

Capstone Project

BATTLE OF NEIGHBORHOODS

Introduction to the

BUSINESS PROBLEM

Dubai is a very dynamic city with many different neighborhoods both in the old/historic part of the city, which existed before discovery of oil and subsequent boost of economy, and the new part of the city full of magnificent masterminds of world renown architects. A lot of international workers or so called expats are moving to Dubai every year, some bringing along their families and loved ones. And this is when the key challenge arises - where to rent/buy an apartment/house?

Each person or a family will have a certain set of criteria for choosing a home, starting from availability of a supermarket in a close proximity, presence of playgrounds/parks and nearby schools/nurseries, to variety of restaurants and cafes around the corner.

The business problem that I have chosen for the project is **identification of the most suitable neighborhood based on criteria provided by the client**, who could be a real estate agent trying to serve the client or the prospective tenants/owners themselves.



DESCRIPTION OF DATA

The main component of data to be used for this project is the data provided by Foursquare API about the neighborhoods in Dubai along with venues of various categories. Neighborhoods will be considered as family friendly if for instance there are nurseries/schools, playgrounds or parks, grocery stores, gyms, recreational facilities located in them.

In addition to the data provided by Foursquare, I also use the geospatial data provided by the Dubai government for the names of registered neighborhoods along with latitude and longitude coordinates.

Description of

METHODOLOGY

The first stage is preparation and cleaning of data in order to ensure that it is ready and valid for the analysis. The data about neighborhoods of Dubai and their corresponding locations had to be imported from a KML file with geospatial information provided by the government of Dubai. I used Beautiful Soup library to parse the file and extract required data. There were overall 226 neighborhoods identified in the city.

Some column types had to be changed in order to be able to map the latitude and longitude on the map.

Next was exploration of venues in the defined neighborhoods using Foursquare API. I had to limit the number of venues per neighborhood to 20, since large numbers were failing due to server failure from the API side. As a result, I obtained a data set of 1119 records, since for some of the neighborhoods there were less than 20 venues listed in Foursquare.

Foursquare provided 200 unique categories of venues in the Dubai neighborhoods. I used One Hot encoding to analyze the frequency of each category within a neighborhood and to identify the most common venues per neighborhood. So for each neighborhood I have identified 10 most common categories of venues.

Next I used k-means clustering technique to segregate the neighborhoods into 8 clusters based on the categories of most common venues that were identified earlier.

By visualizing everything in the map, it is clear that majority of neighborhoods fall under one cluster. 7 other clusters are very narrow and specific to certain types of venues.

RESULTS

As a result of the analysis carried out, I have identified 8 clusters of neighborhoods based on the categories of most common venues in the defined areas. Further review of identified clusters shows that clusters 1 to 7 are very specific and very narrow in terms of variety of venues. **Real estate agents may use these clusters to offer the properties to clients with particular interests:**

Cluster 1 (3 neighborhoods): neighborhoods suitable for lovers of yoga, camping, fishing, food trucks and flowers

Cluster 2 (1 neighborhood): a neighborhood with pet store and yoga studio as most common venue

Cluster 3 (2 neighborhoods): neighborhoods suitable for lovers of pools, yoga, farm life and food courts

Cluster 4 (1 neighborhood): a neighborhood suitable for barbeques and primarily restaurants/food courts as common venues

Cluster 5 (3 neighborhoods): neighborhoods suitable for lovers of mountains, Indian cuisine, flea market and food courts

Cluster 6 (4 neighborhoods): neighborhoods mainly with cafes, yoga studios, fast food restaurants and fishing stores

Cluster 7 (1 neighborhood): a neighborhood with rental car location and yoga studios as most common venues

However, cluster 0 consists of 121 very diverse neighborhoods, which could be used by the real estate agents to filter based on customer preferences. I have used this cluster to further drill down and provide an option to filter the dataset by the provided categories of interest.

For instance, let suppose the client is interested in following categories of venues: 'Playground', 'Indoor Play Area', 'Gym', 'Yoga Studio', 'Pool', 'Park'. By providing these categories to the filter, I have obtained a data set of 78 neighborhoods that have venues of these categories in the 10 most common categories that we have earlier identified. The real estate agent can use this list to identify the properties that can be shown to the clients.

DISCUSSIONS

There is a number of things that should be considered in order to achieve a better result:

1. The number of venues in categories like 'Nursery', 'Child case', ' School' are very minimal in Foursquare for Dubai. Therefore, it would be great to include a database of educational/child care venues into the analysis too. however, at the moment of the analysis such data was not readily available.
2. There are neighborhoods for which no venues were found in Foursquare, which indicates there is a room for improvement in terms of both completeness and accuracy of the results provided by this analysis. This is a matter of time, since Foursquare database is constantly updating.
3. An application can be developed for real estate agencies and prospective tenants/ owners based on this analysis as a back end.

CONCLUSIONS

Overall, analysis based on geospatial data in combination with information provided by such platforms as Foursquare have lots of applications in real life scenarios. This was only one of the scenarios.

For me, as a resident of Dubai, who was searching for an accommodation/house for rent through various means, such analysis would have provided valuable insights and would have helped me in selection of a suitable neighborhood that would meet the demand of every family member with their own interest.