**Introduction**

My hypothetical client is interested in opening a North Sudanese restaurant in Toronto, Canada catering to all foodies who would enjoy it. North Sudanese cuisine is unique in that it shares many dishes with traditional Middle-Eastern food while retaining bold African recipes and presentations most similar to Ethiopian food. It also shares flavour profiles with some Central African dishes, but Ethiopian food is the most similar. An ideal location for such a restaurant would be in an area where several Ethiopian restaurants already exist alongside a few Middle-Eastern restaurants. This ensures ease of discovery by customers who frequent these eateries. Purpose of this project is to help the client find the most suitable area options for this restaurant.

**Data**

1. Foursquare API: pulls every type of venue within chosen proximity of specified locations.
2. Wikipedia Table: Toronto’s Postal Codes with associated Boroughs and Neighborhoods.
3. CSV File: provided by instructors with geographical coordinates for each Postal Code.

**Methodology**

**Import of the Wikipedia table** as a pandas dataframe (df). Data was cleaned and structured. Only cells with assigned Boroughs were processed, cells with *Not Assigned* Boroughs were dropped. Cells with assigned Boroughs but *Not Assigned* Neighbourhoods were given the Borough’s name to replace the missing Neighbourhood name. More than one neighbourhood were allowed to exist within a Postal Code area to help with clustering. Final shape of df was 103 rows by 3 columns.

**CSV File** was imported from a URL as a pandas dataframe. Shape of df rows match on the Postal Codes “rows index” with the first df. Both dfs were merged on Postal Codes. Now df contains all neighbourhoods sharing each Postal Code and Borough, along with the geographical coordinates for each of the Postal Codes within which the neighbourhood groups exist. A folium map was then created to display the neighbourhood groups.

**Foursquare API** was then accessed requesting venue details for each Neighbourhood. A function was defined to return the category of each venue and its coordinate information. This was done for all neighbourhoods in Toronto. Onehot encoding was then performed in order to find the frequency of each categorical venue per neighbourhood. This made it possible to define a function which retreives the Top 10 most common types of venues in each neighbourhood.

**K-Means Clustering** is the most appropriate form of unsupervised machine learning tool for this project. Areas were grouped into clusters with shared centroids. This allows neighbourhoods in different boroughs to be grouped together into the most efficient clusters. A map of the clusters was then displayed to verify results.

**Examining the clusters** it was found that four unique clusters share the category of Ethiopian food amongst their top 10 most common venues.

**Results**

* Ethiopian Food is most common in Clusters 0, 1, 2, and 4.
* Central Toronto dominated the four clusters with only one occurence for both the East and Downtown Toronto Boroughs (in Cluster 0).
* Ethiopian Restaurants AND Falafel Restaurants were amongst Top 10 venues in:
  + Cluster 1, Central Toronto, Roselawn Neighbourhood.
  + Cluster 2, Central Toronto, Lawrence Park Neighbourhood.
  + Cluster 4, Central Toronto, Moore Park and Summerhill East Neighbourhoods.

**Discussion**

Based on the results, both the Roselawn and Lawrence Park neighbourhoods seem to be the best choices for the North Sudanese restaurant.

**Conclusion**

Business problem was identified. Data was extracted, prepared, and processed. Machine learning was utilized using the K-Means Clustering algorithm to provide the best area recommendations to the client.