Analysis of the influence of nearby common venues on the square meter price of properties in Medellín City.

Santiago Velez Gomez

August 16, 2020

1. Introduction

It is known that different factors affect the price of a real estate property offered on the market. In a property, both its own characteristics such as its finishes, distribution of the area or the size in square meters, as well as its external characteristics such as its geographical location in a city, have a direct influence on its commercial value.

In this study, through a two-step analysis, it will be evaluated whether the common venues near properties influence or not their value. Expressed in the price per square meter in the city of Medellín Colombia.

The findings can be used both for construction companies who want to start projects and for buyers and sellers of properties who want to know more about the common venues near the properties that interest them.

The analysis steps mentioned are:

1. Transformation and grouping of data of properties for sale located in Medellín City.
2. Clustering and segmentation by analyzing common venues near the geographic location of properties for each group mentioned in the previous step.

2. Data Section

2.1 Data Sources

Two main data sources are used for the analysis:

1. Real estate property data with its geographic location in Medellín, Colombia. (Aprox. 6100 records of properties in the dataset).

   Each property record includes:

- An unique code.
- Type of offer: (Sale or lease)
- Commercial Value
- Square meter value
- Geographical Coordinates.

The data source is taken from the Medellín Real Estate Observatory, an open data platform of the Medellín Mayor's Office. Url: [Properties-dataset]

Dataframe of real estate properties:

	Property Code	Offer Type	Comercial Value	Mt2 Value	Coordinates
2	76837	1	90000000	1034500.0	[-75.55950273469212, 6.281681757236898]
4	81000	1	100000000	1064000.0	[-75.55848773560459, 6.281669343283677]
7	91380	1	25000000	182500.0	[-75.55812114476444, 6.281473115285412]
13	81003	1	85000000	552000.0	[-75.56071377616377, 6.283044125616279]
16	79042	1	44430000	1234000.0	[-75.56041420318351, 6.283232305993462]
17	79043	1	53022000	1233000.0	[-75.56041420318351, 6.283232305993462]
18	89701	1	150000000	1128000.0	[-75.55928542553175, 6.283084405769118]
21	80991	1	110000000	873000.0	[-75.55805687085824, 6.28303369776205]
22	81626	1	95000000	1131000.0	[-75.55982318923958, 6.270291472838877]
24	91749	1	70000000	1094000.0	[-75.5608635558844, 6.271340524505576]

2. Foursquare API to query common venues near previously classified real estate properties.

   After the dataset of real estate properties is properly set, the study uses the foursquare API to get the common categories of venues surrounding the properties, in order to make a segmentation and classification analysis.

2.2 Feature Selection

For the real estate properties dataset, the main features use for the analysis are:

- Property Code: To uniquely identify a single property.
- Mt2 Value: To represent the value of the property.
- Coordinates: Store the coordinate pair (longitude and latitude) to locate the property geographically.

On the other hand, the features of the dataset of venues for the nearby venue analysis are:

- The venue category name
- Occurrence frequency as a near venue of the evaluated properties.

2.3 Data Cleaning

Looking to work with a homogeneous variable that represents the value of a real estate property, It has been decided to work with the ***price per square meter (Mt2 Value)*** of the properties because it does not involve the size of the property.

Additionally, It has been decided, only to include in the dataset the properties for sale and not for rent (lease). In order to do this, the dataset is filtered to only show the properties with a value of ***1*** in the column ***"Offer type".*** The Offer type codification is ***1*** for property for sale and ***2*** for properties for rent.

3. Methodology

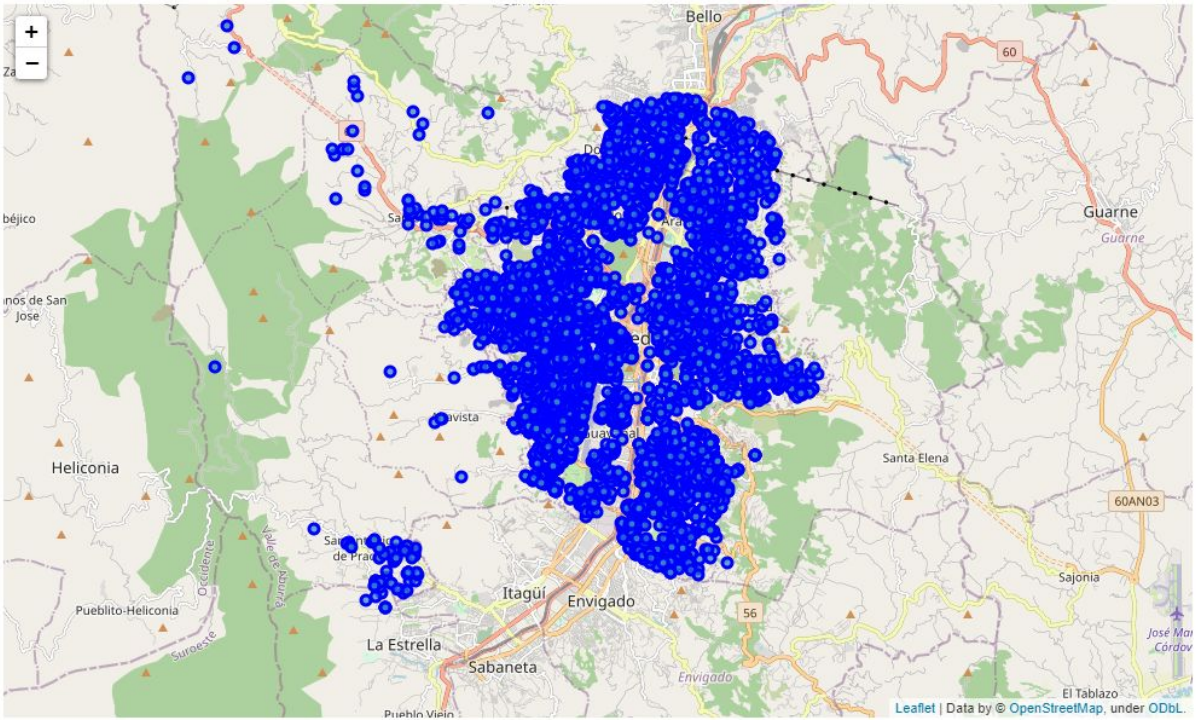
The main focus of the study is the analysis through segmentation and clustering based on categories of the most common places, of subgroups of properties made up of the properties with the highest value per square meter and by the properties with the lowest value, to finally evaluate whether the types of these places influence property value.

To run this analysis, 2 main processes were applied:

3.1 Grouping of properties according to their value per square meter.

The complete property dataset consists of 15,725 records. When the filter is applied to get the properties for sale, the data is reduced to 6149 records.

The following image shows the location of the properties on a map:



Real estate properties located geographically in a Medellín's map.

After exploring the data of a geographic nature, it was decided to divide the dataset into 4 equal parts formed by an analysis of ranges delimited by the quartiles of the values presented in the variable price per square meter (Mt2 Value).
In this way we can have the properties divided into groups, from the properties with the lowest square meter value, to the properties with the highest square meter value.

3.1.1 Interquartile Analysis

After run an interquartile function over the ***Mt2 value*** variable, the following result is obtained:

Interquartile group	Range in Mts2 Value	Description
Q1	0 - 900.000 COP	First 25% of properties
Q2	900000 - 1’412.000 COP	Second 25% of properties
Q3	1’412.000 - 2’000.000 COP	Third 25% of properties
Q4	> than 2’000.000 COP	Last 25% of properties

With this grouping criteria, a new column ***Quartile*** with the interquartile group can be added to the data frame.

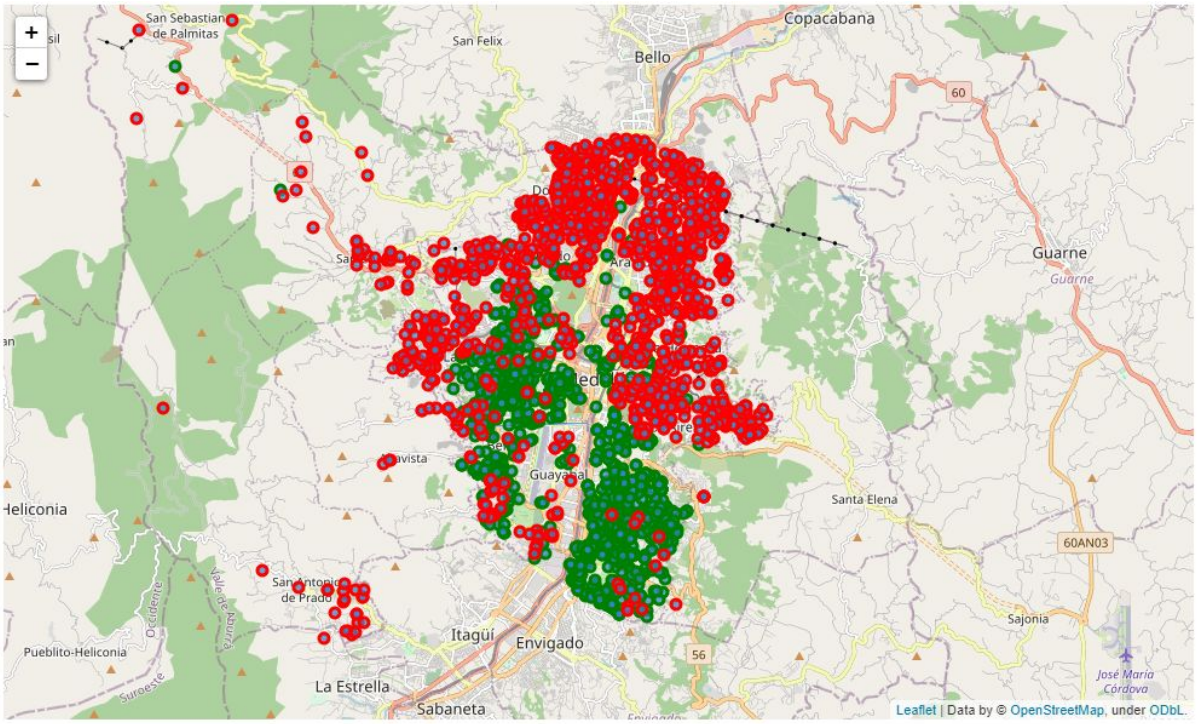
	Property Code	Offer Type	Comercial Value	Mt2 Value	Coordinates	Quartile
2	76837	1	90000000	1034500.0	[-75.55950273469212, 6.281681757236898]	Q2
4	81000	1	100000000	1064000.0	[-75.55848773560459, 6.281669343283677]	Q2
7	91380	1	25000000	182500.0	[-75.55812114476444, 6.281473115285412]	Q1
13	81003	1	85000000	552000.0	[-75.56071377616377, 6.283044125616279]	Q1
16	79042	1	44430000	1234000.0	[-75.56041420318351, 6.283232305993462]	Q2
17	79043	1	53022000	1233000.0	[-75.56041420318351, 6.283232305993462]	Q2
18	89701	1	150000000	1128000.0	[-75.55928542553175, 6.283084405769118]	Q2
21	80991	1	110000000	873000.0	[-75.55805687085824, 6.28303369776205]	Q1
22	81626	1	95000000	1131000.0	[-75.55982318923958, 6.270291472838877]	Q2
24	91749	1	70000000	1094000.0	[-75.5608635558844, 6.271340524505576]	Q2

3.1.2 Interquartile groups Selection

Due to the explorative approach of the project’s analysis and the volume of the data, it has been decided to select only the ***Q1 and Q4 groups***, in order to contrast the properties with the least value per square meter and the higher ones.

To contrast this two groups, this map is presented:

Q1: 1540 rows in RED **R**
Q4: 1488 rows in GREEN **G**



3.1.3 Random sample of 20 properties for each group.

For each representative sample, the analysis aims to observe or discover common venue categories both intragroup and extragroup for Q1 and Q4 properties.

A random sample of 20 rows is taken from Q1 and Q4 group respectively.

3.2 Analysis of segmentation and clustering of property groups

For the analysis of segmentation approach by groups of properties, segmentation based on common places presented by properties is implemented.

For each property of the samples of groups Q1 and Q4, the 10 most common venues categories consulted and calculated in the Foursquare API are obtained.

Specifically, for this calculation, 100 venues within a radius of 200 meters are consulted for each property.

The categories of venues and the frequency of their occurrence are analyzed, with the purpose of obtaining the 10 most common venues categories for each property in each Group (Q1 and Q4).

After this process the next data frames are obtained.

	Property Code	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	80374	Construction & Landscaping	Tram Station	Fish & Chips Shop	Bar	Business Service	Convenience Store	Dessert Shop	Fast Food Restaurant	Food Court	Tourist Information Center
1	80619	Fish & Chips Shop	Bar	Public Art	Dessert Shop	Tram Station	Business Service	Construction & Landscaping	Convenience Store	Fast Food Restaurant	Food Court
2	81013	Argentinian Restaurant	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Fast Food Restaurant	Tram Station	Tourist Information Center
3	82132	Hardware Store	Tram Station	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Fast Food Restaurant	Food Court
4	83024	Motorcycle Shop	Tram Station	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Fast Food Restaurant	Food Court

Q1 properties records with 10 most common venues categories. (first 5 properties)

	Property Code	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	76701	Clothing Store	Sandwich Place	BBQ Joint	Ice Cream Shop	Café	Karaoke Bar	Mexican Restaurant	Men's Store	Chinese Restaurant	Lounge
1	76784	Bakery	Yoga Studio	Eye Doctor	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega	Diner	Donut Shop
2	77550	Park	Seafood Restaurant	Yoga Studio	Clothing Store	Coffee Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega
3	77857	Park	Boutique	Seafood Restaurant	Bar	Coffee Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega
4	78943	Hotel	BBQ Joint	Eye Doctor	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega	Diner	Donut Shop

Q4 properties records with 10 most common venues categories. (first 5 properties)

3.2.1 Segmentation and clustering with k-means Machine learning algorithm

From the data of the common venues, we can submit each group of properties (Q1 and Q4) independently, to a K-means analysis where it will be possible to observe if the algorithm finds coincidences that tend to locate the 20 properties in the same large group.

For each group Q1 and Q4, a k-means analysis is performed with a value of k = 5 to obtain 5 groups (clusters) of properties.

4. Results Section

This section presents the most significant property clusters (because they group more properties) for the subgroups Q1 and Q4.

The main purpose will be to find similarities in common places for these clusters, and later to find the differences between the clusters of the other subgroup (Q1 and Q4 and vice versa).

Finally, a qualitative analysis will be carried out to determine if it is possible that these differences explain to some extent the value of the properties of group Q1 (lowest) and group Q4 (highest).

4.1 Properties group Q1

There is a significant cluster that includes 5 properties:

	Offer Type	Quartile	Cluster Labels	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
3393	1	Q1	1	Tram Station	Seafood Restaurant	Food Court	Scenic Lookout	Public Art	Park	Motorcycle Shop	Hardware Store	Tourist Information Center	Fish & Chips Shop
1587	1	Q1	1	Fish & Chips Shop	Bar	Public Art	Dessert Shop	Tram Station	Business Service	Construction & Landscaping	Convenience Store	Fast Food Restaurant	Food Court
1832	1	Q1	1	Business Service	Convenience Store	Tram Station	Fish & Chips Shop	Bar	Construction & Landscaping	Dessert Shop	Fast Food Restaurant	Food Court	Tourist Information Center
2255	1	Q1	1	Motorcycle Shop	Tram Station	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Fast Food Restaurant	Food Court
2404	1	Q1	1	Scenic Lookout	Park	Tourist Information Center	Tram Station	Fast Food Restaurant	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop

It is observed that for the Q1 group, close to the properties prevail:

- Viewpoints (Scenic Lookout)
- Subway (Train) stations
- Some businesses like bars and construction & landscaping
- tourist sites that are due to locations in the city center.

Types of venues that are associated with the attendance of people due to construction, tourist and public transport activities.

Another cluster, shows a similar behavior with common venues like train stations and construccions.

	Offer Type	Quartile	Cluster Labels	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
1455	1	Q1	2	Fast Food Restaurant	Tram Station	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Food Court	Tourist Information Center
2813	1	Q1	2	Fast Food Restaurant	Tram Station	Fish & Chips Shop	Bar	Business Service	Construction & Landscaping	Convenience Store	Dessert Shop	Food Court	Tourist Information Center

4.2 Properties group Q4

For this group there is a significant cluster of 12 properties of 20:

	Offer Type	Quartile	Cluster Labels	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
12038	1	Q4	1	Hotel	Furniture / Home Store	Italian Restaurant	Beer Garden	Yoga Studio	Donut Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio
5337	1	Q4	1	Plaza	BBQ Joint	Bakery	Salad Place	Beer Garden	Café	Nightclub	Yoga Studio	Deli / Bodega	Colombian Restaurant
13297	1	Q4	1	Café	Italian Restaurant	Salad Place	Hotel	Burger Joint	Colombian Restaurant	Bookstore	Shopping Mall	Ice Cream Shop	Supermarket
13765	1	Q4	1	Hotel	Bar	Café	Italian Restaurant	Restaurant	Burger Joint	Pizza Place	Breakfast Spot	Deli / Bodega	Japanese Restaurant
4747	1	Q4	1	Clothing Store	Sandwich Place	BBQ Joint	Ice Cream Shop	Café	Karaoke Bar	Mexican Restaurant	Men's Store	Chinese Restaurant	Lounge
5250	1	Q4	1	Eye Doctor	Business Service	Donut Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega	Diner	Yoga Studio
14684	1	Q4	1	Hotel	Restaurant	TV Station	Italian Restaurant	Sandwich Place	Yoga Studio	Deli / Bodega	Coffee Shop	Colombian Restaurant	Creperie
3863	1	Q4	1	Restaurant	Bar	Pizza Place	Yoga Studio	Diner	Coffee Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio
12789	1	Q4	1	Yoga Studio	Spa	Food Truck	Fast Food Restaurant	Coffee Shop	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega
7818	1	Q4	1	Hotel	BBQ Joint	Eye Doctor	Colombian Restaurant	Creperie	Cupcake Shop	Dance Studio	Deli / Bodega	Diner	Donut Shop
5775	1	Q4	1	Wine Shop	Ice Cream Shop	Auto Workshop	Dance Studio	Eye Doctor	Colombian Restaurant	Creperie	Cupcake Shop	Deli / Bodega	Diner
6927	1	Q4	1	Gym / Fitness Center	Intersection	Market	Fried Chicken Joint	Gym	Furniture / Home Store	French Restaurant	Food Truck	Fast Food Restaurant	Eye Doctor

It is observed that for the Q4 group, close to the properties prevail venues with a greater variety of commercial categories prevail.
We could find common venues like:

- Hotels.
- Restaurants (of all kinds).
- Healthcare venues.
- Yoga, spa and gyms.

It can be inferred that these venues are associated with activities such as eating out, personal care, and national and international business areas (hotels and restaurants).

5. Discussion

The main difference observed for the clusters of group Q1 with those of group Q4, focuses on the fact that the properties of group Q1 can refer to residential properties located in residential areas with little additional commercial activity, instead, located closer to public transport, areas with industrial activity and with a higher flow of people due to tourism.

On the other hand, the group of properties of the Q4 group that present the highest commercial values for the square meter, present a great variety of commercial venues such as restaurants of various cuisines, health and personal care. Venues valued by users who may present a greater capacity for economic income.

Specifically in a country like Colombia, it can be observed that in general, business centers are located near residential areas, essentially for people with higher incomes, due to the fact that a lower degree of urban conflict can be perceived.

5.1 Recommendations for investors and shareholders

The recommendations for investors and agents in the construction sector is that if they want to build residential properties which they want to sell at a high value according to their price per square meter (2,000,000 COP or higher), they should look for venues categories close, such as those described for the clusters in Q4.

By the contractor, if an investor seeks to create properties with lower price values per square meter (Less than 900,000 COP), in case of using the property for other types of activities, the nearby venues that influence this low price should be those described in the properties of group Q1

6. Conclusion

A qualitative relationship has been found between the categories of nearby common venues and the value of the price per square meter of a real estate property in the city of Medellín.

However, in the variability of the analyzes taking other random samples it is evidenced that other underlying variables may be correlated with the price of the properties, variables that could be found in more in-depth analysis regarding other characteristics of the properties and their surroundings.