“The Battle of the Flavours”

Applied Data Science Capstone by IBM/Coursera

By Mario Rascón Merino (Aerospace Engineer)

December 30, 2019

Table of contents

1. Introduction: Business Case Problem

This project wants to identify suitable locations to open a to craft-beer shop around Madrid downtown. Recently there has been a solid tendency midway between gourmet-like style of food to do-it-yourself artisan way of doing things, which has pushed young people to these kind of shops, running away from industrialized breweries to embrace these authentic traditional flavours.

In almost every district of Madrid there are such places, where not only craft-beer is produced but also sold from other breweries all-around the spanish geography.

We will use data obtained from the different Madrid downtown districts **to find the most suitable location for the shop inside a 3km circle from Madrid City Center**, based on the similar shops around a distance of 500m, and the age distribution of the living population in those districts, **targeting people between a Young Index between de 25% and 75% percentiles**.

Also **the price per square meter** from the different above selected districts, looking for the cheapest one to open the shop as the final criteria to select a place.

2. Data

Based on definition of our problem, factors that will influence our decision are:

- Number of existing craft-beer shops in the neighborhood / district.
- Distance to other craft-beer shops in the neighborhood or district, if any.
- Age distribution in the districts
- Price in euros per square meter of the houses or shops in the districts, real-state.

Following data sources will be needed to extract/generate the required information:

- Centers of districts will be got from Google Maps plus WikiPedia
- Number of craft-beer shops or similar venues in every neighborhood will be obtained using **Foursquare API**
- Coordinate of Madrid center will be obtained using **Geocoder Python Library** of well known Madrid location (Puerta del Sol).

  https://geocoder.readthedocs.io/index.html

- Also distance between points given they coordinates can be reversed with **Geocoder**
- Housing prices and demographic information from Madrid districts will be got from Madrid´s city council web page:

https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Estadistica/Areas-de-informacion-estadistica/Areas-de-informacion-estadistica/?vgnextfmt=default&vgnextoid=9023c9fa0b23a210VgnVCM2000000c205a0aRCRD&vgnextchannel=b65ef78526674210VgnVCM1000000b205a0aRCRD

- From the different excel files got from the city council web page, the information needed from the districts will be grouped and filtered for later use along with the geographical data and venues data in order to complete the business case analysis.

2.1 Data Wrangling and Analysis from Madrid´s City Council

	Distrito	Índice de juventud	ObraNueva (Eur/m2)	ObraUsada (Eur/m2)	lon	lat
0	Ciudad de Madrid	70.179326	2624.77	3387.14	-3.691944	40.418889
1	Centro	51.610072	4521.91	4817.37	-3.707371	40.415347
2	Arganzuela	65.438845	3670.25	3698.09	-3.695403	40.402733
3	Retiro	48.715364	3989.25	4457.96	-3.676729	40.408072
4	Salamanca	47.506659	5674.82	5398.93	-3.677778	40.430000

Table 1. Information from Madrid City Council

With Geocoder we can calculate the distance from the district center coordinates to the center of Madrid downtown.

	Distrito	Índice de juventud	ObraNueva (Eur/m2)	ObraUsada (Eur/m2)	lon	lat	distance_to_center_in_m
0	Ciudad de Madrid	70.179326	2624.77	3387.14	-3.691944	40.418889	1063.914309
1	Centro	51.610072	4521.91	4817.37	-3.707371	40.415347	308.047330
2	Arganzuela	65.438845	3670.25	3698.09	-3.695403	40.402733	1732.631164
3	Retiro	48.715364	3989.25	4457.96	-3.676729	40.408072	2516.251744
4	Salamanca	47.506659	5674.82	5398.93	-3.677778	40.430000	2662.935100

Table 2. District Information with added calculation of distance to the center.

Finally we keep those districts at whose center will be at a minimum of 3km.

With that information arranged we perform some statistical calculations to the data, in order to get the Young Index and Real State prices. For the Young Index we target those districts with higher numbers in the 25%, 50% and 75% quartiles, whereas for the real state we always look for prices below the mean.

These aforementioned conditions will be the constraints posed to our problem along with the condition of not having a craft-beer shop in a radius of 500m around the selected locations.

The statistical analysis of districts show that in a radius of 3km around Puerta del Sol, the Young Index is higher on Moncloa and Arganzuela districts, also for these districts real-state mean prices are under the mean of 4300 eur/m2.

	Índice de juventud	ObraNueva (Eur/m2)
count	7.000000	7.000000
mean	56.383121	4329.911429
std	10.500189	1153.227586
min	44.947219	2624.770000
25%	48.111012	3793.160000
50%	51.610072	3989.250000
75%	65.861602	5098.365000
max	70.179326	5912.310000

Table 3. Statistical data for Madrid Districts. Young Index and Real State data.

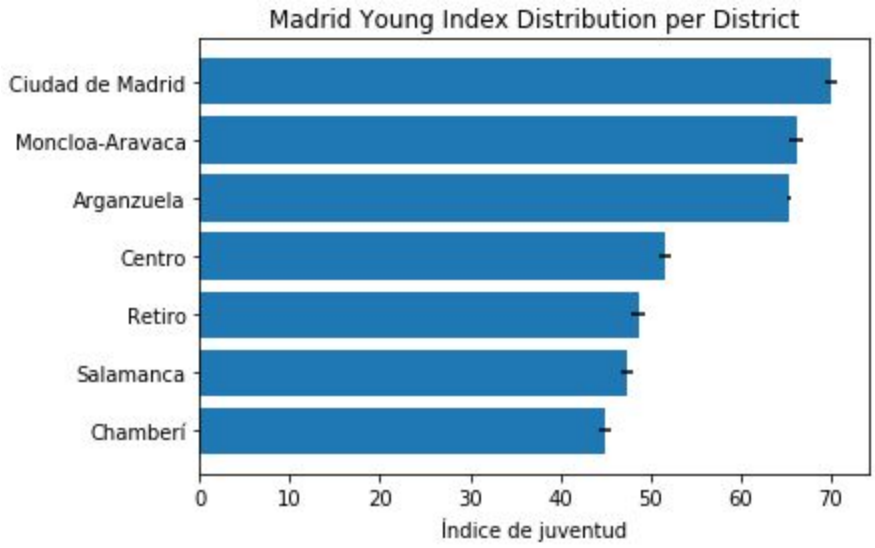

Fig. 1 Young Index Distribution for Madrid Districts at 3km from City Center.

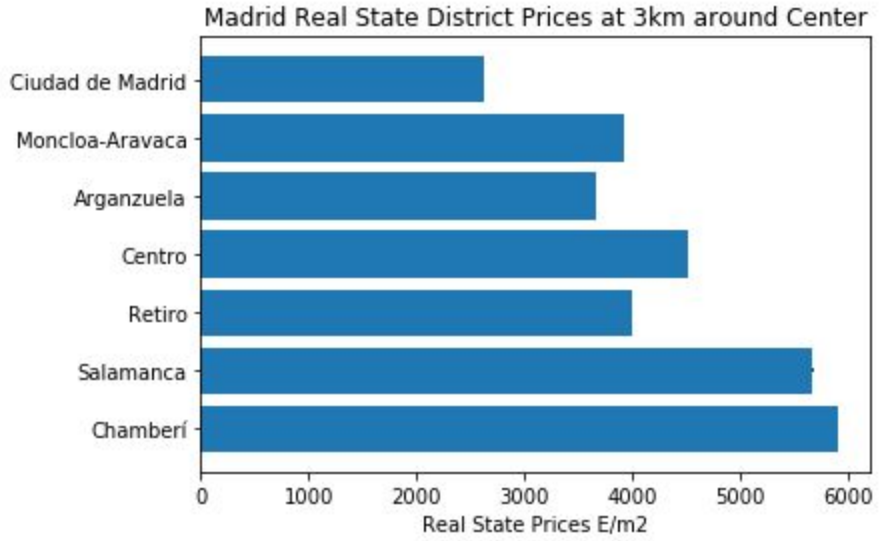

Fig.2 Real State prices fro Districs at 3km from Madrid ciy center

2.2 Data Wrangling and Analysis for Craft-Beer Shops got from Foursquare.

By simply using the Foursquare web page by typing craft-beer in Madrid, it can be retrieved a list of places and recommendations given by other users.

https://es.foursquare.com/user/65416126/list/craft-beer

https://foursquare.com/top-places/madrid/best-places-craft-beer

In our study we get from Foursquare developers page the following codes to get venues related to craft-beer in Madrid, in Spanish language (“Cervecera and Tienda de Cerveza) within a range of 5km around Madrid City Center (“Puerta del Sol”).

Cervecera 50327c8591d4c4b30a586d5d

Tienda de Cerveza 5370f356bcbc57f1066c94c2

We will perform two searches one per category, to get two dataframes which will be merged and cleaned later.

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLn
0	5bf1a840ad910e002c4a046b	La Tienda De La Cerveza - Craft Beer Bar	[{'id': '4bf58dd8d48988d116941735', 'name': 'B...	v-1577026670	False	Calle de las Maldonadas, 5	40.410942	-3.708029	[{'label': 'display', 'l 40.410942, 'lng
1	5905bec2a4236221271945b2	Craft 19 - Craft Beer & Pastrami	[{'id': '4bf58dd8d48988d1c5941735', 'name': 'S...	v-1577026670	False	Calle de Vallehermoso 36	40.434100	-3.708280	[{'label': 'display', 'l 40.4341, 'lng':
2	5b087bb6b9a5a8002c3477cb	Sam Hop Craft Beer	[{'id': '56aa371ce4b08b9a8d57356c', 'name': 'B...	v-1577026670	False	Calle De Lavapies	40.411468	-3.702996	[{'label': 'display', 'l 40.411468, 'lng
3	579905c6498e213c8e2b4f62	Monasterio Craft Beer	[{'id': '50327c8591d4c4b30a586d5d', 'name': 'B...	v-1577026670	False	Calle Rodriguez Sanpedro	40.431570	-3.708656	[{'label': 'display', 'l 40.43157, 'lng'
4	5d87fabc3955b60008e4dc5b	Mayrit craft beer	[{'id': '56aa371ce4b08b9a8d57356c', 'name': 'B...	v-1577026670	False	Quero 33	40.391490	-3.760105	[{'label': 'display', 'l 40.3914896191885

Table 3. Non-Cleaned Craft-Beer shops as given by Foursquare

After cleaning the data we keep the name, categories, address, latitude, longitude and distance to the center of Madrid as the data needed to proceed with our study.

Finally our dataset got from Foursquare will look like:

	name	categories	address	lat	lng	distance
0	Monasterio Craft Beer	Brewery	Calle Rodriguez Sanpedro	40.431570	-3.708656	1677
1	Madrid Shooter Beer	Brewery	Calle Palafox 1	40.430534	-3.700782	1550
2	Beershooter Malasaña	Brewery	Calle La Palma, 69	40.426713	-3.708091	1142
3	Cervezas Eterna	Brewery	Calle de Eraso 14	40.434117	-3.671035	3411
4	house of beer	Brewery	NaN	40.456413	-3.691424	4535
5	Beerhouse	Brewery	Calle Cardenal Cisneros 16	40.430512	-3.701767	1534
6	Bee Beer	Beer Store	C. Augusto Figueroa, 30	40.422077	-3.696958	850
7	Craft Crew Madrid	Beer Store	Barceló, 6	40.426647	-3.698970	1178
8	The Beer Garden Store	Beer Store	C. Cardenal Cisneros, 10	40.430335	-3.701842	1513
9	Sam Hop Craft Beer	Beer Bar	Calle De Lavapies	40.411468	-3.702996	610

Table 4. Foursquare data cleaned for Craft Beer shops

If we joint together the Craft Beer shops downloaded by Foursquare with the Index Young and Real State prices in order to get the actual view for these kind of business.

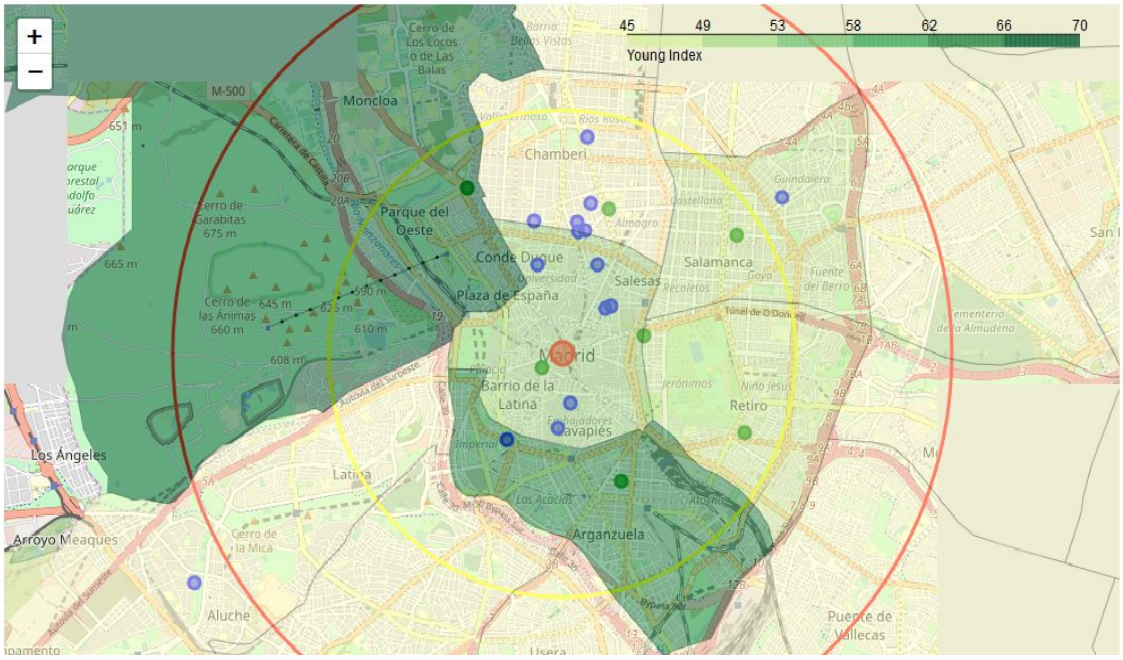

Fig. 3 Craft-Beer shops overlapped to Madrid Districts Young Index

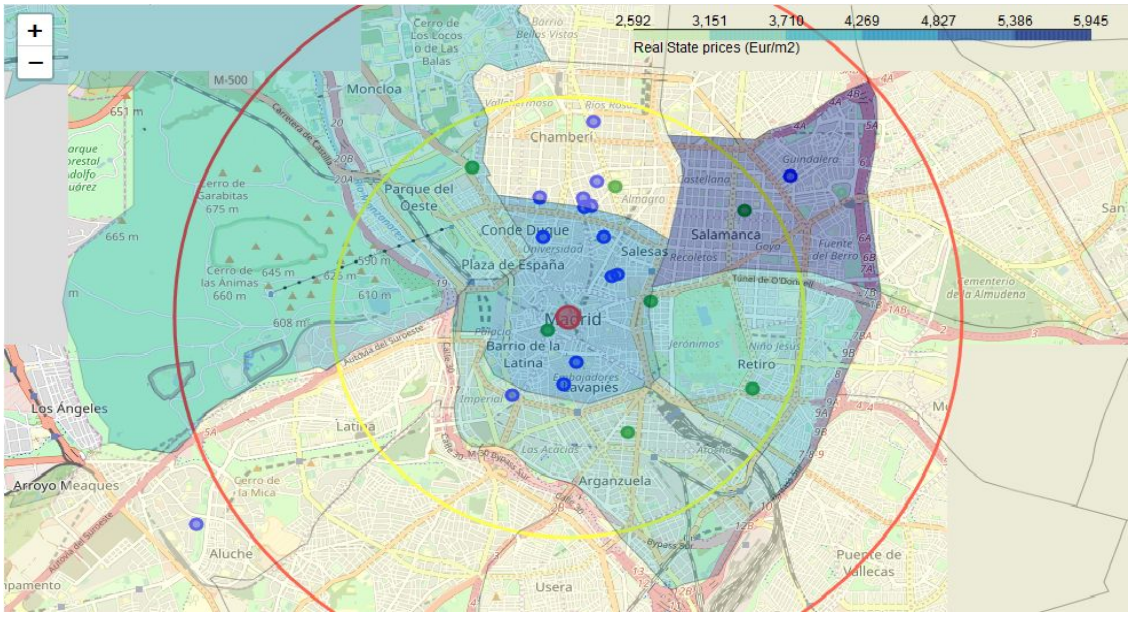

Fig. 4 Craft-Beer shops overlapped to Madrid Districts Real State Prices

2.3 First Conclusions from data wrangling

After the initial data analysis, it looks that the optimum places for our investors in which we can combine a high Young Index and moderate real-state prices are in the border lines around two zones , one between 'Centro' and Moncloa' districts and another between Centro and Arganzuela districts. *Always based on our initial condition of finding this place in a radius of 3km from Madrid City Center*.

3. Methodology

In this project we will direct our efforts on detecting areas of Madrid downtown with low number of Craft Beer shops. We will limit our analysis to an area ~ 3km around the city center.

In first step we have collected the required **data: location and type (category) of every Beer related place within 3km from Madrid center** ("Puerta del Sol). We have also **identified bars or other similar places where beer is sold** (according to Foursquare categorization). But these were removed keeping only breweries and craft-beer shops.

We paired this information with Madrid´s City Council data related to age distribution and real-state prices.

In the methodology section, we will look for the most promising addresses in a 3km circle with a reasonable real-state prices and highest Young Index.

Young Index and Real State prices Chroplet Maps were overlapped to the positions of the identified craft-beer shops identified by Foursquare. Apart from the statistical information, these maps show us a quick situation of the positions of the shops related to these variables.

The second step in our analysis will be calculation and exploration of '**craft beer density**' across different areas of Madrid around in a radius of 3km around the city center. We will use **heatmaps** to identify a few promising areas close to the center where craft-beer shops are concentrated. Also we will paid attention to the Young Index of the living population on those areas and real-state places to identify the best suitable place to open such a shop. Pockets

were to open a new shop will be identified on those zones with low concentration or none shops at all.

In the final step we will focus on selecting two promising pocket areas (Regions Of Interest, ROI) and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no craft beef shops in radius of 500 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis

Moncloa and Arganzuela Districts as Candidates.

It can be seen that this kind of business is not as saturated as we thought, with offers for such a big city like Madrid. Borders of Moncloa and Arganzuela districts look as the most promising places in terms of low density of Craft Breweries combined with moderate real-state prices and mid-age population index so high probability of potential customers. However, there are zones in the border between Centro and Moncloa which are close to the Real Palace or Casa de Campo where there are not commercial places to open such kind of business.

In terms of **turist fluency** and business opportunity the **places near Princesa Street and Calle de Segovia** represent some interesting pocket zones where it could be attractive to open such a shop.

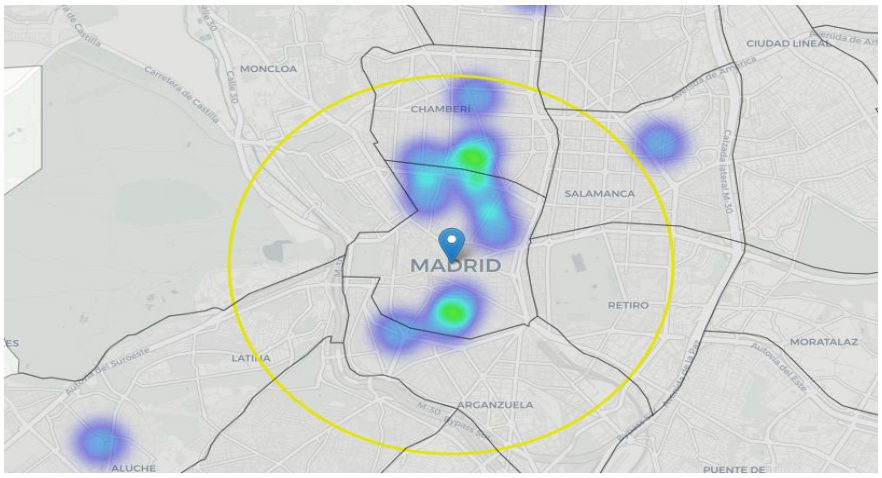

Fig 5. Craft Beer shops density overlapped to the boundaries of the districts

We focus our attention on getting places on those two Regions of Interest, trying to get addresses to open the shop by remembering the condition of not having a similar place in a radius of 500m.

We create the centers of both ROIs getting the coordinates from Geocoder for the addressed “Princesa Street nb 2” and “Calle Manzanares 5”, first inside Moncloa district and the second one inside Arganzuela, but very close to Moncloa limits.

We create points in around both ROI centers at a distance of 250m among them and inside a radius of 500m. These points need to meet the condition of being at a minimum distance of 500m from other craft-beer shops.

Those conditions will lead to the following map distributions for ROI 1 (“Moncloa”), ROI2 (“Arganzuela”).

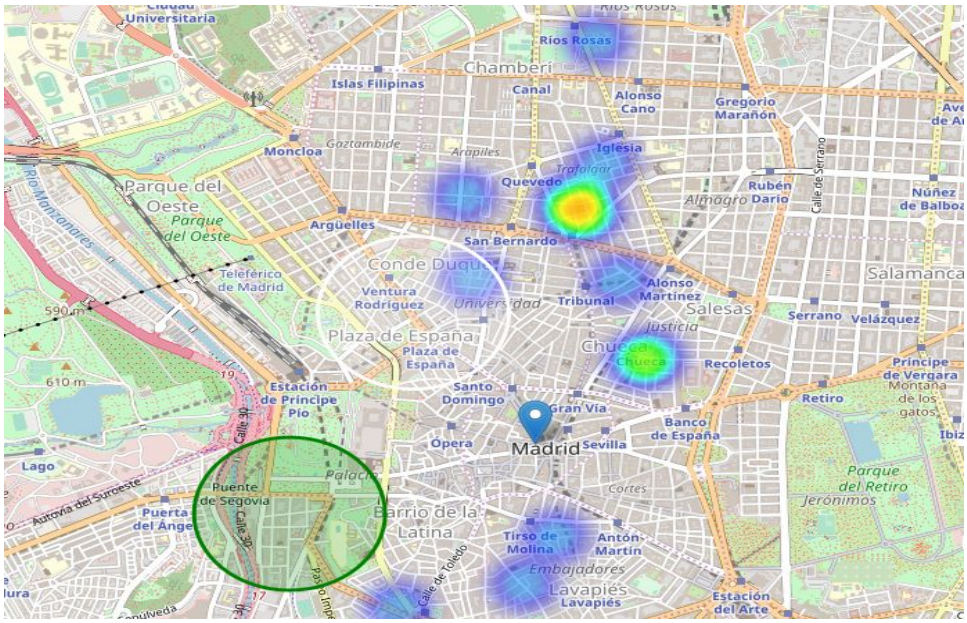

Fig 6. Centers for ROIs 1 (white) and 2 (green) with an influence radius of 500m defined around.

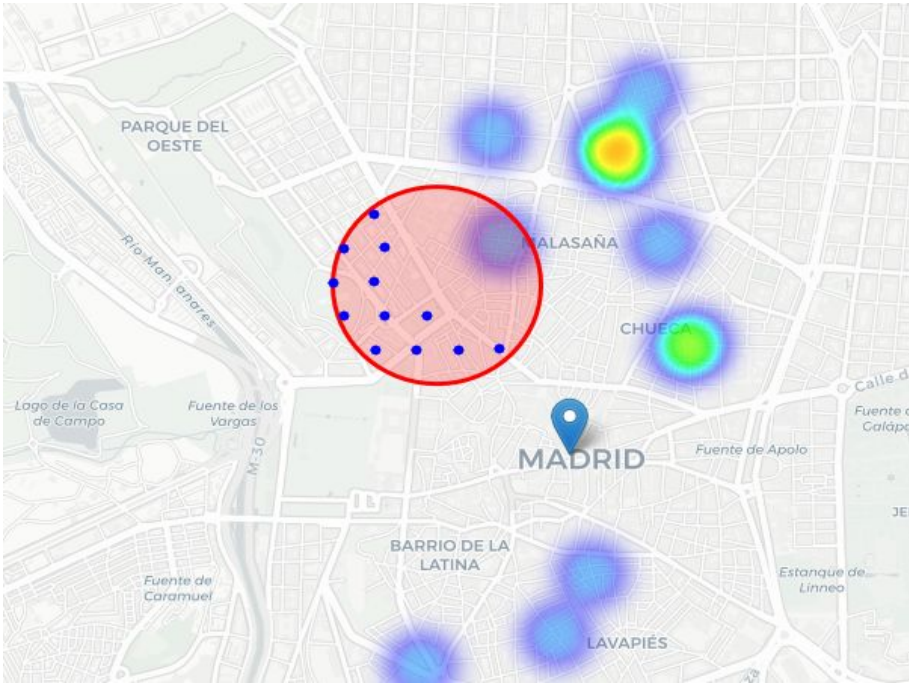

Fig 7. Selected points inside ROI 1 according to 500m distance limit from similar venues.

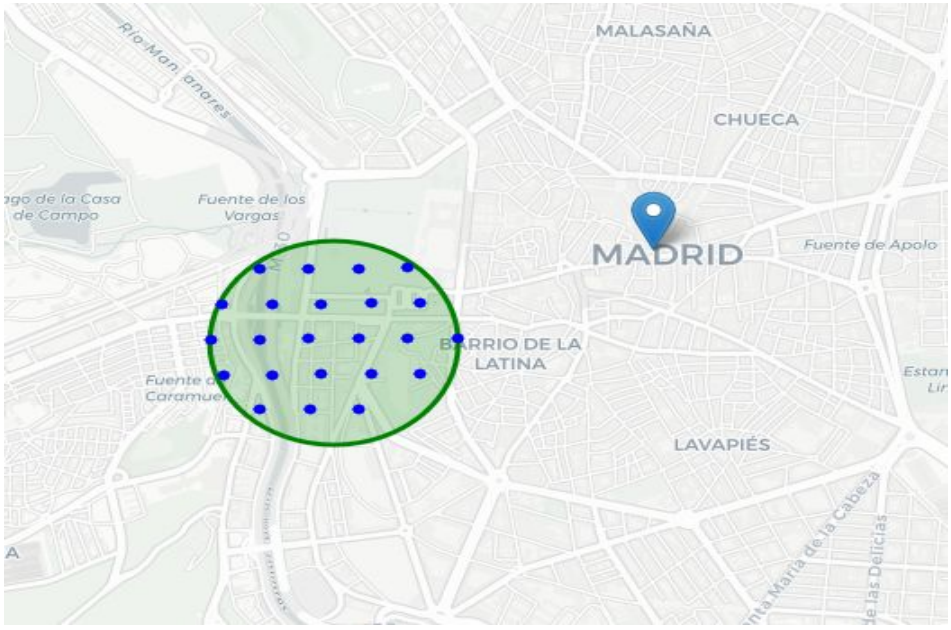

Fig 8. Selected points inside ROI 2 according to 500m distance limit from similar venues.

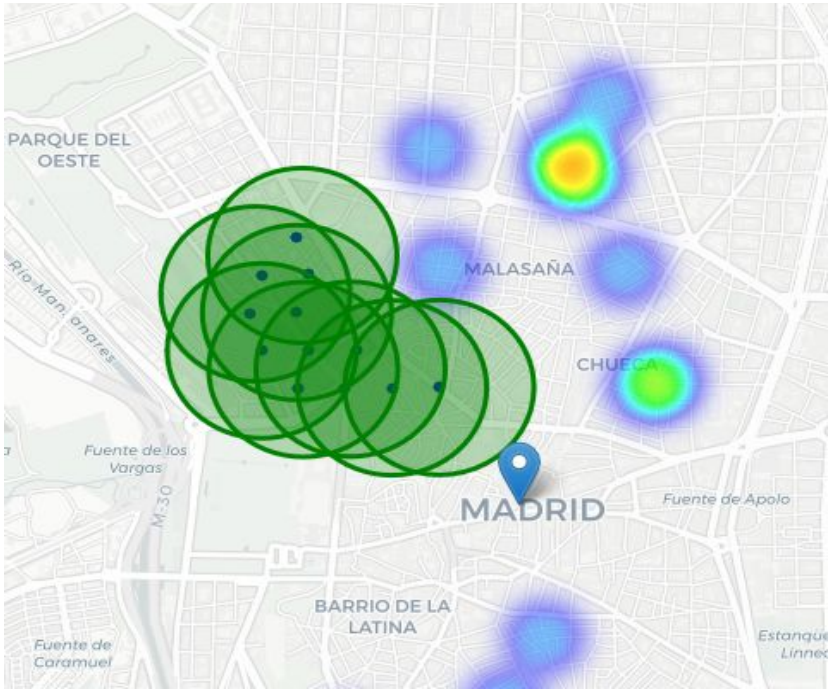

Fig 9. Cluster of points for ROI 1 with 500m limit overlapped.

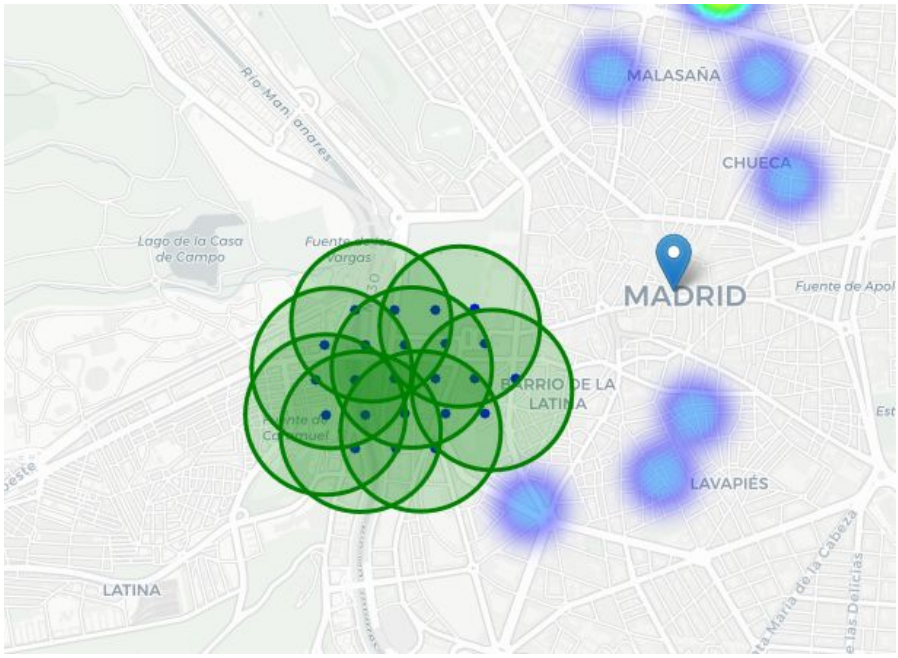

Fig 10. Cluster of points for ROI 2 with 500m limit overlapped.

5. Results and Discussion

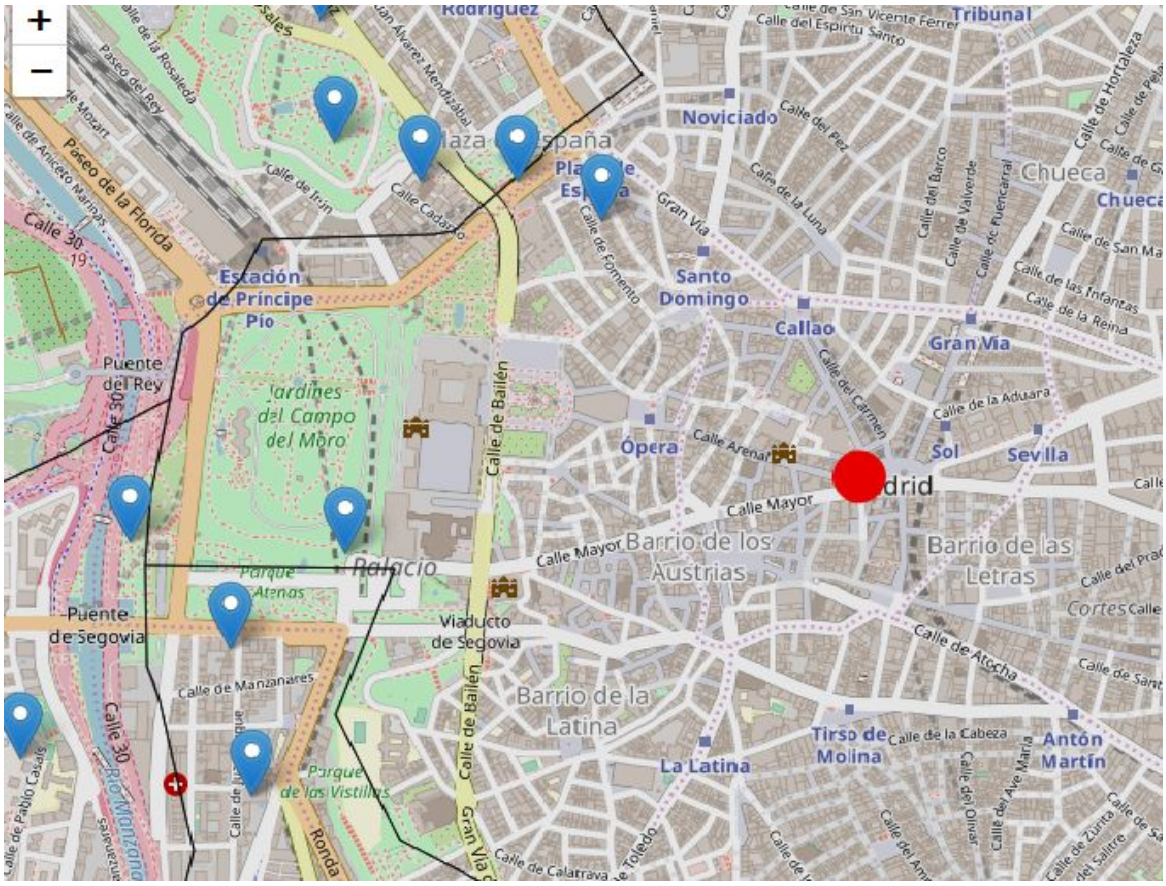

Fig 11. Final positions got by clustering around ROI 1 and ROI 2

This concludes our analysis. We have created 8 addresses per ROI representing centers of places which could be a potential location for a craft-beer shop. However, this study needs further refinement as some of the potential candidates lay-up over public garden addresses or restricted to open such a commercial place because of the lack of locals. Looking at the last map, In the borders of Moncloa district and around calle Manzanares y Arganzuela, arise as highly valuable candidates, because of the above imposed constraints in terms of Young Index and Real State prices. Moreover their good location and good public transport net, make this places good locations to be visited by tourists, then giving chances to be a high profitable place.

As can be seen after the final cleaning of our positions, there can be found suitable places to open craft beer shops in the zones close to the city center, in some reduced areas close on west Madrid close to tourist areas and were real state prices are below the mean. Latter on it

has to be identified if other craft-made products can be sold using this high-valued potential locations.¶

6. Conclusions and way forward

Purpose of this project was to identify Madrid areas with low density of craft beer shops. Surprisingly, the market is not saturated like bars or restaurants, so there exists a high potential for a profitable business of such characteristics.

If we put our attention on districts close to the center with high rate of tourists and whose real-state prices are down the mean, the opportunities to be profitable increase. Also a mid to high value for the Young Index population distribution will be a good indicator to open a shop in those districts as young people tend to drink more beer than old people.