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

Marta Bescansa

May 19, 2020

1. INTRODUCTION

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 decided to gather data on the areas where each building is located.

Once the characterisation of the areas was done, two variables were added to help with the prioritisation of buildings:

- Household income
- Total population

Figure 1. Buildings in the city of Madrid[1]

[1] ref

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

2. DATA

Based on the definition of our problem, three types of variables were gathered:

- Type of area according to the most common venues within 500m of the building.
- Household income in the census section where the building is located.
- Total population of the census section where the building is located.

The data used to build the project was the following:

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

Figure 2. Data gathered for the project

The census section and the population data were previously merged with ArcGIS Pro, generating a file in .json format. Given its importance for policymaking strategies, the visual information generated for this project (household income and population data) will be used for many other purposes by the city council.

All contents were stored in a GitHub account and can be retrieved from there at any time.

3. METHODOLOGY

In this project, we directed our efforts on characterising areas of Madrid that surround several buildings. Particularly, we looked at the type of area based on the most prominent venues in each one, and their economic profile based on rent per household data. Additionally, we added population data to support the decision-making process.

In the first step we downloaded the data on empty buildings from Madrid's Open Data Portal with their coordinates, and we processed it until we obtained a data frame suitable for our analysis. Secondly, each building was located in a map.

In the third step, we used Foursquare API to collect every venue around each building with a radius of 500m and a number of venues limited to 100. Name, coordinates and type were retrieved for each venue. Fourth, we analyzed each of the areas where the buildings are located and found out the ten most common venues in each of them.

In the fifth step we used unsupervised learning K-means algorithm to cluster the areas and create a new data frame that includes the cluster as well as the top 10 venues for each area. With this, we were ready to map our clusters. Sixth, we examined each cluster and determined the discriminating venue categories that distinguish each cluster. Based on the defining categories, we assigned a type to each cluster. Depending on the type of cluster they fell into, the buildings would be converted into cultural, entrepreneurial or education hubs.

The seventh and final step was devoted to the production of visualization tools that can help the city council decide which buildings should be refurbished first, adding two variables. The main variable used was the household income per census section. As this is a social development program, the buildings located in sections with the lowest income would be the first ones to be refurbished. Additionally, population data was displayed to further assist in the decision-making process.

The main outcome was a series of visualizations to support Madrid's city council with the implementation strategy for their project "Refurbishment of empty buildings into cultural, entrepreneurial and education hubs".

3.1 Data retrieval of public empty buildings in Madrid

In the first step downloaded the data on public empty buildings from Madrid's Open Data Portal with their coordinates and processed it until we obtained a data frame suitable for our analysis.

3.2 Geolocation of buildings

Secondly, each building was located in a map making use of the folium library.

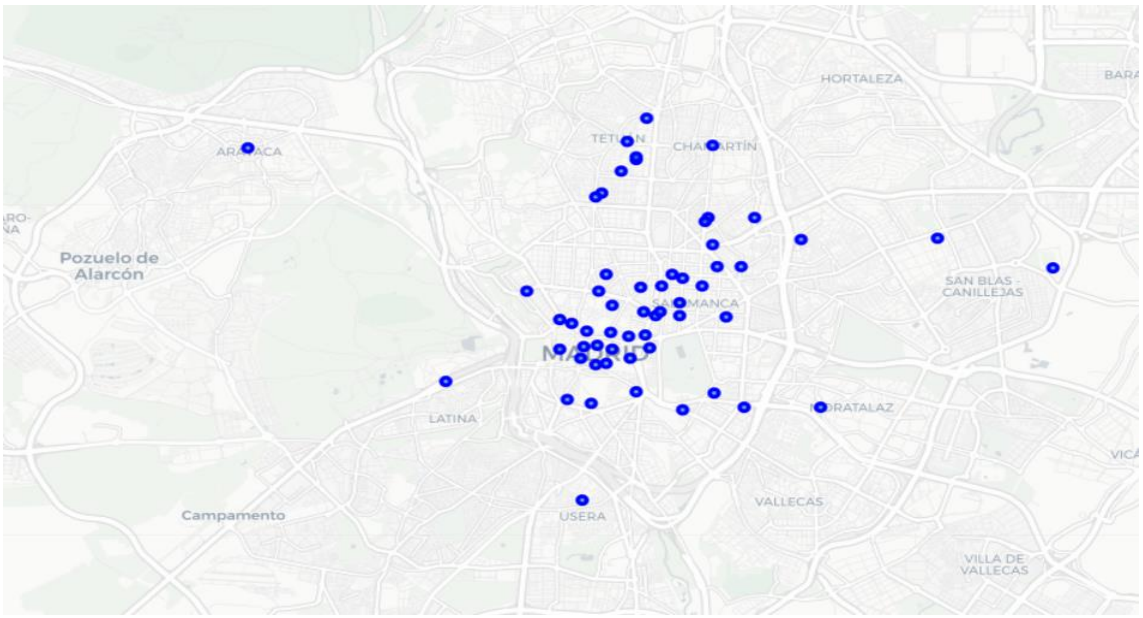

Figure 3. Location of buildings

3.3 Data retrieval from Foursquare API

In this step, we used the Foursquare API to collect every venue around each building with a radius of 500m and a number of venues limited to 100. Name, coordinates and

type were retrieved for each venue. We also built a graph to see how many venues were located in the surroundings of each building.

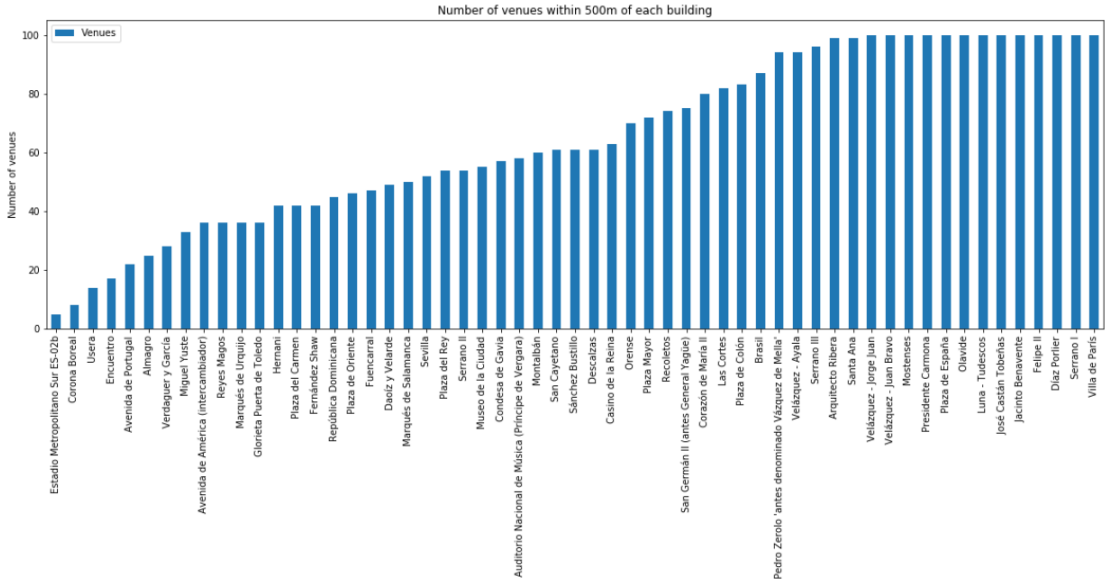

Figure 4. Venues per building

3.4 Analysis of each area

Afterwards, we analyzed each of the areas where the buildings were located and found out the ten most common venues in each of them.

	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

Figure 5. Top 10 most common venues around each building

3.5 Area clustering

We had some common patterns in the areas so we decided to use unsupervised learning K-means algorithm to cluster them. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

We ran the elbow method to ensure how many clusters were optimal and then ran k-means. Finally, we created a new data frame that included the cluster as well as the top 10 venues for each area. With this, we were ready to map our clusters.

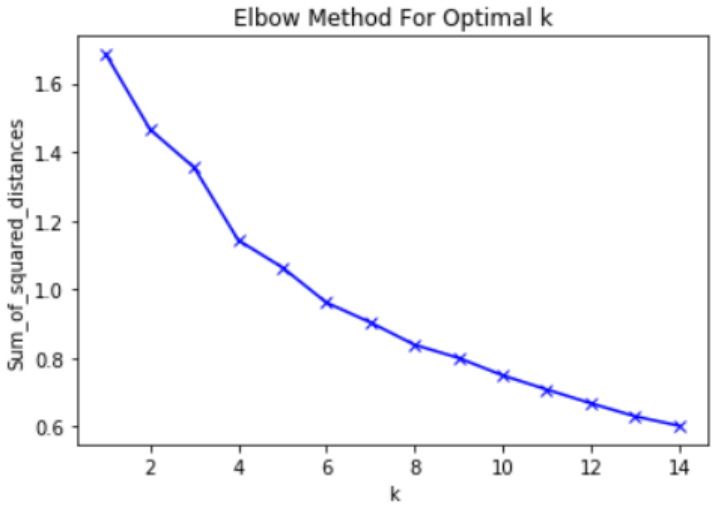

Figure 6. Elbow method

The elbow method revealed that 5 was a good number of clusters, so we ran k-means accordingly. After its completion, we were finally ready to map our clusters.

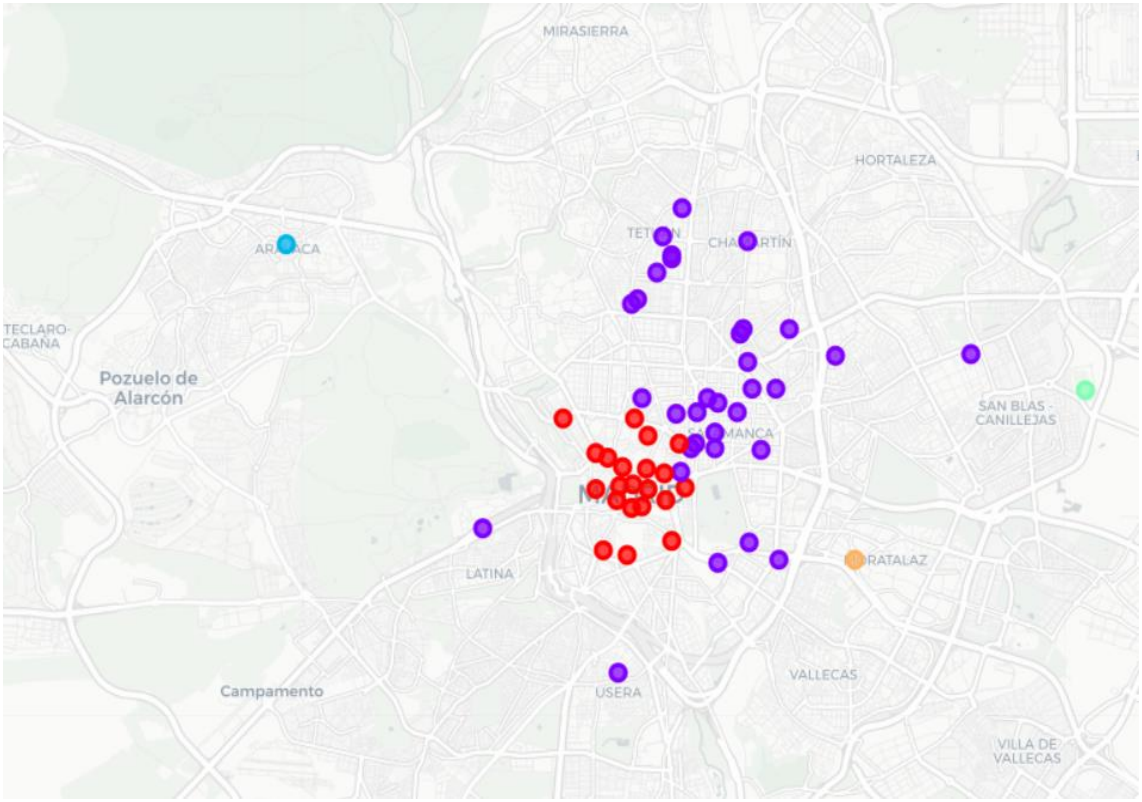

Figure 7. Result of the k-means clustering method

3.6 Examination of the clusters

In this step, we examined each cluster and determined the discriminating venue categories that distinguished each cluster. Based on the defining categories, we assigned a name to each cluster. After having a close look at the data, the following categories were assigned.

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

3.7 Production of visualization maps

Once we had characterised each cluster, we needed more variables to decide which buildings should be refurbished first. As this was a social development program, the variable used was the household income data available per census section. Thus, the buildings located in sections with the lowest income would be the first ones to be refurbished.

Then, we built a map that allowed us to identify the buildings (with its corresponding cluster tag) that were located in areas with lower household income (latest data available: year 2016)

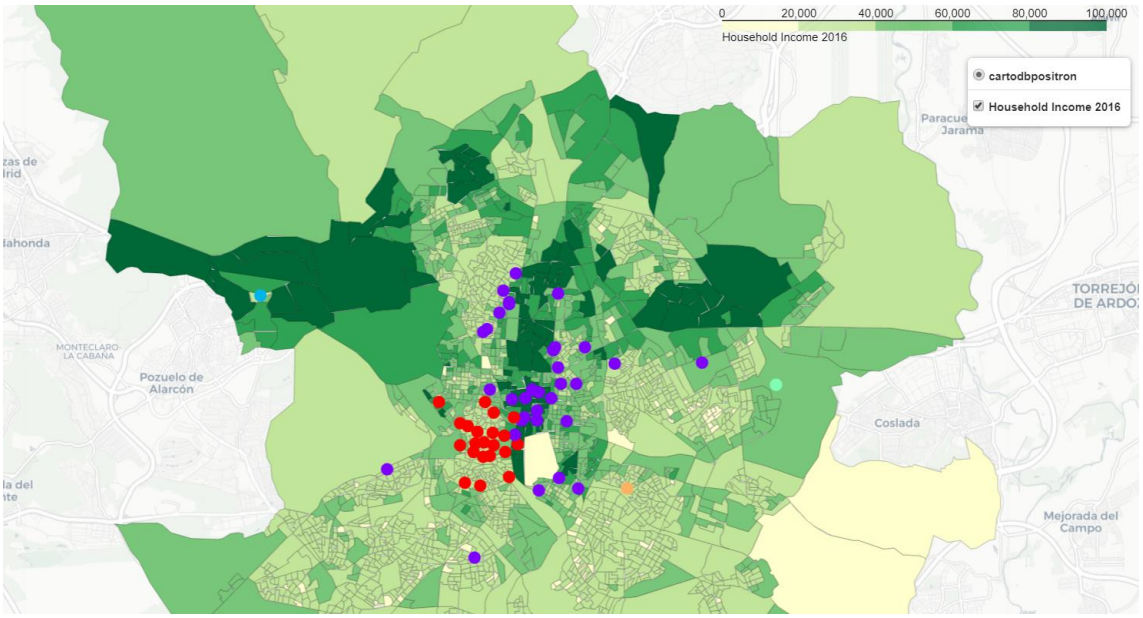

Figure 8. Household income data and clusters on a map

Additionally, we could add population data to help us evaluate where there was a higher number of potential users.

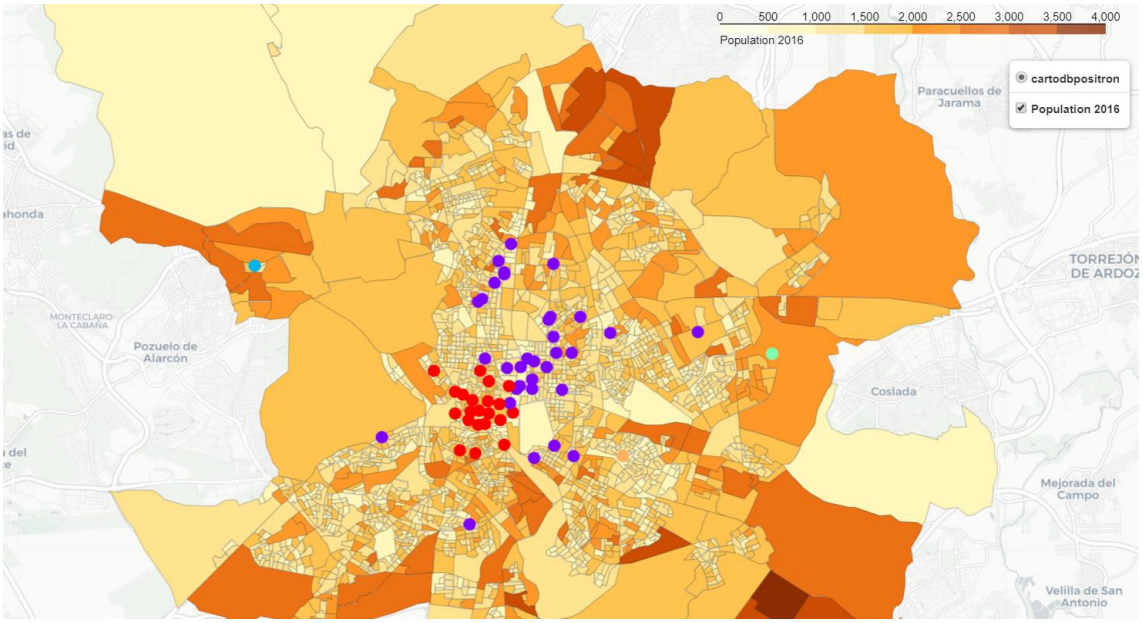

Figure 9. Population data and clusters on a map

RESULTS AND DISCUSSION

Regarding the clustering process, the information contained in the Foursquare API is limited for Spain. Although the clustering results were 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 have fallen into a cluster type that did not represent them.

On the other hand, the production of the variable regarding household income was based on very accurate data. The use of this type of data is extremely useful for public investment policies. In the present project, it served a basis for the prioritization of buildings to be refurbished. A total of 23 buildings fell into lower-income census sections and will be the first ones to be remodeled.

Further decisions could 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.

In general, the folium library is very acceptable for the visualization of data. However, its functionalities are limited. For this reason, I created a simple app with ArcGIS Online to produce a more user-friendly app. Here is the link.

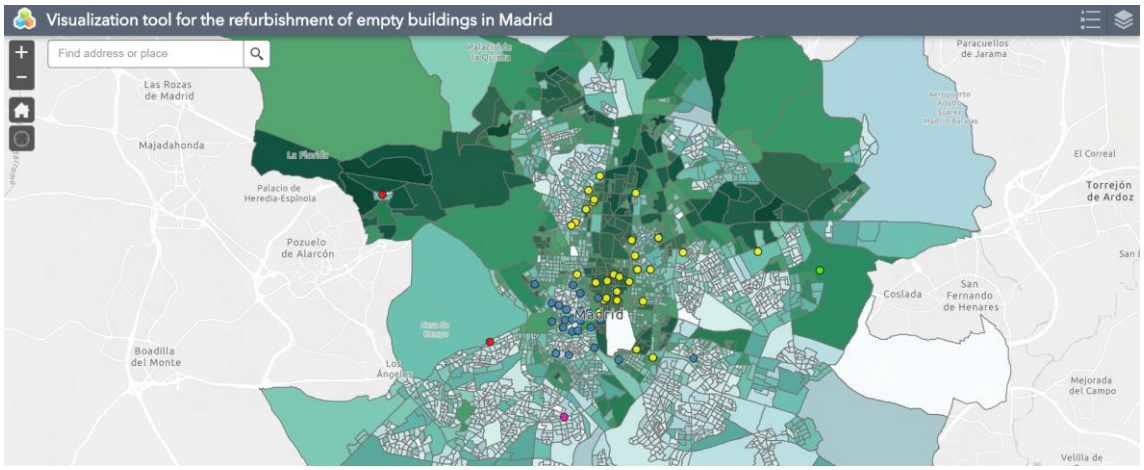

Figure 10. Screenshot of the created app with ArcGIS Online

CONCLUSION

We 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 was 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.