IBM COURSERA DATA SCIENCE PROFESSIONAL CERTIFICATE – CAPSTONE PROJECT

SELECTING TRAVEL DESTINATION FOR THE CLIENT OF A TRAVEL AGENCY

1. INTRODUCTION

In 2019 the global travel and tourism industry was estimated to be worth $ 1.74 trillion[1] with the Europe, Middle East and Africa region accounting for the largest chunk at $ 644 billion. Travel agencies play a critical role in this industry by providing travel and tourism-related services to the business and leisure travellers on behalf of accommodation or travel suppliers. To perform this function well travel agencies must be able to give recommendations to clients who require them on where to travel, places to see taking into consideration the clients' interests and tastes.

In this particular case a travel agency has approached me to use my services as a Data Scientist to help give a recommendation to one of their clients an American family based in South Africa who are returning to South Africa after a vacation back in the United States. The family is a coffee loving family want to have a two-day stopover in any West African city and savour the local coffee scene. However, as they have only 2 days to spend they will need a location where the coffee shops are in close proximity to each other so they can visit as many as possible with ease. The travel agency has narrowed down to three locations, Accra the capital of Ghana, Lagos and Abuja the commercial and administrative capitals respectively of Nigeria. The reason for the selections are Nigeria and Ghana are the biggest economies in West Africa[2] and the three cities boast of international airports, numerous hotels and other amenities.

This type of analysis will be useful for other travel agencies in helping make recommendations to their clients.

2. DATA DESCRIPTION

The data to be used in this project will be location data of the cities and coffee shops in those cities. Data will be sourced from Foursquare.com a popular location data service platform used by a lot of companies like Uber, Apple, Snap Chat, Spotify, Jet Blue, MasterCard etc. To access the required data an account was opened with foursquare.com and the data retrieved using their foursquare API.

3. METHODOLOGY

Using the Foursquare API I accessed the location data for venues in the 3 cities namely Accra in Ghana, Lagos and Abuja in Nigeria. I used the coffee shops category ID of 4bf58dd8d48988d1e0931735 to narrow down the search results to only coffee shops as this was the specific criteria to be used.
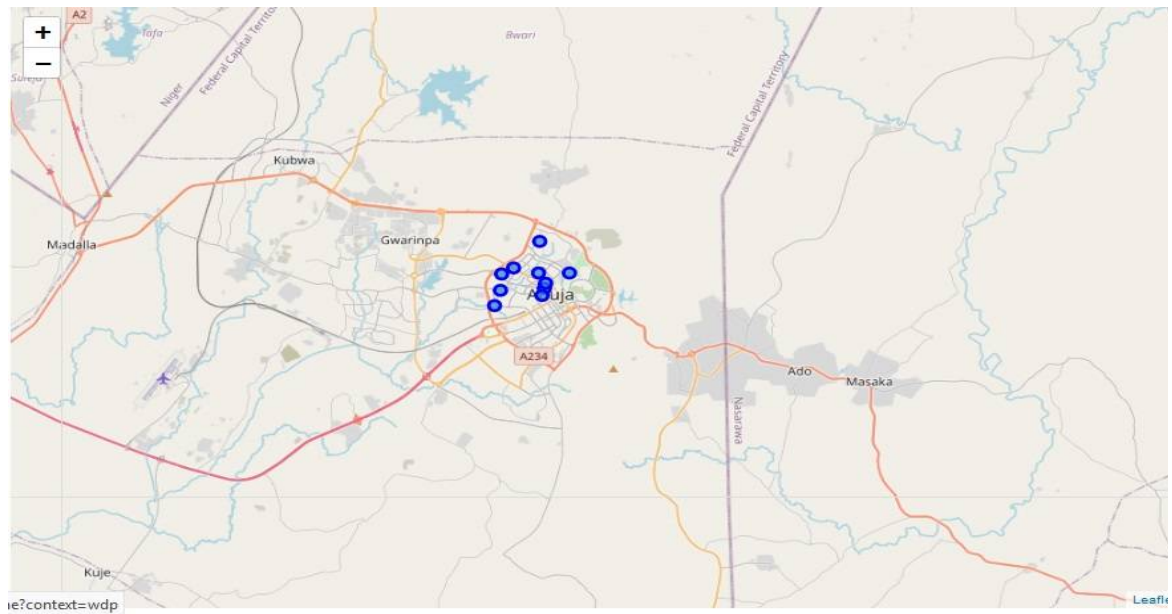
---

[1] www.statista.com

[2] www.countryeconomy.com
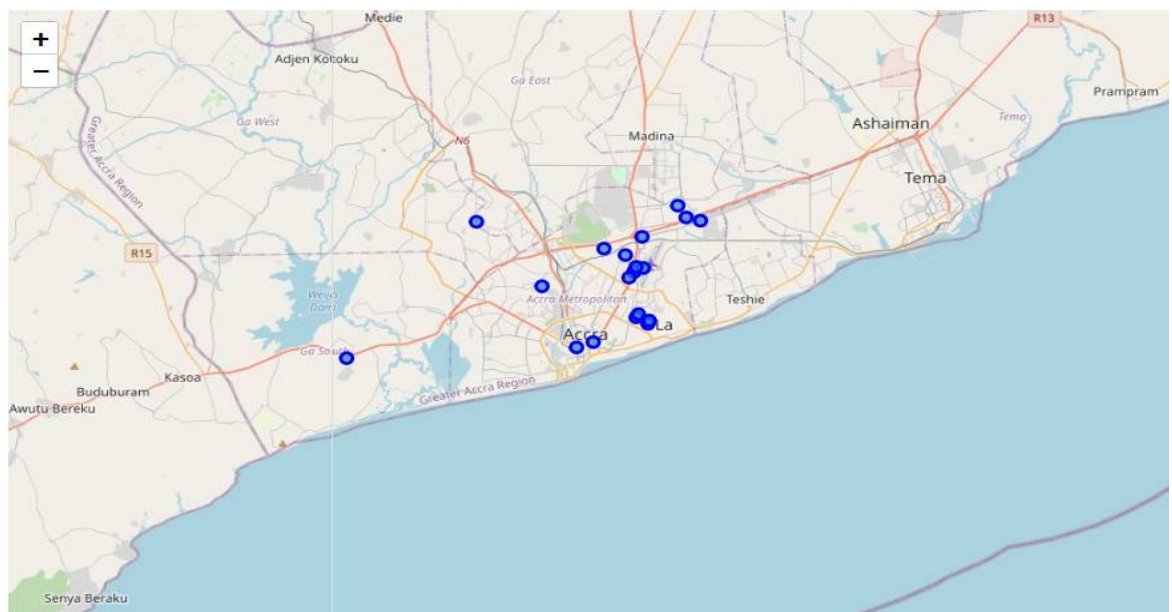
The results were now visualized by plotting on maps. As the aim of the project is to select a location where the coffee shops are in close proximity to each other the next step was to determine the central coffee shop location by calculating the mean coordinate in each location and then estimate how far away the other shops are from the central location by calculating the average of the distance from each coffee shop to the mean in each city.

4. RESULTS

From visually assessing the maps below you can see that yes there are coffee shops in each location but one cannot estimate the density of their locations by visual inspection alone.

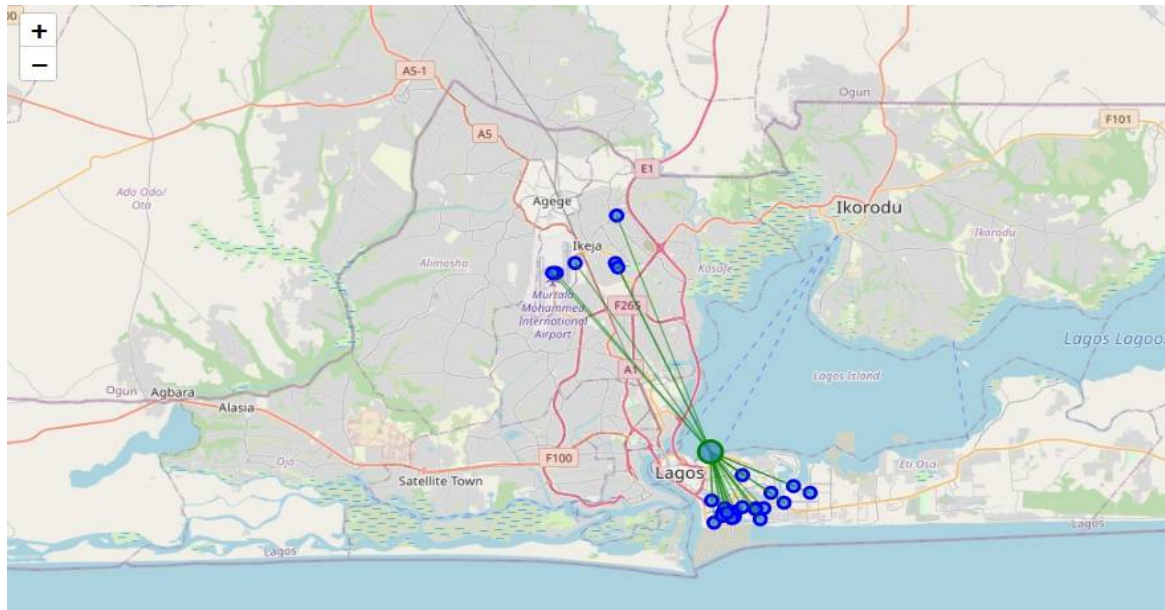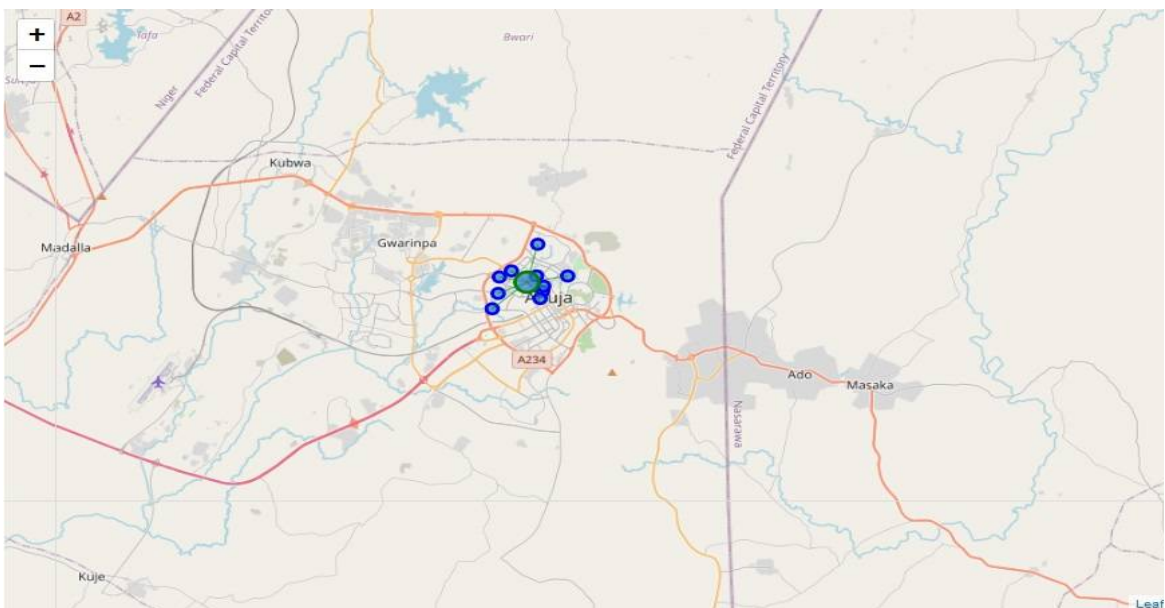**Lagos**

Abuja

Accra

The following maps were plotted after calculating the mean coordinate and average distance to mean coordinate.
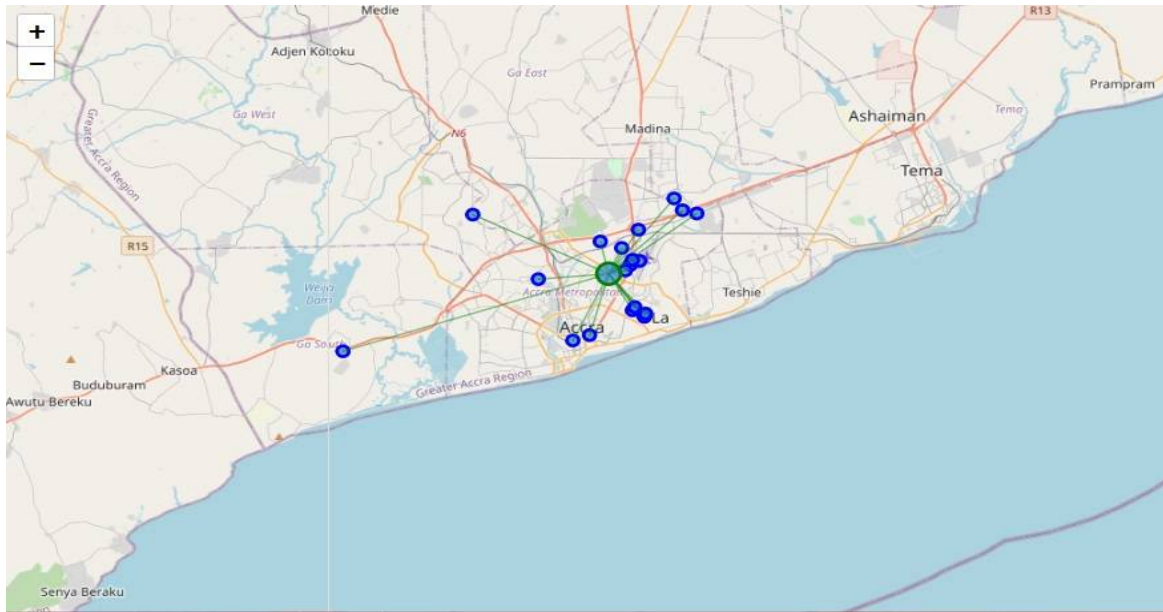
Lagos

Abuja

Accra

Below are the results from the calculation of mean distance to mean coordinates from least to highest.

City	Mean Distance from Mean coordinates
Abuja	0.01616734229575901
Accra	0.04085511761575744
Lagos	0.06385700846991325

5. DISCUSSION AND CONCLUSION

It is worthy to note that while Foursquare has a limit of 100 results per query none of the 3 cities were able to return a 100 results. This points to the fact that the coffee shop market is still a burgeoning one in the region.

However coffee shops are available and based on the analysis I will recommend that the travel agency select Abuja to their clients as Abuja of all the 3 have coffee shops in closest proximity to each other based on the mean distance from mean coordinates for all 3 cities.