COURSERA CAPSTONE

Capstone Project – The Battle of Neighborhoods - Quito's bests: Neighborhood classification and clustering

Author: Andrés Reyes

1. **Introduction**
   Description of the Problem and Background

   The city of Quito is the biggest city in Ecuador, therefore there is a lot areas to choose from when selecting a place to live in the city. With varying budgets and needs, people have some difficulties to find a suitable neighborhood to accommodate them and their families, especially if they are new in the city.

   The key question is: How can we find a convenient and enjoyable place today in Quito? Certainly, we can use available real estate apps and Google but the idea is to use the tools and methodology learned during the course.

   A potential newcomer aspiring to buy or rent a property would like to become knowledgeable about the current pricing and the proximity to green areas, schools, medical facilities, restaurants to make a conscious decision.

   Target audience
   People who look for a suitable guide to buy or rent properties in Quito.

   Stakeholders
   Buyers
   Sellers
   Government of Quito

2. **Data Section**
   The following sources of data are used for the Capstone Project development:

   - Update list of Quito's neighborhoods with geodata. Source: https://maps.google.com.ec/maps/ms?msid=208848417386158528016.0004b488410d2799413a2&msa=0
   - Goverment of Quito geodata for parks and public transportation. Source: http://gobiernoabierto.quito.gob.ec/
   - Venues for each Quito's neighborhood (for clustering). Source: https://es.foursquare.com/
   - Venues for restaurants and recreational places (from Foursquare)

   The data will be used as follows:

   - Use Foursquare and geopy data to map top 10 venues for all Quito neighborhoods and clustered in groups (as per Course LAB).

- Use foursquare and geopy data to map the location of parks and restaurants, separately and on top of the above clustered map in order to * be able to identify the venues and amenities near each one.
- Use Foursquare and geopy data to map the location of rental places, in some form, linked to the parks and restaurants.
- Data will be searched in open data sources if available, from real estate sites if open to reading, libraries, etc.

This data is going to be processed to cluster the city neighborhoods, to establish each neighborhood top revenues and to answer the following question: How venues distribute among Quito Metropolitan Area neighborhoods and around the parks?

3. Methodology

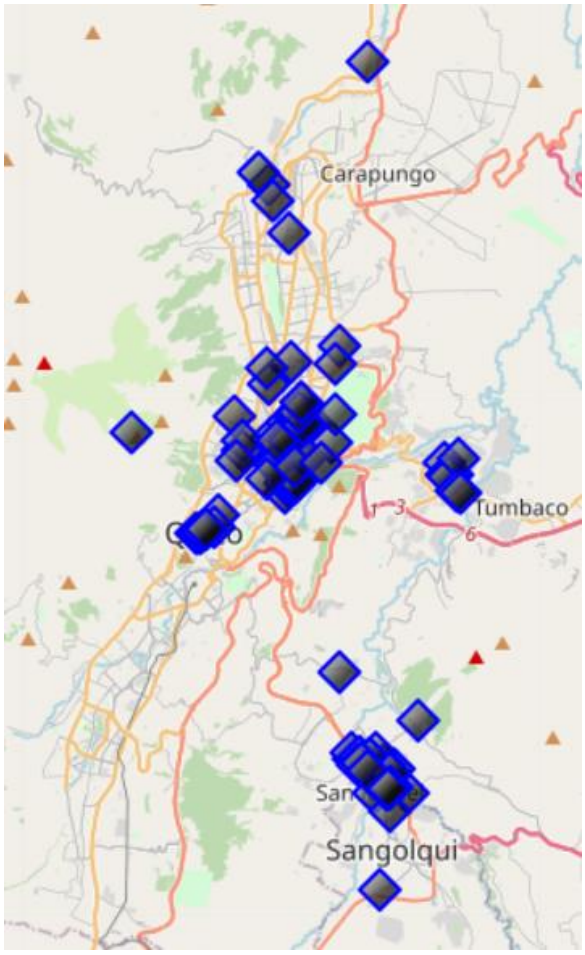
This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a bus station, rental price and similar venues in Quito. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very convenient for the user.

Processing these DATA and its mapping will allow to answer the key questions in order to make a decision:

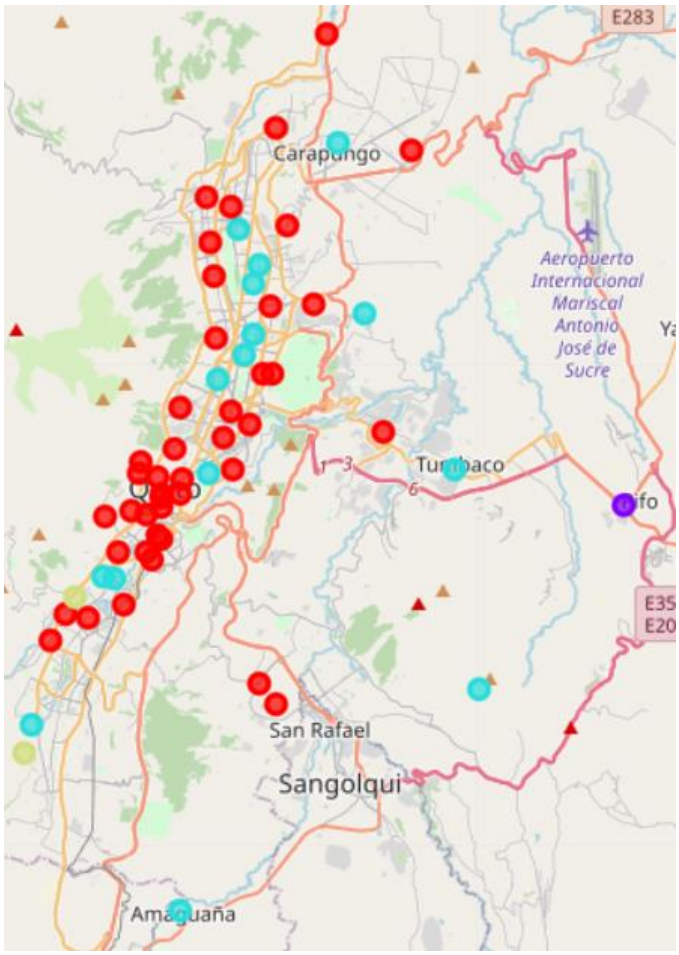
- What is the cost of available rental places that meet the demands?
- What is the cost of rent around a mile radius from each subway metro station?
- What is the area of Quito with the best conditions that meet criteria established?
- What are the venues of the two best places to live?
- How venues distribute among Quito neighborhoods?

Quito's venues and facilities:



4. Results

All the required information is consolidated to make the neighbourhood review and selection in one map. In the image below the neighbourhoods are clustered according to k-means (4) processing.



5. Discussion

Case of Study

In general, with this report a new resident of Quito can check all venues like parks, restaurants, groceries, etc. So, it's possible to have a clear view about the city's facilities and its distribution and make a correct choice of the future neighborhood to live in.

Cluster 0 contains the neighborhoods that have the most facilities in the city.

The neighborhoods with the most facilities are evenly distributed along the north and the south parts of the cities.

About the course

This Capstone project presented a great opportunity to practice and apply the Data Science tools and methodologies learned in the previous courses.

The IBM Data Science Professional Certificate is an outstanding starting point to become a professional Data Scientist and to continue exploring and creating examples of practical cases.

6. Conclusions

The decision of a buyer is influenced by the familial needs, personal biases. So, based upon the findings summarized in the results and discussion sections, following conclusions can be made:

- While making recommendations to a prospective client, it is imperative to know his/ her immediate needs and requirements besides the budget. This would help to catch his/ her attention.
- Knowledge about the most recent market prices can be very helpful for the client and can help him take a decision.