Coursera Capstone

Identifying real estate opportunities in Madrid, Spain

Juan Carlos Martinez **May 2020**

Business Problem

- Job market, environmental and health conditions are changing the game of real estate
- The objective of this capstone is to find underpriced neighborhoods in Madrid
- Which areas could show an opportunity with the latest global changes?

Data

- Data required:
  - List of neighborhoods in Madrid. This defines the scope of this project which is confined to the city of Madrid, the capital city of the country of Spain
  - Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data
  - Venue data for the different neighborhoods, in order to see the similitude of the different areas
  - Housing prices for the different neighborhoods. We will use the price per square meter to compare economic situation of the real state sector in the different areas
  - Neighborhoods geometry and position, to be able to divide the city in its different parts

.

Data

- Source:
  - Wikipedia, https://es.wikipedia.org/wiki/Anexo:Barrios_administrativos_de_Madrid
  - Geocoder package for latitude and longitude coordinates
  - Foursquare API for venue data
  - Idealista, https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/venta/madrid-comunidad/madrid-provincia/madrid/
  - Madrid statistics web, https://datos.madrid.es

.

Methodology

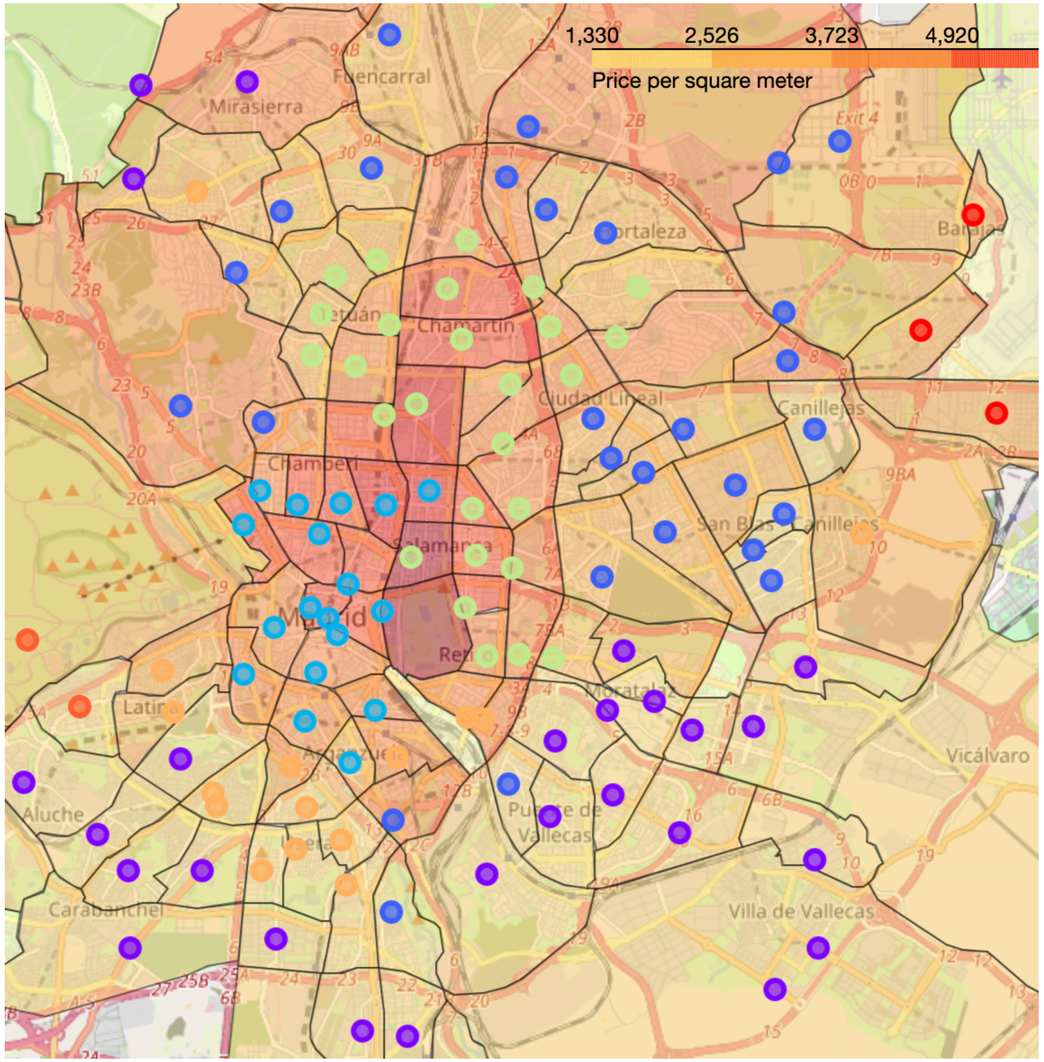
- Web scraping Wikipedia page for neighborhoods list
- Web scraping Idealista page for neighborhoods prices
- Web scraping Madrid statistics web page for neighborhoods geometries
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by neighborhood and get the frequency for each venue type
- Cluster data by k-means algorithm and elbow method
- Visualize neighborhoods and prices in Folium
- Analyze data to obtain market opportunities

.

Results

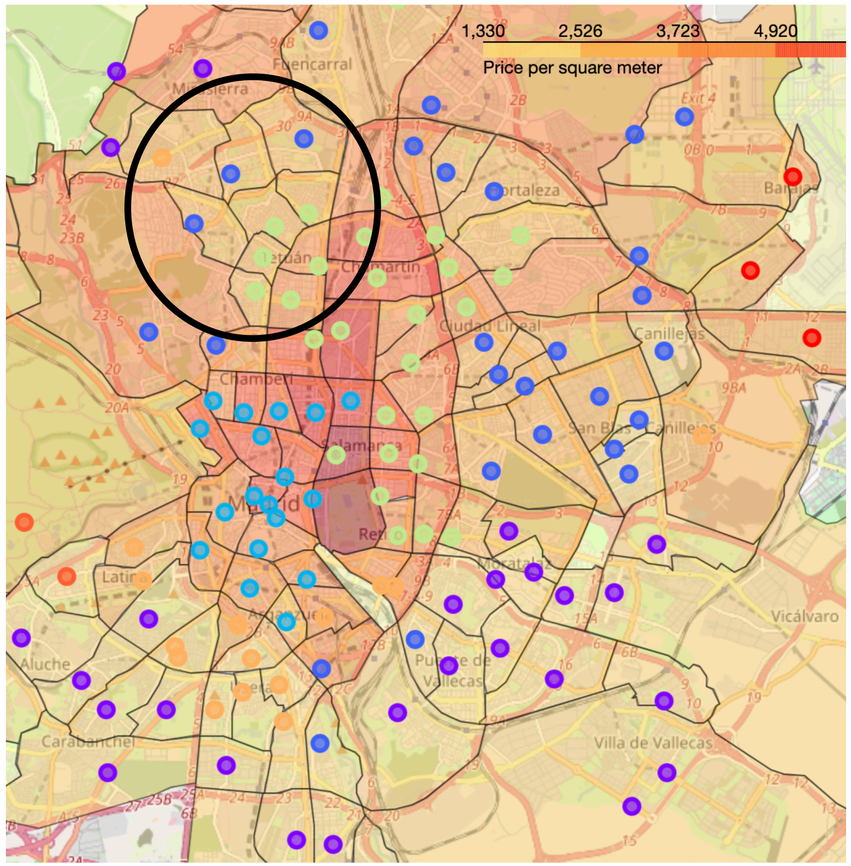
- Elbow method determined Kc=9

- Cluster 0: ‘Airport’ neighborhoods (Red in the map)
- Cluster 1: Outdoor South neighborhoods. Residential areas with a lot of grocery stores, supermarkets and bars. (Purple in the map)
- Cluster 2: Outdoor North neighborhoods. Residential areas with supermarkets, Spanish restaurants and green areas (Dark blue in the map)
- Cluster 3: Center neighborhoods. Area with a lot of hotels, restaurants and squares. (Light blue in the map)
- Cluster 4: Only one neighborhood: El Cañaveral - Los Berrocales. The algorithm made the cluster considering a lot of venues related to tolls and road businesses. (Turquoise in the map)
- Cluster 5: Only one neighborhood: Casco Histórico de Vicálvaro. The algorithm made the cluster considering a lot of venues related to tolls and the presence of the zoo. (Tourquoise-green in the map)
- Cluster 6: North neighborhoods. Area with a lot of different cuisine restaurants, cafes, and boutiques. (Light green in the map)
- Cluster 7: South neighborhoods: Area with a lot of parks, grocery stores and Spanish restaurants. (Orange in the map)
- Cluster 8: ‘Theme park neighborhoods’. Main venues related to the theme park. (Dark orange in the map)



Results

- Main opportunities found in the northwest area. Neighborhoods with prices lower than the districts surrounding them:



	Neighborhood	Cluster Labels	Price	Max	Max_period	Year change	Cluster mean price	Price/Max
2	Tres Olivos - Valverde	2	2698	2895	dic 2010	0.035	3103.000000	0.931952
3	Apóstol Santiago	2	2890	2963	mar 2020	0.008	3103.000000	0.975363
4	Pinar del Rey	2	2863	3238	nov 2010	-0.047	3103.000000	0.884188
5	Valdezarza	2	3090	3448	mar 2019	-0.075	3103.000000	0.896172
6	Las Tablas	6	4192	4324	jul 2019	0.006	4478.103448	0.969473
7	Canillas	6	3242	3692	nov 2009	0.005	4478.103448	0.878115
8	Bellas Vistas	6	3513	3688	may 2019	-0.024	4478.103448	0.952549
9	Berruguete	6	3253	3330	nov 2009	0.022	4478.103448	0.976877
10	Valdeacederas	6	3252	3541	oct 2008	0.027	4478.103448	0.918385
11	Ventilla-Almenara	6	3383	3494	feb 2020	-0.014	4478.103448	0.968231

Discussion

- Clusters clearly divides Madrid in 4 main areas: North/South and center/ surroundings.
- Opportunities found in the northwest area of Madrid
- Key metrics identified to select best neighborhood depending on the investors needs:
    - Max_period
    - Year change

.

Conclusion

- Analysis of the city was performed showing the different areas in terms of venues types
- 10 Neighborhoods found subject of investment
- Future line of research: Input more variables in the model and properties data to find the best market opportunity.

.

Thank you


thank you
danke
謝謝
ngiyabonga
рахмат
teşekkür ederim
спасибо
gracias
tapadh leat
Баярлалаа
dank je
misaotra
matondo
paldies
grazzi
mahalo
faafetai lava
vinaka
спасибі
blagodaram
mersi
barka
welalin
tack
spas
kia ora
kiitos
dankie
dhanyavad
nanni
nandri
хвала
asante
manana
obrigada
tenki
bedankt
hvala
mauruuru
köszönöm
akun
dankon
ačiū
bayarlalaa
grăcie
enkosi
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
murakoze
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
taiku
sukriya
kop khun krap
sagolun
sobodi
dēkuji
obrigado
mèsi
ありがとう
arigatô
takk
dakujem
trugarez
grazie
tanemirt
najis tuke
didi madloba
kam sah hamnida
rahmat
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
мерси
merci
ευχαριστώ
তোমাকে ধন্যবাদ
감사합니다
xièxie