



# **COURSERA CAPSTONE PROJECT ANALYSIS OF NEIGHBORHOODS IN MADRID**

**by VITALII BARDASHEVSKYI**

# I. INTRODUCTION

Madrid is the capital and most populous city of Spain with more than 3 million inhabitants.

Also, it is the political, economic and cultural center of the country with many opportunities and advantages. In addition, it is the second-largest city of the European Union.

All these facts make Madrid attractive to investments and a very interesting place for launching new business

## PROBLEM DEFINITION

Group of investors is interested in launching a new chain of sport food stores. For this purpose they would like to get the following information:

- what districts (hence neighborhoods) of Madrid are most populous (with the most quantity of inhabitants per square kilometer)
- how many gym, fitness, athletic (and all other related) venues are concentrated in most populous districts (hence neighborhoods) of Madrid\*
- visualization of location all these venues on the map of Madrid
- list of these venues with name and location

All this information would be helpful in choosing the best locations for launching sport food stores. Because main targeting auditory – people who visit gyms, fitness and other sports clubs. List of such venues would be useful in case of cooperation and building partnership relationships with them to promote sport food products

So the main goal is:

» to determine neighborhoods of Madrid with the highest density of target venues (gym, fitness direction) along with the list of these venues and then visualize this information

\*Analysis takes into account only the most populous districts of Madrid

## II. DATA DESCRIPTION

For implementing our analysis we use the following tools and sources of information:

1. Information about the population density of Madrid by neighborhood: Wikipedia <https://en.wikipedia.org/wiki/Madrid> ("Población por distrito y barrio". Banco de datos. Ayuntamiento de Madrid); [https://en.wikipedia.org/wiki/Districts\\_of\\_Madrid](https://en.wikipedia.org/wiki/Districts_of_Madrid) (Data as of 1st January 2017. Source: Área de Gobierno de Economía, Hacienda. Subdirección General de Estadística. Padrón Municipal de Habitantes ("Anexo:Distritos de Madrid", Wikipedia, la enciclopedia libre (in Spanish), 2019-10-06, retrieved 2019-12-23)
2. Dataset that contains a list of target venues with name, category and location coordinates: Foursquare API <https://foursquare.com/>
3. JSON files with geographical information about districts and neighborhoods of Madrid so that we can visualize on map: GitHub <https://github.com/>
4. Location coordinates (latitude and longitude) of neighborhoods of Madrid: <https://www.mapcoordinates.net/en>

### III. METHODOLOGY

Initially, we should obtain the data from Wikipedia sources and clean it. We do it by scraping with Python package *pandas*

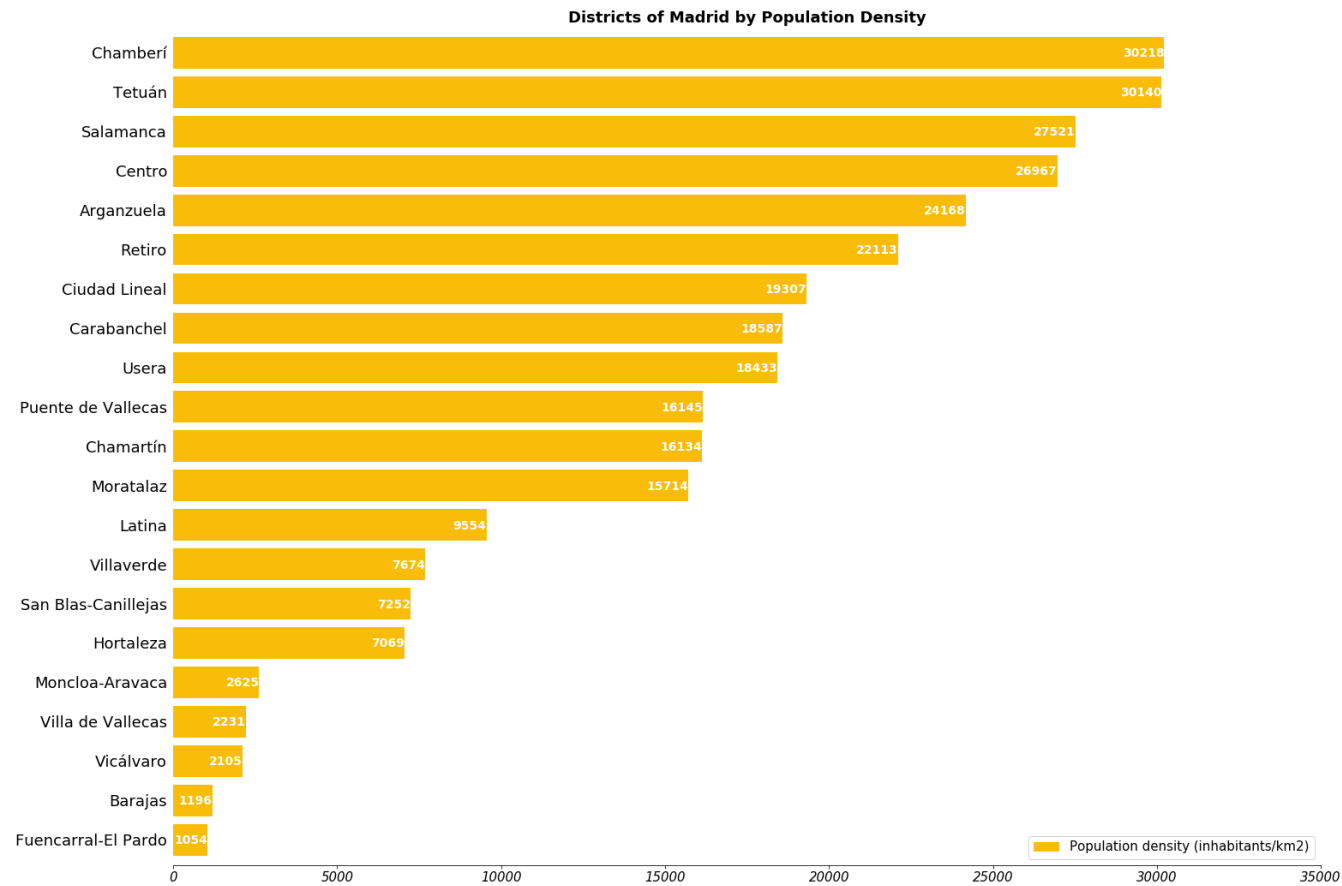
The first table contains actual information about population (1 Jan 2020), area (ha) and names of districts.

The second table contains along with already mentioned above (but information about the population is outdated) names of neighborhoods relatively to each district

After several manipulations over the tables (creating columns with area measured in square kilometers, cleaning column with the names of neighborhoods, joining two tables, adding columns with district's latitude and longitude, dropping non-useful columns) we obtain only one table (its head and a few rows are represented below)

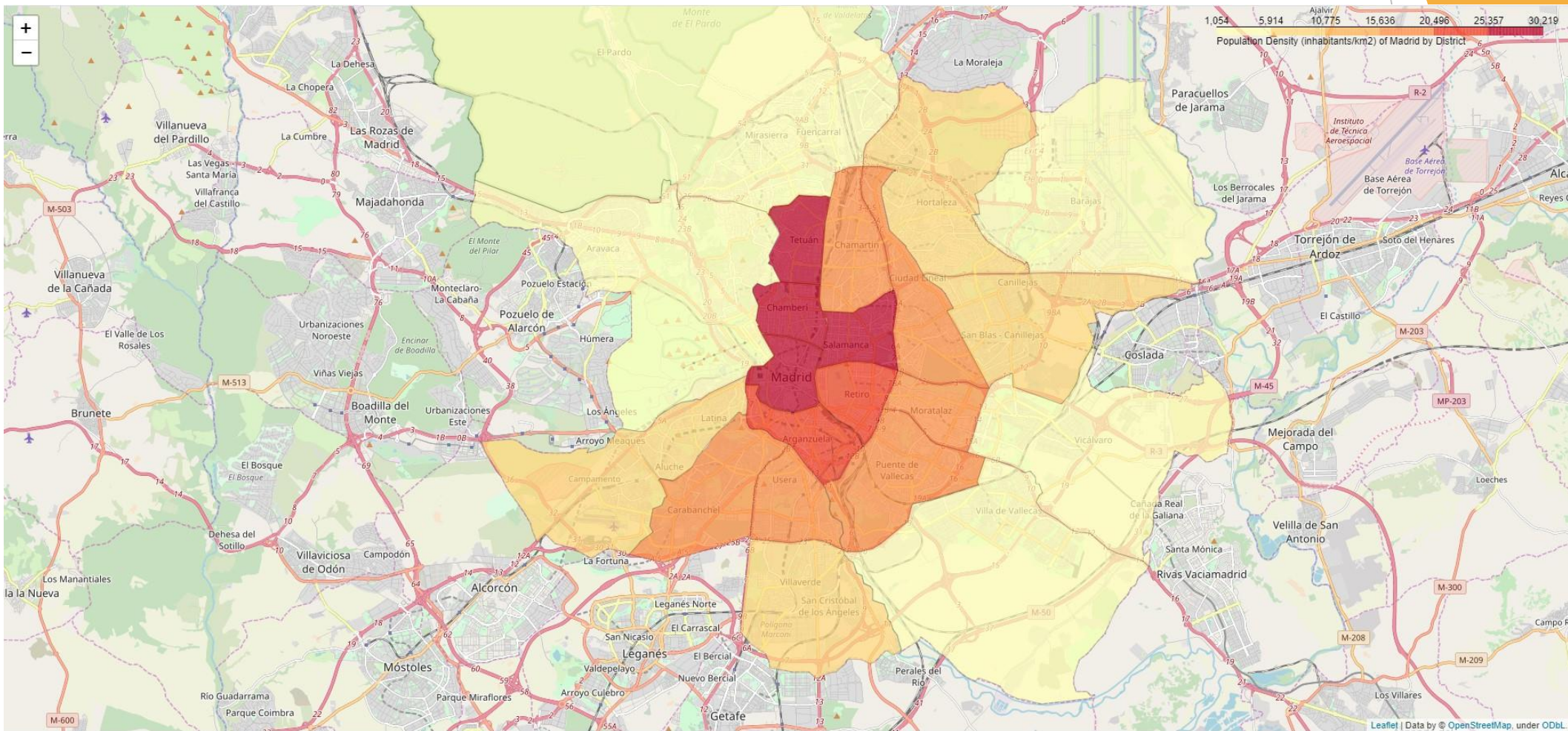
	District	Population	Neighborhoods	Area (km2)	Population Density (inhabitants/km2)	Latitude	Longitude
0	Centro	140991	Palacio, Embajadores, Cortes, Justicia, Universidad, Sol	5.2282	26967	40.417552	-3.707941
1	Arganzuela	156176	Imperial, Acacias, Chopera, Legazpi, Delicias, Palos de Moguer, Atocha	6.4622	24168	40.396954	-3.697289
2	Retiro	120873	Pacífico, Adelfas, Estrella, Ibiza, Jerónimos, Niño Jesús	5.4662	22113	40.411150	-3.676057
3	Salamanca	148405	Recoletos, Goya, Fuente del Berro, Guindalera, Lista, Castellana	5.3924	27521	40.427045	-3.680602
4	Chamartín	148039	El Viso, Prosperidad, Ciudad Jardín, Hispanoamérica, Nueva España, Castilla	9.1755	16134	40.458987	-3.676129

Here we use the power of Python library *matplotlib* to make a visual description of our dataset

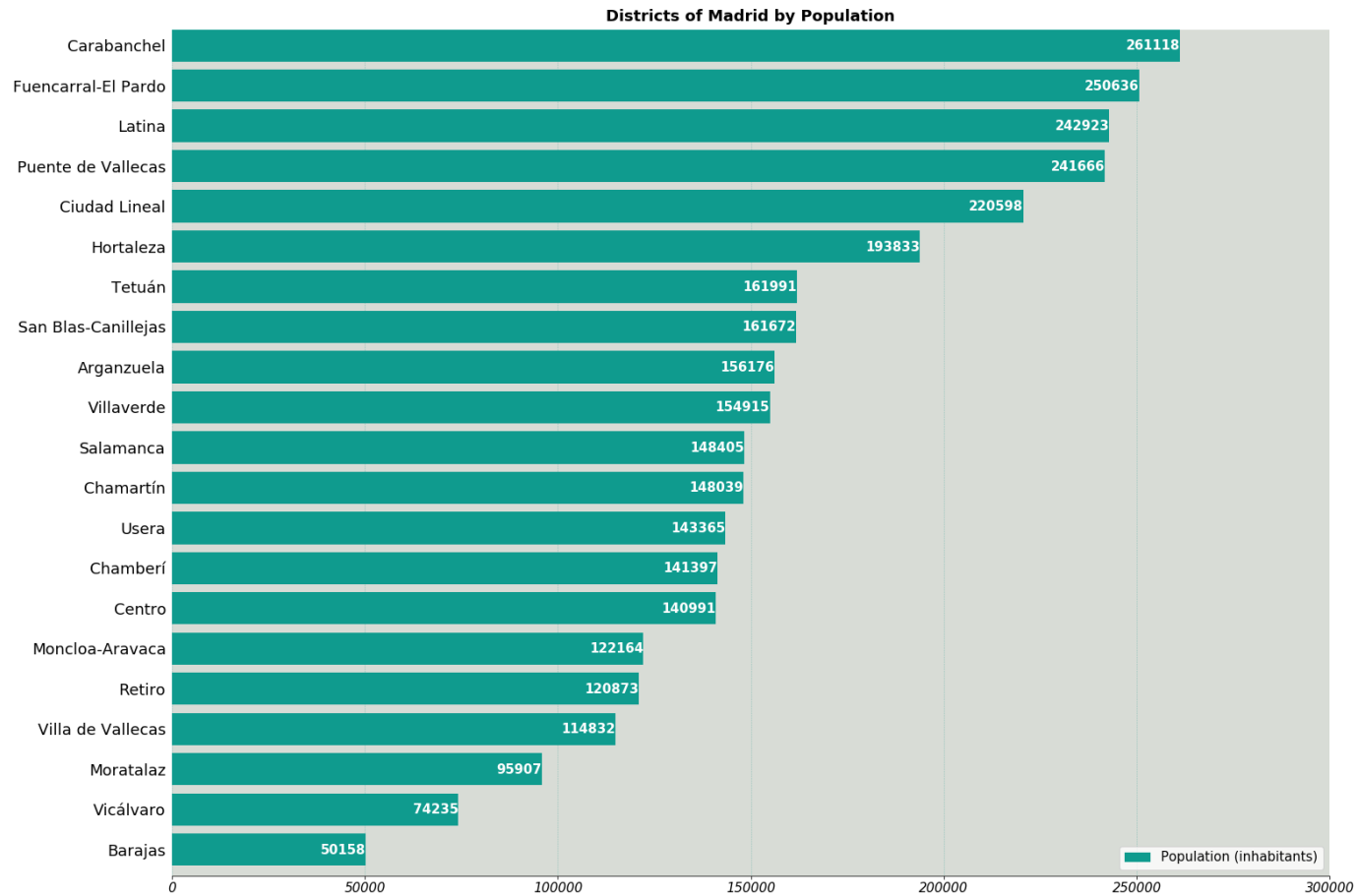




With *folium* package and JSON file that contains geographical data about districts of Madrid we can provide the information of previous slide displaying it on an interactive map

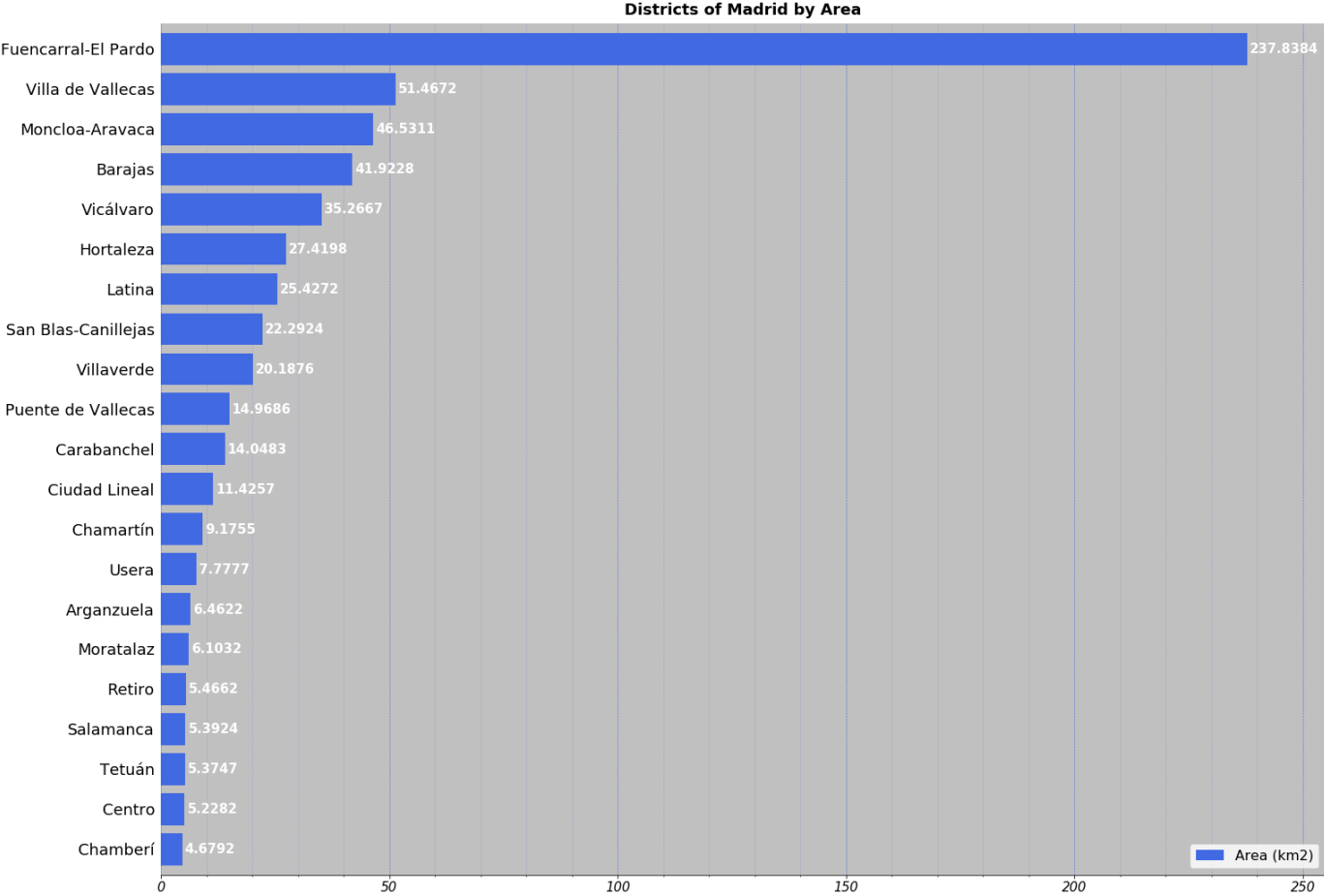


In addition two other bar charts with descriptive information about districts of the capital of Spain. Here is one





Here is another



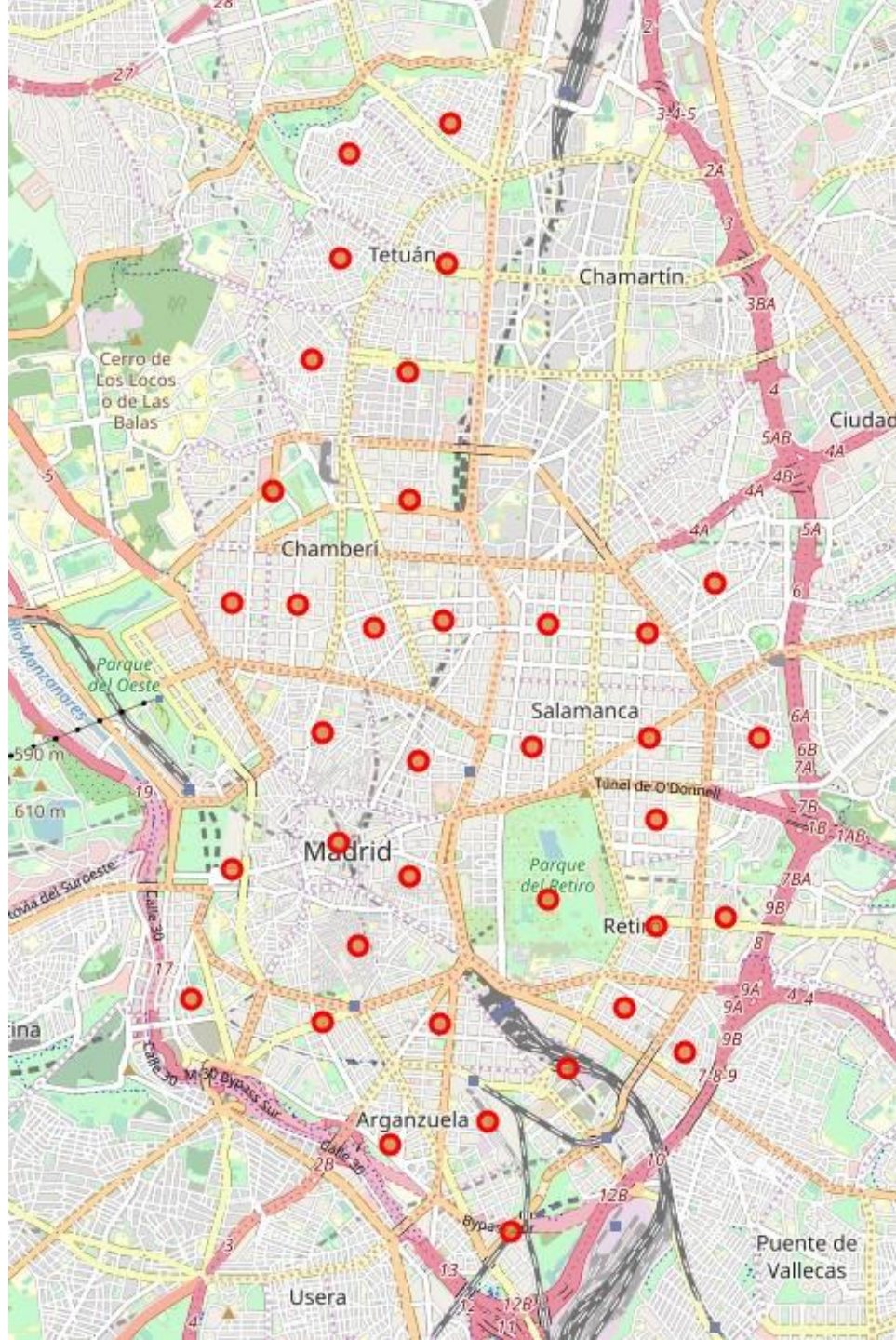
Returning to our first two visualizations with the yellow bar chart and the map that are described the population density by district we can select the six districts with the highest indicator. And from now these ones will be in the focus of our interest:

- » Chamberí
- » Tetuán
- » Salamanca
- » Centro
- » Arganzuela
- » Retiro

But to solve the problem for the better way and to obtain more exact result we should make research not only of these districts but dive a little bit deeper and analyze all neighborhoods relevant to them

For this reason, we make a new database that will be useful in our future work (this table along with all considered neighborhoods pointed on the map are represented on the next slide)

	District	Neighborhood	Latitude	Longitude
0	Chamberí	Gaztambide	40.434936	-3.714709
1	Chamberí	Arapiles	40.434764	-3.708315
2	Chamberí	Trafalgar	40.433061	-3.701018
3	Chamberí	Almagro	40.433663	-3.694286
4	Chamberí	Ríos Rosas	40.442570	-3.697629
5	Chamberí	Vallehermoso	40.443189	-3.710823
6	Tetuán	Bellas Vistas	40.452791	-3.707013
7	Tetuán	Cuatro Caminos	40.451847	-3.697800
8	Tetuán	Castillejos	40.459879	-3.694067
9	Tetuán	Almenara	40.470196	-3.693638
10	Tetuán	Valdeacederas	40.467911	-3.703423
11	Tetuán	Berruguete	40.460271	-3.704345
12	Salamanca	Recoletos	40.424364	-3.685743
13	Salamanca	Goya	40.425028	-3.674419
14	Salamanca	Fuente del Berro	40.424977	-3.663700
15	Salamanca	Guindalera	40.436356	-3.668019
16	Salamanca	Lista	40.432694	-3.674509
17	Salamanca	Castellana	40.433324	-3.684132
18	Centro	Palacio	40.415329	-3.714709
19	Centro	Embajadores	40.409687	-3.702509
20	Centro	Cortes	40.414779	-3.697584
21	Centro	Justicia	40.423276	-3.696680
22	Centro	Universidad	40.425409	-3.705989
23	Centro	Sol	40.417256	-3.704452
24	Arganzuela	Imperial	40.405800	-3.718685
25	Arganzuela	Acacias	40.404075	-3.705957
26	Arganzuela	Chopera	40.395098	-3.699482
27	Arganzuela	Legazpi	40.388628	-3.687779
28	Arganzuela	Delicias	40.396784	-3.690038
29	Arganzuela	Palos de Moguer	40.403976	-3.694602
30	Arganzuela	Atocha	40.400687	-3.682391
31	Retiro	Pacífico	40.405175	-3.676750
32	Retiro	Adelfas	40.401903	-3.670958
33	Retiro	Estrella	40.411762	-3.666998
34	Retiro	Ibiza	40.418953	-3.673725
35	Retiro	Jeronimos	40.413144	-3.684218
36	Retiro	Niño Jesús	40.411094	-3.673720





The next step is to explore the six most populous districts of Madrid with all neighborhoods relevant to them.

To implement this task we appeal to Foursquare API and obtain all venues of each neighborhood (within a 500-meter radius), then we convert it to dataframe. The resulting table comprises of 2293 rows and 7 columns. It includes information about many different categories of venues that are far beyond the scope of our analysis.

So we slice it, drop (delete) duplicates and get the dataset with 45 rows and 7 columns that include only information about targeting venues (its fragment is represented below)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Gaztambide	40.434936	-3.714709	Smartclub	40.434545	-3.712652	Gym
1	Gaztambide	40.434936	-3.714709	Club Deportivo José Valenciano	40.432614	-3.714972	Boxing Gym
3	Trafalgar	40.433061	-3.701018	Fightland	40.435452	-3.702157	Boxing Gym
4	Almagro	40.433663	-3.694286	Metropolitan Abascal	40.437688	-3.693865	Gym / Fitness Center
5	Almagro	40.433663	-3.694286	Fitup Chamberí	40.436280	-3.697453	Gym
6	Vallehermoso	40.443189	-3.710823	GoFit Vallehermoso	40.440798	-3.709032	Gym / Fitness Center
7	Vallehermoso	40.443189	-3.710823	Canal de Isabel II Area Deportiva	40.442743	-3.705524	Athletics & Sports
8	Cuatro Caminos	40.451847	-3.697800	Holmes Place Premium Fitness Clubs	40.452965	-3.692915	Gym / Fitness Center
9	Cuatro Caminos	40.451847	-3.697800	Paidesport	40.449942	-3.693264	Gym
10	Castillejos	40.459879	-3.694067	HealthCity	40.457513	-3.694797	Gym
11	Castillejos	40.459879	-3.694067	Holiday Gym Cuzco	40.460451	-3.688372	Gym / Fitness Center
12	Almenara	40.470196	-3.693638	Básico Sport Center CDM Antonio Díaz Miguel	40.471220	-3.695356	Gym / Fitness Center
13	Almenara	40.470196	-3.693638	Centro Nacional de Entrenamiento de Esgrima	40.472376	-3.691622	Gym / Fitness Center
14	Berruguete	40.460271	-3.704345	Benefits Body Mind	40.462986	-3.702698	Gym / Fitness Center
15	Berruguete	40.460271	-3.704345	Playa Victoria, polideportivo	40.458726	-3.702130	Gym / Fitness Center

## IV. RESULTS

Now we can group our resulting table and get the quantity of targeting venues in each considered neighborhood.

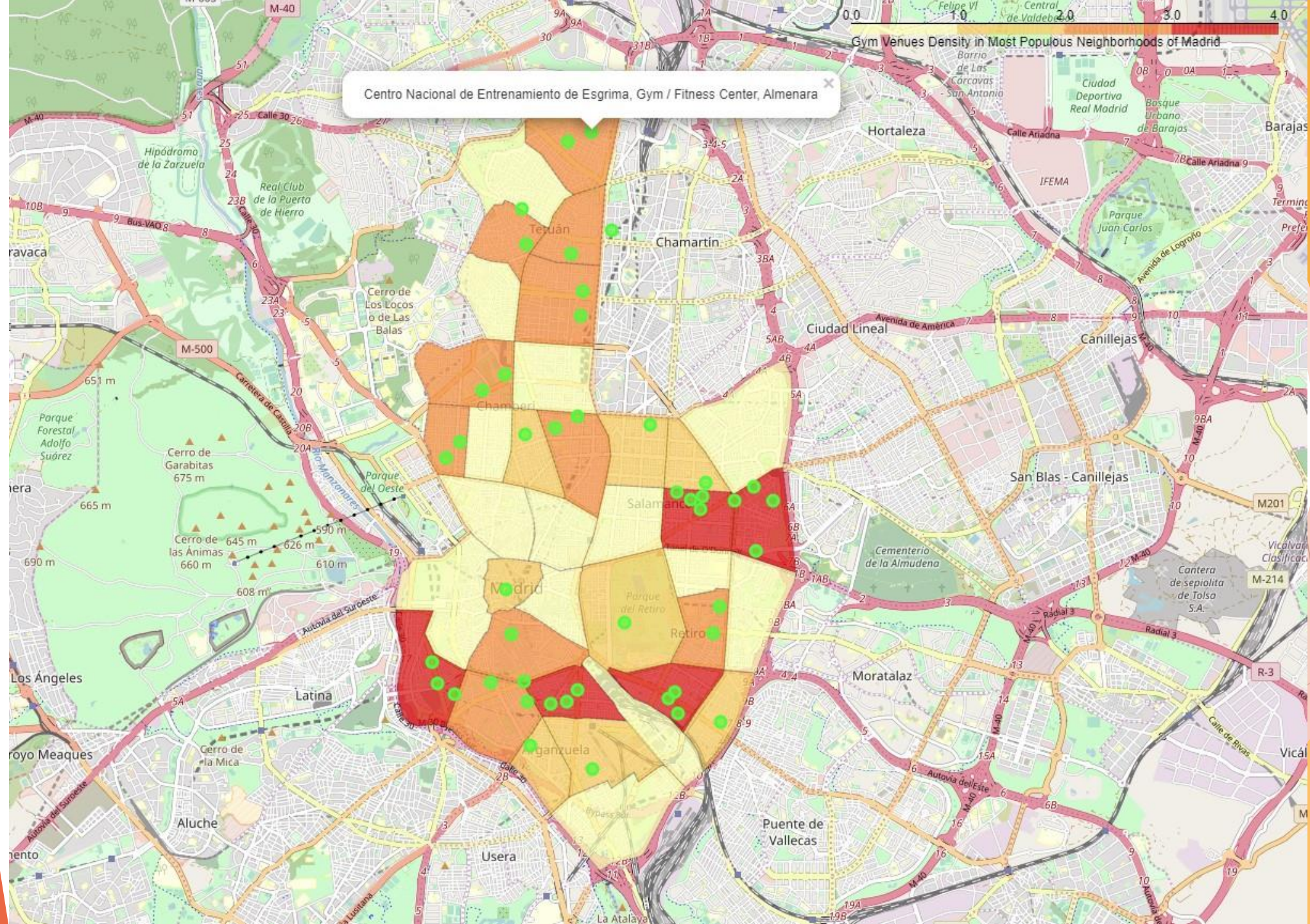
And again using *folium* and another JSON file\* that contains geographical information about neighborhoods of Madrid we provide the information that was determined as our main goal, that is: to determine neighborhoods of Madrid with the highest density of target venues (gym, fitness direction).

The screenshot of visual solution is represented on the next slide. This is an interactive map\*. Light green color points mean targeting venues. After clicking on any of those points we can see a pop-up label with the name of the venue, category and neighborhood. Red/orange color areas mean neighborhoods with the higher density of targeting venues.

The list of all targeting venues (precisely speaking its fragment) is represented on the previous slide. It is a table with 45 rows with information about the venue, its location coordinates, category and relevant neighborhood.

\*JSON file was edited to fit only for analyzed neighborhoods







## V. DISCUSSION

Based on the results of this project it is obvious that the most attractive to launch new sport food stores are the following neighborhoods of Madrid: Fuente del Berro, Goya, Pacífico, Palos de Moguer, Imperial. Because there are the highest density of targeting venues, such as gyms, fitness centers and so on. And people who visit such kind of venues – main targeting auditory.

To make more exact conclusion it would be better to provide analysis more deeply, including commercial property rental prices (and commercial property prices), foot-traffic, attendance rate of our targeting venues, reviews, perform risk assessment, etc.

However, all these activities are beyond the scope of this project...

Wish all the success and fulfillment towards the goal!