



IBM Data Science

Applied Data Science Project

**Peer-graded Assignment: Capstone Project**

**The Battle of Neighborhoods (Week 2)**

Vasileios Fafoutis

13.12.2020

## **Index**

**p.3     Scope and data sources of the analysis**

**p.4     Methodology**

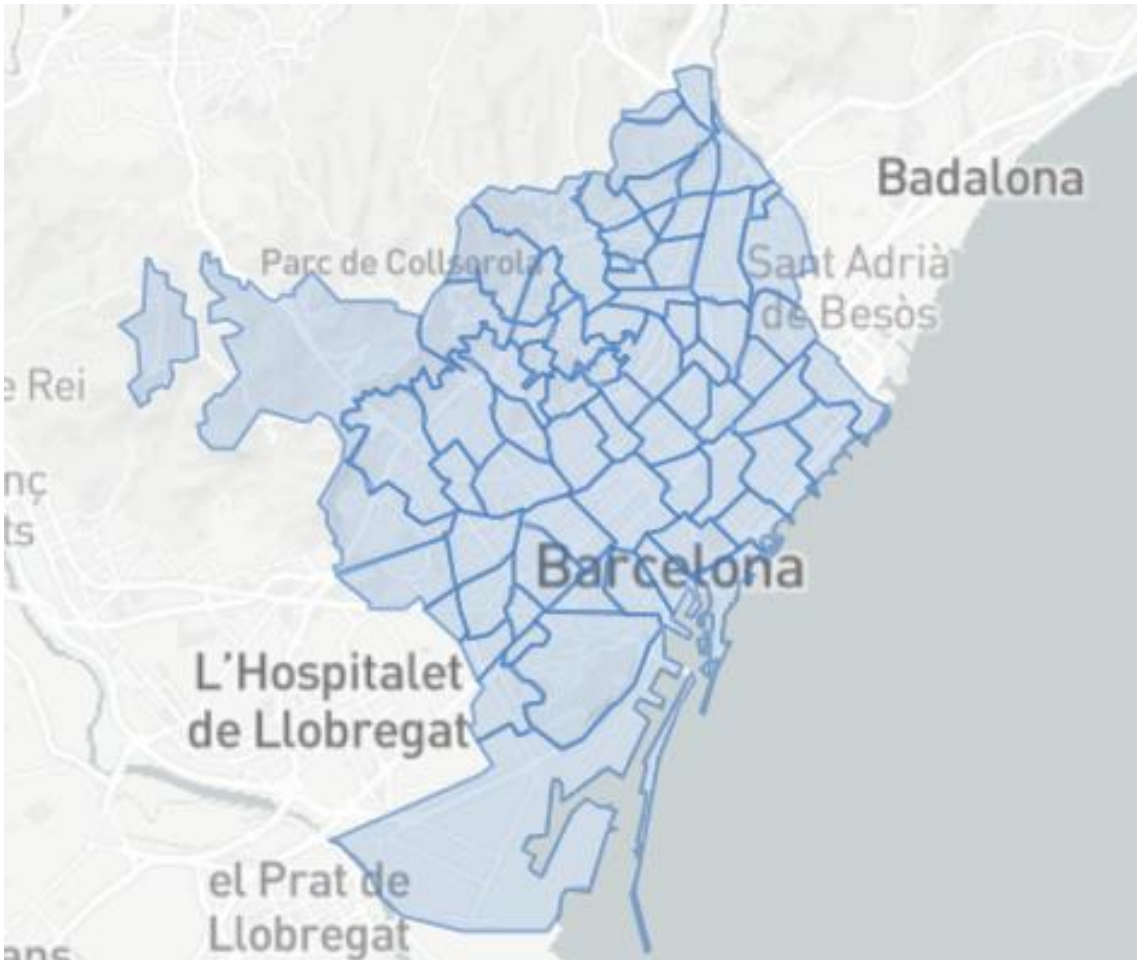
**p.5     Results and conclusions**



# 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/catalog\\_dades\\_obertes/dades-delinqeencials/](https://mossos.gencat.cat/ca/els_mossos_desquadra/indicadors_i_qualitat/dades_obertes/catalog_dades_obertes/dades-delinqeencials/)

[https://ca.wikipedia.org/wiki/Districtes\\_i\\_barris\\_de\\_Barcelona](https://ca.wikipedia.org/wiki/Districtes_i_barris_de_Barcelona)

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

A screenshot of the Barcelona Open Data portal. It displays a table with columns for 'Average monthly rent (€/month)' and 'Average rent per surface (€/m2 per month)'. The table lists various neighborhoods and their corresponding rental data.

A screenshot of the Gencat Dades delinqeencials portal. It shows a table with columns for 'Delictes', 'Delictes per 1000', and 'Delictes per 10000'. The table lists various types of crimes and their frequency.

A screenshot of the Wikipedia page for 'Districtes i barris de Barcelona'. It provides a list of the 73 neighborhoods of Barcelona, grouped by district.

A screenshot of the Foursquare Developers page. It features a 'Welcome to Foursquare Developers' message and a list of features and resources for developers.

## Methodology

## 1. Python libraries and data retrieval

[illegible]

## 4. Venues exploration

[illegible]

## 2. Data preprocessing

	Address	Lat	Long	District	District	Neighborhood	Neighborhood
3	el Raval, Ciutat Vella, Barcelona, Spain	41.379516	1.932056	1	Ciutat Vella	1	el Raval
4	el Barri Gòtic, Ciutat Vella, Barcelona, Spain	41.383395	1.978912	1	Ciutat Vella	2	el Barri Gòtic
5	la Barceloneta, Ciutat Vella, Barcelona, Spain	41.380053	1.959927	1	Ciutat Vella	3	la Barceloneta
6	Sant Pere, Santa Caterina i la Ribera, Ciutat Vella, Barcelona, Spain	41.388322	1.271411	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera
4	el Fort Pienc, Eixample, Barcelona, Spain	41.396925	1.821235	2	Eixample	5	el Fort Pienc
...	...	...	...	...	...	...	...
68	Diagonal Mar i el Front Marítim del Poblenou, ...	41.405228	2.213362	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou
69	el Besòs i el Maritim, Sant Martí, Barcelona, ...	41.414879	2.201357	10	Sant Martí	70	el Besòs i el Maritim
70	Provençals del Poblenou, Sant Martí, Barcelona, ...	41.411948	2.204125	10	Sant Martí	71	Provençals del Poblenou
71	Sant Març, Sant Martí, Barcelona, ...	...	...	...	...	...	...
72	la Vinya, Sant Martí, Barcelona, ...	...	...	...	...	...	...

73 rows × 7

## 5. Dataframes merging (all data) and standardization

Final steps to create the database for clustering

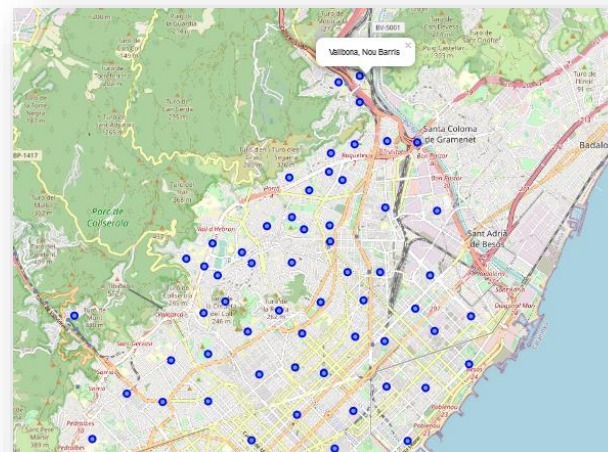
```
In [59]: #Creating the database for clustering
Barcelona_df2=Barcelona_df2.drop(columns=['Neighborhood'])
Barcelona_df2
```

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	0.00
1	-0.260635	0.529486	-0.494437	1.200080	0.000000	0.00	0.00
2	0.076823	-0.507536	-0.927017	1.169415	0.000000	0.00	0.00
3	-0.066167	-0.045034	-0.371765	2.120056	0.010000	0.02	0.00
4	-0.199434	0.167466	0.644474	1.111981	0.027778	0.00	0.00
...	...	...	...	...	...	...	...
68	-0.002537	2.254972	-0.939930	-0.392205	0.000000	0.00	0.00
69	-0.015406	-0.512343	-0.320114	-0.057480	0.000000	0.00	0.00
70	-0.086186	-0.076765	-0.416960	-0.167409	0.000000	0.00	0.00
71	-0.337850	-0.512343	0.628978	-0.015899	0.000000	0.00	0.00
72	-0.056873	-0.512343	-0.016663	0.065963	0.000000	0.00	0.00

73 rows x 8 columns

### 3. Neighborhood coordinates retrieval and map exploration



## 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).**

