COURSERA CAPSTONE FINAL PROJECT

THE BATTLE OF NEIGHBORHOODS IN MADRID

INTRODUCTION

This project is a guide for any investor who wants to open a food venue in Madrid and needs to do a business plan to decide, in the first place, which kind of food venue should open, and in the second place, which would be the best borough/neighborhood to open it.

Through this project we will remark some important factors that the investors should keep in mind. It´ll be done by exploring the different boroughs in Madrid, with some important and general characteristics as the population or the average income per borough, and some specific characteristics as the most common food venues in each borough.

This project doesn´t pretend to develop a magic algorithm with which the investor knows what place he should open and where, because the are many other variables that should be analyze to make this decision.

This project gives a general overview of the socialeconomic situation of the boroughs in Madrid, and the distribution of the categories of food venues in each borough. This will be the first step and the base to develop a stronger business plan.

DATA

-House pricing dataset (precio_vivienda.csv): It was obtained from the national statistic institute in Spain (INE). I´ll only use the column "Renta_INE", which gives the average income for each borough.

-Population distribution in Madrid dataset (Poblacion_barrios_madrid.csv): It´ll give us the total population of each borough, and the distribution by age. It was downloaded from the the official website of the Madrid town hall

-Date of the datasets: January 2019

-Getting latitude and longitude data of the boroughs via Geocoder package

-Fourthsquare API: One of the most power ful APIs to get located dat. It´ll be used to get the information about the food venues in each borough

METHODOLOGY:

EXTRACTING THE DATA:

Python packages that have been used for the project:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geocoder – To retrieve Location Data
- Requests – Library to handle http requests
- Seaborn- data visulization library based on matplotlib
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

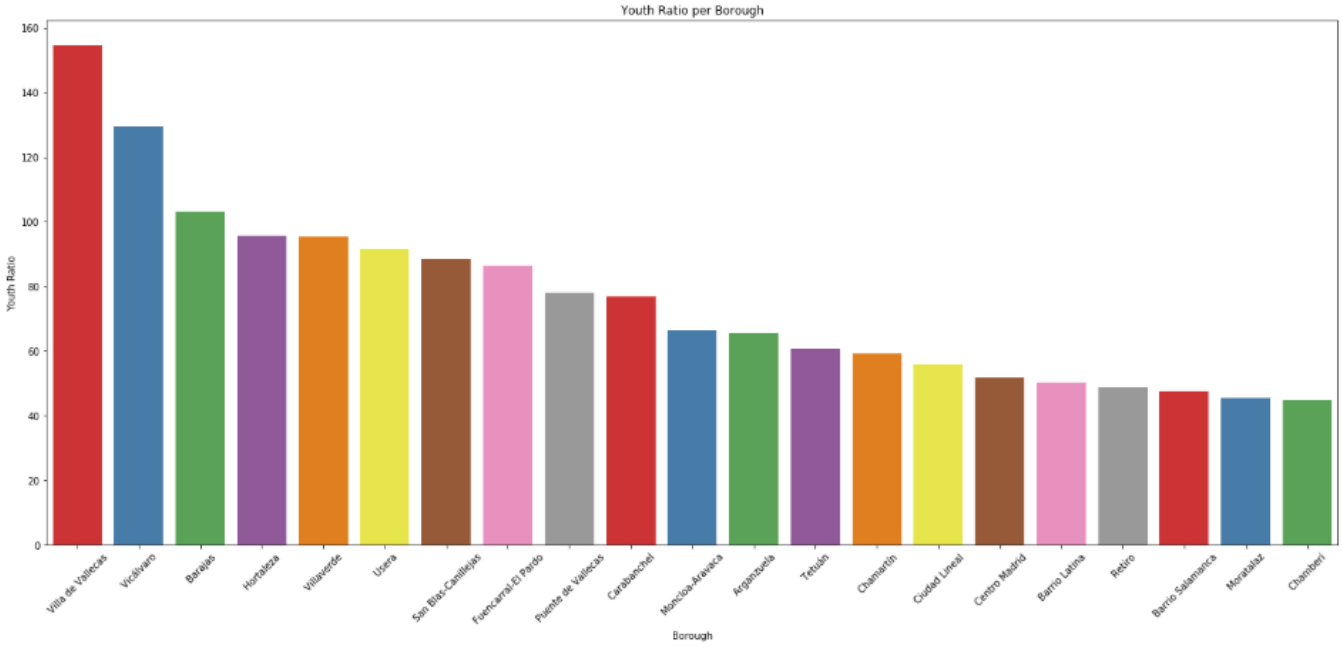
GETTING, EXPLORING AND ANALYZING THE DATA

This is the dataframe after loading, cleaning and joining the two datasets

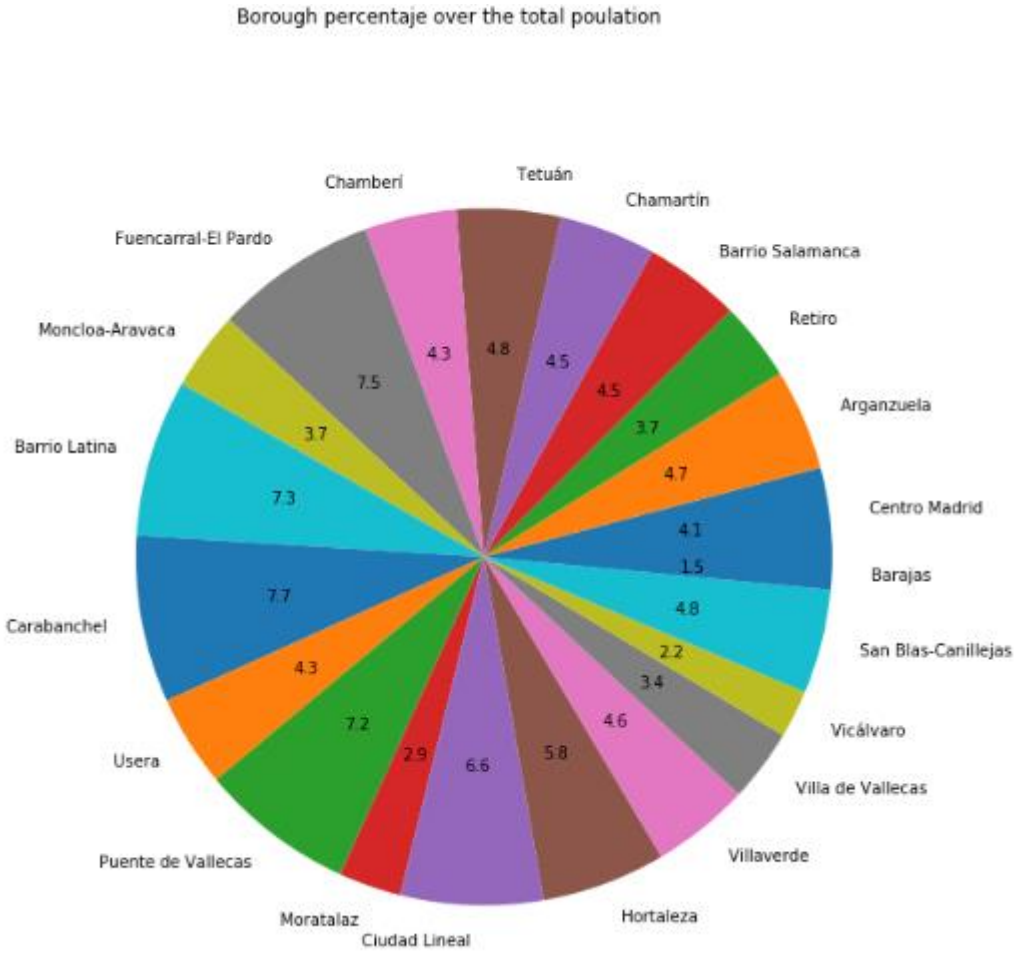
Borough	Population	0-15 years	16-64 years	More than 65 years	More than 80 years	Youth Ratio	Renta_INE
Centro Madrid	134881	8.28	75.67	16.05	5.52	51.61	16147
Arganzuela	153830	12.61	68.12	19.27	6.48	65.44	17306
Retiro	119379	12.55	61.70	25.76	8.73	48.72	21504
Barrio Salamanca	146148	11.35	64.76	23.89	8.91	47.51	24433
Chamartín	145865	13.82	62.91	23.28	8.32	59.35	25969
Tetuán	157937	11.72	68.95	19.33	7.42	60.64	14970
Chamberí	139448	10.87	64.95	24.18	8.86	44.95	22499
Fuencarral-El Pardo	246021	17.97	61.23	20.80	6.22	86.38	18573
Moncloa-Aravaca	119423	14.52	63.58	21.90	7.64	66.28	22152
Barrio Latina	238154	12.44	62.77	24.79	8.99	50.19	12232
Carabanchel	253040	14.76	65.99	19.24	7.29	76.72	10872
Usera	139501	15.63	67.29	17.08	6.93	91.50	9395
Puente de Vallecas	234770	13.94	68.20	17.86	7.02	78.01	9545
Moratalaz	94609	11.88	62.07	26.05	10.29	45.58	13944
Ciudad Lineal	216270	12.76	64.41	22.83	8.83	55.87	15408
Hortaleza	188267	17.78	63.61	18.61	5.74	95.56	18277
Villaverde	148883	16.59	66.01	17.40	6.64	95.33	9756
Villa de Vallecas	110436	19.30	68.22	12.48	3.67	154.65	11925
Vicálvaro	72126	18.68	66.86	14.45	4.87	129.28	11695
San Blas-Canillejas	158166	15.53	66.90	17.57	7.08	88.42	13404
Barajas	48972	18.33	63.91	17.77	4.41	103.17	17641

Some visualizations of this dataframe, it´s a good starting point for a firstanalysis:

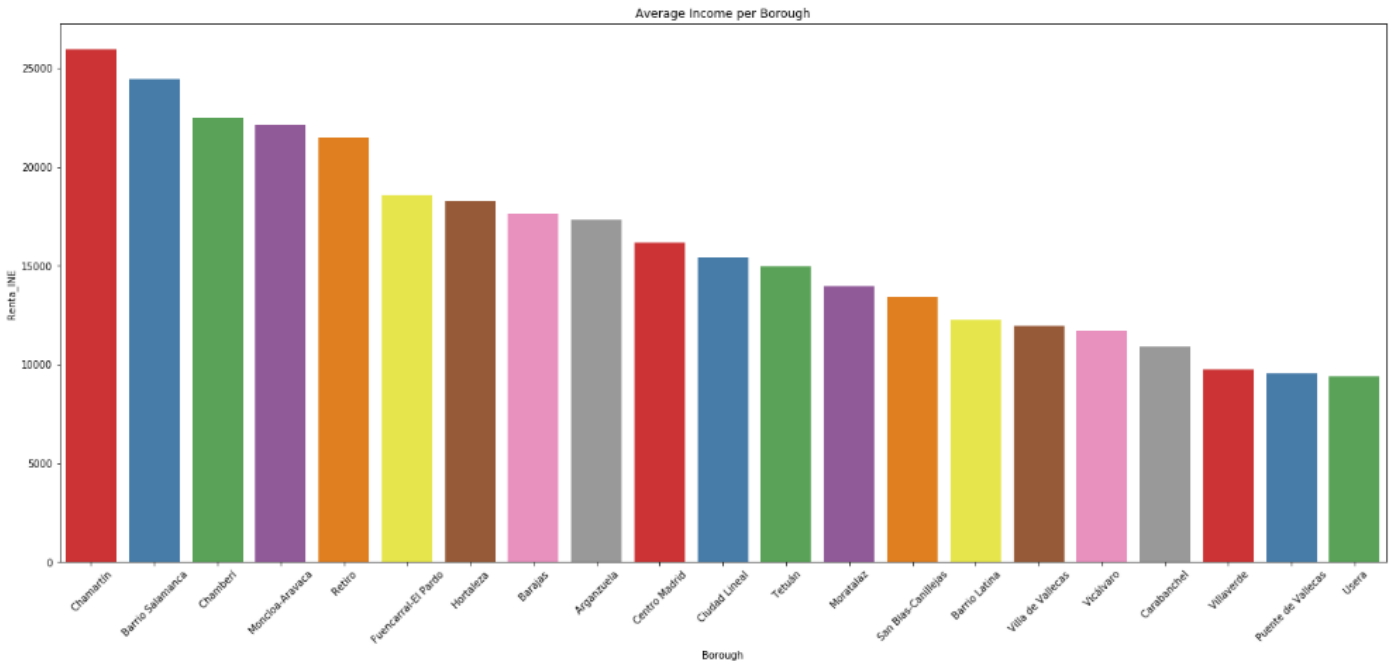
-Youth ratio per borough:



-Distribution of the population (%):



-Average income per borough



Analysis of the charts:

-Youth Ratio is calculated as (0-15 years population / More than 65 years population * 100).
This graph shows us that except for some exceptions, the boroughs with a higher Youth Ratio are those that are in the outskirts of the city
This might be a factor that the investor should keep in mind, depending on the kind of food venue he chooses. If he pretend to open a fast food venue, Youth Ratio can be an important point of decision.

-The population of each borough can be an important factor. If our investor doesn´t like the risk, he maybe should consider this graph to avoid boroughs as Moratalaz, Barajas or Vicálvaro, which have a very low percentage of the total population

-The average income per borough can also be helpful in order to orientate the decision. We can see that the five boroughs with more income are also part of the boroughs with a lower Youth Ratio, and are mostly the boroughs that are located in the center of the city

FOURTHSQUARE API

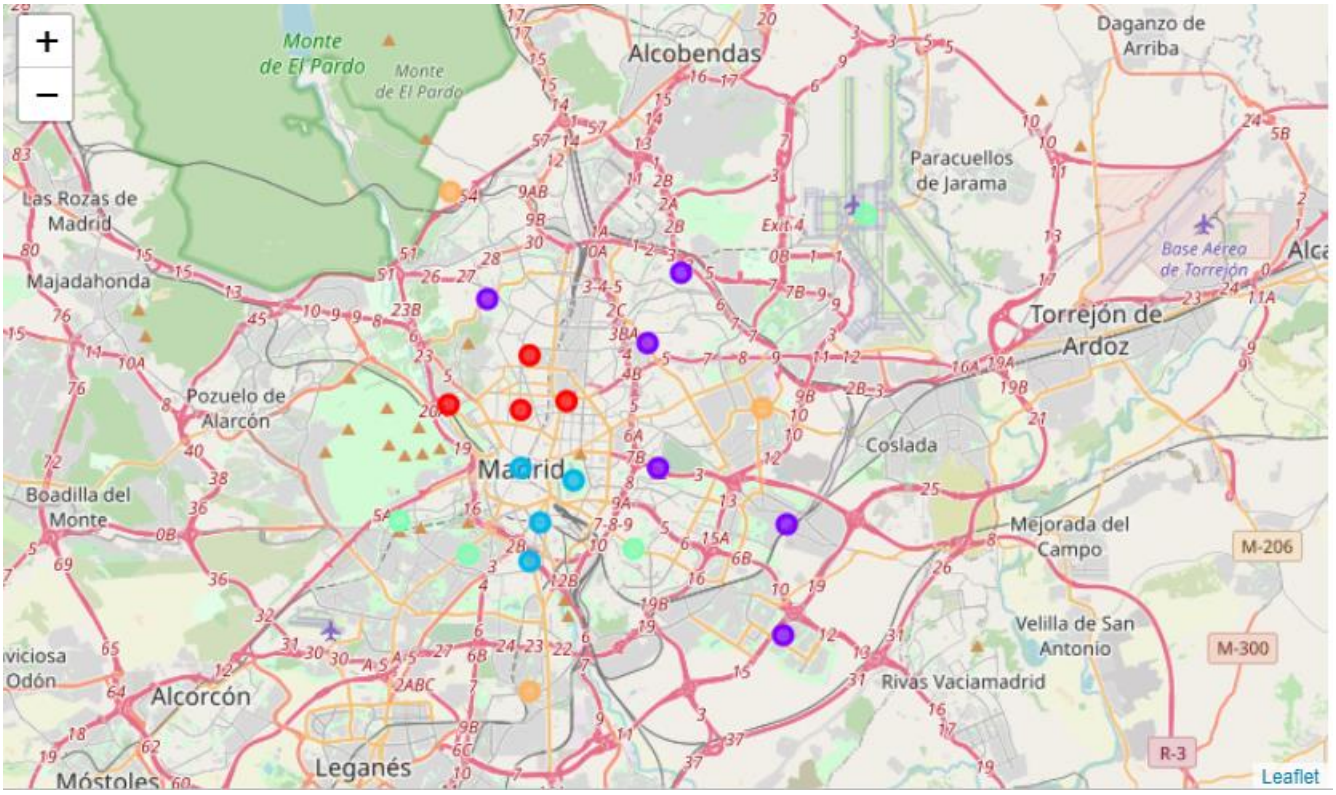
After obtaining latitude and longitude for each neighborhood with the library Geocoder, we use these coordinates to make queries to the Fourthsqueare API, in order to obtain the details of the food venues in each borough. We got a maximum amount of 150 food venues for each borough, due to the settled limit, but it will be enough to have an idea about the kind of food venue categories in each borough.

This is the top fifteen food venue categories in Madrid:

```
Coffee                  381
Spanish Restaurant      346
Bakery                  314
Restaurant              265
Tapas Bar               227
Pizza Place              88
Fast Food Restaurant     86
Diner                    77
Asian Restaurant         69
Snack Place              65
Burger Joint             62
Italian Restaurant       56
Deli / Bodega            54
Breakfast Spot           51
Ice Cream Shop           46
Name: Venue Category, dtype: int64
```

CLUSTERING:

Run K-means algorithm to cluster the boroughs in 5 groups. The variables that are taking into account for the algorithm were: population, youth ratio, average income, longitude and longitude of each borough.



ANALYSIS BY CLUSTERS:

CLUSTER 1

```
final_df[final_df['Cluster']==0]
```

	Borough	Population	Youth Ratio	Renta_INE	Latitude	Longitude	Food Venues	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
3	Barrio Salamanca	146148	47.51	24433	40.43630	-3.68560	149	Bakery	Spanish Restaurant	Coffee	Restaurant	Tapas Bar	Japanese Restaurant	Italian Restaurant	Mexican Restaurant	Market	Mediterranean Restaurant
6	Chamartín	145865	59.35	25969	40.45000	-3.70000	150	Bakery	Restaurant	Spanish Restaurant	Coffee	Tapas Bar	Italian Restaurant	Gastropub	Pizza Place	Donut Shop	American Restaurant
7	Chamberí	139448	44.95	22499	40.43404	-3.70379	150	Bakery	Coffee	Restaurant	Spanish Restaurant	Tapas Bar	Italian Restaurant	Asian Restaurant	Market	Gastropub	Fast Food Restaurant
11	Moncloa-Aravaca	119423	66.28	22152	40.43547	-3.73170	148	Coffee	Tapas Bar	Restaurant	Spanish Restaurant	Bakery	Italian Restaurant	Diner	Paella Restaurant	BBQ Joint	Gastropub

- Central boroughs
- Have the highest incomes of the city
- Have aged populations, they are between the boroughs with lower youth ratios

CLUSTER 2

```
final_df[final_df['Cluster']==1]
```

	Borough	Population	Youth Ratio	Renta_INE	Latitude	Longitude	Food Venues	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
8	Ciudad Lineal	216270	55.87	15408	40.45349	-3.65434	146	Spanish Restaurant	Restaurant	Bakery	Tapas Bar	Burger Joint	Indian Restaurant	Coffee	Fast Food Restaurant	Gastropub	Asian Restaurant
10	Hortaleza	188267	95.56	18277	40.47444	-3.64110	145	Spanish Restaurant	Bakery	Coffee	Restaurant	Pizza Place	Diner	Tapas Bar	Snack Place	American Restaurant	Gastropub
12	Moratalaz	94609	45.58	13944	40.41667	-3.65000	149	Bakery	Tapas Bar	Coffee	Spanish Restaurant	Restaurant	Seafood Restaurant	Pizza Place	Deli / Bodega	Ice Cream Shop	Mexican Restaurant
16	Tetuán	157937	60.64	14970	40.46667	-3.71667	148	Spanish Restaurant	Bakery	Restaurant	Coffee	Pizza Place	Tapas Bar	Breakfast Spot	Diner	Snack Place	American Restaurant
18	Vicálvaro	72126	129.28	11695	40.40000	-3.60000	146	Spanish Restaurant	Tapas Bar	Coffee	Bakery	Restaurant	Pizza Place	Fast Food Restaurant	Food Truck	Burger Joint	Diner
19	Villa de Vallecas	110436	154.65	11925	40.36695	-3.60146	141	Coffee	Fast Food Restaurant	Pizza Place	Restaurant	Tapas Bar	Burger Joint	Food Truck	Cafeteria	American Restaurant	Asian Restaurant

- It forms a belt around the north and east of the city
- It contains the three boroughs with the youngest population
- Only one borough is in the top ten list of the highest average income

CLUSTER 3

```
final_df[final_df['Cluster']==2]
```

	Borough	Population	Youth Ratio	Renta_INE	Latitude	Longitude	Food Venues	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	Arganzuela	153830	65.44	17306	40.40021	-3.69618	149	Bakery	Coffee	Tapas Bar	Spanish Restaurant	Restaurant	Argentinian Restaurant	Mexican Restaurant	Asian Restaurant	Ice Cream Shop	Seafood Restaurant
5	Centro Madrid	134881	51.61	16147	40.41670	-3.70356	150	Bakery	Coffee	Tapas Bar	Spanish Restaurant	Restaurant	Market	Asian Restaurant	Diner	Ice Cream Shop	Italian Restaurant
14	Retiro	119379	48.72	21504	40.41317	-3.68307	149	Bakery	Spanish Restaurant	Coffee	Restaurant	Tapas Bar	Asian Restaurant	Italian Restaurant	Ice Cream Shop	Dessert Shop	Seafood Restaurant
17	Usera	139501	91.50	9395	40.38866	-3.70035	146	Coffee	Spanish Restaurant	Bakery	Restaurant	Tapas Bar	Chinese Restaurant	Argentinian Restaurant	Asian Restaurant	Pizza Place	Mexican Restaurant

- Boroughs located in the center and south of the city
- Except Usera, the rest of the boroughs have aged populations and medium-high incomes
- Usera is the borough with the lowest average income ratio

CLUSTER 4

```
final_df[final_df['Cluster']==3]
```

	Borough	Population	Youth Ratio	Renta_INE	Latitude	Longitude	Food Venues	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
1	Barajas	48972	103.17	17641	40.491810	-3.569480	127	Coffee	Spanish Restaurant	Fast Food Restaurant	Snack Place	Restaurant	Breakfast Spot	Tapas Bar	Deli / Bodega	Diner	Sandwich Place
2	Barrio Latina	238154	50.19	12232	40.401064	-3.751336	147	Coffee	Spanish Restaurant	Bakery	Tapas Bar	Fast Food Restaurant	Restaurant	Chinese Restaurant	Snack Place	Seafood Restaurant	Diner
4	Carabanchel	253040	76.72	10872	40.390940	-3.724200	148	Coffee	Spanish Restaurant	Tapas Bar	Restaurant	Bakery	Chinese Restaurant	Mediterranean Restaurant	Burger Joint	Diner	Peruvian Restaurant
13	Puente de Vallecas	234770	78.01	9545	40.392350	-3.659700	145	Coffee	Bakery	Spanish Restaurant	Restaurant	Tapas Bar	Fast Food Restaurant	Pizza Place	Diner	Japanese Restaurant	Argentinian Restaurant

- 4 boroughs in the outskirts of the city, located in the south except Barajas
- Three of them (Carabanchel, Puente de Vallecas y Barrio Latina) are three of the boroughs with more population, but also with less average incomes

CLUSTER 5

```
final_df[final_df['Cluster']==4]
```

	Borough	Population	Youth Ratio	Renta_INE	Latitude	Longitude	Food Venues	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
9	Fuencarral-El Pardo	246021	86.38	18573	40.49840	-3.7314	110	Bakery	Restaurant	Coffee	Spanish Restaurant	Burger Joint	Breakfast Spot	Diner	Pizza Place	Cafeteria	Mediterranean Restaurant
15	San Blas-Canillejas	158166	88.42	13404	40.43451	-3.6099	147	Coffee	Tapas Bar	Restaurant	Spanish Restaurant	Fast Food Restaurant	Deli / Bodega	Asian Restaurant	Italian Restaurant	Sandwich Place	Pizza Place
20	Villaverde	148883	95.33	9756	40.35000	-3.7000	145	Coffee	Spanish Restaurant	Restaurant	Bakery	Tapas Bar	Fast Food Restaurant	Diner	Deli / Bodega	Snack Place	Falafel Restaurant

- It contains 3 boroughs that are far away from the city center
- The three of them are in the top ten youth ratio list
- Only Fuencarral-El Pardo, which is the only borough of the cluster located in the north, as a medium-high average income

FINAL CONCLUSION:

As I advanced in the introduction, this project was not think to provide a close conclusion or a best solution about which is the “magic place” to open a food venue.

I have analyzed the social and economic situation of the neighbors and the categorization of the food venues I each of those. Every investor should adapt this analysis to his own interest and possibilities.

For example, let´s imagine I am a regular investor, with no excessive money, and I don´t want to take any risk because I´m investing an important part my savings in a venue food. Based on the analysis done, maybe I would focus in the cluster 4. I could open classic food venue(coffee or a typical Spanish Restaurant), in Latina, Carabanchel, o Puente de Vallecas.
These boroughs are the most populated of Madrid, the rent probably will be cheaper because it´s in the outskirts of the city, and these food venues that I just named have no risk because they are the most common food venues there, probably because they have aged populations, which are more likely to go to a classic food venue than a fancy new restaurant.

Of course, this example is a superficial analysis and the investor will have to analyze another variables , but this is a good starting point to delimit the options, and doing a deeper analysis from there.