FINDING THE BEST LOCATION FOR A NEW BUSINESS

Introduction

Buenos Aires is the capital and largest city of Argentina. This financial, industrial, and commercial hub has the highest population and average income of the country. As it is a large financial center, the main public and private banks are located there.

One of the main private banks wants to create a new digital bank, splitting an area and moving some of its employees to the new one. As a digital, low-cost bank, face-to-face customer service won´t be provided and it won't be necessary to stay located at the financial center. However, stakeholders want their young employees to maintain social life and the benefits of the local area, in a less crowded place. They also need to locate the new bank near a subway station, based on employee preferences and travelling costs. Finally, it must be located near hotels and cafes, so stakeholders from all over the country or region could take advantage of the location.

Data

We will use Foursquare location data to solve the problem. It's a very useful tool and it provides information from any kind of venues. Also, datasets provided by Buenos Aires' government will be utilized as input for the Foursquare Developers' API. It seems the best way to find a solution will be clustering stations thus making recommendations based on the results.

Estaciones de Subte

Mapa de las estaciones de subte de la Ciudad.

Previsualización

Explorador de Datos

Tabla | Gráfico | Mapa | 87 registros. | « | 1 – 87

long	lat	id	estacion	linea
-58.3989...	-34.6357...	1	CASEROS	H
-58.4009...	-34.6293...	2	INCLAN ...	H
-58.4023...	-34.6230...	3	HUMBE...	H
-58.4047...	-34.6152...	4	VENEZU...	H

Source: https://data.buenosaires.gob.ar/dataset/subte-estaciones/archivo/199.4

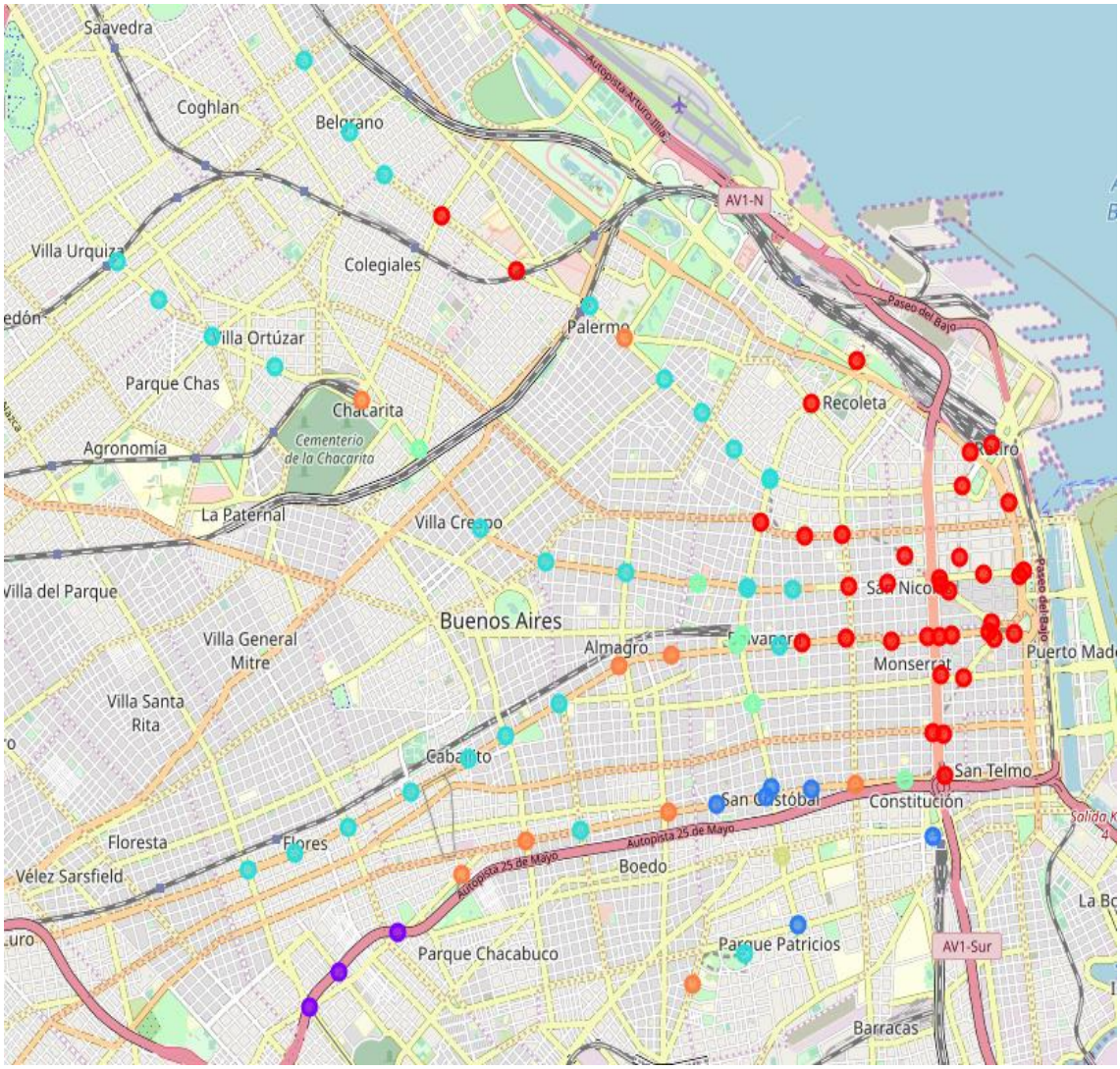
Methodology

First, at business understanding stage we clarify the problem we are trying to solve. In this case, we are looking the better location for a new bank attending to stakeholders' requirements. These are: far from the epicenter of the city, near to a subway station, and located near cafes and hotels. We make an analytic approach to determine how we can use data to answer which is the better location for this new bank.

After completing those steps, we start the process of requiring, collecting, understanding and preparation of the data. We need data from subway station locations gathered from Government's public dataset, and information from Foursquare API. We explore all the dataset and create a map of Buenos Aires Subway Stations. Next, we analyze each station and venues from surrounding area applying one hot encoding, grouping venue categories using inferential statistical (mean), and sorting by most common venue within 500 meters from each station. We deploy an unsupervised machine learning algorithm, k-means clustering, because it gives us similar patterns or characteristics to find the right place to establish the new bank.

Results and key findings

After doing some analysis, we merge all data and visualize the different clusters:



We also examine clusters and show some rows from Cluster 0 (or RED):

	estacion	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4
5	9 DE JULIO	0	Café	Hotel	Coffee Shop	
6	FACULTAD DE MEDICINA	0	Coffee Shop	Hotel	Italian Restaurant	
7	TRIBUNALES - TEATRO COLÓN	0	Coffee Shop	Theater	Hotel	
12	RETIRO	0	Hotel	Coffee Shop	Argentinian Restaurant	
13	LAVALLE	0	Coffee Shop	Café	Hotel	
14	DIAGONAL NORTE	0	Coffee Shop	Café	Hotel	
15	AV. DE MAYO	0	Hotel	Café	Spanish Restaurant	
16	MORENO	0	Spanish Restaurant	Hotel	Argentinian Restaurant	

Discussion

We observe a concentrated red area (Cluster 0) that matches almost every location requirement. This area is full of cafes, hotels and subway stations, even though that area is recognized as the epicenter of the city, where we could find almost every Government office, banks and companies' headquarters. Fortunately, we can find two red points far from the epicenter of the city. They are CARRANZA and OLLEROS, where employees could enjoy the same benefits from previous location plus being far from most crowded zone. We strongly recommend placing the new bank near CARRANZA or OLLEROS to meet the requirements of stakeholders.

Conclusion

Area near stations from Cluster 0 (or RED) clearly seems to be the best positioned to develop the new business. We find CARRANZA and OLLEROS stations at Palermo neighborhood, which are distant from the others, the best suited place for bank's stakeholders and employees.

Appendix

Python notebook code: https://github.com/estebancarboni/IBM-Introduction-to-Machine-Learning/blob/master/Unsupervised%20Learning%20Final%20Project.ipynb