

# **IBM Capstone Project — The Battle of Neighborhoods**

---

**The best location for an African restaurant in Calgary, AB  
Canada**



# Introduction

- **1.1 Description of the Problem**

- The population of Calgary has grown considerably over the last decades. Calgary is very diverse. The city is expanding, and new communities are being created and built all the time. As communities increase in Calgary, the number of restaurants is increasing too but not at the same rate as the new communities.
- Although there are many fine restaurants in Calgary — Asian, Middle Eastern, Latin and American restaurants but it's very hard to find good place to dine in the finest of African cuisine that has combination of Nigerian, Ghanaian, Cameroonian, Senegalese and more.

- **1.2 Discussion of the Background**

- Due to the increase of Africans immigrating to Calgary, opening an African restaurant in Calgary right now will be very lucrative. Calgary demography is so big, the available data for demography, communities and ethnicity will be used in determining the best community a new African restaurant should be planted

- **1.3 Target Audience**

- Calgary has a multicultural sense because of its diversity. As such, there is a shortage in the high-end African-inclined restaurant. The target audience is broad, it ranges from Africans, Caribbean, Calgarians, tourists, and those who are passionate about African food.

## 2. Data acquisition and cleaning

- **2.1 Data sources**

- The first data I checked is a Wikipedia page that described the demographics of Calgary data that shows the population of different ethnicities' and the progression over the years - [Ethnicity of Calgary](#). The second data I used to get Calgary communities and its coordinates with the postal codes is a Wikipedia page that list all of Canada's postal codes and its coordinates - [Postal Codes of Canada](#)

- **2.2 Data cleaning**

- A table with Postal code, Borough, Neighborhood, Latitude and Longitude data from the [Postal Codes of Canada](#) page was scraped and downloaded to the dataframe. The Borough data is then filtered to only show Calgary. Because some of the data in Calgary is showing NaN(null values), I dropped the rows that has NaN in any of its column.
- The Foursquare API will be used to obtain number of restaurants and other point of interest and their type and location in every community in Calgary.



# 3. Methodology

- **Data Exploration**
- **Calgary Data:** From the [Postal Codes of Canada](#) Wikipedia data that comprises of all the postal codes in Canada, I filtered the data by Calgary so that only Calgary data will be used in my project. Also the datatype for latitude and longitude of the communities were changed from object to float, all the rows with null values (NaN) were removed from the data, this now becomes the final Calgary data to be used to the project.

```
In: df_Calgary_final['Latitude'] = pd.to_numeric(df_Calgary_final['Latitude'],errors='coerce')
df_Calgary_final['Longitude'] = pd.to_numeric(df_Calgary_final['Longitude'],errors='coerce')
df_Calgary_final = df_Calgary_final.dropna()
df_Calgary_final.head()
```

	Postal Codes	Borough	Communities	Latitude	Longitude
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.964320
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158
10	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600
11	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600
19	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100

# Foursquare API

Then using foursquare API I explored the top 100 venues that are within a 1500 metres radius of Calgary. To do this, the function `getNearbyVenues` is used and it's created to repeat the same process for all communities in Calgary. The number of venues returned for each communities is then explored and the number of the unique categories for each community is then counted.

```
36]: print('There are {} uniques categories.'.format(len(Calgary_area_venues['Venue Category'].unique())))
```

```
There are 57 uniques categories.
```

```
40]: Calgary_area_venues_unique_count = Calgary_area_venues['Venue Category'].value_counts().to_frame(name='Count')
Calgary_area_venues_unique_count
```

```
40]:
```

	Count
Pub	204
Coffee Shop	204
Restaurant	170
Steakhouse	170
Italian Restaurant	136
Hotel	136
Café	102
Park	102
Performing Arts Venue	68
Diner	60



## Foursquare API

- Then I checked and explored the venues in each community

```
# move neighborhood column to the first column
fixed_columns = [Calgary_area_onehot.columns[-1]] + list(Calgary_area_onehot.columns[:-1])
Calgary_area_onehot = Calgary_area_onehot[fixed_columns]
Calgary_area_onehot.head()
```

[illegible]

# Clustering

- For this section, the communities in Calgary will be clustered based on the processed data obtained from the foursquare API.
-



## 4. Result

---

- The following are the highlights of the 5 clusters I did
  1. Pubs, Cafe, Coffee Shops are popular in the South East and South West Calgary.
  2. Restaurants are very popular in all around Calgary especially in South West Calgary.
  3. There is no African restaurants in any community in Calgary.



# Discussion and Conclusion

---

- It is very important to note that Clusters 2 and 3 are the most viable clusters to create a brand African Restaurant. Their proximity to other amenities and accessibility to station are paramount. These 2 clusters do not have top restaurants that could rival their standards if they are created. Although there are no African restaurants in Calgary, North East area is the best place for an African restaurant because it has more concentration of Africans
- In conclusion, this project would have had better results if there were more data in details showing numbers of Africans in each community, crime data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).

-