Magical Kenya: An exposition of key locations to travellers

1. Introduction

1.1 Background

Kenya is a country in the Eastern part of Africa. It is a popular tourist destination that boasts of spectacular sites including a coastline on the Indian Ocean that bequeaths it beautiful sandy beaches and rich marine life; amazing and varied wildlife; the Great Rift Valley, which is a series of trenches that extend across the Middle East and Africa; world heritage sites; lakes; forests; and friendly people, peace and tranquility.

For being only one of eleven countries that sit on the equator, Kenya enjoys a favorable warm equatorial climate throughout the year which helps to galvanize its attraction to tourists from around the world. The country has also made a name on the global map for being home to outstanding athletes and runners.

Tourism is a key pillar of its economy and with so much to offer visitors, Kenya aggressively markets itself to both foreign and local tourists. In 2018, international visitors were 2.03 million while domestic tourists were 3.98 million.

1.2 Audience

These figures represent people, people moving into and around the country with the objective of experiencing fun and beauty. And these objectives always are accompanied by the need for venues for food, drink and accommodation.

This project therefore aims to explore and segment locations that are found around Kenya and act as a guide to people in movement especially to an area they are unfamiliar with whether local tourists or foreigners. The objective is to give a feel of familiarity and knowledge of what to find in various parts of Kenya at a glance of the map.

The results of this project would be useful as a visual aid for travellers as well as by those promoting travel to particular destinations around the country to targeted customers.

By empowering travellers with familiarity and knowledge, it is expected that travellers can do better plans and get to visit more places with less hassle or guesswork while spending more. This in turn would lead to more satisfied travellers and more business to the country.

2. Data

In order to explore and segment the locations, data needed included names of key towns, geographical coordinates of these towns, and the many venues that exist in or near these towns.

The towns, also called neighborhoods are the places that travellers would go to. This data was sourced from Wikipedia's list of cities and towns in Kenya [1].

Below is a screenshot of the Wikipedia page, my interest is the City/Town column.

№	City/ Town	Status	Population	County	Ref
	Nairobi Metro	Metro	6,547,547	Nairobi Kajiado Machakos Kiambu	[1]
1.	Nairobi	City	3,375,000	Nairobi	[2]
2.	Mombasa	City	1,500,000	Mombasa	[3]
3.	Kisumu	City	1,268,826	Kisumu	[4]
4.	Nakuru	Municipality	500,000	Nakuru	[4]
5.	Eldoret	Municipality	400,000	Uasin Gishu	[4]
6.	Kehancha	Municipality	256,086	Migori	[4]

Geographical coordinates for the towns were determined through geopy as below.

	Neighborhood	Latitude	Longitude
0	Mombasa	-4.039015	39.648391
1	Nairobi	-1.283253	36.817245
2	Kisumu	-0.102911	34.754176

Venues that exist within and near these towns were determined using the Foursquare API via code on Jupyter Notebook. Below are examples of venues near Malindi.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malindi	-3.216599	40.116593	Karen Brixen Cafe Malindi	-3.209104	40.117532	Café
1	Malindi	-3.216599	40.116593	Osteria del Chianti, Malindi	-3.219514	40.125946	Italian Restaurant

3. Methodology

The data consisted of the key travel destinations (towns) of Kenya. Their latitude and longitude was determined using the geopy library. Jupyter Notebook and Python were used the scripting platform. The resultant code and Jupyter Notebooks were shared via Github.

3.1 Coverage

For each town, a radius of 10,000 meters was considered. This long radius was considered since the towns venues are fairly spread out, unlike the dense settlements in cities.

3.2 Venues

Venues in each of the towns were retrieved using Foursquare API. Below is a count of the number of venues returned for the towns.

Neighborhood	Venue
Aberdare	100
Amboseli	37
Arabuko Sokoke	68
Bamburi	90
Baringo	25
Bisanadi	4
Bogoria	63

Out of all the venues found in all the towns, unique categories were 194. Using one-hot encoding methodology, each of these venue categories were converted into features and it was then determined how many of each of these categories each town had. This quantification was represented as a mean.

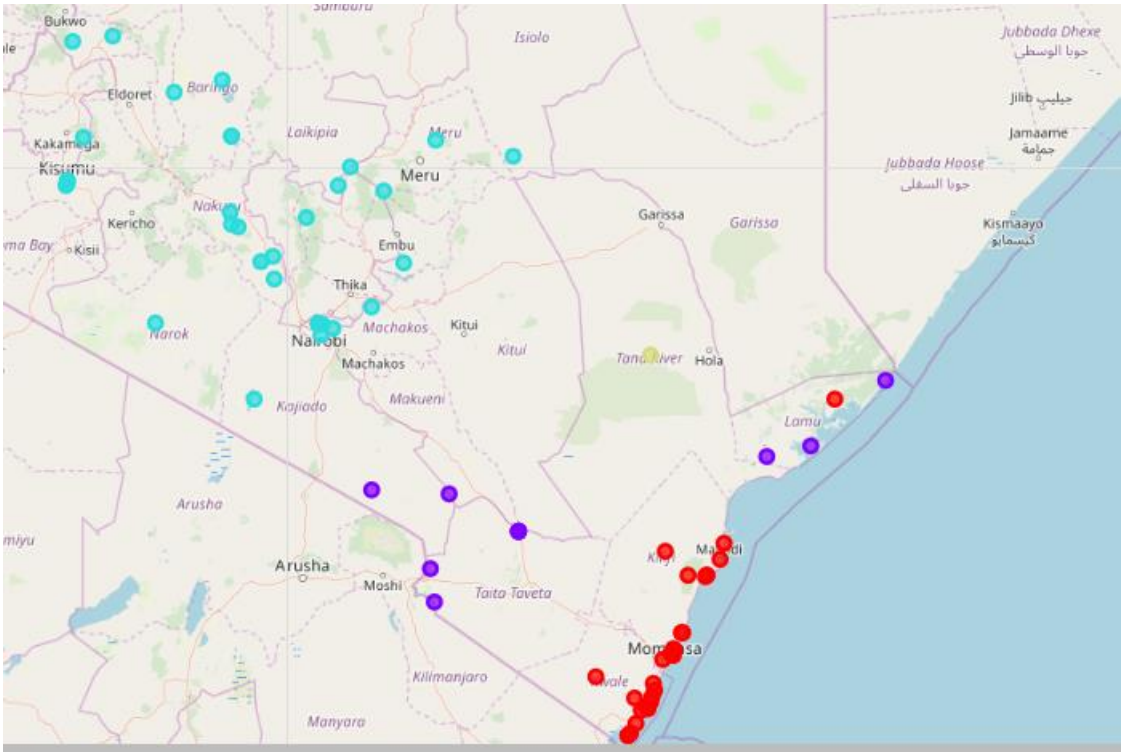
	Neighborhood	Zoo Exhibit	ATM	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	..
0	Aberdare	0.00	0.000000	0.060000	0.000000	0.00	0.000000	0.00	0.00	0.00	.
1	Amboseli	0.00	0.027027	0.027027	0.027027	0.00	0.000000	0.00	0.00	0.00	.
2	Arabuko Sokoke	0.00	0.044118	0.029412	0.014706	0.00	0.000000	0.00	0.00	0.00	.
3	Bamburi	0.00	0.044444	0.033333	0.011111	0.00	0.000000	0.00	0.00	0.00	.
4	Baringo	0.00	0.000000	0.080000	0.000000	0.00	0.000000	0.00	0.00	0.00	.

For each town, I then determined the most popular types of venues, the top 10 venue categories to be precise.

	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	Aberdare	Hotel	Resort	Lounge	Café	African Restaurant	Shopping Mall	Restaurant	Bar	Coffee Shop	BBQ Joint
1	Amboseli	Hotel	Resort	Café	Park	Scenic Lookout	Mountain	Italian Restaurant	Tourist Information Center	Campground	RV Park

3.3 Segmenting the venues

To segment the venues and thereby bring out the similarities between them, k-means clustering algorithm was used. Four clusters were generated and visualized on a map using the Folium library.



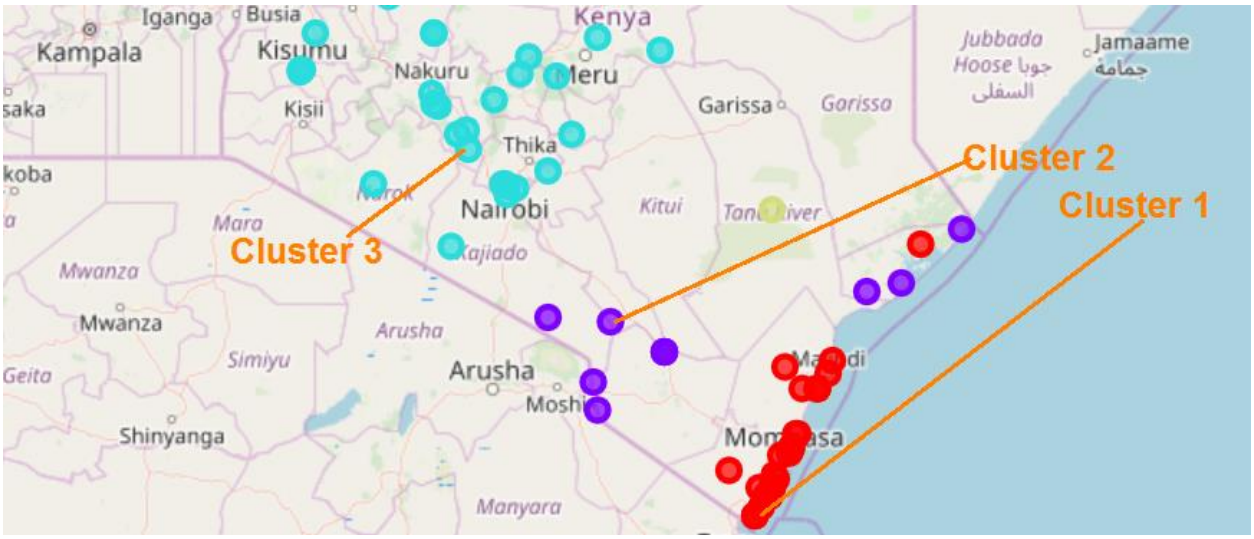
4. Results

The segmentation had results as flows:

Cluster 1 consisted of seafood restaurants, beaches and hotels which were placed right next to the ocean.

Cluster 2 consisted of hotels, scenic lookouts, mountains, forests and parks.

Cluster 3 consists of hotels, shopping malls, lounges, night clubs and national parks.



5. Discussion

The k-means clustering did achieve a fair amount of segmenting Kenya. This can definitely be improved by availing more data on towns. Such data would include a more comprehensive list of the town names complete with grading of the towns to give an analyst a feel of the size of the town or its level of economic activity.

The towns were spread apart, so the standard 10,000 meters that I used for all towns can be varied to be shorter for towns that are more densely populated, and made longer for the towns where settlements are spread out.

6. Conclusion

Kenya is a beautiful country where you find a wide variety including beaches, hotels, restaurants, national parks, forests, and museums.

With improvements to the quality of data, data mining methods such as k-means clustering may be relied on to do real-time display and visualization of information to interested parties, including tourists.

7. References

https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Kenya_by_population