

IBM Applied Data Science Capstone

Opening a New Shopping Mall in BOGOTA – COLOMBIA

By: Miguel Santana H.

June 2020



Introduction

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Bogota – Colombia and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Bogota - Colombia to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Bogota - Colombia, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Bogota - Colombia. This project is timely as the city is currently suffering from oversupply of shopping malls.

To solve the problem, we will need the following data:

- List of Districts in Bogota. This defines the scope of this project which is confined to the city of Bogota, the capital city of the country of Colombia in South America.

- Latitude and longitude coordinates of those Districts. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://es.wikipedia.org/wiki/Anexo:Barrios_de_Bogot%C3%A1) list of Districts (“Localidades”) in Bogota, with a total of 19 Districts. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the Districts using Python Geocoder package which will give us the latitude and longitude coordinates of the Districts.

After that, we will use Foursquare API to get the venue data for those Districts. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization.