The best spot to open a new Hotel in Bogota

Paulo Mauricio Sanchez B

March, 2019

Outline

1 Introduction

2 Data

3 Methodology

4 Results and Discussion

5 Conclusion

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.

- Besides the 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.

A picture of Bogota: La Candelaria

Data: Localities chosen

- 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.
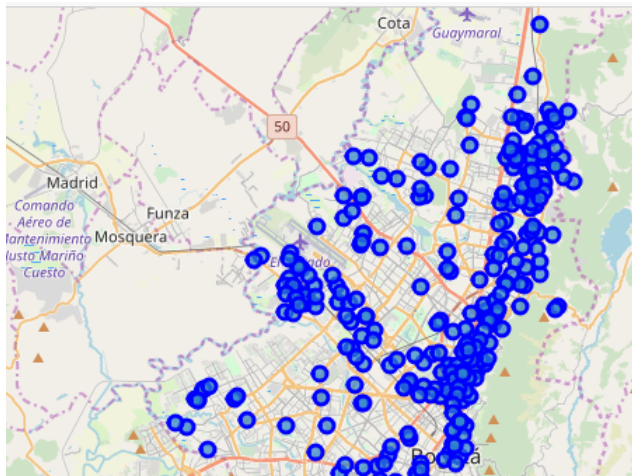
Data: Localities chosen

- 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.

Example of dataset

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

Plotting the data

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.

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 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.

After of characterizing the clusters identified, we select the cluster number four because:

1. t contains neighbourhoods near the airport as well as some neighbourhoods near or in the financial district in Chapinero
2. it contains all the neighbourhoods in which the hotel is the first venue.

Locality	Neighborhood	1st most common venue	2nd most common venue
Usaquen	Capri	Shopping mall	Restaurant
Usaquen	Bellavista	Music store	Pizza place
Usaquen	Ginebra	Spa	Restaurant
Usaquen	La glorieta	Shopping mall	Mobile phone shop
Usaquen	Las delicias del carmen	Mobile phone shop	Dessert shop
Usaquen	San Gabriel	Sandwich place	Bakery
Usaquen	Santa ana occidental	Hotel	Park
Usaquen	Santa Barbara	Shopping mall	Pizza place
Usaquen	Santa Barbara alta	Shopping mall	Pizza place
Usaquen	La calleja	Park	Market
Usaquen	Santa bibiana	Hotel	Restaurant
Chapinero	Chico reservado	Hotel	Burger joint
Chapinero	Bellavista	Music store	Pizza place
Chapinero	Seminario	Restaurant	Bakery
Chapinero	Palomar	Dog run	Tea food
Chapinero	Chico norte	Hotel	Burger joint
Chapinero	Chico norte II	Hotel	Burger joint
Chapinero	Chico occidental	Hotel	Burger joint
Teusaquillo	Acevedo tejada	Restauant	Diner
Teusaquillo	El recuerdo	Hotel	Fast food restaurant
Teusaquillo	Gran america	Latin America restaurant	Restaurant
Teusaquillo	Quinta paredes	Bakery	Pizza place
Teusaquillo	Ortezal	Hotel	Bakery e
Teusaquillo	Ciudad salitre surorinetal	Restaurant	Hotel
La candelaria	Santa bárbara	Shopping mall	Pizza place
Fontibon	Fontibon centro	Shopping mall	Restaurant
Fontibon	La giralda	Hotel	Food and drink shop
Fontibon	Florencia	Mexican restaurant	Shopping mall
Fontibon	Capellania	Hotel	Gym
Fontibon	Cofradia	Public art	Restaurant

Discussion

We believe the neighbourhoods in the cluster previously identified
are the correct spots to open a new hotel.

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 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 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.