IBM Capstone Project - The Battle of Neighborhoods

The Selection of Neighborhoods for Staying in a suitable location in Madrid

Applying Data Science using Python & Machine learning to identify the suitable place within surrounding areas

By Debesh Roy
17Aug2020

1. Introduction

Due to the present era of Globalization, people are moving around the globe, knowing or without knowing much about the destination and surrounding neighborhoods, where they are planning to stay for some time – months to years.

2. Background

Madrid is the capital and most populous city of Spain. The city has almost 3.3 million inhabitants and a metropolitan area population of approximately 6.5 million. It is the second-largest city in the European Union (EU), surpassed only by Berlin, and its monocentric metropolitan area is the second-largest in the EU, surpassed only by Paris. The municipality covers 604.3 km2 (233.3 sq mi).

Madrid is a city of elegant boulevards and expansive, manicured parks and many more. It's renowned for its rich repositories of European art, including the Prado Museum's works by Goya, Velázquez and other Spanish masters. The heart of old Hapsburg Madrid is the portico-lined Plaza Mayor, and nearby is the baroque Royal Palace and Armory, displaying historic weaponry.

In this project we will try to find a suitable locations for staying, who is planning to relocate to Madrid for some time, based on his/her personal preferences of the surrounding neighborhoods and available venues.

We will use the data science mechanism to generate more information about the nearby venues and whereabouts, which will help stakeholders or interested people to take the decision and plan accordingly.

3. Data

The related information about Madrid got downloaded from Wikipedia, i.e.
https://en.wikipedia.org/wiki/Districts_of_Madrid

We extracted info like District, Area, Population, Density, corresponding Wards from the Wiki, and definitely the Venue & Category details using Foursquare URL https://foursquare.com and corresponding APIs.

3.1.District list

	Number	Name	Area (KM Sqr.)	Population	Population density(/KM Sqr.)
0	1.0	Centro	5.23	131928	25225.24
1	2.0	Arganzuela	6.46	151965	23523.99
2	3.0	Retiro	5.47	118516	21666.54
3	4.0	Salamanca	5.39	143800	26679.04
4	5.0	Chamartín	9.18	143424	15623.53
5	6.0	Tetuán	5.37	153789	28638.55
6	7.0	Chamberí	4.68	137401	29359.19
7	8.0	Fuencarral-El Pardo	237.84	238756	1003.85
8	9.0	Moncloa-Aravaca	46.53	116903	2512.42
9	10.0	Latina	25.43	233808	9194.18

3.2.District wise ward count

	Name	Number
0	Arganzuela	7
1	Barajas	5
2	Carabanchel	7
3	Centro	6
4	Chamartín	6
5	Chamberí	6
6	Ciudad Lineal	9
7	Fuencarral-El Pardo	8
8	Hortaleza	6
9	Latina	7
10	Moncloa-Aravaca	7

3.3.District wise ward list

	Number	Name	Ward
0	1.0	Centro	Palacio
1	1.0	Centro	Embajadores
2	1.0	Centro	Cortes
3	1.0	Centro	Justicia
4	1.0	Centro	Universidad

Also using the “geopy” library and got the Latitude & Longitude for all districts and wards.

4. Methodology

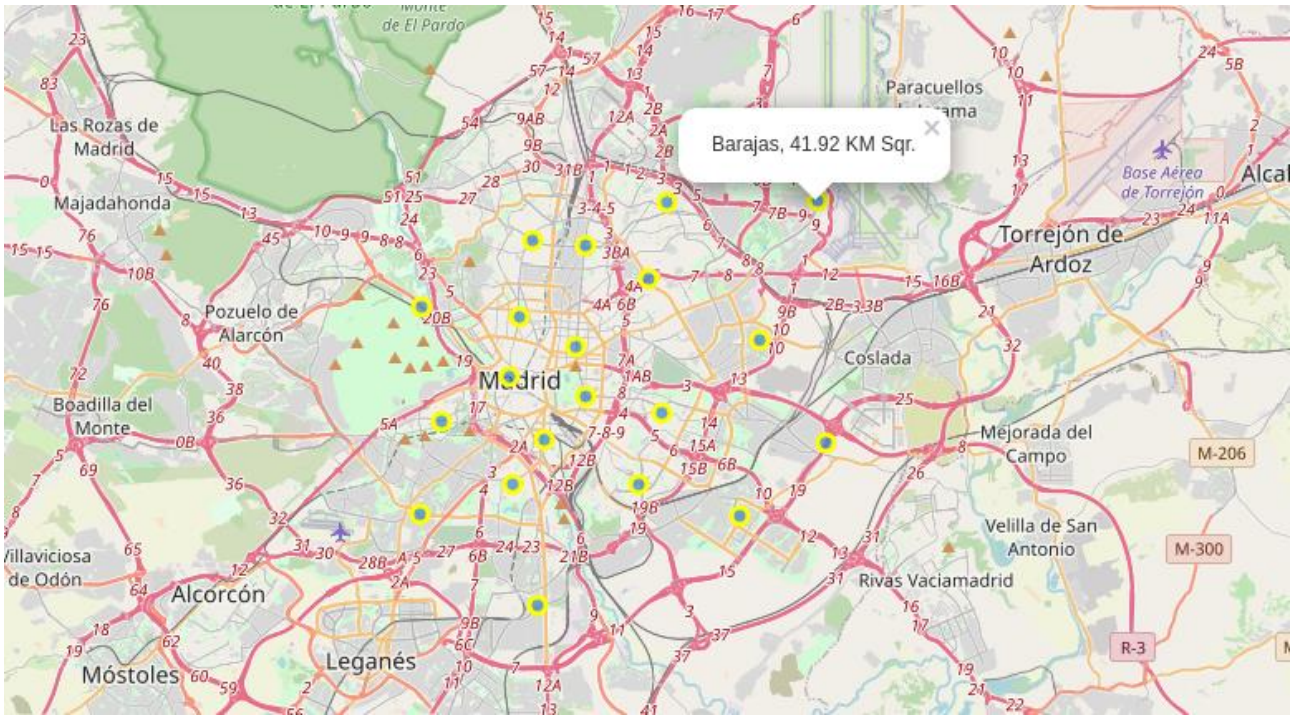
The base of the project is Python, where we extensively used multiple important libraries, i.e. pandas, numpy, geocoder, folium, and many more. Each one of the used library is having dedicated purpose.

Used pandas’ “read_html” to read the information from wiki URL. Used DataFrame feature extensively to handle data, i.e. transforming, cleansing, etc.

This data handling helps to identify the District wise Ward wise Venue and the Category.

	Name	Category	Venue
475	San Blas-Canillejas	Italian Restaurant	Piazza Italia
476	San Blas-Canillejas	Medical Center	Centro de Salud Los Alpes
477	San Blas-Canillejas	Bank	BBVA oficina 2273
478	San Blas-Canillejas	Pizza Place	Domino's Pizza
479	San Blas-Canillejas	Soccer Stadium	Estadio Wanda Metropolitano
480	San Blas-Canillejas	Yoga Studio	Yoga Dhairyam
481	San Blas-Canillejas	None	Animal Party
482	San Blas-Canillejas	Pet Service	Centro Veterinario Los Alpes I
483	San Blas-Canillejas	School	CEIP Julián Marías
484	San Blas-Canillejas	Playground	Parque Sofía con Manchester
485	San Blas-Canillejas	Office	MAPFRE Oficina
486	San Blas-Canillejas	Cosmetics Shop	Peluqueria Marco Aldany

Used Florium map feature, which will help interested stakeholder to identify the surrounding venues, categories, neighborhoods, etc.



5. Results

Using the Python and corresponding libraries, we identified that there are 21 districts in Madrid, along with the corresponding area, population and density.

	Number	Name	Area (KM Sqr.)	Population	Population density(/KM Sqr.)
0	1.0	Centro	5.23	131928	25225.24
1	2.0	Arganzuela	6.46	151965	23523.99
2	3.0	Retiro	5.47	118516	21666.54
3	4.0	Salamanca	5.39	143800	26679.04
4	5.0	Chamartín	9.18	143424	15623.53
5	6.0	Tetuán	5.37	153789	28638.55
6	7.0	Chamberí	4.68	137401	29359.19
7	8.0	Fuencarral-El Pardo	237.84	238756	1003.85
8	9.0	Moncloa-Aravaca	46.53	116903	2512.42
9	10.0	Latina	25.43	233808	9194.18
10	11.0	Carabanchel	14.05	243998	17366.41
11	12.0	Usera	7.78	134791	17325.32
12	13.0	Puente de Vallecas	14.97	227595	15203.41
13	14.0	Moratalaz	6.10	94197	15442.13
14	15.0	Ciudad Lineal	11.43	212529	18593.96
15	16.0	Hortaleza	27.42	180462	6581.40
16	17.0	Villaverde	20.19	142608	7063.30
17	18.0	Villa de Vallecas	51.47	104421	2028.77
18	19.0	Vicálvaro	35.27	70051	1986.14
19	20.0	San Blas-Canillejas	22.29	154357	6924.94
20	21.0	Barajas	41.92	46876	1118.23

We also identified that each district is having on an average 6~7 wards.

	Name	Number
0	Arganzuela	7
1	Barajas	5
2	Carabanchel	7
3	Centro	6
4	Chamartín	6
5	Chamberí	6
6	Ciudad Lineal	9
7	Fuencarral-El Pardo	8
8	Hortaleza	6
9	Latina	7
10	Moncloa-Aravaca	7
11	Moratalaz	6
12	Puente de Vallecas	6
13	Retiro	6
14	Salamanca	6
15	San Blas-Canillejas	8
16	Tetuán	6
17	Usera	7
18	Vicálvaro	4
19	Villa de Vallecas	3
20	Villaverde	5

And the corresponding wards for each district

	Number	Name	Ward
0	1.0	Centro	Palacio
1	1.0	Centro	Embajadores
2	1.0	Centro	Cortes
3	1.0	Centro	Justicia
4	1.0	Centro	Universidad
...	...	...	...
126	21.0	Barajas	Alameda de Osuna
127	21.0	Barajas	Aeropuerto
128	21.0	Barajas	Casco Histórico de Barajas
129	21.0	Barajas	Timón
130	21.0	Barajas	Corralejos

6. Discussion

Well, the purpose of this Python notebook is to identify the suitable location to stay. The tool will not suggest the Good or Bad neighbor, but based on personal interest, whether you are looking for School or College, trade off between Shopping Mall and Supermarket, preference between Candy Store and Ice Cream Shop, looking for Restaurant or Bar, and many more.

Along with the data point, from the Florium map one can also identify the near around venues and approx. distances.

7. Conclusion

It’s always preferred to have a careful consideration before selecting a neighborhood, where one is going to stay for some time. Though this report will not decide the best place but it’s up to individual choice to finalize.

After going through the detailed analysis between districts, wards and areas, once can easily get a fair knowledge on the surroundings and can conclude. Based on that, s/he can either put on advertise on the local media, or just can pay a visit to feel the ambience of the place.