

Capstone Project – Final Report - Applied Data Science Capstone

Selecting a place to build a new shopping mall in Quito, Ecuador

Introduction

The construction of a new shopping mall can be a great opportunity for potential clients to hang out with friends, have something to eat, buy some fancy jeans, and so on. Shopping malls provide us with the opportunity to do a variety of tasks at the same time in a single place (wife can go to buy a perfume while husband goes grocery shopping). Malls can be very helpful that way because all the family's potential needs can be met in one sole location.

Business problem: If a property developer is looking to build/open a new shopping mall in Quito, Ecuador, what parish/neighborhood should he/she choose?

Target audience: This capstone project could be useful to property developers and investors in general looking to open a new shopping mall in Ecuador's capital, Quito.

Data required

- List of parishes that belong to Quito, Ecuador (the reason I did not choose neighborhoods is that in Ecuador, neighborhoods tend to be smaller areas than in other country's cities). This data will be obtained/scraped from Wikipedia.
- Geographical coordinates (latitude and longitude) of those parishes, which is needed to plot the parishes on a map. This data will be obtained with the help of python's geocoder package.
- Venues data, in this case about the existing shopping malls in Quito, in order to construct our clusters with this and the parishes data. The source of this data will be Foursquare's API.

Methodology

First, we need the list of parishes that belong to Quito, Ecuador. The list is available in Wikipedia at the URL https://en.wikipedia.org/wiki/Category:Parishes_of_Quito_Canton. We'll web scrap this list from the internet by using the BeautifulSoup package. Then we need to get the geographical coordinates (latitude and longitude), to do so, we'll use the Geocoder package. After gathering the data, we will construct a data frame with it using the Pandas library, and then visualize the parishes' location in a map using the Folium package. This allows us to validate that the geographical coordinates data returned by Geocoder are correctly plotted into Quito's map.

Next up, we'll use the Foursquare API to get the top 100 venues within a radius of 2000 meters of each parish (registration to a Foursquare Developer Account and creation of an app within the platform is required). Then we make an API call to Foursquare passing in the geographical coordinates of the parishes. Foursquare will return a JSON file with the venue data. Then, from this file we'll filter only the venues with the type "shopping mall", extract the venue name, venue category, and coordinates. Then, we'll skim each parish by grouping them and observing the frequency of occurrence of each venue category per parish. By doing so, we are also preparing the data for the clustering process.

Finally, we'll perform clustering on the data by using K-means clustering. We'll cluster the parishes into 3 clusters/groups based on their frequency of occurrence for "Shopping Mall". The results will hopefully allow us to identify a list of parishes with the higher concentration of shopping malls (saturated parishes), a list with a moderate concentration of shopping malls (being this list the answer to our business problem, since these parishes are the ones that have the highest commercial

potential), and a list with very low/inexistent concentration of shopping malls (not enough commercial potential to work with).

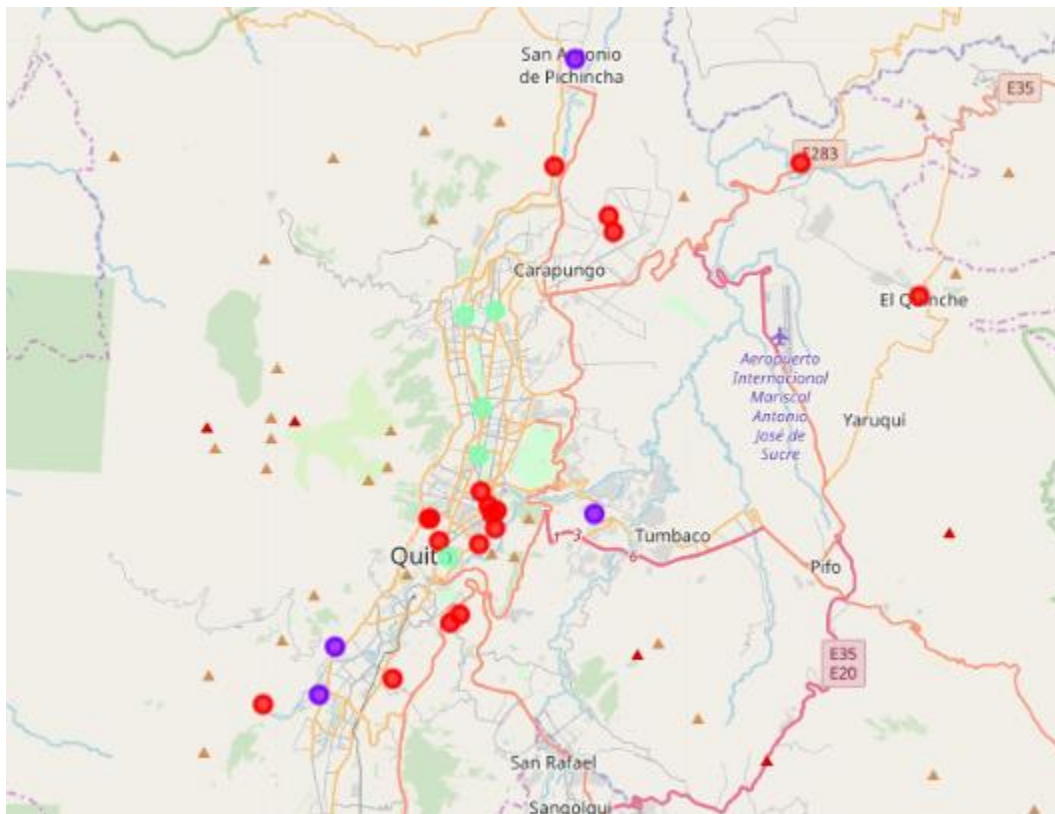
Risk identified: Since Foursquare is not that popular in Ecuador, the amount of customer reviews/check-ins stored on its databases is surely lower than in other countries, which makes this study very prone to outlying data in two forms: parishes with too little information, and parishes with “too much information” (parishes where “Foursquare enthusiasts” have checked-in at every single store he/she visits within the parish).

Study results

After running the process explained previously on a Jupyter Notebook, K-means clustering show that we can categorize Quito's parishes into 3 clusters, based on the frequency of occurrence for the venue type “shopping mall”:

- Cluster 0: Parishes with a low number of shopping malls (red dots).
- Cluster 1: Parishes with a high number of shopping malls (purple dots).
- Cluster 2: Parishes with a medium number of shopping malls (green dots).

The clustering results are shown in the map below, where cluster 0 parishes are painted in red, cluster 1 parishes are painted in purple, and parishes in cluster 2 are painted in green:



And the lists of parishes belonging to each cluster are shown below. The “Shopping Mall” columns show this venue’s concentration per parish:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
27	Santa Prisca, Quito	0.01	0	-0.214632	-78.504560
17	La Vicentina	0.00	0	-0.216500	-78.487930
16	La Magdalena, Quito	0.00	0	-0.246774	-78.495721
15	La Libertad, Quito	0.00	0	-0.285230	-78.580380
14	La Floresta, Quito	0.00	0	-0.209530	-78.480970
26	Santa Bárbara, Quito	0.00	0	-0.203210	-78.482360
11	Guápulo	0.01	0	-0.200000	-78.483330
10	Guayllabamba	0.00	0	-0.052990	-78.349640
20	Puengasí	0.00	0	-0.250000	-78.500000
9	González Suárez	0.00	0	-0.202104	-78.480342
7	El Quinche	0.00	0	-0.109960	-78.298580
22	San Blas, Quito	0.01	0	-0.205620	-78.508800
23	San Marcos, Quito	0.00	0	-0.273918	-78.524543
4	Chillogallo	0.00	0	-0.075415	-78.432173
24	San Roque, Quito	0.01	0	-0.205620	-78.508800
2	Calderón, Quito	0.01	0	-0.205600	-78.508800
1	Benalcázar	0.00	0	-0.082714	-78.430165
8	El Salvador, Quito	0.01	0	-0.193713	-78.486892
19	Pomasqui	0.00	0	-0.054460	-78.455420

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
21	San Antonio de Pichincha	0.058824	1	-0.008370	-78.446840
0	Alfaro, Quito	0.045455	1	-0.280763	-78.556043
6	Cumbayá	0.043011	1	-0.203060	-78.438040
18	Nono, Ecuador	0.062500	1	-0.260566	-78.549491

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
12	Itchimbia	0.020000	2	-0.221923	-78.500875
5	Cotacollao, Quito	0.028169	2	-0.118390	-78.494050
3	Chaupicruz	0.030000	2	-0.158160	-78.486320
25	San Sebastián, Quito	0.017857	2	-0.116312	-78.480754
13	Iñaquito	0.030000	2	-0.177577	-78.488000

Discussion

The study shows that according to Foursquare data, most shopping malls are concentrated in clusters 1 and 2. These clusters would be of no use for our purposes, since these parishes are already saturated with shopping malls and competition would be fierce.

As I mentioned before, since Foursquare is not that popular in Ecuador, the amount of customer reviews/check-ins stored on its databases is surely lower than in other cities, which makes this study

very prone to outlying data in two forms: parishes with too little information, and parishes with “too much information” (for instance, parishes where “Foursquare enthusiasts” have checked-in at every single store he/she visits within the parish). Since I live in Quito, I can, for instance, infer that the “San Antonio de Pichincha” and “Nono” parishes on cluster 1 are outliers, surely due to a case of “too much information” registered.

On the other hand, these results allow us to identify Cluster 0 as the list with a low number/concentration of shopping malls. This is the list that encapsulates the answer to our business problem, since these parishes are the ones that have the highest commercial potential. Opening a shopping mall in one of these parishes will represent a great opportunity since there is a mix of commercial potential and not too much to no competition from existing malls.

Conclusion

This brief study recommends property developers to capitalize on these findings to open new shopping malls within parishes identified in cluster 0. These parishes are the ones that have the highest commercial potential since there is not much competition from existing malls yet.

Nevertheless, caution should be taken into consideration when making a decision based on the data provided by the Foursquare API about Quito, since the amount of customer reviews/check-ins in Ecuador cities is surely lower than in other countries where the Foursquare app is better known.