Capstone project - Battle of the Neighborhood - Week 2

Applied Data Science Capstone by IBM/Coursera

Original work developed by Bilal AL-Soudi 7-2-2020

Introduction: Business Problem

Whether you are single or relocating to the UAE (United Arab Emirates) with your family, there are certain areas that continue to attract expats in their droves.

Capital city Abu Dhabi and its neighbor Dubai are the hotspots of the UAE for expat activity. Skyscrapers, a thriving hospitality industry, and the inner-city beaches are attractive places to reside.

Top of the list for many families moving to a new location is **quality schooling**, and the UAE provides high-quality education all around. Many nations have schools for their expats, and there are plenty of international schools to choose from.

Closeness to facilities such as recreational areas is also important. There are many areas to choose from in both cities, depending on what your family needs are. Not all residential areas are in high-rise buildings, but this might best suit you and your family.

We'll be focusing on **Abu Dhabi as its the capital of the UAE** and the second-largest city after Dubai. If you choose to live here, there are ample options and we'll demonstrate how an **XPATs can pick the neighborhood** of his/her choice based on **private school**, s **rating** and **curriculum** and of course **near by venues**.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed towards best possible selection of best locations.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing venues nearby schools (any type of venues)
- number of private schools in general
- number of private schools based on ratings i.e. Outstanding, Very Good, Good , Acceptable, Weak, Very Weak
- number of private schools based on curriculums i.e. American , English , Indian, Japanese, German, Canadian..etc

Following data sources will be needed to extract/generate the required information:

- Will be using **GEOPY python library** to generate Latitude and Longitude for the Abu Dhabi Island
- Will be using the **GEOCODER python library** to geocode coordinates (Latitude and Longitude for schools in Abu Dhabi.
- Downloaded csv file for all Abu Dhabi localities (Neighborhoods) along with their coordinates (Latitudes and Longitudes) from https://www.geopostcodes.com/Abu_Dhabi
- Used a free handy tool**(WEBHARVY)** https://www.webharvy.com/ to scrape multiple pages to get school's related ratings and curriculums to a csv file.
- number of restaurants and their type and location in every neighborhood will be obtained using **FOURSQUARE API**

- Write near by venues and venue's details returned via FOURSQUARE APIs to unique separate csv files. The files will read later to data \ frames for continued testing. This is necessary since FOURSQUARE has limitation in making daily calls to their API's.

Methodology

In this project we will direct our efforts on detecting areas of Abu Dhabi with Private schools that have high ratings i.e. Outstanding and Very Good which will lead to the neighborhood to be considered living in since schools fees reflects its overall ratings i.e. school with Outstanding and Very Good ratings are relatively more expensive than Schools with Ratings of Good and/or Acceptable.

We will cluster the Schools based on its curriculum i.e. American, English, Indian, German, Canadian, ..etc. This is important since it will reflect demographic of people living in the hosted neighborhood. Here we will use the One Hot Encoding technique and using Kmeans clustering technique and later show clusters on a map.

Also will look into the venues near by the schools and will show how to manipulate the data for the purpose of plotting meaningful bar charts to be used as tools for the purpose of building a case in making a decision to consider a neighborhood to live in.

We will also calculate total venue's average rating. The cumulative average will provide more decision power for Xpats or parents towards considering the locality (Neighborhood) to live in.

Results and Discussion
Its important to know that the city of Abu Dhabi is really an Island and it is the capitol city of the UAE. Also, its the Capitol city of the emirates of Abu Dhabi. The emirates of Abu Dhabi is one of Seven emirates making up UAE : Dubai, Abu Dhabi, Fujaira, Ejman, Um AL Quain, Ras Al Khaima, and Sharja.

The Island is where the government resides. It’s a world class city that exercise up to date technological trends and as such attracts many professionals (XPATS) from all over the globe. Those are our stakeholders who we are targeting to provide them with minimum and sufficient analysis to assist them in deciding to pick the right neighborhood.

Also its important to remember that the term community/locality resembles neighborhoods. The term Boroughs is not used.

The XPATs(Residents) will target the Private schools as oppose to public schools which is attended by the locals(citizens) and thus our analysis is centered around private schools and nearby venues. This project will also focus on identifying neighborhoods where private schools resides in with rating of "Outstanding" and "Very Good".

We first map out all the private schools in the emirates of Abu Dhabi including the island of Abu Dhabi and our first observation shows that the private schools are centered in the Island of Abu Dhabi and the city of Allain. This indicates that XPAT will surly target the two cities.

Next step will focus on private schools with rating of "Outstanding" and "Very Good". So, we filter the desired schools out and cluster them based on the curriculum offered by these schools. Will adopt the One Hot Encoding technique to convert the categorical values i.e. "American", "English", Indian"..etc into numerical values and then adopt the Kmeans clustering technique to cluster the schools into three clusters.

The top three localities (Neighborhoods) with most schools with rating of "Outstanding" and "Very Good" will be found in AlBateen, Khalifa City and Mohamed bin Zaid. This will conclude that these neighborhoods will mostly be targeted by XAPTS.

Next, we pick "ALbateen" Locality since it has the highest number of schools with ratings of outstanding/very good among the other localities and will fetch the nearby venues. We categorize them and lookup the statistics i.e. how many hospitals or restaurants (Indians, Italians...etc)

Once we get the venues , we then fetch the related venue details i.e. ratings/likes/tips. Here we will calculate the average ratings of all venues surrounding each schools in the Albateen locality.

Finally we show all venues surrounding the Abateen schools.

Conclusion

Purpose of this project was to identify Abu Dhabi's most popular localities based on the private school's ratings and curriculum. ALso we factor in venues and its details surrounding these schools which will be very imperative to be considered in selecting the area/communities. This project is a corner stone for further enhancements to account for further criteria.

We applied One Heat Encoding along usage of Kneans clustring technique to cluster schools based on curriculum. The schools can be as well clustered per ratings for added value.

In this project we spend lots of efforts and time towards compiling data via scraping and geocoding coordinates. Illustration is very important to demonstrate to Stakeholders and thus relied heavilly on making sure the school's localities and its venue's latitudes and Loogitudes are available for the sake of mapping them on maps using the folium library.

Also demonstrated bar charts to effectively show how the communities and Schools compare to each other.