

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

Francesco Fusaro

April, 2021

Content

- **Introduction/Business problem**
- **Data**
- **Methodology**
- **Results and Discussion**
- **Summary**

Introduction/Business problem

- Key question:
 - “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?”
- Approach/Value proposition:
 - Combination of segmentation of city’s neighborhood by socio-economic development (represented by neighborhood’s venue recommendation retrieved from FourSquare) with housing and demographic development data
- Business interest
 - Real estate investors and real estate agencies

Data

- Sources:
 - Demographic data/neighborhoods:
<http://www-2.munimadrid.es/CSE6/control/menuCSE>
 - Pricing data:
 - 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/>
 - 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/>
 - Location Data:
 - Venues Explore. In FourSquare/developers. Retrieved March 18, 2021, from

Data

- Housing prices and demographic development in Madrid

	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 x 25 columns

Consolidated data covering actual rent and sale price, historical development, population development for 134 areas and neighborhoods

Data

- Madrid venues data

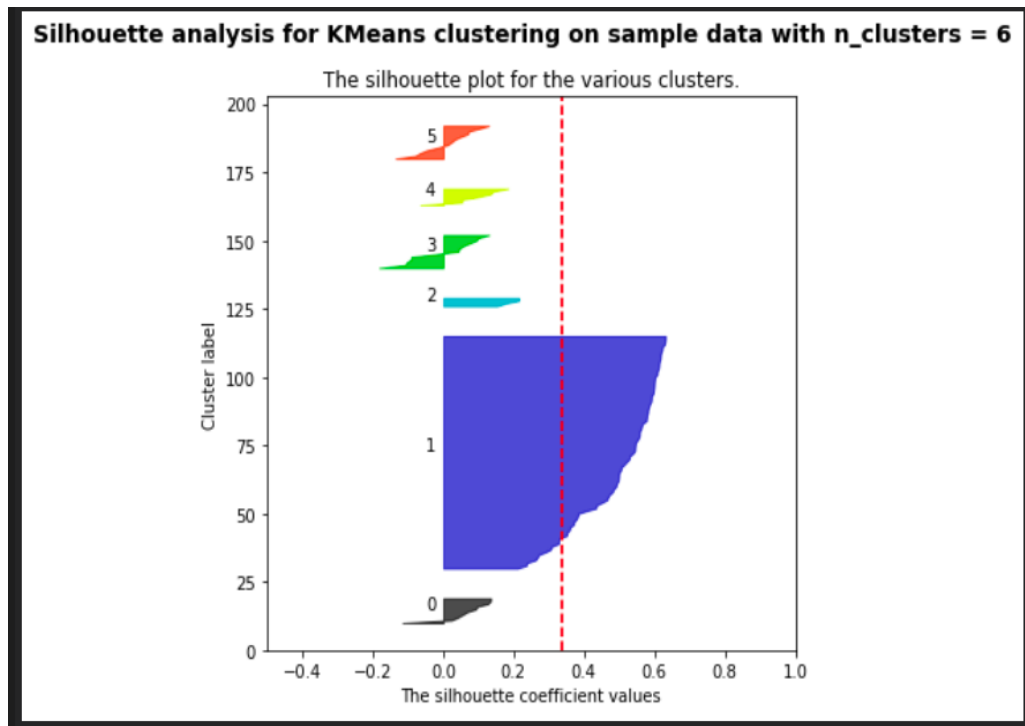
	Neighborhood	Accessories Store	Airport	Airport Terminal	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	...	Used Bookstore	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Vietnamese Restaurant	Whisky Bar	Wine Bar	Wine Shop	Women's Store	Yoga 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	Zofio	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 x 281 columns

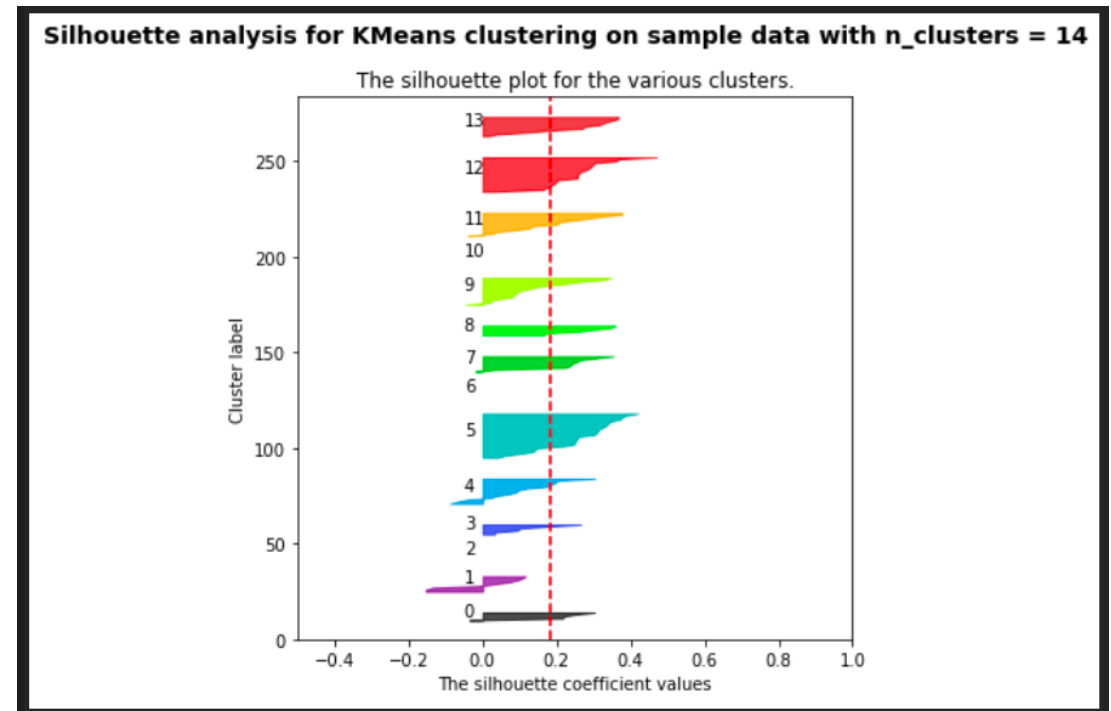
Venue category frequencies in 133 neighborhoods and areas across 280 different categories and ~5000 different entries

Methodology

- Automating grouping into clusters using K-means algorithm



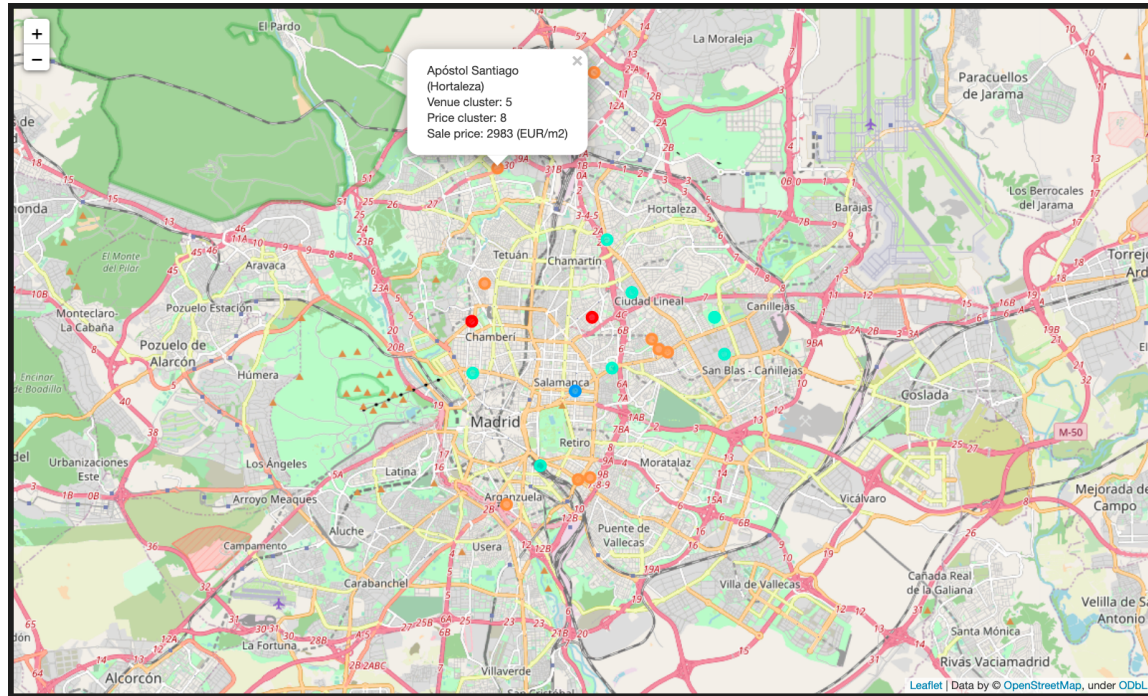
Venues: 6 clusters



Housing prices: 14 clusters

Results and Discussion:

Combined segmentation analysis



Neighborhoods with lower price range in given venue cluster						
	Neighborhood	District	Label_venue	Label_price	Sale price (EUR/m2)	Rent price (EUR/m2)
95	Prosperidad	Chamartín	0	11	4228.0	14.40
123	Vallehermoso	Chamberí	0	4	4953.0	15.20
54	Goya	Salamanca	2	5	5599.0	16.50
35	Comillas	Carabanchel	3	12	2565.0	12.20
112	Simancas	San Blas-Canillejas	3	12	2633.0	11.80
125	Ventas	Ciudad Lineal	3	12	2663.0	12.00
44	El Salvador	San Blas-Canillejas	3	7	3311.0	12.20
17	Atocha	Arganzuela	3	4	3915.0	14.15
16	Atalaya	Ciudad Lineal	3	11	4132.5	14.40
108	San Juan Bautista	Ciudad Lineal	3	11	4151.0	13.30
96	Pueblo Nuevo	Ciudad Lineal	5	12	2506.0	12.10
99	Quintana	Ciudad Lineal	5	12	2909.0	12.50
36	Concepción	Ciudad Lineal	5	9	2942.0	13.10
10	Apóstol Santiago	Hortaleza	5	8	2983.0	10.80
18	Bellas Vistas	Tetuán	5	4	3440.0	14.30
31	Chopera	Arganzuela	5	4	3641.0	14.00
81	Pacífico	Retiro	5	11	3976.0	14.20
2	Adelfas	Retiro	5	4	4077.0	13.80
63	Las Tablas	Fuencarral-El Pardo	5	11	4130.0	12.00

Identification of 19 neighborhoods as opportunities for investment.
Opportunities need to be confirmed on case by case basis due to sale price overlaps in price clusters.

Conlusions

- Combination of venue segmentation with prices segmentation has led to the identification of 19 neighborhoods (out of more than 130) as potential investment opportunities for further screening.
- The underlying segmentation by venue categories was able to correctly group neighborhoods agreeing with Madrid's situation. The segmentation quality might be improved by reducing the granularity of the original venue categories retrieved from FourSquare
- The segmentation by prices was able to cluster neighborhoods that appear consistent with Madrid's actual real estate market situation. However overlapping clusters in terms of sale and price range are observed. It remains to be investigated how the selection of lower number of features and/or clusters would impact the results.