The Battle of Neighborhoods in Chapinero, Bogotá, Colombia

The best place to place a restaurant

Choosing the optimal place to open a restaurant

❑ In this project, we will try to find the best location for a restaurant.

❑ Specifically, this report will target stakeholders who are interested in opening a restaurant in Chapinero, Bogotá, Colombia.

❑ We will try to choose locations with a low density of restaurants closer to the neighborhoods of interest.

Explore Dataset

▶ The data was extracted from web pages, in which the latitude and longitude data for each of the neighborhoods can be extracted, in addition to allowing the data to be extracted directly in csv format.

▶ The extraction, cleaning, handling and adjustment of the data was carried out previously, in order to emphasize the analysis of the project and not the operation of the data; for the same reason, the data can be shown already loaded in a pandas dataframe and in its correct structure.

▶ The data is composed of 30 neighborhoods, each with its location in the sector (Borough) and with their respective latitude and longitude values.

2021/6/29

Structure of the data used

Out[1]:

|	Borough	Neighborhood	Latitude	Longitude
0	Chapinero	Sucre	4.632268	-74.066803
1	Chapinero	Cataluna	4.637752	-74.062538
2	Chapinero	Marly	4.639927	-74.063965
3	Chapinero	Chapinero Central	4.644474	-74.065406
4	Chapinero	Chapinero Norte	4.650260	-74.064364

The first 5 rows for the data set obtained are shown here.

2021/6/29

Map of Chapinero with their respective neighborhoods

2021/6/29
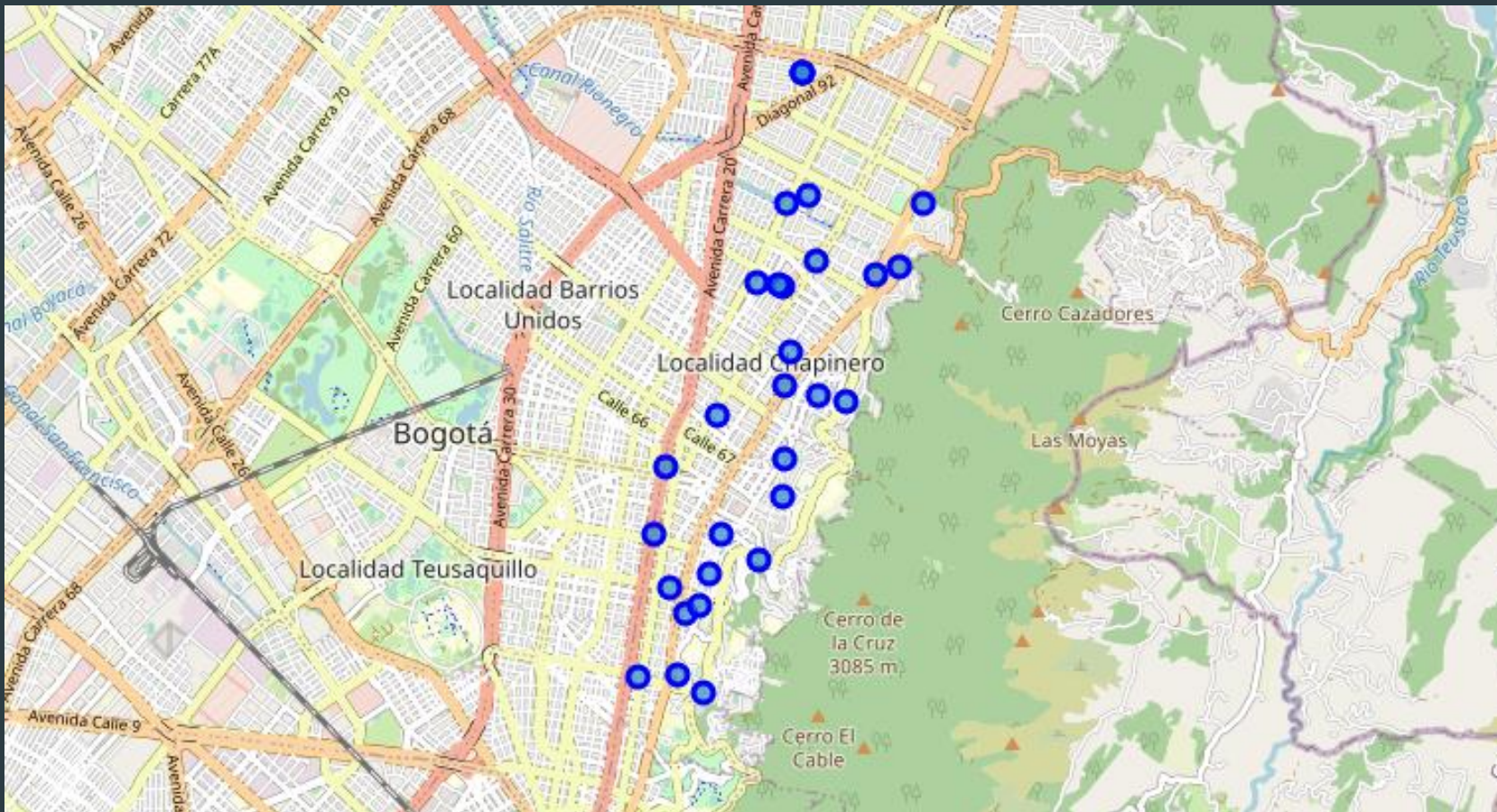
Elbow method to determine the number of clusters

The Elbow Method showing the optimal k
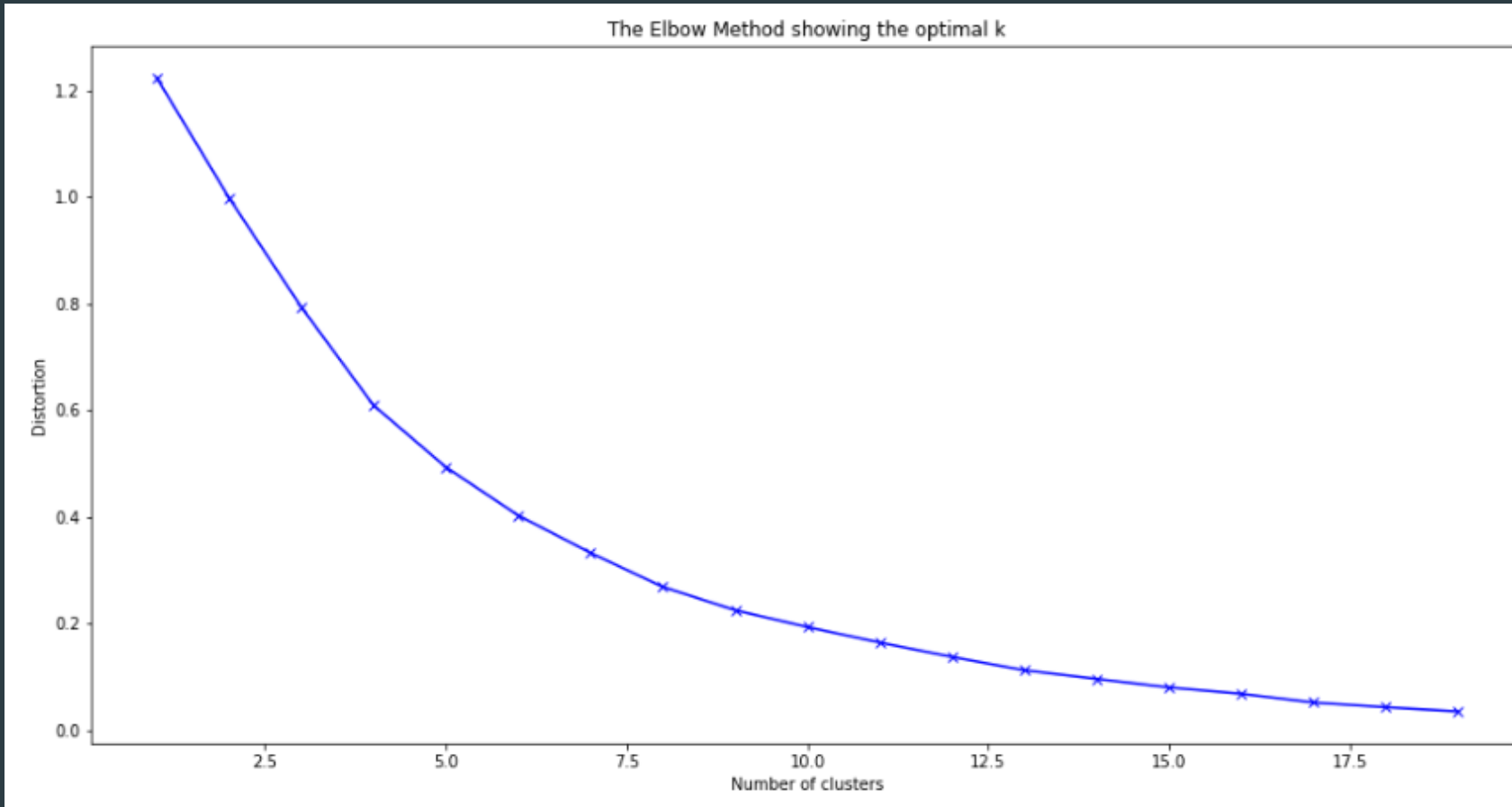
Distortion curve vs number of clusters, to obtain the optimal k.

2021/6/29

Grouping the neighborhoods into 5 clusters

```
In [27]:  # set number of clusters
          kclusters = 5

          # run k-means clustering
          kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(chapinero_grouped_clustering)

          # check cluster labels generated for each row in the dataframe
          kmeans.labels_[0:10]

Out[27]:  array([0, 0, 4, 4, 4, 4, 4, 0, 0, 0], dtype=int32)
```

K-means algorithm applied to 5 clusters.

2021/6/29

Top 10 venues for each neighborhood with their respective cluster label

Out[28]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Chapinero	Sucre	4.632268	-74.066803	4	Fast Food Restaurant	Italian Restaurant	Pizza Place	Café	Bar	Bakery	Music Venue	Sandwich Place	Coffee Shop	Mexican Restaurant
1	Chapinero	Cataluna	4.637752	-74.062538	4	Bar	Fast Food Restaurant	Burger Joint	Restaurant	Café	Nightclub	Bakery	Latin American Restaurant	Vegetarian / Vegan Restaurant	Market
2	Chapinero	Marly	4.639927	-74.063965	4	Bar	Restaurant	Café	Burger Joint	Fast Food Restaurant	Caribbean Restaurant	Nightclub	Latin American Restaurant	Bakery	Tapas Restaurant
3	Chapinero	Chapinero Central	4.644474	-74.065406	4	Restaurant	Gay Bar	Bar	Caribbean Restaurant	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Coffee Shop	Bakery	Sandwich Place	Arepa Restaurant
4	Chapinero	Chapinero Norte	4.650260	-74.064364	4	Nightclub	Sandwich Place	Vegetarian / Vegan Restaurant	Department Store	Theater	Coffee Shop	Soccer Field	Bar	Spanish Restaurant	Mexican Restaurant

The first 5 rows of the neighborhoods with their cluster label and their top 10 most visited venues.

2021/6/29

Map of Chapinero with their neighborhoods in 5 clusters

- ▶ Red cluster: Cluster 1
- ▶ Purple cluster: Cluster 2
- ▶ Blue cluster: Cluster 3
- ▶ Green cluster: Cluster 4
- ▶ Orange cluster: Cluster 5

2021/6/29
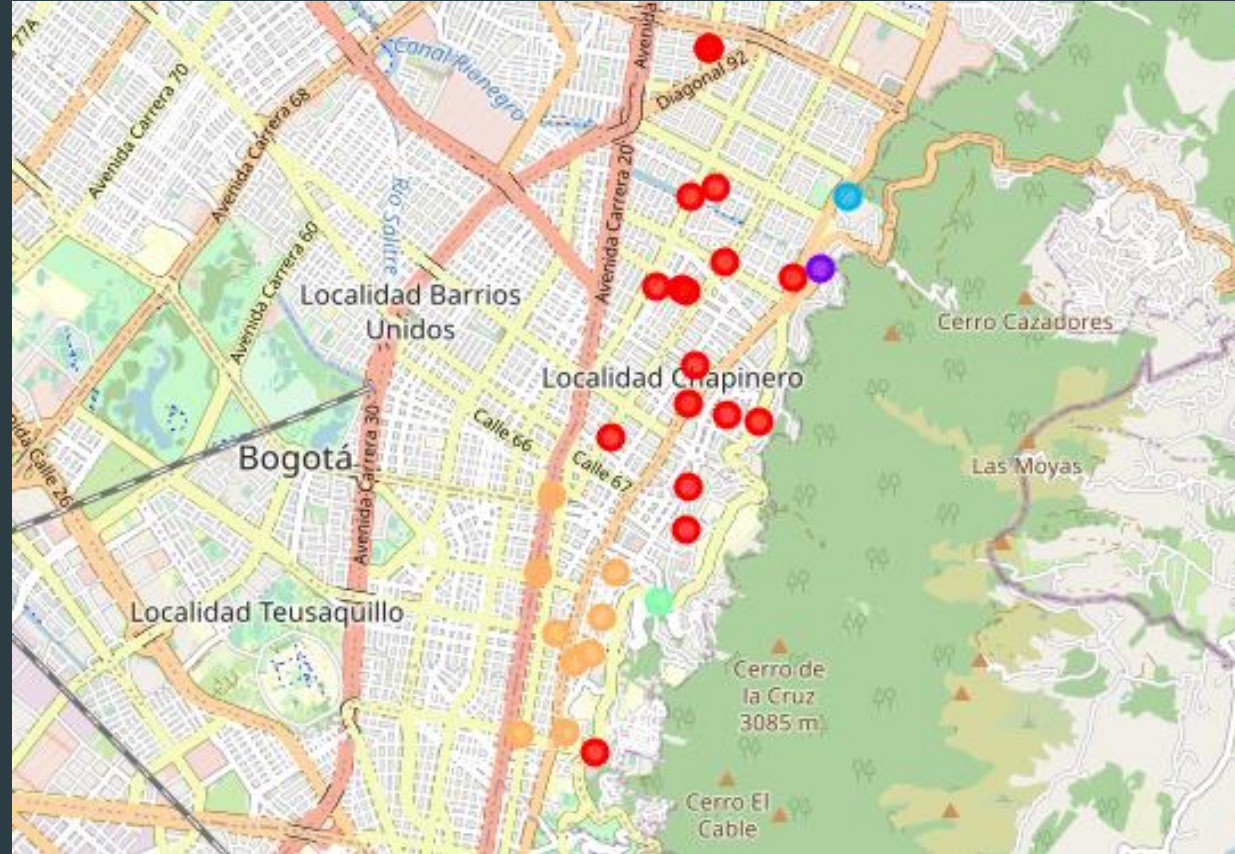
Results and Discussion

▶ Cluster 1 and 5, very high demand, high competition, large flow of people daily, many associated restaurants, cafes and bars.

▶ Cluster 4, high demand, medium competition, medium flow of people daily, only one associated neighborhood, a cafeteria and a restaurant within the top 10 most visited places.

▶ Cluster 2, high demand, low competition, high flow of people daily, only one partner neighborhood, several gyms, resorts and related yoga studios.

▶ Cluster 3, high demand, low competition, high flow of people daily, only one associated neighborhood, several museums, parks, squares and related pharmacies.

2021/6/29

Conclusion

▶ In places within or close to clusters 1 and 5 you can choose the place, to have a very high demand, but there is also high competition, since the sector is quite commercial.

▶ The sector close to cluster 4 generates a lot of uncertainty, due to the varied number of places and some restaurants, cafes and bakeries that would generate some kind of competition in the market.

▶ This project yielded two quite favorable results to facilitate the decision-making of the location of the new restaurant. On the one hand, if you choose a place close to cluster 3, you can take advantage of the advantage of the sector, serving food low in fat, vegetarian, dietetic, high in protein, among others, to considerably increase demand. On the other hand, if you choose a place close to cluster 2, you can take advantage of the cultural sector, for special dates and for family or group meals.

2021/6/29