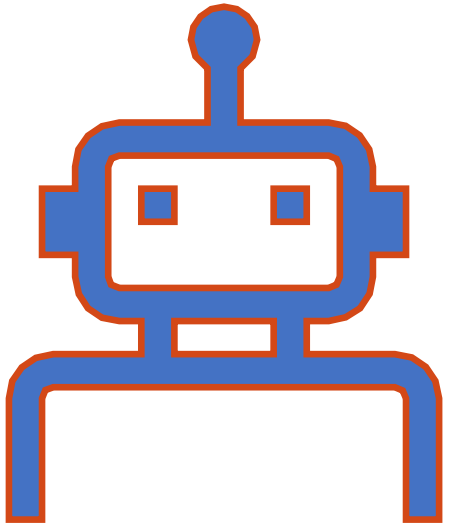
A WAY TO FIND THE BEST PLACE

CLUSTERING THE NEIGHBOURHOODS OF BOGOTA

Richard Alejandro Mora Perilla



INTRODUCTION

- Normally when people travel they make a plan or a routine to be able to know certain places of the site to which they are going to travel, however, prior to this, people usually look for those relevant places to which they would like to go according to the opinions or comments of the people who are near there.

- The problem to be solved is that of which sites could be known according to the good opinions of people from places close to that site, in order to make a decision before making your trip also this data collection can also help to recognize a good place to live in the city.

DATA DESCRIPTION

We require geolocation data for both Bogota. The city's zip codes serve as a starting point. Using the zip codes we use we can find the most popular towns, districts, places and their categories of places.

- **Bogota**

To derive our solution, We scrape our data from https://es.wikipedia.org/wiki/Anexo:Localidades_de_Bogot%C3%A1

This wikipedia page has information about all the localities.

1. *town* : Name of the localitie.
2. *post_code* : Postal codes.

FOURSQUARE API DATA

We will need data about different venues. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

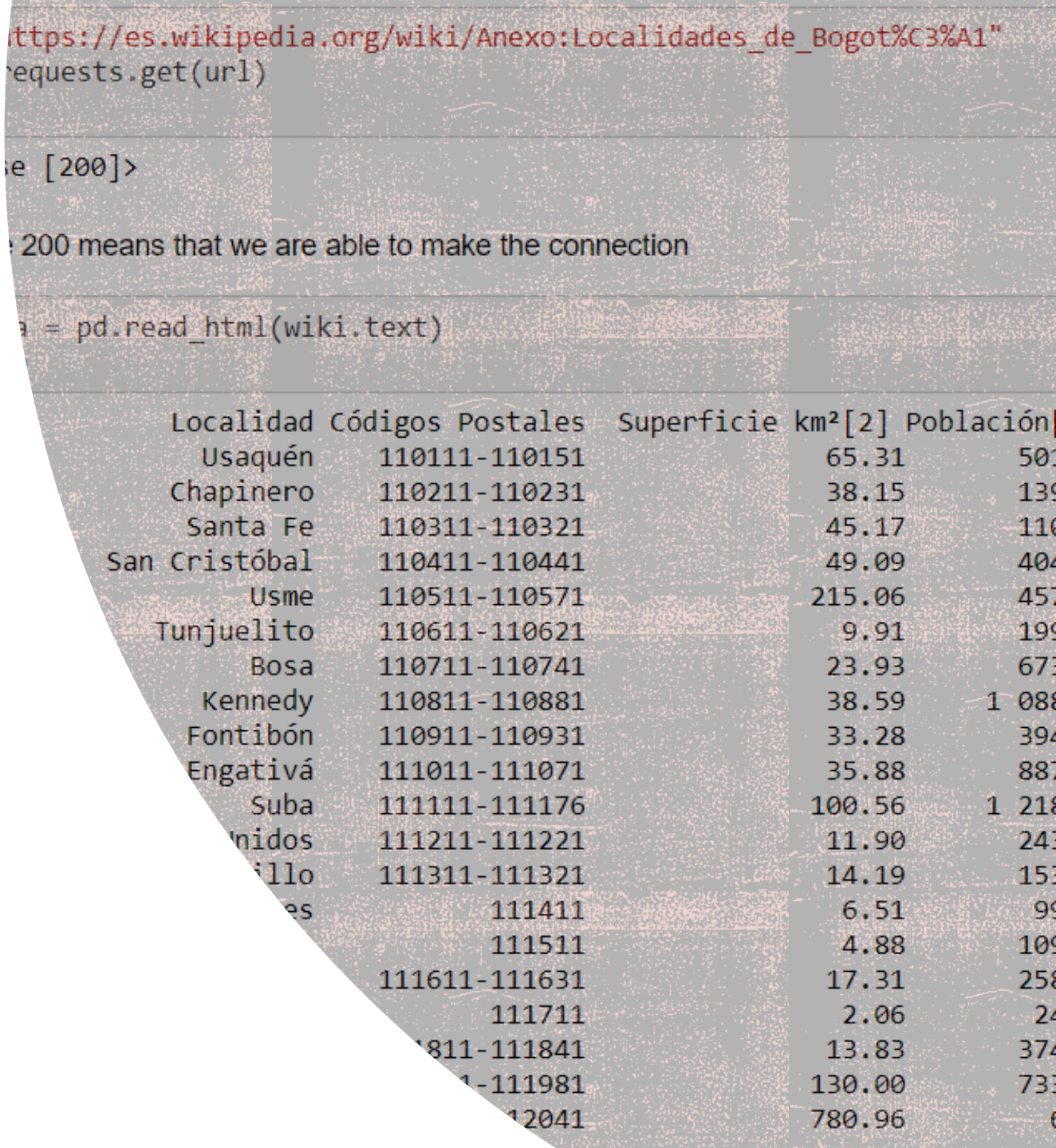
Neighborhood Neighborhood Latitude Neighborhood Longitude Venue Name of the venue e.g. the name of a store or restaurant Venue Latitude Venue Longitude Venue Category

We will be creating our model with the help of Python so we start off by importing all the required packages.

In [1]:
```
import pandas as pd
import requests
import numpy as np
import matplotlib.cm as cm
import matplotlib.colors as colors
import folium
from sklearn.cluster import KMeans
```

METHODOLOGY

- **Localities**
- We begin to start collecting and refining the data needed for the business solution.
- **Data Collection**
- To get the localities in Bogota, we start by scraping the list of areas of Bogota wiki page.



- **Data Preprocessing**
- we remove the spaces in the column titles and then we add _ between words.

```
umns=lambda x: x.strip().replace(" ", "_"), inplace=True)
```

	Códigos_Postales	Superficie_km²[2]	Población[3]	Densidad_hab/km²
	110111-110151	65.31	501 999	7 686.4
	110211-110231	38.15	139 701	3 661.88
	110311-110321	45.17	110 048	2 436.3
	110411-110441	49.09	404 697	8 243.98
	110511-110571	215.06	457 302	2 126.39
	110611-110621	9.91	199 430	20 124.11
	10711-110741	23.93	673 077	28 126.91
	0811-110881	38.59	1 088 443	28 205.31
	911-110931	33.28	394 648	11 858.41
	1-111071	35.88	887 080	24 723.52
	11176	100.56	1 218 513	12 117.27
	1221	11.90	243 465	20 459.24
	21	14.19	153 025	10 784
		6.51	99 119	15 225.65
		4.88	109 176	22 372.12
		31	258 287	14 921.25
			24 088	11 693.2
			374 246	27 060.44
			733 859	5 442.83
		531		8.36

We need to limit the postal codes of the localities to only one.

In [9]:
```
df['post_code'] = df['post_code'].str.split('-',expand=True)
df
```

Out[9]:

	town	post_code
0	Usaquén	110111
1	Chapinero	110211
2	Santa Fe	110311
3	San Cristóbal	110411
4	Usme	110511
5	Tunjuelito	110611
6	Bosa	110711
7	Kennedy	110811
8	Fontibón	110911
9	Engativá	111011
10	Suba	111111
11	Barrios Unidos	111211
12	Teusaquillo	111311
13	Los Mártires	111411
14	Antonio Nariño	111511
15	Puente Aranda	111611
16	La Candelaria	111711
17	Rafael Uribe Uribe	111811
18	Ciudad Bolívar	111911
19	Sumapaz	112011

We change the column names for athe english language

In [8]:
```
df.columns = ['town','post_code']
df
```

Out[8]:

	town	post_code
0	Usaquén	110111-110151
1	Chapinero	110211-110231
2	Santa Fe	110311-110321
3	San Cristóbal	110411-110441
4	Usme	110511-110571
5	Tunjuelito	110611-110621
6	Bosa	110711-110741
7	Kennedy	110811-110881
8	Fontibón	110911-110931
9	Engativá	111011-111071
10	Suba	111111-111176
11	Barrios Unidos	111211-111221
12	Teusaquillo	111311-111321
13	Los Mártires	111411
14	Antonio Nariño	111511
15	Puente Aranda	111611-111631
16	La Candelaria	111711
17	Rafael Uribe Uribe	111811-111841
18	Ciudad Bolívar	111911-111981
19	Sumapaz	112011-112041

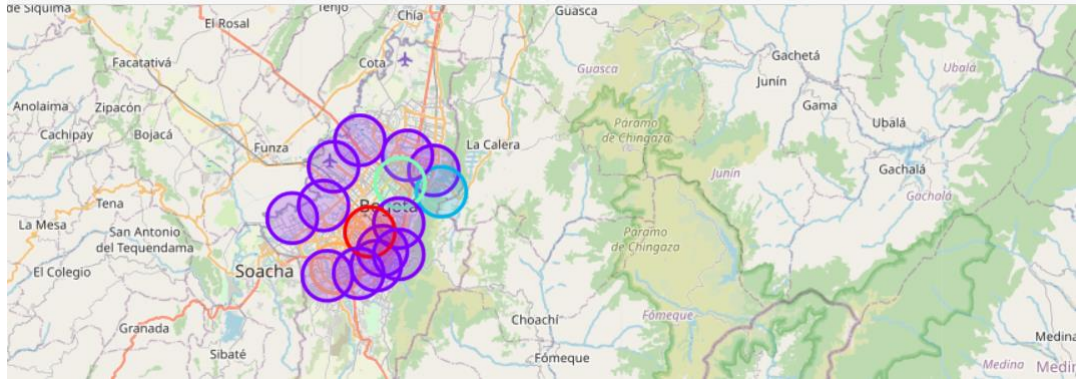
VISUALIZING THE CLUSTERS

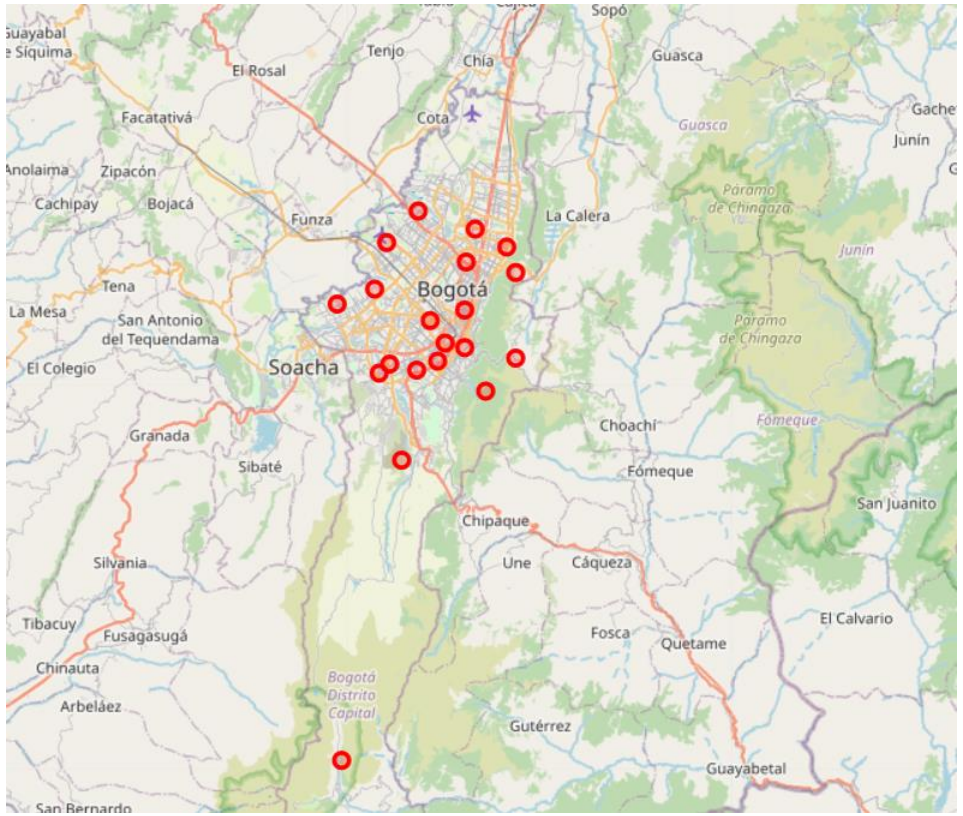
```
ters = folium.Map(location=[bogota_lat_coords, bogota_lng_coords], zoom_start=12)

range(clusters)
+ x + (i*x)**2 for i in range(clusters)]
rray = cm.rainbow(np.linspace(0, 1, len(ys)))
= [colors.rgb2hex(i) for i in colors_array]

colors = []
 lon, poi, cluster in zip(bogota_data['latitude'], bogota_data['longitude'], bogota_data['town'], bogota_data['
l = folium.Popup('Cluster ' + str(int(cluster)) + '\n' + str(poi) , parse_html=True)
um.CircleMarker(
[lat, lon],
radius=20,
popup=label,
color=rainbow[int(cluster-1)],
fill=True,
fill_color=rainbow[int(cluster-1)]
).add_to(map_clusters)

ters
```





EXAMINING OUR CLUSTERS

Cluster 2

```
In [42]: 1 bogota_data.loc[bogota_data['Cluster Labels'] == 2, bogota_data.columns[[1] + list(range(5, bogota_data.shape[1]))]]
```

Out[42]:

	post_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
1	110211	Mountain	Vegetarian / Vegan Restaurant	Gastropub	Cocktail Bar	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store	Cosmetics Shop	Cultural Center

Cluster 3

```
In [43]: 1 bogota_data.loc[bogota_data['Cluster Labels'] == 3, bogota_data.columns[[1] + list(range(5, bogota_data.shape[1]))]]
```

Out[43]:

	post_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
11	111211	Mexican Restaurant	Latin American Restaurant	Vegetarian / Vegan Restaurant	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store	Cosmetics Shop	Cultural Center

Cluster 4

```
In [44]: 1 bogota_data.loc[bogota_data['Cluster Labels'] == 4, bogota_data.columns[[1] + list(range(5, bogota_data.shape[1]))]]
```

Out[44]:

	post_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
5	110611	Restaurant	G								

Cluster 5

```
In [45]: 1 bogota_data.loc[bogota_data['Cluster Labels'] == 5, bogota_data.columns[[1] + list(range(5, bogota_data.shape[1]))]]
```

Out[45]:

	post_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
15	111611	BBQ Joint	Motorcycle Shop	Seafood Restaurant	Bakery	Vegetarian / Vegan Restaurant	Department Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store

Cluster 1

```
In [41]: 1 bogota_data.loc[bogota_data['Cluster Labels'] == 1, bogota_data.columns[[1] + list(range(5, bogota_data.shape[1]))]]
```

Out[41]:

	post_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	110111	Hotel	French Restaurant	Café	Asian Restaurant	Pub	Steakhouse	Bar	Restaurant	Lounge	Nightclub
6	110711	Movie Theater	Ice Cream Shop	Shopping Mall	Pharmacy	Clothing Store	Cocktail Bar	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
7	110811	Burger Joint	Health Food Store	Shopping Mall	Beer Garden	Vegetarian / Vegan Restaurant	Department Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
8	110911	Airport Lounge	Coffee Shop	Café	Pizza Place	Duty-free Shop	Donut Shop	Cosmetics Shop	Gift Shop	Fried Chicken Joint	Cafeteria
9	111011	Pizza Place	Multiplex	Gym / Fitness Center	Park	Bar	Bakery	Seafood Restaurant	Shop & Service	Pub	Toy / Game Store
10	111111	Coffee Shop	Restaurant	Gastropub	Fast Food Restaurant	Park	Seafood Restaurant	Ice Cream Shop	Pub	Cupcake Shop	Cocktail Bar
12	111311	Brewery	Café	Restaurant	Cocktail Bar	Burger Joint	New American Restaurant	Bookstore	Park	Performing Arts Venue	Peruvian Restaurant
13	111411	Shopping Mall	Restaurant	Mobile Phone Shop	Boutique	Department Store	Clothing Store	Duty-free Shop	Drugstore	Cocktail Bar	Coffee Shop
14	111511	Department Store	Clothing Store	Restaurant	Sandwich Place	Deli / Bodega	Pie Shop	Pizza Place	Mobile Phone Shop	BBQ Joint	Salon / Barbershop
16	111711	Café	Italian Restaurant	History Museum	Cocktail Bar	Hostel	Restaurant	Food	Comfort Food Restaurant	Hotel	Burger Joint
17	111811	Neighborhood	Burger Joint	Mexican Restaurant	Pharmacy	Cupcake Shop	Cocktail Bar	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
18	111911	Park	Fast Food Restaurant	Auto Garage	Seafood Restaurant	Vegetarian / Vegan Restaurant	Deli / Bodega	Cocktail Bar	Coffee Shop	Comfort Food Restaurant	Concert Hall

- The towns of Bogota have a diverse number of places to which you could go, the number of towns is few, despite the grouping that it can be seen that Bogota is a multicultural area, with few towns but with a great variety of activities that They can be done depending on what the person is looking for, its restaurants are divided mainly into typical Colombian food, French, Chinese, Italian and sometimes thematic restaurants, it has many museums and green areas that people seem to like very much, many bars and for people looking to exercise all towns have at least one gym.

CONCLUSION

- The purpose of this project was to explore the localities of the city of Bogota and see how attractive it is for tourists, people who live and also for people who would like to live in Bogota. We explore the city based on the zip codes of the localities and then we extrapolate the common places present in each of the neighborhoods and finally we conclude with the grouping of similar neighborhoods.
- We were able to see that each of the towns in the city of Bogota has a wide variety of experiences to offer that are unique in their own way. The cultural diversity is quite evident, which also gives the feeling of inclusion thanks to its section on multinational culture.
- Not all towns seem to offer a vacation getaway or romantic getaway with many places to explore, beautiful landscapes, and a wide variety of cultures. But if there is a large amount that could end up in a great experience to spend a short vacation with a pleasant memory not only for its museums, restaurants and parks, but also for its culture, tourist sites and the wonderful Colombian coffee.

THANKS!

REFERENCES

1. The Battle of Neighbourhood — A Tale of Two Cities by Thomas George
2. Foursquare API
3. ArcGIS API
4. https://es.wikipedia.org/wiki/Anexo:Localidades_de_Bogot%C3%A1