Capstone Project
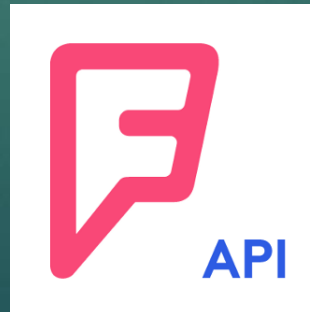
MOVING AROUND THE CITY

Data acquisition and cleaning

- The location of the city is obtain from the Geopy library.

- The data of every hotel in the city it's getting trough the Foursquare API.

  - Foursquare provides Name, location, latitude and longitude position, distance, city, state, country, categories, among others.

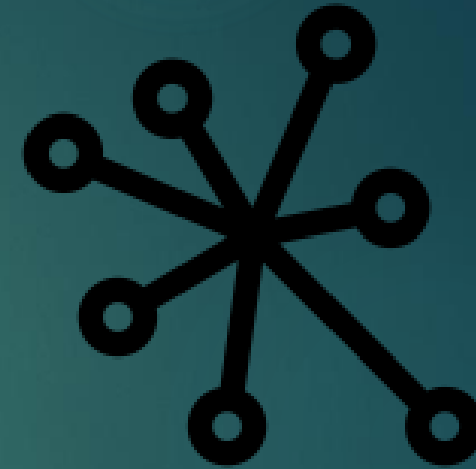 - We clean the information and get into a Dataframe with only the columns Name, categories, latitude and longitude.

	name	categories	lat	lng
0	Hotel Nutibara	Hotel	6.252248	-75.567424
1	Hotel 33 Conquest	Bed & Breakfast	6.239903	-75.582979
2	Hotel GHL Comfort San Diego	Hotel	6.231251	-75.569698
3	Hotel El Porton De San Joaquin	Hotel	6.247125	-75.587399
4	Gran Hotel ****	Hotel	6.251117	-75.562975
5	Hotel Mediterráneo Medellín	Hotel	6.247575	-75.589297
6	Hotel Sauces del Estadio	Hotel	6.257425	-75.585480
7	Hotel el Rami	Hotel	6.246856	-75.591606
8	Hotel Tryp Medellin	Hotel	6.258960	-75.586973
9	Hotel Plaza Mayor	Bed & Breakfast	6.244671	-75.571328

API

Methodology

To start our analysis, we use the K-means clustering that is one of the simplest and popular unsupervised machine learning algorithms.

This algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

After we import the libraries that are needed, we define the number of centroids "K" we need, in the first test we used 4 clusters and added the cluster to the table.
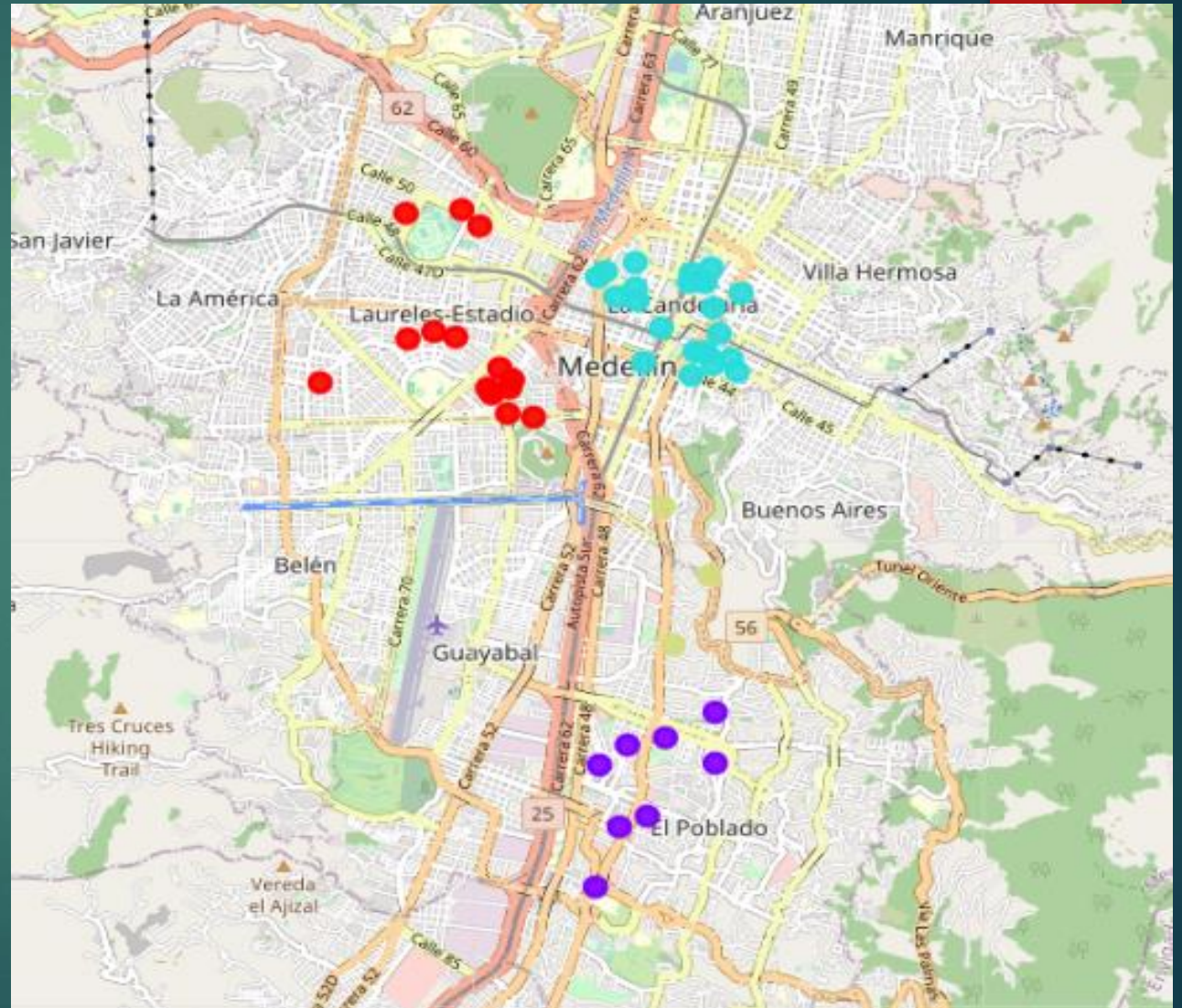
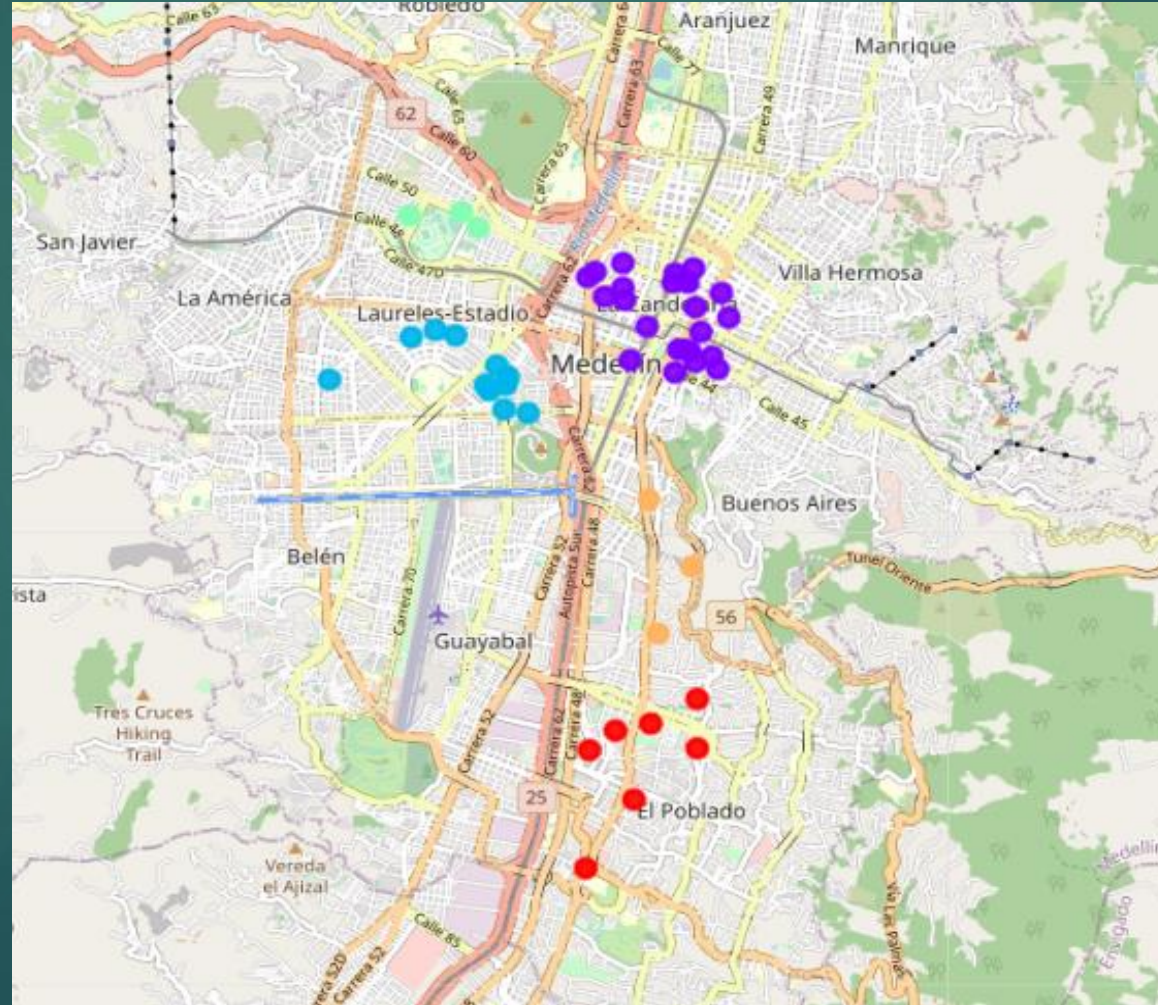|	name	categories	lat	lng	Cluster_labels
9	Hotel Tryp Medellin	Hotel	6.258960	-75.586973	0
20	Hotel Linton	Hotel	6.245569	-75.565846	0
26	Hotel Spiwak	Hotel	6.243015	-75.582685	0
8	Hotel Plaza Mayor	Bed & Breakfast	6.244671	-75.571328	0
1	Hotel 33 Conquest	Bed & Breakfast	6.239903	-75.582979	0
37	Hotel Gallery	Hotel	6.243773	-75.563290	0
6	Hotel el Rami	Hotel	6.246856	-75.591606	0
46	Hotel Poblado Boutique Express	Hotel	6.207089	-75.565102	1
2	Hotel GHL Comfort San Diego	Hotel	6.231251	-75.569698	1
5	Gran Hotel ****	Hotel	6.251117	-75.562975	1

Methodology

Show a map with folium and the coordinates of Medellin city, then added the colors to each cluster and used the latitude and longitude of each hotel to get their position in the map.

Analyzing the map with 4 clusters I realized that the cluster number 1(Red) have 3 points widely separated, so I decide to add 1 cluster more.
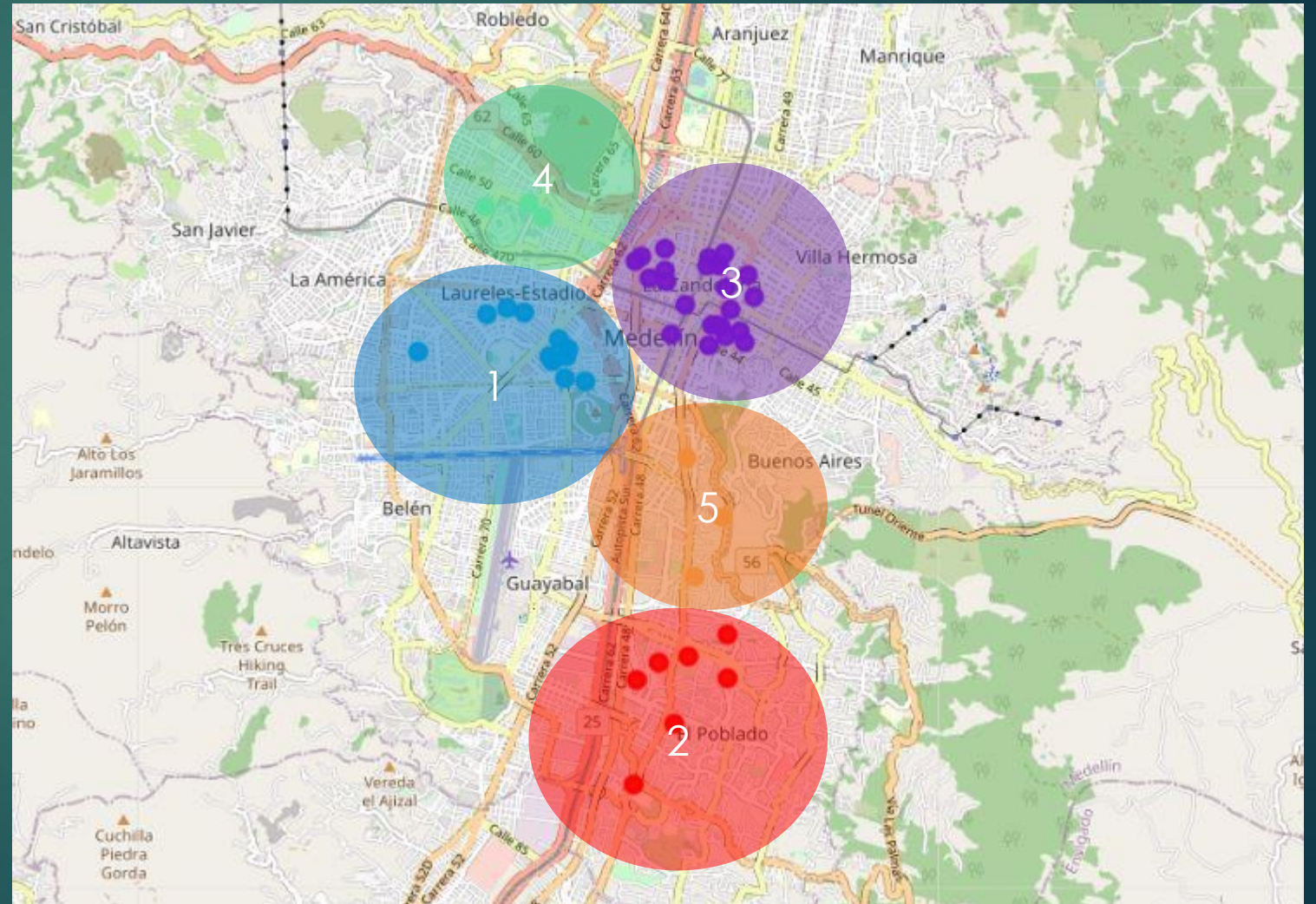
Methodology

Repeating the previous steps and get our map with 5 clusters

Results

Now we can examine each cluster and determine by the position of your work or visiting place, which cluster of hotels can help you to minimize the time of transportation.

Results

Cluster 1

	name	categories	lat	lng	Cluster_labels
46	Hotel Poblado Boutique Express	Hotel	6.207089	-75.565102	0
29	Hotel Loyds	Hotel	6.208798	-75.572760	0
43	Hotel Portón Medellín	Hotel	6.195577	-75.575512	0
30	Hotel Estelar Blue	Hotel	6.202240	-75.570989	0
31	Hotel Du Parc	Hotel	6.209457	-75.569427	0
13	Hotel Park 10 Medellin	Hotel	6.211936	-75.565126	0
35	Hotel Site	Hotel	6.207023	-75.575165	0

Cluster 2

	name	categories	lat	lng	Cluster_labels
0	Hotel Nutibara	Hotel	6.252248	-75.567424	1
48	Hotel El Cid	Hotel	6.253493	-75.565544	1
25	Hotel Suite 45	Hotel	6.247427	-75.564903	1
32	Hotel Boyaca	Hostel	6.253976	-75.572095	1
34	Hotel Internacional	Building	6.249727	-75.565319	1
23	Hotel Medellín Kapital	Hotel	6.252699	-75.575360	1
37	Hotel Gallery	Hotel	6.243773	-75.563290	1
38	Hotel Eupacla	Hotel	6.252389	-75.566443	1
41	Hotel Normandia	Hotel	6.252097	-75.566075	1
44	Hotel Botero Plaza	Hotel	6.253021	-75.567160	1
45	Plaza Hotel Ayacucho	Hotel	6.250777	-75.573787	1
47	Hotel Concept	Building	6.248837	-75.562311	1
33	Hotel San Ignacio	Bed & Breakfast	6.244961	-75.563835	1
21	Hotel El Deportista	Hotel	6.244425	-75.565392	1
24	Hotel Medellin Kapital	Hotel	6.253113	-75.574619	1
49	Metro Hotel	Bed & Breakfast	6.245617	-75.566905	1
19	Hotel 47 Medellin Street	Hotel	6.245654	-75.566438	1
18	Hotel Junimar	Bed & Breakfast	6.251630	-75.572072	1
5	Gran Hotel ****	Hotel	6.251117	-75.562975	1
16	Hotel Conquistadores	Housing Development	6.250402	-75.571851	1
15	Hotel Punto Uno	Hotel	6.243423	-75.567210	1
11	Hotel San Miguel	Hotel	6.247911	-75.569818	1
20	Hotel Linton	Hotel	6.245569	-75.565846	1
8	Hotel Plaza Mayor	Bed & Breakfast	6.244671	-75.571328	1

Results

Cluster 3

	name	categories	lat	lng	Cluster_labels
1	Hotel 33 Conquest	Bed & Breakfast	6.239903	-75.582979	2
3	Hotel El Porton De San Joaquin	Hotel	6.247125	-75.587399	2
4	Hotel Mediterráneo Medellín	Hotel	6.247575	-75.589297	2
42	Hotel Agualongo - Pasto	Hotel	6.242217	-75.584806	2
6	Hotel el Rami	Hotel	6.246856	-75.591606	2
40	Hotel Capilla del Mar	Hotel	6.241770	-75.584420	2
39	Hotel Guadalajara BugA	Hotel	6.241770	-75.584420	2
22	Hotel Platinum Plaza	Hotel Bar	6.239616	-75.580797	2
36	Hotel madison	Residential Building (Apartment / Condo)	6.244201	-75.583692	2
12	Hotel Casa Laurel	Bed & Breakfast	6.242704	-75.599238	2
28	hotel villa de los angeles	Hotel Pool	6.242344	-75.582764	2
27	Hotel Estelar Gran Playa Manzanillo	Hotel	6.243076	-75.582708	2
26	Hotel Spiwak	Hotel	6.243015	-75.582685	2

Cluster 4

	name	categories	lat	lng	Cluster_labels
10	Hotel Manager Obelisk	Hotel	6.258616	-75.591831	3
7	Hotel Sauces del Estadio	Hotel	6.257425	-75.585480	3
9	Hotel Tryp Medellin	Hotel	6.258960	-75.586973	3

Cluster 5

	name	categories	lat	lng	Cluster_labels
14	Hotel Dann Carlton Belfort	Hotel	6.218177	-75.568700	4
2	Hotel GHL Comfort San Diego	Hotel	6.231251	-75.569698	4
17	LQ Hotel by La Quinta	Hotel	6.224661	-75.565661	4

Discussion

- This project is based on the Foursquare API, so our accuracy depends on the update and use of this application in the city, so we compare the project information with a simple research of hotels with google maps in Medellin city, it's provide very similar information, but it's more updated Google maps with some new hotels.

- This project works if you will move in a taxi or uber to minimize the time and distance with the cluster, but also there is an alternative way to move like the Metro system that is around the city.

Conclusion

▶ Based on the information we represent with this clustering analysis, although it's not the most updated, because we find more new hotels with Google, this project still can help someone to look for a place to stay with the clustering results and minimize the time to get around a place in the city.

▶ Also this project can help to identify the zones in the city that don't have hotels, it can be an opportunity to build a new one in a different place.

▶ The methodology and code of this project can be implemented in any city that has information with Foursquare API, the analysis can be very helpful especially if the city is recently updated in the Foursquare data.