Where to deploy an ATM[1]?

Marcelo Quiroga

September, 2020

1. Introduction

The city of Cochabamba is a small city in the center of Bolivia with a population near to 1 million; this city has the service of different banking entities which have branches and ATMs in the city. Approximately 30 years ago ATMs began to be deployed in this city, along with the inevitable growth of businesses and places of interest, ATMs were accompanying this growth to provide cash to users in a safe and reliable way. Despite the existence of POS for the payment of services or online payments recently deployed, ATMs continue to provide service to customers and their use will be recurrent in the medium term.

The relevance of these devices is given because *"ATMs were designed as simple cash withdrawal terminals. However, with the creation of first multifunctional self-service devices with the deposit function, the role of the banking self-service channel had to be reconsidered. Financial institutions quickly realized that cash-in/cash-out ATMs could be a valid touchpoint for end-customers. The user is granted with the ability to top-up a bank account or an e-wallet, pay fines, fees, utility bills, and various taxes via an ATM".*[2]

1.1. Business Problem

The costs of acquisition, deployment and maintenance of ATMs make it necessary to evaluate where they will be implemented. Poor geographic location results in excessive costs and poor service for end-customers.

1.2. Interest

Any bank can take advantage of the data provided by geolocation tools, which provide updated information on companies or businesses used or recommended by consumers in real time. This data is useful for every financial institution that is considering expand its ATMs or branch locations.

---

[1] Automated Teller Machine

[2] https://www.bs2.lt/en/news-about-products/how-to-deploy-seamless-payment-infrastructure-across-atm-and-kiosk-fleets/

1.3. Scope

In the city of Cochabamba there are different banks, some with many clients and some with few clients. This project is limited to evaluating the availability of ATMs of a single bank[3], however this analysis can be extended to other entities.

2. Data

2.1. Data Sources

The initial source of data will be the current location of the ATMs, this is restricted to the city of Cochabamba and data from other cities within Bolivia will not be used. The data source is found on the bank's website within the branches section where it is necessary to use the filter "Current Location" and "Category". The total data obtained will be recorded as a Python dictionary where the name of the site and the current address are recorded.

```
bank_addresses

{'HIPERMAXI CIRCUNVALACION': 'Hipermaxi Circunvalacion',
 'AG. AMERICA': 'Banco Ganadero Central',
 'AG. LA CANCHA': 'Calle Honduras',
 'HIPERMAXI PRADO': 'Avenida Ballivian 753',
 'OF. CENTRAL': 'Correos',
 'TORRES SOFER': 'Torres Sofer',
 'BLANCO GALINDO': 'Rotonda Peru',
 'SERVICIO DE CAMINOS': 'Avenida Eliodoro Villazón',
 'SURTIDOR EL CRISTO': 'Clinica Los Angeles',
 'IC NORTE': 'Avenida América 817',
 'AMERICA Y MELCHOR': 'Melchor Perez Olguin',
 'HIPER MAXI JUAN DE LA ROSA': 'Hipermaxi Juan Rosa',
 'AEROPUERTO': 'Aeropuerto',
 'U. CATOLICA': 'Plaza Tarija'}
```

Figure 1 – ATM addresses

By creating a dataframe with Pandas and using the GeoPy library, these addresses will be used to obtain the latitude and longitude corresponding to each ATM. It was necessary to verify and correct the addresses for some ATMs more than once because the results of GeoPy were not always accurate, in some cases when the address was not known by GeoPy it was necessary to use a close address (50 meters) to get as close as possible to the actual location of the ATM.

We worked with Foursquare to find the greatest number of venues as possible on a specific area. It is necessary to consider that in a small city (Cochabamba) the number of venues is not always a large number.

[3] https://www.bg.com.bo/

3. Methodology

The area of analysis is an area of 4 Km. radius in the city of Cochabamba, the reason for that is because the most venues are not outside of this area. A second point to consider is the center of this area, it was necessary to relocate this point another than the actual 'center' of the city, this is because the most venues of interest could be stay outside if we use the real 'center' of the city.

The next image shows the area of interest; this is the area where the venues where founded:

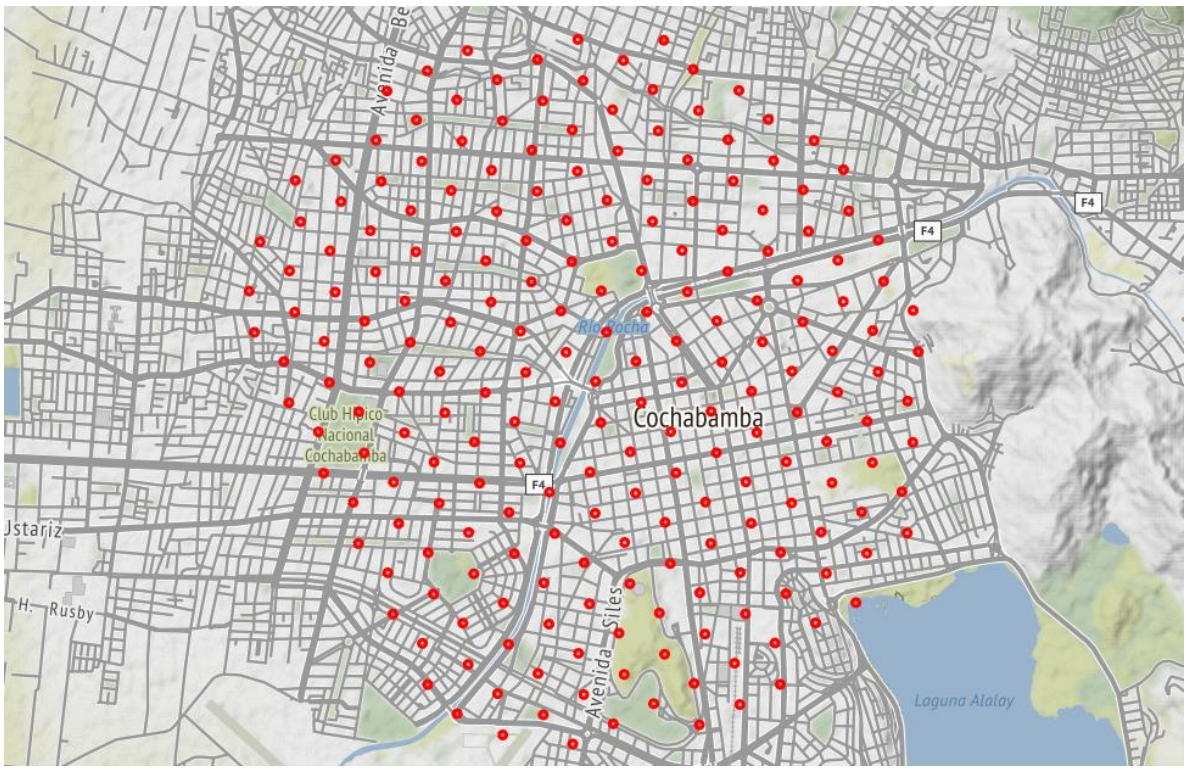


Figure 2 – Area of interest in Cochabamba

The total points of this area is 218, every point has the location data (latitude, longitude) and those where founded using GeoPy library. With this base area defined the next step was search for venues using the Foursquare app. In this process the search tries to find venues around 300 meters of every point (218).

The total venues founded were 835 but is necessary to redefine this number because there are duplicated venues, once the duplicated venues were deleted the total

venues was reduced to 342. May be this number not generate a big expectation but is so useful for this research.

In the next image all the venues founded are added to the map and we can see some areas that has more density than others in the distribution of venues.

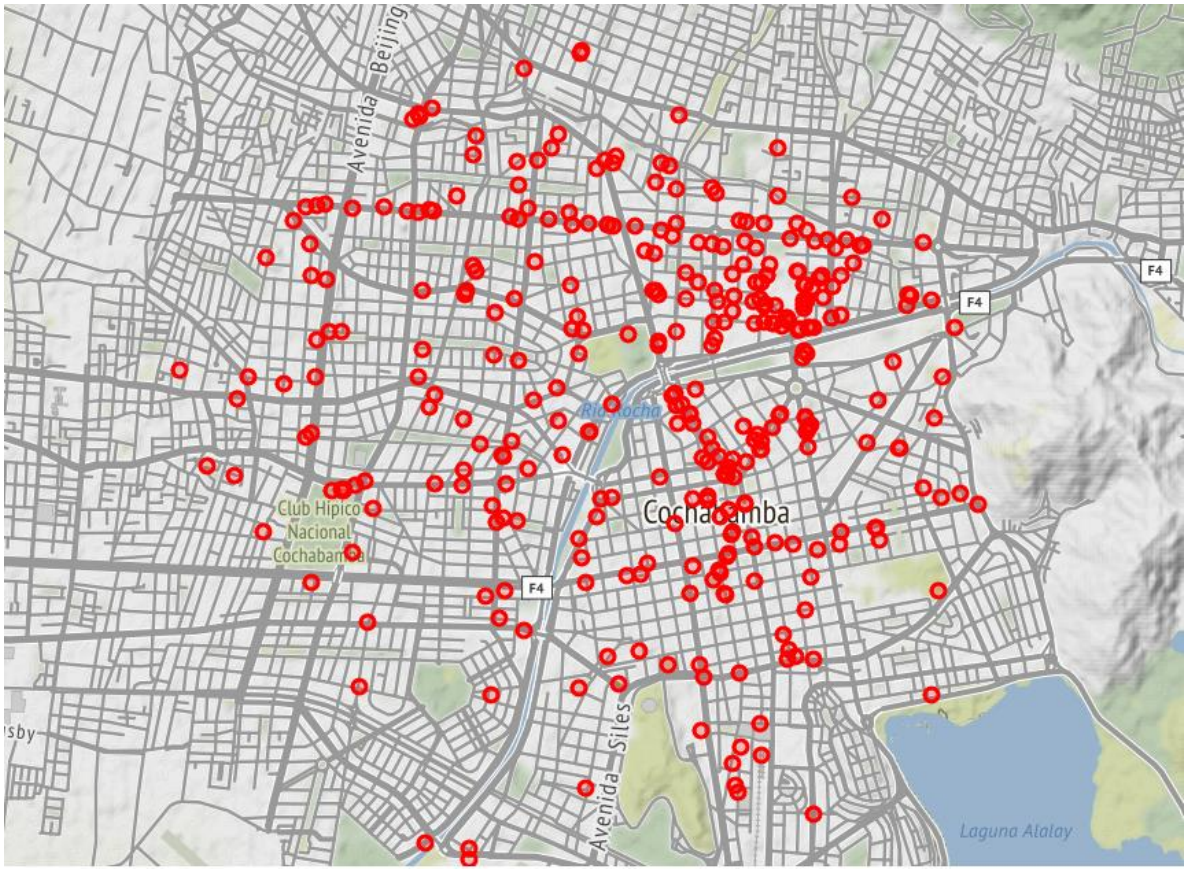


Figure 3 – Total Venues founded

The other side to consider is the actual locations of ATMs, the information needed is the location (latitude, longitude) of that points. In this process the bank has 14 ATMs deployed in Cochabamba, the addresses were founded in the web page of the institution[4]. Again the transformation of addresses to location data was archived with GeoPy library. It was necessary to verify and correct the addresses for some ATMs more than once because the results of GeoPy were not always accurate.

[4] https://www.bg.com.bo/sucursales/

	Office	Latitude	Longitude
0	HIPERMAXI CIRCUNVALACION	-17.363888	-66.165877
1	AG. AMERICA	-17.372293	-66.162408
2	AG. LA CANCHA	-17.401421	-66.152221
3	HIPERMAXI PRADO	-17.384373	-66.159049
4	OF. CENTRAL	-17.392794	-66.158624
5	TORRES SOFER	-17.384613	-66.151078
6	BLANCO GALINDO	-17.393905	-66.170656
7	SERVICIO DE CAMINOS	-17.379576	-66.122469
8	SURTIDOR EL CRISTO	-17.378686	-66.164538
9	IC NORTE	-17.372670	-66.151056
10	AMERICA Y MELCHOR	-17.371528	-66.175702
11	HIPER MAXI JUAN DE LA ROSA	-17.376970	-66.172585
12	AEROPUERTO	-17.409904	-66.173863
13	U. CATOLICA	-17.373754	-66.142962

Figure 4 – ATM addresses

At this point we have two datasets: venues locations and ATM locations (14), if we have in mind these datasets we can image the most simple relation between them: the distance of every venue to every ATM. So the next step is to calculate those distances.

Is not possible to calculate those distances using location data, it is necessary transform that information to obtain the distance in kilometers. So it was defined some functions on Python to translate between World Geodetic System - WGS84 (latitude, longitude coordinates) and Universal Transverse Mercator - UTM (X,Y coordinates).

Applying those transformations and calculating the distances, it was possible to define a new dataset that correlated the venues and ATMs (14), the next figure shows the last four rows:

	Venue	Venue Latitude	Venue Longitude	HIPERMAXI CIRCUNVALACION	AG. AMERICA	AG. LA CANCHA	HIPERMAXI PRADO	OF. CENTRAL	TORRES SOFER	BLANCO GALINDO	SERVICIO DE CAMINOS	SURTIDOR EL CRISTO	IC NORTE	AMERICA Y MELCHOR	HIPER MAXI JUAN DE LA ROSA	AEROPUERTO	U. CATOLICA
338	Eterna store	-17.404624	-66.155805	13717	10792	1538	6709	3973	6718	5841	13306	8928	10572	12519	10484	5932	10886
339	La Cueva	-17.411785	-66.156042	15992	13088	3598	9025	6269	9032	7441	14916	11164	12905	14565	12533	5633	13114
340	Krel Electrónica	-17.412540	-66.156540	16206	13310	3883	9256	6497	9303	7544	15201	11368	13170	14724	12695	5511	13398
341	Celutron	-17.407650	-66.150720	15108	12150	2092	8059	5461	7543	7715	12779	10433	11456	14204	12176	7309	11364

Figure 5 – Venues and ATM correlated

With this dataset, it was applied the K-Means algorithm, the number of clusters (k value) defined was 14 because there are 14 ATMs, so in theory each ATM should be at the center of each cluster. By the moment we will see the number of venues in each cluster:

N° clusters	Mean	Minimum	Maximum
14	24.4	9	51

Figure 6 - clusters

4. Results and Discussion

The first approach in the relation of venues and ATMs locations is observed in the image below. It is clear that some ATMs are too far of venues cloud, the ATM situated on east side is an outliner and has none venue near of him but this is because in the process of search venues that location was not considered. In the south we can see another ATM that has not many venues near of him but this is the airport of the city so that location is a good one.

On the west there is not any ATM; in the center and in the north of the city is where you find the most venues and of course the most ATM deployed.

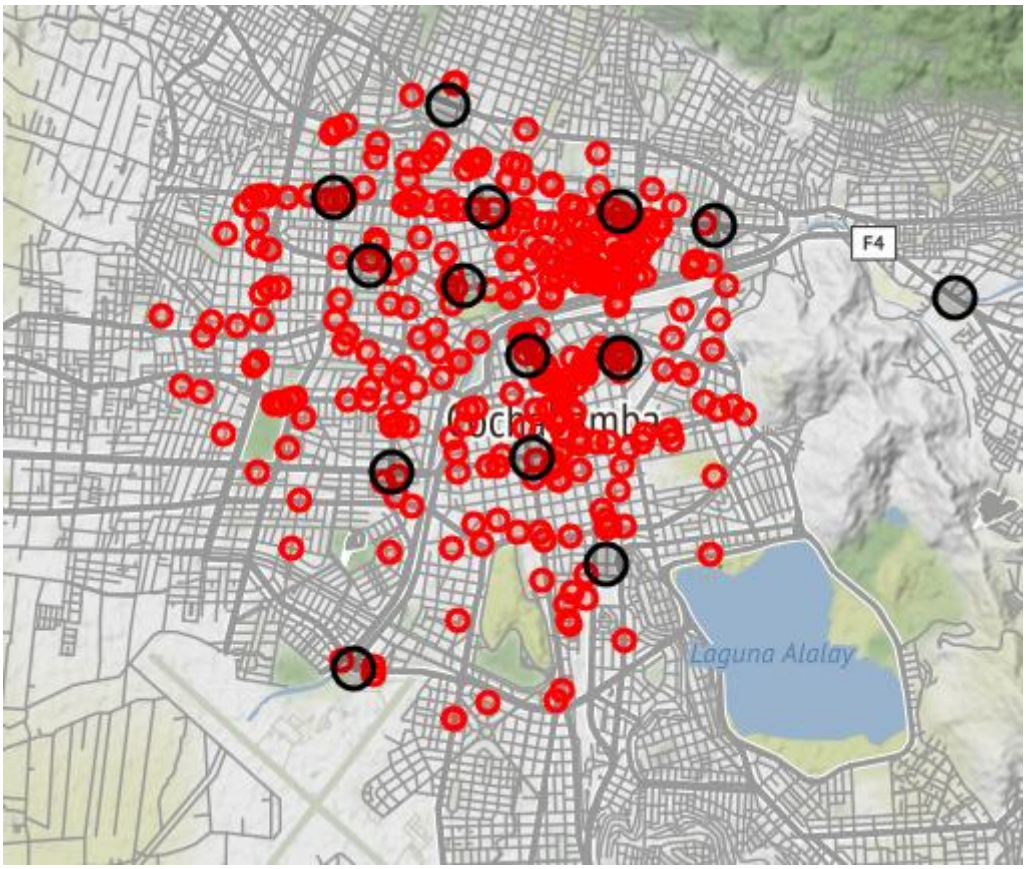
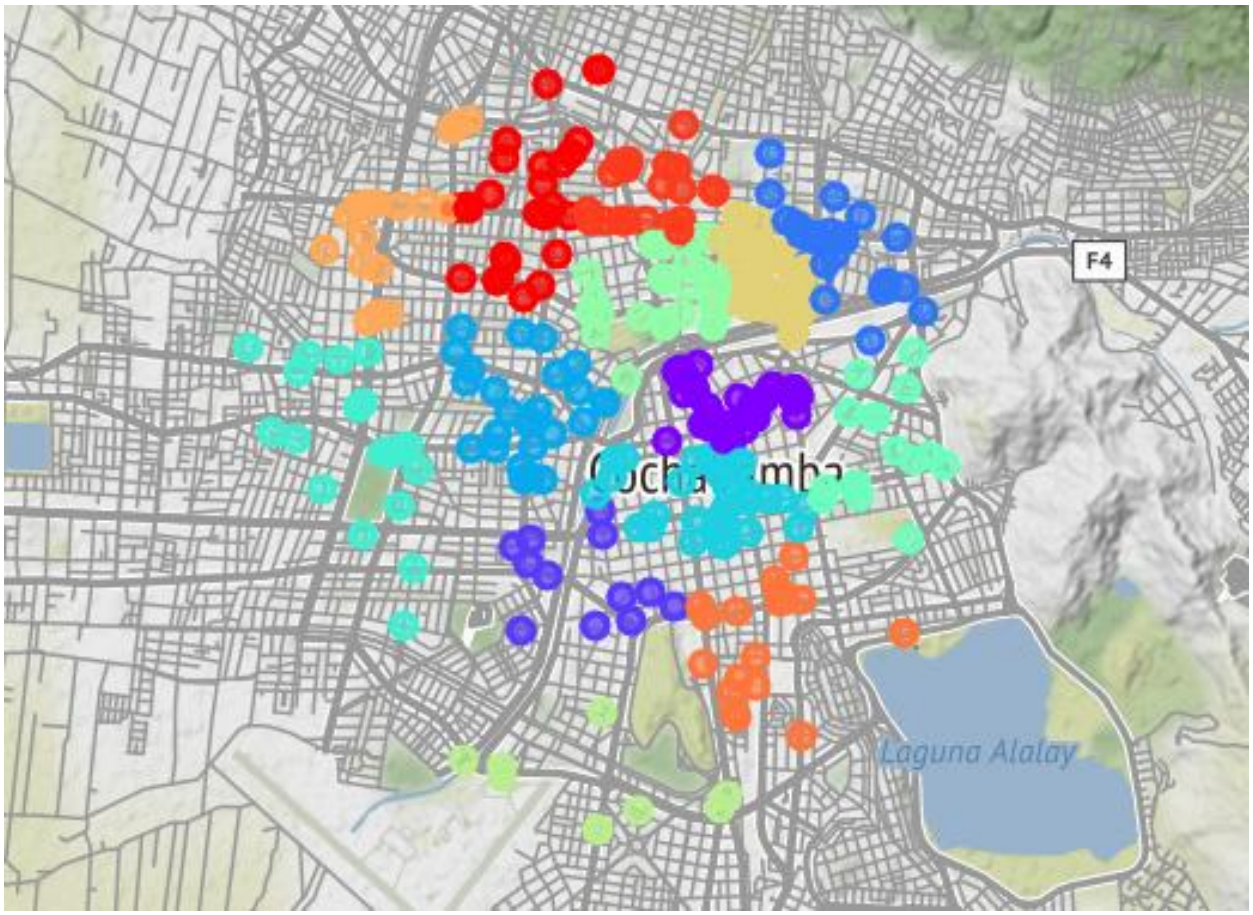
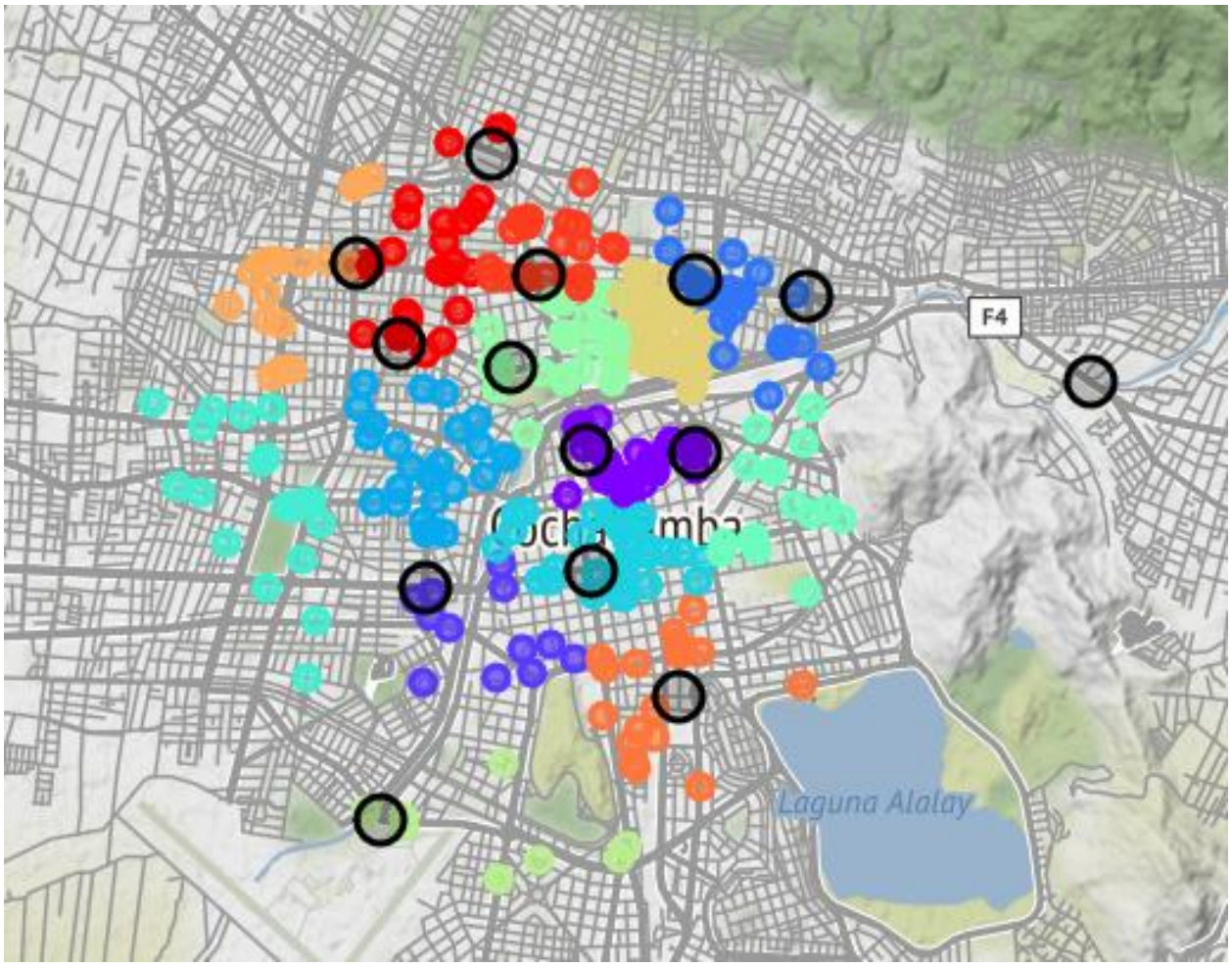


Figure 7 – ATMs vs Venues

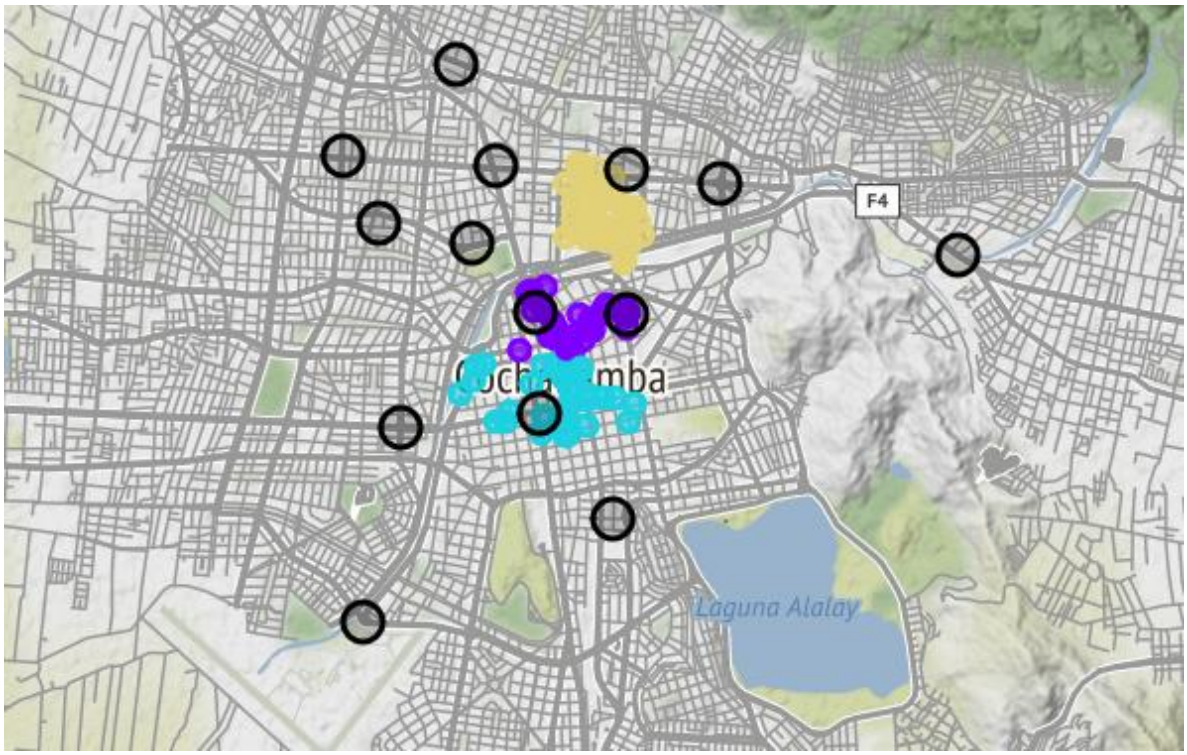
The results of K-means clustering are in the next figure, you can see that k-means was so clever to group the venues as the best possible way, in fact, the only variable used was the distance, but this distance is related to the ATMs, this distance *is not the distance between venues*. Is good to see how k-means works, giving us a results that are not expected.



Now let's aggregate the actual ATMs to this clusters. Evaluating this chart is possible to suggest a better distribution of ATMs near to the centroids of the cluster but this has to be evaluated by the stakeholders because it is possible that the current provision obeys other rules that only the bank knows.



Finally, another point of evaluation is the number of elements in each cluster, the three largest clusters contains 51, 38 and 34 venues, for this reason it is possible that the ATM requires greater maintenance tasks, and for this reason is recommended relocate some ATMs or add a new one.



5. Conclusion

In this study we look for best locations of actual or new ATMs, using the Foursquare data we compare the actual disposition of ATMs and the distributions of venues frequented by the people in Cochabamba city.

Using an ML algorithm, that used the geographical distances between the venues, it was possible to create groups or zones that could be the best place to deploy new ATMs.

It was recommended to move or deploy new ATMs considering the clusters presented by k-means, some clusters are very long so the recommendation is valid.

This study could be applied to other financial institutions with ATMs (or branches) that need to be relocated or that need to deploy new ones.