

IBM Data Science Capstone Project Report

Introduction and Business Problem

For any brick and mortar business a strategic location is key for success, it could literally mean profit or losses.

The project will focus on answering the following question: Which neighborhoods could be the best options to opening a new café in Buenos Aires?

Most influential factors to answer this could be Neighborhood competition, Potential demand, Neighborhood Density, Average Income, and so forth.

That's why the project will gather data about it, process it and come up with the best possible answers to it.

Project itself would be useful for any entrepreneur/franchise that wants to open a new business unit. This project will focus on a particular case for a cafe in Buenos Aires, but it could be replicated for other cities and business.

Data and Sources

- Population of each Neighborhood (Statistics and Census - Buenos Aires City)
- Average Family Income per Neighborhood (Statistics and Census - Buenos Aires City)
- Number of Café in each Neighborhood (Foursquare API)

Related Links:

<https://www.estadisticaciudad.gob.ar/eyc/>

**Note: It's assumed that data it's up to date, complete and there isn't errors on it*

Methodology

First step is to search on the web for related information about the described problem and useful data available.

Second, collect data and wrangle it to required format. Make some plots to get a sense of relationships and basic understanding about the problem.

Then with everything in place, make clusters based on selected features. Analyze them, and look for relationships.

After all, select a cluster based on explained criteria. And explore different neighborhoods within it to get a final conclusion

Results

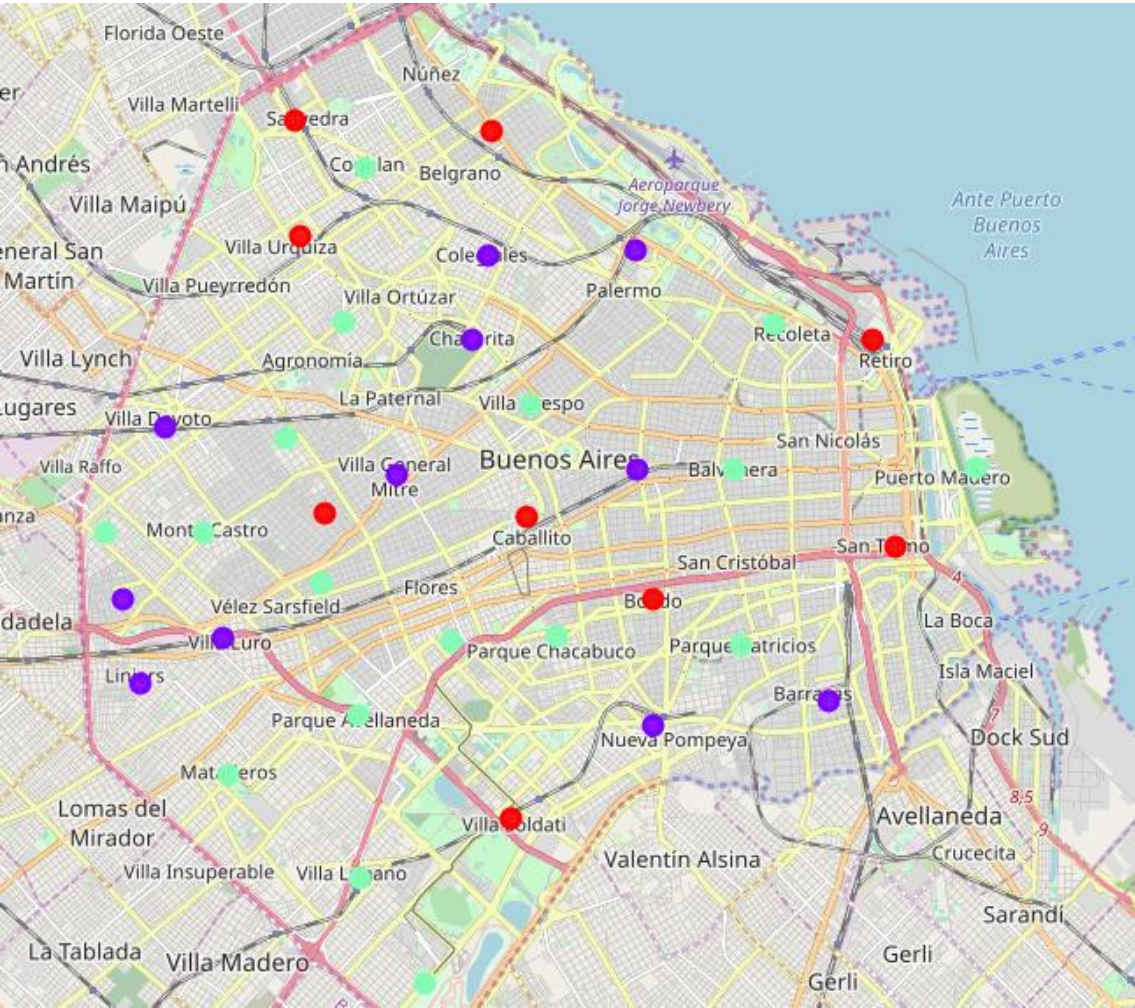


Figure 1: Result of K-means Clustering
Cluster 0 = Red ; Cluster 1 = Purple ; Cluster 2 = Turquoise

Main Characteristics of each Cluster

```
[ ] # Compare each cluster
clust_an.groupby('Cluster Labels').mean()
```

	Habitantes	Ingreso	Café
Cluster Labels			
0	0.269494	0.540356	0.333333
1	0.273616	0.433625	0.155844
2	0.240647	0.343990	0.235294

From the table we can infer that:

- Cluster 0 High Income, High Café number.
- Cluster 1 Medium to High Income, Lower Café number.
- Cluster 2 Lower Income.

Discussion

After observing the results, cluster number 1 is identified as the most promising.

We proceed to analyze the best possible candidates within it.

From the data we observe that each of these neighborhoods has a low supply of Cafes in relation to its population.

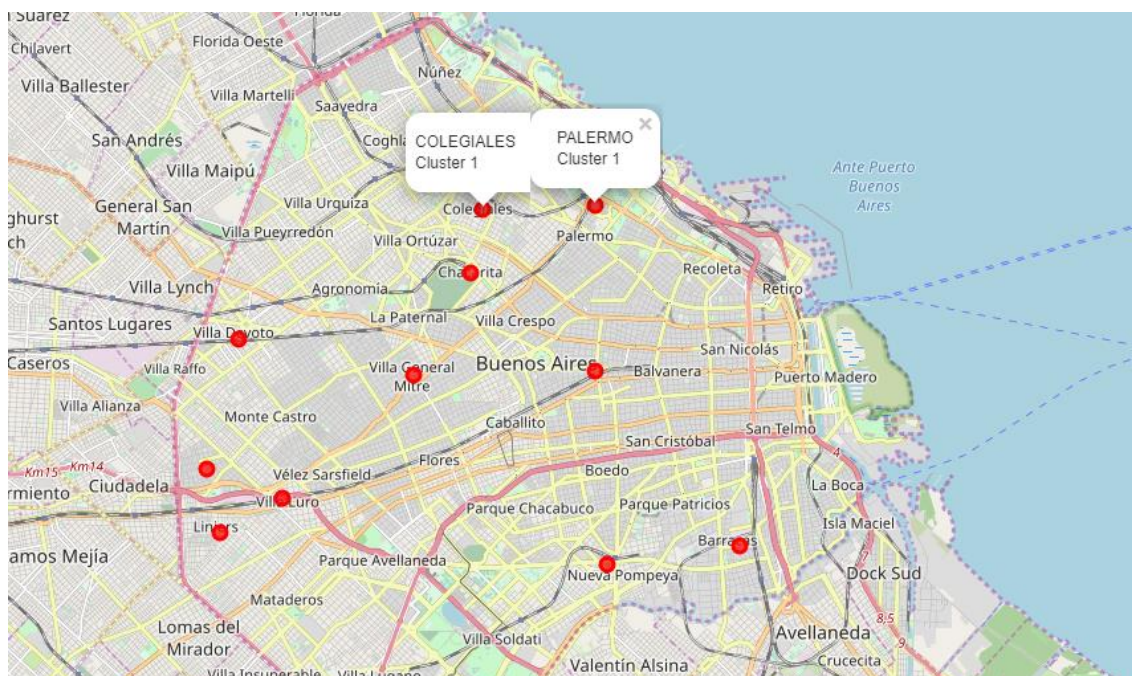
Given this, we decided to order the neighborhoods belonging to the cluster based on their purchasing power to define the 2 main candidates.

```
# Select the top 2 candidates from the cluster, ordered by Income
clust_an.loc[clust_an['Cluster Labels'] == 1].sort_values('Ingreso', ascending=False).head(2)
```

	Habitantes	Ingreso	Café	Neighborhood	Cluster Labels
25	1.000000	1.00000	0.142857	PALERMO	1
7	0.203814	0.89796	0.285714	COLEGIALES	1

The most promising locations to start a new Café would be "Palermo" or "Colegiales"

Let's visualize them:



Conclusion

As mentioned at the beginning, location is key for the business, and the project shows a basic workflow to start narrowing down best possible options.

Although then much later analysis is required to carry out the project regarding different issues such as rental costs, operations, etc.

It is important to start with a solid foundation to maximize the likelihood of research for optimal location.