

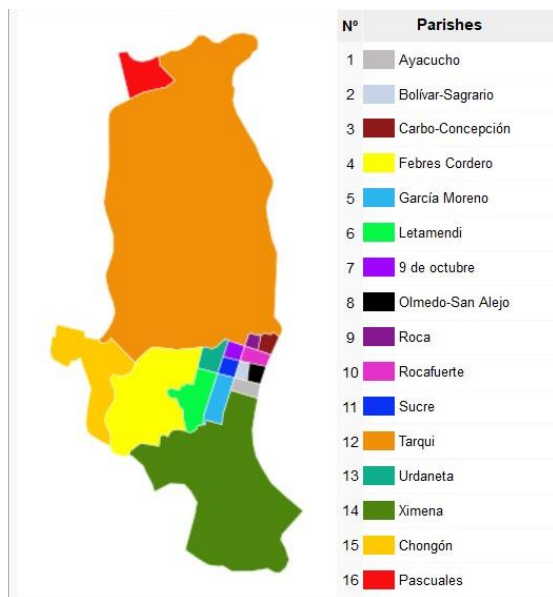
Clustering urban parishes in Guayaquil, depending on consumption trends in hotel establishments.

Applied Data Science Capstone in Python

Gabriel Mantilla Saltos.

Introduction

Important hotel companies from abroad, wish to perform a geo-statistical analysis of urban parishes in the city of Guayaquil. Guayaquil is the largest city in Ecuador and occupies a primary place in the national economy. For this, it will use the Foursquare platform, to measure consumption trends in each area, such as: what types of places exist (parks, shopping centers, restaurants, etc.), which of these places are well attended, provided by hotels in the area, etc. For this, the analyst who was hired by these hotel companies, obtained relevant information from each parish, and it turns out that the parishes of the city center, are very close. While the parishes of the north and east, are well separated. The information obtained refers to the names, geographical coordinates, neighborhoods and commercial places of each parish. With this information, we will study if the parishes are grouped, by commercial trends, without taking into account the distance between them.

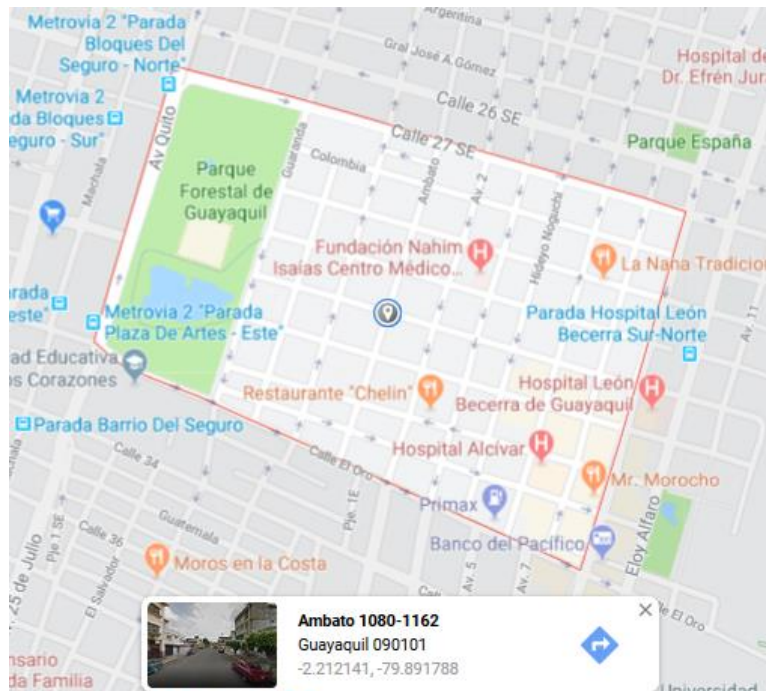


Data Source

A csv database was created, with the Postcode, Parish Borough, Neighborhood, Latitude, Longitude columns. The postal codes were obtained from the following address, with the respective names of the parishes:

<http://www.forosecuador.ec/forum/ecuador/educaci%C3%B3n-y-ciencia/149665-c%C3%B3digo-postal-de-guayaquil-ecuador-todos-los-c%C3%B3digos-by-parishes>

Then, the geographical coordinates were obtained, looking for each name in the google maps, clicking on the position found.



And the information of each Neighborhood, representing the neighborhoods or main sectors (Borough) of each parish, were obtained from the following place:

https://es.wikipedia.org/wiki/Sectores_de_Guayaquil.

I share the database (Guayaquil.csv), and the repository of how to open it, from the outside.

<https://github.com/gabriemanclub/Segmentation-and-Clustering-of-Guayaquil-s-Parishes>

gabriemancub / Segmentation-and-Clustering-of-Guayaquil-s-Parishes
Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

Important hotel companies from abroad, wish to perform a geo-statistical analysis of urban parishes in the city of Guayaquil. Guayaquil is the largest city in Ecuador and occupies a primary place in the national economy. For this, it will use the Foursquare platform, to measure consumption trends in each area, such as: what types of places exist...

[Manage topics](#)

13 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find File Clone or download

gabriemancub Add files via upload Latest commit 48fc85a 7 hours ago

Data section.ipynb	Add files via upload	7 hours ago
Guayaquil.csv	Add files via upload	yesterday
README.md	Initial commit	yesterday

README.md

Segmentation-and-Clustering-of-Guayaquil-s-Parishes

Important hotel companies from abroad, wish to perform a geo-statistical analysis of urban parishes in the city of Guayaquil. Guayaquil is the largest city in Ecuador and occupies a primary place in the national economy. For this, it will use the Foursquare platform, to measure consumption trends in each area, such as: what types of places exist (parks, shopping centers, restaurants, etc.), which of these places are well attended, provided by hotels in the area, etc. For this,

Below is the complete table that will be used for the parish grouping analysis.

	Postcode	Parish	Borough	Neighborhood	Longitude	Latitude
0	090101	Ayacucho	Downtown Guayaquil	Ayacucho	-79.890106	-2.210537
1	090102	Bolívar	Southeast Guayaquil	Bolívar (Sagrario)	-79.890021	-2.201835
2	090103	Carbo	Southeast Guayaquil	Carbo (Concepción)	-79.880388	-2.187644
3	090104	Febres Cordero	Southeast Guayaquil	Abel Gilbert, Batallón del Suburbio, Estero Sa...	-79.923126	-2.195226
4	090105	García Moreno	Southeast Guayaquil	García Moreno	-79.900845	-2.21422
5	090106	Letamendi	Southeast Guayaquil	Letamendi, Puerto Lisa.	-79.911136	-2.212051
6	090107	Nueve De Octubre	Central Guayaquil	Nueve de Octubre Oeste.	-79.893237	-2.189603
7	090108	Olmedo	Southeast Guayaquil	Olmedo (San Alejo).	-79.885288	-2.203112
8	090109	Roca	Southeast Guayaquil	Roca	-79.885523	-2.186729
9	090110	Rocafuerte	Southeast Guayaquil	Rocafuerte	-79.886083	-2.194642
10	0090111	Sucre	Southeast Guayaquil	Sucre	-79.897066	-2.201498
11	090112	Tarqui	North Guayaquil	Acuarela, Comegua, Brisas del Rio, Limonal, Ju...	-79.905822	-2.125677
12	090113	Urdaneta	Southeast Guayaquil	Garay	-79.906624	-2.19664
13	090114	Ximena	South Guayaquil	Centenario, Cuba, Del Astillero, Floresta, Gua...	-79.894642	-2.212882
14	090115	Chongon	West Guayaquil	Puerto Azul Norte, Puerto Azul Sur	-80.080779	-2.233158
15	090116	Pascuales	North West Guayaquil	Bastión Popular, Las Orquideas Este, Las Orquí...	-79.93242	-2.071827

Methodology

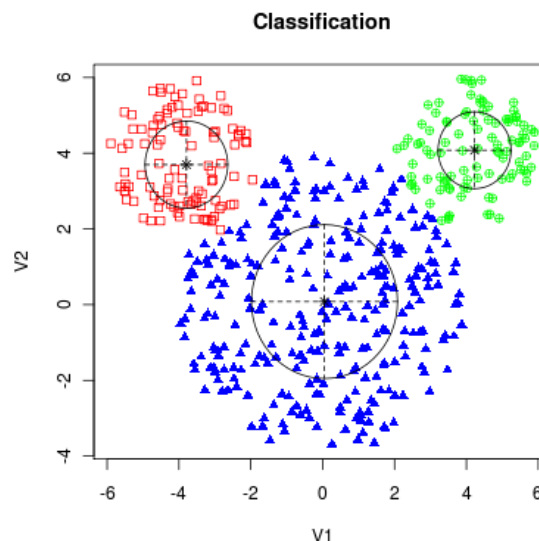
K-means Clustering

As a technique of grouping variables, cluster analysis is similar to factor analysis; but, while the factorization is rather not very flexible in some of its assumptions (linearity, normality, quantitative variables, etc.) and always estimates the distance matrix in the same way, the agglomeration is less restrictive in its assumptions (not demanding linearity, nor symmetry, allows categorical variables, etc.) and supports various methods of estimating the distance matrix.

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.

The analysis of K mean conglomerates is especially useful when a large number of cases are available. There is the possibility of using the technique in an exploratory way, classifying the cases and iterating to find the location of the centroids, or only as a classification technique, classifying the cases from known centroids supplied by the user. When it is used as an exploratory technique, it is common for the user to ignore the ideal number of clusters, so it is convenient to repeat the analysis with a different number of clusters and compare the solutions obtained; In these cases, the hierarchical cluster analysis method with a subsample of cases can also be used.

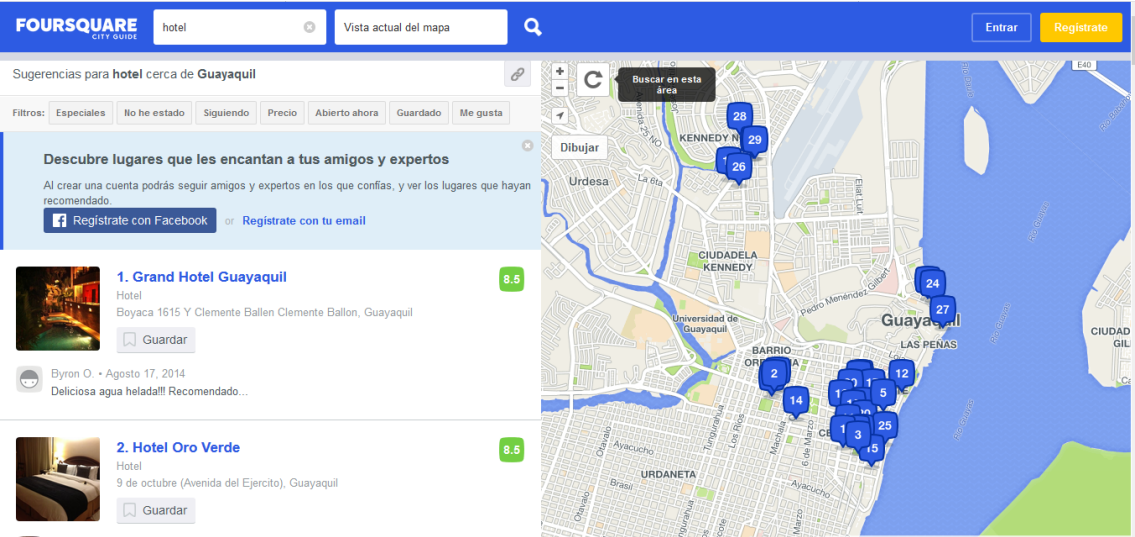
To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids



Foursquare platform.

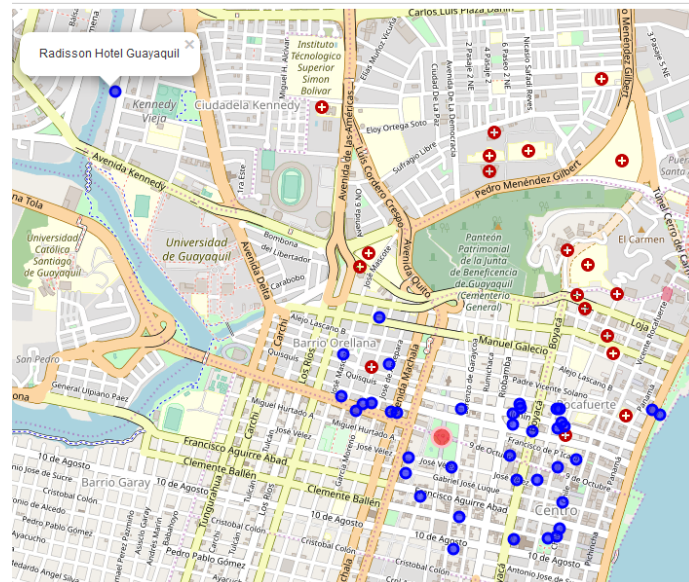
Its mission of the company in create the most reliable and independent data platform for understanding how people move through the real world. The application allows you to find the best restaurants, bars, hotels and more. The application is based on collecting commercial information, food services, entertainment, and the most visited places in an area that has all these geo-referenced places. Which are then rated by users, who give tips

on these places, and so other users can find places, products that not everyone knows, tickets to hidden places, wi-fi passwords, and so on. Then the users can know where other users have been and where they want to go. Through a username and password, created in the platforms like developers, it is possible to use this information obtained on the platform, which will allow analysts to find patterns of tastes and positive ratings in the places they have visited.

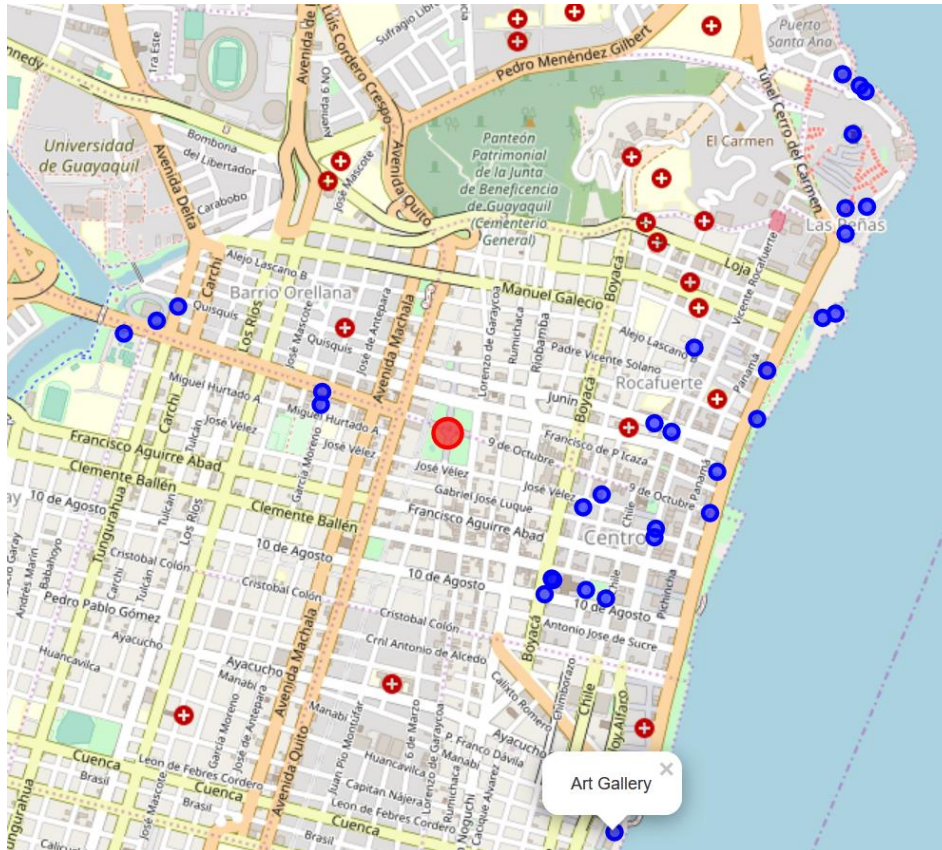


Results

In the following map, we can visualize about 40 hotels that are within a radius of 2.5 km away, with respect to the Centenario Park, which is the Red dot that we visualize on the plot.



Below are the busiest places in the area, as we can see an Art Gallery, which even now that place is called, The University of the Arts in Guayaquil. There are parks, shopping centers, coffee shops, bars and everything a bit.

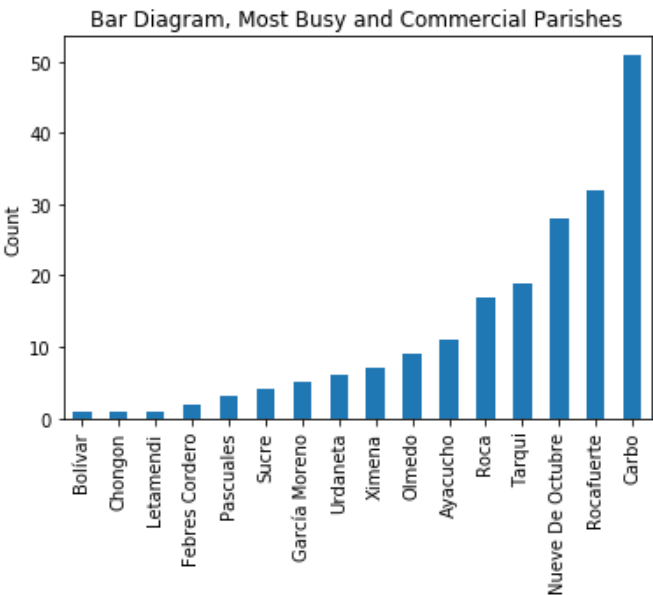


If we explore the same analysis, but taking the coordinates found in each Parish, we could explore the 5 busiest places or trends, ordered from highest to lowest.

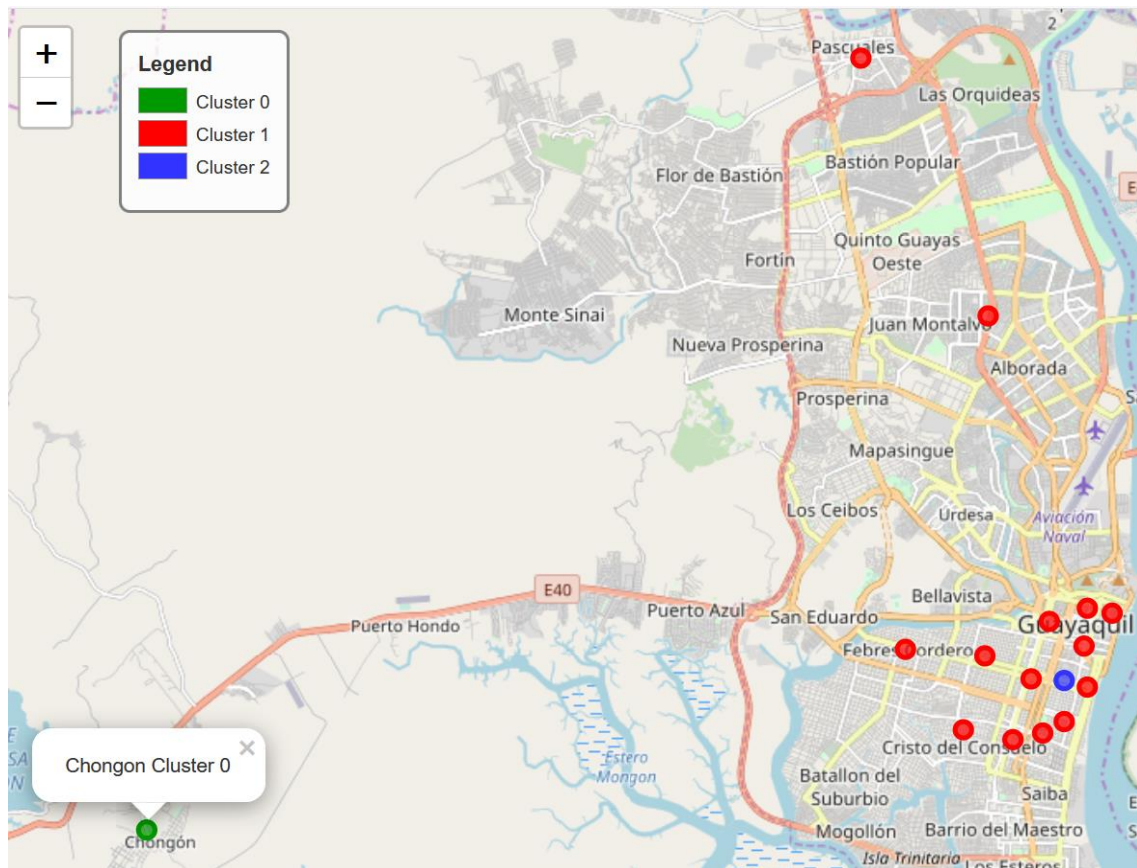
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
García Moreno	Breakfast Spot	Southern / Soul Food Restaurant	Sandwich Place	Food	Department Store
Letamendi, Puerto Lisa.	Chinese Restaurant	Vegetarian / Vegan Restaurant	Department Store	Diner	Falafel Restaurant
Acuarela, Comegua, Brisas del Río, ...	Coffee Shop	Shopping Mall	American Restaurant	Burger Joint	Food Court
Roca	Hotel	Seafood Restaurant	Pharmacy	General Entertainment	Hotel Bar
Rocafuerte	Hotel	Steakhouse	Restaurant	Bakery	Café
Bolívar (Sagrario)	Market	Vegetarian / Vegan Restaurant	Food Truck	Diner	Falafel Restaurant
Centenario, Cuba, Del Astillero, FL...	Park	Theater	Seafood Restaurant	Music Venue	Restaurant
Abel Gilbert, Batallón del Suburbio...	Pharmacy	Gift Shop	Food Court	Dessert Shop	Diner
Puerto Azul Norte, Puerto Azul Sur	Playground	Deli / Bodega	Dessert Shop	Diner	Falafel Restaurant
Carbo (Concepción)	Seafood Restaurant	Hotel	Bar	BBQ Joint	Bakery
Sucre	Seafood Restaurant	Diner	Candy Store	Vegetarian / Vegan Restaurant	Food Court
Nueva de Octubre Oeste.	Seafood Restaurant	Hotel	Park	Restaurant	BBQ Joint
Olmedo (San Alejo).	Seafood Restaurant	Art Gallery	Pet Store	Coffee Shop	Juice Bar
Ayacucho	Seafood Restaurant	Breakfast Spot	Playground	Restaurant	Fast Food Restaurant
Garay	Seafood Restaurant	Dessert Shop	Rest Area	BBQ Joint	Fast Food Restaurant
Bastión Popular, Las Orquídeas Este...	South American Restaurant	Diner	Chinese Restaurant	Food Truck	Falafel Restaurant

The following bar plot, we show that parishes are the most visited, if we talk in tourist, commercial terms. Pedro Carbo Parish, this city in downtown Guayaquil, is where the nightclubs are located, the Noria of Guayaquil, which is the largest roller coaster in South America. There is even Cerro Santana, and the Malecon 2000, being one of the most tourist places in Guayaquil. The 2000 boardwalk receives approximately 1.6 million tourists per month. This graphic is interesting, because the Bolivar parish is almost south west of Guayaquil, very close to the Center. But there are small businesses in that area, it is a commercial area, such as selling clothes, shoes and things for the home, but there are

no representative restaurants or bars in that area. Chongón Parish instead is located west of the city, about 20 km away, this is a more industrialized area, there are parks and hotels in this area, and it is known by tourists, for the attraction of knowing the monkey of Chongón.



Below is a map, with the coordinates of the parishes in Guayaquil, the objective of the colors, represents how the parishes are grouped commercially, that being the characteristic of this study. We can see that the Red dots, like the Pascual parish, are grouped in the cluster 1, it is interesting to observe that, because, although they are very separated from each other of the same color, because they have similar characteristics. On the other hand, the parish of Chongón and Bolívar are grouped in clusters 0 and 2, because it is possible that the activities carried out in that place may differ from the commercial activities of the other parishes.



Discussion

It is necessary, for a business feasibility analysis, to take into account other theoretical aspects, when studying an area, such as not locating your business where the demand for hotels is very high, where these hotels offer the same service that me. On the other hand, if the center is a very busy area, in the northern part of the city, there are shopping centers and places with wider streets to traffic, being a modern attraction for tourists like Cinemas and interesting Discotheques, with a higher purchasing power, than by on the other hand there are not a majority of hotels.

Conclusion

In this analysis, we can affirm that in the center of the city, although it is a more commercial place, we do not have land or places available to build a hotel. Although, the market trend is being located better to the north of the city, with wider, recreational areas, and with a better tourist aspect. In addition, the competition is lower, where we would have the opportunity to charge a little more for our service, due to the purchasing power of the area. Meanwhile, that the analyzes of starting a new business suggest that analyzing the type of area would be an essential step in order to gather information about the type of commercial area in which we intend to install the new hotel. For example if we have the legal permission to build a hotel in that area, if the streets are crowded by pedestrians, and if the capital gain in the area is increasing.