RELATIONSHIP BETWEEN THE STRATUM AND THE LOCATION OF SITES IN BOGOTÁ

David Mora

October, 2020

1. Introduction

1.1. Context

My hometown, Bogotá, has a population of close to 8 million inhabitants and is geographically divided into 20 towns, each with different characteristics, at the same time, each town is geographically divided by neighborhoods for a total of 1922 neighborhoods. In addition to this geographical division, Bogotá has a division by strata, each area of the city is classified depending on the characteristics of the dwellings, the urban environment of the area and the urban context. This is how the city is subdivided into 6 socio-economic strata, 1 being the lowest and 6 the highest. This in order to identify areas of action and distribute the cost of public services; where the highest strata subsidize the lowest and these at the same time can access educational or health benefits given the stratification. This has allowed the city to quickly identify vulnerable sectors and this is how, among other things, it has managed to guarantee free of charge to strata 1 and 2 the minimum vital consumption of water (the first and only city in Colombia). It is necessary to add that with Based on the significant number of architectural, cultural, and historical heritage in the city of Bogotá, the heritage layer 1 was created, which provides the aforementioned benefits to the owners of those properties.

According to The Economist, on an economic level, Bogotá stands out for its economic strength associated with the size of its production, the facilities for creating companies and doing business, financial maturity, the attraction of global companies and the quality of its human capital. It is the main market in Colombia and the Andean Region, and the first destination for foreign direct investment that reaches Colombia (70%). It has the highest nominal and per capita GDP in the country, contributing most of the national total (24.5%), and is the sixth city by size of GDP in Latin America (about USD 106 678 million) with a nominal GDP per capita of USD 11,594 and a GDP PPP per capita of USD 20,120, it is also the largest business platform in Colombia where most of the high impact ventures occur. The above reasons make Bogotá a business objective for world-renowned brands.

Bogotá is considered a multicultural city because it has received inhabitants from all regions of the country, either because they seek better opportunities for their businesses, a greater educational offer or because of the armed conflict that has hit rural regions of Colombian territory since the 1950s. . The people who come to Bogotá for reasons of forced displacement are people who, for the most part, do not have enough resources to survive in the city, so it is imperative to live in the lower-income neighborhoods where the payment of public services is the highest. low possible but living conditions are more difficult.

1.2. Problem and stakeholders

1.2.1. Problem 1: International Brands in Bogotá

The decision of where to put a point of sale depends on many variables, among which we can name the target audience, rental price, price of public services, perception of security, the economic nature of the neighborhood, among other aspects. These aspects are closely related in Bogotá to the location stratum of the place, for example in a high stratum location the rent and service will be more expensive but there will be access to clients with better economic capacity. Any company that is starting its activities in Bogotá is interested in knowing where to locate its sales or work points.

1.2.2. Problem 2: Location of places in Bogotá according to the opinion of users in contrast to their stratum

For the Bogotá mayor's office and the institutions that monitor the quality of life of its inhabitants, such as the Secretary of Habitat, the Ministry of Health and the Bogotá Chamber of Commerce, among others, it is important to know the opinion of its citizens and identify if there is a trend. in the location of the places with the best quality of service and if these places have the variety of adequate services to meet the needs that increase the quality of life of citizens.

The quality of life indicators that are taken into account to give the quality of life index according to the Mercer firm (https://mobilityexchange.mercer.com/quality-of-living-reports) that placed Bogotá in position 128 Among more than 450 cities are the following: consumer goods, economic environment, accommodation, medical and health considerations, natural environment, political and social environment, public services and transportation, recreation, schools and education, sociocultural environment. Of the above indicators, I consider that they are related to the information that can be obtained from FourSquare are the following consumer goods, economic environment, socio-cultural environment and recreation.

2. Acquisition and cleaning of data

For the two problems raised, the following information is required: map of Bogotá, stratum of each block in Bogotá, places in Bogotá and for the first problem, the above would be additionally needed, a list of internationally recognized brands.

2.1. Map of Bogotá

It was obtained from the page https://gist.github.com/john-guerra/727e8992e9599b9d9f1dbfdc4c8e479e. This is described by means of a GeoJson file that contains the outline of all the departments of Colombia described by polygons, only the Bogota polygon was selected.

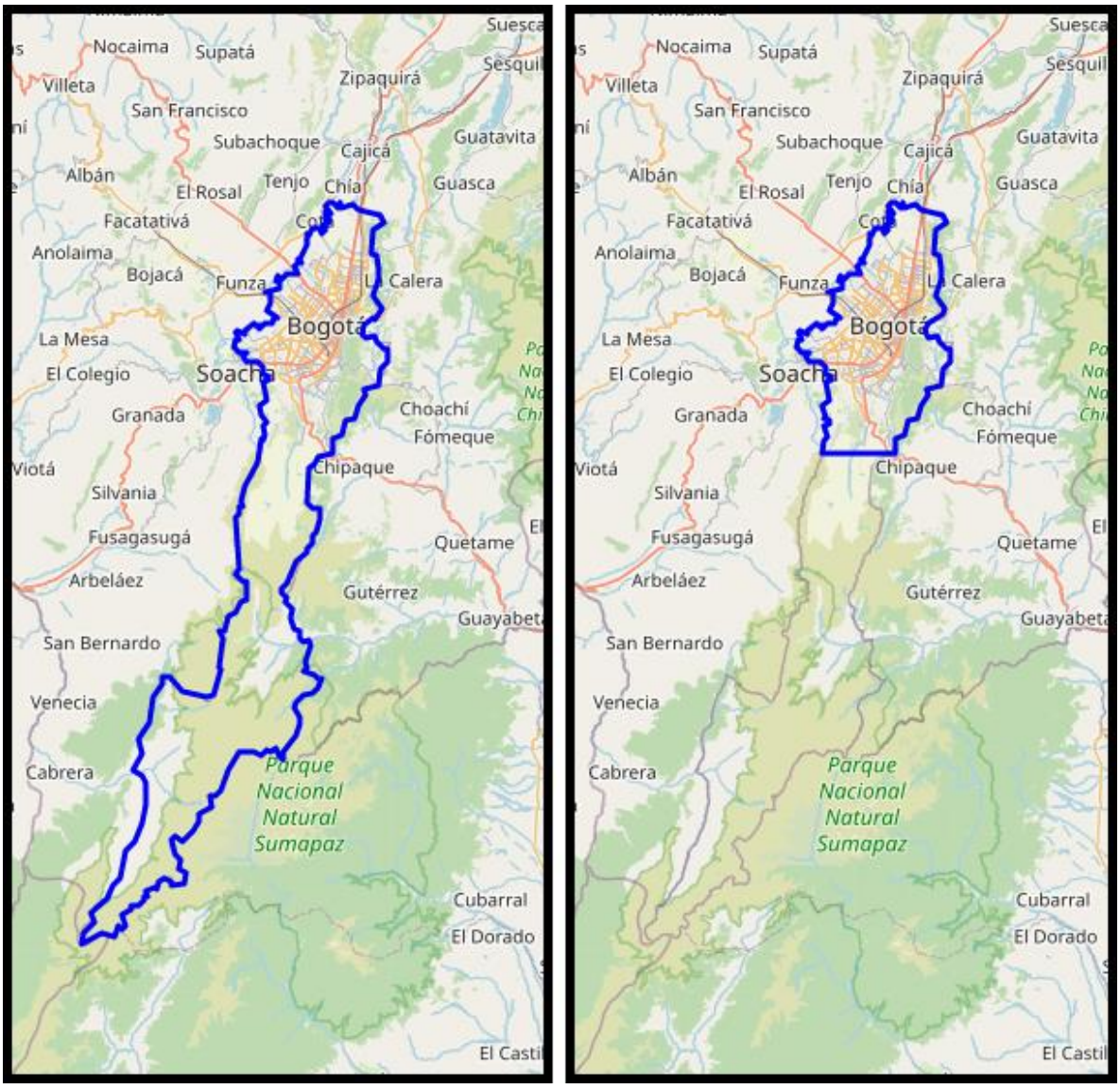

Figure 1. Map of the contour of Bogotá, on the left the complete contour of Bogotá, on the right the contour that was used, the southern zone corresponding to a rural zone was cut out.

In figure 1 on the left you can see the outline of Bogotá. Given that more than 60% of Bogotá is a rural area and our analysis focuses on the urban area, the map was cut at latitude 4.459243 in such a way that a rural area smaller than the urban area was preserved, this reduced map of Bogotá is you can see in figure 2 on the right.

2.2. Stratum of each block in Bogotá

This information can be found in a GeoJson file obtained from the Bogotá open data page, in particular in the section https://datosabiertas.bogota.gov.co/dataset/manzana-estratificacion-bogota-d-c. This file contains each block of Bogotá consigned in polygons that are associated with the stratum of each respective block, in figure 2 you can see some blocks, one of them with its stratum.

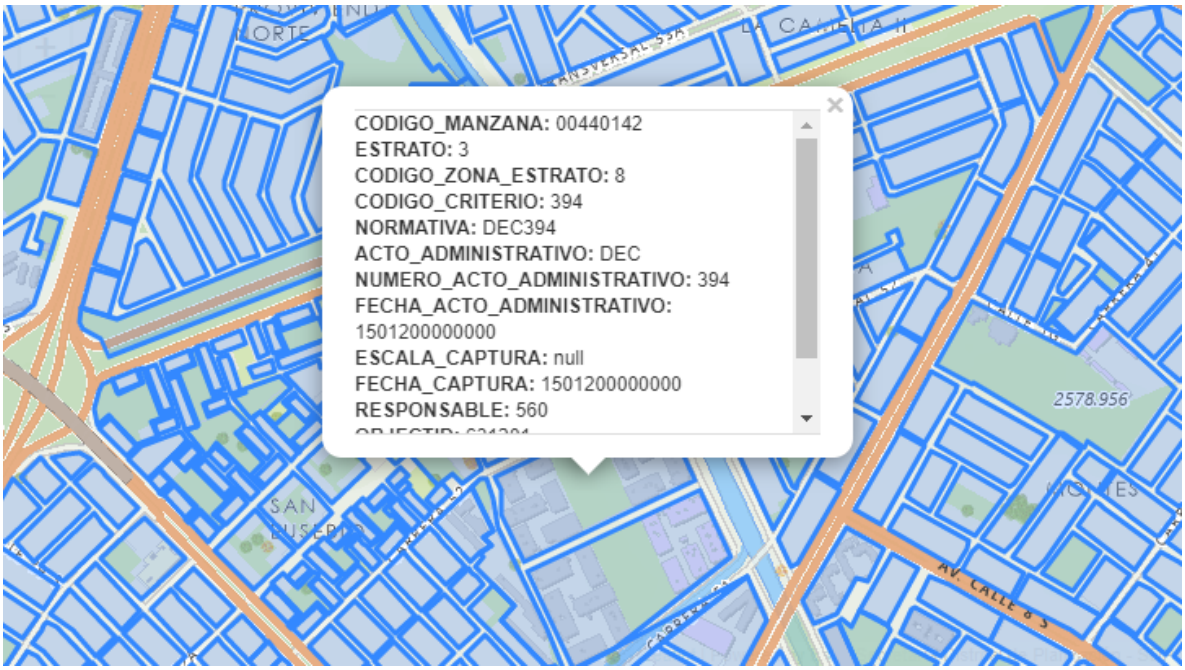

Figure 2. Map with the polygons of each block in Bogotá and their information including the stratum.

2.3. Places of Bogotá

A dataframe was formed with the information of 7488 places in Bogotá that was obtained through regular and premium requests to Four Square, the dataframe contains the following columns: name, category, latitude, longitude, id, point number, stratum of the place, rating , number of qualifiers, number of photos, number of opinions.

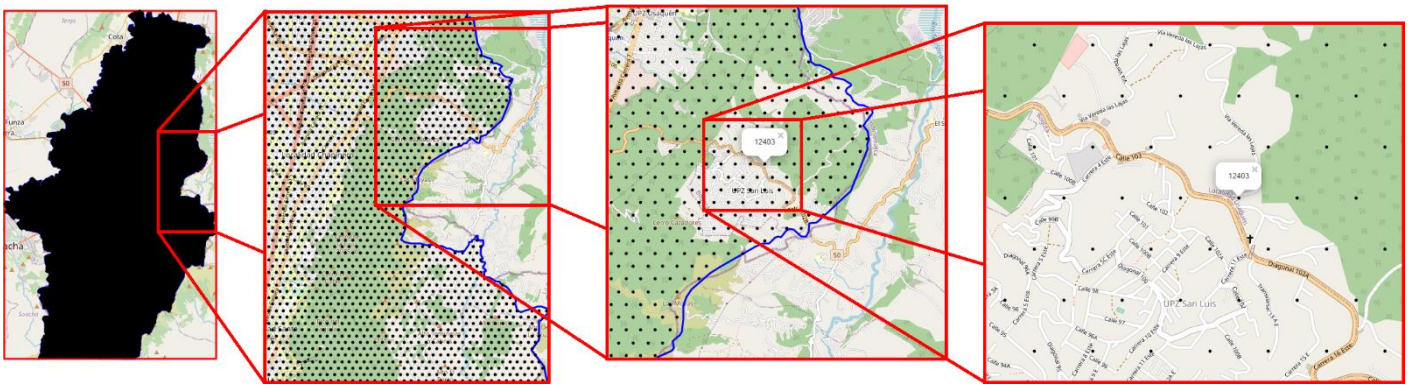

Figure 3. Grid of points on which requests for places were made to Four Square in Bogotá.

Requests were made at locations given by a triangular grid (it was chosen in this way to optimize the number of locations requested) on the Bogotá contour map with a radius of 129m. In figure 3 you can see the grid arranged over the area within the contour of Bogotá and in figure 4 you can see an area in which the places obtained in the points of the grid are shown, each point is numbered and each place found is associated with the named point in the PointNum column. The radius and the distance between points on the grid were configured in such a way that as many places as possible were obtained.

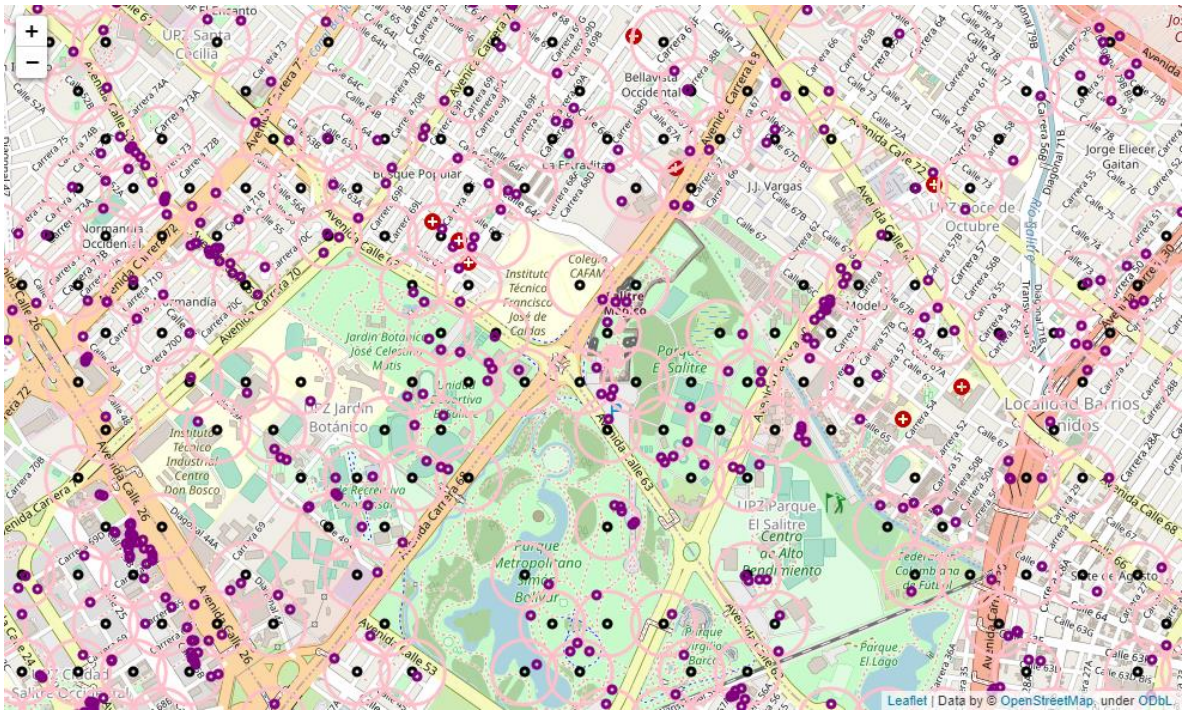

Figure 4. Map of a region of Bogotá, in purple the places found in a pink circle centered on the points of the black grid.

Using the location of the obtained places and the information of the stratum of each block (numeral 2.2), the stratum in which the place is located was established, in some cases the locations were not right, remaining on avenues or sidewalks, in that case it was assigned the stratum of the closest block. A clean-up was made of places that presented unusual locations, for example sports equipment stores or a mc'donnals in the middle of a forest on the outskirts of the city. In figure N you can see the column titles and the first and last five rows of the data frame obtained.

	name	categories	lat	lng	id	NumPunto	Estrato	CRat	Rat	CFot	COpi
0	El Fogón Del Llano	Cajun / Creole Restaurant	4.468930	-74.127996	4f4a782de4b0cccedd61818b	134	2	0.0	0.0	0.0	0.0
1	Plaza de Usme	Plaza	4.469970	-74.126363	4f4a3862e4b0eddb52dd97bb	135	2	0.0	0.0	12.0	0.0
2	Juan Valdez	Café	4.477763	-74.122131	537260f7498e7c622a9bf03a	245	1	0.0	0.0	0.0	0.0
3	agroparque los soches	Field	4.483056	-74.091950	51ebe330498eecb9a833ac50	367	1	0.0	0.0	3.0	0.0
4	Lacteos Santa Monica	Health Food Store	4.490626	-74.148808	5bd1f68e12c8f0002cc8b39c	445	2	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...
7483	TÍPICA sopas y picadas	Steakhouse	4.815070	-74.031742	53d41873498e67ffa980e88d	14903	6	12.0	6.4	13.0	5.0
7484	Club Chalaneria Guadalupe	Stables	4.817814	-74.069090	563e6a79498e41debf3b8dab	14926	6	0.0	0.0	0.0	0.0
7485	Al Limite Paintball	Paintball Field	4.818875	-74.032603	4ca77d9ba6e08cfa63508594	14944	6	37.0	8.2	51.0	13.0
7486	Refugio Hípica - Club Guaymaral	Stables	4.820947	-74.039281	503913f2e4b0cc82fd815e42	14978	6	0.0	0.0	11.0	1.0
7487	Bogotá Tennis Club	Tennis Court	4.820655	-74.036786	4bc217ed74a9a5935cb9d2f6	14979	6	55.0	8.0	36.0	8.0

7488 rows × 11 columns

Figure 5. Dataframe obtained after requests to Four Square.

Inspecting the values of the Rat column, it was observed that in most cases where CRat is zero, Rat was also zero, this is because when no rating has been given FourSquare does not give any value, in that case it was decided set the value to zero. It was observed that 14 cases despite not having

qualifications, a value was given in the Rat column, this may be due to the fact that FourSquare imports said information from other sources in which the number of qualifiers is not available or it is simply information that was lost for some unknown reason.

2.4. List of internationally recognized brands

Two sources were scraped: GLOBAL 2000 The World's Largest Public Companies (https://www.forbes.com/global2000/) and About World Top Companies Value (https://www.value.today/world/world-top -1000-companies), the 2000 and 1000 brand names with the highest profits were obtained worldwide respectively. At least 15 brands that I recognize as international did not appear in the listings obtained in the scraping, for example: KFC, Buffalo Wings, Burger King, Zara, Lego, Reebok, Subway, to name a few, so they were exchanged with brands that were previously verified they are not in Bogotá.

In both lists it was necessary to eliminate some acronyms and words such as S.A., LTD., .Com, PLC, inc, incorporated, limited, group, company, in order to avoid wrong matches in the process.

Using the dataframe conformed in section 2.3, the names of the brands that coincided in one of the two lists were selected, some coincidences with non-international brands were presented due to the name of some brands but these were cleaned, an example of this the word Santander a It is internationally recognized as a Spanish bank and in Colombia it is the surname of a hero of the wars of independence, consequently it is the name of a department in Colombia, which is why it is widely used in the names of typical food restaurants in that department. A data frame was reached with 334 places that belong to international brands in Bogotá and the information for each place.

3. Methodology and results

3.1. Problem 1: International Brands in Bogotá

From the dataframe obtained in 2.4 it was determined that the 334 places are made up of 27 brands and 29 different categories. A grouping by stratum was made to determine the number of places in each stratum, these values can be seen in figure 6. It can be seen that the highest percentage of international places are located in strata 3 and 4 (68.8%) , followed by strata 5 and 6 (27.6%), and in lower percentages in strata 1 and 2 (3.6%) where it is noteworthy that in stratum 1 no international brand was found.

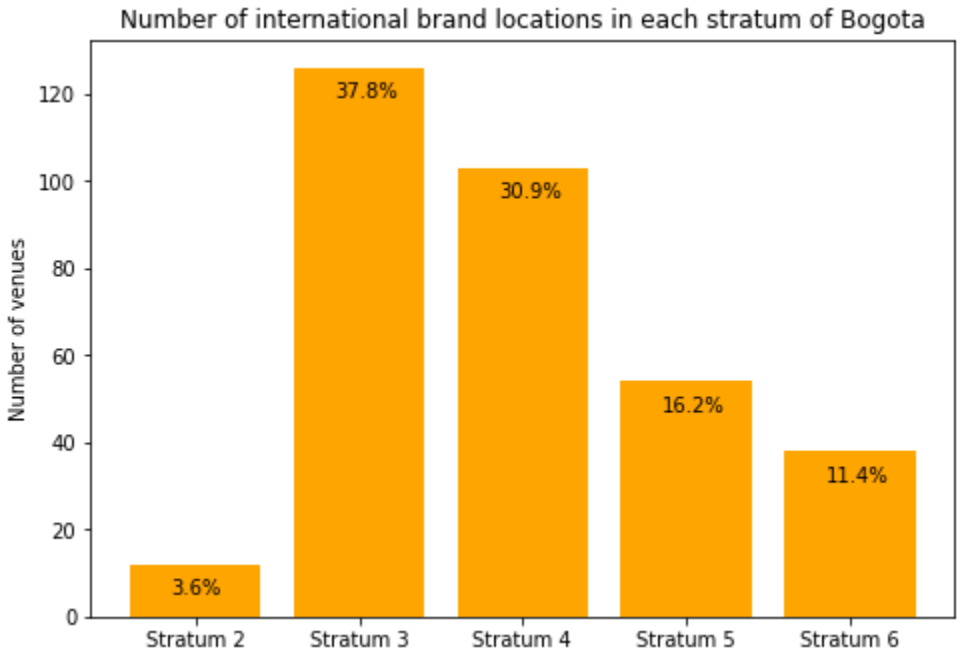

Figure 6. Bar chart by strata of international brand locations in Bogotá.

When observing the histogram of the number of places that each brand has with 20 bars in figure 7, it can be seen that about half of the brands have between 1 and 5 places, which was also seen when doing the descriptive statistics of the grouping by brands.

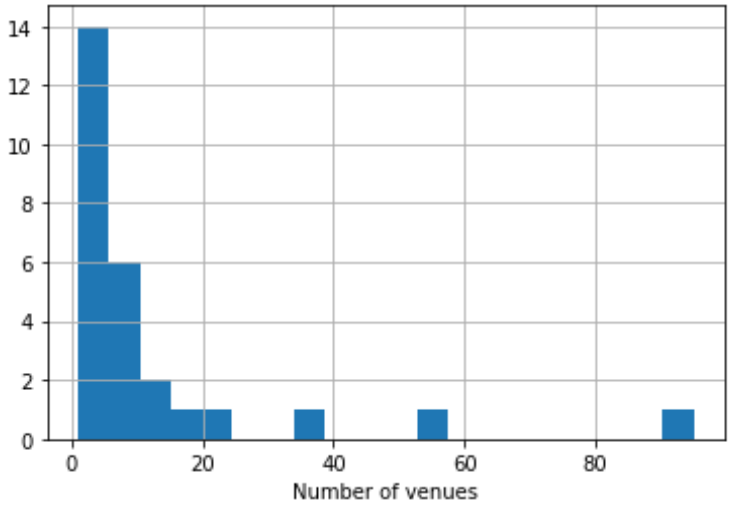

Figure 7. Histogram of the number of international brand locations in Bogotá

A grouping by brands was made in which it was found that there are brands with a single location up to brands with 95 locations, it was found that at least 25% of brands have a single location, 50% have 4 or fewer locations and the 75% have 12.5 locations or less.

Since it seeks to know in which stratum international brands prefer to be located, a grouping by brand was also made and the average of each stratum was obtained, in this way brands such as

Subway, Dunkin Donuts or McDonald's would be represented by a single place and stratum in the who decide to locate. Figure 8 shows the distribution of international brands by strata when taking the average stratum by brand and in this it can be seen that international brands do not have any point in the lowest stratum of the city (stratum 1), they prefer to be located in stratum 4 and there are more brands that are located in stratum 5 or 6 than in strata 2 or 3.

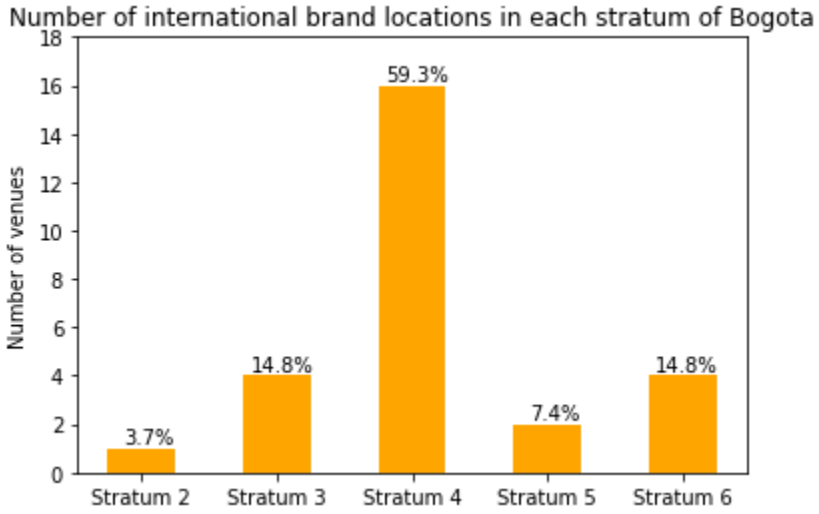

Figure 8. Number of international brands in each stratum.

Then we proceeded to make an analysis by quartiles in the number of places that each brand has.

3.1.1. First Quartile

In figure 9 you can see the distribution by strata of brands that have only one place in the city, this corresponds to 10 brands, that is, 37% of international brands.

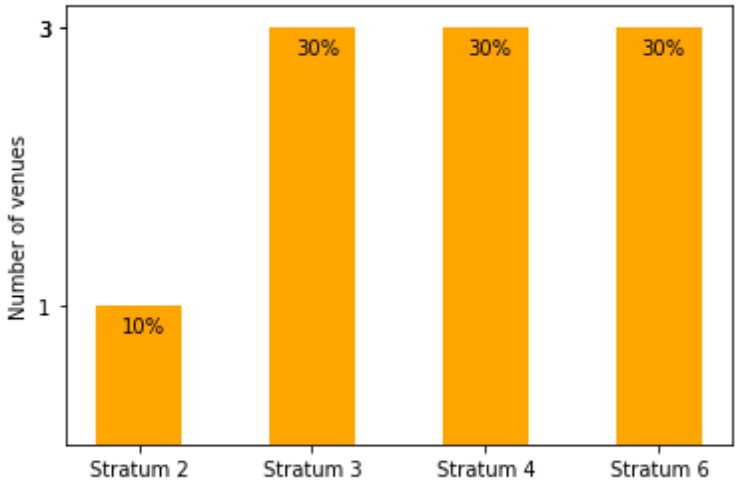

Figure 9. Number of locations per stratum when international brands have only one location in Bogotá.

The previous graph shows us that when international brands have a single point in the city, they prefer to be located in the middle and upper strata, in this way they have the possibility of attracting customers with higher income.

3.1.2. Second and Third Quartile

Figure 10 shows the distribution by brand strata that have between four and twelve locations in the city, these correspond to 8 brands of the 27 international brands found in Bogotá, these 8 brands accumulate a total of 59 locations.

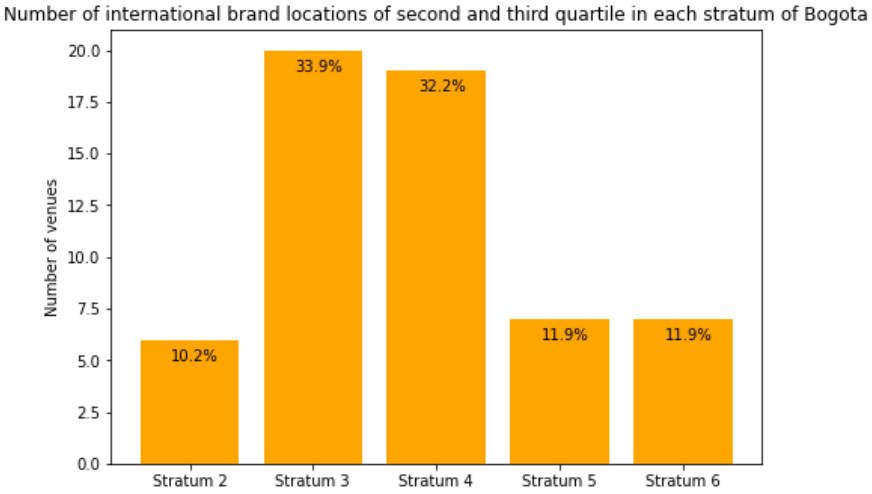

Figure 10. Number of locations per stratum when international brands have between 4 and 12 locations in Bogotá.

In figure 10 we can see that the brands that have between four and twelve places in the city tend to be located mainly in middle strata (3 and 4) and in high strata (23.8% in strata 5 and 6)

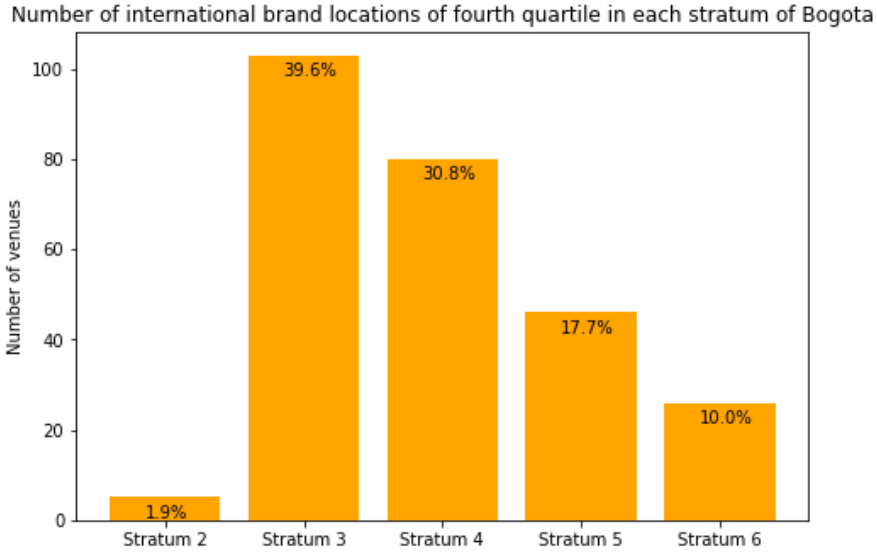

Figure 11. Number of locations per stratum when international brands have more than 12 locations in Bogotá.

3.1.3. Fourth Quartile

Here you can see the behavior of the international brands that have the most places in Bogotá. The fourth quartile is made up of brands that have more than 13 locations, these correspond to 7 brands that have a total of 260 locations, in figure 11 you can see the number of locations that brands with 13 or more locations have in each stratum In Bogotá, it can be observed that the majority of brands in this quartile have their stores located mainly in the middle strata (stratum 3) and medium-high (stratum 4), 39.6% and 30.8% respectively, and to a lesser extent in high strata, strata 5 and 6 (27% in adding both) and a minority in stratum 2 (1.9%).

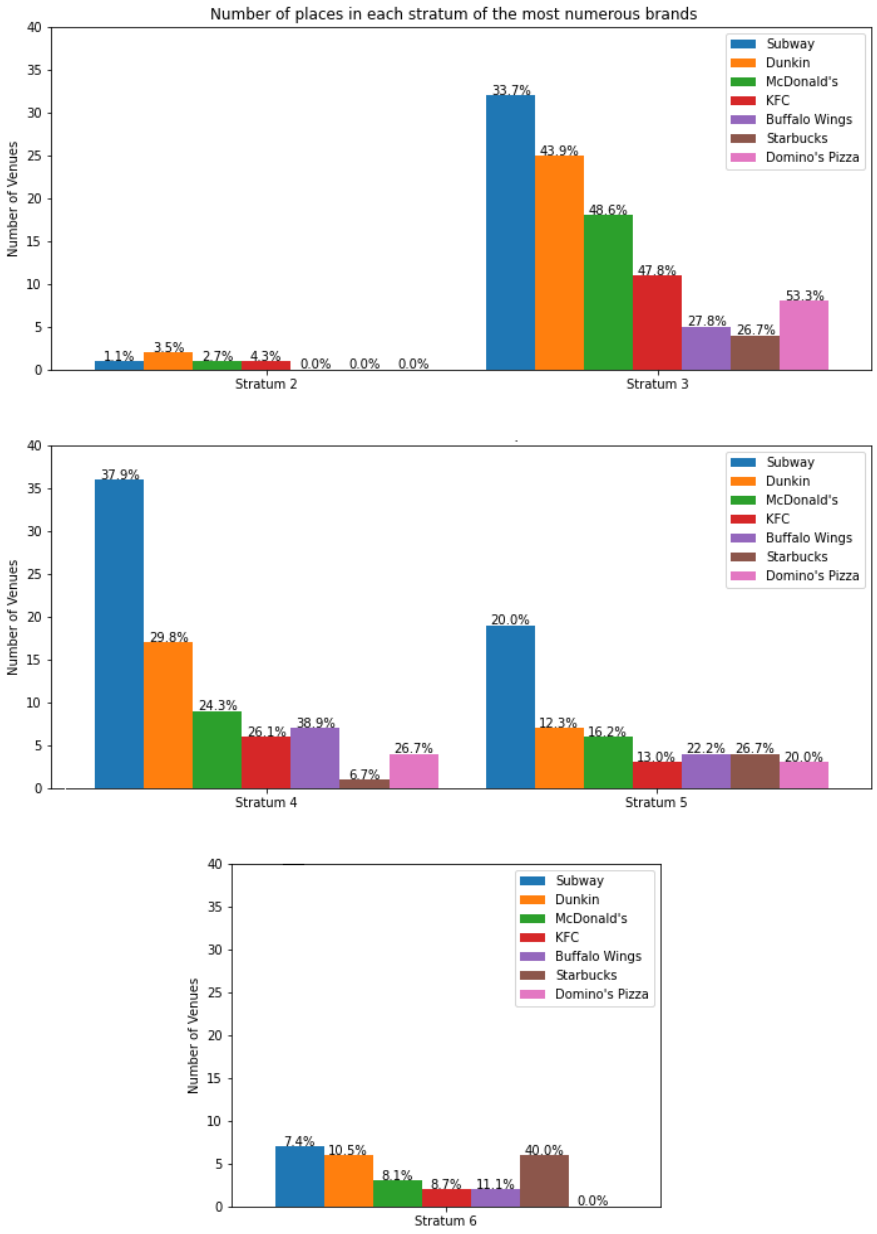

Figure 12. Number of locations in each stratum of international brands with more than 12 locations in Bogotá

Figure 12 shows the percentage of places that each of the international brands have in each stratum with more than 12.5 places in Bogotá. It can be seen that the majority of international brands with more than 13 locations in Bogotá prefer to be located in stratum 3 and gradually decrease the number of locations they have in each stratum as it is higher, only Subway and Buffalo Wings have more points in stratum 4 than stratum 3 and Starbucks is the only one that has more points of sale in stratum 6 than in the rest of the strata.

3.2. Problem 2: Location of places in Bogotá according to the opinion of users in contrast to their stratum

It began by grouping the dataframe obtained in section 2.3 by strata. The distribution of places by strata in Bogotá can be seen in figure 13. When grouping by categories, it was found that there are 357 different categories, the first 5 correspond to food sales sites (restaurant, hamburger point, bakery, fast food restaurant, pizza place).

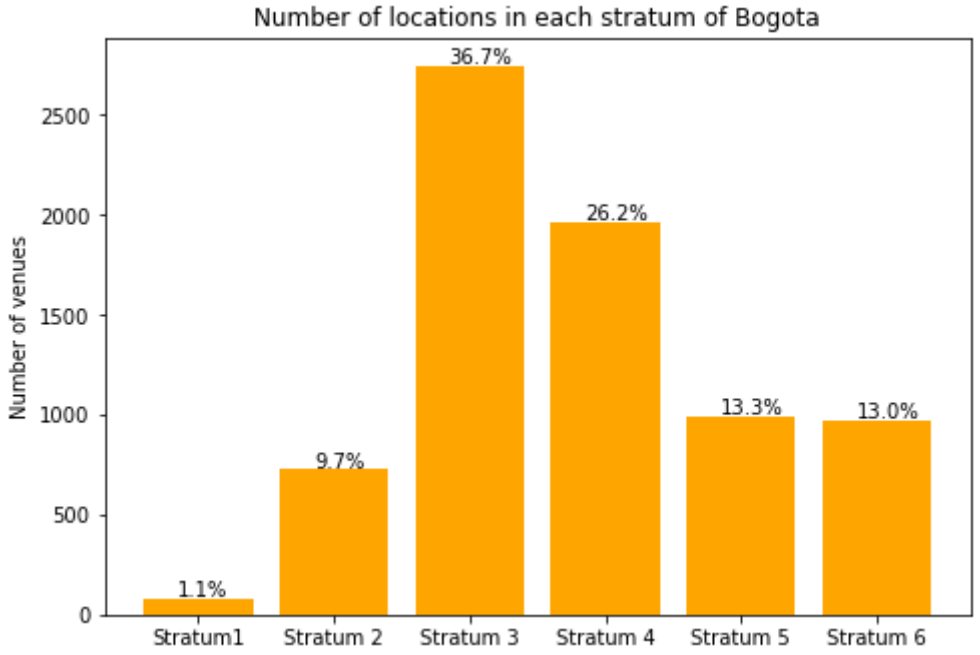

Figure 13. Bar diagram of the number of places in each stratum of all the places found in Bogotá

Recall that from the dataframe obtained in section 2.3 the last 4 columns correspond to the image that users have about the place, understand by image how much users are willing, for example, to recommend it, to comment on it, to give an independent rating whether it is high or low or to upload a photo of the place; These columns have the following abbreviations: number of ratings (CRat), average rating (Rat), number of photos (CFot) and number of opinions (COpi). For this work, for simplicity, it will be taken into account that a given opinion, an uploaded photo or a rating provided have the same weight. In that order of ideas, the CRat, CFot and COpi columns can have values from zero onwards and the Rat column can only have values between 0 and 10.

Histograms and box plot were made for each image qualification column, these can be seen in figure 14. It can be seen that in all cases the bars that cover the first interval have the most data, this implies that there are many places that have quantities at zero or close to zero, in addition to that,

it can be seen that in the CRat, CFot and COpi columns the horizontal axes appear with the order of thousands but with minimum height bars, also in the box diagram You can see that there are outliers in the thousands in those same columns, this implies that there are a few places that have a lot of ratings, photos and opinions. With this in mind, it was decided to organize our rating analysis into three categories: when the three values are zeros, when there are 1 or 2 zeros, and when there are no zeros in the CRat, CFot and COpi columns.

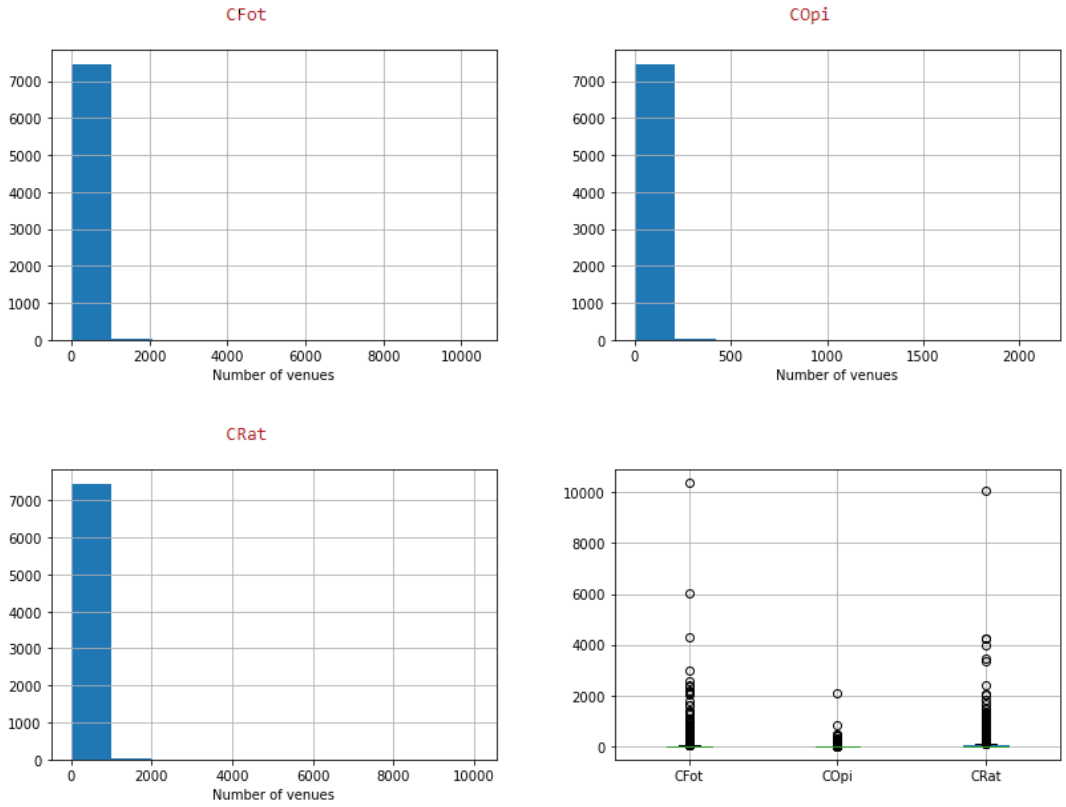

Figure 14. Histograms and box plot of the columns of appreciation of the users of the places of Bogotá

3.2.1. All values are zeros

The rows of the dataframe that had all the elements with zero value in the columns of appreciation of the places of Bogotá were selected, leaving only 750 places (10.02% of the total places), then they were grouped by strata and the distribution shown in graph 15.

Given the condition of ratings of these places, it will be considered that they are the worst rated since no user or the owners themselves have taken the task of rating the place. It can be seen that these places are mainly in stratum 3, followed by stratum 2 and stratum 4 and to a lesser extent in strata 1, 5 and 6, this indicates that the places that do not have any type of qualification are found to a greater extent in the middle and lower strata.

Clustering was not applied to this data group as there is no way to distinguish their behavior in terms of their scores.

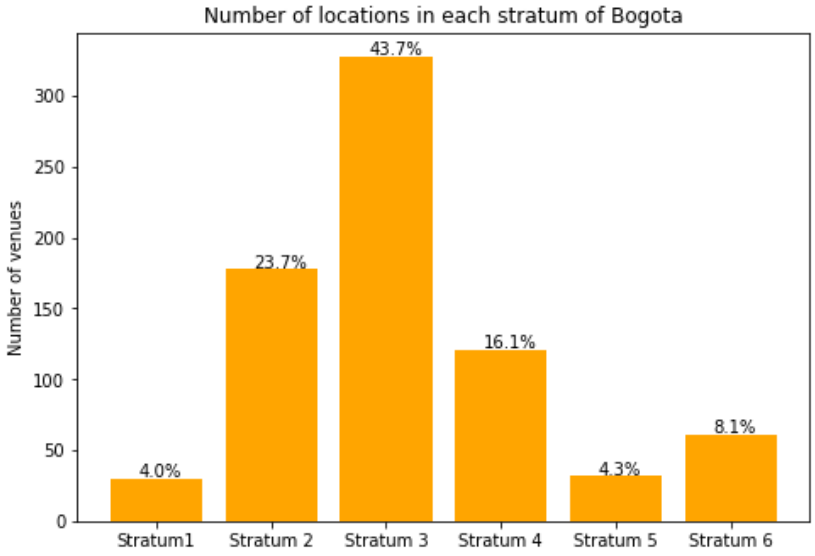

Figure 15. Bar diagram by stratum of the places that have the appreciation columns at zero.

3.2.2. There are 1, 2 or 3 values at zero

The rows of the dataframe that had 1, 2 or 3 zeros in the columns of appreciation of the places of Bogotá were selected, then they were grouped by strata and the distribution shown in graph 16 was obtained. This dataframe was left with a total of 2424 records (32.37% of all places). There were 414 places with a zero (5.53% of the total), 902 places with 2 zeros (12.05% of the total) and 1108 places with 3 zeros (14.80% of the total).

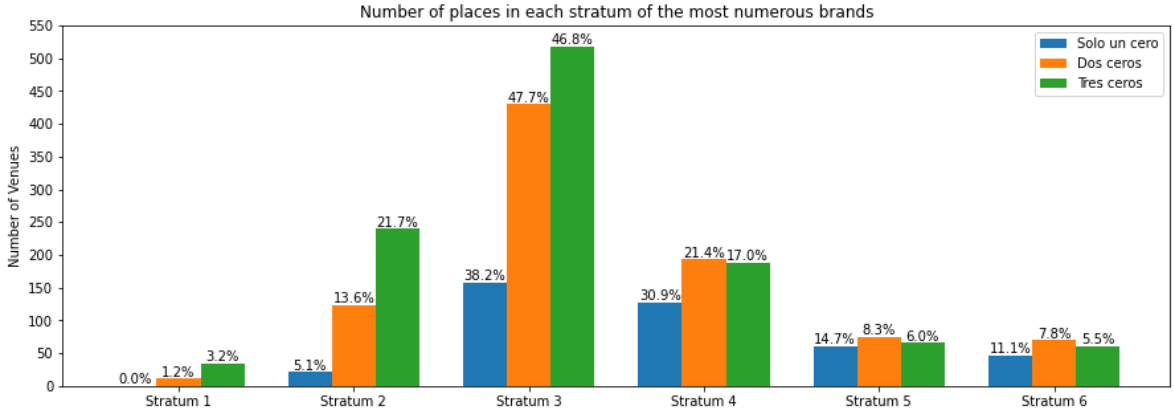

Figure 16. Bar diagram by stratum of the places that have 1, 2 or 3 zeros at the appreciation columns at zero.

Figure 17 shows the distribution of the sum of the values in each stratum, it can be seen that in total stratum 3 predominates first and stratum 4 second and stratum 2 third and to a lesser extent strata 1, 5 and 6.

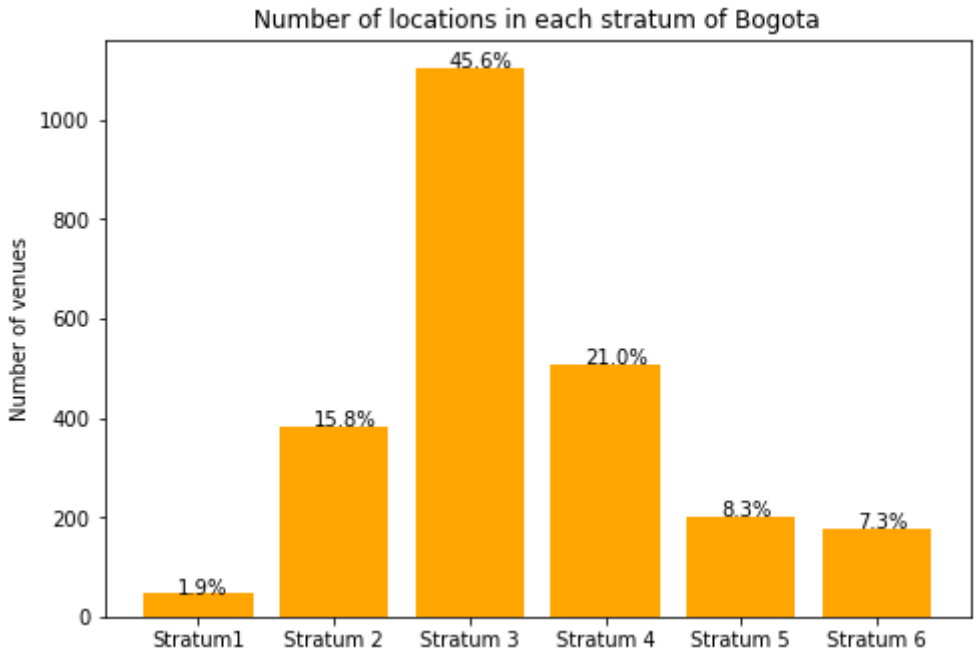

Figure 17. Bar diagram by stratum of the places that have 1, 2 or 3 zeros in the appreciation columns.

When observing the maximum values in the evaluation columns, it was found that there are places with a high number of opinions (35), photos (306) and ratings (69) compared to the respective means. It is also worth noting that the average stratum was 3.39, when observing the statistical description of the appreciation columns in figure 18, we can see from which values are considered outliers. The SumC row corresponds to the statistics of the sum of the CRat, COpi and CFot columns, the RanIQ column is the interquartile range, AtipLevelInf the lower value of the non-outlier range, AtipLevelSup the upper value of the non-outlier range.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelnf	AtipLeveSup
CRat	2424.0	2.337046	5.881679	0.0	0.0	0.0	0.0	69.0	0.0	0.0	0.0
COpi	2424.0	0.851073	1.704699	0.0	0.0	0.0	1.0	35.0	1.0	-1.5	2.5
CFot	2424.0	3.580858	9.931714	0.0	1.0	2.0	4.0	306.0	3.0	-3.5	8.5
SumC	2424.0	6.768977	12.266753	0.0	2.0	4.0	8.0	315.0	6.0	-7.0	17.0

Figure 18. Statistical description of the columns of appreciation of places with 1,2 or 3 zeros

3.2.3. There is no value at zero

The rows of the dataframe that did not have zero values in the columns of appreciation of the places of Bogotá were selected, in this way there are 4313 records (57.61% of the total of places), then they were grouped by strata and the distribution shown in Graph 19.

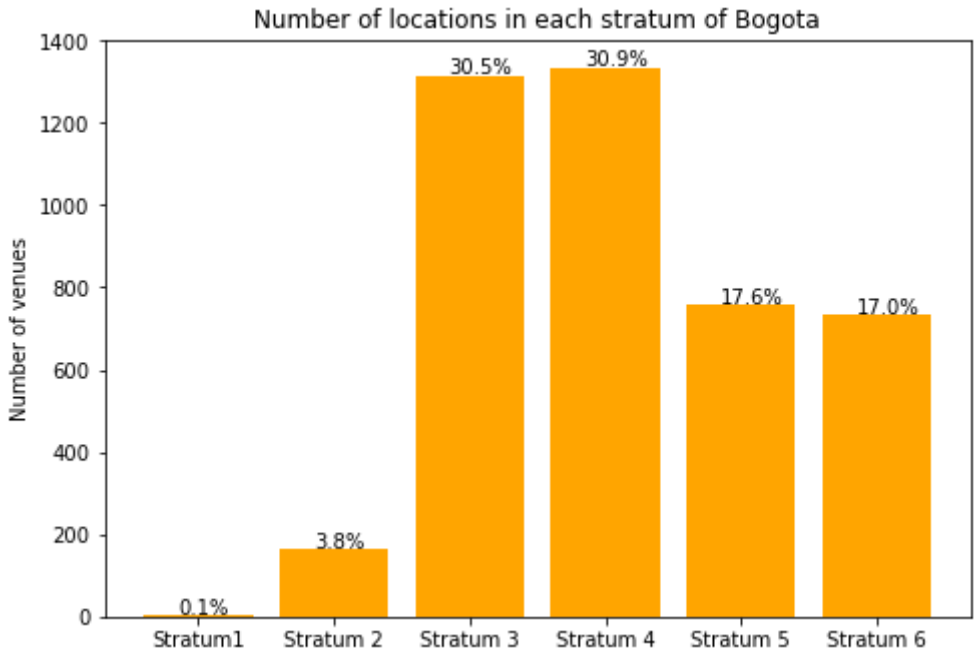

Figure 19. Bar diagram by stratum of the places that do not have zeros in the appreciation columns.

It was found that there were places whose values were very high compared to the average values, reviews (2105), photos (10368) and ratings (1050). It is also worth noting that the average stratum was 4.13, when observing the statistical description of the appreciation columns in figure 20 we can see from which values are considered atypical values from now on. The SumC row corresponds to the statistics of the sum of the CRat, COpi and CFot columns, the RanIQ column is the interquartile range, AtipLevelInf the lower value of the non-outlier range, AtipLevelSup the upper value of the non-outlier range.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelInf	AtipLeveSup
CRat	4314.0	85.280482	254.912375	1.0	13.0	25.0	68.0	10050.0	55.0	-69.5	150.5
COpi	4314.0	17.537552	46.885198	1.0	3.0	6.0	16.0	2105.0	13.0	-16.5	35.5
CFot	4314.0	54.800649	244.082387	1.0	6.0	13.0	36.0	10368.0	30.0	-39.0	81.0
SumC	4314.0	157.618683	526.173577	4.0	24.0	46.0	123.0	22523.0	99.0	-124.5	271.5

Figure 20. Statistical description of the appreciation columns of the places that do not have zeros

3.2.4. Classification of places and Clustering

The places in Bogotá were classified based on the appreciation columns given by the users, in the lowest category are the places that have zero in all the appreciation columns. It was decided to unify the dataframes of 1, 2 or 3 zeros and the one that did not have any zero in the appreciation columns because in both dataframes there were similar values in the sum column SumC, this implies that there are places that we could classify in the same category regardless of whether the place is in the 1, 2 or 3 zero dataframe or the dataframe that has no zero value. By unifying these dataframes, one

with a total of 6738 records was obtained. In figure 21 we can see the descriptive statistics of said dataframe, the SumC row corresponds to the statistics of the sum of the CRat, COpi and CFot columns, the RanIQ column is the interquartile range, AtipLevelInf the lower value of the interval of non-atypical values , AtipLevelSup the upper value of the non-outlier range.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelInf	AtipLeveSup
CRat	6738.0	55.441526	207.839831	0.0	0.0	13.0	37.0	10050.0	37.0	-55.5	92.5
COpi	6738.0	11.534580	38.372908	0.0	1.0	3.0	9.0	2105.0	8.0	-11.0	21.0
CFot	6738.0	36.374295	196.927087	0.0	2.0	6.0	20.0	10368.0	18.0	-25.0	47.0
SumC	6738.0	103.350401	427.246973	0.0	6.0	23.0	67.0	22523.0	61.0	-85.5	158.5

Figure 21. Statistical description of the columns of appreciation of the places that do not have all zeros.

Because there are very large values in each parameter of appreciation of the users, it was decided to divide the classification of the places into those that had an extremely high amount compared to the other places (places with very high amounts are isolated when using the algorithm clustering, clusters were only observed with between 1 to 10 places). The dataframe segmentation criterion was taken as the sum of the CRat, COpi and CFot columns. Places with 159 or more in sum are considered outliers and are considered in the highest ranking. The number of places in this classification was 868 and when reviewing some of these places, the Nemesio Camacho soccer stadium (main and largest stadium in Bogotá), El Dorado airport (the largest airport in Colombia), shopping centers. , restaurants of great fame and places open to the public such as the cycle path, public transport stations or public parks.

With the dataframe resulting from excluding the outliers, the unsupervised machine learning algorithm of clustering was applied from 3 to 7 clusters, having as input the columns CRat, Rat, CFot and COpi. For each number of clusters applied, the means of the sum of the appreciation columns of each cluster were obtained and its standard deviation of the means was calculated in order to find the highest value, this in order to find the cluster number in which the cluster means are further apart. It was determined that the greatest dispersion occurred with 4 clusters. Figure 22 shows the standard deviation peak of the SumC means.

Number of Clusters	Mean Estratum	Mean Count	Mean SumC
3	0.49	1127.39	50.0299
4	0.44	858.81	50.6050
5	0.39	938.66	49.8859
6	0.36	825.79	49.4438
7	0.35	783.96	48.7309

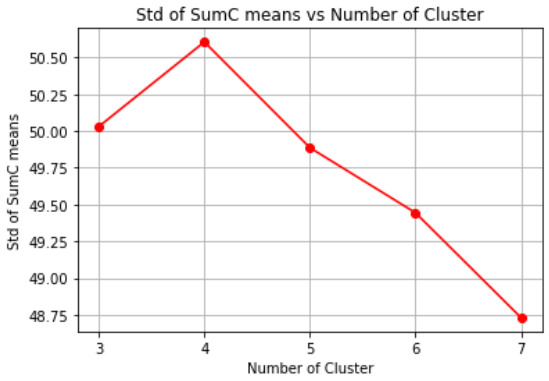

Figure 22. On the left, table of standard deviations of the means for each number of clusters. On the right, graph the standard deviations of the mean of SumC for each number of clusters.

In figure 23 you can see the table with the averages of the appreciation columns, the sum of the appreciation columns (SumC) and the stratum for each cluster. It can be seen that cluster zero corresponds to the places with the lowest values in appreciation, and the values are ascending until reaching cluster 4.

Cluster4	Rat	CFot	COpi	CRat	SumC	Estrato	Count
0	0.000000	3.264957	0.862714	0.000000	4.127671	3.262821	1872
1	6.992837	5.380679	2.699959	13.147360	21.227998	3.972984	2443
2	7.215855	18.505253	8.033429	32.208214	58.746896	4.125119	1047
3	7.351378	36.444882	16.358268	65.574803	118.377953	4.236220	508

Figure 23. Table with the averages of the opinion columns, the top column of the opinion columns, the stratum and the count of each cluster when 4 clusters were made.

With the above, we could already define a classification for the places in Bogotá based on the appreciation given by users. In the table in figure 24 we can find this classification. The categories were assigned from the lowest to the highest in alphabetical order, starting with category A, which corresponds to the lowest, up to category F, being the highest.

Clasification	Description	Count	Average Stratum	Std Straum	Kurtosis
A	All Zeros Venues	750.0	3.173333	1.197163	0.469789
B	Cluster 0 Label	1872.0	3.262821	1.102616	0.543549
C	Cluster 1 Label	2443.0	3.972984	1.101004	-0.670343
D	Cluster 2 Label	1047.0	4.125119	1.140515	-0.907376
E	Cluster 3 Label	508.0	4.236220	1.179755	-1.001810
F	Highly Appreciated Places	868.0	4.354839	1.181709	-0.962634

Figure 24.Classification of the places of Bogotá in 6 categories.

It can be seen in figure 25 that there is a direct relationship between the average stratum and the classification that the places received, as the places are classified in a higher category the stratum is higher, this relationship cannot be taken strictly since it is being taken the average stratum of the places in each category and the standard deviation compared to the number of strata tells us that the data are not distributed in a small interval, on the contrary it tells us that 68.2% of the places in each category are an interval of approximately plus and minus around the average stratum.

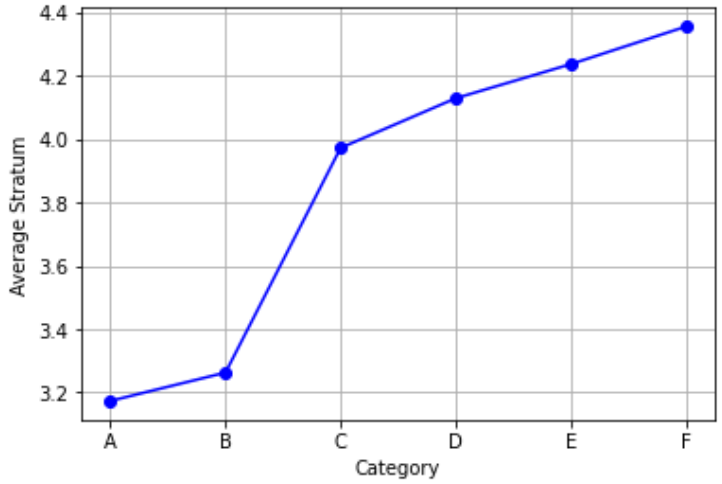

Figure 25. Average stratum vs category

4. Discussion

It should be taken into account that the database that foursquare manages is not completely adjusted to reality in terms of the number of places that they have if registered, many places of international brands and in general that are of my knowledge do not appear in the database foursquare data.

It should be mentioned that there were difficulties due to the lack of understanding under which parameters foursquare determines which places shows the requests for places near a specific point (latitude, longitude), an example of this is that by looking directly in the city-guide From foursquare the Bancolombia bank in Bogotá, there are three records, while in the scan by positions (latitude, longitude) of the triangular grid that covered the entire city, only one record appeared, a case that most likely occurs with many other places and brands.

For the two previous reasons, it must be taken into account that the results and conclusions found cannot be considered completely accurate, but we can affirm that there is a trend towards them.

4.1. Problem 1. International Brands in Bogotá

There was an imaginary that international brands would always tend to be located in the highest strata in order to find customers with the highest possible payment capacity, with the trends found it can be seen that this is not the case, if there are companies that They seek to be among the highest but it is observed that this trend occurs when they have only one point of sale and also depends on the product and business model of the brand.

An important phenomenon that was not raised within the framework of this project are shopping centers. Due to the improvements in infrastructure in the city, shopping centers have been built in middle and lower class neighborhoods, this allows many brands and users to enter contact without having to travel long distances, this is evidenced by the fact that most of the brands found in strata 2 and 3 correspond to fast foods such as Subway, McDonald's and Dunkin 'Donnuts are located in shopping centers.

A great difficulty was encountered by not finding a national open database that contains a list of international brands that operate in the country, so it was necessary to resort to lists of international brands of which a low percentage was found in Bogotá, since they contain many companies that do not work in Colombia.

A study more reduced to a brand category would allow identifying benefits in choosing the location of a business.

4.2. Problem 2. Location of places in Bogotá according to the opinion of users in contrast to their stratum

It was expected to find a more marked trend that the places with the best appreciations on the part of the users were located in the highest strata, if this trend was observed but it was not so marked, this expected trend could be diminished because it was not He thought of public access places that meet the needs of the population for free and in some way subsidized by the city, examples of this are parks, squares, theaters, stadiums, shopping centers and the airport, among many others that despite Having a very high amount of photos, opinions and ratings are located in places that do not depend on the stratum, seen in this way we can consider that the city has done an effective job in distributing places of public interest in all strata.

Based on this analysis of the places in Bogotá and the appreciation given by its users, improvement plans can be drawn up in the attention to the needs of the users and in the services provided by the places according to the location of the place. This can improve the perception that users have. A more reduced analysis of the places in Bogotá to a particular category, together with the appreciation that users have about it, would allow us to identify in which locations it is more convenient to start a business or a public work.

5. Conclusions

This report presented two problems related to the location of places in Bogotá and the stratum in which those places are located.

The first questioning led us to find a trend in the decisions that international brands make when locating their points of sale in the city, when the brand proposes to have only one point, they prefer in 60% of cases to locate it in a middle stratum high (4.1 on average), brands with between 4 and twelve points were mainly located in the middle stratum (3.2 on average), brands with more than 12 points were located in the medium-high stratum (3.9 on average), it is also worth noting that no international brand has points in the lower stratum of the city. The

The second questioning sought to see if there is a relationship between the stratum in which the places are located and the appreciation that users have. The places were classified into six categories, the lowest being A, the highest category F and the intermediate categories B, C, D and E that were determined by the clustering algorithm according to their number of photos, opinions, number of ratings and average rating of the place, it was found that the number of clusters that best separated these quantities was four clusters. From this classification, the average stratum of the places in each category was obtained, resulting in the lowest category (A) having the lowest stratum

on average (3.17 middle stratum), and as the category increases the stratum increases, the Highest category (F) has the highest average stratum (4.35 upper middle stratum). It was expected to see a more marked trend such that in the lowest category the stratum was approximately 2 and in the highest category the stratum was close to 6.

The trends were not found as marked as expected because there are public places such as theaters, parks, squares, shopping centers and the airport, among others, that provide a pleasant and highly remembered experience to the public regardless of the stratum in which it is located the place.

It can be said that if there is a tendency for the places with the best appreciation by users to be found in the highest strata, but this is not the rule.