Finding similar neighbourhoods for a thriving restaurant.

Adedamola Adesoye

July 2, 2021

INTRODUCTION

Lagos State, the smallest in area of Nigeria's 36 states, is arguably the most economically important state of the country despite being the smallest in area. There within the boundaries of the Eti-Osa Local Government Area (LGA) lies the state's main business and financial centre, Victoria Island (VI). It is also one of the most exclusive and expensive areas to reside in the state.

To earn my IBM Data Science Professional Certificate, I will look to solve an abstract problem pertaining to a restaurant in VI.

BUSINESS PROBLEM

For quite some time, R.S.V.P, an American restaurant, has thrived in Lagos State. Its board of directors believes they know the recipe for its success and are looking to replicate this elsewhere. Aside from the recipes for its sought-after delicacies, another ingredient for this success recipe is its strategic location in VI. "Where else in Lagos can we open a new restaurant with a similar neighbourhood to that of our current restaurant in VI?" is the question the board is asking and that I will attempt.

DATA

To find suitable spots, I will comb the 20 local government areas (LGAs) of Lagos and their subdivions. To do this I will need:

- A list of all LGAs and wards in Lagos state, sourced from the Lagos state Wikipedia page
- Geospatial coordinates of all these wards using Google's geocoding API
- All nearby venues of each ward using Foursquare's Places API.

METHODOLOGY

- After importing the dependencies, I scraped the Lagos state Wikipedia page for the table below.

LGA	Wards
Agege	Isale/Idimangoro; Iloro/Onipetesi; Oniwaya/Papa-Uku; Agbotikuyo/Dopemu; Oyewole/Papa Ashafa; Okekoto; Keke; Darocha; Tabon Tabon/Oko Oba; Orile Agege/Oko Oba; Isale Odo
Ajeromi/Ifelodun	Ago Hausa; Awodi-Ora; Wilmer; Olodi; Tolu; Temidire I; Ojo Road; Layeni; Alaba Oro; Mosafejo; Temidire II
Alimosho	Shasha/Akowonjo; Egbeda/Alimosho; Idimu/Isheri Olofin; Akesan; Ikotun/Ijegun; Egbe/Agodo; Igando/Egan; Ipaja North; Ipaja South; Ayobo/Ijon Village (Camp David); Pleasure/Oke-Odo; Abule-Egba/Aboru/Meiran/Alagbado
Amuwo-Odofin	Amuwo-Odofin Housing Estate, Mile 2; Festac 1; Festac II; Kirikiri; Amuwo; Ijegun; Satellite; Irede; Ibeshe; Igbologun; Festac III
Apapa	Apapa I (Marine Rd. and environs); Apapa II (Liverpool Rd. and environs); Apapa III (Creek Rd. Tincan/Snake Island; Apapa IV (Pelewura Crescent and environs); Ijora-Oloye; Olodan St. Olojowou St/Alh. Dogo Olatokunbo St. Iganmu; Gaskiya & environs; Afolabi Alasia Str. and environs; Malu Road and environs; Sari and environs

- I converted these data, inserted them in a dataframe and cleaned the dataframe to make sure each row represents a ward and its LGA. This resulted in a dataframe with 363 rows and 2 columns. I also renamed the “Wards” column: “Neighbourhood”.

Out[4]:

	LGA	Neighbourhood
0	Agege	Isale
1	Agege	Idimangoro
2	Agege	Iloro
3	Agege	Onipetesi
4	Agege	Oniwaya
...	...	...
358	Surulere	Iresaadu
359	Surulere	Iregba
360	Surulere	Iwofin
361	Surulere	Ilajue
362	Surulere	Mayin

363 rows × 2 columns

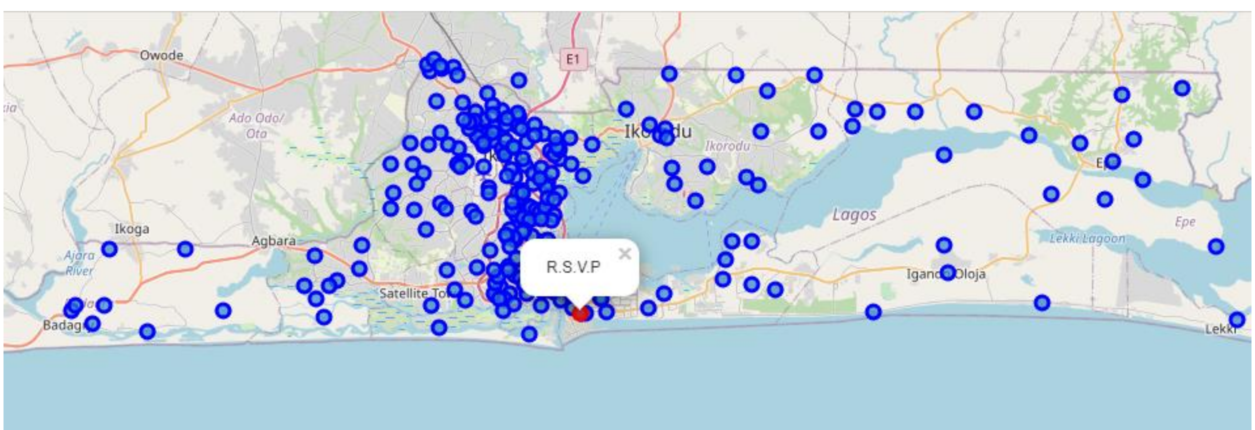
- Using Google’s Geocoding API, I imported the geospatial coordinates of each Neighbourhood and incorporated them with the dataframe. For 95 neighbourhoods, the API returned coordinates clashing with those already registered and for 3 others, none at all. I retained the other neighbourhoods.

Out[203]:

	LGA	Neighbourhood	Latitude	Longitude
0	Agege	Isale	6.61401	3.32504
1	Agege	Idimangoro	6.60338	3.32001
2	Agege	Iloro	6.60873	3.31749
3	Agege	Onipetesi	6.60925	3.32981
4	Agege	Oniwaya	6.61756	3.32034
...	...	...	...	...
260	Surulere	Gambari	6.51041	3.3442
261	Surulere	Iresaapa	8.05633	4.34624
262	Surulere	Iregba	8.0328	4.51089
263	Surulere	Iwofin	8.16415	4.44908
264	Surulere	Ilajue	6.49829	3.34857

265 rows × 4 columns

Here is a map visualization of these coordinates in blue circle and R.S.V.P in red using Python's Folium library.



- Using the Foursquare's Places API, I got a list of all venues within a 1500m radius of each neighbourhood's geospatial coordinates.

	LGA	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agege	Isale	6.614014	3.325041	KFC	6.620788	3.317968	Fast Food Restaurant
1	Agege	Isale	6.614014	3.325041	Shoprite Ikeja	6.614340	3.331319	Shopping Mall
2	Agege	Isale	6.614014	3.325041	Cement Bus Stop	6.607652	3.318337	Bus Station
3	Agege	Isale	6.614014	3.325041	Mango Busstop	6.606112	3.317628	Bus Station
4	Agege	Isale	6.614014	3.325041	Access Bank Plc Ret Shop - Aluminum Village (011)	6.609436	3.314277	Bank

- I did the same with the coordinates of R.S.V.P restaurant.

	LGA	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	R.S.V.P	R.S.V.P	6.428207	3.421662	R.S.V.P	6.428207	3.421662	Restaurant
1	R.S.V.P	R.S.V.P	6.428207	3.421662	Casper & Gambini's	6.429813	3.418668	Modern European Restaurant
2	R.S.V.P	R.S.V.P	6.428207	3.421662	Spice Route	6.429557	3.419823	Asian Restaurant
3	R.S.V.P	R.S.V.P	6.428207	3.421662	Viceroy Restaurant	6.427469	3.423786	Indian Restaurant
4	R.S.V.P	R.S.V.P	6.428207	3.421662	Craft Gourmet	6.433277	3.420752	Mediterranean Restaurant

I will digress for a few paragraphs to explain what **K-Means clustering algorithm is** and **how I intend to use it** before I continue with the **next step**.

What is K-Means clustering?

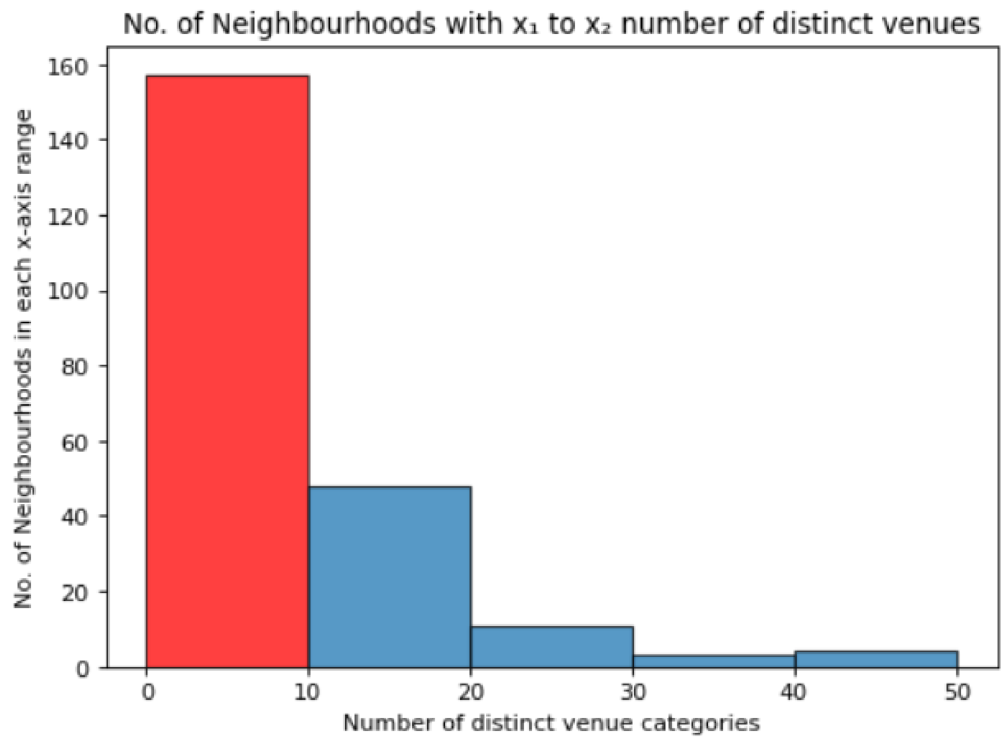
K-Means clustering is an algorithm that takes in a dataset and groups every data point according to how similar they are to other data points in that group. These groups are called clusters (which literally means a group of similar things). The goal of the algorithm is to increase similarity within clusters and dissimilarities between clusters. The similarities between these data points is derived from the properties (or columns) of the data points.

How I intend to use it.

I intend to create a dataframe with each neighbourhood being a data point and its properties being the ten most common venue categories within its 1500m radius.

Neighbourhood	1st Most Common Venue Cat.	2nd Most Common Venue Cat.	3rd Most Common Venue Cat.	4th Most Common Venue Cat.	5th Most Common Venue Cat.	6th Most Common Venue Cat.	7th Most Common Venue Cat.	8th Most Common Venue Cat.	9th Most Common Venue Cat.	10th Most Common Venue Cat.

While exploring the venues dataset, I discovered that most of the neighbourhoods (157 in number) returned less than 10 nearby venues or less than 10 distinct venue categories.



Including these in the final dataset will skew the clustering results because these neighbourhoods will have a top ten most common venue categories when they truly do not.

For example, Coker returned only 5 nearby venues (4 distinct venue categories) while Iddo returned 10 venues but only 8 distinct venue categories.

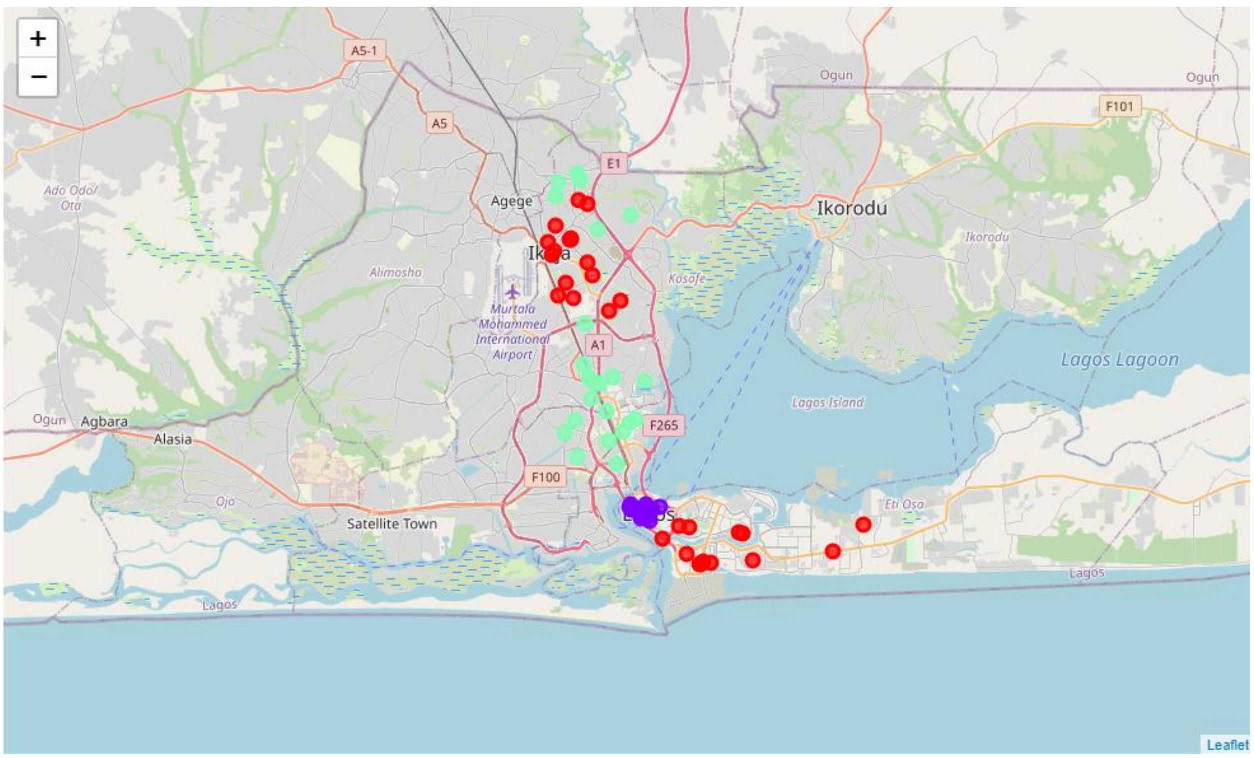
	LGA	Neighbourhood	Venue	Venue Category
1	Lagos Mainland	Iddo	Iddo terminus	Train Station
2	Lagos Mainland	Iddo	Access Bank Plc Ret Shop - Iddo, Ebute Metta)	Bank
3	Lagos Mainland	Iddo	Mr Biggs	Fast Food Restaurant
4	Lagos Mainland	Iddo	idumota	Pharmacy
5	Lagos Mainland	Iddo	Consolidated Breweries	Liquor Store
6	Lagos Mainland	Iddo	CFAO Motors Nig. Ltd	Rental Car Location
7	Lagos Mainland	Iddo	Tantalizers - Oyingbo	Fast Food Restaurant
8	Lagos Mainland	Iddo	Indomie HQ	Noodle House
9	Lagos Mainland	Iddo	Mr Biggs, Apapa Road	Fast Food Restaurant
10	Lagos Mainland	Iddo	Premium Seafoods	Fish Market
1	Surulere	Coker	tank and tummy festac	Fast Food Restaurant
2	Surulere	Coker	C. N. Okoli Transport	Bus Station
3	Surulere	Coker	Odiche House	Bakery
4	Surulere	Coker	Kelin Court Hotel	Hotel
5	Surulere	Coker	Villa-Park Hotel & Resorts	Hotel

As seen below, Iddo's 9th and 10th most common venue categories are Cocktail Bar and Coffee Shop when there is none nearby while Coker has none of the 5th to 10th nearby.

	LGA	Neighbourhood	1st Most Common Venue Cat.	2nd Most Common Venue Cat.	3rd Most Common Venue Cat.	4th Most Common Venue Cat.	5th Most Common Venue Cat.	6th Most Common Venue Cat.	7th Most Common Venue Cat.	8th Most Common Venue Cat.	9th Most Common Venue Cat.	10th Most Common Venue Cat.
0	Lagos Mainland	Iddo	Fast Food Restaurant	Bank	Rental Car Location	Liquor Store	Fish Market	Noodle House	Pharmacy	Train Station	Cocktail Bar	Coffee Shop
1	Surulere	Coker	Hotel	Fast Food Restaurant	Bakery	Bus Station	Cricket Ground	Dog Run	Diner	Dessert Shop	Department Store	Deli / Bodega

Next step

I excluded all of the 157 neighbourhoods and was down to 66 neighbourhoods for clustering. After feeding the algorithm the remaining dataset, the 3 different clusters of these neighbourhoods are visualized below.



RESULTS

Aside from Victoria Island Phase I, where R.S.V.P is located, there are 27 wards with similar neighbourhoods to R.S.V.P. These are visualized on the map below.

```
array(['Victoria Island I', 'Victoria Island II', 'Ilasan Housing Estate',
       'Ikate and environs', 'Ilado', 'Ikoyi I', 'Ikoyi II', 'Obalende',
       'Ikeja', 'Agidingbi', 'Alausa', 'Ipodo', 'Seriki Aro',
       'Adekunle Vill.', 'Adeniyi Jones', 'Onigbongbo',
       'Military Cantonment', 'Gra', 'Police Barracks', 'Opebi', 'Allen',
       'Ajao Estate', 'Kosofe', 'Oko-Awo', 'Onikan', 'Lafiaji', 'Maroko',
       'Alasia'], dtype=object)
```

Of these neighbourhoods are Victoria Island II, Ikoyi I & II, Alausa, Government Reserved Area of Ikeja. These are some of the most luxurious areas in Lagos state befitting a restaurant of R.S.V.P's opulence.

DISCUSSION

This, of course, does not imply that those zones are actually optimal locations for a new restaurant! The purpose of this analysis was to only provide areas with similar neighbourhoods to that of R.S.V.P's. Many other factors can be considered like competition in the area, land availability, etc.

With all these said, this analysis can be greatly improved upon. Granted I am a fledgling data science enthusiast, the data available was both unreliable and insufficient.

Foursquare API returned mostly outdated venue locations. Many existing venues were missing in the API's returns. For example, it returned 12 Pharmacy stores and 21 banks in Lagos State, the state with the largest Gross Domestic Profit (GDP) in Africa's largest economy. Even with a radius of 1500m, most of the reference locations returned less than ten venues. This should be due to OpenStreetMap, Foursquare API's data source, being mostly crowdsourced and few volunteers in the country.

CONCLUSION

The purpose of this project was to identify the most similar neighbourhoods to that of R.S.V.P. The nearest venues to all wards in Lagos state were sourced from Foursquare API. The dataset was sieved to retain wards with at least 10 distinct venues. Finally, K-means clustering algorithm was used to identify all neighbourhoods similar to that of R.S.V.P.

Stakeholders, based on specific characteristics of these neighborhoods, will make final decision on optimal location. These characteristics could be attractiveness of each location, levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.