

# Refurbishment of empty buildings into cultural, entrepreneurial and education hubs in Madrid

Applied Data Science Capstone by IBM/Coursera

# Table of contents

- ▶ Introduction: Business Problem
- ▶ Data
- ▶ Methodology
- ▶ Analysis
- ▶ Results and Discussion
- ▶ Conclusion

# INTRODUCTION: BUSINESS PROBLEM

The city council of Madrid (Spain) counts with 57 empty buildings of its property. As part of the social development strategy, these buildings will be converted into cultural, entrepreneurial and education hubs over the next four years. In order to know which venues should be reconverted sooner, or what type of use should be given to each building, the council has decided to gather data on the areas where each is located.



Once the characterisation of the areas is done, two variables will be added to help with the prioritisation of buildings:

- Household income
- Total population

With this data-based approach, the council will count with reliable information about the buildings and their surroundings. This project thus aims at creating visualization tools that will support the design of an implementation plan.

# DATA

| NAME                                     | FORMAT | SOURCE                                  |
|--|--------|---|
| Location of empty buildings              | .json  | Open Data Portal of Madrid City Council |
| Venues around buildings                  | .json  | Foursquare API                          |
| Census sections of the city of Madrid    | .shp   | Regional Government of Madrid           |
| Household rent per census section 2016   | .xlsx  | Spanish Statistic Institute             |
| Total population per census section 2016 | .xlsx  | Spanish Statistic Institute             |

The census section and the population data have been previously merged with ArcGIS Pro, generating a file in .json format.

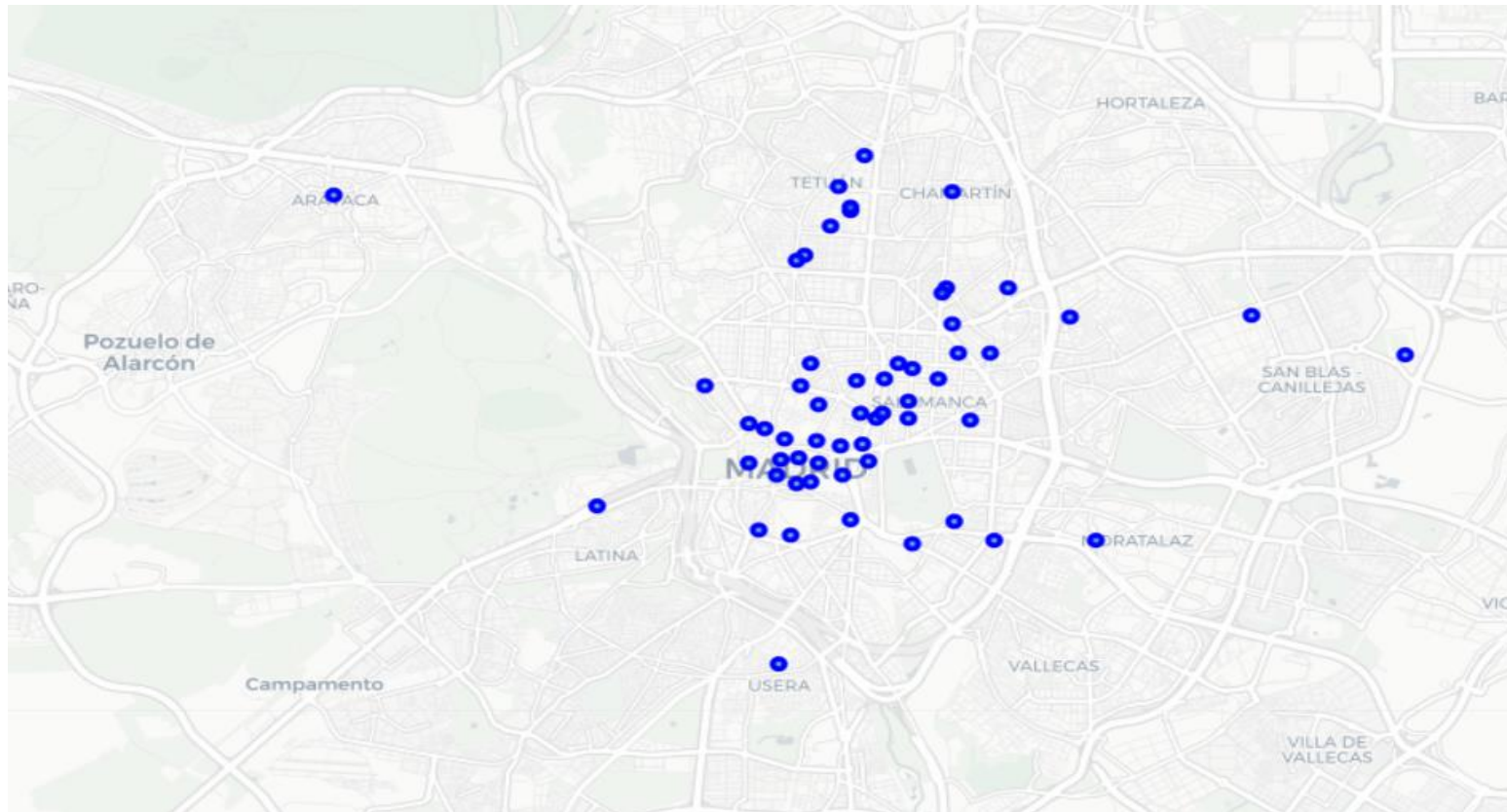
All contents are stored in a GitHub account and can be retrieved from there.

# METHODOLOGY

1. Download the **data on empty buildings** from **Madrid's Open Data Portal** with their **coordinates** and process it until a suitable data frame is obtained for the analysis.
2. **Location of each building** on a map.
3. Gathering of every venue around each building with a **radius of 500m** and a number of venues **limited to 100** with the Foursquare API. Name, coordinates and type will be retrieved for each venue.
4. Analysis of each of the areas where the buildings are located and collection of the **ten most common venues** in each of them.
5. Application of the **K-means** algorithm to cluster the areas. Create a new data frame that includes the cluster as well as the **top 10 venues** for each area.
6. Examination of each cluster and assign a **type to each cluster**. Depending on the type of cluster the buildings are in, they will be converted into **cultural, entrepreneurial or education hubs**.
7. **Production of visualization tools** that can help the city council decide which buildings should be refurbished first. Two variables will be added: **household income and population data per census section**.

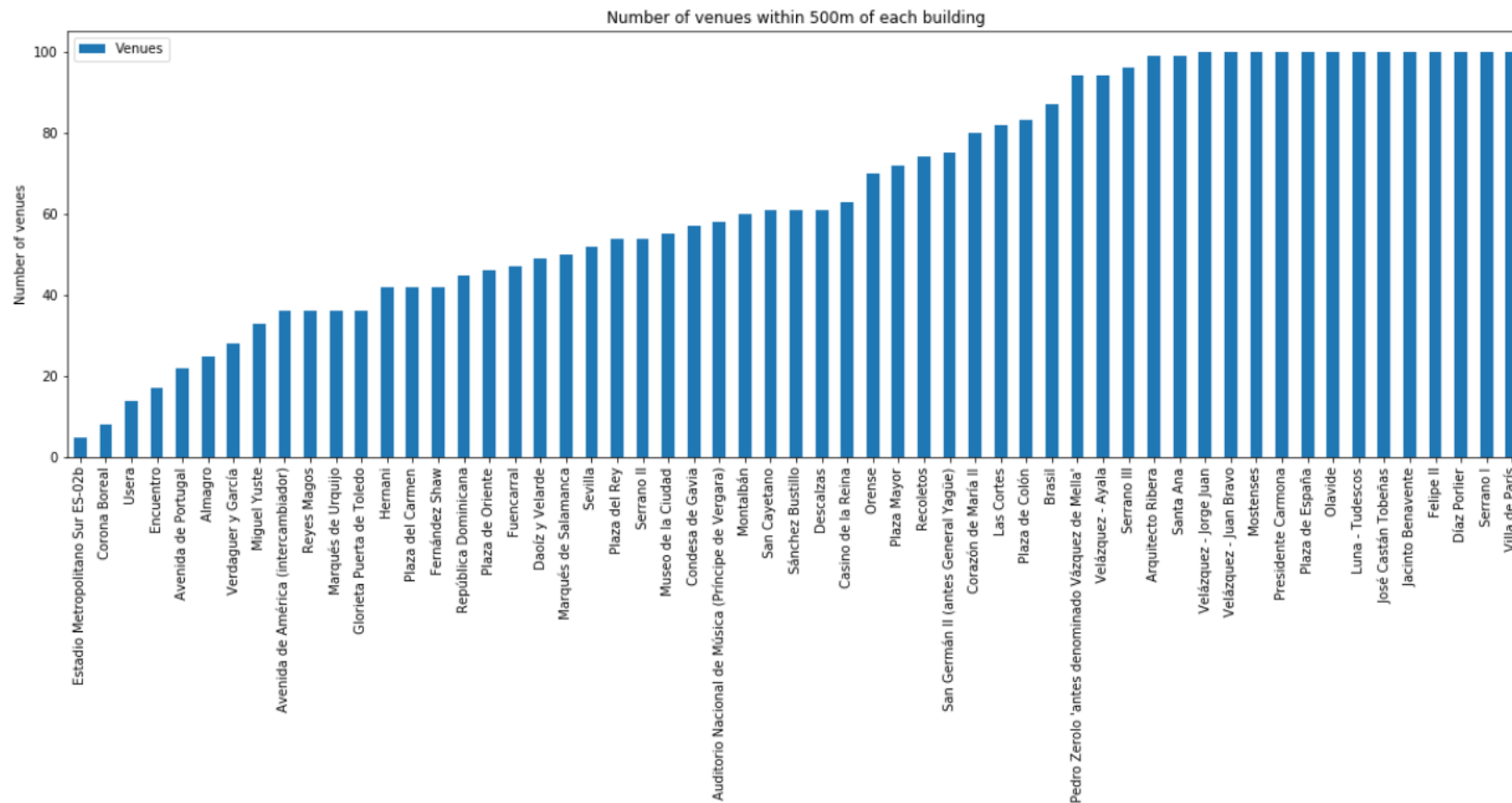
# ANALYSIS

Location of the empty buildings with Folium



# ANALYSIS

Data retrieval on the venues around the buildings with the Foursquare API





# ANALYSIS

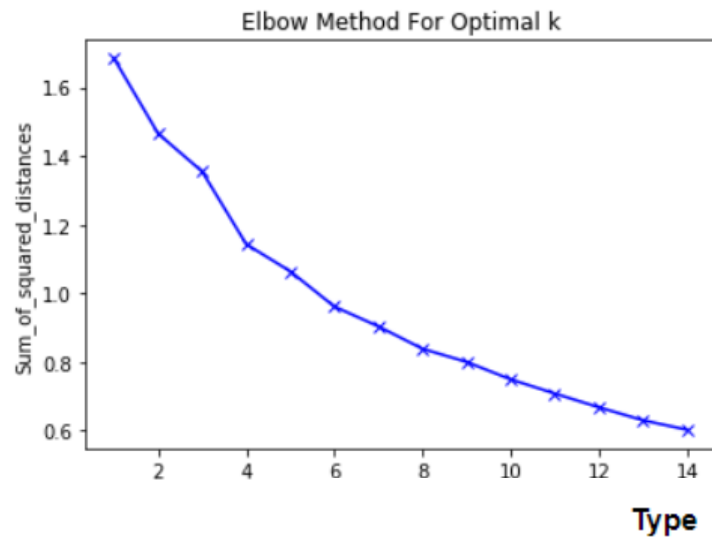
Data wrangling to obtain the ten most popular venues around each buildings

|   | Name   | 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 |
|---|--|-----------------------|-----------------------|-----------------------|--------------------------|-----------------------|-------------------------------|--------------------------|-----------------------|-----------------------|------------------------|
| 0 | Almagro  | Spanish Restaurant    | Italian Restaurant    | Plaza                 | Mediterranean Restaurant | Hotel                 | Café                          | French Restaurant        | Bistro                | Restaurant            | Wine Shop              |
| 1 | Arquitecto Ribera                                | Restaurant            | Bar                   | Spanish Restaurant    | Plaza                    | Hotel                 | Vegetarian / Vegan Restaurant | Tapas Restaurant         | Italian Restaurant    | Bookstore             | Ice Cream Shop         |
| 2 | Auditorio Nacional de Música (Príncipe de Ver... | Spanish Restaurant    | Restaurant            | Bar                   | Grocery Store            | Plaza                 | Nightclub                     | Tapas Restaurant         | Japanese Restaurant   | Café                  | Supermarket            |
| 3 | Avenida de América (intercambiador)              | Spanish Restaurant    | Indian Restaurant     | Café                  | Hotel                    | Seafood Restaurant    | Supermarket                   | Mediterranean Restaurant | Metro Station         | Gastropub             | Sushi Restaurant       |
| 4 | Avenida de Portugal                              | Spanish Restaurant    | Grocery Store         | Athletics & Sports    | Stadium                  | Restaurant            | Pizza Place                   | Supermarket              | Farmers Market        | Tapas Restaurant      | Bar                    |

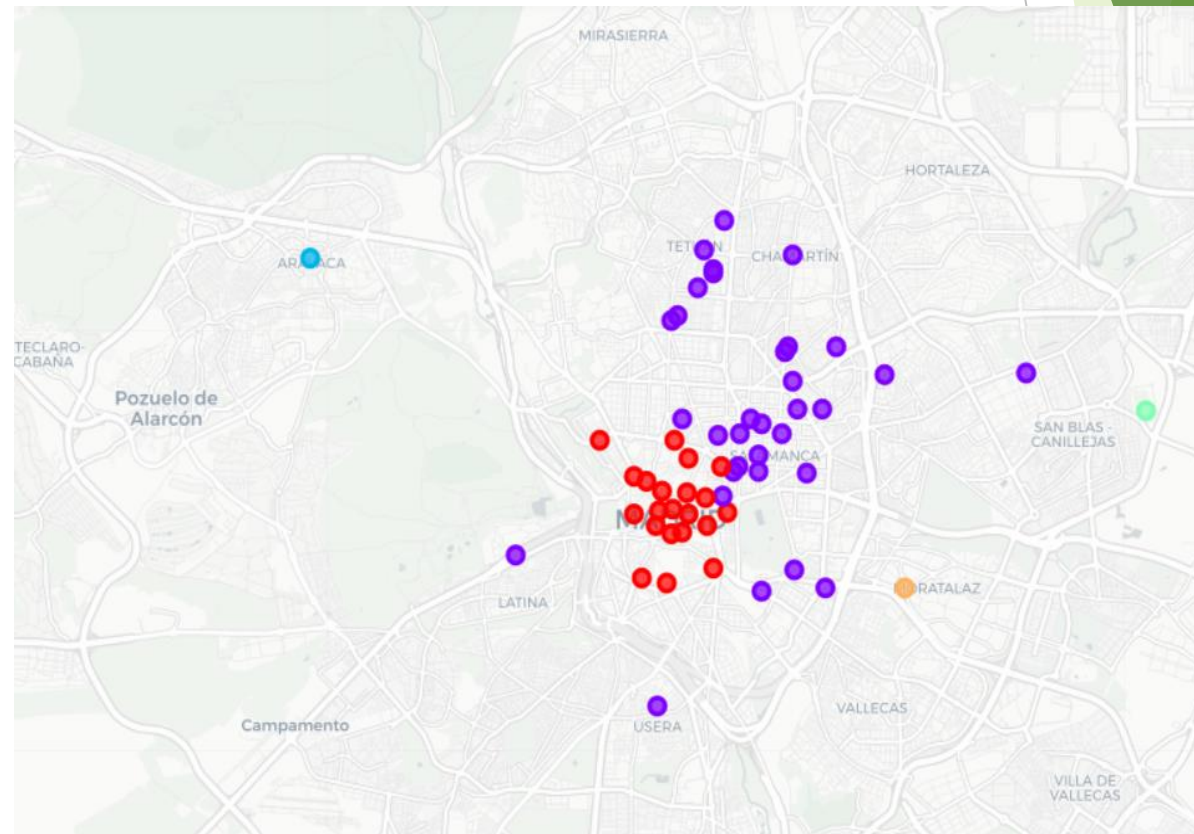


# ANALYSIS

Application of K-Means to generate 5 clusters and classification of each one

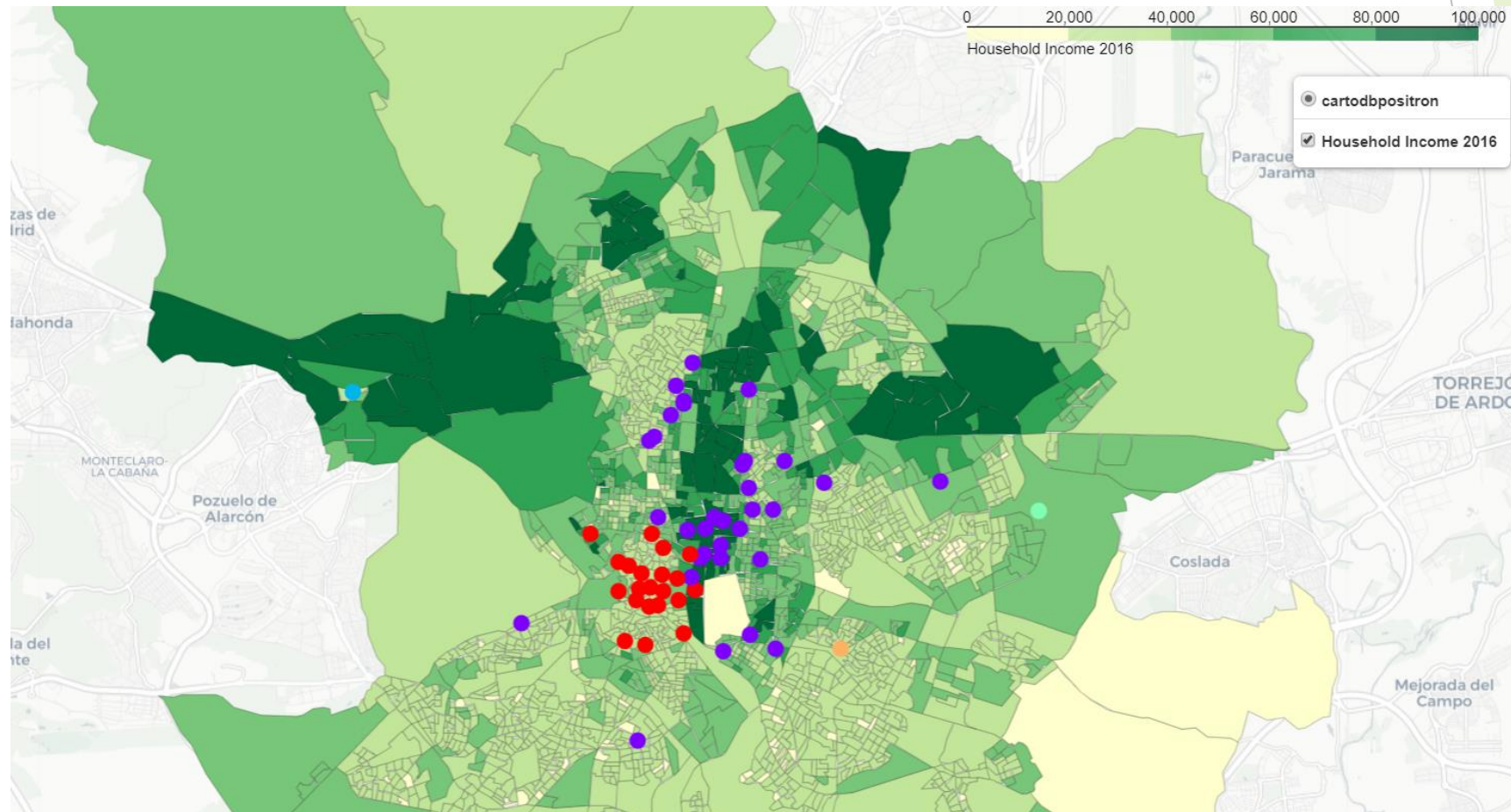


|           |                    |
|-----------|--------------------|
| Cluster 0 | Entertainment area |
| Cluster 1 | Park area          |
| Cluster 2 | Residential area   |
| Cluster 3 | Airport area       |
| Cluster 4 | Coffee shops area  |



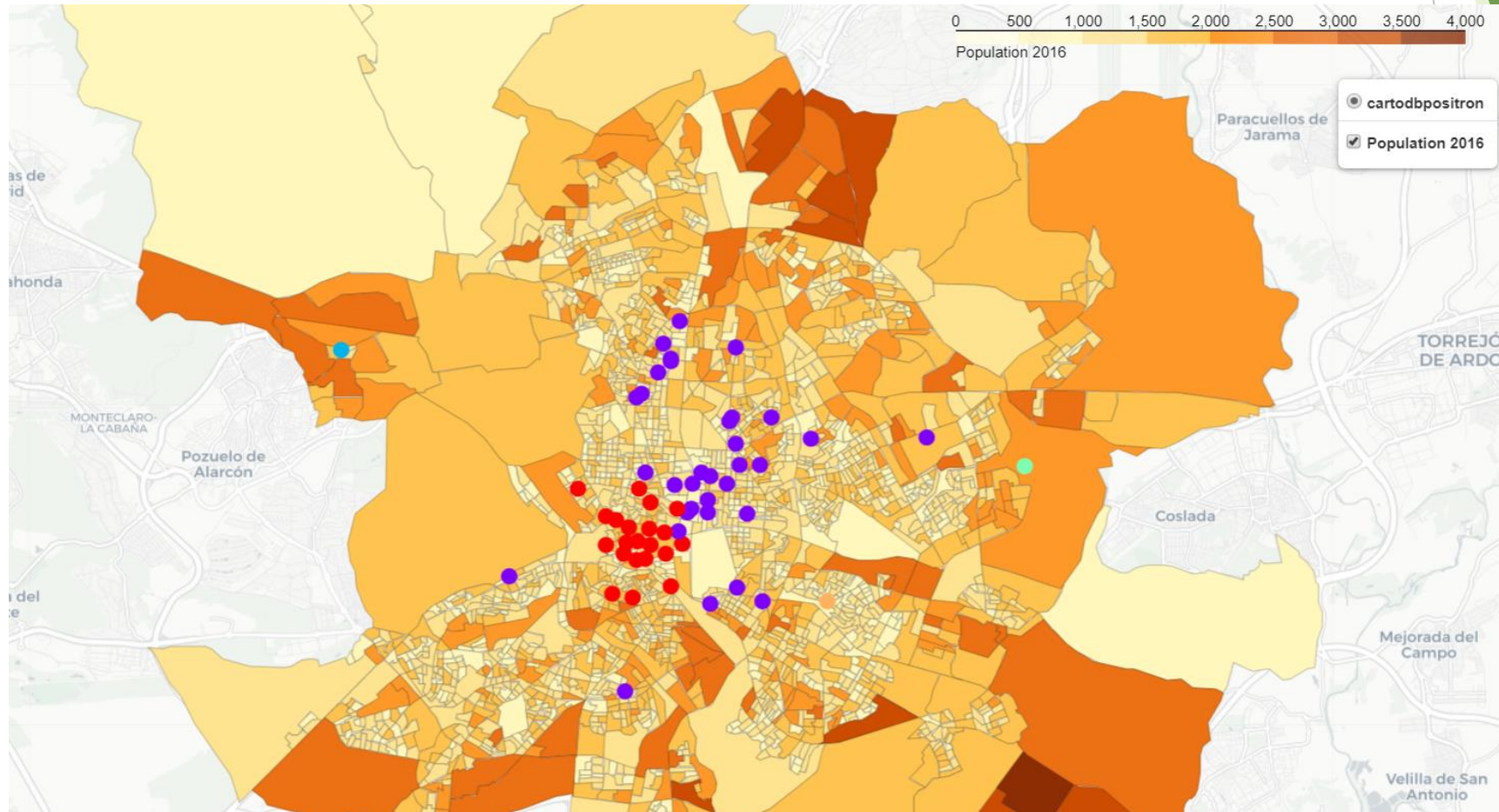
# ANALYSIS

## Production of visualization maps with additional data - Household income 2016



# ANALYSIS

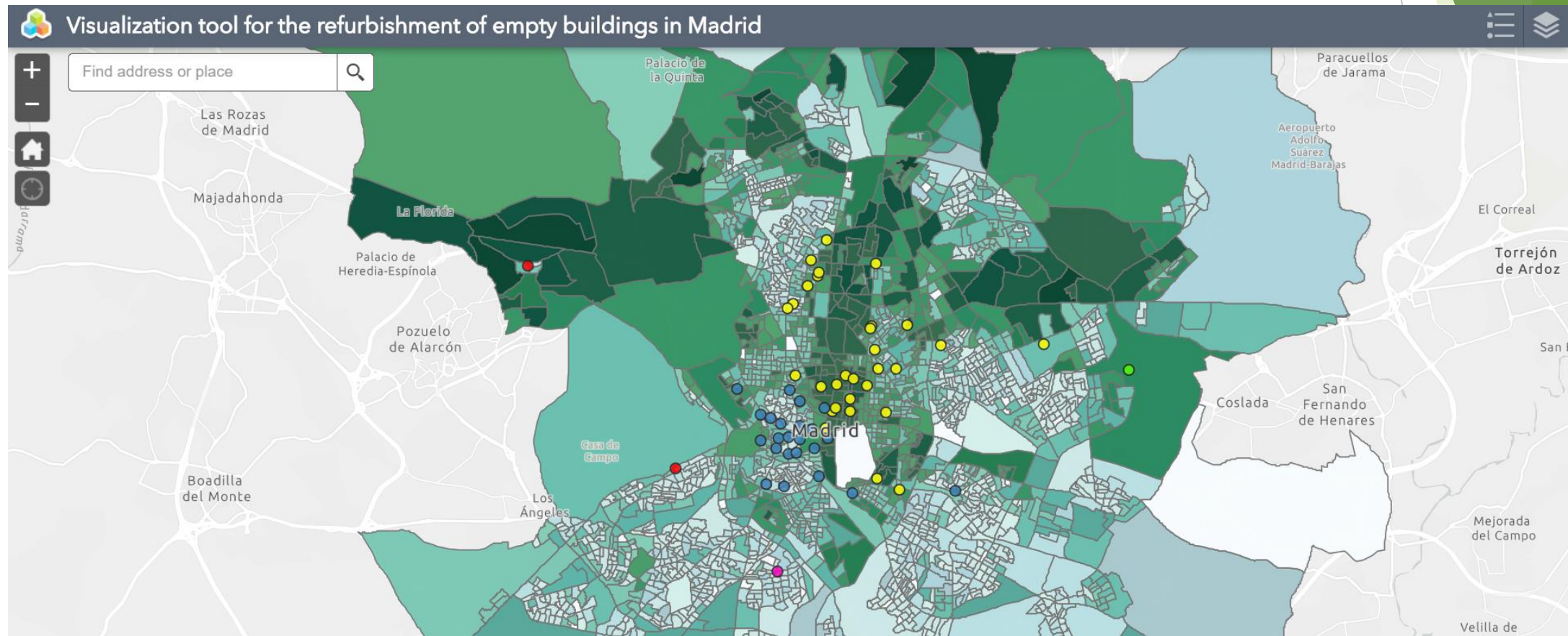
## Production of visualization maps with additional data - Population 2016





# ANALYSIS

## Elaboration of a user-friendly application with ArcGIS Online



# RESULTS AND DISCUSSION

1. Regarding the **clustering** process, the information contained in the Foursquare API is limited for Spain. Although the clustering results are not entirely wrong, a more accurate picture of each area could have been obtained with comprehensive data. This obviously affects the decision-making process, as buildings may fall into a cluster type that does not represent them.
2. On the other hand, the production of the variable regarding **household income** is based on very accurate data. The use of this type of data is extremely useful for public investment policies. In the present project, it serves a basis for the prioritization of buildings to be refurbished. A total of 23 buildings fall into lower-income census sections and will be the first ones to be remodeled.
3. Further decisions can be taken with these visualization tools. Thanks to the display of **population data**, it is possible to predict which venues will have a larger number of users. Thus, this variable could also be used for the prioritization of buildings.
4. In general, the folium library is very acceptable for the visualization of data. However, its functionalities are limited. For this reason, I decided to create a simple app with ArcGIS Online to produce a more user-friendly product.

# CONCLUSIONS

We have built different visualizations that will be used by Madrid's council to design the implementation strategy for their project "Refurbishment of empty buildings into cultural, entrepreneurial and innovation hubs".



Although the underlying data coming for the Foursquare API is not comprehensive enough for Spanish cities, the developed workflow illustrates how useful it is to produce visual data as a support tool for policymaking processes in local governments.

