Identification of real investment opportunities by crossed segmentation analysis of socio-economic development indicators and housing prices

Author: Francesco Fusaro

Date: April, 2021

Table of Contents

1 Introduction/Business problem

1.1 Background

Real estate investors are continuously looking for new areas in which to acquire properties. Obviously, the expectation is to ensure a good income from the rents generated by the property and to maximize the ratio between income derived from the rents and the initial property investment. From a real estate investor perspective, it is therefore interesting to identify at an early-stage, areas in which rents are expected to increase due to the area's positive economic-value development, while property prices do not yet (fully) reflect the development. As a side effect, due to its increase in market value over time, a property located in an area with such development prospects might also lead to higher benefits if the property is sold later on.

1.2 Business related key question

We are therefore trying to answer the following question of a real estate investor: "Which area of the city is expected to have higher ratios between the future income derived from renting housing space (rents) and the initial property investment?"

1.3 Approach/Value proposition

To answer this question, we will identify city areas for recommendation, whose current economic-value development indicates an increase of the rents level over time, but by today's comparison with area with similar economic-value, features lower property prices.

On one hand, we will try to assess the economic-value development of a city area by leveraging FourSquare location data to segment and compare the areas based on their most common venue categories according to popular places and venues recommended by FourSquare users. As further indication of area development we will consider the demographic evolution (population increase or decrease). On the other hand, current and historical property price and rents data will be used to segment the city areas in terms of property price level and rents. For more details on the data refer to Section 2.

By combining insights from economic-value and social development data with price and cost data we expect to be able to provide valuable data driven information to answer the question above.

We will prototype the approach on the neighborhoods of the city of Madrid, Spain

1.4 Business interest

Real estate investors and real estate agencies in Madrid would be equally interested in a product providing guidance for selecting areas in a city in order to have higher ratios between the future income derived from renting housing space (rents) and the initial property investment.

2 Data

2.1 Data sources

2.1.1 Geographical and demographic information

The city of Madrid is subdivided into 21 districts (distritos), which are subdivided into 131 neighborhoods (barrios administrativos) [1]. To characterize districts and neighborhood geographically and demographically the following information was retrieved from [1] and [2]:

- list of districts
- list of neighborhoods
- neighborhood surface area

Demographic evolution of each neighborhood was obtained from the statistical information database of the city of Madrid [3].

2.1.2 Prices of housing for sales and housing for rent

The City of Madrid's statistical database [3] contains actual and historical data about m2/€ housing selling prices and m2/€ housing rents for the last 5 years. However, whereas selling prices are available at the neighborhood level, rents are only reported at the district level.

The real estate web portal *idealista.com* has a report section which offers the possibility to consult Madrid area housing selling prices and rents at the neighborhood level [4],[5]. Quarterly variation, year on year variation, and all times maxima are also reported. The *idealista.com* data were used since we require data at the neighborhood level.

2.1.3 Madrid neighborhood location data (Places and venues in Madrid)

The `venues/explore` endpoint in the FOURSQUARE *Places API* was used to explore and retrieve recommended venues and their categories in each of Madrid's neighborhoods [6]. The search area was defined through the longitude/latitude of each neighborhood and a search radius of 700 m. (This corresponds to a surface area of 153 ha., which corresponds to the 3rd quartile of the distribution of the surface area of Madrid's neighborhoods).

The result of each was further processed to determine the frequency of each venue category in each of Madrid's neighborhoods (see section 2.2). The neighborhood's venue category frequency was used to try to segment neighborhoods in clusters of different economic-value development.

2.2 Data preparation

2.2.1 Geographical and demographic information

2.2.1.1 Madrid districts and neighborhoods subdivision

The name of Madrid's districts and their further subdivision in neighborhoods was obtained from [2]. A dataframe with the following columns was created (Figure 1):

- Neighborhood
- District
- Neighborhood surface area

	Neighborhood	District	Area (km2)
0	Palacio	Centro	1.471
1	Embajadores	Centro	1.032
2	Cortes	Centro	0.592
3	Justicia	Centro	0.742
4	Universidad	Centro	0.947
...	...	...	...
126	Alameda de Osuna	Barajas	1.961
127	Aeropuerto	Barajas	25.132
128	Casco Histórico de Barajas	Barajas	0.609
129	Timón	Barajas	9.595
130	Corralejos	Barajas	4.633

131 rows × 3 columns

Figure 1: mad_neighborhoods - Madrid's districts and neighborhoods subdivision. There are 21 districts which are subdivided in 131 neighborhoods.

2.2.1.2 Madrid's neighborhood demographic information

Demographic evolution data of each neighborhood in the years 2018, 2019, 2020 was downloaded from the statistical information database of the city of Madrid [3], namely:

- Neighborhood name
- Surface area
- Population density for the (years 2018-2020)
- Population (years 2018-2020)

The data were cleaned to ensure consistent data format. Moreover, two additional columns with the relative change of the population density with respect to the previous year were created:

- dDensity(2020)
- dDensity(2019)

The relative change for year *n* was calculated as:

$$dDensity(n) = \frac{(Density(n) - Density(n-1)}{Density(n-1)} \times 100 \qquad (2\text{-}1)$$

where *n* is the year 2020 or 2019.

The resulting dataset were merged with the one containing Madrid's districts and neighborhood subdivision obtained from Wikipedia [2], (see Section 2.2.1.1). (Note that prior to the merge some of the neighborhood names obtained from Wikipedia needed to be aligned to the ones from City of Madrid's database and the surface area from Wikipedia was dropped). A snapshot of the resulting dataframe is shown in Figure 2.

	Neighborhood	District	Area (Ha)	Density 2018 (Inh/Ha)	Population 2018 (Inh)	Density 2019 (Inh/Ha)	Population 2019 (Inh)	Density 2020 (Inh/Ha)	Population 2020 (Inh)	dDensity 2019 rel (%)	dDensity 2020 rel (%)
0	Palacio	Centro	146.99	153.17	22515	155.95	22923	160.51	23593	1.814977	2.924014
1	Embajadores	Centro	103.37	431.74	44630	437.82	45259	455.13	47048	1.408255	3.953680
2	Cortes	Centro	59.19	177.93	10531	177.13	10484	181.98	10771	-0.449615	2.738102
3	Justicia	Centro	73.94	224.20	16578	231.98	17153	243.72	18021	3.470116	5.060781
4	Universidad	Centro	94.80	325.91	30897	334.64	31725	352.50	33418	2.678654	5.337079
...	...	...	...	...	...	...	...	...	...	...	...
126	Alameda de Osuna	Barajas	197.03	98.69	19446	99.10	19526	100.59	19820	0.415442	1.503532
127	Aeropuerto	Barajas	2962.61	0.61	1794	0.62	1851	0.64	1900	1.639344	3.225806
128	Casco Histórico de Barajas	Barajas	54.94	133.53	7336	137.70	7565	139.84	7683	3.122894	1.554103
129	Timón	Barajas	509.45	23.06	11750	24.32	12388	25.23	12853	5.464007	3.741776
130	Corralejos	Barajas	468.25	16.04	7510	16.32	7642	16.56	7754	1.745636	1.470588

131 rows × 11 columns

Figure 2: mad_demographics – Evolution of demographic figures in the 131 neighborhoods of Madrid in the years 2018-2020.

2.2.2 Prices of housing for sales and housing for rent

Data for prices of housing for sales and housing for rent in the neighborhoods of Madrid were obtained from [4] and [5], respectively. Data were cleaned to ensure a consistent data format and data for housing sell price and housing rent prices were stored in two separate datasets.

2.2.2.1 Housing rent prices

Data for housing rent prices are shown in Figure 3 and are organized in the following columns:

- District
- Neigborhood
- Rent price (EUR/m2), February 2021
- Monthly rent variation (Monthly var rent (%))
- Quarterly price variation (Quarterly var rent (%))
- Yearly price variation (Yearly var rent (%))
- Historical maximum rent (Max rent EUR/m2)
- Variation with respect to maximum price (Max var rent (%))
- Year in which maximum price was achieved (Max rent year)

	District	Neighborhood	Rent price (EUR/m2)	Monthly var rent (%)	Quarterly var rent (%)	Yearly var rent (%)	Max rent (EUR/m2)	Max var rent (-)	Max rent year
0	Centro	Justicia	17.3	-1.0	-3.7	-14.4	20.9	-17.1	2020.0
1	Centro	Cortes	16.1	-0.7	-5.9	-13.7	19.6	-17.7	2018.0
2	Centro	Embajadores	15.9	-0.6	-3.6	-17.2	19.2	-17.2	2020.0
3	Centro	Universidad	16.7	0.0	-3.2	-16.7	20.0	-16.7	2020.0
4	Centro	Palacio	16.1	-0.9	-4.8	-11.6	18.2	-11.6	2020.0
...	...	...	...	...	...	...	...	...	...
125	Villaverde	Butarque	10.3	3.1	-1.9	-3.0	16.4	-37.0	2019.0
126	Villaverde	Ángeles	10.3	-4.4	-3.7	-4.0	11.4	-9.8	2020.0
127	Villaverde	Los Rosales	10.4	-0.4	-3.8	-5.3	11.8	-11.4	2020.0
128	Villaverde	Villaverde Alto, Casco Histórico de Villaverde	10.6	-1.4	-2.6	-2.6	11.8	-10.0	2019.0
129	Villaverde	San Cristóbal	NaN	NaN	NaN	NaN	NaN	NaN	NaN

130 rows × 9 columns

Figure 3: mad_rents_feb - Variation of prices of housing for rent in Madrid's neighborhoods.

2.2.2.2 Housing sale prices

Data for housing sell prices are shown in Figure 3 and are organized in the following columns:

- District
- Neigborhood
- Sale price (EUR/m2), February 2021
- Monthly sale price variation (Monthly var sale (%))
- Quarterly sale variation (Quarterly var sale (%))

- Yearly sale price variation (Yearly var sale (%))
- Historical maximum sale price (Max sale EUR/m2)
- Variation with respect to maximum price (Max var sale (%))
- Year in which maximum price was achieved (Max sale year)

	District	Neighborhood	Sale price (EUR/m2)	Monthly var sale (%)	Quarterly var sale (%)	Yearly var sale (%)	Max sale (EUR/m2)	Max var sale (-)	Max sale year
0	Centro	Justicia	5707.000	1.2	-1.7	-1.7	6120.0	-6.7	2019.0
1	Centro	Cortes	5229.000	1.5	3.0	-2.1	5481.0	-4.6	2018.0
2	Centro	Embajadores	4162.000	0.0	-1.7	-7.3	4489.0	-7.3	2020.0
3	Centro	Universidad	5051.000	0.3	-1.0	-4.4	5497.0	-8.1	2020.0
4	Centro	Palacio	4764.000	-0.1	0.8	1.8	5073.0	-6.1	2019.0
...	...	...	...	...	...	...	...	...	...
125	Villaverde	Butarque	2.092	0.4	-1.4	-2.0	2472.0	-15.4	2010.0
126	Villaverde	Ángeles	1.764	0.6	-0.7	-2.2	2487.0	-29.1	2010.0
127	Villaverde	Los Rosales	1.851	0.4	2.2	1.5	2299.0	-19.5	2010.0
128	Villaverde	Villaverde Alto, Casco Histórico de Villaverde	1.626	1.5	0.7	0.0	2691.0	-39.6	2009.0
129	Villaverde	San Cristóbal	1.308	-1.2	-3.4	-10.0	1727.0	-24.3	2010.0

130 rows × 9 columns

Figure 4: mad_sale_feb - Variation of prices of housing for sale in Madrid's neighborhoods.

2.2.2.3 Handling of missing housing sale and housing rent data

For few neighborhoods, some housing sale price data and housing rent data entries were not available in [4] and [5], respectively and therefore needed to be estimated from the available data. The corresponding estimation methods are summarized in Table 1 and Table 2.

Neighborhood	**District**	**Missing rent data**	**Missing rent data estimation assumption**
Atocha	Arganzuela	All housing rent data	District median
Fuentelarreina	Fuencarral-El Pardo	Quarterly and yearly variation	Same as monthly variation
Cuatro Vientos	Latina	All housing rent data	Same as Campamento (Latina)
El Plantío	Moncloa-Aravaca	All housing rent data	Same as Valdezarza (Moncloa-Aravaca)
Orcasitas	Usera	quarterly and yearly variation	Same as monthly variation
Orcasur	Usera	Yearly variation	Same as maximum variation
San Fermín	Usera	Yearly variation	Same as maximum variation
Pradolongo	Usera	Yearly variation	Same as maximum variation
Fontarrón	Moratalaz	Yearly variation	Same as maximum variation
Vinateros	Moratalaz	Yearly variation	Same as maximum variation
Pavones	Moratalaz	All housing rent data	Same as Fontarrón (Moratalaz)

Horcajo	Moratalaz	All housing rent data	Same as Marroquina (Moratalaz)
Apóstol Santiago	Hortaleza	Yearly variation	Same as maximum variation
San Cristóbal	Villaverde	All housing rent data	Same as Villaverde Alto
Santa Eugenia	Villa de Vallecas	All housing rent data	Same as maximum variation
El Cañaveral	Vicálvaro	All housing rent data	Same as Valdebernardo (Vicálvaro)
Hellín	San Blas-Canillejas	All housing rent data	Same as Amposta (San Blas-Canillejas)
Amposta	San Blas-Canillejas	Yearly variation	Same as maximum variation
Arcos	San Blas-Canillejas	Yearly variation	Same as maximum variation
Atalaya	Ciudad Lineal	Quarterly variation	Same as monthly one

Table 1: Summary of missing housing rent data and corresponding assumptions for their estimation.

Neighborhood	District	Missing sale data	Missing sale data estimation assumption
Atocha	Arganzuela	District median	District median
Cuatro Vientos	Latina	Same as Campamento (Latina)	Same as Campamento (Latina)
Pavones	Moratalaz	Same as Fontarrón (Moratalaz)	Same as Fontarrón (Moratalaz)
Horcajo	Moratalaz	Same as Marroquina (Moratalaz)	Same as Marroquina (Moratalaz)
Atalaya	Ciudad Lineal	All housing sale data	District median

Table 2: Summary of missing housing sale data and corresponding assumptions for their estimation.

2.2.3 Consolidation of neighborhood demographic data with neighborhood prices data

The housing price and rent datasets presented in Sections 2.2.2.1 - 2.2.2.3 were merged with the neighborhood demographic dataset from Section 2.2.1.2.

Note that the housing for sale and housing for rent price data [4],[5] does not fully follow the official neighborhood boundaries. On one hand, few neighborhoods are missing, as they are not relevant from a real estate point of view (e.g., the neighborhood “Aeropuerto” whose boundaries are drawn around Madrid’s international airport). These neighborhoods were dropped from the consolidated demographic and prices data. On the other hand, few areas interest have separate entries although

officially they are part of one of the 131 neighborhoods. Those areas were added to the consolidated prices and demographic data.

A summary of the adjustments done to consolidate prices data and demographic data into one dataset are summarized in Table 3 and Table 4. The latter also reports how the surface area and population densities for the years 2018-2020 for the added areas was estimated. (The missing absolute population data was obtained by multiplication of the surface area and the population density).

A snapshot of the consolidated prices and demographic dataset is shown in Figure 5. The dataset contains 134 neighborhoods and areas.

Neighborhood dropped from consolidated demographic and prices data		
Name	***District***	***Reason***
El Pardo	Fuencarral-El Pardo	Park area, no urban area
El Goloso	Fuencarral-El Pardo	Military base, no urban area
Aeropuerto	Barajas	Airport, no urban area
Casco Histórico de Vicálvaro	Vicálvaro	Mainly non urban area; urban area covered by Ambroz

Table 3: Neighborhoods dropped from consolidated Madrid demographic and pricing data

Areas added to consolidated demographic and prices data			
Area Name	***Neighborhood***	***Surface Area***	***Population Densities***
Pau de Carabanchel	Cuatro Vientos (Latina) and Buenavista (Carabanchel)	See [7]	Same as Cuatro Vientos
Arroyo del Fresno	Mirasierra	See [8]	Same as Mirasierra
Las Tablas	Valverde	Median of all neighborhoods	Same as Valverde
Montecarmelo	El Goloso, Mirasierra	Median of all neighborhoods	Take Mirasierra, as El Goloso is a military base
Sanchinarro	Valdefuentes	Median of all neighborhoods	Take Valdefuentes
Virgen del Cortijo - Manoteras	Valdefuentes	Median of all neighborhoods	Take Valdefuentes
Ambroz	Casco Histórico de Vicálvaro	Median of all neighborhoods	Take Valderrivas

Table 4: Neighborhoods added to consolidated Madrid demographic and pricing data.

	Neighborhood	District	Area (Ha)	Density 2018 (Inh/Ha)	Population 2018 (Inh)	Density 2019 (Inh/Ha)	Population 2019 (Inh)	Density 2020 (Inh/Ha)	Population 2020 (Inh)	dDensity 2019 rel (%)	...	Max rent (EUR/m2)	Max var rent (-)	Max rent year	Sale price (EUR/m2)	Monthly var sale (%)	Quarterly var sale (%)	Yearly var sale (%)	Max sale (EUR/m2)	Max var sale (-)	Max sale year
0	Palacio	Centro	146.99	153.17	22515.0000	155.95	22923.0000	160.51	23593.0000	1.814977	...	18.2	-11.6	2020	4764.0	-0.1	0.8	1.8	5073.0	-6.1	2019
1	Embajadores	Centro	103.37	431.74	44630.0000	437.82	45259.0000	455.13	47048.0000	1.408255	...	19.2	-17.2	2020	4162.0	0.0	-1.7	-7.3	4489.0	-7.3	2020
2	Cortes	Centro	59.19	177.93	10531.0000	177.13	10484.0000	181.98	10771.0000	-0.449615	...	19.6	-17.7	2018	5229.0	1.5	3.0	-2.1	5481.0	-4.6	2018
3	Justicia	Centro	73.94	224.20	16578.0000	231.98	17153.0000	243.72	18021.0000	3.470116	...	20.9	-17.1	2020	5707.0	1.2	-1.7	-1.7	6120.0	-6.7	2019
4	Universidad	Centro	94.80	325.91	30897.0000	334.64	31725.0000	352.50	33418.0000	2.678654	...	20.0	-16.7	2020	5051.0	0.3	-1.0	-4.4	5497.0	-8.1	2020
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
129	Las Tablas	Fuencarral-El Pardo	135.64	68.97	9355.0908	70.37	9544.9868	72.00	9766.0800	2.029868	...	13.2	-9.0	2020	4130.0	0.0	0.0	-0.9	4324.0	-4.5	2019
130	Montecarmelo	Fuencarral-El Pardo	135.64	45.84	6217.7376	47.05	6381.8620	48.69	6604.3116	2.639616	...	13.9	-10.4	2018	4373.0	0.4	-0.7	-4.6	4630.0	-5.6	2019
131	Sanchinarro	Hortaleza	135.64	32.40	4394.7360	34.54	4685.0056	36.96	5013.2544	6.604938	...	13.5	-7.6	2019	4286.0	0.7	2.3	-3.5	4538.0	-5.6	2019
132	Virgen del Cortijo - Manoteras	Hortaleza	135.64	32.40	4394.7360	34.54	4685.0056	36.96	5013.2544	6.604938	...	14.1	-12.6	2019	3678.0	-0.4	-4.5	9.2	3914.0	-6.0	2018
133	Ambroz	Vicálvaro	135.64	282.66	38340.0024	282.07	38259.9748	282.20	38277.6080	-0.208731	...	13.0	-6.3	2019	2068.0	3.3	2.3	-1.5	2468.0	-16.2	2011

134 rows × 25 columns

Figure 5: mad_demo_prices - Consolidated prices and demographic dataset for Madrid's neighborhoods and areas.

2.2.4 Madrid neighborhood location data (Places and venues in Madrid)

In order to determine the frequency of the most recommended venue categories in each of Madrid's neighborhoods the following steps were applied:

1. Determine latitude and longitude data for each neighborhood/area in the consolidated demographic prices dataset using *Geopy* (*https://geopy.readthedocs.io/en/stable/*) package in conjuction with *Nominatim* geocoder[1]
2. Use *FourSquare Places* API `explore` endpoint to retrieve top recommended venues in each neighborhood/area by specifying the corresponding coordinates [6]. The venues were searched in a radius of 700 m and a limit of 100 venues per neighborhood was set.
3. Create a dataset containing the retrieved venues, their categories and their neighborhoods
4. Convert venue categories from categorical variables (e.g., 'Restaurant', 'Shop') in indicator variables using the `pandas.get_dummies()` method
5. Use the `pandas.groupby().sum()` method to group by neighborhood and sum the occurrences of each venue category in each of the neighborhoods. For normalization divide each entry by the total number of unique venue categories

A snapshot of the dataset with the frequency of the venue category in each of Madrid's neighborhood is shown in Figure 6. The dataset contains 280 venue categories and 133 neighborhoods/areas (one less than the demographic data and prices dataset, as for the 'El Cañaveral" neighborhood no venue recommendation was found in the specified radius).

	Neighborhood	Accessories Store	Airport	Airport Terminal	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	...	Used Bookstore	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Vietnamese Restaurant	Whisky Bar	Wine Bar	Wine Shop	Women's Store	Yoga Studio
0	Abrantes	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	Acacias	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.014286	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.003571	0.0	0.0
2	Adelfas	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	Alameda de Osuna	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	Almagro	0.0	0.0	0.0	0.003571	0.0	0.0	0.0	0.003571	0.003571	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
128	Virgen del Cortijo - Manoteras	0.0	0.0	0.0	0.003571	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
129	Vista Alegre	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
130	Zofío	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
131	Águilas	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
132	Ángeles	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

133 rows × 281 columns

Figure 6: mad_onehot_grouped – Frequency of venue categories in Madrid's neighborhoods according to venues recommended by FourSquare users

Finally, by sorting each row of the venue category frequency dataset, a dataset containing the top venue category in each neighborhood can obtained (see

[1] Plotting the locations on a map using the longitude and latitude retried from *Nominatim*, showed that in few cases the returned location was not within the City of Madrid. The coordinates of these locations were obtained from http://www.google.com/maps.

	Neighborhood	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	...	11th Most Common Venue	12th Most Common Venue
0	Abrantes	Pizza Place	Soccer Field	Bakery	Gym / Fitness Center	Fast Food Restaurant	Athletics & Sports	Restaurant	Yoga Studio	Fish & Chips Shop	...	Farm	Farmers Market
1	Acacias	Spanish Restaurant	Bar	Art Gallery	Tapas Restaurant	Coffee Shop	Pizza Place	Theater	Restaurant	Market	...	Pub	Café
2	Adelfas	Bar	Café	Grocery Store	Fast Food Restaurant	Spanish Restaurant	Bakery	Supermarket	Hotel	Gym	...	Sandwich Place	Farmers Market
3	Alameda de Osuna	Spanish Restaurant	Plaza	Bakery	Gym	Restaurant	Smoke Shop	Hotel	Bistro	Bar	...	Fried Chicken Joint	Scenic Lookout
4	Almagro	Restaurant	Spanish Restaurant	Plaza	Bar	Hotel	Japanese Restaurant	Coffee Shop	Mediterranean Restaurant	Italian Restaurant	...	Lounge	French Restaurant
...	...	...	...	...	...	...	...	...	...	...	...	...	...
128	Virgen del Cortijo – Manoteras	Spanish Restaurant	Burger Joint	Cosmetics Shop	Gastropub	Sushi Restaurant	Cafeteria	Restaurant	Bar	Mediterranean Restaurant	...	Train Station	Park
129	Vista Alegre	Grocery Store	Fast Food Restaurant	Coffee Shop	Plaza	Bakery	Pizza Place	Pub	Convenience Store	Comedy Club	...	Spanish Restaurant	Breakfast Spot
130	Zofío	Spanish Restaurant	Park	Theater	Beer Garden	Athletics & Sports	Asian Restaurant	Fish Market	Farm	Farmers Market	...	Fish & Chips Shop	Yoga Studio
131	Águilas	Bar	Spanish Restaurant	Shopping Mall	Restaurant	Gym Pool	Smoke Shop	Café	Tapas Restaurant	Convenience Store	...	Train Station	Seafood Restaurant
132	Ángeles	Bar	Restaurant	Spanish Restaurant	Grocery Store	Pet Store	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	...	Fish Market	Exhibit

133 rows × 21 columns

Figure 7: mad_venues_sorted – Most common venues categories by Madrid's neighborhood according to FourSquare users.

2.3 Feature Selection for segmentation into clusters

2.3.1 Madrid neighborhood location data (Places and venues in Madrid)

To segment Madrid's neighborhoods into clusters according to their venues and places, the dataset with the frequency of venue categories in Madrid's neighborhoods (see Section 2.2.4) was used as is.

2.3.2 Madrid neighborhood demographic and prices data

To segment Madrid's neighborhoods into clusters according to prices and demographic development the following features were selected from the demographic data and prices dataset (see Section 2.2.3):

- Population density for the year 2018
- dDensity(2020)
- dDensity(2019)
- Rent price (EUR/m2), February 2021
- Quarterly price variation (Quarterly var rent (%))
- Yearly price variation (Yearly var rent (%))
- Historical maximum rent (Max rent EUR/m2)
- Variation with respect to maximum price (Max var rent (%))
- Year in which maximum price was achieved (Max rent year)
- Sale price (EUR/m2), February 2021
- Quarterly sale price variation (Quarterly var sale (%))
- Yearly sale price variation (Yearly var sale (%))
- Historical maximum sale price (Max sale EUR/m2)
- Variation with respect to maximum price (Max var sale (%))
- Year in which maximum price was achieved (Max sale year)

3 References

[1] Anexo:Distritos de Madrid. In *Wikipedia*. Retrieved March 18, 2021, from https://es.wikipedia.org/wiki/Anexo:Distritos_de_Madrid

[2] Anexo:Barrios administrativos de Madrid. In *Wikipedia*. Retrieved March 18, 2021, from https://es.wikipedia.org/wiki/Anexo:Barrios_administrativos_de_Madrid

[3] http://www-2.munimadrid.es/CSE6/control/menuCSE

[4] Price evolution of housing for sale in Spain. In *idealista.com*. Retrieved March 18, 2021 from https://www.idealista.com/en/press-room/property-price-reports/

[5] Price evolution of housing for rent in Spain. In *idealista.com*. Retrieved March 18, 2021 from https://www.idealista.com/en/press-room/property-price-reports/rent/

[6] Venues Explore. In FourSquare/developers. Retrieved March 18, 2021, from https://developer.foursquare.com/docs/api-reference/venues/explore/

[7] La Peseta (Madrid). In *Wikipedia*. Retrieved March 18, 2021, from https://es.wikipedia.org/wiki/La_Peseta_(Madrid)

[8] Arroyo del Fresno. Retrieved March 18, 2021, from https://www.idealista.com/news/inmobiliario/vivienda/2011/11/23/365358-asi-es-arroyo-del-fresno-el-pau-mas-desconocido-y-privilegiado-de-madrid-video