IBM APPLIED DATA SCIENCE PROJECT

By Oluwatoyin Odeniyi

OPENING A NEW SHOPPING MALL IN ALGIERS, ALGERIA

Introduction:

This project seeks to identify areas where establishing a new shopping mall in the capital of Algeria which is Algiers,would work best for the investors and business owners and the government, by extension also the citizens of the city.

Business/Problem:

The demand for the retail industry to establish itself firmly in Algiers is growing daily as the Algeria's economic development as impacted positively on the income of the populace. Also, due to strong fiscal position of the country, the purchasing power of citizens has been on the increase. However, the number of malls in the state capital has not been enough to cater to the growing needs of the populace as regards the retail market.

The challenge or problem for the retail industry in Algiers is to meet up with the domestic demand for retail goods which can be sold in and thru shopping malls, in areas where they are sure to make sales and this will in turn ensure profitability for the investors or business owners and overall increased revenue for the government.

This project seeks to explore data insights which will enable potential investors and business owners and the government to identify the best areas to build shopping malls (or in the case of the government, to encourage both local and foreign investors on the best areas to build shopping malls). It is also a source of information for the government as to which areas in neighbourhoods in Algiers would interest new investors and business owners when it comes to the retail industry. This would overall impact positively on the revenue which can be generated thru tax and import duties.

Audience and Stakeholders:

The audience and stakeholders for this project are the local and international investors looking for profitable opportunities to diversify their investments. Also shop owners within and outside the country looking for ways to grow and increase their customer base. Stakeholders in the government who are open to new ways of generating revenue for the city(in the form of sale of land, tax and other avenues).

Data source, data collection and data cleaning:

This project sources and integrates data from the Wikipedia: https://en.wikipedia.org/wiki/Category:Suburbs_of_Algiers. Another data source is Foursquare API. This section describes each of these data sources and provides examples of the data.

The Wikipedia page provided suburbs in Algiers as well as populated areas in the city. The data included provinces in an alphabetical order and grouped as such.

As the data was obtained from a Wikipedia page, it was scraped and the Beautiful soup package was used to parse the data. Then the data was appended into a list and a pandas dataframe created.

Foursquare API,a social networking service which provides a mobile app that allows each user to search for venues close by and see information and reviews from other users was used to get location data. Users feed information to Foursquare passively as the app tracks users' locations. Users also do so actively whenever they enter venue names, locations, and reviews. This project will access Foursquare venue data for the selected Suburbs in Algiers to obtain neighbourhoods, specifically popular venues. The Foursquare venue data will particularly seek to identify popular areas that have unique categories. These data will then be used for subsequent categorization of neighborhoods to provide insight to places where shopping malls are present or totally absent.

Methodology:

Foursquare API used to get the venue data for the neighbourhoods:

The Foursquare application programming interface (API) was accessed(with my client id and client secret) to obtain the venues in the Algiers neighbourhoods . The interface was used specifically to:

Get the top 100 venues with a radius of 2500 metres(this large measurement was used as Algiers has a wide expanse of land, especially desert land).

The venues were converted into a dataframe including venue latitude, longitude and category.

Unique venues were curated and the results was checked to see if any has 'shopping mall' in their names.

Each neighbourhood was analysed and a new dataframe was created to show shopping malls only.

The methodology also included exploring the neighbourhoods and segmenting them into clusters. To do this, a machine learning algorithm, K-means clustering was used. Clustering (or cluster analysis) is the task of breaking up a set of objects into groups called clusters. Inside each group, there were similar objects(neighborhoods), and those which were dissimilar, were grouped differently. The k-means algorithm is the simplest of machine learning algorithms. In this project, it was used to show neighbourhoods which had shopping malls and their numbers:

1.k-means clustering was run on a set number of 5 clusters and merged into a new dataframe.

2.The new dataframe was grouped with the locations of the neighbourhoods

Results:

This results section provides an overview of the outcomes of the methodology and their relevance to the original problem of identifying the best Neighborhoods to build a new shopping mall in the city of Algiers.

Majority of the shopping malls in Algiers is concentrated in the Neighborhood of Mohammedia, Algiers as shown in cluster 1 with Dar El Beida Neighborhood following it behind,having the next highest number of malls, in cluster 3. These two Neighborhoods would not be best for opening a new shopping mall as there is a huge likelihood it would not thrive and grow due to competition. Also more of the initial capital would have to be expended on promos and paid adverts to drive in customers. This would not help a new mall's overall profitability in such Neighborhoods. In cluster 0, we can see the large number of Neighborhoods which have no shopping mall whatsoever. Shopping malls can be opened in Draria,Saoula,El Bria, Baba Hassen and other Neighborhoods in the cluster. Such areas are best for developers or investors to open a shopping mall as there would be no competition and so turnover on investment is feasible. Neighborhoods in cluster 2 and cluster 5, might not be best for opening a shopping mall as they already have shopping malls present, though not as many as clusters 1 and 3. Overall, this data analysis of the city of Algiers in Algeria as regards the best areas to open a new shopping mall, shows that Neighborhoods in cluster 0 are ideal for opening a new shopping mall and it is highly recommended especially as it has no shopping malls therein.

Consideration was made for identifying top venues in the Neighborhoods of Algiers where there were venues well-visited like Beaches, restaurants, Electronics store, Theatre and so on. These places would make any of the neighbourhoods identified, ideal for a shopping mall.

Discussion:

The result above shows the importance of quality in depth data analysis before setting up a business in an area. Algiers is a well-populated city, however shopping malls are scarce. According to the resulting clusters, shopping malls are even scarce in some areas where there are top venues like gyms, restaurants, beach, piers and so on. This is an unexpected development as it shows the city has not been explored by mall chains which have branches spread around the world. Algiers has relatively been untapped by the retail industry considering the increased income of the populace, which is reflected by the top 100 venues seen in the neighbourhoods. For those businesses to exist as top venues, then there is every reason for investors and business people to explore the neighbourhoods in cluster 0 and in fact all the other clusters(with consideration for possible competition) for opening new malls. The coordinates (obtained by using the Foursquare API)also show that distance to the top venues is close enough to each other and can encourage heavy traffic from those places to a new shopping mall.