
3	Bronx	Fieldston	40.895437	-73.905643	0	0.0
4	Bronx	Riverdale	40.890834	-73.912585	0	0.0

As we can see, after the execution of the algorithm we obtain the following result, where we have for each neighborhood the classification attributed to it, we still have the information of the geographic coordinates as we can see above. With this information we can make a visualization on the map of such neighborhood, and distinguishing the 3 classes created with different colors for better understanding.

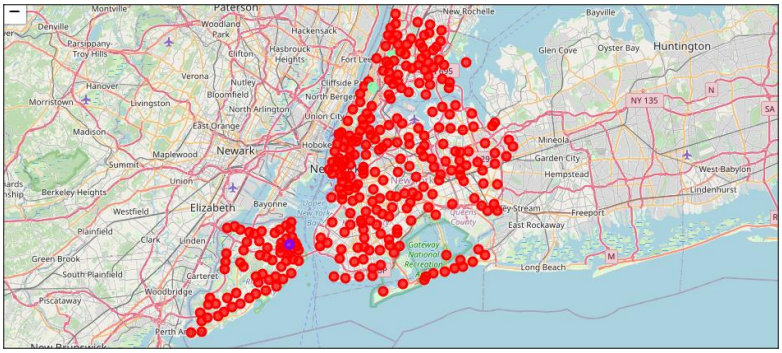


5. Results

After performing the neighborhood classification procedure, where this procedure consisted of first obtaining data from New York City neighborhoods, and then using the Foursquare platform, make requests to its API to obtain the venues of the previously mentioned neighborhoods. Remember that these requests were made with the parameters of 100 venues and a 500m radius, so in the end we had a dataset that listed the venues by neighborhood. The next step was to group this information by neighborhood and calculate its means and then apply the kmeans clustering algorithm.

This algorithm was used taking into account that the desired number of clusters would be 3, where we would classify them among neighborhoods with a low rate of African restaurants, neighborhoods with a medium rate of African restaurants and neighborhoods with a high rate of African restaurants in nearby areas.

With this application made on the dataset, it was possible to obtain the desired classifications for the neighborhoods of the city of New York, we can review the visualization of this classification on the map once more in order to make an interpretation of the results obtained, the map in question is illustrated below:



As we can see in the graphic above, we have dots painted in three different colors, namely red, green and purple. Remember that these points are the neighborhoods of New York City and the colors represent which cluster this neighborhood is part of. The red color represents cluster 0, this cluster is the classification of the neighborhood with the lowest or zero number of restaurants with African cuisine. The purple color represents the neighborhoods with an average number of restaurants with African cuisine and finally the green color represents the neighborhoods with the largest number of restaurants with African cuisine.

Cluster 1 - represented by 0 in the dataset:

	Neighborhood	Longitude	Cluster Labels	African Restaurant
0	Wakefield	-73.847201	0	0.000000
1	Co-op City	-73.829939	0	0.000000
2	Eastchester	-73.827806	0	0.000000
3	Fieldston	-73.905643	0	0.000000
4	Riverdale	-73.912585	0	0.000000

The image above illustrates the neighborhoods classified with the lowest number of restaurants with African cuisine, it should be noted that the image shows only the first 5 restaurants on this list. In total, we have 304 neighborhoods on this list

Cluster 2 - represented by 1 in the dataset:

	Neighborhood	Longitude	Cluster Labels	African Restaurant
305	Fox Hills	-74.08174	1	0.2

In cluster 2, we have the list of neighborhoods with an average African restaurant index.

Cluster 3 - represented by 2 in the dataset:

	Neighborhood	Longitude	Cluster Labels	African Restaurant
14	University Heights	-73.910416	2	0.041667
105	Central Harlem	-73.943211	2	0.066667

In cluster 3, we have the list of neighborhoods with an average index of African cuisine restaurants.

6. Conclusion

In this project, your main objective was to help the Uhuro group to carry out a great task of analyzing which neighborhood in New York City they could start with in their expansion process. The Uhuro group is a very prestigious restaurant group on the African continent and has already received several awards in this area, and this fact has attracted investors to the group and intend to take the group's services to other corners of the world. It was with this assumption of expansion that the group decided to hire our data science services so that they could make a concise analysis about the places to install the establishment. After collecting all the reverent data for the neighborhoods of New York City and crossing this same data with the venues next to each one, we had results that classified these neighborhoods into 3 different groups, the neighborhoods with a small index of restaurants with African cuisine, neighborhoods with an average rate of restaurants with African cuisine and neighborhoods with a high rate of restaurants with African cuisine. With these results it was possible to verify that New York City has several neighborhoods with a low rate of restaurants with African cuisine, making it even easier to indicate the neighborhoods with medium and high rates or the neighborhoods less recommended to carry out the installation of the restaurant. With this, we conclude that there is a wide variety of neighborhoods where the Uhuro group can install its service, where in these locations they have no close competitors, thus being able to have the advantage of having a unique and differentiated service.