OPTIMAL LOCATION FOR A COLOMBIAN RESTAURANT IN MADRID

ANIBAL JOSÉ CEBALLOS HURTADO

The optimal location of a Colombian restaurant in Madrid needs to be assessed. To make it possible, some processes have to be accomplished

Extracting the main data from the Madrid's City Town Hall and the Foursquare API app

Get the data

Work the data

Making a segmentation by district and population main features in Madrid (Clustering).

Analyzing the results and extracting conclusions based on them.

Extract insights from the data

The data used contained information about the nationality and number of citizens living in each district

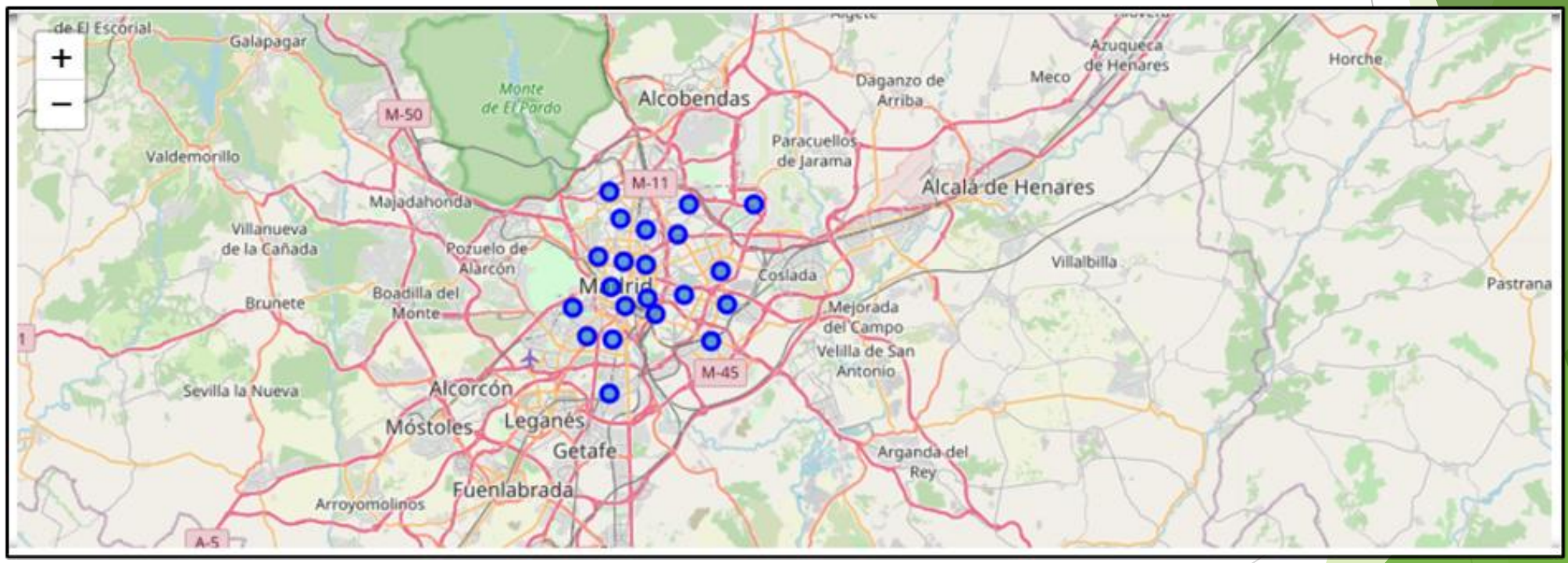
	Country of Origin	Total Ciudad de Madrid	Centro	Arganzuela	Retiro	Salamanca	Chamartín	Tetuán	Chamberí	Fuencarral-El Pardo	...	Usera	Puente de Vallecas	Moratalaz	Ciudad Lineal
0	Rumanía	45036.0	815.0	754.0	480.0	753.0	680.0	1466.0	597.0	1830.0	...	2241.0	4784.0	1286.0	2888.0
1	China	37276.0	1506.0	1356.0	564.0	755.0	652.0	1988.0	816.0	1733.0	...	9207.0	3602.0	564.0	1960.0
2	Ecuador	23953.0	647.0	741.0	265.0	619.0	380.0	1395.0	453.0	632.0	...	1806.0	3290.0	491.0	2471.0
3	Venezuela	23359.0	1563.0	913.0	638.0	1564.0	933.0	1310.0	794.0	1428.0	...	875.0	1829.0	480.0	1858.0
4	Colombia	22616.0	998.0	717.0	483.0	803.0	551.0	822.0	659.0	999.0	...	1752.0	1733.0	482.0	1792.0
5	Marruecos	21909.0	1101.0	390.0	184.0	322.0	280.0	1393.0	320.0	930.0	...	942.0	3437.0	258.0	1011.0
6	Italia	20306.0	3030.0	1219.0	840.0	1817.0	1060.0	1194.0	1640.0	1195.0	...	412.0	704.0	310.0	1258.0
7	Perú	18829.0	563.0	521.0	253.0	612.0	419.0	965.0	567.0	805.0	...	1131.0	2079.0	668.0	1726.0
8	Paraguay	18682.0	364.0	474.0	237.0	521.0	657.0	3311.0	584.0	1024.0	...	727.0	1354.0	360.0	1619.0
9	República Dominicana	17511.0	365.0	654.0	204.0	344.0	322.0	2272.0	443.0	589.0	...	1202.0	1989.0	223.0	1581.0
10	Honduras	15981.0	149.0	228.0	232.0	332.0	337.0	755.0	317.0	863.0	...	1115.0	2483.0	281.0	1062.0
11	Bolivia	14930.0	284.0	407.0	182.0	342.0	315.0	576.0	280.0	401.0	...	2827.0	1573.0	227.0	1086.0
12	Filipinas	12628.0	1344.0	640.0	142.0	578.0	661.0	4473.0	771.0	442.0	...	225.0	418.0	76.0	590.0
13	Portugal	9860.0	769.0	372.0	262.0	695.0	534.0	590.0	509.0	693.0	...	283.0	597.0	173.0	612.0
14	Francia	9561.0	1606.0	455.0	370.0	968.0	554.0	387.0	699.0	366.0	...	66.0	138.0	85.0	599.0
15	Ucrania	9453.0	152.0	214.0	133.0	220.0	176.0	221.0	149.0	312.0	...	428.0	912.0	239.0	348.0
16	Brasil	9324.0	677.0	309.0	244.0	431.0	280.0	567.0	322.0	361.0	...	410.0	604.0	201.0	421.0
17	Bulgaria	7842.0	262.0	137.0	115.0	113.0	123.0	245.0	74.0	316.0	...	396.0	760.0	285.0	780.0
18	Estados Unidos de América	6791.0	1637.0	406.0	385.0	749.0	389.0	300.0	657.0	297.0	...	86.0	139.0	232.0	252.0
19	Reino Unido	5915.0	1274.0	324.0	256.0	550.0	466.0	329.0	501.0	313.0	...	81.0	142.0	59.0	285.0
20	Bangladesh	5886.0	2742.0	381.0	33.0	32.0	21.0	210.0	48.0	27.0	...	175.0	649.0	23.0	179.0
21	Cuba	5725.0	435.0	243.0	111.0	169.0	161.0	257.0	175.0	305.0	...	301.0	513.0	118.0	483.0
22	Polonia	5487.0	246.0	173.0	117.0	138.0	130.0	220.0	151.0	184.0	...	298.0	424.0	88.0	266.0
23	Argentina	5061.0	588.0	262.0	186.0	323.0	206.0	271.0	291.0	261.0	...	230.0	319.0	74.0	317.0
24	Alemania	4707.0	592.0	199.0	175.0	494.0	442.0	245.0	367.0	607.0	...	60.0	93.0	39.0	240.0
25	Nicaragua	4697.0	82.0	107.0	101.0	174.0	217.0	248.0	111.0	215.0	...	291.0	601.0	212.0	523.0
26	México	4503.0	644.0	213.0	173.0	656.0	263.0	266.0	359.0	263.0	...	89.0	133.0	38.0	217.0
27	El Salvador	3228.0	71.0	58.0	82.0	82.0	91.0	135.0	81.0	115.0	...	205.0	575.0	102.0	264.0
28	Rusia	2522.0	178.0	117.0	89.0	215.0	138.0	132.0	113.0	127.0	...	92.0	166.0	50.0	141.0
29	Chile	2520.0	285.0	124.0	84.0	212.0	76.0	122.0	161.0	161.0	...	69.0	172.0	80.0	187.0
30	Senegal	2106.0	507.0	235.0	24.0	31.0	18.0	63.0	18.0	37.0	...	62.0	270.0	11.0	52.0
31	India	1704.0	138.0	67.0	40.0	126.0	142.0	136.0	72.0	71.0	...	74.0	93.0	14.0	142.0
32	Países Bajos	1615.0	218.0	77.0	90.0	134.0	103.0	78.0	127.0	68.0	...	48.0	74.0	21.0	99.0
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
34	Resto de países (1)	35113.0	3278.0	1287.0	882.0	2203.0	1497.0	1781.0	1606.0	1991.0	...	1285.0	2813.0	500.0	2009.0

To access the Foursquare API data, the raw data had to be converted into a fashion easily used by the Foursquare API

A second data frame was created where the districts names were included. Also, the latitude and longitude values of each district were added

	District	Latitude	Longitude
0	Centro	40.415347	-3.707371
1	Arganzuela	40.402733	-3.695403
2	Retiro	40.408072	-3.676729
3	Salamanca	40.43	-3.677778
4	Chamartin	40.453333	-3.6775
5	Tetuán	40.460556	-3.7
6	Chamberí	40.432792	-3.697186
7	Fuencarral-El Pardo	40.478611	-3.709722
8	Moncloa-Aravaca	40.435151	-3.718765
9	Latina	40.402461	-3.741294
10	Carabanchel	40.383669	-3.727989
11	Usera	40.381336	-3.706856
12	Puente de Vallecas	40.398204	-3.669059
13	Moratalaz	40.409869	-3.644436
14	Ciudad Lineal	40.45	-3.65
15	Hortaleza	40.469457	-3.640482
16	Villaverde	40.345925	-3.709356
17	Villa de Vallecas	40.3796	-3.62135
18	Vicálvaro	40.4042	-3.60806
19	San Blas-Canillejas	40.426001	-3.612764
20	Barajas	40.470196	-3.58489

With the data processed, it was possible to create a map of the city with the districts included as blue dots



The nearby venues by districts were extracted along with their frequencies of occurrence

V E N U E S GROUPED BY DISTRICTS

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro	40.415347	-3.707371	Plaza Mayor	40.415527	-3.707506	Plaza
1	Centro	40.415347	-3.707371	The Hat Madrid	40.414343	-3.707120	Hotel
2	Centro	40.415347	-3.707371	La Taberna de Mister Pinkleton	40.414536	-3.708108	Other Nightlife
3	Centro	40.415347	-3.707371	Mercado de San Miguel	40.415443	-3.708943	Market
4	Centro	40.415347	-3.707371	Plaza Santa Cruz	40.415063	-3.705661	Plaza

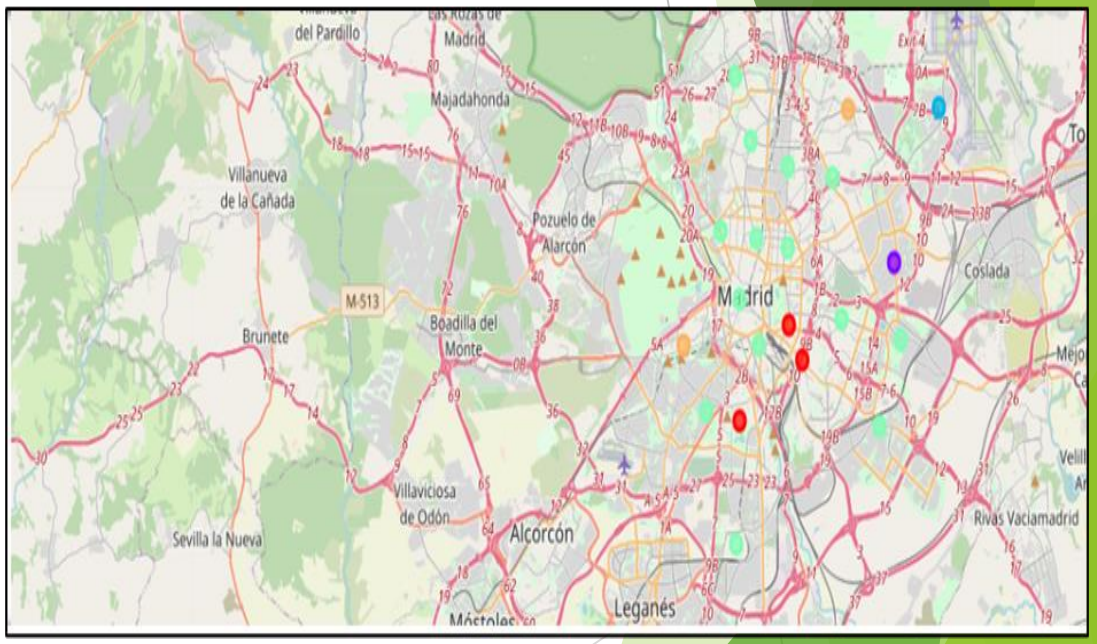
VENUES FREQUENCIES BY OCURRENCE

```
----Carabanchel----
                venue  freq
0               Plaza   0.1
1              Bakery   0.1
2  Fast Food Restaurant   0.1
3           Nightclub   0.1
4        Soccer Field   0.1
```

A dataframe containing the most common venues by district was created

	District	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	Restaurant	Spanish Restaurant	Bakery	Grocery Store	Tapas Restaurant	Gym / Fitness Center	Falafel Restaurant	Burger Joint	Hotel	Plaza
1	Barajas	Hotel	Restaurant	Spanish Restaurant	Coffee Shop	Tapas Restaurant	Supermarket	Mexican Restaurant	Brewery	Breakfast Spot	Plaza
2	Carabanchel	Metro Station	Soccer Field	Burger Joint	Mobile Phone Shop	Pizza Place	Plaza	Nightclub	Bakery	Fast Food Restaurant	Tapas Restaurant
3	Centro	Plaza	Tapas Restaurant	Spanish Restaurant	Hostel	Coffee Shop	Bistro	Ice Cream Shop	Café	Cocktail Bar	Mexican Restaurant
4	Chamartin	Spanish Restaurant	Restaurant	Grocery Store	Tapas Restaurant	Bakery	Coffee Shop	Japanese Restaurant	Gastropub	Park	Pizza Place
5	Chamberí	Spanish Restaurant	Restaurant	Bar	Café	Brewery	Japanese Restaurant	Tapas Restaurant	Mexican Restaurant	Plaza	Italian Restaurant
6	Ciudad Lineal	Spanish Restaurant	Gastropub	Supermarket	Restaurant	Burger Joint	Argentinian Restaurant	Pizza Place	Cocktail Bar	Café	Gym / Fitness Center

After gathering these information, clusters were made and could be graphed.



This is a sample of what final clusters looked like

District	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	9th Most Common Venue	10th Most Common Venue
Centro	3	Plaza	Tapas Restaurant	Spanish Restaurant	Hostel	Coffee Shop	Bistro	Ice Cream Shop	Café	Cocktail Bar	Mexican Restaurant
Arganzuela	3	Restaurant	Spanish Restaurant	Bakery	Grocery Store	Tapas Restaurant	Gym / Fitness Center	Falafel Restaurant	Burger Joint	Hotel	Plaza
Retiro	3	Spanish Restaurant	Tapas Restaurant	Supermarket	Museum	Grocery Store	Gym	Mediterranean Restaurant	Food & Drink Shop	Pizza Place	Burger Joint
Fuencarral-El Pardo	3	Clothing Store	Fast Food Restaurant	Burger Joint	Italian Restaurant	Tapas Restaurant	Sporting Goods Shop	Pizza Place	Restaurant	Chocolate Shop	Kebab Restaurant
Hortaleza	3	Breakfast Spot	Supermarket	Pizza Place	Plaza	Bar	Food	Restaurant	Chinese Restaurant	Spanish Restaurant	Pub
Chamberí	3	Spanish Restaurant	Restaurant	Bar	Café	Brewery	Japanese Restaurant	Tapas Restaurant	Mexican Restaurant	Plaza	Italian Restaurant
Ciudad Lineal	3	Spanish Restaurant	Gastropub	Supermarket	Restaurant	Burger Joint	Argentinian Restaurant	Pizza Place	Cocktail Bar	Café	Gym / Fitness Center
Moncloa-Aravaca	3	Restaurant	Spanish Restaurant	Bar	Pizza Place	Tapas Restaurant	Mediterranean Restaurant	Ice Cream Shop	Italian Restaurant	Japanese Restaurant	Pub
Salamanca	3	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Seafood Restaurant	Coffee Shop	Burger Joint	Supermarket	Tapas Restaurant	Clothing Store	Mexican Restaurant
Vicálvaro	3	Pizza Place	Spanish Restaurant	Beer Bar	Breakfast Spot	Café	Camera Store	Restaurant	Fast Food Restaurant	Tapas Restaurant	Sandwich Place
Chamartin	3	Spanish Restaurant	Restaurant	Grocery Store	Tapas Restaurant	Bakery	Coffee Shop	Japanese Restaurant	Gastropub	Park	Pizza Place
Barajas	3	Hotel	Restaurant	Spanish Restaurant	Coffee Shop	Tapas Restaurant	Supermarket	Mexican Restaurant	Brewery	Breakfast Spot	Plaza
Usera	3	Spanish Restaurant	Seafood Restaurant	Chinese Restaurant	Bubble Tea Shop	Mobile Phone Shop	Noodle House	Asian Restaurant	Café	Theater	Fast Food Restaurant
Tetuán	3	Spanish Restaurant	Grocery Store	Coffee Shop	Chinese Restaurant	Brazilian Restaurant	Supermarket	Breakfast Spot	Motorcycle Shop	Farmers Market	Clothing Store

CONCLUSIONS

According to data obtained from The Madrid City Hall's Web Portal, there is a considerable Colombian population registered in town which ranks five with 22.618 K as per survey. Furthermore, it can be seen the Colombian citizens are mostly located in the districts Caranbachel (3395 inhabitants), Ciudad Lineal (1792 inhabitants), Latina (1786 inhabitants), Usera (1752 inhabitants) and Puente de Vallecas (1733 inhabitants).

Regarding the districts outlined above, Carabanchel which belong to Cluster 1 is the one with the highest Colombian population and is also considered one of the most diverse neighborhoods in the country, with a large population of immigrants. Furthermore, if we carry out a deeper look of it about its most popular venues, it can be observed there isn't any Colombian and/or Latino restaurants; there are just Tapas and Fast-Food Restaurant. As a consequence, it could be a good opportunity to open a Colombian restaurant and, in the mid-term, some other Latino restaurants owing to the huge potential of the district.

Regarding the other districts with a considerable Colombian citizens population like Ciudad Lineal, Latina, Usera and Puente de Vallecas located in the clusters 1, 4, 5 its population are mostly Latinos, mixed with some other Europeans and Asians. Besides, having analyzed their most popular venues, there are several Fast Food, Argentinian, and South American restaurants. As a consequence, in these clusters, it can be seen the existing restaurants matches the population nationalities and food preferences.

If someone might be interested to open a new Colombian restaurant in the city or any kind of Latino restaurant, it would only be necessary to find a place where there are similar restaurants like the one considered to be opened, make a market research, and find similar clusters of population in the city that don't have them yet or have very few venues like the one to be created.

In conclusion, taking into consideration the explanations given above as well as the data, it is highly possible that the clusters 1, 4 and 5 could be a right place to open a Colombian restaurant. As explained above, the same logic could apply to open other type of restaurants or businesses in any other area of the city. It is just necessary to examine the existing businesses in the target area and study the population, then compare these 2 factors with the same ones in areas where there are existing businesses like the one to be opened, and finally verify if the matching is correct.

Finally, there is always room for improvement and hence the solution given above can be also improved for better results depending upon the data available.