CLUSTERING BARCELONA

Ian Riera Smolinska

Clustering Barcelona

- Cities are divided in administrative way. Can we divide it so it represents the demographic characteristics?
- Is there any correlation between the different demographic parameters?
- Find potential sources for modelling and predicting.

Data acquisition and cleaning

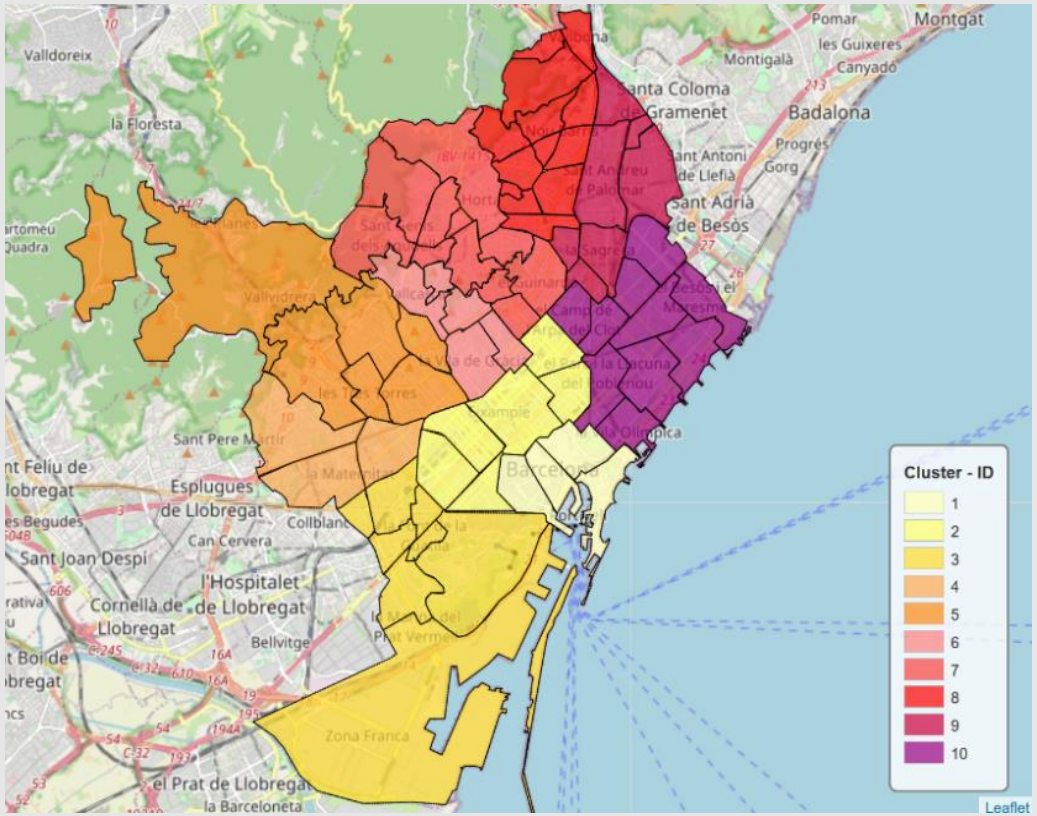
- Statistics form the Barcelona city council about the different parameters. https://www.bcn.cat/estadistica/castella/dades/barris/tpob/pad/ine/a2019/ine17.htm

- Foresquare API.

- COVID-19 information https://aspb.shinyapps.io/COVID19_BCN/

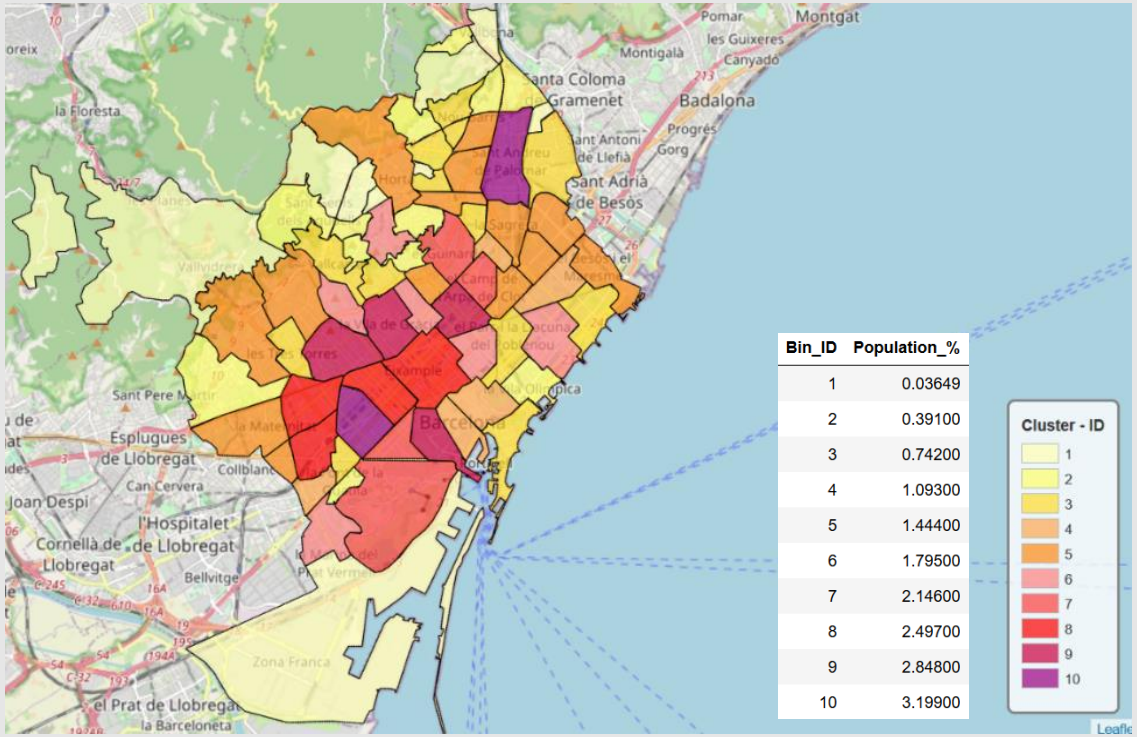
- 73 neighborhoods, 8 different parameters to analyze

Current administrative division



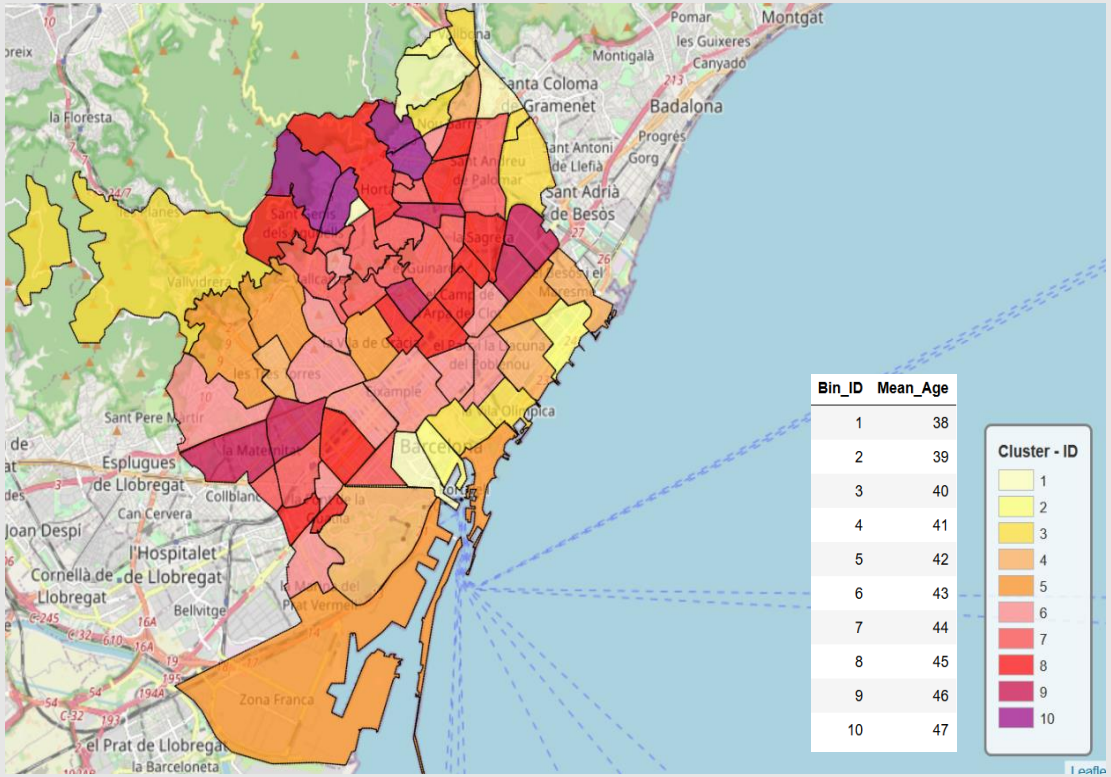
Number	District	Size km²	Neighborhoods
1	Ciutat Vella	4.49	La Barceloneta, El Gòtic, El Raval, Sant Pere, Santa Caterina i la Ribera
2	Eixample	7.46	L'Antiga Esquerra de l'Eixample, La Nova Esquerra de l'Eixample, Dreta de l'Eixample, Fort Pienc, Sagrada Família, Sant Antoni
3	Sants-Montjuïc	21.35	La Bordeta, la Font de la Guatlla, Hostafrancs, la Marina de Port, la Marina del Prat Vermell, El Poble-sec, Sants, Sants-Badal, Montjuïc*, Zona Franca - Port*
4	Les Corts	6.08	les Corts, la Maternitat i Sant Ramon, Pedralbes
5	Sarrià-Sant Gervasi	20.09	El Putxet i Farró, Sarrià, Sant Gervasi - la Bonanova, Sant Gervasi - Galvany, les Tres Torres, Vallvidrera, Tibidabo i les Planes
6	Gràcia	4.19	Vila de Gràcia, el Camp d'en Grassot i Gràcia Nova, la Salut, el Coll, Vallcarca i els Penitents.
7	Horta-Guinardó	11.96	El Baix Guinardó, El Guinardó, Can Baró, El Carmel, la Font d'en Fargues, Horta, la Clota, Montbau, Sant Genís dels Agudells, la Teixonera, La Vall d'Hebron.
8	Nou Barris	8.04	Can Peguera, Canyelles, Ciutat Meridiana, La Guineueta, Porta, La Prosperitat, les Roquetes, Torre Baró, la Trinitat Nova, El Turó de la Peira, Vallbona, Verdum, Vilapicina i la Torre Llobeta
9	Sant Andreu	6.56	Baró de Viver, Bon Pastor, El Congrés i els Indians, Navas, Sant Andreu de Palomar, La Sagrera i Trinitat Vella
10	Sant Martí	10.80	El Besòs i el Maresme, el Clot, El Camp de l'Arpa del Clot, Diagonal Mar i el Front Marítim del Poblenou, el Parc i la Llacuna del Poblenou, Poblenou, Provençals del Poblenou, Sant Martí de Provençals, La Verneda i la Pau, la Vila Olímpica del Poblenou

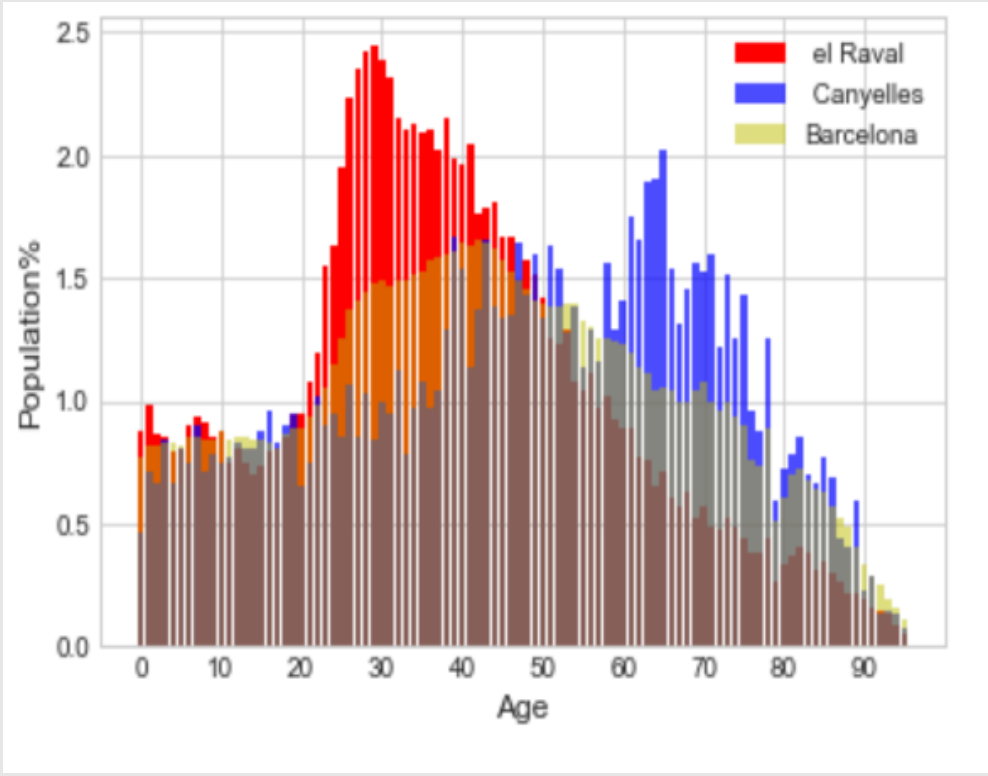
Population based division



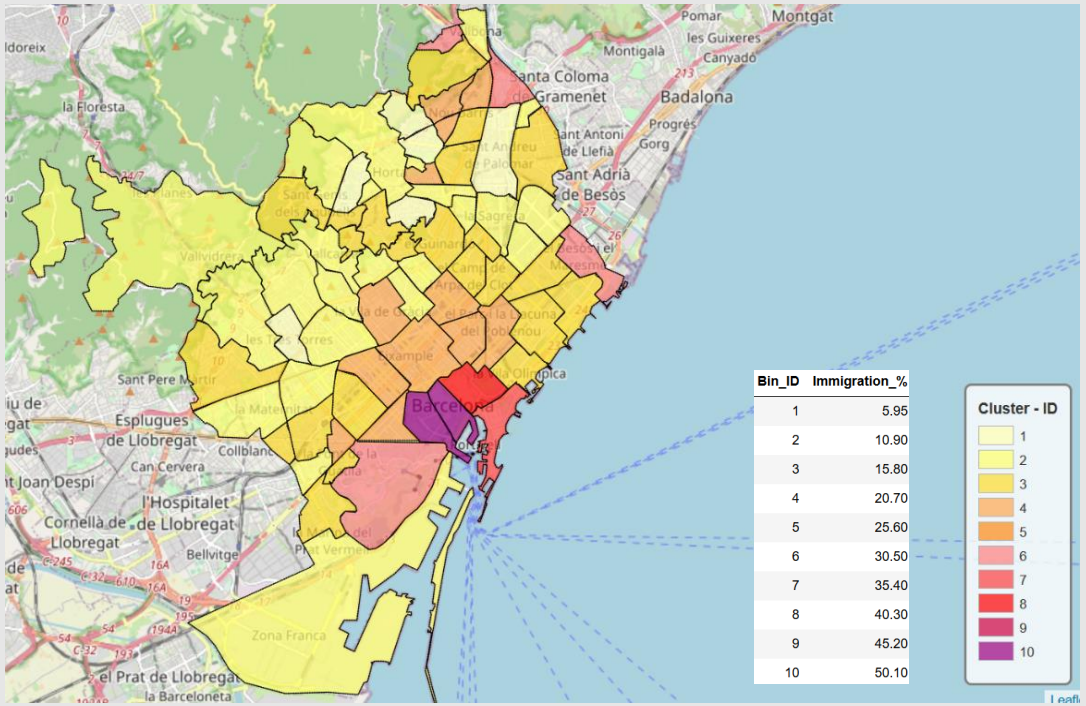
Bin_ID	Population_%
1	0.03649
2	0.39100
3	0.74200
4	1.09300
5	1.44400
6	1.79500
7	2.14600
8	2.49700
9	2.84800
10	3.19900

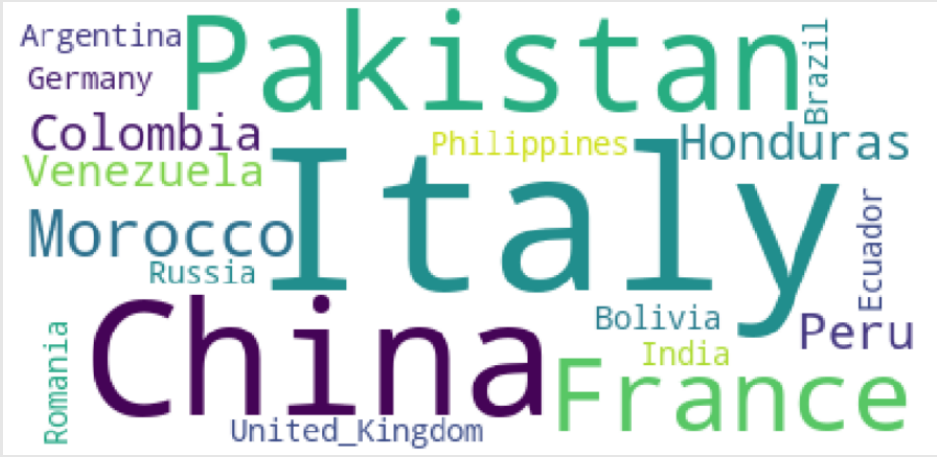
Age based division



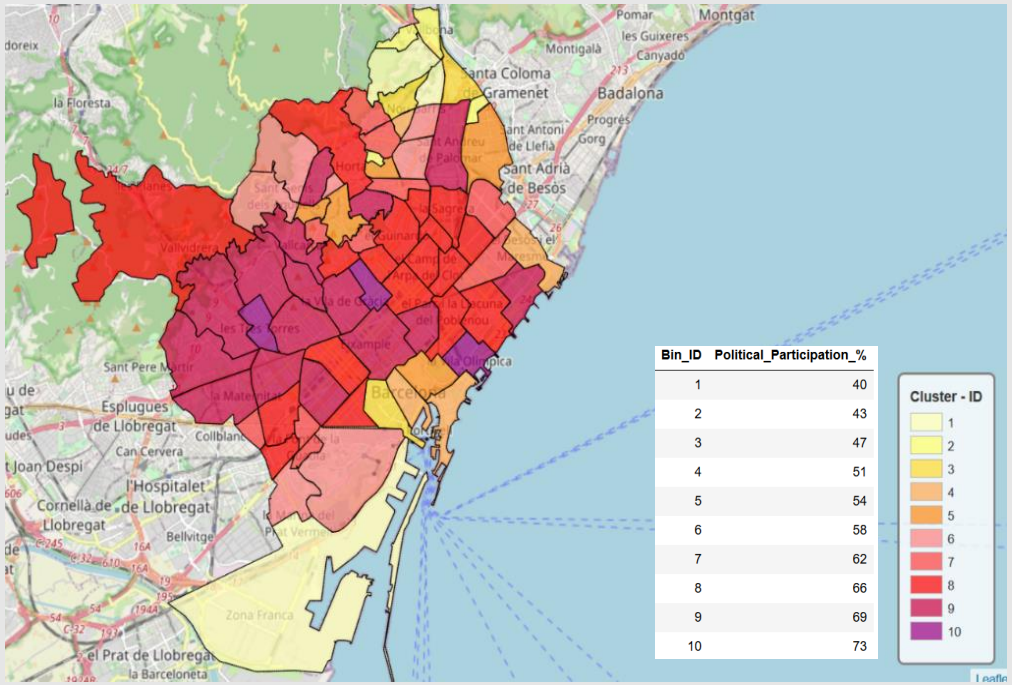


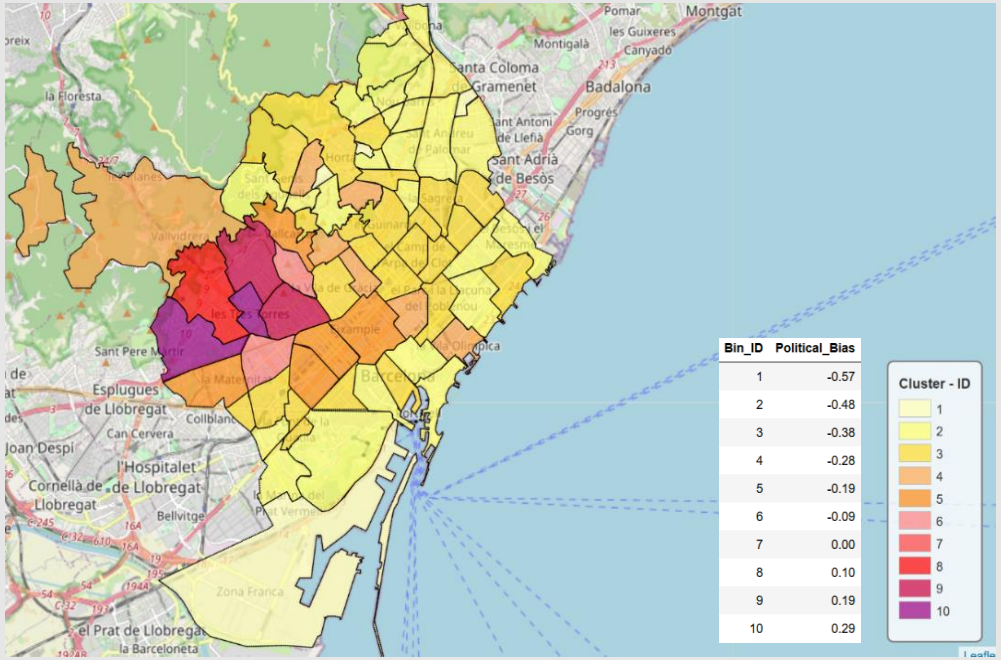
Nationality based division





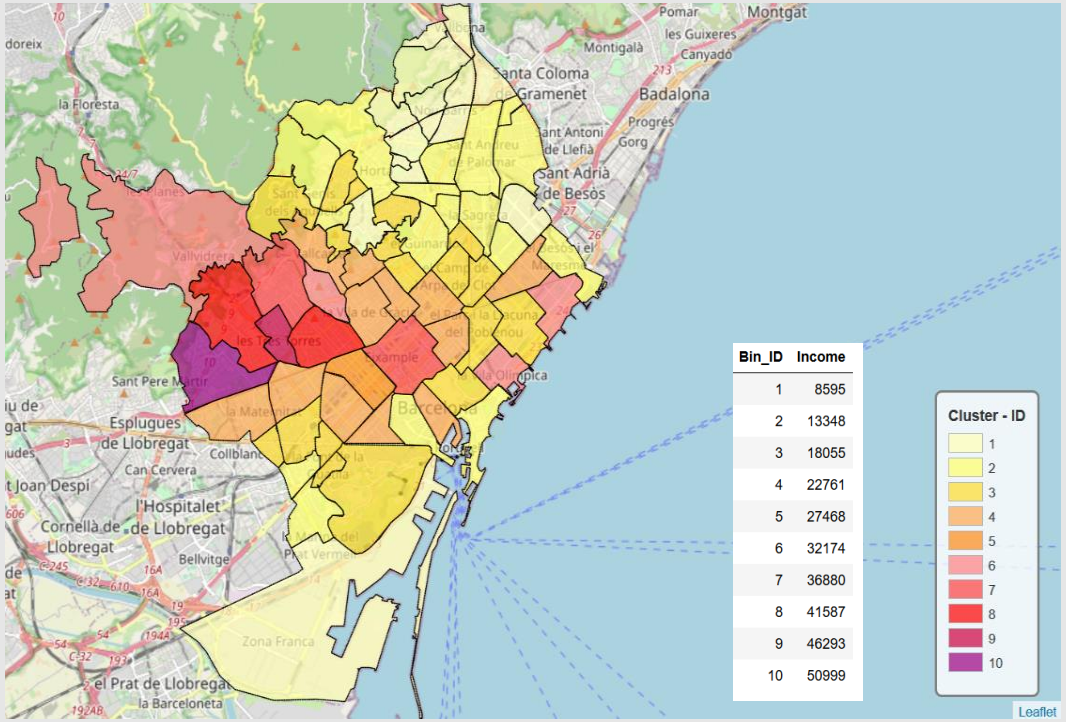
Politically based division





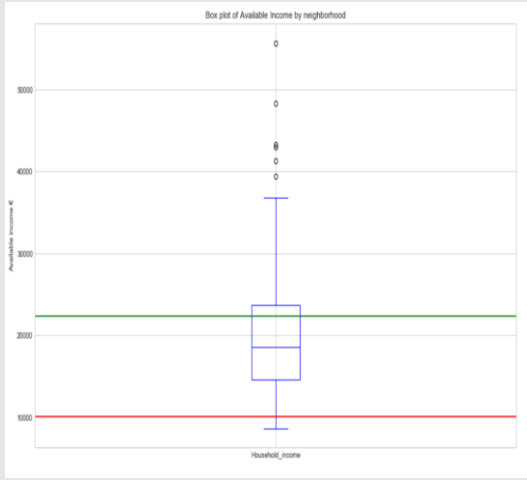
Being left political wing = -1 and right =+1

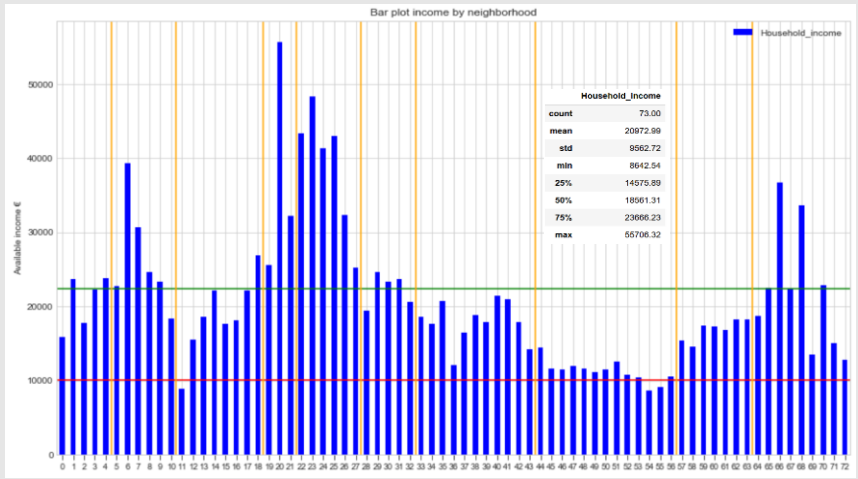
Household income division



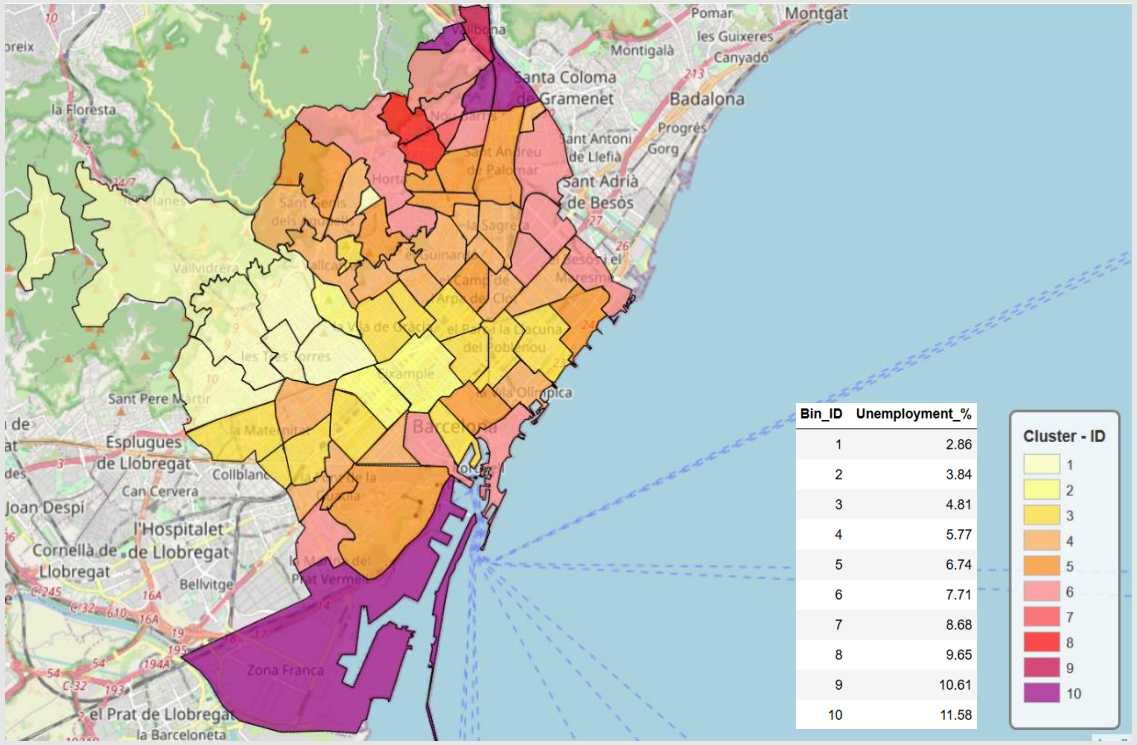
Bin_ID	Income
1	8595
2	13348
3	18055
4	22761
5	27468
6	32174
7	36880
8	41587
9	46293
10	50999

Cluster - ID
1
2
3
4
5
6
7
8
9
10



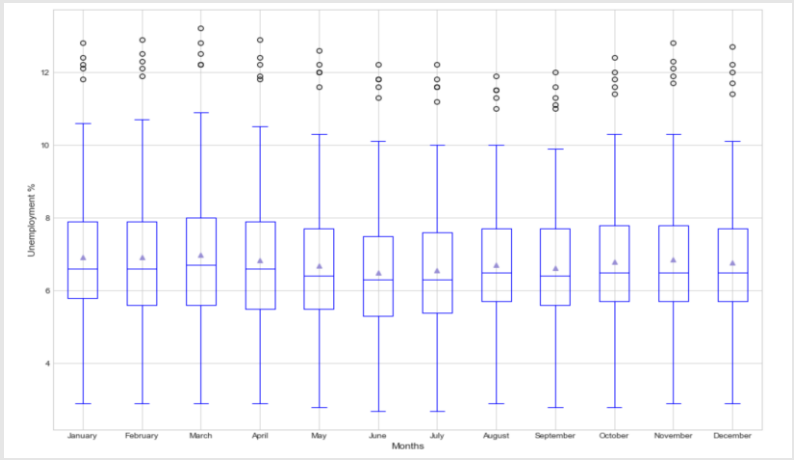


Unemployment based division



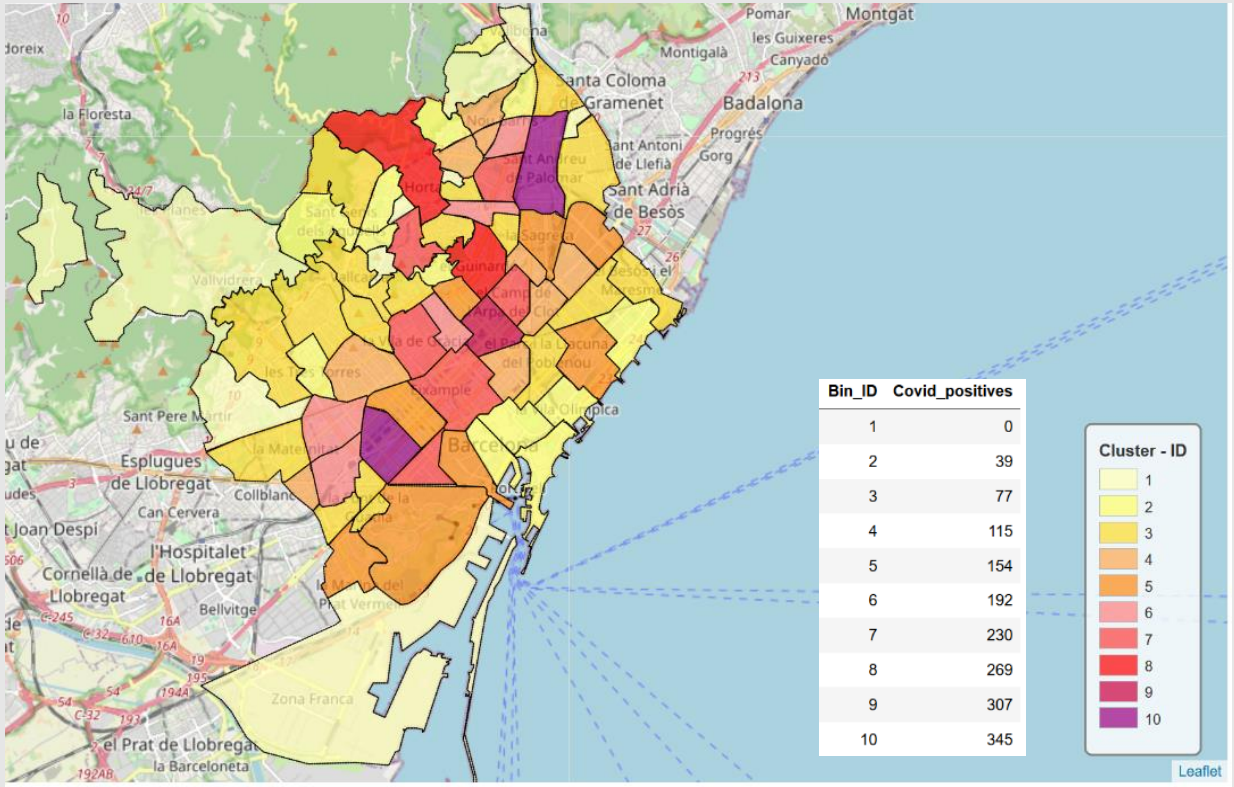
Bin_ID	Unemployment_%
1	2.86
2	3.84
3	4.81
4	5.77
5	6.74
6	7.71
7	8.68
8	9.65
9	10.61
10	11.58

Cluster - ID
1
2
3
4
5
6
7
8
9
10



	January	February	March	April	May	June	July	August	September	October	November	December
count	73.00	73.00	73.00	73.00	73.00	73.00	73.00	73.00	73.00	73.00	73.00	73.00
mean	6.92	6.92	6.99	6.83	6.68	6.50	6.56	6.70	6.62	6.78	6.86	6.78
std	2.17	2.20	2.27	2.18	2.15	2.09	2.05	1.96	1.96	2.07	2.14	2.08
min	2.90	2.90	2.90	2.90	2.80	2.70	2.70	2.90	2.80	2.80	2.90	2.90
25%	5.80	5.60	5.60	5.50	5.50	5.30	5.40	5.70	5.60	5.70	5.70	5.70
50%	6.60	6.60	6.70	6.60	6.40	6.30	6.30	6.50	6.40	6.50	6.50	6.50
75%	7.90	7.90	8.00	7.90	7.70	7.50	7.60	7.70	7.70	7.80	7.80	7.70
max	12.80	12.90	13.20	12.90	12.60	12.20	12.20	11.90	12.00	12.40	12.80	12.70

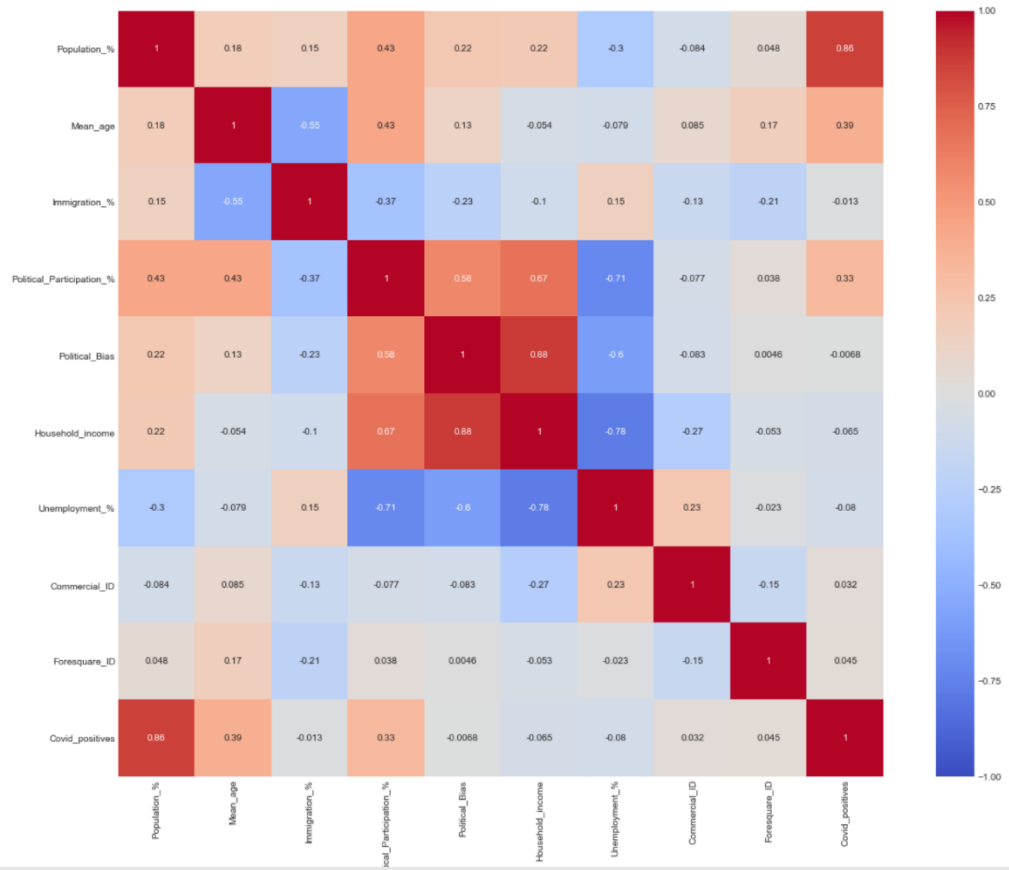
Covid-19 division



Bin_ID	Covid_positives
1	0
2	39
3	77
4	115
5	154
6	192
7	230
8	269
9	307
10	345

Cluster - ID
1
2
3
4
5
6
7
8
9
10

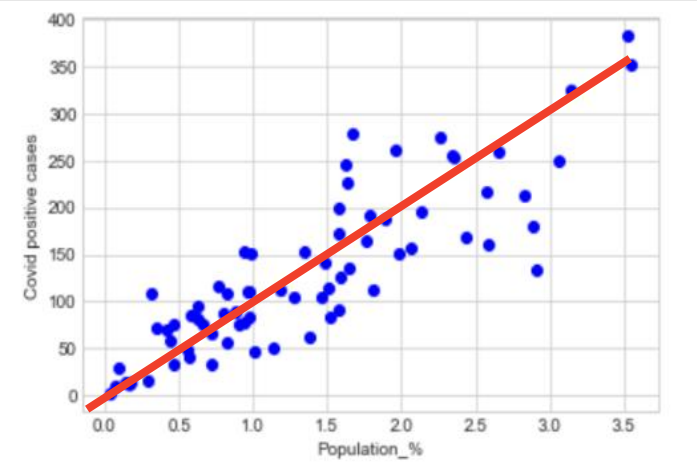
Correlation matrix



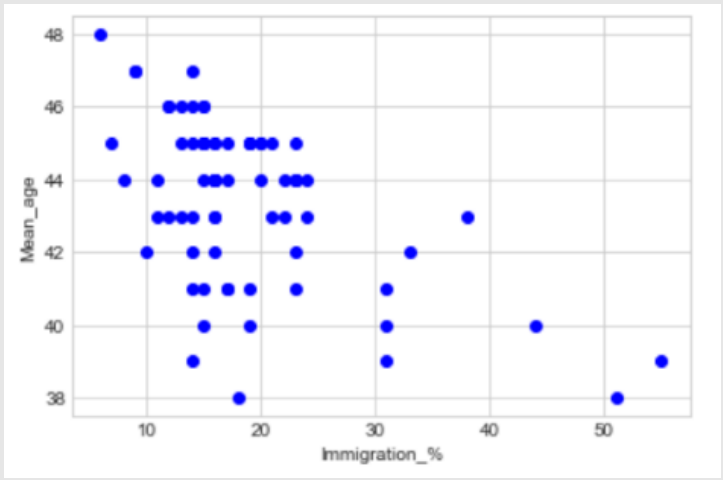
By category and considering correlations absolute values over 0.5, we can observe:

- Population %: Logically, we observe a high correlation with the amount of positive cases in covid.
- Mean_age: There is a weak negative correlation between mean age and immigration %.
- Political Participation: We can see a mild correlation of political participation with political bias and household income. There is also a negative correlation with unemployment %.
- Political bias: There is a high correlation between the political bias and the household income. There is also a negative correlation with unemployment %.
- Household income: In addition to the already mentioned, there is a logical strong negative correlation with the unemployment %.
- In terms of Foresquare in relation to official data, there is absolute non correlation, and therefore we can discard Foresquare as a good option for clustering and representing the city.

Results (I)

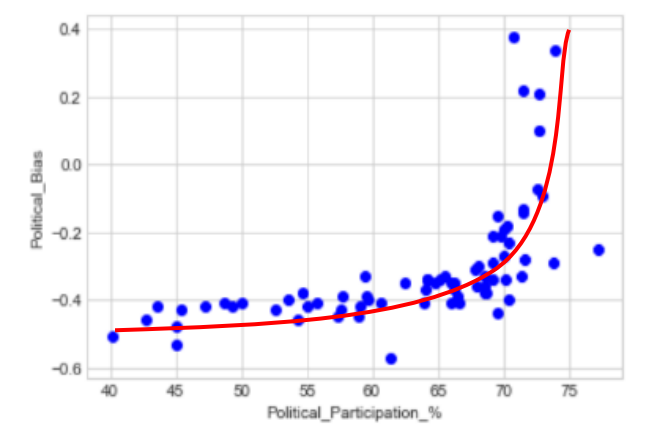


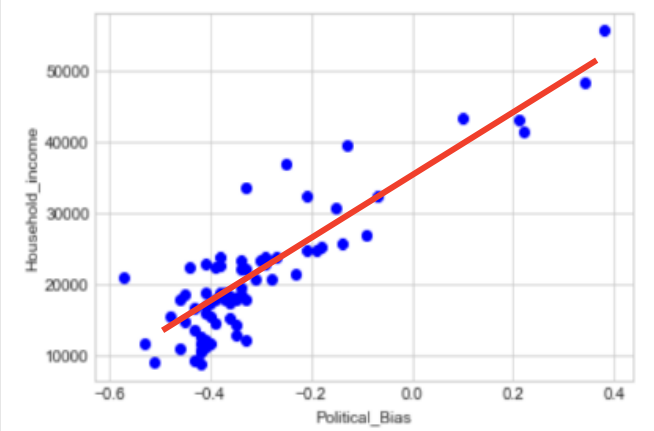
We can observe almost a lineal relationship.

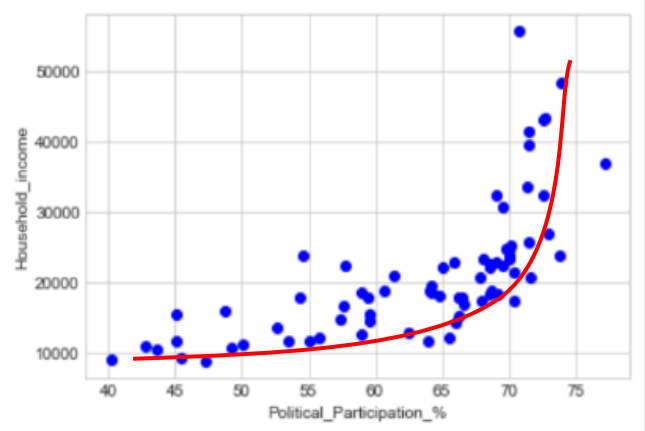


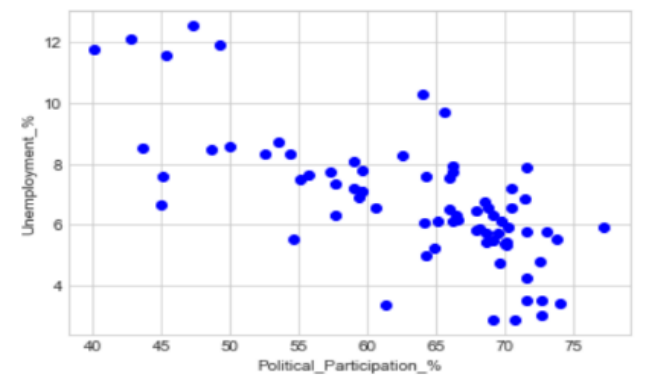
Tendency is not as clear, although the maximums draw a diagonal upper bound.

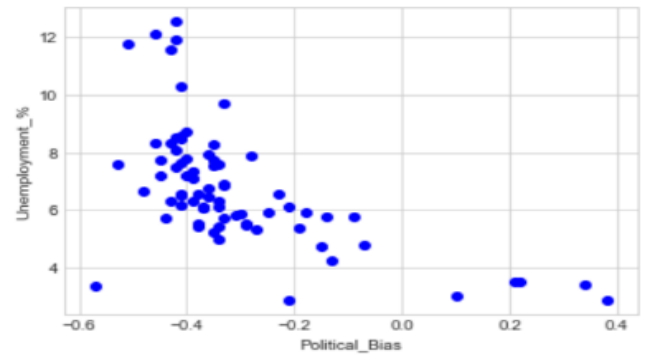
Results (II)

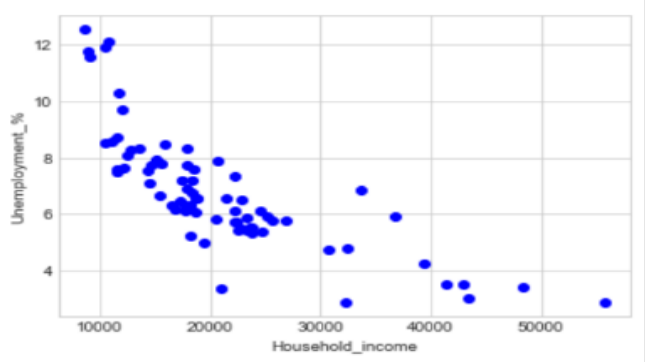

Political_Bias
Political_Participation_%


Household_income
Political_Bias


Household_income
Political_Participation_%


Unemployment_%
Political_Participation_%


Unemployment_%
Political_Bias


Unemployment_%
Household_income

Conclusion and future directions

- Barcelona hardly can be clustered based in all the demographic parameters, as for each of them you get a different painting.

- Correlation was found between the tuple household income and unemployment, the political participation and the political bias.

- Foresquare data not suitable for clustering

❑ A modelling could be applied to estimate the potential number of votes a political party could get in each neighborhood.

❑ This project aims to be repeated 2021 to observe the effects of the coronavirus and test if models based on the mentioned correlations would provide accurate predictions.