Search for a bakery location in Barcelona

Carlos Moreno Visa

- ***Barcelona (BCN)*** capital from Catalonia (Spain) is the second city in Spain, one of the most populated in the EU (1,7M) , olympic city 1992.
- Global leading tourist, business and cultural centre with extensive transportation network, nice climate, lively/friendly character from citizens.
- **"Gourmet" bakery** is a profitable business if "Location" and rest of KPI (product, price, promotion, process, persons, plan) plan are well selected.
- **Location parameters:** choose 1-5 neighborhoods(from 73), avoid too much competition & excellence **proximity** (more than 400 bakeries, some "best"), priorize high **population/density**, **family income** rates higher than average and store **rental prices** not in top range in any case.

Introduction

- ***Foursquare API: info about bakeries and neighborhood venues***
- ***BCN Urban Statistics***

  (**Population, Area and Density**) **by neighbordhoods** (data 2018)

  http://www.bcn.cat/estadistica/angles/dades/barris/timm/tterr/sup418.htm

  (**Family income**) (data 2017)

  http://www.bcn.cat/estadistica/angles/dades/economia/renda/rdfamiliar/evo/rfbarris.htm

  (Price registered contracts and rental housing -**Rental€m2**) (data 2018, 4Q summed up)

  http://www.bcn.cat/estadistica/angles/dades/barris/timm/ipreus/habllo/ls2018.htm

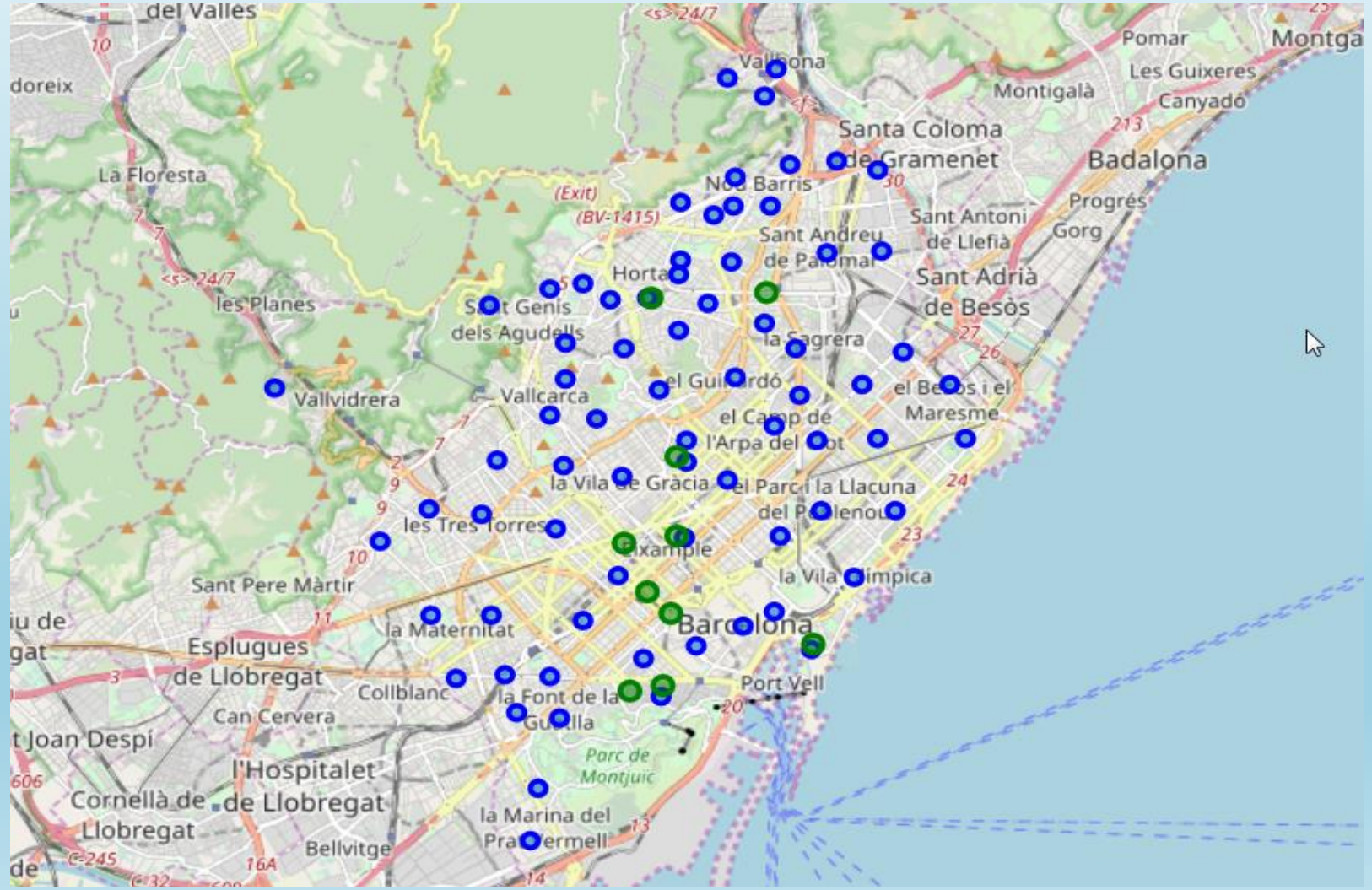
- **Latitude and longitude by neighborhood (Wikipedia+Geohack)**
- **“The ten best bakeries in Barcelona” El Periodico, January 11-2019**

https://www.elperiodico.com/es/onbarcelona/visitar/20190111/mejores-panaderias-barcelona-7233367

Data

Methodology

- First after loading data we display a BCN map with 73 **neighborhood centers** (blue) and 10 “best bakeries” (green)
- Neighborhood “live” center is not geographical center.



Methodology

M1 Bakeries analysis

M2 Neighborhood & Clustering analysis

M3 Neighborhood similarity

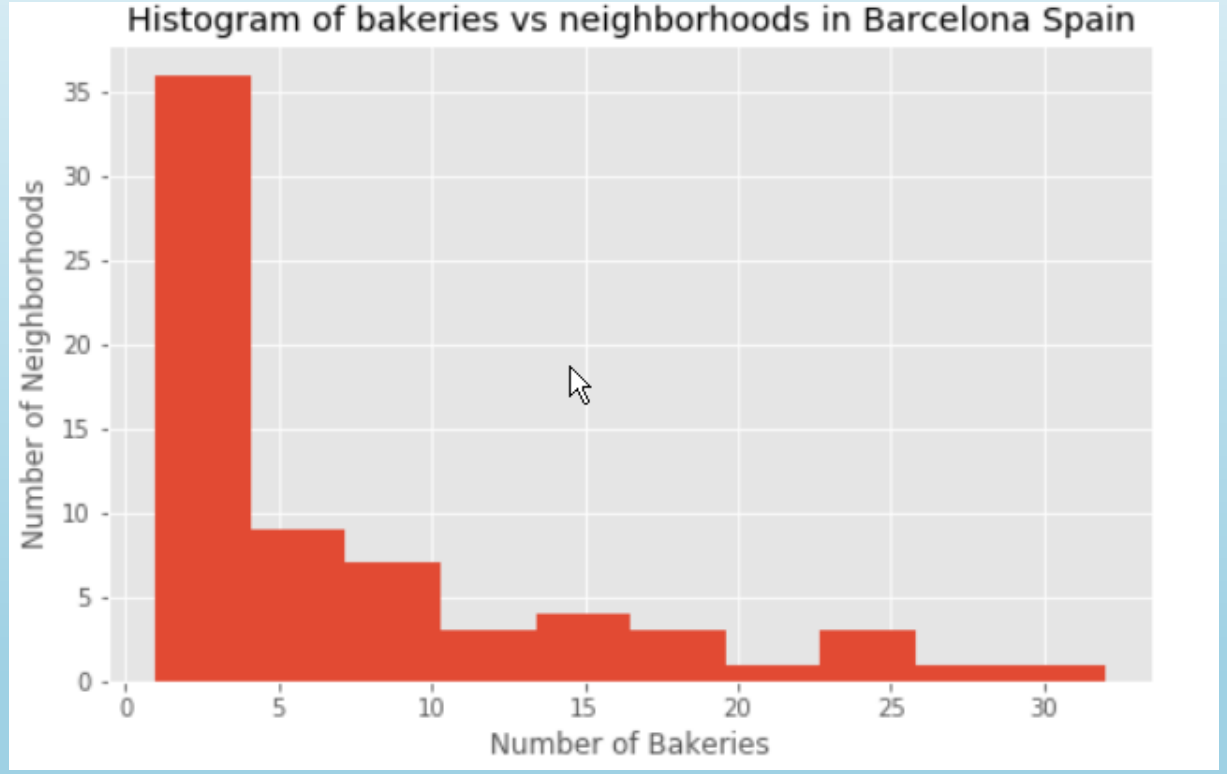
1-5 Neighborhoods

- Displaying & Analysing Bakeries
- Neighborhood variable analysis
- Getting FS Venues
- First Discarding neighborhoods

- Cleaning and grouping FS bakeries by neighborhood (min(1),max(32),mean(7,36), 5 neighborhoods without any bakery returned -radius=500m -)..

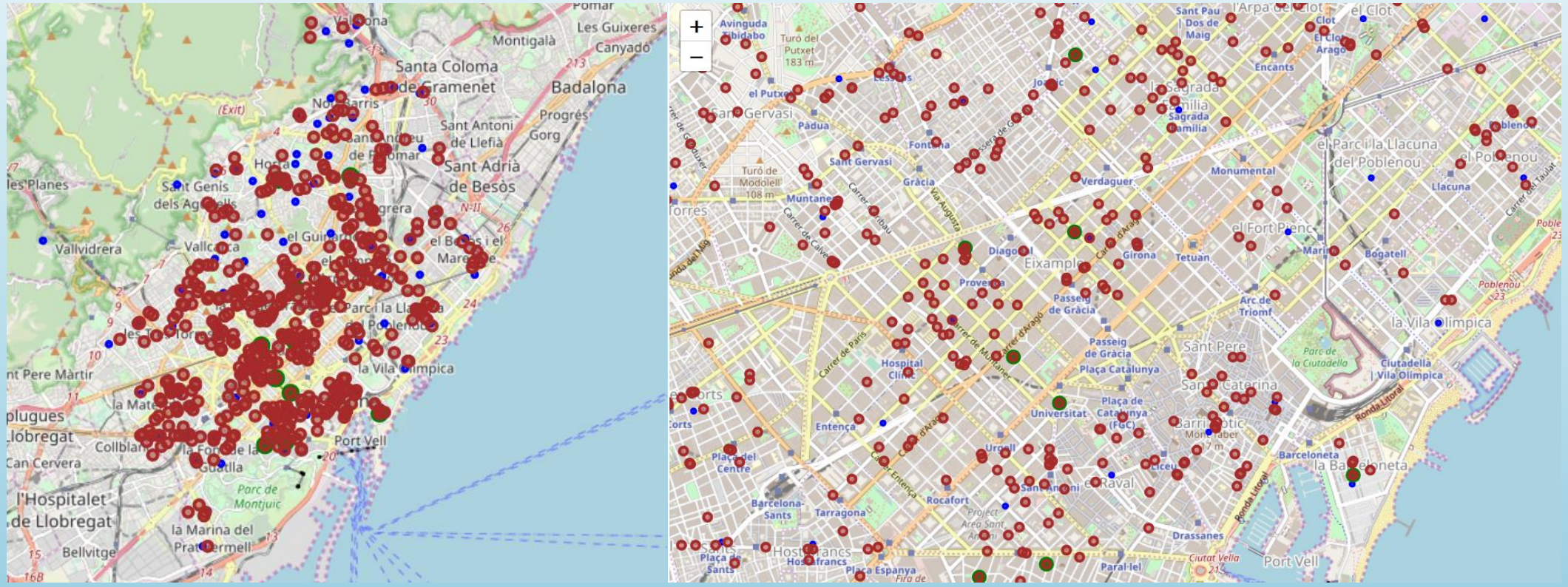
First discarding set: 5 highest bakeries-populated neighborhoods

Nobkrys	Neighborhood
32.0	Antiga Esquerra Eixample
27.0	Dreta Eixample
24.0	les Corts
24.0	Raval
23.0	Barri Gotic



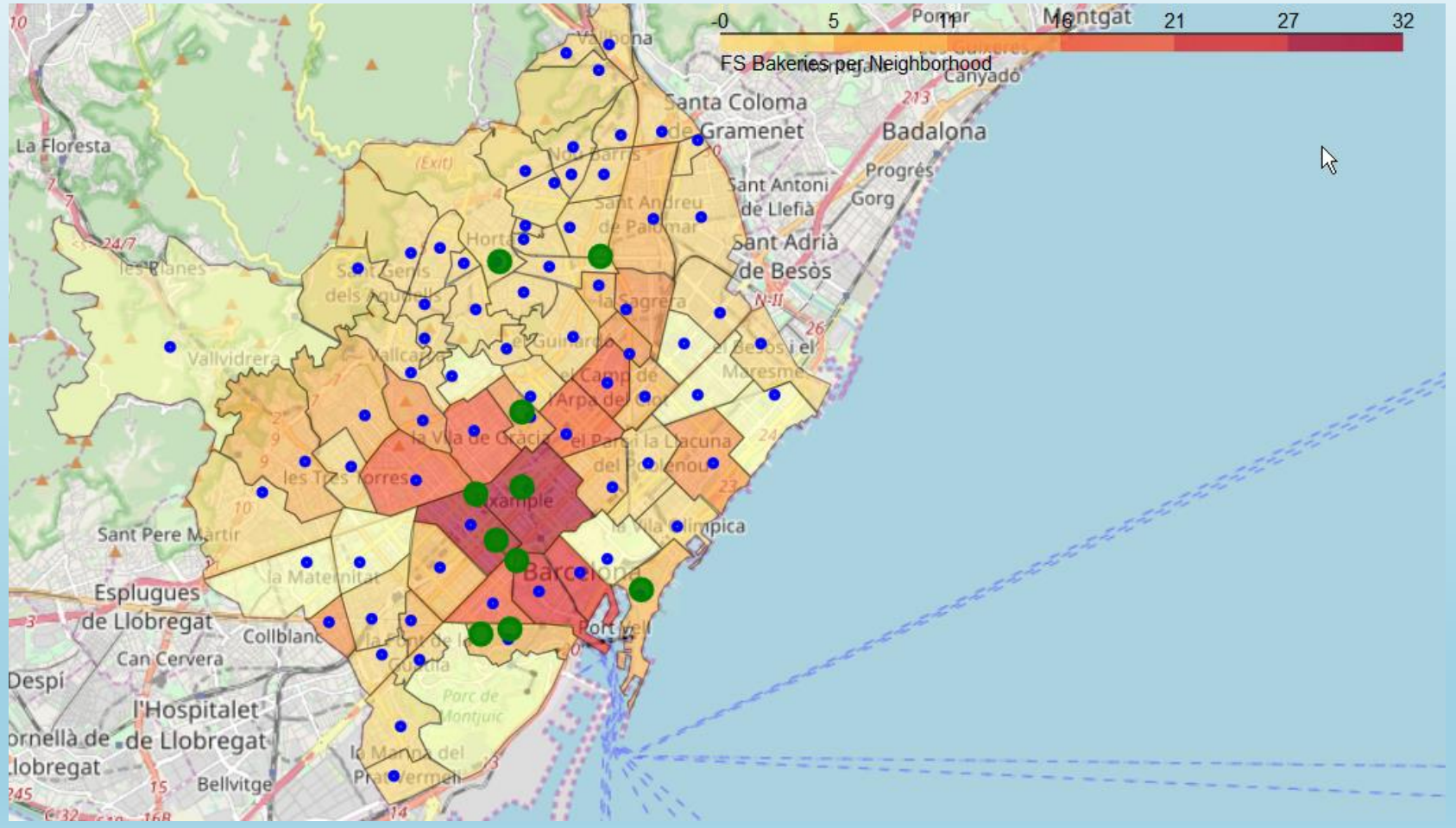
M1.1 Bakeries analysis- Displaying & Analysing Bakeries (1)

- BCN map including bakeries(brown), best bakeries(green) and neighborhood centers(blue)



M1.1 Bakeries analysis- Displaying & Analysing Bakeries (2)

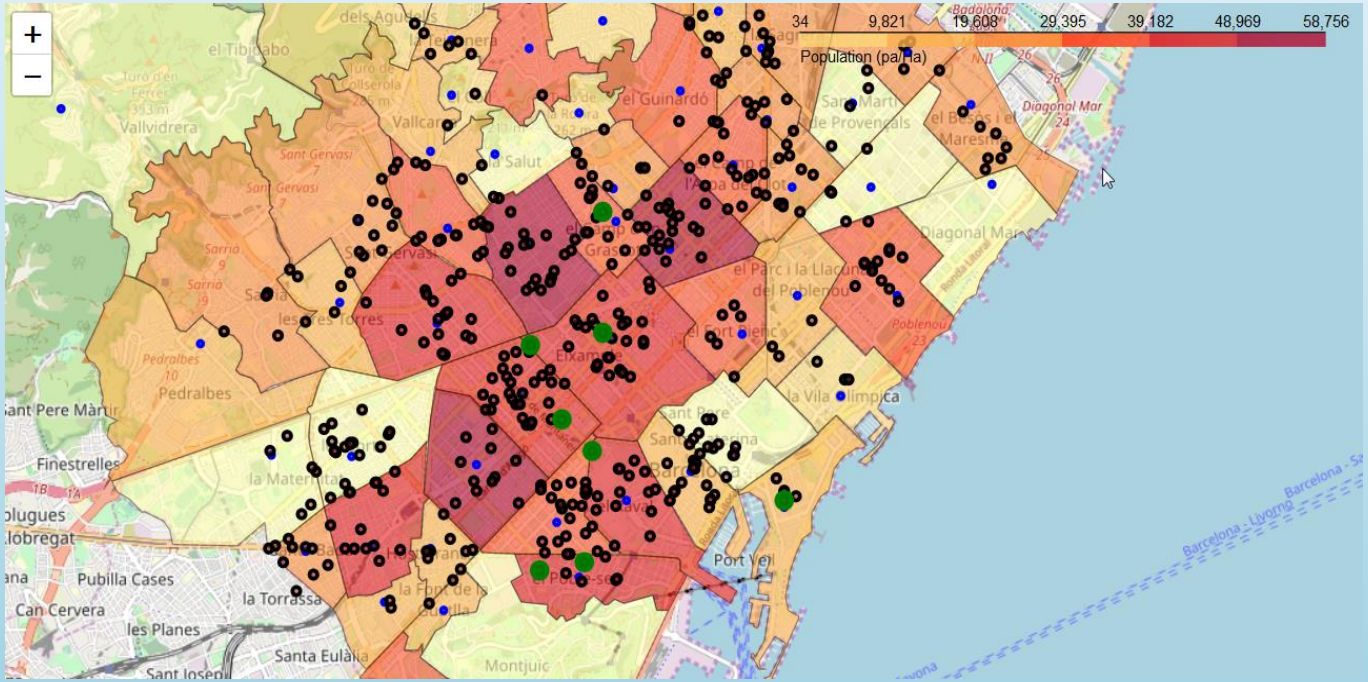
- BCN choropleth map based on bakeries number by neighborhood including best bakeries(green) and neighborhood centers(blue)



M1.1 Bakeries analysis- Displaying & Analysing Bakeries (3)

- Detected 0,72 correlation Family income vs Residential rental€m2 month

. 2 new discard sets: **low_population** (<quartile 40%), **low_density**(<quartile 15%)



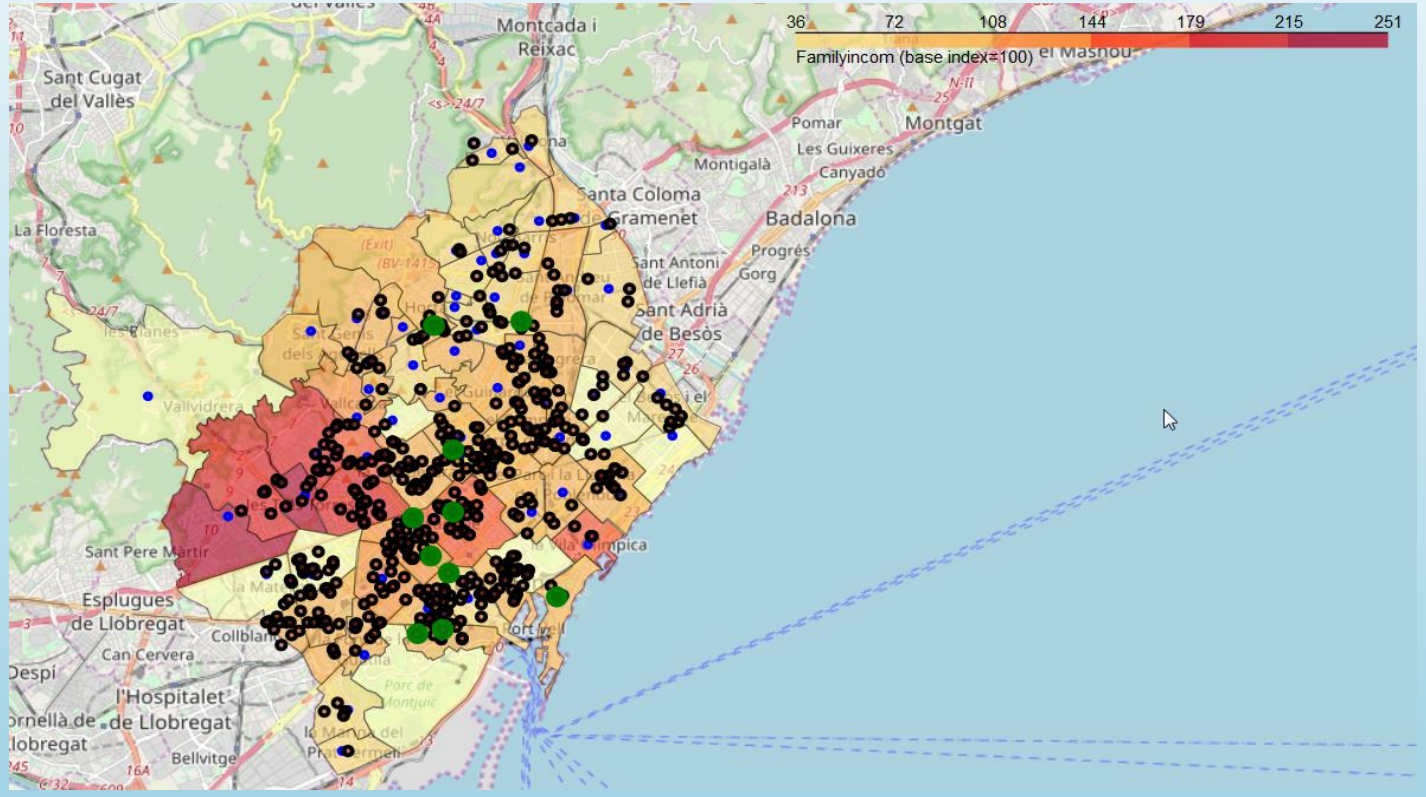
Number of discarded neighborhood by low population : 29

	Neighborhood	Population
41	la Clota	610
11	la Marina del Prat Vermell	1149
55	Vallbona	1372
46	Can Peguera	2271
57	Baro de Viver	2539
53	Torre Baro	2856

- BCN choropleth map based on population, bakeries(black), best bakeries(green) and neighborhood centers(blue). Higher bakeries concentration on higher populated neighborhoods.

M1.2 Neighborhood variable analysis (1)

- We create new discard sets on **low_family income** (<quartile 30%) and **too_high_rental** price (>quartile 75%, maximum 10)

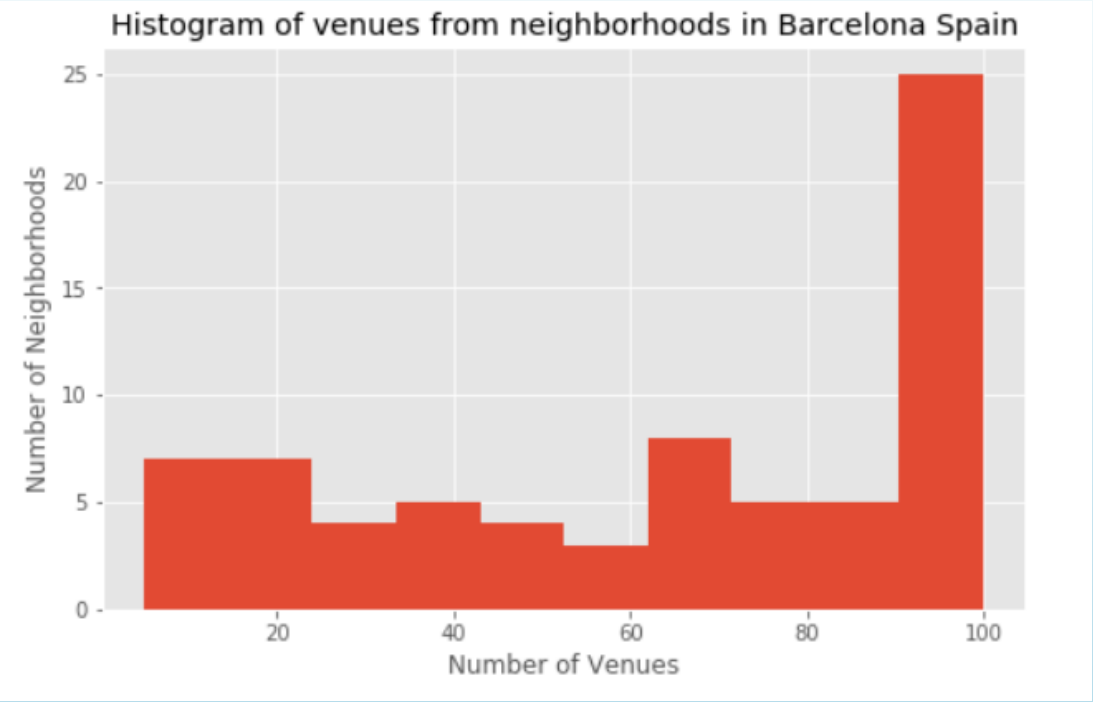


Number of discarded neighborhoods by too high rental : 10 from 17

	Neighborhood	Rental€m2
22	Sarria	17.2
66	Vila Olimpica	16.9
2	Barceloneta	16.9
23	Tres Torres	16.4
24	Sant Gervasi - la Bonanova	16.1

- BCN choropleth map based on family income. Lower bakery concentration on lower family income neighborhood (exc. 2 dark red neighborhoods)

M1.2 Neighborhood variable analysis (2)



- About 4700 Venues in 292 venue categories.
  25 neighborhoods have 100 venues(FS limit)
  About 13 neighborhoods less than 20 venues.

- Assume low touristic movement on the last ones and create a new discard set (**low_movement**)

```
Number of discarded neighborhoods by low_movement :  11
```

Neighborhood	Venue
Sant Genis dels Agudells	5
la Trinitat Vella	5
Vallvidrera, Tibidabo, Les Planes	6
Vallbona	7
Ciutat Meridiana	8

M1.3 Getting FS neighborhood venues

- We create a new discard set (**bbdiscard**) including neighborhoods with more than one “best bakery”
- M1 discard resulting set is generated using a “multiple” outer join with all discard sets created previously. A filtering candidate column “Selected” is added to original neighborhoods table.

```
Initial number of candidates end Step 1 is :  23
```

	IdNeig	Neighborhood	IdBor	Borough	Population	Area	Density	Familyincom	Rental€m2	Latitude	Longitude	Selected
4	5	Fort Pienc	2	EIXAMPLE	32016	92.9	344.7	106.5	13.1	41.395675	2.183703	CAN
5	6	Sagrada Familia	2	EIXAMPLE	51539	105.1	490.4	101.8	13.7	41.403561	2.174347	CAN
8	9	Nova Esquerra Eixample	2	EIXAMPLE	58180	133.8	434.9	110.2	13.9	41.383389	2.149000	CAN
9	10	Sant Antoni	2	EIXAMPLE	38345	80.1	478.7	104.2	13.2	41.378010	2.159490	CAN
14	15	Hostafrancs	3	SANTS-MONTJUIC	15904	41.0	387.7	99.0	13.5	41.375556	2.143056	CAN
15	16	la Bordeta	3	SANTS-MONTJUIC	18530	57.7	321.4	79.0	12.2	41.370494	2.137097	CAN
16	17	Sants - Badal	3	SANTS-MONTJUIC	23987	41.1	584.3	81.0	13.0	41.375278	2.126667	CAN
17	18	Sants	3	SANTS-MONTJUIC	41127	109.8	374.6	99.0	13.1	41.375730	2.135250	CAN

M1.4 First discarding neighborhoods

- Construct a neighborhood grouped table with dummy variables from column “Venue category”
- Values shown are the mean of the frequency of occurrence for each category.

	Neighborhood	Accessories Store	African Restaurant	American Restaurant	Amphitheater	Antique Shop	Arcade
0	Antiga Esquerra Eixample	0.000000	0.0	0.000000	0.0	0.00	0.000000
1	Baix Guinardo	0.000000	0.0	0.000000	0.0	0.00	0.000000
2	Barceloneta	0.000000	0.0	0.000000	0.0	0.00	0.000000
3	Baro de Viver	0.000000	0.0	0.000000	0.0	0.00	0.000000
4	Barri Gotic	0.000000	0.0	0.000000	0.0	0.00	0.000000
5	Besos i Maresme	0.000000	0.0	0.000000	0.0	0.00	0.000000
6	Bon Pastor	0.014085	0.0	0.028169	0.0	0.00	0.000000
7	Camp d'en Grassot i Gracia Nova	0.000000	0.0	0.000000	0.0	0.00	0.000000

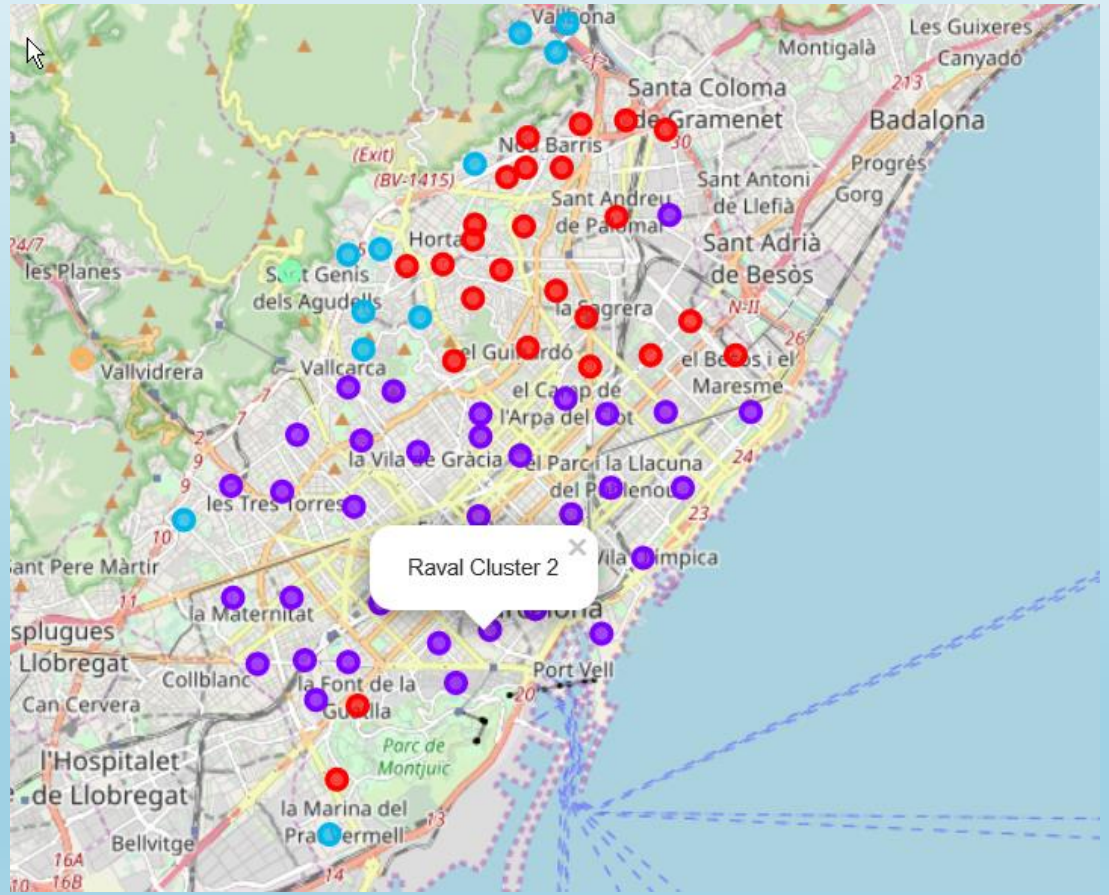
```
----Antiga Esquerra Eixample----
                      venue  freq
0                     Hotel  0.10
1        Spanish Restaurant  0.08
2             Cocktail Bar  0.06
3       Japanese Restaurant  0.05
4  Mediterranean Restaurant  0.05


----Baix Guinardo----
                 venue  freq
0                  Bar  0.07
1   Tapas Restaurant  0.06
2          Restaurant  0.04
3  Spanish Restaurant  0.04
4  Italian Restaurant  0.04
```

Neighborhood	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	10th Most Common Venue
Antiga Esquerra Eixample	Antiga Esquerra Eixample	Hotel	Spanish Restaurant	Cocktail Bar	Japanese Restaurant	Mediterranean Restaurant	Pizza Place	Burger Joint	Sandwich Place	Bakery	Tapas Restaurant
Baix Guinardo	Baix Guinardo	Bar	Tapas Restaurant	Restaurant	Italian Restaurant	Spanish Restaurant	Hotel	Japanese Restaurant	Supermarket	Gym	Grocery Store
Barceloneta	Barceloneta	Tapas Restaurant	Paella Restaurant	Seafood Restaurant	Mediterranean Restaurant	Bar	Burger Joint	Spanish Restaurant	Ice Cream Shop	Restaurant	Wine Bar
Baro de Viver	Baro de Viver	Spanish Restaurant	Supermarket	Plaza	Metro Station	Asian Restaurant	Chinese Restaurant	Salon / Barbershop	Café	Track Stadium	Restaurant
Barri Gotic	Barri Gotic	Tapas Restaurant	Plaza	Spanish Restaurant	Bar	Wine Bar	Cocktail Bar	Ice Cream Shop	Coffee Shop	Hotel	Vegetarian / Vegan Restaurant

- Print each neighborhood along with 5 most common venue categories plus freq and create a neighborhood indexed table displaying top 10 venues categories

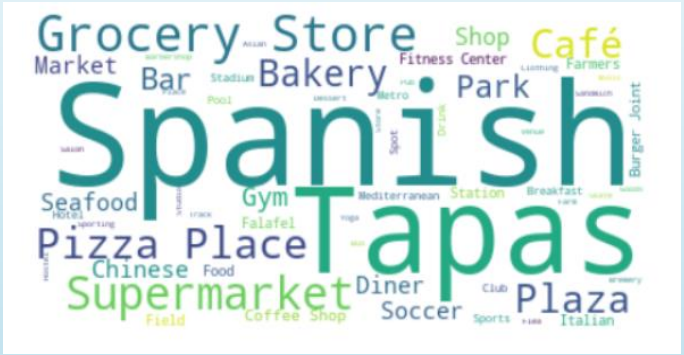
M2 Neighborhood and clustering analysis(1)

- ***K-means*** algorithm from Scikit-learn library for Python is used to cluster BCN neighborhoods into **5** clusters (other values tested did not add semantic value)
- Barcelona colored-cluster dotted map displays different neighborhood clusters



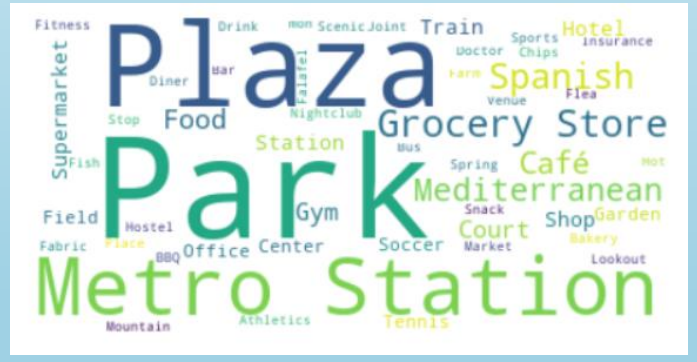
M2 Neighborhood and clustering analysis(2)

- **Word Clouds** help us to better understand clusters nature (based on 10 most freq.venues table)

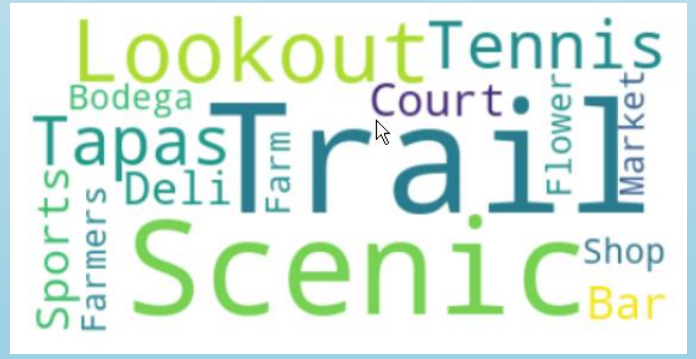


Cluster 1

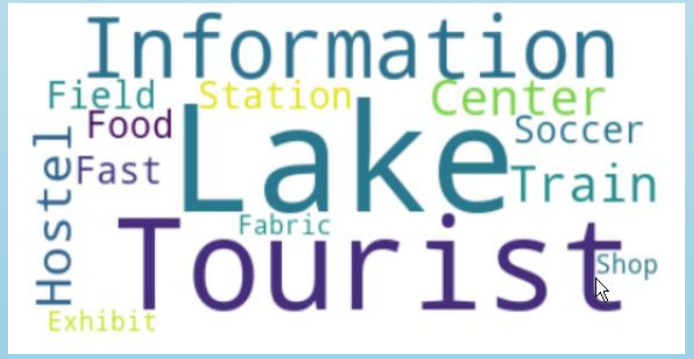
Cluster 2



Cluster 3



Cluster 4



Cluster 5

M2 Neighborhood and clustering analysis(3)

Cluster 1 — 25 neighborhoods. Well grouped(2 exc). High venue density. Spanish and tapas restaurant, Supermarket, Grocery store, Plaza, Bakery and Café most noticeable→ → Residential, low tourist neighborhoods

Cluster 2 — 35 neighborhoods. Well grouped(1 exc). Very high venue density. Downtown, coastal and west areas of BCN. Spanish restaurant,Hotel, Tapas, Café, Italian rest, Bakery and different specialized restaurants-> High population and tourism (people flow)

Cluster 3 — 11 neighborhoods. Spread along mountain line from SW to N, far from sea and downtown. Lower venue density. Park, Plaza and Metro Station most noticeable→ → Mostly Residential, well connected, Nature & Sports facilities.

Cluster 4 — 1 neighborhood (S.Genis dels Agudells). Very low density. Trails, Scenic Lookouts, Tennis courts most noticeable→ Low population. Apart from cluster 5 the closest to mountain line(includes words as *Farm* and *Farmers*)

Cluster 5 — 1 neighborhood formed by 3 isolated surrounding residential areas inside green mountain area close to city. Lowest venue density. Lake, Tourist information, Train, Hostel, Station most noticeable→ Lowest population

* Candidates from Stage 1 ∩ Cluster 2 = Candidates from Stage 2 (18)

M2 Neighborhood and clustering analysis(4)

- We determine most similar neighborhoods to the profile of a *“location model bakery”* selected from our list of “*10 best bakeries*”.
- **Similarity** using *Pearson correlation* on a neighborhood matrix composed by “simple-scaling normalized” values (population, density,area, family income, rental€m2) and the mean of the frequency of occurrence of each venue category

Neighborhood	Population	Area	Density	Familyincom	Rental€m2	Accessories Store	African Restaurant	American Restaurant	Amphitheater	Antique Shop	...
Raval	0.818288	0.576984	0.741742	0.492393	0.958904	0.00	0.0	0.0	0.0	0.0	...
Fort Pienc	0.550292	0.488177	0.589937	0.736515	0.897260	0.01	0.0	0.0	0.0	0.0	...
Sagrada Familia	0.885854	0.552286	0.839295	0.704011	0.938356	0.00	0.0	0.0	0.0	0.0	...
Nova Esquerra Eixample	1.000000	0.703100	0.744309	0.762102	0.952055	0.00	0.0	0.0	0.0	0.0	...
Sant Antoni	0.659075	0.420914	0.819271	0.720609	0.904110	0.00	0.0	0.0	0.0	0.0	...

Neighborhood	Raval	Fort Pienc	Sagrada Familia	Nova Esquerra Eixample
Raval	1.000000	0.964186	0.987568	0.984956
Fort Pienc	0.964186	1.000000	0.976190	0.973048
Sagrada Familia	0.987568	0.976190	1.000000	0.990938
Nova Esquerra Eixample	0.984956	0.973048	0.990938	1.000000

- After matrix transposition and correlation we get a table we can filter by our selected best bakery and top end number of candidates to obtain sorted results.

Neighborhood	Raval
Sagrada Familia	0.987568
Vila de Gracia	0.986378
Nova Esquerra Eixample	0.984956
Sants	0.983052
Camp de l'Arpa	0.977581

M3 Neighborhood similarity and final Top ranking

- **Top 5 neighborhoods** for a bakery location in BCN found according to initial parameter given. Answer is a starting point for final “street level” exploration.
- Crossing external information about available store rental in selected/s neighborhoods and zooming in some of project maps including bakeries distribution will help in a great way to get our target.
- **Further research and improvements:**
  - *Checking actual bakeries data vs Foursquare returned (limited by 500m radius parameter) and non coincident-coordinates overlapping.*
  - *Using actual store rental prices by neighborhood (creating specific dataset from most relevant real estate agencies)*
  - *Removing 100 returning events by call (paid option).*
  - *Using candidate sets intersection instead of joining “discarding” sets (Stage 1) and compare results*
  - *Including Stage 1 variables into clustering*
  - *Replacing “best-bakery model” neighborhood (stage 3) by a “n-best-bakeries” model for calculating an “average ranking” of rankings by tuple “best-bakery-neighborhood” would provide different perspectives and better generalization.*
  - *Analysis on same-owner bakery shop groups if reliable bakery guild data available.*

Results and Discussion

- A “funnel” 3-step methodology extensible to other finding-business-location cases has been proposed.
- It applies to public and Foursquare data sources and combines some discarding rules assumptions based on initial business problem parameters and a mix of tools and machine learning algorithms: interactive geodata maps, choropleth, wordclouds, histogram, descriptive statistics, k-means clustering and Pearson correlation similarity).
- Decision support systems(DSS) based on data analysis and AI are widely spread. Automatic and semi-autonomous systems will even have more impact in our future lives.

Conclusion