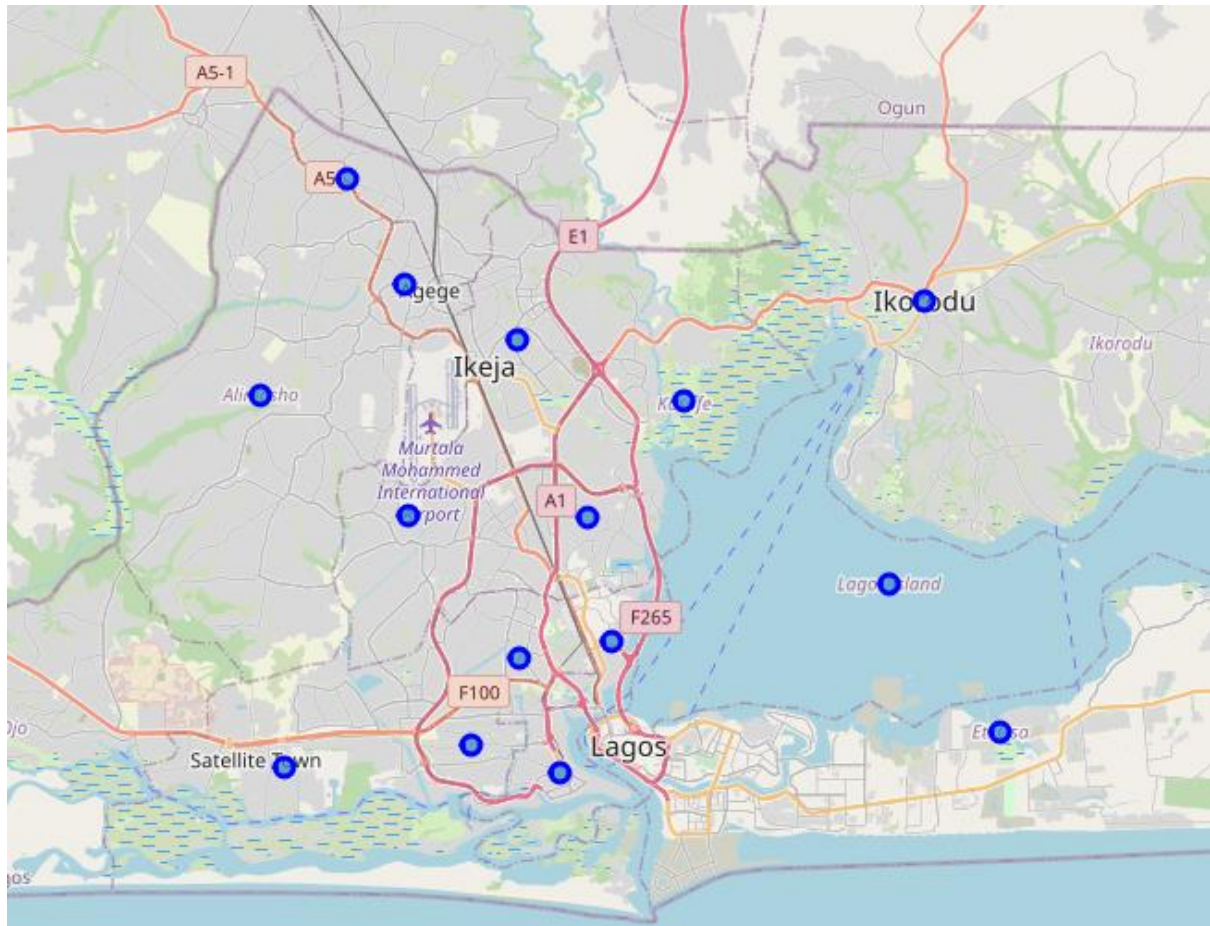


Capstone Project - The Battle of Neighbourhoods (Lagos, Nigeria)

Applied Data Science Capstone by IBM on Coursera



By

Adebayo Amodu

1. Introduction: Problem Background

This project deals with discussing the neighbourhoods of **Lagos city, The commercial capital of Nigeria** (The Most populous black nation) and the commercial hub of West Africa. It is the 4th wealthiest city in Africa, fourth place behind Johannesburg, Cairo and Cape Town.

This project would help people planning to start-up business in the city find the most ideal location.

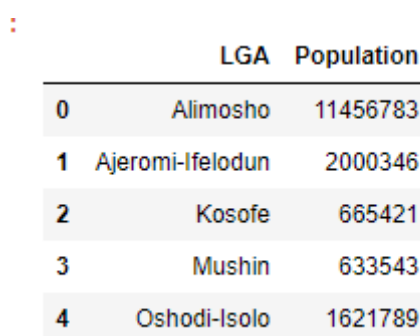
The Foursquare API is used to access the venues in the neighbourhoods. Since, it returns less venues in the neighbourhoods, we would be analysing areas for which countable number of venues are obtained. Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task. Folium visualization library can be used to visualize the clusters superimposed on the map of the city. These clusters can be analysed to help small scale business owners select a suitable location for their need such as Hotels, Restaurants or Pizza shops.

2. Data Requirements

The Dataset is the Wikipedia page of Lagos state [link](#)

We shall explore Lagos city through its respective Local Government Areas (LGA) or Boroughs. The above link is a web page that shows the respective boroughs in Lagos State and each population figure. It's a public Wikipedia data page.

First, we'd save the weblink for The Lagos data, then get the source code HTML data from the Website and use The BeautifulSoup module to parse it. The next step is to get the table that contains the data we want to scrape from. Then finally, we'd iterate through each link of the table and append the text parts.



The image shows a screenshot of a Wikipedia table with the following data:

	LGA	Population
0	Alimosho	11456783
1	Ajeromi-Ifelodun	2000346
2	Kosofe	665421
3	Mushin	633543
4	Oshodi-Isolo	1621789

Fig 1: Lagos State and each population

We then creating the data frame, we create a dictionary and append the LGA and corresponding Population data to it. Then we would attach a column of longitude and latitude by using the geopy.nominatim library to get each LGA latitude and longitude data. See image below.

	LGA	Population	latitude	longitude
0	Alimosho	11456783	6.584343	3.257631
1	Ajeromi-Ifelodun	2000346	6.455122	3.335946
2	Kosofe	665421	6.581974	3.414836
3	Mushin	633543	50.518229	3.291654
4	Oshodi-Isolo	1621789	6.54001	3.312415

Fig 2: Lagos Neighbourhoods Dataset

Foursquare API

A total of 1144 venues data have been obtained from Foursquare. The resultant venues dataset, (shown in Fig) is used for the analysis process.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alimosho	6.584343	3.257631	Ipaja market ipaja	6.602248	3.255585	Market
1	Alimosho	6.584343	3.257631	Wings Restaurant	6.576023	3.325612	Restaurant
2	Alimosho	6.584343	3.257631	Tantalisers	6.649299	3.265609	Burger Joint
3	Alimosho	6.584343	3.257631	De Grange suites & bar	6.602309	3.267038	Bar
4	Alimosho	6.584343	3.257631	National Stadium, Surulere	6.593562	3.304300	Athletics & Sports

Fig 3: Lagos Venues Dataset

The following steps are taken to analysed the data:

- We scrape the web page using the beautiful soup library
- Follow by using Foursquare API calls to retrieve geolocation data and then fetch the text data using the requests library
- We than convert it from JSON to Pandas data frame using the json_normalize module
- We shall use the folium library to render the maps and plot these via The Matplotlib library. Then we shall explore respective boroughs and analyse each area as a location for Start-up based on the aforementioned parameters.
- After this, we shall select our top location.

3. Methodology

Now, with the neighbourhood's data we also have the most popular venues in each neighbourhood obtained using Foursquare API. A total of **1144 venues** have been obtained in the whole city and **135 unique categories**. But as seen we have multiple neighbourhoods

with less than 10 venues returned. In order to create a good analysis let's consider only the neighbourhoods with more than 10 venues.

We can perform one hot encoding on the obtained data set and use it find the 10 most common venue category in each neighbourhood. Then clustering can be performed on the dataset. Here K - Nearest Neighbor clustering technique have been used. To find the optimal number of clusters silhouette score metric technique is used.

The clusters obtained can be analysed to find the major type of venue categories in each cluster. This data can be used to suggest business people, suitable locations based on the category.

4. Analysis

Looking into the dataset we found that there are neighbourhoods with less than 20 venues which can be remove before performing the analysis to obtain better results. The following plot shows only the neighbourhoods from which 20 or more than 10 venues were obtained.

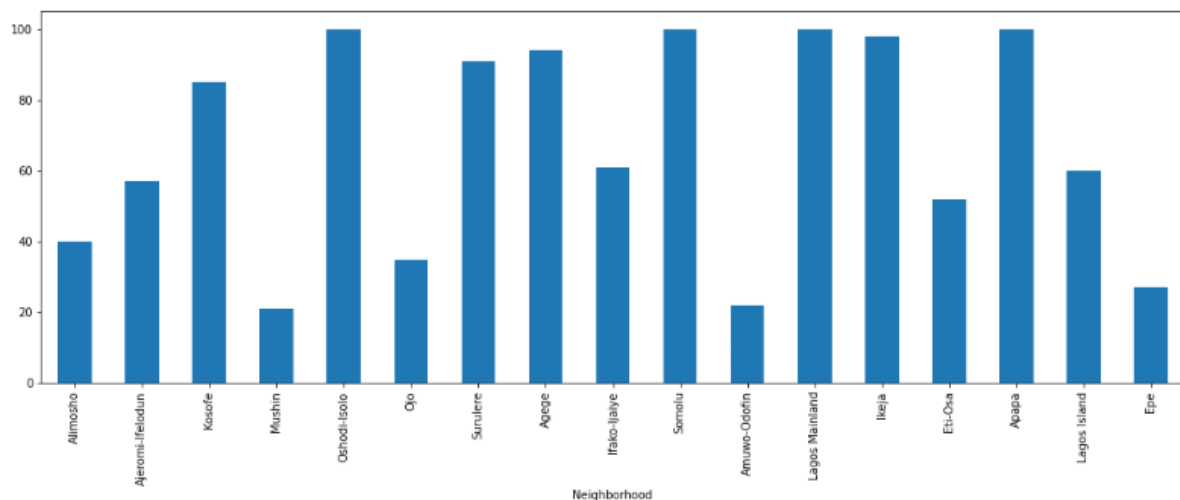


Fig 4: Filtered Neighbourhood Dataset

Next, we will perform one hot encoding on the filtered data to obtain the venue categories in each neighbourhood. Then group the data by neighbourhood and take the mean value of the frequency of occurrence of each category.

(17, 133)

	Neighborhood	African Restaurant	Airport	Airport Food Court	Airport Lounge	American Restaurant	Arcade	Art Gallery	Arts & Entertainment	Asian Restaurant	Athletics & Sports
0	Agege	0.031915	0.021277	0.010638	0.031915	0.010638	0.0	0.000000	0.0	0.000000	0.010638
1	Ajeromi-Ifelodun	0.052632	0.000000	0.000000	0.000000	0.000000	0.0	0.035088	0.0	0.035088	0.000000
2	Alimosho	0.000000	0.025000	0.025000	0.075000	0.025000	0.0	0.000000	0.0	0.000000	0.025000
3	Amuwo-Odofin	0.045455	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000
4	Apapa	0.050000	0.000000	0.000000	0.000000	0.010000	0.0	0.030000	0.0	0.010000	0.010000

Fig 5: Mean of frequency of occurrence of each category

The above dataset is used to obtain the top 10 most common venues in each neighbourhood i.e. the 10 venues with the highest mean of frequency of occurrence. sample for the first 5 neighbourhoods is shown in fig 6

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agege	Fast Food Restaurant	Shopping Mall	Hotel	Convenience Store	Bar	Lounge	Pizza Place	Restaurant	Park	Chinese Restaurant
1	Ajeromi-Ifelodun	Hotel	African Restaurant	Lounge	Pizza Place	Fast Food Restaurant	Chinese Restaurant	Nightclub	Shopping Mall	Movie Theater	Park
2	Alimosho	Fast Food Restaurant	Airport Lounge	Hotel	Restaurant	Pizza Place	Burger Joint	Shopping Mall	Lounge	Boutique	Soup Place
3	Amuwo-Odofin	Hotel	Fast Food Restaurant	Park	Beach	Gym	Food	Flea Market	Mobile Phone Shop	Movie Theater	Diner
4	Apapa	Hotel	Lounge	African Restaurant	Restaurant	Ice Cream Shop	Nightclub	Art Gallery	Pizza Place	Chinese Restaurant	Café

Fig 6: Ten Most Common Venues in each Neighbourhood

The resultant dataset can be used for the clustering algorithm. Here, the K-Nearest Neighbour (KNN) clustering algorithm is used. It is an unsupervised machine learning technique that clusters the given data into K number of clusters. For optimal result we need to select the best value for K. Here, the silhouette score is used to find the best value for K. A range of values from 2 to 10 was considered, KNN clustering was performed on the dataset and the silhouette score was calculated and plotted on a line plot. From the plot we can see that a K value of 2 provides the best score. This K value is used for the K-Means Clustering Technique. The K-Means labels obtained were included in the top neighbourhood's dataset for examining the characteristics of each cluster.

5. Results and Discussion

Using the clusters and the top venue categories let's visualize the top 5 venue category in each Cluster for comparison.

Cluster 1

The top venue categories in Cluster 1 are Hotel, African Restaurant, Lounge, Pizza Place, Fast food restaurant.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Ajeromi-Ifelodun	Hotel	African Restaurant	Lounge	Pizza Place	Fast Food Restaurant	Chinese Restaurant	Nightclub	Shopping Mall	Movie Theater	Park
6	Surulere	Hotel	Lounge	Pizza Place	Shopping Mall	Nightclub	Department Store	Park	Fast Food Restaurant	Bar	African Restaurant
11	Lagos Mainland	Lounge	Shopping Mall	Hotel	Department Store	Pizza Place	African Restaurant	Nightclub	Bar	Fast Food Restaurant	Art Gallery
14	Apapa	Hotel	Lounge	African Restaurant	Restaurant	Ice Cream Shop	Nightclub	Art Gallery	Pizza Place	Chinese Restaurant	Café

Cluster 2

The top venue categories in Cluster 2 are Restaurant, Hotel, Café, Supermarket and Spa

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
16	Epe	Restaurant	Hotel	Café	Supermarket	Spa	Dessert Shop	Discount Store	Campground	Farm	Skating Rink

Cluster 3

The top venue categories in Cluster 3 are Fast Food Restaurant, Airport Lounge, Hotel, Restaurant and Pizza Place.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alimosho	Fast Food Restaurant	Airport Lounge	Hotel	Restaurant	Pizza Place	Burger Joint	Shopping Mall	Lounge	Boutique	Soup Place

Cluster 4

The top venue categories in Cluster 4 are Fast Food Restaurant, Shopping Mall, African Restaurant, Bar, Nightclub.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Kosofe	Fast Food Restaurant	Shopping Mall	African Restaurant	Bar	Nightclub	Pizza Place	Restaurant	Convenience Store	BBQ Joint	Department Store
4	Oshodi-Isolo	Fast Food Restaurant	Shopping Mall	African Restaurant	Hotel	Pizza Place	Bar	Bakery	Nightclub	Restaurant	Department Store
9	Somolu	Fast Food Restaurant	Shopping Mall	Pizza Place	African Restaurant	Bar	Nightclub	Department Store	Restaurant	Lounge	Convenience Store

Cluster 5

The top venue categories in Cluster 5 are Soccer Field, Train Station, French Restaurant, Home Service and Bagel Shop.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Mushin	Soccer Field	Train Station	French Restaurant	Home Service	Bagel Shop	Nightclub	Plaza	Electronics Store	Shopping Mall	Basketball Stadium

Cluster 6

The top venue categories in Cluster 6 are Restaurant, Mediterranean Restaurant, Eastern European Restaurant, Supermarket and Nature Preserve.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Ojo	Restaurant	Mediterranean Restaurant	Eastern European Restaurant	Supermarket	Nature Preserve	Hardware Store	Castle	Diner	Fast Food Restaurant	Garden Center

Cluster 7

The top venue categories in Cluster 7 are Hotel, Lounge, Fast Food Restaurant, Department store and Beach.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Eti-Osa	Hotel	Lounge	Fast Food Restaurant	Department Store	Beach	Chinese Restaurant	Spa	Pizza Place	Convenience Store	Ice Cream Shop
15	Lagos Island	Hotel	Lounge	Bus Station	Ice Cream Shop	Shopping Mall	Chinese Restaurant	Convenience Store	Department Store	Beach	Fast Food Restaurant

Cluster 8

The top venue categories in Cluster 8 are Hotel, Fast food Restaurant, Park, Beach and Gym.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Amuwo-Odofin	Hotel	Fast Food Restaurant	Park	Beach	Gym	Food	Flea Market	Mobile Phone Shop	Movie Theater	Diner

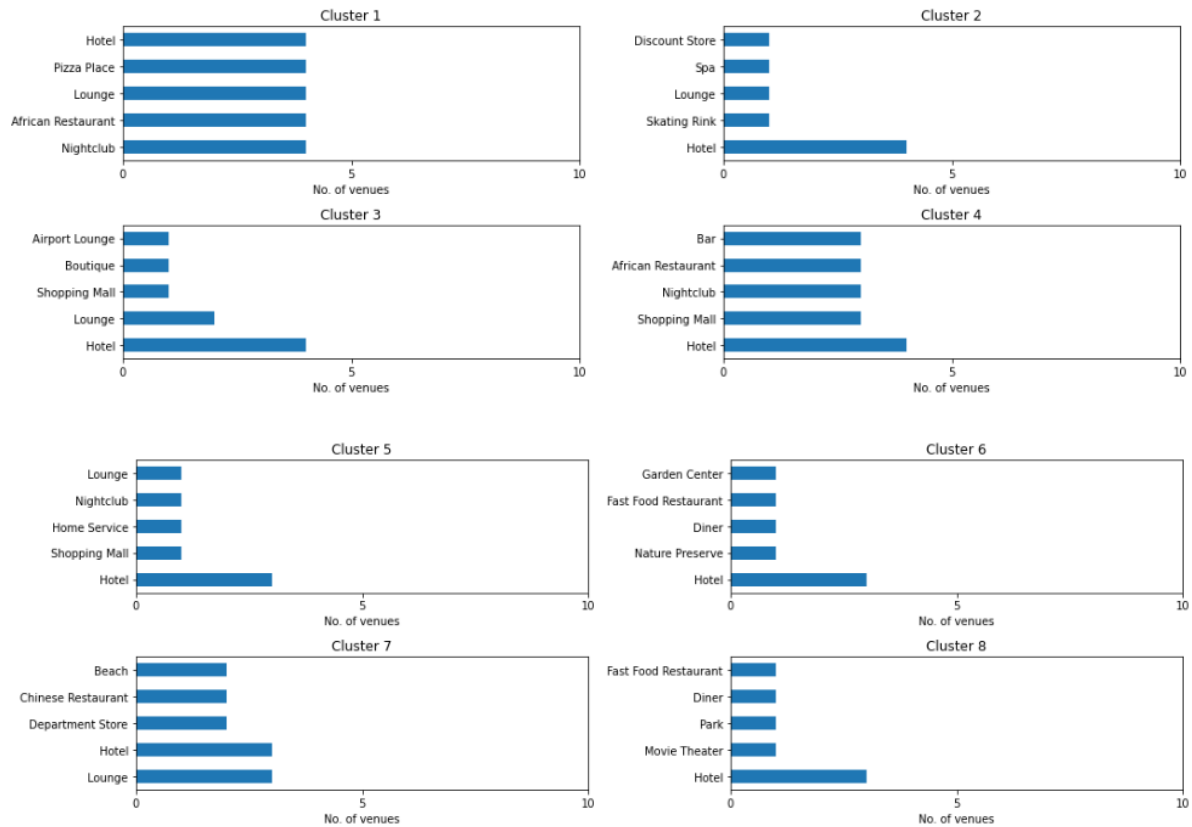
Cluster 9

The top venue categories in Cluster 9 are Fast food Restaurant, Shopping mall, Hotel, Convenience Store and Bar.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Agege	Fast Food Restaurant	Shopping Mall	Hotel	Convenience Store	Bar	Lounge	Pizza Place	Restaurant	Park	Chinese Restaurant
8	Ifako-Ijaiye	Fast Food Restaurant	Shopping Mall	Hotel	Burger Joint	Bus Station	Restaurant	Chinese Restaurant	Beer Garden	Convenience Store	Pizza Place
12	Ikeja	Fast Food Restaurant	Hotel	Pizza Place	Shopping Mall	Bar	Convenience Store	Restaurant	Lounge	BBQ Joint	Department Store

6. DISCUSSION

Now that we have the clusters and the top venue categories we visualize the top 5 venue category in each Cluster for comparison.



This plot can be used to suggest valuable information to Business persons. Let's discuss a few examples considering they would like to start the following category of business.

1. Ice Cream Shop

The only neighbourhoods cluster with the highest number of ice cream shop is 4 while the rest has few or none. Hence opening one in others will be a likely good move.

2. Pizza Place

Just like ice cream shop, only neighbourhoods cluster 3 highest number of Pizza Place while the rest has few or none.

3. African Restaurant

Since Lagos is one of the biggest cities in Africa, once would expect all neighbourhoods cluster to have good number of African restaurants, but base on this analysis only cluster 2 and 7 have.

7. Conclusion

Purpose of this project was to analyse the neighbourhoods of Lagos, Nigeria and create a clustering model to suggest personal places to start a new business based on the category. The neighbourhood's data was obtained from Wikipedia and the Foursquare API was used to find the major venues in each neighbourhood. But we found that many neighbourhoods had less than 10 venues returned. In order to build a good Data Science model, we filtered out these locations. The remaining locations were used to create a clustering model. The best number of clusters was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster.

A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. Both these can be used by stakeholders to decide the location for the particular type of business.

A possible shortfall of this exercise.

A major drawback of this project was that the Foursquare API returned only few venues in each neighbourhood. As a future improvement, better data sources can be used to obtain more venues in each neighbourhood. This way the neighbourhoods that were filtered out can be included in the clustering analysis to create a better decision model.

Thank you

I hope you found the project useful and interesting. Feel free to contact me if you have any queries or suggestions.

This notebook was developed by [Adebayo Amodu](#).