Analysis to locate a Colombian food restaurant in Manhattan: Final Project

Capstone Project - The Battle of Neighborhoods (Week 2)
https://github.com/ElkinSn/Coursera_Capstone/blob/main/FinalProject.ipynb

ELKIN R. SANTAFÉ R.
Jan. 19 / 2021

INTRODUCTION

BUSINESS PROBLEM

Colombia is a country that within Latin America enjoys having a unique cultural richness, which is reflected in its gastronomy. The ethnic diversity of Colombia manifests itself in Colombian gastronomy. In Colombian food there's a rich mix of Spanish and Indigenous ingredients with African, Arab and Spanish preparation.

The main problem to be solved is to establish the potential of a Colombian food restaurant in Manhattan. We want the restaurant to be unique, and stand out strongly with the neighboring restaurants through the diversity in the menu. However, New York is known to have a particular Latino influence that is seen not only in the arts but also in food.

BUSINESS PROBLEM

The menu will design including the most typical Colombian plates that everybody should try:

Ajiaco

Hormiga culona

Lechona

Bandeja Paisa

Sancochos

Tamales

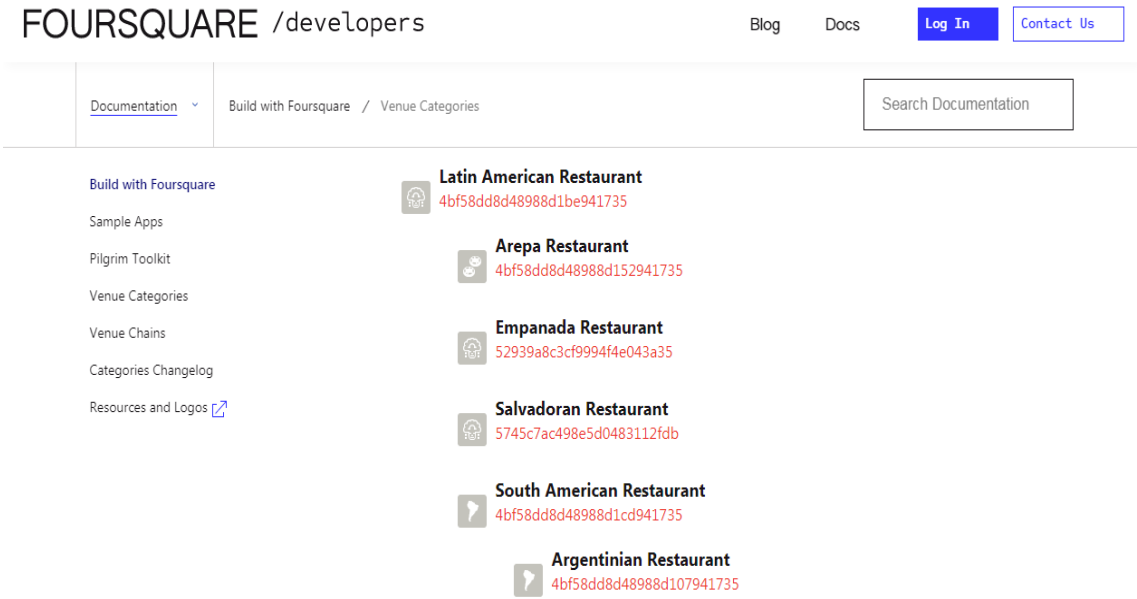
Arepas and Empanadas

Patacón with everything

DATA

DATA

Map of the five boroughs of New York (and New Jersey on the left and Long Island on the right). Area of interest is indicated.



ID codes for restaurants from FOURSQUARE

METHODOLOGY

METHODOLOGY

- **Preparing data base**
  - Importing Libraries
  - Downloading files and placing them on the server
  - Transforming the data into a *pandas* data frame
- **Making the base map**
  - Using *geopy* library to get the latitude and longitude values of New York City.
  - Creating a map of New York with neighborhoods
- **Integrating the information of FOURSQUARE venues**
  - Looking for Colombian restaurants
  - Looking for Latino Restaurants
- **Generating competitors map**
- **Grouping the information by neighborhoods**.
- **Analyzing by neighborhood**.
  - Raking the venues
- **Clustering the neighborhood**
  - Clustering process
  - Generating new table
  - Generating clusters map
  - Generating tables by cluster (with the top 10 of venues)
- **Analyzing the information (from tables)**

METHODOLOGY

The Machine Learning analysis was implemented in the clustering section where using the *Kmeans* criteria for grouping the neighborhoods. A clustering process generated different tables (ranked on 10) with the top ten of venues related to Latino restaurants. The idea is located the cluster with the less offer of Latino food.

RESULTS

Preparing data base

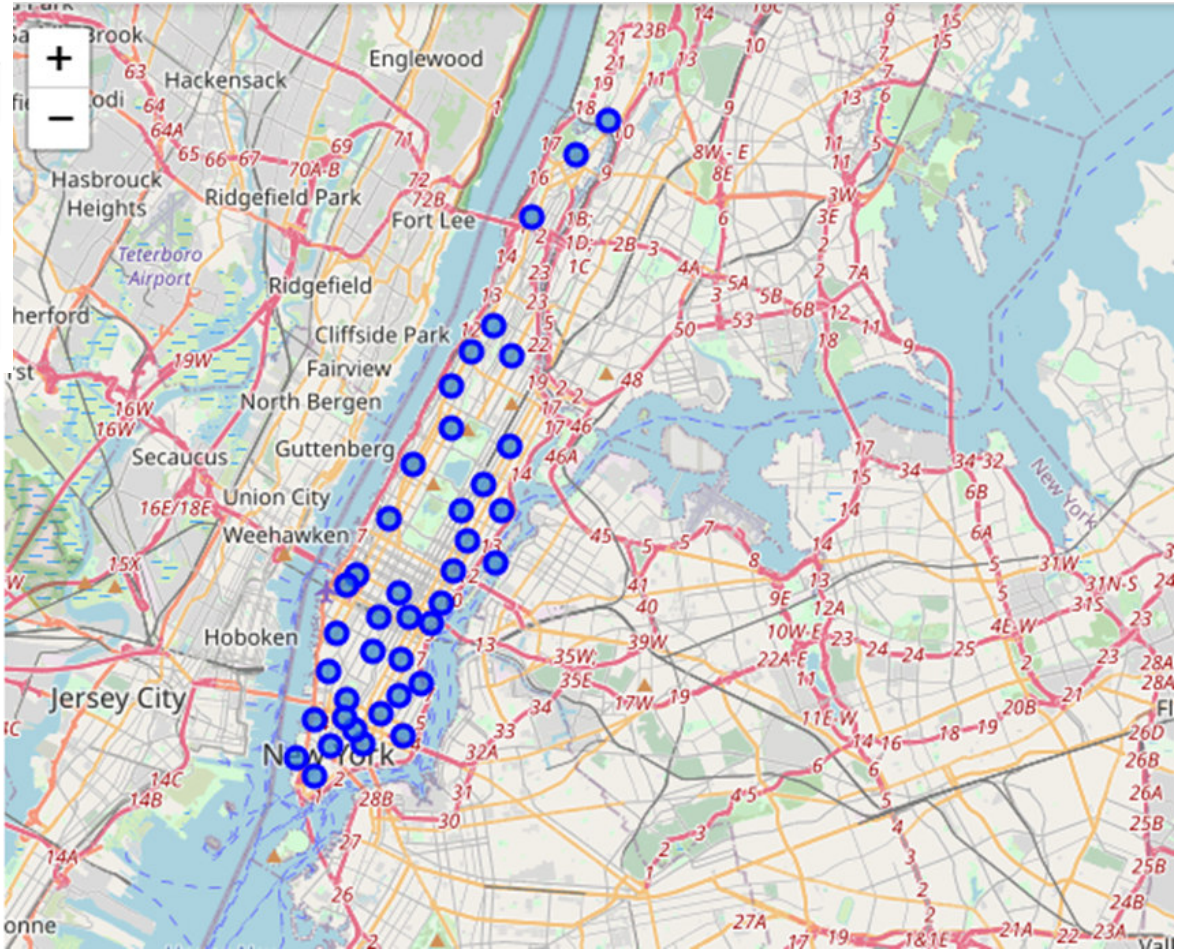
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Table result of preparing data base

Making the base map

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Table with coordinates of neighborhoods



Map with neighborhoods

Integrating the information of FOURSQUARE venues

>> Looking for Colombian Restaurants

```
[ ]  #https://developer.foursquare.com/docs/resources/categories
     #Colombian Restaurant: 58daa1558bbb0b01f18ec1f4
     neighborhoods = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(d
     newyork_venues_colombian = getNearbyVenues(names=neighborhoods['Neighborhood'], lati
     newyork_venues_colombian.head()
```

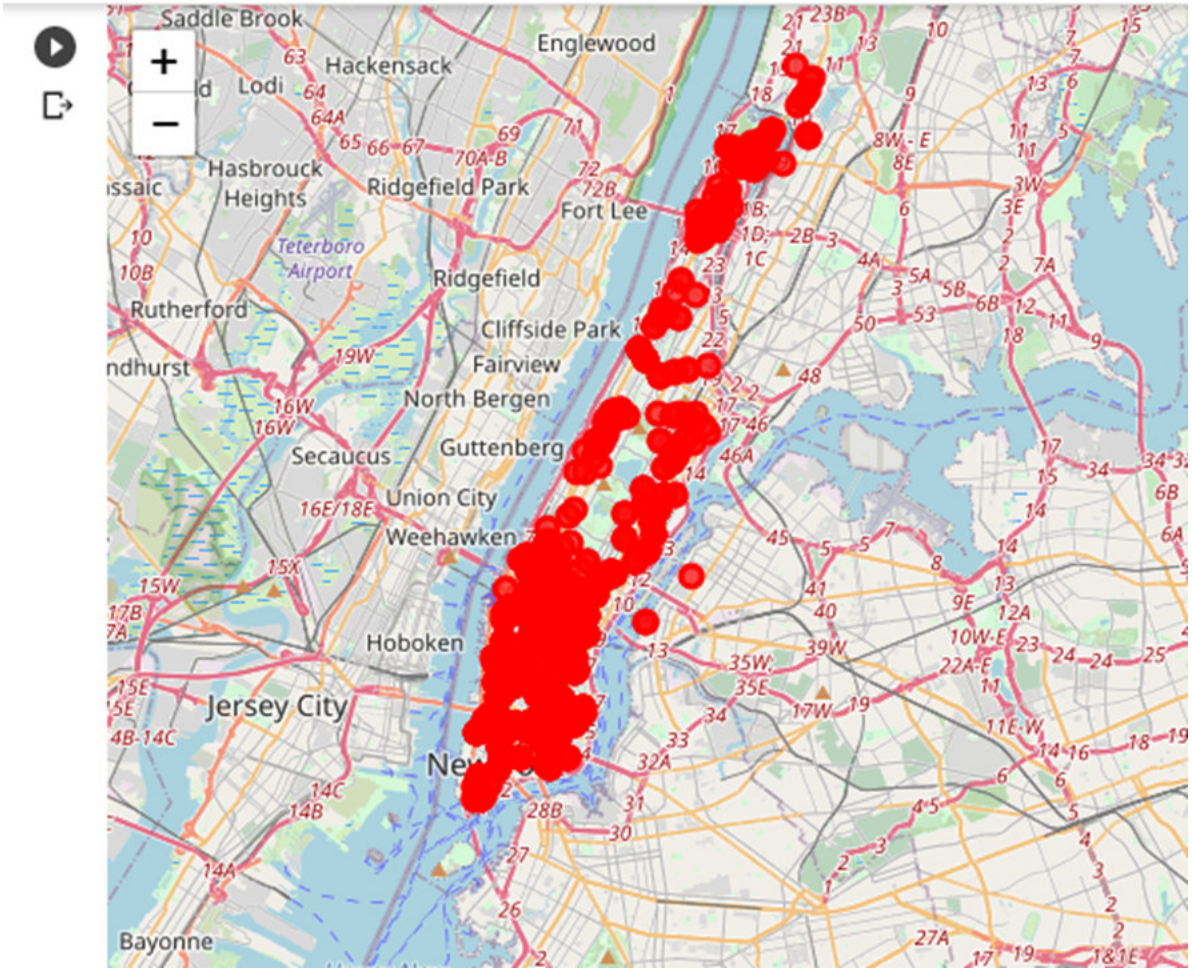
```
https://api.foursquare.com/v2/venues/search?&client_id=QSQC42XYF50MXN2PNVIVQ5ASUGUY5
{'meta': {'code': 200, 'requestId': '6004a6ecba07166a6ac3254e'}, 'response': {'venue
[]
Empty DataFrame
Columns: []
Index: []
```

Results of searching Colombian food

Results of searching Latino food restaurants

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Hola España	40.879160	-73.904553	Latin American Restaurant
1	Marble Hill	40.876551	-73.91066	La Caridad	40.869219	-73.903219	Latin American Restaurant
2	Marble Hill	40.876551	-73.91066	Leche y Miel	40.883742	-73.901857	Latin American Restaurant
3	Marble Hill	40.876551	-73.91066	La Tia	40.878808	-73.905187	Latin American Restaurant
4	Marble Hill	40.876551	-73.91066	Empanadas Monumental	40.883476	-73.901964	Empanada Restaurant

Generating competitors map



High amount of Latino food restaurants

Grouping the information by neighborhoods

East Village	40	40	40	40	40	40
Financial District	8	8	8	8	8	8
Flatiron	32	32	32	32	32	32
Gramercy	25	25	25	25	25	25
Greenwich Village	35	35	35	35	35	35
Hamilton Heights	11	11	11	11	11	11
Hudson Yards	34	34	34	34	34	34
Inwood	36	36	36	36	36	36
Lenox Hill	12	12	12	12	12	12
Lincoln Square	20	20	20	20	20	20
Little Italy	41	41	41	41	41	41
Lower East Side	39	39	39	39	39	39
Manhattan Valley	16	16	16	16	16	16
Manhattanville	10	10	10	10	10	10
Marble Hill	14	14	14	14	14	14

What has been done now is to generate a new table that groups the information by neighborhoods

Analyzing by neighborhood.

	Neighborhood	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Bakery	Bar	Brazilian Restaurant	Café	Caribbean Restaurant	Churrascaria	Cuban Restaurant	Deli / Bodega	Emp Resta
0	Battery Park City	0.111111	0.111111	0.000000	0.000000	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.000000	0.1
1	Carnegie Hill	0.090909	0.090909	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
2	Central Harlem	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.1
3	Chelsea	0.086957	0.130435	0.000000	0.043478	0.000000	0.043478	0.000000	0.043478	0.000000	0.043478	0.000000	0.0
4	Chinatown	0.076923	0.102564	0.000000	0.000000	0.000000	0.051282	0.000000	0.000000	0.000000	0.000000	0.025641	0.0
5	Civic Center	0.222222	0.166667	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
6	Clinton	0.042553	0.085106	0.000000	0.000000	0.000000	0.170213	0.000000	0.000000	0.000000	0.000000	0.000000	0.1
7	East Harlem	0.058824	0.058824	0.000000	0.000000	0.058824	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.0
8	East Village	0.050000	0.075000	0.000000	0.000000	0.000000	0.075000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
9	Financial District	0.000000	0.125000	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	0.000000	0.1
10	Flatiron	0.125000	0.093750	0.000000	0.031250	0.000000	0.062500	0.000000	0.031250	0.000000	0.031250	0.031250	0.0

What has been done now is to generate tables with the categories associated to each neighborhood for applying a ranking

Raking the venues

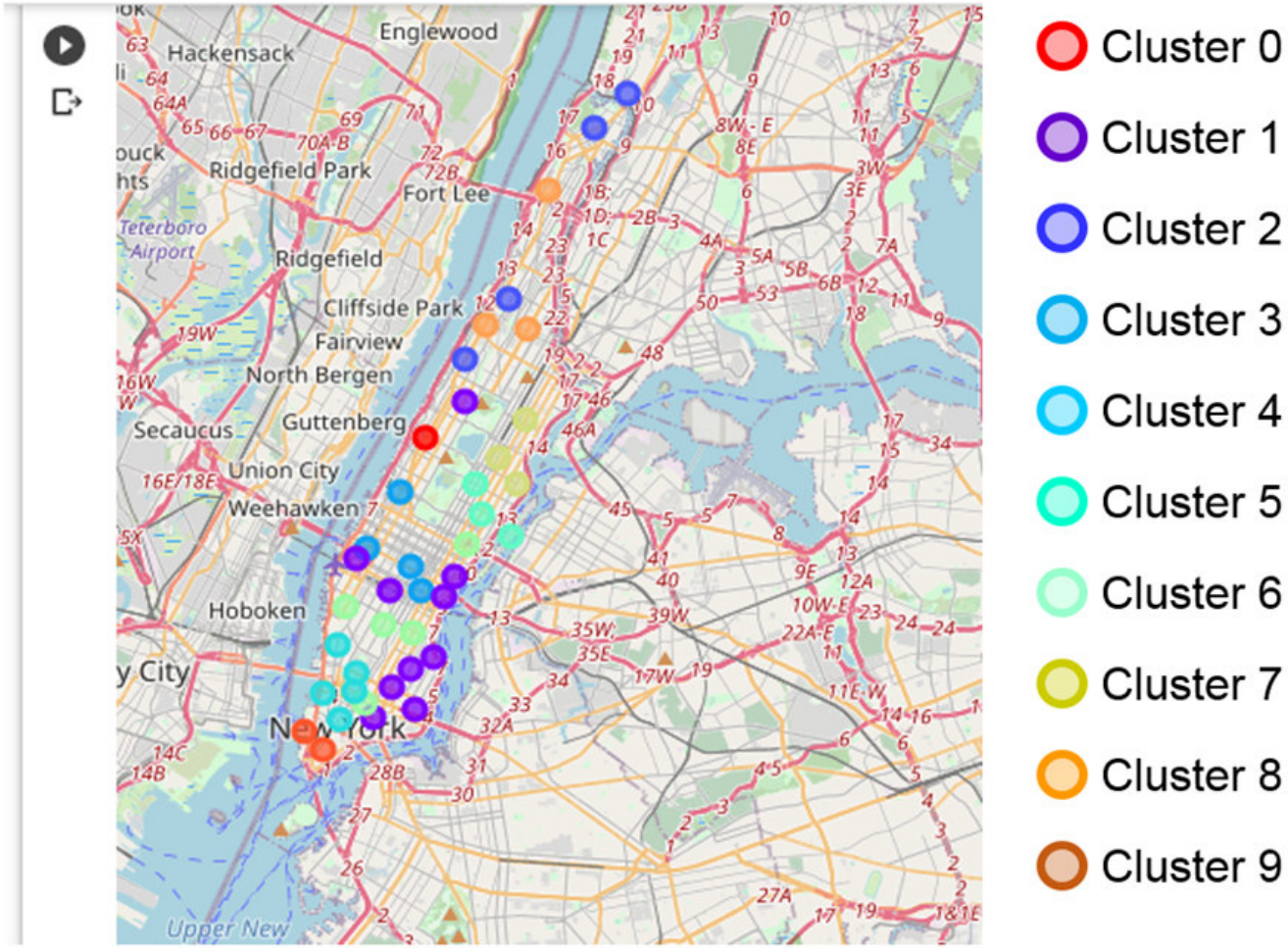
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Latin American Restaurant	Arepa Restaurant	Brazilian Restaurant	Empanada Restaurant	Middle Eastern Restaurant	Peruvian Restaurant	South American Restaurant	Argentinian Restaurant	Bar	Café
1	Carnegie Hill	Latin American Restaurant	Peruvian Restaurant	Arepa Restaurant	Argentinian Restaurant	Food Truck	Empanada Restaurant	Deli / Bodega	Cuban Restaurant	Churrascaria	Caribbean Restaurant
2	Central Harlem	Latin American Restaurant	Food Truck	Spanish Restaurant	Empanada Restaurant	Wine Bar	Caribbean Restaurant	Deli / Bodega	Cuban Restaurant	Churrascaria	Café
3	Chelsea	Latin American Restaurant	Peruvian Restaurant	Argentinian Restaurant	Arepa Restaurant	Bakery	Brazilian Restaurant	French Restaurant	Caribbean Restaurant	Cuban Restaurant	Mexican Restaurant
4	Chinatown	Latin American Restaurant	Peruvian Restaurant	Argentinian Restaurant	Arepa Restaurant	South American Restaurant	Brazilian Restaurant	Mexican Restaurant	Tapas Restaurant	Venezuelan Restaurant	Deli / Bodega

Raking the best venues by neighborhood

Clustering the neighborhoods

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Latin American Restaurant	Tapas Restaurant	Lounge	Empanada Restaurant	Wine Bar	Churrascaria	Food Truck	Deli / Bodega
1	Manhattan	Chinatown	40.715618	-73.994279	1	Latin American Restaurant	Peruvian Restaurant	Argentinian Restaurant	Arepa Restaurant	South American Restaurant	Brazilian Restaurant	Mexican Restaurant	Tapas Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	8	Latin American Restaurant	Empanada Restaurant	South American Restaurant	Seafood Restaurant	Caribbean Restaurant	Arepa Restaurant	Italian Restaurant	Spanish Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	2	Latin American Restaurant	Venezuelan Restaurant	Spanish Restaurant	Lounge	Empanada Restaurant	Wine Bar	Churrascaria	Deli / Bodega
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Latin American Restaurant	Empanada Restaurant	Wine Bar	Churrascaria	French Restaurant	Food Truck	Deli / Bodega	Cuban Restaurant

Clustering the neighborhoods



Clustering the neighborhoods

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Upper West Side	South American Restaurant	Peruvian Restaurant	Latin American Restaurant	Brazilian Restaurant	Spanish Restaurant	Caribbean Restaurant	Empanada Restaurant	Deli / Bodega	Cuban Restaurant	Churrascaria

Neighborhoods and restaurants in cluster 0

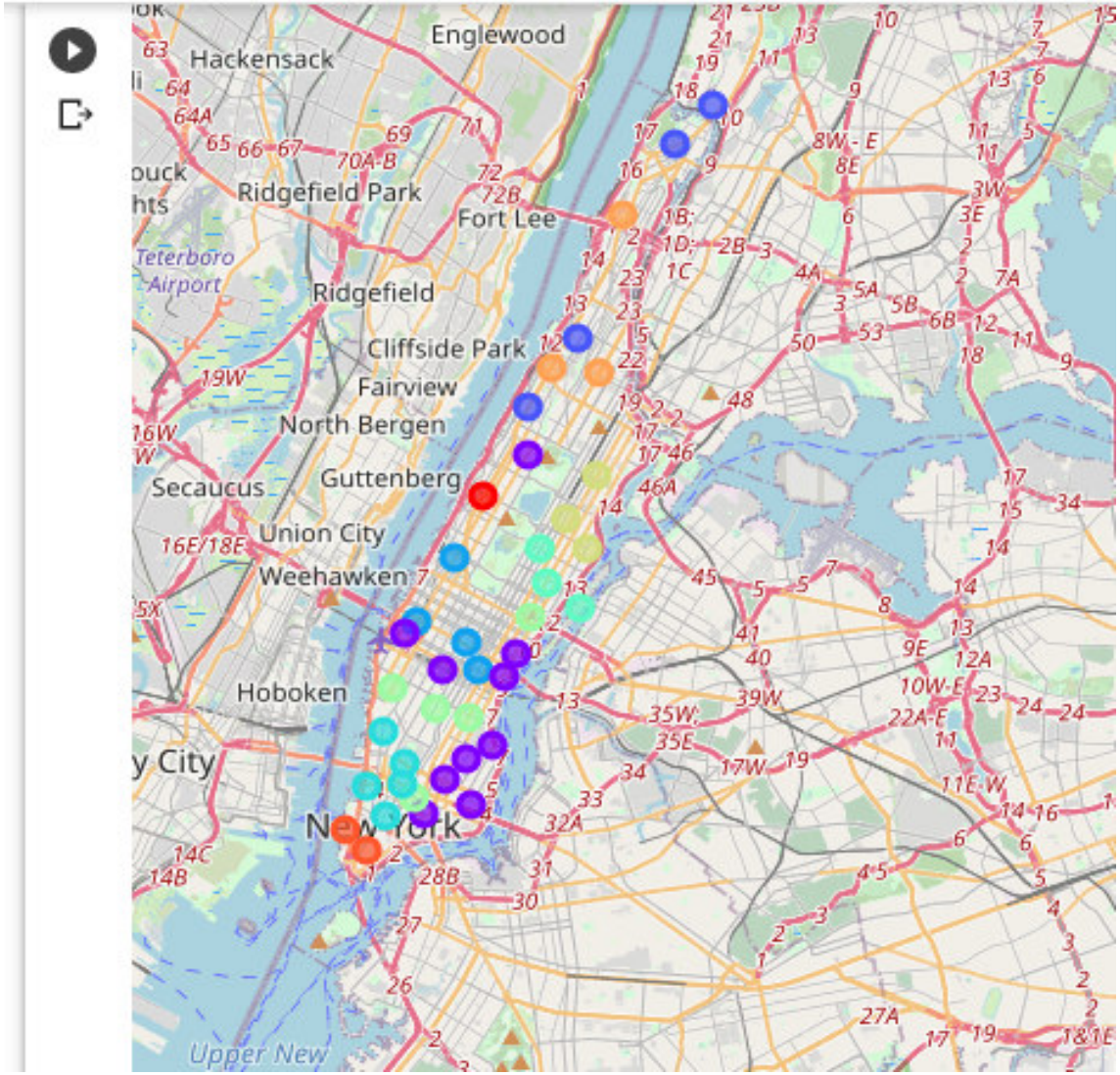
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
28	Battery Park City	Latin American Restaurant	Arepa Restaurant	Brazilian Restaurant	Empanada Restaurant	Middle Eastern Restaurant	Peruvian Restaurant	South American Restaurant	Argentinian Restaurant	Bar	Café
29	Financial District	Latin American Restaurant	Peruvian Restaurant	Argentinian Restaurant	Empanada Restaurant	South American Restaurant	Brazilian Restaurant	Middle Eastern Restaurant	Churrascaria	Deli / Bodega	Cuban Restaurant

Neighborhoods and restaurants in cluster 9

BUSINESS PROBLEM

Colombia is a country that within Latin America enjoys having a unique cultural richness, which is reflected in its gastronomy. The ethnic diversity of Colombia manifests itself in Colombian gastronomy. In Colombian food there's a rich mix of Spanish and Indigenous ingredients with African, Arab and Spanish preparation.

The main problem to be solved is to establish the potential of a Colombian food restaurant in Manhattan. We want the restaurant to be unique, and stand out strongly with the neighboring restaurants through the diversity in the menu. However, New York is known to have a particular Latino influence that is seen not only in the arts but also in food.



Cluster 0

Cluster 1

Cluster 2

Cluster 3

Cluster 4

Cluster 5

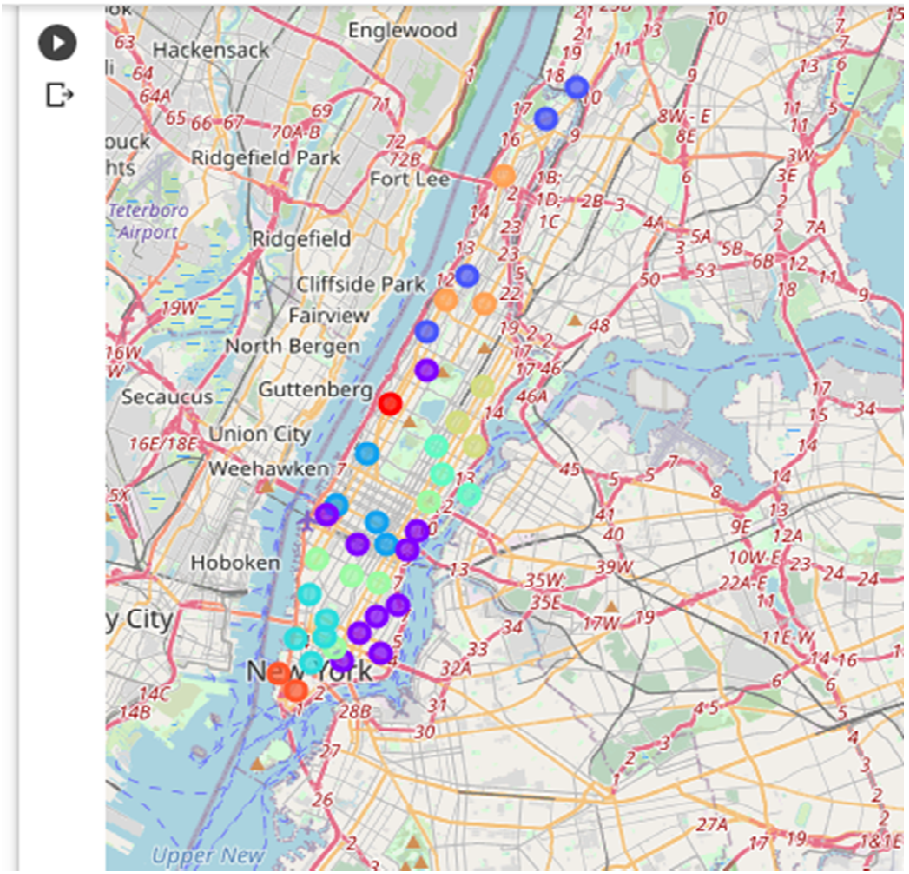
Cluster 6

Cluster 7

Cluster 8

Cluster 9

DISCUSION



- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
- Cluster 7
- Cluster 8
- Cluster 9

MENU

- Ajiaco
- Hormiga culona
- Lechona
- Bandeja Paisa
- Sancochos
- Tamales
- Patacón with everything
- Arepas and ~~Empanadas~~

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Upper West Side	South American Restaurant	Peruvian Restaurant	Latin American Restaurant	Brazilian Restaurant	Spanish Restaurant	Caribbean Restaurant	Empanada Restaurant	Deli / Bodega	Cuban Restaurant	Churrascaria

DISCUSION

When analyzing the clusters, what can be observed is that the area with the greatest potential for development chosen is the one with the least amount of neighborhoods. In this case, we have chosen cluster 0. Although there is only one neighborhood, this is positive given that the level of competition is lower. In this way, Colombian food can be introduced on a smaller scale but benefiting from voice to voice. Within cluster 0, what stands out the most is that in the top 10 of the restaurants there is a Latin American restaurant and a Caribbean restaurant, which could include Colombian food in their menu, but could not go into many details given the wide range which may be the gastronomy of these areas. It is also striking that there is an Empanadas restaurant, which suggests direct competition with one of the products included in the menu, which should be removed from the food menu.

CONCLUSIONS

- For the final IBM Capstone project it was chosen to work a small part of the marketing study to open a Colombian food restaurant in Manhattan.

- Within the implemented program, the unsupervised learning algorithm *Kmeans* was used to perform clustering support. Folio was also used to develop the maps and information from the FOURSQUARE database to generate categories and rank data.

- As a result of the study, it was possible to identify an area of Manhattan (Upper West Side) as a potential to open this restaurant.

- In the identified area, 2 indirect competitors were detected (one for Latin American food and another for Caribbean food) which could include Colombian food in their menus, but it is based on the fact that because the offer is very wide, very possibly they have discarded many of the dishes that they intend to offer.

- In the identified area, 1 direct competitor was detected (empanada restaurant) which is specialized in a single product that is also found in the menu that is being proposed. This suggests that such a product should exit the menu, at least initially.