The Battle Of Neighborhoods

Table of Contents

1 Introduction

This project is aimed at analyzing the best areas to buy a house in Madrid. The analysis will be based on the number of restaurants, schools, shops or metro lines available in the different neighborhood as well as on the potential renting profit on the area. Where the annual renting profit is computed as the renting price per month multiplied by the months in a year and divided by the sale price.

This problem could be of interest for anyone who is considering buying a house in a big city, either for living or investing.

2 Data

In order to find the best neighborhoods in Madrid to buy a house, the following data will be used:

- List of venues provided by foursquare. The following categories will be considered:
  - Food (Restaurants)
  - Food & Drink Shop
  - Clothing Store
  - Bus Stop
  - Metro Station
  - Arts & Entertainment
  - Park
  - Athletics & Sports
  - Hospital
  - School
  - University
  - Bank
  - Pharmacy
- Sale price per square meter in the different districts and Neighborhoods of Madrid. This information will be obtained from https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Estadistica/Areas-de-informacion-estadistica/Edificacion-y-vivienda/Mercado-de-la-vivienda/Precios-de-la-vivienda/?vgnextfmt=default&vgnextoid=bf281b47a277b210VgnVCM1000000b205a0aRCRD&vgnextchannel=22613c7ea422a210VgnVCM1000000b205a0aRCRD
- Rent price per month and per square meter in the different districts of Madrid. Note that the rent price is not available per neighborhood and therefore, it will be assumed the same rent price for all the neighborhoods within each district.
- List of Neighborhoods and their coordinates obtained from the following geojson: https://github.com/codeforamerica/click_that_hood/blob/master/public/data/madrid.geojson

3 Methodology

The methodology followed in this project is detailed hereafter:

- Process madrid.geojson to obtain the list of Neighborhoods in Madrid
- Compute the reference point of each Neighborhood as the centroid of each polygon

- Using foursquare, look for the list of hospitals, schools, parks, restaurants, metro and bus stations, etc. within 5 minutes-walk (500 meters distance) from the reference position (refer to section 2 for the list of categories to be considered for the analysis).
- Store the data provided by foursquare in a pandas dataframe
- Process the resulting dataframe to obtain the number of venues of each type available in each neighborhood
- Using Kmeans and based on the number of venues available, classify the neighborhoods in 5 different groups. Display them in a folium map.
- Check the obtained groups to identify the most suitable groups (i.e. those with good communications, several schools and parks, at least one hospital, and with a relevant number of restaurants and shops in the area).
- Process the rent and sales database, merge them with the previous generated dataframe (containing the neighborhoods location, venues per category and cluster labels) and display the rent and sale prices in a folium map.
- Compute the renting profit as the ratio between the rent (multiplied by 12) and the sale price.
- Filter out the non-suitable group(s) and identify the most profitable neighborhoods
- Recommend one or two neighborhoods to buy a house based on the neighborhood features and the potential renting profit.

4 Results

To be completed

5 Discussion

To be completed

6 Conclusion

To be completed