IBM Data Science

Applied Data Science Project

Peer-graded Assignment: Capstone Project

The Battle of Neighborhoods (Week 2)

Vasileios Fafoutis

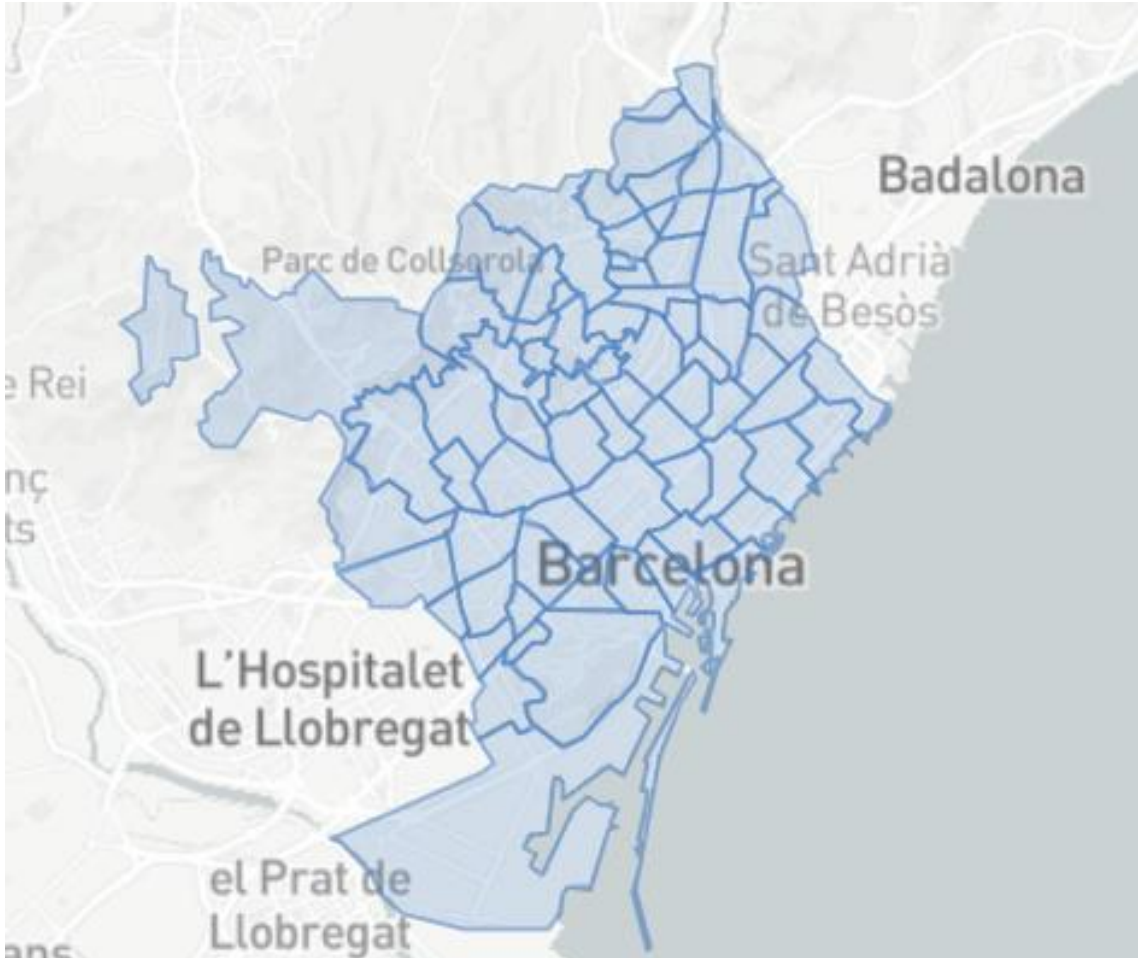
13.12.2020

Index

Scope and data sources of the analysis

Objective:

To cluster Barcelona city (Spain) 73 neighborhoods according to various data (mean rent price, surface, population density, most common venues).



Data sources:

Foursquare / Developer, Barcelona city hall open data, Police Force of Catalunya, Wikipedia, Github (geojson)

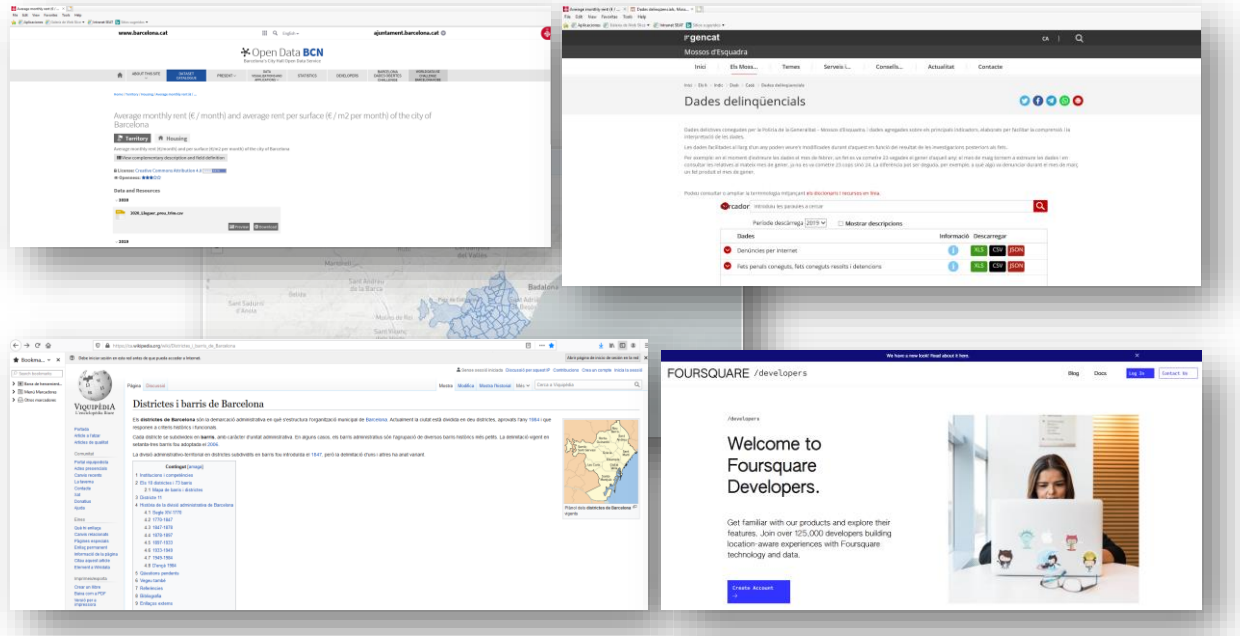
https://developer.foursquare.com/developer/

https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-mercat-immobiliari-lloguer-mitja-mensual

https://mossos.gencat.cat/ca/els_mossos_desquadra/indicadors_i_qualitat/dades_obertes/cataleg_dades_obertes/dades-delinqueencials/

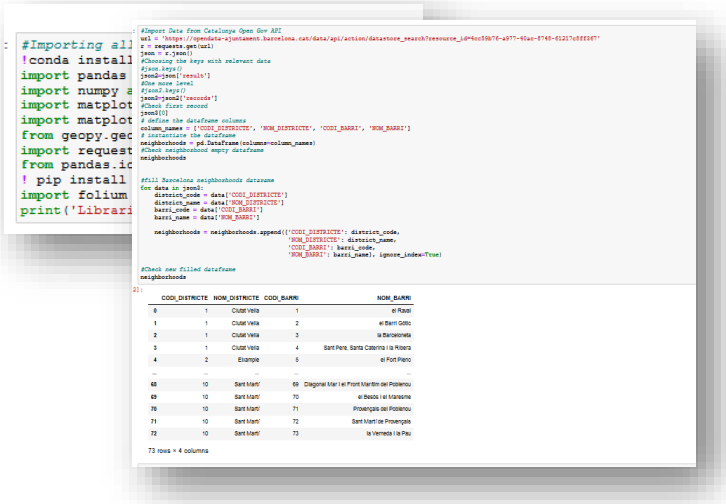
https://ca.wikipedia.org/wiki/Districtes_i_barris_de_Barcelona

https://github.com/martgnz/bcn-geodata/blob/master/barris/barris.geojson

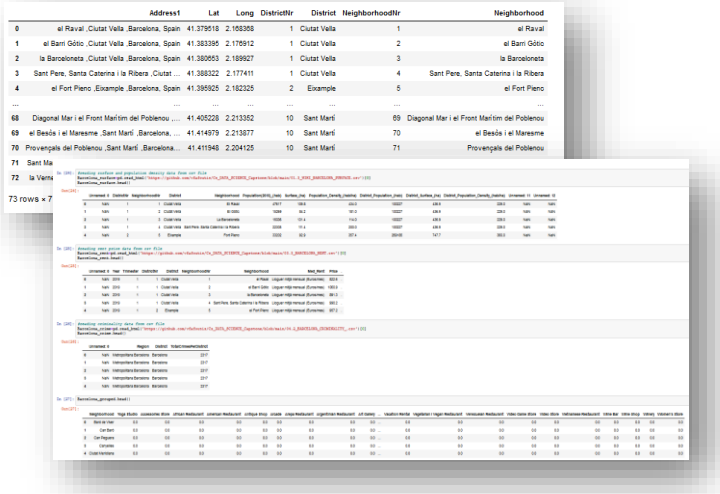


Methodology

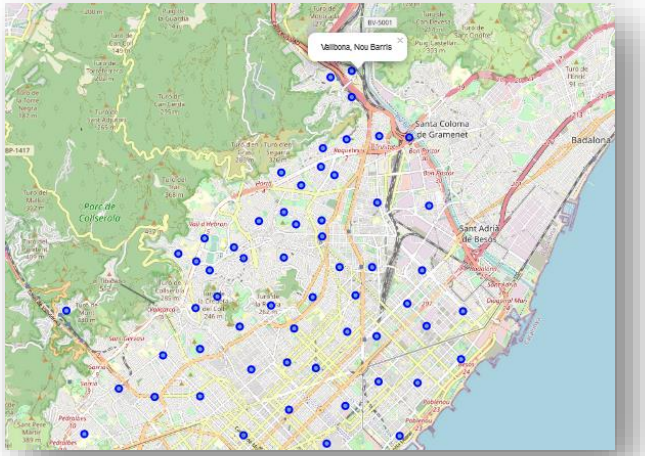
1.Python libraries and data retrieval



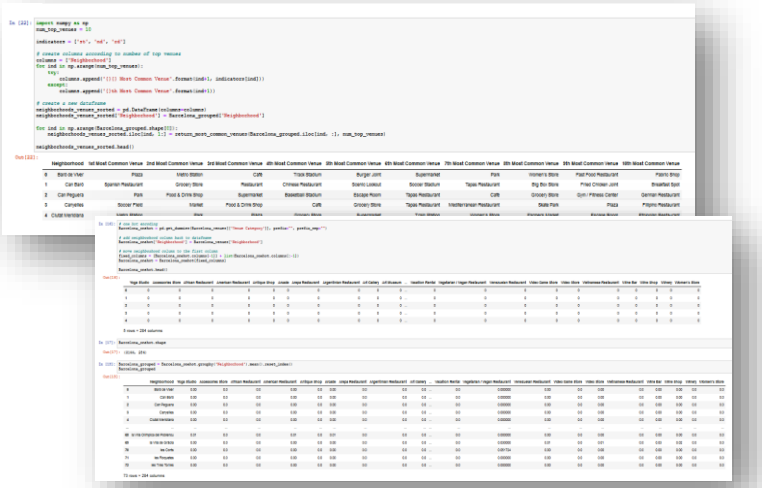
2. Data preprocessing



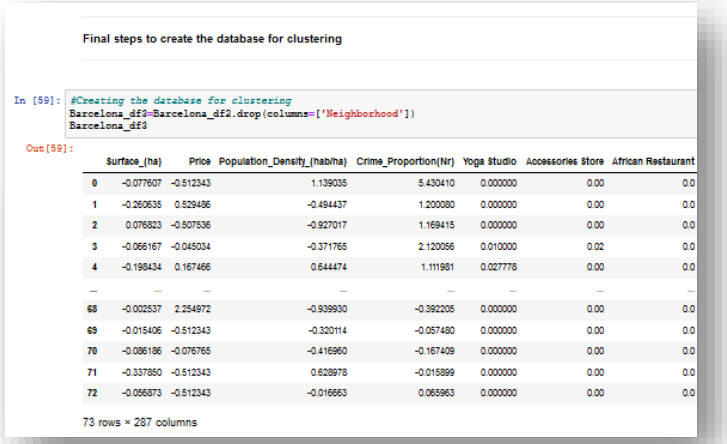
3. Neighborhood coordinates retrieval and map exploration



4. Venues exploration



5. Dataframes merging (all data) and standardization



Final steps to create the database for clustering

```
In [59]: #Creating the database for clustering
         Barcelona_df3=Barcelona_df2.drop(columns=['Neighborhood'])
         Barcelona_df3
```

Out[59]:

	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	Yoga Studio	Accessories Store	African Restaurant
0	-0.077607	-0.512343	1.139035	5.430410	0.000000	0.00	0.0
1	-0.260635	0.529486	-0.494437	1.200080	0.000000	0.00	0.0
2	0.076823	-0.507536	-0.927017	1.169415	0.000000	0.00	0.0
3	-0.066167	-0.045034	-0.371765	2.120056	0.010000	0.02	0.0
4	-0.198434	0.167466	0.644474	1.111981	0.027778	0.00	0.0
...	...	...	...	...	...	...	...
68	-0.002537	2.254972	-0.939930	-0.392205	0.000000	0.00	0.0
69	-0.015406	-0.512343	-0.320114	-0.057480	0.000000	0.00	0.0
70	-0.086186	-0.076765	-0.416960	-0.167409	0.000000	0.00	0.0
71	-0.337850	-0.512343	0.628978	-0.015899	0.000000	0.00	0.0
72	-0.056873	-0.512343	-0.016663	0.065963	0.000000	0.00	0.0

73 rows × 287 columns

6. K means clustering

Results: 73 Barcelona neighborhoods are divided in 5 clusters in a meaningful way. Further improvement recommendations are also detected (please see report).

Cluster 4 — Less dense urban area

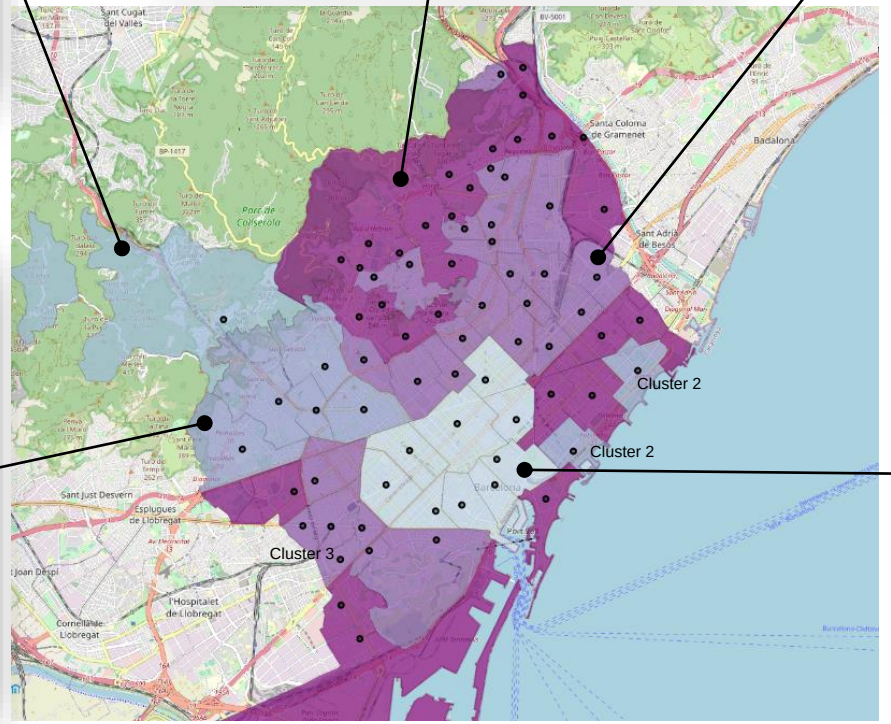
Rent: ●●○○○
Pop. Density: ●●○○○
Criminality: ●●○○○
Common venues: Rest., bakeries

Cluster 1 — Barcelona's Peak

Rent: ●●●●○
Pop. Density: ●○○○○
Criminality: ●○○○○
Common venues: Rest., Rail station

Cluster 3 — Dense urban area

Rent: ●●○○○
Pop. Density: ●●●●○
Criminality: ●●●○○
Common venues: Restaurants



Cluster 2 — High class area

Rent: ●●●●●
Pop. Density: ●●○○○
Criminality: ●●●○○
Common venues: Rest., bakeries

Cluster 0 — City Center

Rent: ●●●○○
Pop. Density: ●●●●○
Criminality: ●●●●●
Common venues: Hotels, Bars, Rest.