THE BEST SPOT TO OPEN A NEW HOTEL IN BOGOTA

PAULO MAURICIO SÁNCHEZ B

March, 2019

1 INTRODUCTION

The hotel I am working is currently engaged in expanding their business to new cities in Colombia. Right now we have a big operation in the northern part of the country but the vast majority of new visitors is going to the inner cities like Bogota and Medellin.

Our CEO is interested in opening a new hotel in Bogota because is the capital city of the country and receives every year a lot of visitor not only for leisure purposes but also people on business trips. The main problem is that the city is so big and of all the possible locations we have to select the most profitable one.

Besides my CEO and some other employees of the hotel I think this work is important also for external public interested in finding out what are the neighbourhoods or localities with the highest concentration of restaurant and hotels in Bogota. I am going to cluster the neighbourhoods of Bogota in order to find the group of neighbourhoods more suitable for opening a new business.

This work is composed of six sections being the first one this introduction. The second section is going to explain the data used, thrid section is going to show the methodology used. Fourth section will present the results, fifth section is going to discuss about some possible recommendations and finally section six is going to conclude.

2 DATA

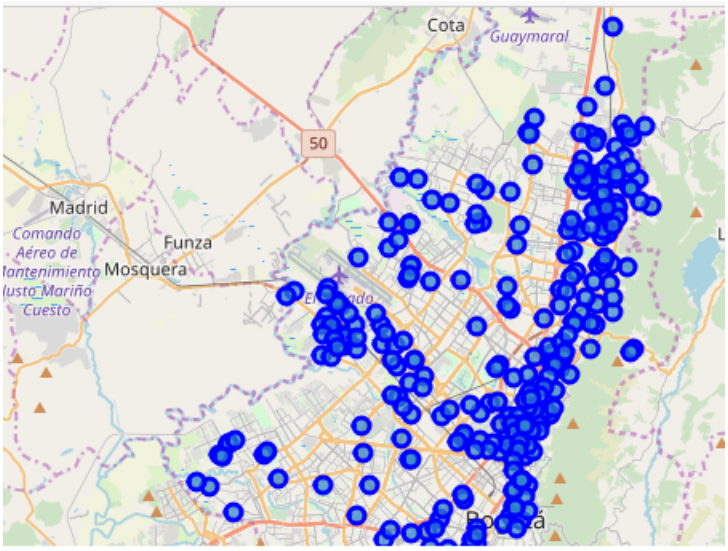
The data used for this research is get from wilkipedia, you can visit the page Localities of Bogota. Bogota has 20 localities and inside these localities there are UPZ which are major areas that contain neighborhoods. Not all localities have a relevance for tourism because some of then are dangerous places and do not have important venues nearby. I chosen the six more relevant localities for tourism: Usaquen, Chapinero, Fontibon, Teusaquillo, La Candelaria and Puente Aranda.

The information about each locality is also in Wilkipedia Example of locality data.

- Usaquen is known for being place of entrepreneurial areas but also is the heart of some of the most iconic neighbourhoods in the city.
- Chapinero is a locality known for the party and gastronomy. With bars and cultural events everywhere is recognized also for being the heart of the financial district in Bogota.
- Fontibon is important because the airport is part of this locality. Besides there are some big business complexes.
- Teusaquillo is known for being the locality of artists. In this locality is located the most recognized university of the country.
- La candelaria is the heart of the city because the streets are colonial and is home of the majority of museums as well of the presidential house.

This data was collected in a csv file and upload to Github. Before import to python was cleaned and manipulated in the csv file.

Locality	UPZ	Neighborhood
Chapinero	Gran Chapinero	Chapinero Central
⋮	⋮	⋮
Usaquen	Paseo de los libertadores	Canaima



The data from Github is uploaded in Python and in order to geolocate each neighbourhood is identified the latitude and longitude. After we have the coordinates we look in four square the most common venues within a radius of 500.

3 METHODOLOGY

In order to identify which neighbourhoods are the more appropriate for opening a new hotel we decided to cluster the data. The algorithm used was kmeans.

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

K-means clustering algorithm

1. Start with initial guesses for cluster centers (centroids)
2. For each data point, find closest cluster center (partitioning step)
3. Replace each centroid by average of data points in its partition
4. Iterate 1+2 until convergence

Write $x_i = (x_{i1}, ...x_{ip})$:
If centroids are $m_1, m_2, ...m_k$, and partitions are
$c_1, c_2, ...c_k$, then one can show that K-means converges to a *local* minimum of
Euclidean distance

$$\sum_{k=1}^{K} \sum_{i \in c_k} ||x_i - m_k||^2$$

(within cluster sum of squares)

The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means due to the name. Applying the 1-nearest neighbor classifier to the cluster centers obtained by k-means classifies new data into the existing clusters.

The critical point about this algorithm is the selection of a starting point or number of cluster. Because we are conscious of the variety and diversity of the neigborhoods we decided to use 10 clusters. Altough we do not validate the number of clusters in a technical fashion we run the exercise using different numbers of clusters and our results were robust because less than 15 clusters sub-identify the neighbourhoods.

4 RESULTS

We are going to show the cluster that in our characterization was the most appropriate for opening our hotel. This cluster include neighbourhoods that have as most important venue hotels but are also near shopping malls and restaurants. We decided to discard the other clusters because according to the most common venues appear to contain primarily residential areas. Besides the cluster we choose it contains neighbourhoods near the airport as well as some neighbourhoods near or in the financial district in Chapinero. Another important factor for the decision of this cluster as the right one is that among the others it contains all the neighbourhoods in which the hotel is the first venue.

Locality	Neighborhood	1st most common venue	2nd most common venue	3rd most common venue
Usaquen	Capri	Shopping mall	Restaurant	Mobile phone shop
Usaquen	Bella suiza	Restaurant	Fast food restaurant	Gym
Usaquen	Bellavista	Music store	Pizza place	Pharmacy
Usaquen	Ginebra	Spa	Restaurant	Donut shop
Usaquen	La glorieta	Shopping mall	Mobile phone shop	Food and drink shop
Usaquen	Las delicias del carmen	Mobile phone shop	Dessert shop	Department store
Usaquen	San Gabriel	Sandwich place	Bakery	Art museum
Usaquen	Santa ana occidental	Hotel	Park	Cafe
Usaquen	Santa Barbara	Shopping mall	Pizza place	Pub
Usaquen	Santa Barbara alta	Shopping mall	Pizza place	Pub
Usaquen	La calleja	Park	Market	Theater
Usaquen	Santa bibiana	Hotel	Restaurant	Steakhouse
Chapinero	Chico reservado	Hotel	Burger joint	Bakery
Chapinero	Bellavista	Music store	Pizza place	Pharmacy
Chapinero	Seminario	Restaurant	Bakery	Shoe store
Chapinero	Palomar	Dog run	Tea food	Furniture
Chapinero	Chico norte	Hotel	Burger joint	Bakery
Chapinero	Chico norte II	Hotel	Burger joint	Bakery
Chapinero	Chico occidental	Hotel	Burger joint	Bakery
Teusaquillo	Acevedo tejada	Restauant	Diner	Chinese restaurant
Teusaquillo	El recuerdo	Hotel	Fast food restaurant	Pizza place
Teusaquillo	Gran america	Latin America restaurant	Restaurant	Hotel
Teusaquillo	Quinta paredes	Bakery	Pizza place	Hotel
Teusaquillo	Ortezal	Hotel	Bakery	Pizza place
Teusaquillo	Ciudad salitre surorinetal	Restaurant	Hotel	Japanese restaurant
La candelaria	Santa bárbara	Shopping mall	Pizza place	Pub
Fontibon	Bahia solano	Moroccan restaurant	Restaurant	Office supplies
Fontibon	Fontibon centro	Shopping mall	Restaurant	Mobile phone shop
Fontibon	La cabaña	Shopping mall	BBQ joint	Yoga studio
Fontibon	La giralda	Hotel	Food and drink shop	Park
Fontibon	Bohios	Park	Gym	Pizza place
Fontibon	Florencia	Mexican restaurant	Shopping mall	Restaurant
Fontibon	La estacion	Shopping mall	Seafood restaurant	Mobile phone shop
Fontibon	Capellania	Hotel	Gym	Breakfast spot
Fontibon	Cofradia	Public art	Restaurant	Art museum

5 DISCUSSION

We believe the neighbourhoods in the cluster previously identified are the correct spots to open a new hotel. However we did some research and some of the neighbourhoods are residential this fact do not exclude the fact that these neighbourhoods are near the hot spots. As far as we could research these neighbourhoods are well classified mainly because:

1. The airport is near the neigborhood identified in Fontibon
2. The most important convention center (Corferias) is also at the heart of one neighbourhood.
3. The stadium and the most important university is also near the neighbourhoods in the cluster.
4. Thecolonial down-town in La Candelaria is also part of the neighbourhoods in the cluster.
5. Maini spots in Chapinero and Usaquen are also included.

The algorithm used (k-means) was the simplest but the results were good. The work is open to the possibility to use more sophisticated algorithms as well as to use a correct method to identify the number of cluster t use as initial guess.

Although we did not include all the localities in the city we are confident that the six localities included are representative. However we believe an inclusion of the complete list of neighbourhoods and localities could be an interesting exercise in order to demonstrate the robustness of our exercise.

6 CONCLUSION

The profits of the project are above the scope of this research. However we now know the correct spots to start a new project in Bogota. Further research shows that the neighbourhoods identified are the appropriate ones because contain the most important places in the city.

The geo-coded data as well as the machine learning technique proved to be useful because was a good classifier for our exercise.