

TO LOCATE THE BEST LOCATION FOR A BUSINESS IN MADRID

Victor Ortuño Crespo

DATA ORIGIN

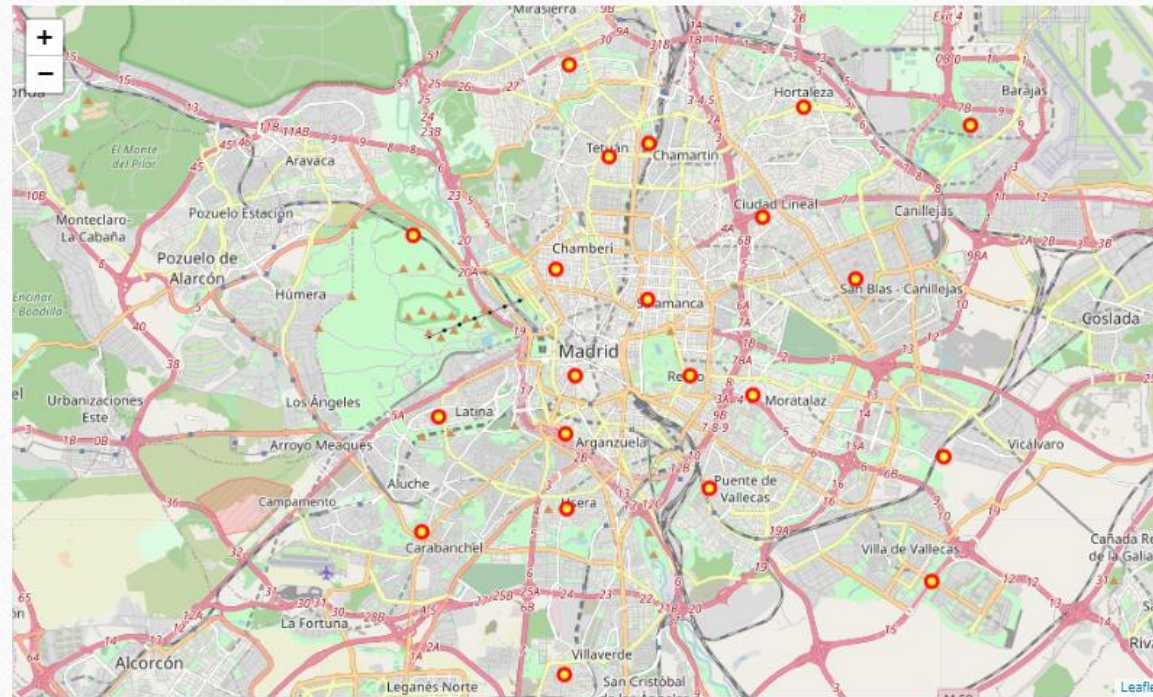
- **The official foreign population statistics offered by Madrid's City Hall website:** You will find the excel document in my GitHub 'CAPSTONE' folder. This information is critical, not only because we got the names of all the 'neighborhood' we are investigating, but because before we open our restaurant, we must familiarize ourselves with the foreign population in that region. Even though its obvious, we cannot invest in a project destined for Latin people, including typical foods, music and culture-based inside jokes, if they are found halfway across the Autonomous Region.
- **Google Maps:** Considering I was unable to find a document containing all the coordinated I needed, I had to pull them from Google maps. Every coordinate is an exact replica of the information in their database.
- **Foursquare:** This tool has been made available for all students (not exclusively). From this database, I pulled all the pertinent metadata of each venue in a 500m radius and filtered it to fit my needs. For example, venue location (Lat, Lng), type of venue, name, etc...

FOREIGN
POPULATION
DATA

	Nationality	Madrid City	Centro	Arganzuela	Retiro	Salamanca	Chamartin	Tetuan	Chamberi	Fuencarral-El Pardo	Moncloa-Aravaca	Latina	Carabanchel	Ueera
0	Rumania	45036	815	754	480	753	680	1468	597	1830	991	4904	5873	2241
1	China	37276	1508	1356	564	755	652	1988	816	1733	960	2554	4398	9207
2	Ecuador	23953	647	741	265	619	380	1395	453	632	387	2194	3674	1806
3	Venezuela	23359	1563	913	638	1564	933	1310	794	1428	630	1448	1870	875
4	Colombia	22618	996	717	483	803	551	822	659	999	454	1786	3395	1752
5	Maruecos	21909	1101	390	184	322	280	1393	320	930	342	1539	2223	942
6	Italia	20306	3030	1219	840	1817	1060	1194	1640	1195	710	826	915	412
7	Perú	18829	563	521	253	612	419	965	567	805	368	2026	2425	1131
8	Paraguay	18682	364	474	237	521	657	3311	584	1024	636	2061	2152	727
9	República Dominicana	17511	365	654	204	344	322	2272	443	589	536	1501	1607	1202
10	Honduras	15981	149	228	232	332	337	755	317	863	335	2021	2870	1115
11	Bolivia	14930	284	407	182	342	315	576	280	401	225	1458	2625	2827
12	Filipinas	12628	1344	640	142	578	661	4473	771	442	568	629	400	225
13	Portugal	9960	769	372	262	695	534	590	509	693	365	533	658	283
14	Francia	9561	1608	455	370	968	554	387	699	366	347	196	188	66
15	Ucrania	9453	152	214	133	220	176	221	149	312	168	1745	1251	428
16	Brasil	9324	677	309	244	431	280	567	322	361	234	1159	1596	410
17	Bulgaria	7842	262	137	115	113	123	245	74	316	170	996	988	395
18	Estados Unidos de América	6791	1637	406	385	749	389	300	657	297	428	207	125	86
19	Reino Unido	5915	1274	324	256	550	466	329	501	313	304	211	170	81
20	Bangladesh	5886	2742	381	33	32	21	210	48	27	14	257	410	175
21	Cuba	5725	435	243	111	169	161	257	175	305	133	395	650	301
22	Polonia	5487	246	173	117	138	130	220	151	184	133	801	706	296
23	Argentina	5061	588	262	186	323	208	271	291	261	188	284	426	230
24	Alemania	4707	592	199	175	494	442	245	367	607	247	140	111	60
25	Nicaragua	4697	82	107	101	174	217	248	111	215	137	430	556	291
26	México	4503	644	213	173	656	283	266	359	263	184	205	199	89
27	El Salvador	3228	71	58	82	82	91	135	81	115	72	226	413	205
28	Rusia	2522	178	117	89	215	138	132	113	127	102	186	131	92
29	Chile	2520	285	124	84	212	76	122	161	161	85	146	164	69
30	Senegal	2106	507	235	24	31	18	63	18	37	14	40	239	62
31	India	1704	138	67	40	126	142	136	72	71	31	102	74	74
32	Países Bajos	1615	218	77	90	134	103	78	127	68	61	47	89	48

Puente de Vallecas	Moratalaz	Ciudad Lineal	Hortaleza	Villaverde	Villa de Vallecas	Vicalvaro	San Blas-Canillejas	Barajas
4784	1286	2888	1466	3646	3384	2606	2929	661
3602	564	1960	1104	1236	685	472	972	190
3290	491	2471	401	2017	498	439	1015	138
1829	480	1858	1434	909	762	321	1486	314
1733	482	1792	910	1618	740	384	1282	258
3437	258	1011	426	3372	1655	802	649	333
704	310	1258	1109	330	427	189	786	337
2079	668	1726	603	1280	564	338	810	106
1354	360	1619	583	870	217	199	581	151
1989	223	1581	359	1881	296	151	889	103
2483	281	1062	349	1040	387	165	596	62
1573	227	1086	479	737	222	159	454	71
418	76	590	300	140	25	24	152	30
597	173	612	727	302	262	168	565	191
138	85	599	1802	68	59	52	357	197
912	239	348	282	1649	256	152	390	56
604	201	421	388	417	223	81	308	91
760	285	780	283	516	261	269	662	92
139	232	252	206	54	51	33	101	57
142	59	285	319	61	49	27	120	74
649	23	179	27	489	57	32	78	2
513	118	483	248	362	175	88	325	78
424	88	266	194	506	164	273	213	62
319	74	317	270	162	119	50	157	75
93	39	240	327	39	46	25	130	89
601	212	523	186	203	69	58	147	29
133	38	217	244	60	77	19	158	43
575	102	264	113	244	39	43	189	28
166	50	141	195	120	67	33	93	37
172	80	187	112	90	46	25	105	14
270	11	52	38	248	86	31	69	13
93	14	142	97	99	19	6	146	15
74	21	99	117	41	22	11	55	35

NEIGHBORHOOD MAP



TOP 10 VENUES PER NEIGHBORHOOD

- This will show us the path to the most popular venues.
- The next step is to marge the data and plot a map containing the assigned K cluster per region.

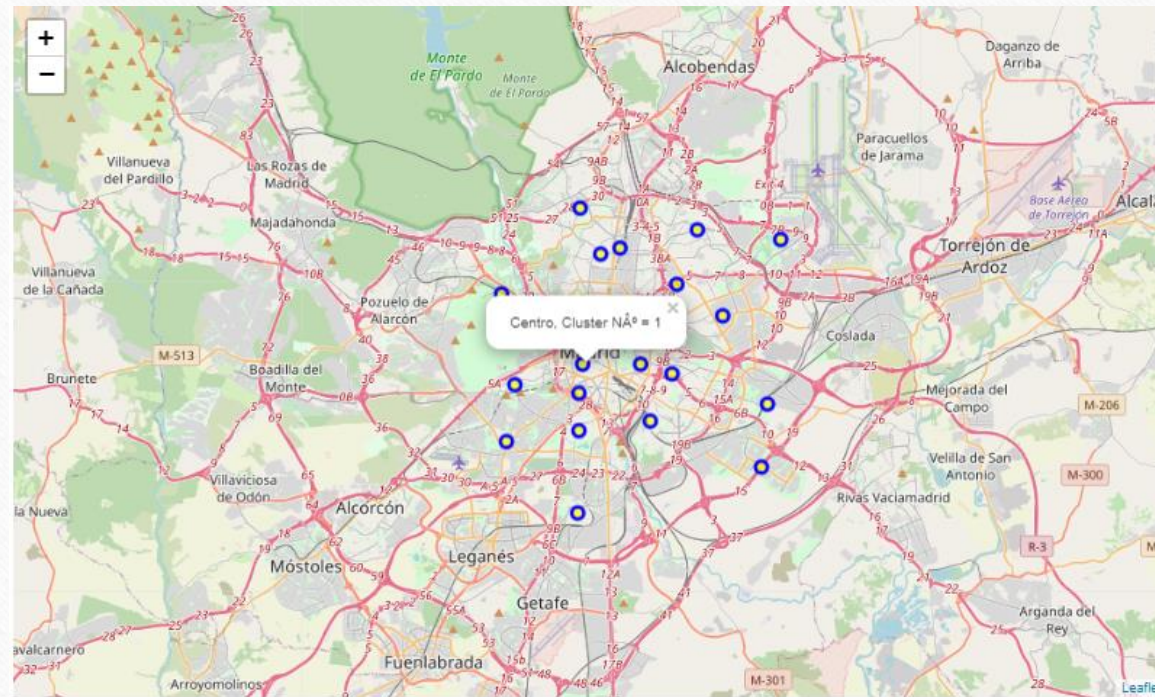
Out[981]:

	Neighborhood	1	2	3	4	5	6	7	8	9	10
0	Arganzuela	Spanish Restaurant	Park	Bar	Grocery Store	Plaza	Playground	Beer Garden	Tapas Restaurant	Hotel	Restaurant
1	Barajas	Hostel	Coffee Shop	Food & Drink Shop	Café	Tapas Restaurant	Mexican Restaurant	Supermarket	Restaurant	Wine Shop	Dessert Shop
2	Carabanchel	Bar	Gym / Fitness Center	Gastropub	Mobile Phone Shop	Convenience Store	Cosmetics Shop	Creperie	Deli / Bodega	Department Store	Farmers Market
3	Centro	Tapas Restaurant	Spanish Restaurant	Bar	Hotel	Plaza	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Café	Hostel	Dessert Shop
4	Chamartin	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Tapas Restaurant	Hotel	American Restaurant	Nightclub	Paella Restaurant	Salad Place	Pizza Place

[982]: `sorted_venues.shape`

Out[982]: (21, 11)

MAP WITH K CLUSTER REGIONS



CLUSTERS THAT ANSWERED MY QUESTION

- CLUSTER 1:

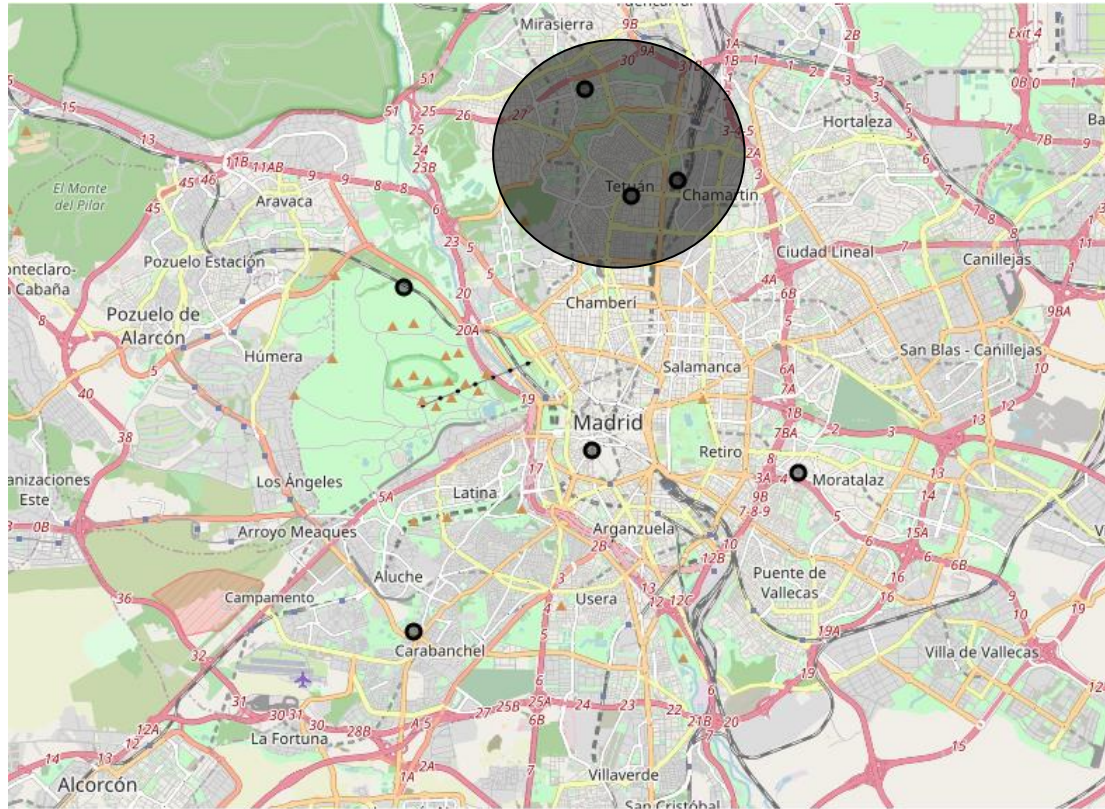
There's a larger variation of countries and regions in this cluster. We can mainly find Latin American countries, occupying 5/12 rows in this cluster. The top 3 most popular venues in each region are related to the food industry, such as: 'Spanish Restaurants', 'Tapas Restaurants', and 'Grocery Stores'.

- CLUSTER 2:

The country density decreases again, however, it's not all bad new, Honduras is the only participant in this cluster, making this a very useful data set. Moreover, several of the top 10 venues in this region are related to food.

- CLUSTER 3:

I found cluster 3 to be just as lucky, as Ecuador is the sole participant of this cluster. There's a lower density of food-related venues in this region, however, we can find a couple of Bars and Pubs in the data.



CONCLUSION (Pt.1)

By clustering clusters 1, 2, and 3, we get the ultimate trio. Most of these clusters have food-related venues as their top 5. In cluster 1, we find a high density of Latin countries, occupying 5/12 rows in the data frame. To make the conclusion easier to visualize, I made a map containing the Latin counties in Clusters 1, 2, and 3:

CONCLUSION

(Pt. 2)

In the code, the map is interactive, and we can see how in the northern region we find foreigners from Peru, Colombia, and Brazil. Madrid's subway system offers fast and easy connectivity between Madrid Centro and Chamartín. Considering all of these factors simultaneously, I conclude that the best location of our restaurant is in the region of Chamartín. It's far enough from the Center, so it won't be shadowed by other venues with higher reputation, but it's close enough to the Center, allowing the subway system to do its magic. Moreover, because our main theme is based around Colombia, and Chamartín's main foreign population are Colombians, this creates an added value.