**Capstone Final Project**

**Yazid M.**

**Introduction/Business Problem**

My Final Capstone project is to help people who are planning to open a new African restaurant in London. The idea of this capstone project is to help those people to know the best location in London to have a good revenue at the end. The problem that they face is the big area of London which has 8 different areas as following: South East, South West, West, North, East Central, East, West Central, and North West. At the end of this project, I will determine for them the best area (which one of the 8 areas of London) to be used in London, then I will explore the neighborhoods of that best area in order to have a view of the venue categories and determine the group or groups of neighborhoods to open the restaurant.

**Data Will be Used**

To answer those questions, I will get the postcodes of London from wikipedia in order to create the dataframe of London. This dataframe will include different kind of data for example the PostCode, the Name of the Neighborhood, the Area, and the Area Category to determine the 5 areas of London and the process of cleaning the data will be used to clean all venues not included in the PostCode column. Then, I will be combining the data of current restaurants in London areas with the data will be used from the Foursquare location data to extract the venues information for the areas of interest and grouping the neighborhood in 5 clusters based on different numbers of facilities that are there in that area. At the end of this project, I will be going to use folium library to create the necessary maps in order to give a visual view to those people of the best location to be used.

**Methodology**

In this project, after gathering the required data to built the appropriate dataframe from different data sources as explained above, I will proceed in data analysis to find the best area and neighborhood to open the restaurant. From the Foursquare data I will take into consideration the top common venues in that area, and this can be implemented through frequency statistics and aggregations per area and derive the distributions per area and per venue category. At that point we had to choose the best area of London and we selected the area with the minimum number of restaurants and this will be implemented in two stages the first stage is to selected the top 20 restaurants in the area and the next stage is to select the top 10 restaurant then I will proceed in K-means clustering method in order to group our neighborhoods into 5 clusters based on the venues nearby. Geospatial map view will be used to represent the results. The last step is exploring the derived clusters to choose the cluster that there is a probable opportunity to open restaurant after determining the optimal neighborhood based on the least percentage of restaurants and the distance between the new restaurant to other restaurants in the area and other entertainment venues nearby if any.

**Discussion**

To conclude the observation while working on this project in order to determine the best area to open the restaurant in London, I can summary the observation as following:

* There were lots of restaurants and food industries in London.
* There were lots of special non geolocation postcodes that correspond to food and other facilities in the data.
* Many factories used in this project to determine the best location/area to open the restaurant such as the number of food restaurants nearby, distance from central London and if there are other facility venues nearby.

**Results**

The following are the results from analyzing the above 5 clusters :

* Pubs, Cafe, Coffee Shops are popular shops in the South East London.
* The Italian Restaurants are very popular in the South East London area (Especially in Southwark and Lambeth areas).
* Lewisham area is the most condensed area of Africans in the South East Area.

**Conclusion**

From the results we can conclude that Clusters 2 and 3 are the most viable clusters to create an African Restaurant. Their proximity to other amenities and accessibility to station are paramount. The proximity to resources needed is paramount as Lewisham and Lambeth are not far out from Peckham (under Southwark). More factors can affect the results if included in this project, for example, more facilities near the area we will open the African restaurant, the traffic access, hotels, and also the feedback of people on other African restaurants if any.