A SHOPPING MALL FOR SANTO DOMINGO ESTE, DOMINICAN REPUBLIC

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

28.JAN.2021

José Luis D'Andrade
Distrito Nacional, Dominican Republic

Business Problem

High density of malls in Distrito Nacional.

Shopping centers in Distrito Nacional have experienced a decline and many are vacant and disused.

Santo Domingo Este is the second metropolitan city.

Santo Domingo Este has a strong economy.

Where in the city of Santo Domingo Este would be the best location to build a new shopping mall?

Description of the data

I. Neighbourhoods in Santo Domingo Este

The scope of this project is constrained to the city of Santo Domingo Este, the second most important municipality of the province of Santo Domingo.

II. Latitude and longitude coordinates of neighbourhoods

Geocoding is the process of transforming a description of a location, such as an address, or a name of a place, to a location on the earth's surface. The resulting locations are output as geographic features with attributes, which can be used for mapping or spatial analysis..

III. Venue data

Data of businesses in the vicinity of the geocoded neighbourhoods.

Data Sources, APIs and Python Libraries

Government. We have obtained information from Oficina Nacional de Estadisticas, https://www.one.gob.do/ , Dominican Republic's head department in charge of statistics.

Foursquare API. **Venue data** of neighbourhoods. Foursquare has one of the largest databases of places and is used by over 125,000 developers.

Python Libraries. We will get geographical coordinates using Python Geocoder package which will give us coordinates of the neighbourhoods. Other libraries to be used:

Pandas: For creating and manipulating vectors and matrices.

Folium: For data visualization, to visualize the neighborhood cluster distribution.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

Methodology

CRISP-DM methodology for mining data

Google API and Foursquare API to get venue data

Machine Learning techniques like one hot encoding

k-means clustering

Visualization of clusters in a map

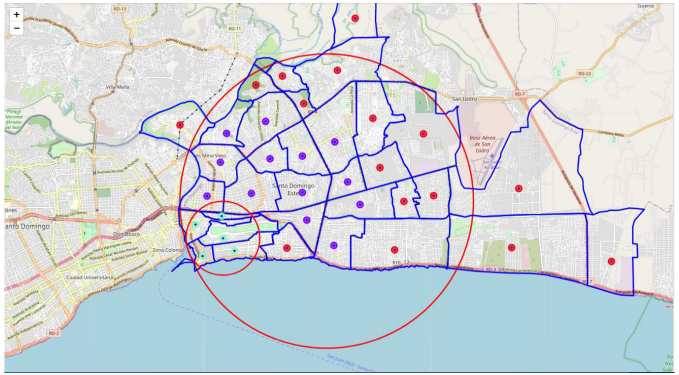
Results

34 neighbourhoods

3 km radius search

1841 venues found

119 unique venue categories



21 neighbourhoods account for 25 shopping locations

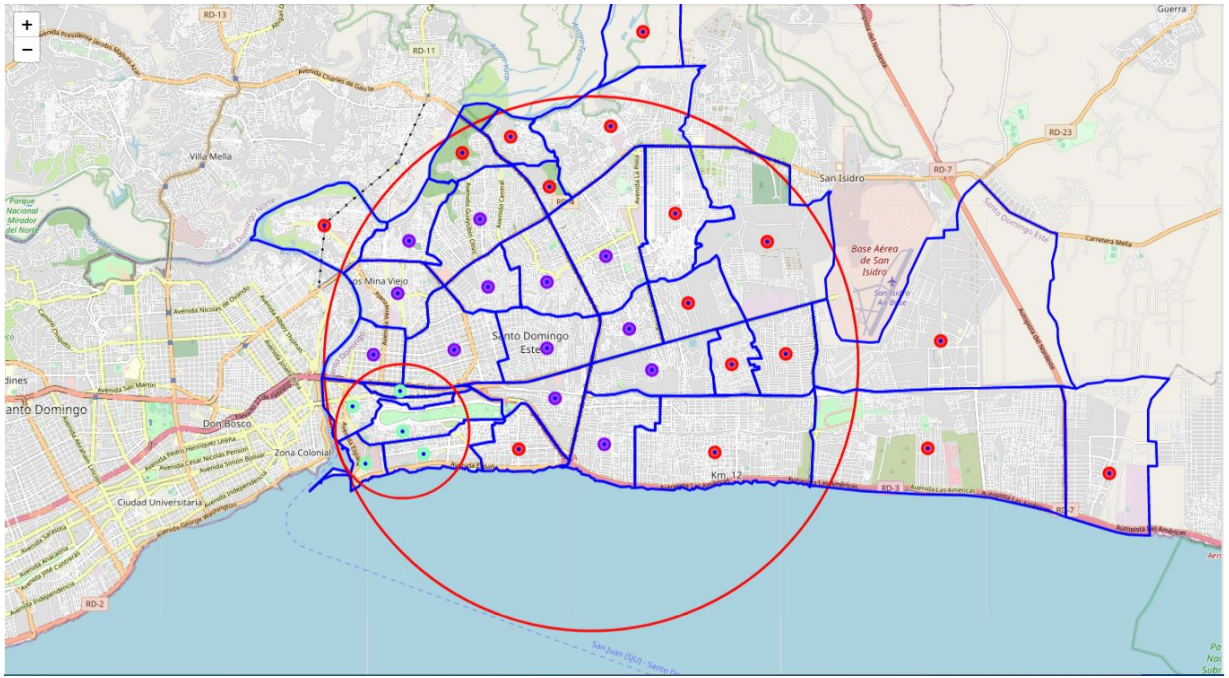
3 clusters in total

5 neighbourhoods have no shopping locations

Discussion

A good number of shopping locations are in the central area of Santo Domingo Este, with the highest number in cluster 1 and almost the same moderate number in cluster 0.

It seems there is opportunity and high potential areas to open new shopping malls as there is no competition from existing malls in Cluster 2 (smaller circle).



Recommendation

This project recommends property developers to open new shopping malls in neighbourhoods in cluster 2 where there is no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 0 with moderate competition.

Shopping malls in cluster 0 and 1 are likely suffering from intense competition due to high concentration of shopping locations.

Recommendation

One other observation is the surrounding areas in cluster 2. A Museum, Aquatic Park, Aquarium and Racetrack are among the attractions in the area that are a walking distance away or a few minutes driving. **And, better yet, there is unbuilt land.**

Conclusion

Our findings can help stakeholders to capitalize on opportunities on high potential locations

Neighbourhoods in cluster 2 are the most suitable locations to open a new shopping mall