ANALYZING SOCIOECONOMIC INDICATORS FOR RELOCATION RECOMMENDATIONS IN VALENCIA

By Yasser Fuentes-Edfuf

BUSINESS UNDERSTANDING

- Valencia is one of the principal business hub of Spain.
- Cost of living in Valencia is higher than in other Spanish cities, but the socioeconomic differences between the different neighborhoods of the city can be used to find affordable places to live.
- A Company wants to analyze socioeconomic data from Valencia to offer insights to prospective employees to help them find accommodation in the city

BUSINESS UNDERSTANDING

The key indicators employed to analyze Madrid's neighborhoods will be:

- Population
- Average income
- Crime level
- Amenities in the neighborhood
- Real estate and rent prices (per square meter)

The neighborhoods will be segmented and classified according to those features

BUSINESS UNDERSTANDING

Our audience:

- New employees from the company wishing to move to Madrid and know a Little bit about the city before moving in.
- Company's management expects to understand the rationale behind the recommendations made.
- The general public could be also benefitted from this information.

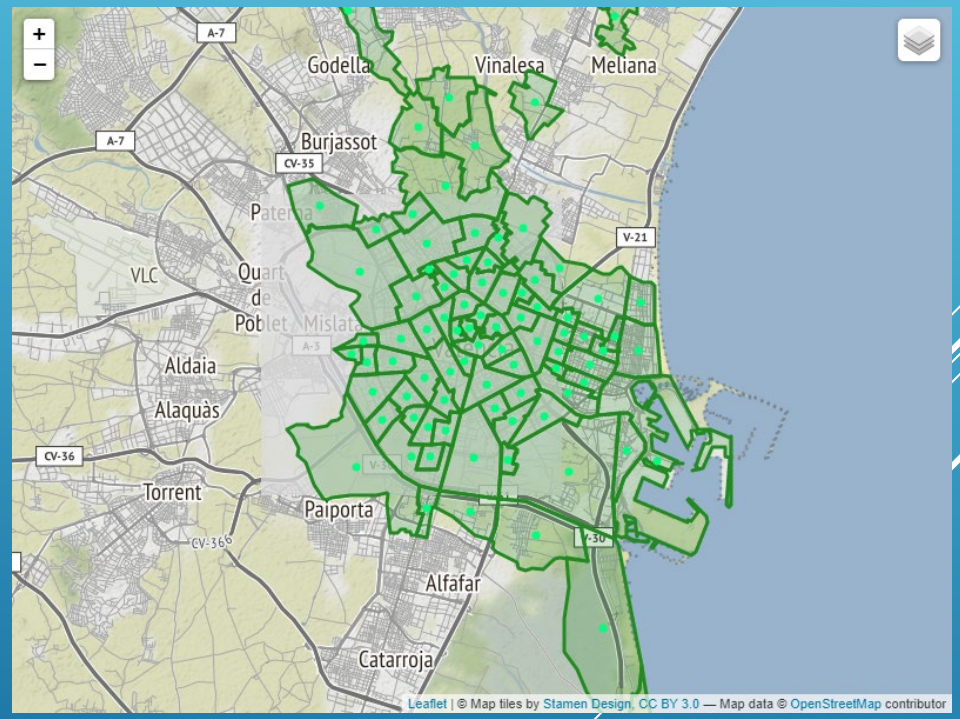
BUSINESS UNDERSTANDING

Success criteria:

- The project will be considered successful if a list of Valencia neighborhoods based on socioeconomic and business diversity in the neighborhood can be presented to the client to inform its prospective employees of their living choices in the city.

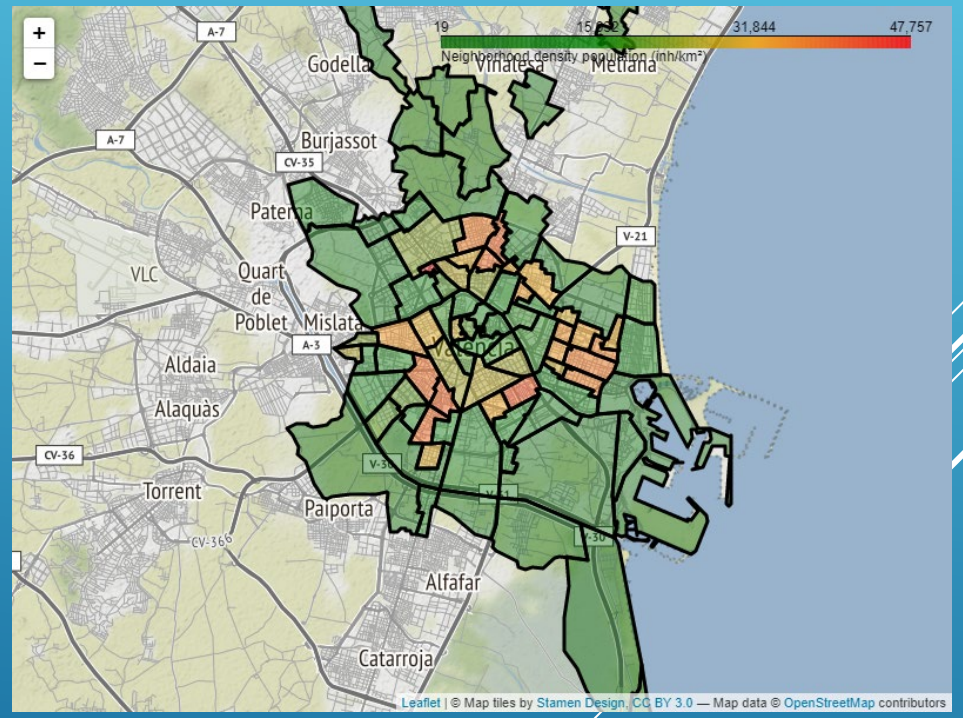
DATA ETL –GEOGRAPHICAL DATA

- We will be using an API from mapas.valencia.es for districts and neighborhoods. Those files are in Geography Markup Language File (.gml) and in EPSG:25830.
- We will transform them into a much more convenient geojson format. We will be using geopandas, a pandas like module able to work with shapes.
- We will project the shapes of the neighborhoods and districts to calculate their areas.and centroids foreach neighborhood shape.



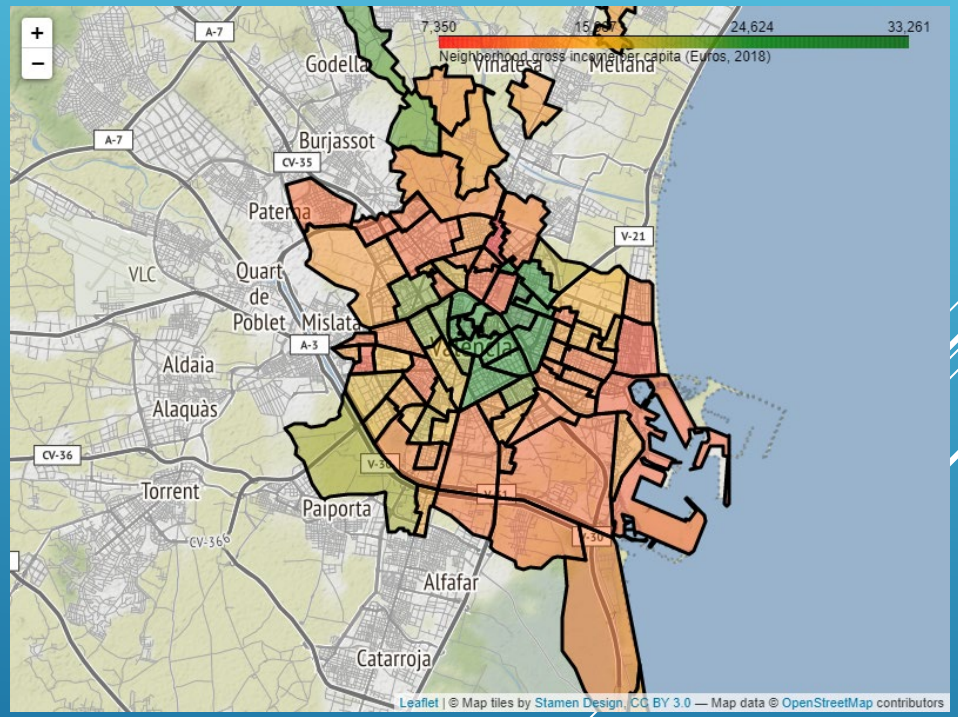
DATA ETL –POPULATION DATA

- Population data for the city of Valencia is taken from the data bank of the City Council of Valencia, here.
- The information extracted is the number of inhabitants per neighborhood block.
- We aggregate the population. and iterate over each Neighborhood of Valencia



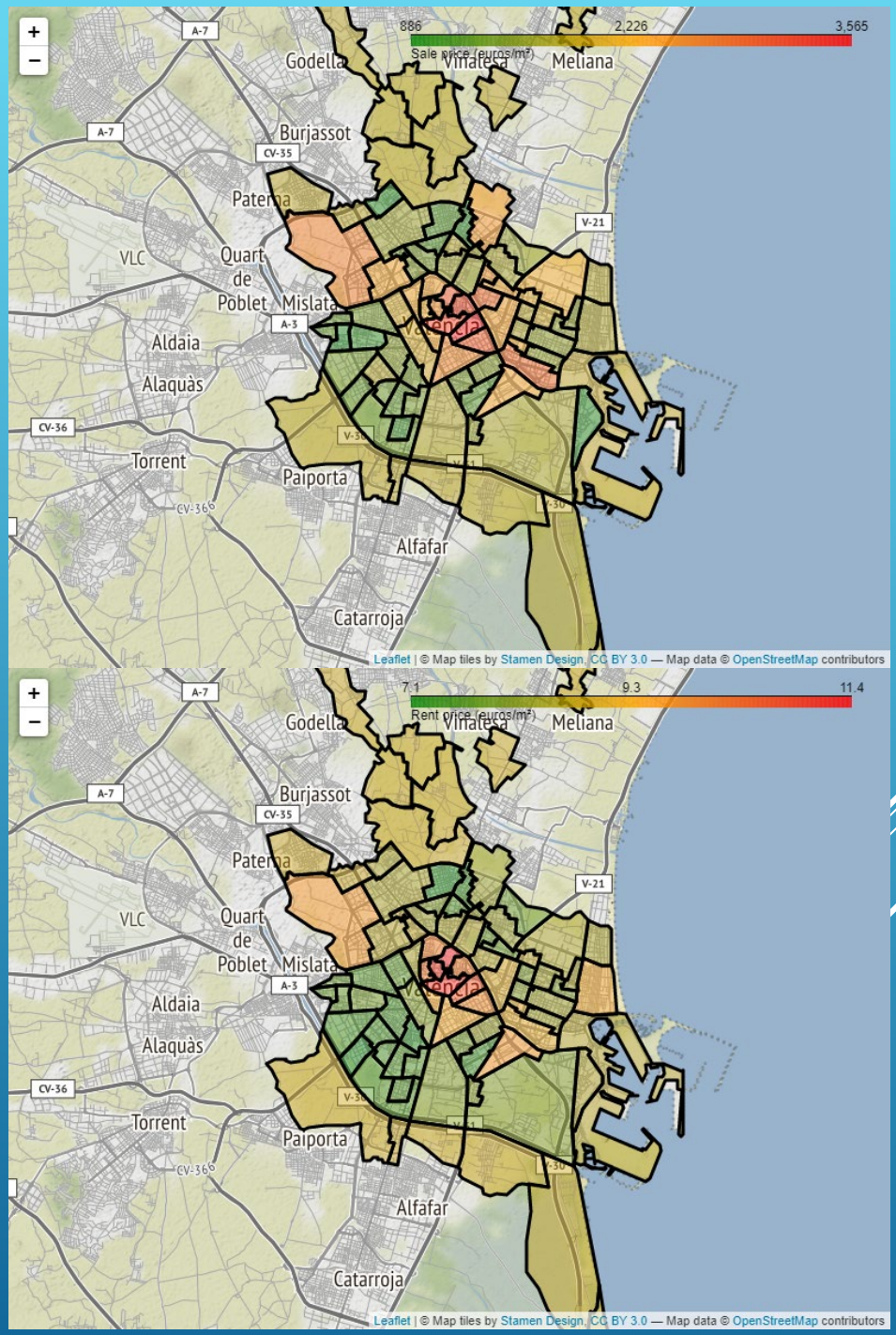
DATA ETL –INCOME DATA

- Data corresponding to the mean gross income per person available it taken from Spain's Instituto Nacional de Estadistica, INE, (National Statistics Institute).
- The downloaded file is a json with all the district data for the city of Valencia.
- This data was taken at a Neighborhood level



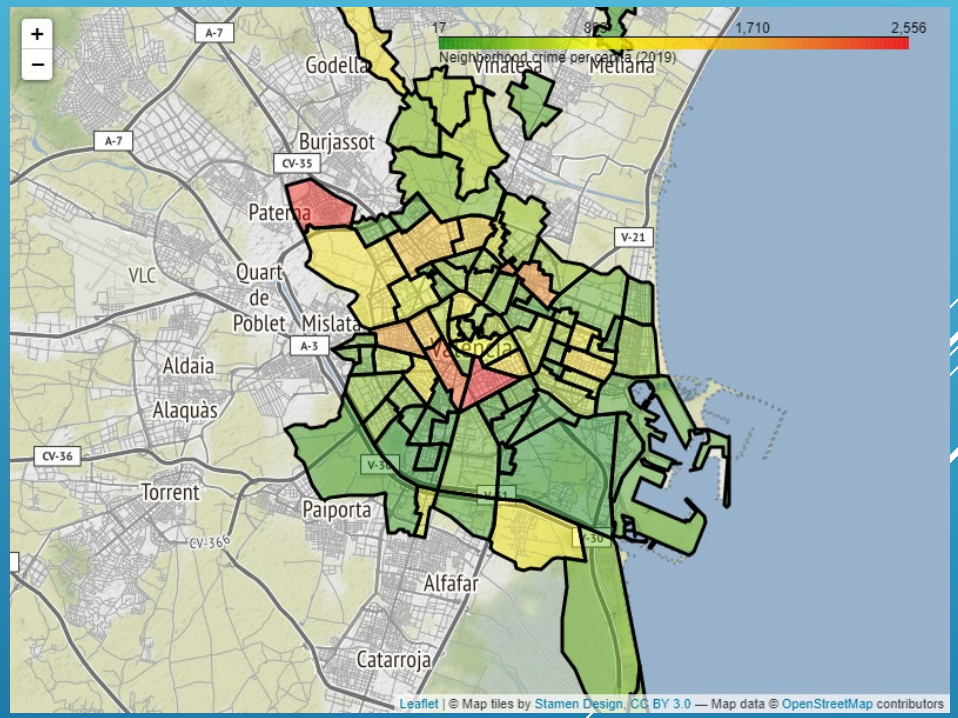
DATA ETL –REAL STATE DATA

- The information for the average price of the square meter at a neighborhood level will be scraped from Idealista data website.
- We collect data regarding sale and rent price per sqr meter



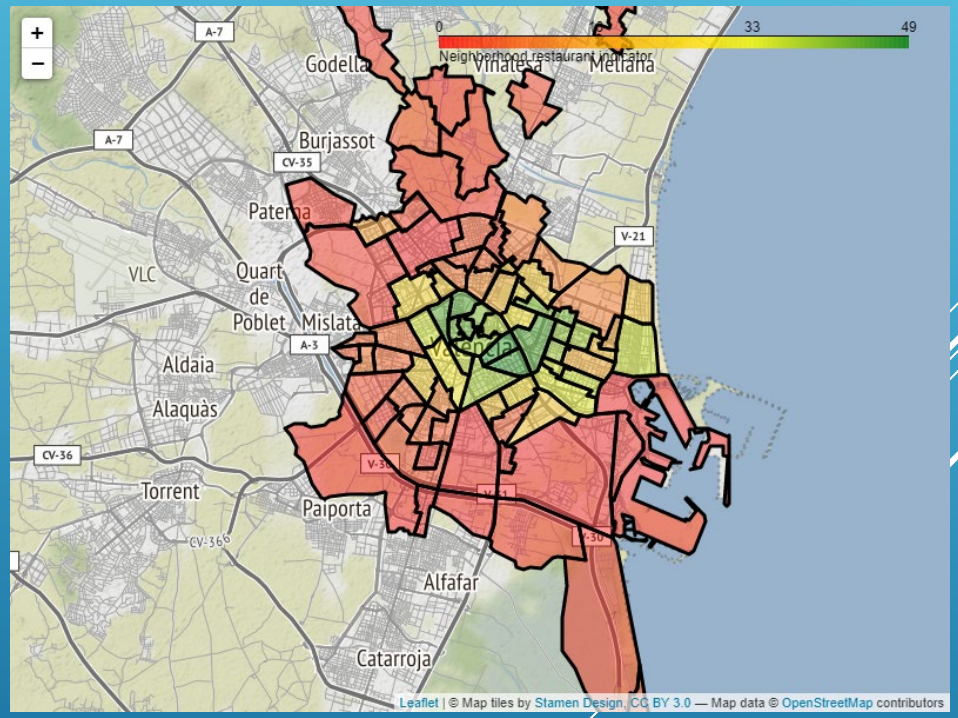
DATA ETL –CRIME DATA

- Crime data for 2019 is taken from a statistical review of the city of Valencia from the Valencia council here (SP).
- Data disaggregated at district level
- We transform the data by district dividing it by the population of each neighborhood

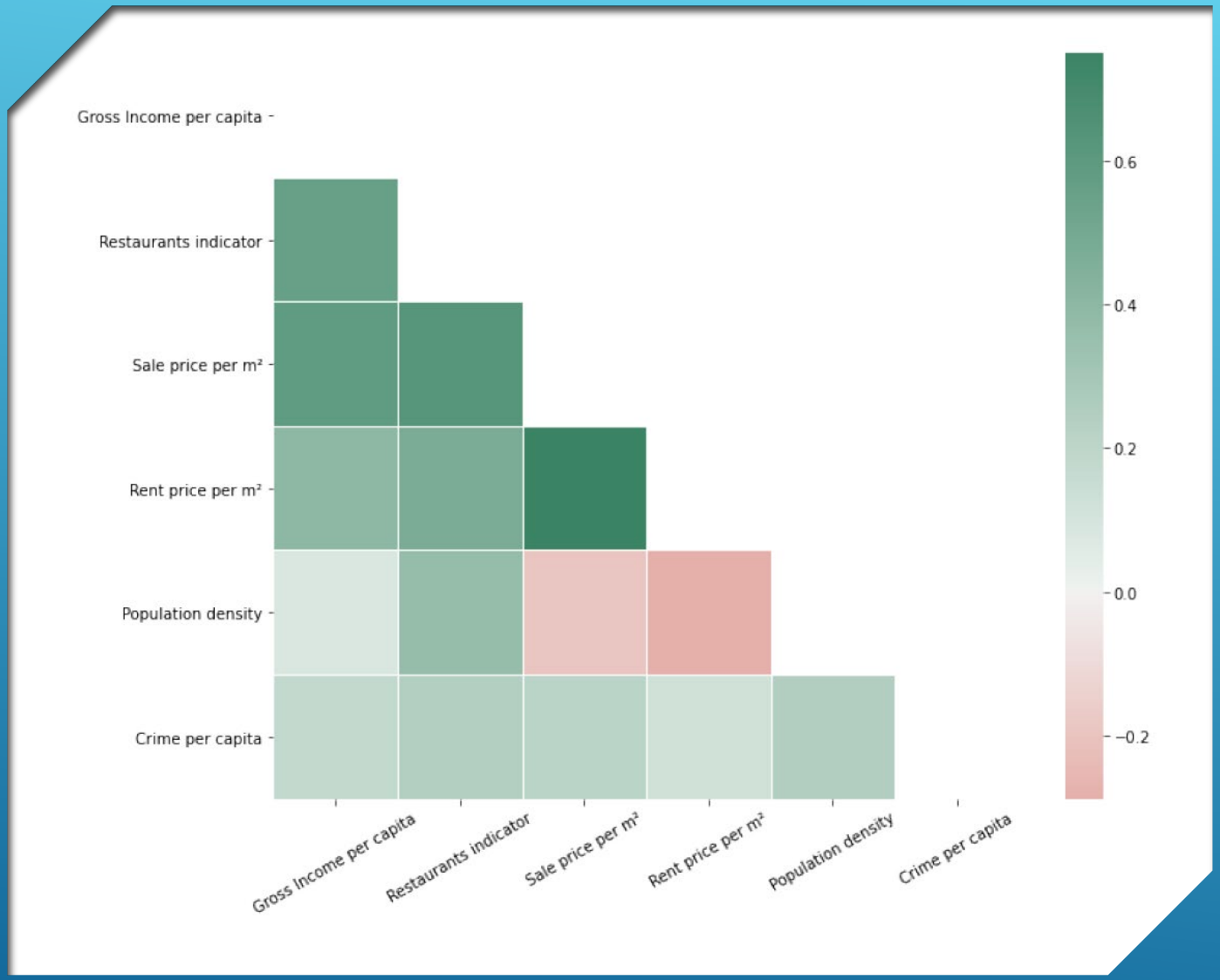


DATA ETL –AMENITIES DATA

- The number of venues in each neighborhood and district will be taken from Foursquare usingtheir API. This data will be used as a proxy of economic activity in each neighborhood.
- Restaurants can be used as a proxy of socioeconomic activities in a neighborhood in absence of other data[1]



[1]"Predicting neighborhoods' socioeconomic attributes using restaurant data" by Lei Dong, Carlo Ratti, and Siqi Zheng. PNAS July 30, 2019 116 (31) 15447-15452.



DATA ETL – CORRELATION

- Correlation is between Sale and Rent prices, most expensive real estate for buying is also expensive for renting.
- We find a correlation between amenities (Restaurant indicator) with gross income per capita. Also, this two indicators have a positive correlation with sales and renting prices. Richer people live in expensive areas where usually have more commodities.
- It's interesting that the population density is no so correlated with the gross income data, but it is inversely correlated with sale and rent prices per square meter: expensive areas are less populated than the cheap ones.
- There is a correlation between population density and the restaurants indicator: more people in the area need more restaurants in the area.
- We notice also that the crime per capita indicator has no great correlation with any of the other indicators. This means that crime is smoothly distributed among Valencia neighborhoods.

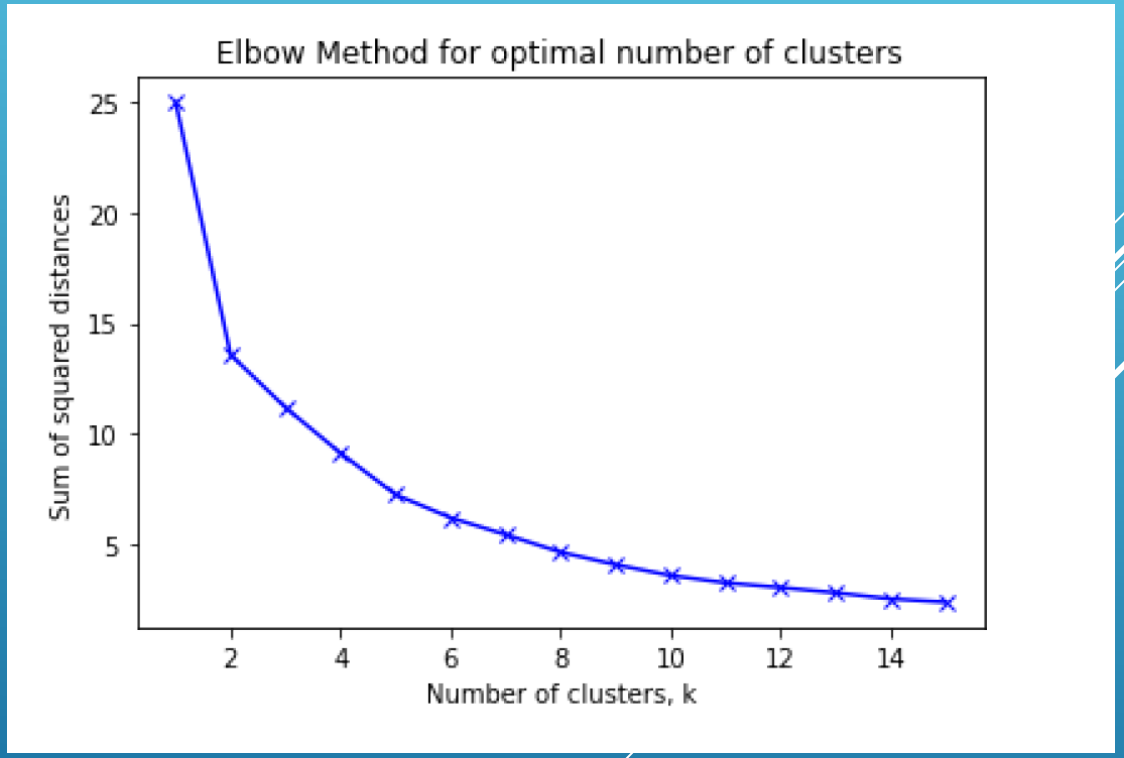
MODELING

Used a K-means clustering algorithm with the following features:

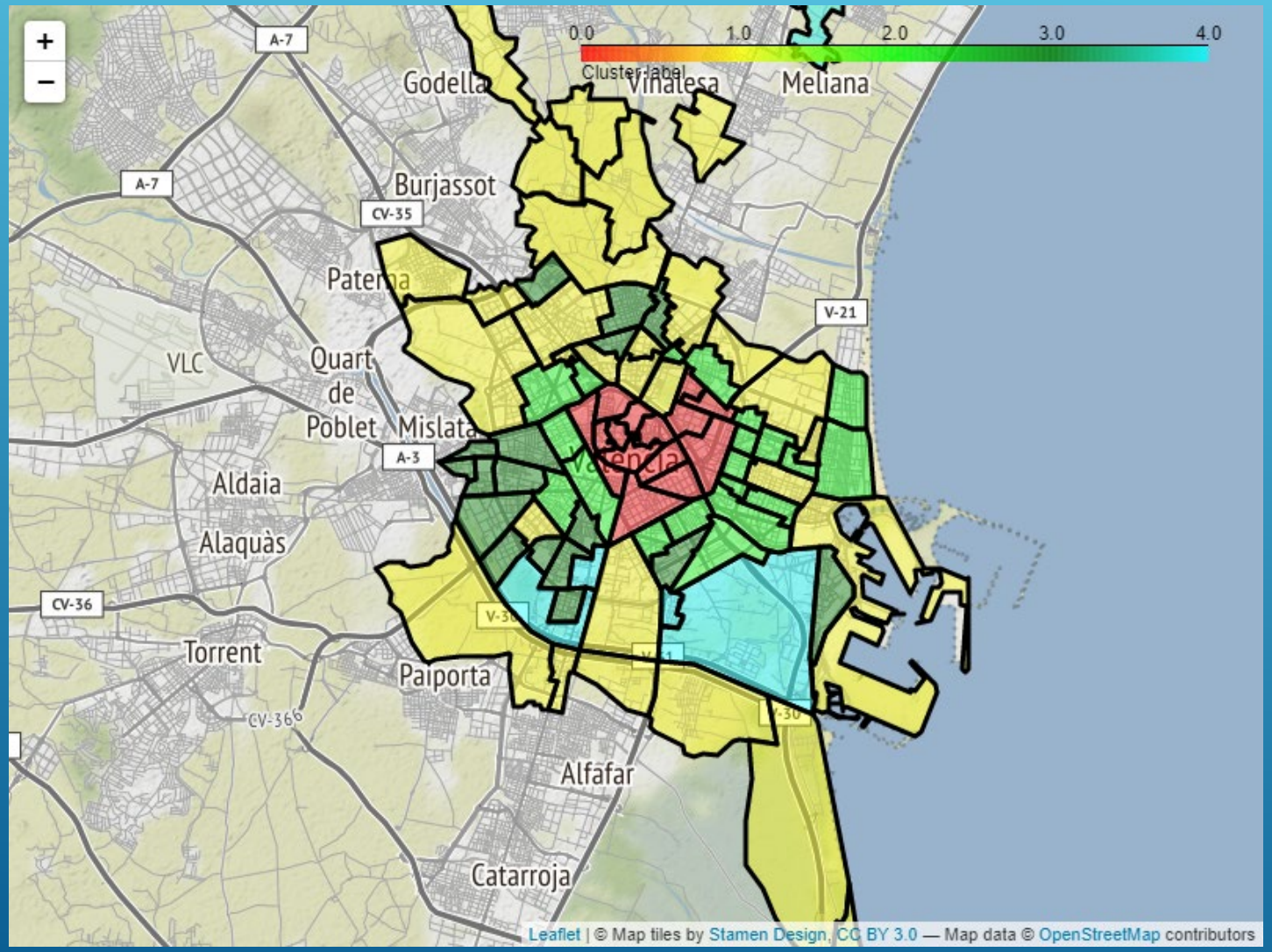
- Real Estate selling prices per square meter.
- Real Estate renting prices per square meter.
- Mean income per person.
- Population density in inhabitants per square kilometer.
- Number of arrests per capita.
- Number of restaurants in the neighborhood.

Data normalized using MinMaxScalermethod.

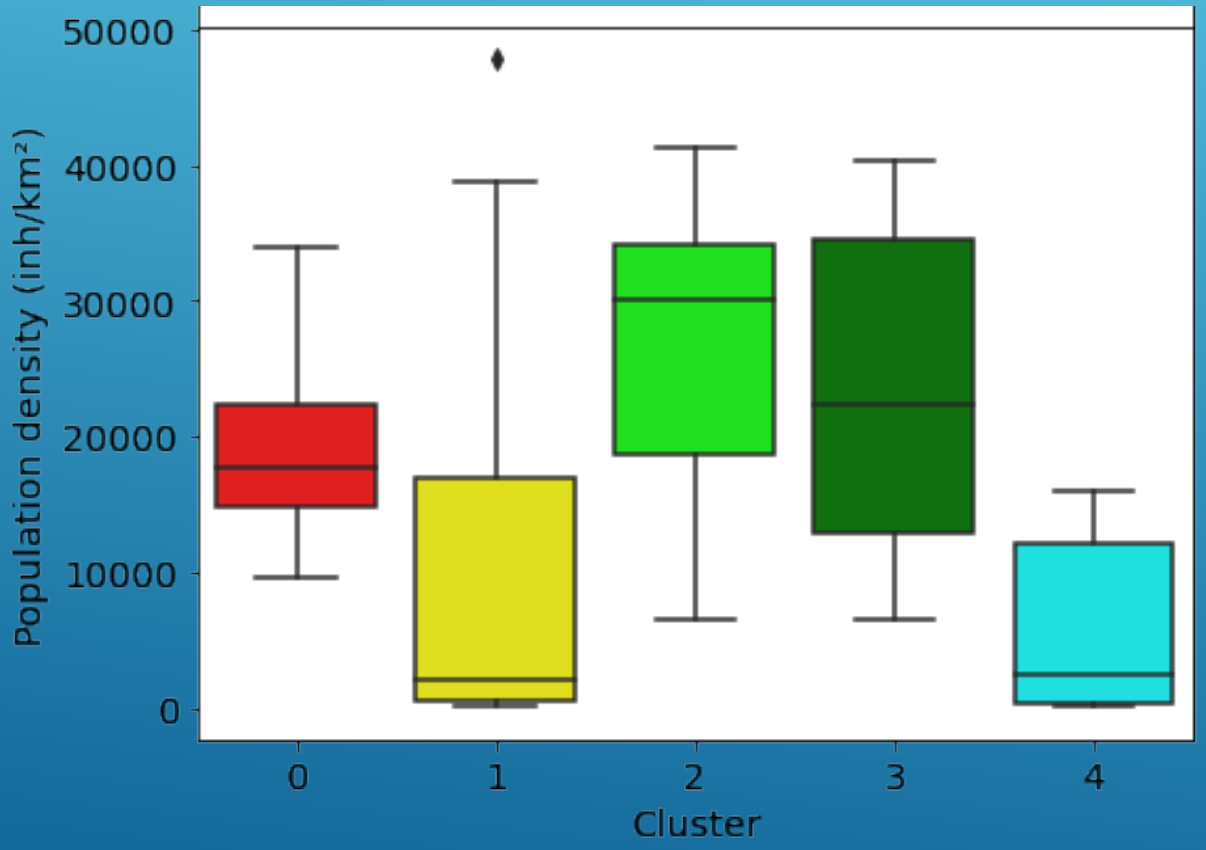
Optimal number of clusters calculated using the elbow method.

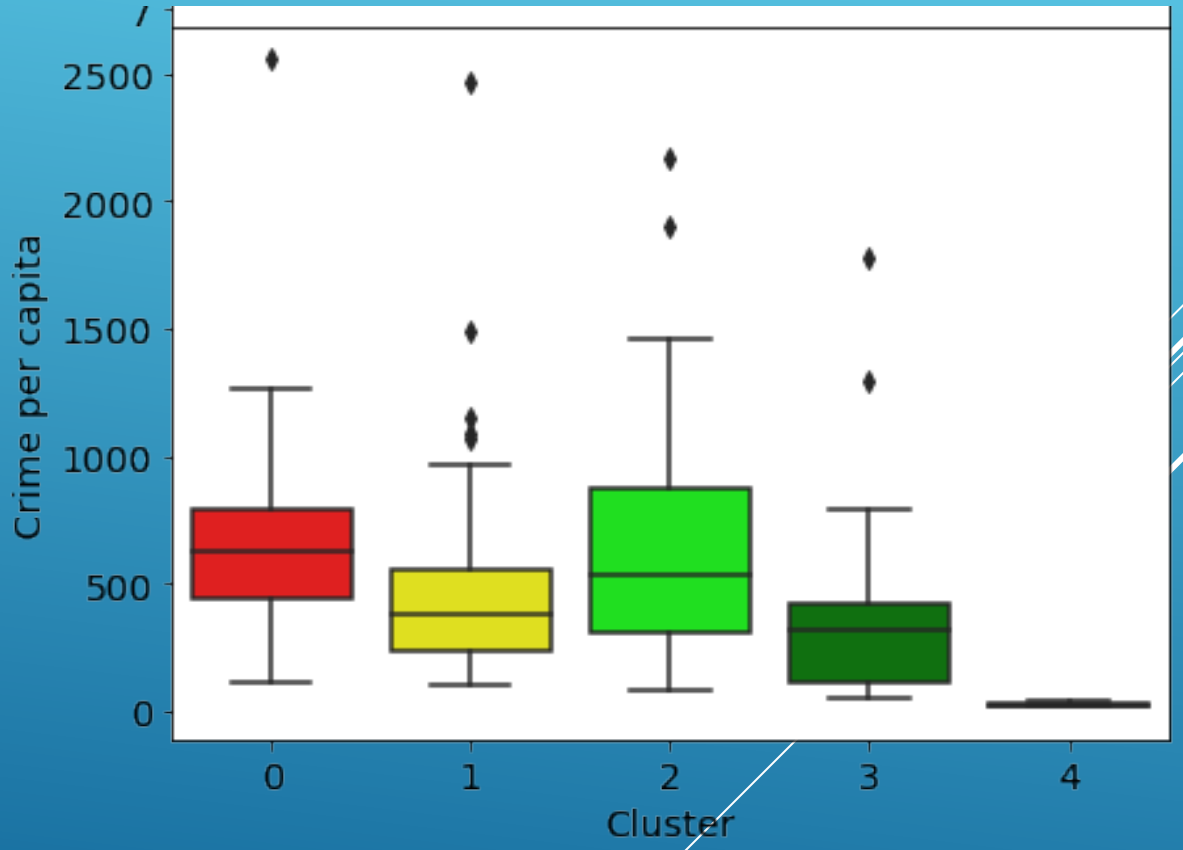


MODELING

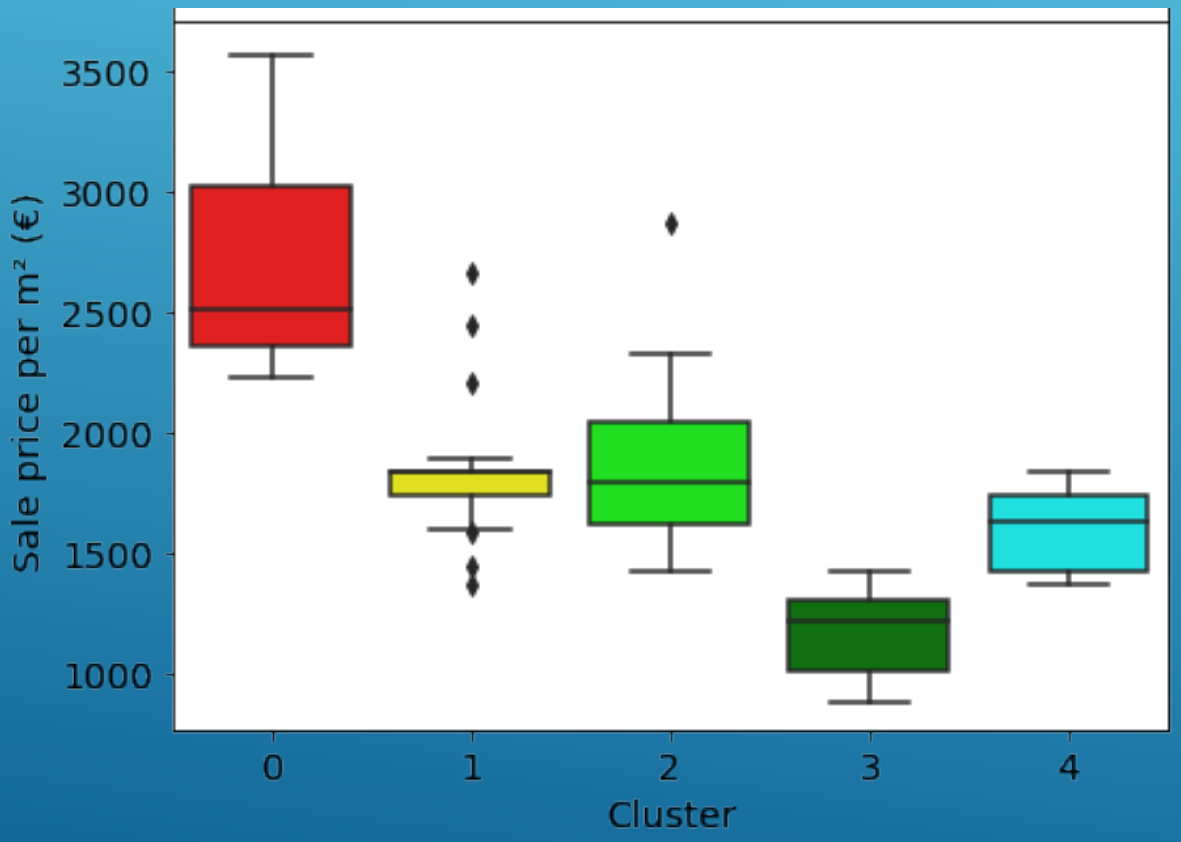

0.0
1.0
2.0
3.0
4.0
Cluster Label
Godella
Vinalesa
Meliana
Burjassot
Paterna
Quart de Poblet
Mislata
VLC
Aldaia
Alaquàs
Torrent
Paiporta
Alfafar
Catarroja
A-7
CV-35
V-21
A-3
CV-36
Leaflet | © Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors

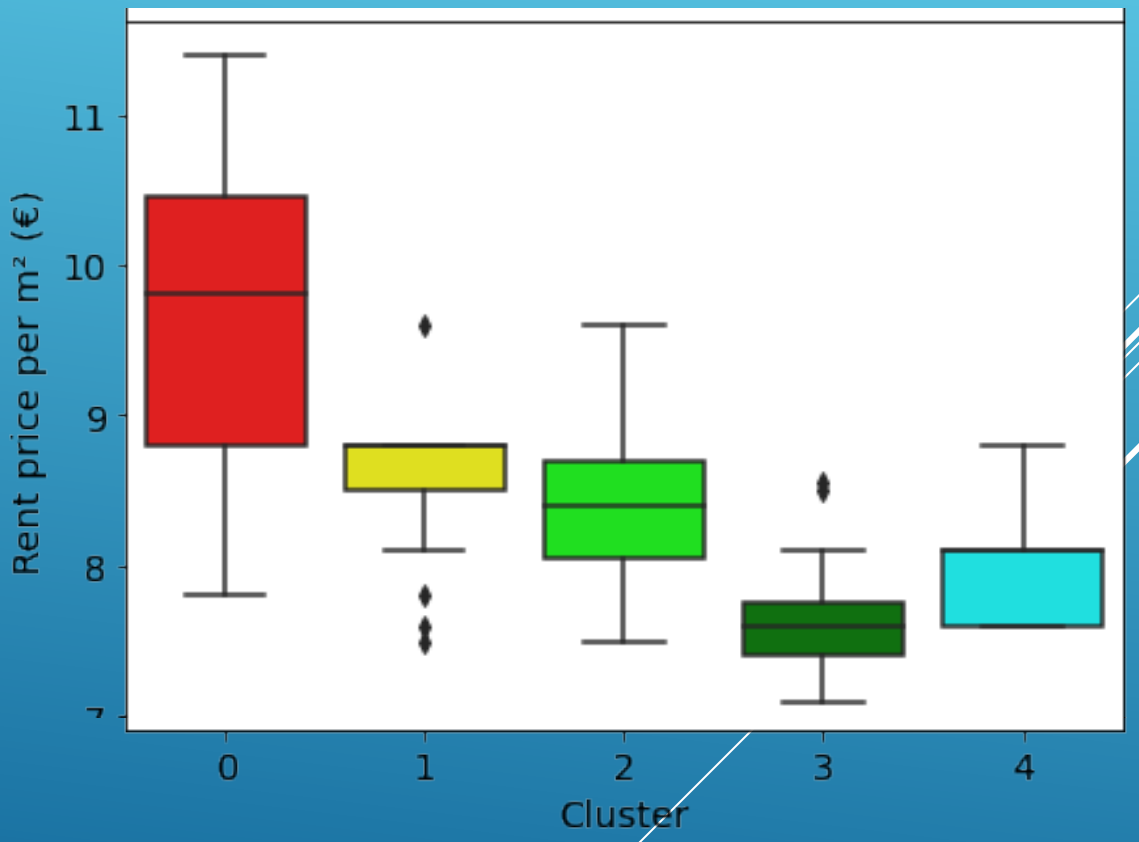
EVALUATION


Population density (inh/km²)
50000
40000
30000
20000
10000
0
0
1
2
3
4
Cluster

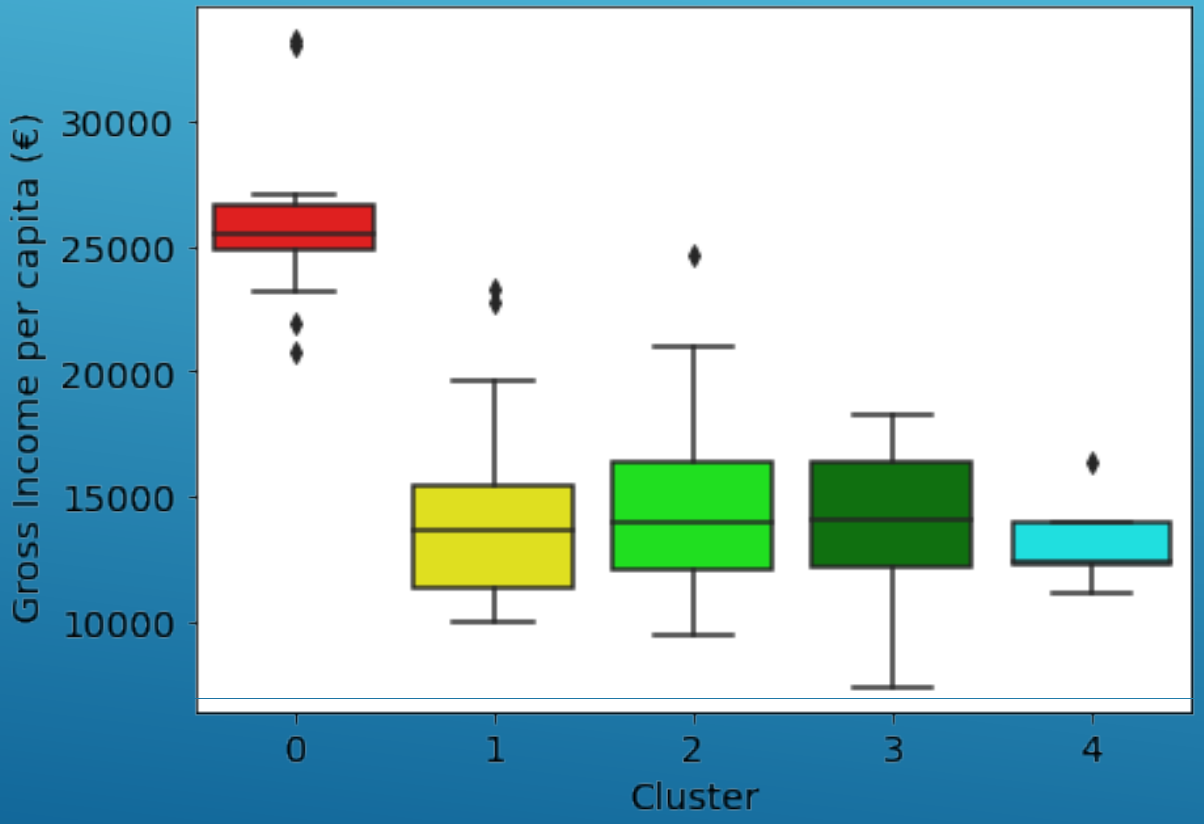

Crime per capita
2500
2000
1500
1000
500
0
0
1
2
3
4
Cluster

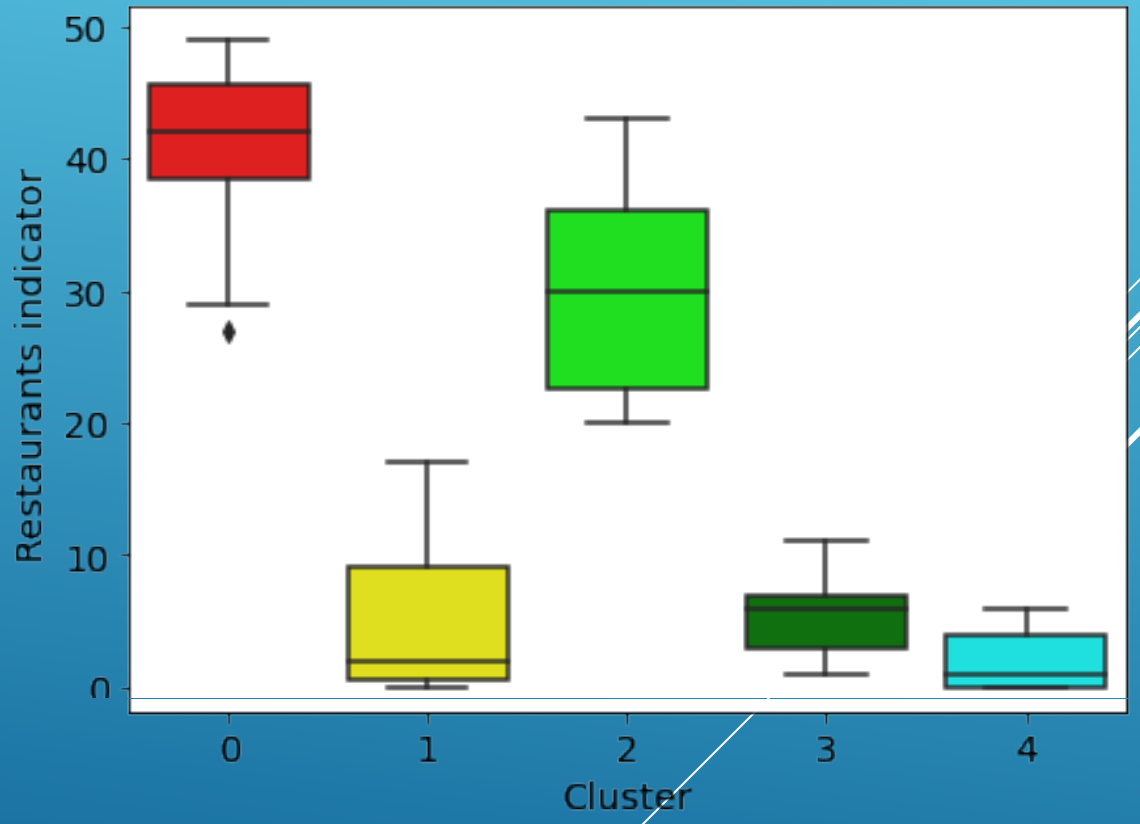
EVALUATION


Sale price per m² (€)
3500
3000
2500
2000
1500
1000
0
1
2
3
4
Cluster


Rent price per m² (€)
11
10
9
8
0
1
2
3
4
Cluster

EVALUATION


Gross Income per capita (€)
30000
25000
20000
15000
10000
0
1
2
3
4
Cluster


Restaurants indicator
50
40
30
20
10
0
0
1
2
3
4
Cluster

EVALUATION

1. Neighborhoods in the first cluster represents regular outer neighborhoods, with the lowest prices on sale and rent and few commodities. This neighborhoods are heavy populated but safe.

2. Second cluster represent richer parts of the city, at city center. All neighborhoods it contains have the highest prices on sale and rent, thus, its population is the wealthier of the city. This clusters has an high quantity of commodities.

3. The third cluster is composed only by three neighborhoods, and is defined by the lower density/crime among all. This is due to that those neighborhoods are almost vast green areas.

4. In the forth cluster represents the second richer parts of the city. Although income per capita and sale/rent prices are in the mean values, this cluster has many commodities and it is heavy populated,We can state that people living in this cluster is most conformed by the middle class of Valencia inhabitants.

5. The fifth cluster show lower class neighborhoods, with few amenities. Notice that the median rent in this neighborhoods is higher than the sale. Provably it is due to that most of people there are not landlords but renters.

CONCLUSION

- **Classify the neighborhoods of Valencia based on socioeconomic and business diversity in order to give information about living conditions in Valencia**
- **Data taken from several official sources, realtors and Foursquare**
- **The features were used to segment the neighborhoods into three clusters**

FUTURE WORK

- Historical data for analysis:
  – A derivative can be perform on the time series of the indicators, thus the momentum of its evolution can help to predicted the tendencies

- Selection of neighborhoods by assign a custom weight on the indicators:
  – By this manner, reader could select the ones more interesting to him by, for example, minimize the crime and sale prices and maximize the commodities

- Custom POIs to improve the segmentation
  – Adding a POI and explore new segmentations minimizing the distance to it