Table of Contents

(NB: Most Tables in this document are partial representations of the whole and meant for illustrative purposes)

# Section 1: Introduction/Business Problem & Data Used

Introduction/Business Problem

We are Africa Tours, and our goal is to boost tourism within Africa. We believe that there is a shift occurring in Africans' minds due to the decolonization movement. As a result, they are becoming less focused on the allure of western countries, and starting to open their eyes to what Africa has to offer. This is leading to an increased desire to travel the continent, especially amongst the younger generation (23 to 40). However, there is a lack of information on the Internet about African cities and what there is to see and do in them. This information gap results in young African travellers either going abroad or staying local.

Data Used

Africa Tours seeks to close this information gap by leveraging website, coordinate, and Foursquare data. The website data will come from a backpacking site and be used to identify all the major cities in Africa.

http://www.hostels247.com/hostel_247blog_Countries__in__Africa/:
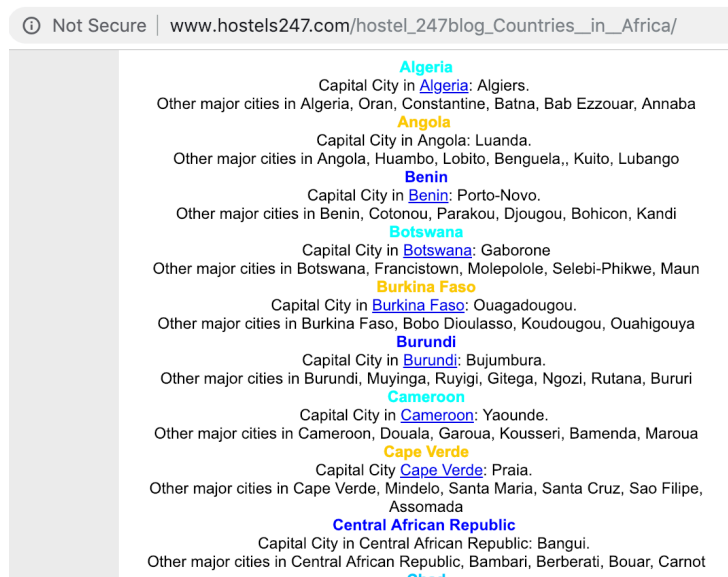
(**figure 1.0**)

It will be turned into a database that looks like the following:

	Country	City	Type
0	Algeria	Algiers	CC
1	Algeria	Oran	MC
2	Algeria	Constantine	MC
3	Algeria	Batna	MC
4	Algeria	Bab Ezzouar	MC
5	Algeria	Annaba	MC
6	Angola	Luanda	CC
7	Angola	Huambo	MC

$$(1.1)$$

Where CC and MC stand for capital city and major city respectively. The coordinate data will be combined with the above table to produce a dataset such as the following:

	Country	City	Type	Latitude	Longitude
0	Algeria	Algiers	CC	28	2.99998
1	Algeria	Oran	MC	35.7033	-0.649298
2	Algeria	Constantine	MC	36.3645	6.60826
3	Algeria	Batna	MC	35.5544	6.17675
4	Algeria	Bab Ezzouar	MC	36.722	3.18567
5	Algeria	Annaba	MC	36.8982	7.75493
6	Angola	Luanda	CC	-8.82727	13.244
7	Angola	Huambo	MC	-12.7765	15.732

$$(1.2)$$

Then the Foursquare data will be used to retrieve information on the top venues in the listed cities via a json file like the one below.

```
'venue': {'id': '4ef936d330f894d869ff6fd3',
 'name': "Jardin d'essais d'El Hamma",
 'location': {'address': 'Rue Hassiba Ben Bouali - B.P. 141, Hamma - El Anasser',
  'lat': 36.7487193385486,
  'lng': 3.0758285522460938,
  'labeledLatLngs': [{'label': 'display',
    'lat': 36.7487193385486,
    'lng': 3.0758285522460938}],
  'distance': 5836,
  'postalCode': '16000',
  'cc': 'DZ',
  'country': 'الجزائر',
  'formattedAddress': ['Rue Hassiba Ben Bouali - B.P. 141, Hamma - El Anasser',
   '16000',
   'الجزائر']},
 'categories': [{'id': '4bf58dd8d48988d163941735',
   'name': 'Park',
   'pluralName': 'Parks',
   'shortName': 'Park',
```

$$(1.3)$$

Afterwards, the relevant information from the json file will be turned into a dataframe that looks like this:

|	Country	City	Type	Venue Name	Venue Category
0	Algeria	Algiers	CC	Casbah İstanbul	Turkish Restaurant
1	Algeria	Algiers	CC	Best Night Hotel	Hotel
2	Algeria	Algiers	CC	Salad box Algerie	Salad Place
3	Algeria	Algiers	CC	Jardin d'essais d'El Hamma	Park
4	Algeria	Algiers	CC	The Crystal Lounge	Restaurant
5	Algeria	Algiers	CC	Taste Of India	Indian Restaurant
6	Algeria	Algiers	CC	Le Tantra	French Restaurant
7	Algeria	Algiers	CC	LEONARD	Restaurant
8	Algeria	Algiers	CC	Piano Piano	Lounge
9	Algeria	Algiers	CC	Cosmopolitain	Bar

$$(1.4)$$

The dummied (See figure 2.0) version of the above table will then be used to cluster the cities according to the venues they have in common. For example:

```
cluster_dict['cluster1']
```

	Country	City	Type	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	Algeria	Oran	MC	Hotel	Restaurant	Café	Ice Cream Shop	Boat or Ferry	Sandwich Place	Breakfast Spot	Shopping Mall	Burger Joint	Seafood Restaurant
12	Egypt	Cairo	CC	Historic Site	Lounge	Café	Italian Restaurant	Egyptian Restaurant	Theater	Hotel	Middle Eastern Restaurant	Coffee Shop	Pastry Shop
13	Egypt	Alexandria	MC	Café	Seafood Restaurant	Coffee Shop	Ice Cream Shop	Sandwich Place	Middle Eastern Restaurant	Historic Site	Restaurant	Egyptian Restaurant	Syrian Restaurant
29	Morocco	Rabat	CC	Café	Historic Site	Restaurant	Hotel	Moroccan Restaurant	French Restaurant	Diner	Coffee Shop	Dessert Shop	Pizza Place
30	Morocco	Casablanca	MC	Café	Coffee Shop	Hotel	Fast Food Restaurant	French Restaurant	Burger Joint	Pub	Tapas Restaurant	Gastropub	Restaurant

(1.5)

These clusters will then be analysed and named based on their distinctive features. This will allow for easier identification of cities on the map of Africa based on type. For example:
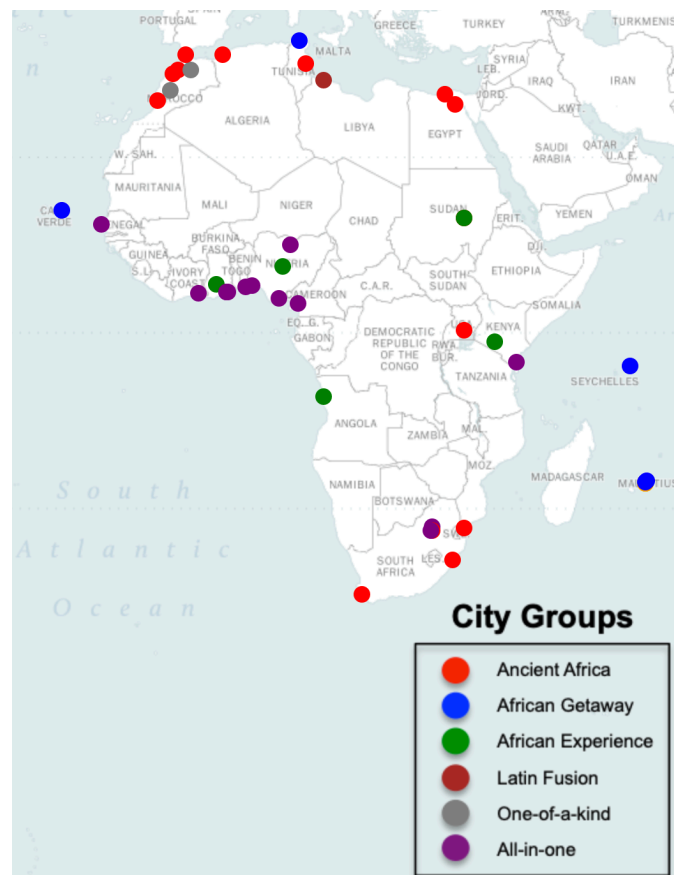
(1.6)

If a customer is planning a tour, they could visit a city from each cluster to ensure a unique experience of a lot of what Africa has to offer. Or they could visit multiple cities in a single cluster if they like one type of city. Either way, they can get the best experience for their budget.

Potential travellers will be able subscribe to our service and gain access to our database, which will enable them to confidently plan their exploration of Africa.

Section 2: Methodology

In order to cluster the cities, machine learning was used in the form of a k-means clustering algorithm. After the venues for each city were put into a dataframe (see figure 1.5), they were assigned dummy variables so that they could be passed into the unsupervised learning model. For example:

City	Restaurant	Roof Deck	Rugby Pitch	Salad Place
Algiers	0	0	0	0
Algiers	0	0	0	0
Algiers	0	0	0	1
Algiers	0	0	0	0
Algiers	1	0	0	0

(2.0)

Then each venue was grouped by city and its average calculated to see how many times it occurs in that city.

City	Accessories Store	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Terminal	American Restaurant
Algiers	0.0	0.021739	0.000000	0.0	0.021739	0.021739	0.021739
Oran	0.0	0.000000	0.000000	0.0	0.000000	0.025641	0.000000
Bab Ezzouar	0.0	0.027778	0.000000	0.0	0.027778	0.027778	0.000000
Luanda	0.0	0.065574	0.000000	0.0	0.000000	0.000000	0.016393
Cotonou	0.0	0.027778	0.000000	0.0	0.027778	0.000000	0.000000
Gaborone	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.025641

(2.1)

Thereafter, venues were sorted into city_venues_sorted, a data frame that shows the **top 20** most frequently occurring venues for each city. As well as its coordinates, and the country it is in.

City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Algiers	Hotel	Restaurant	Mediterranean Restaurant	Coffee Shop	French Restaurant
Oran	Hotel	Restaurant	Café	Ice Cream Shop	Boat or Ferry
Bab Ezzouar	Hotel	Mediterranean Restaurant	Coffee Shop	Restaurant	Lake
Luanda	Restaurant	Pizza Place	African Restaurant	Lounge	Hotel
Cotonou	Resort	Hotel	Restaurant	Shopping Mall	Hotel Bar

$$(2.2)$$

In order to add countries to city_venues_sorted, it was merged on the 'City' column with part of another dataframe, afrimain_df3, to create city_venues_final:

Country	City	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Algeria	Algiers	36.7	3.15	Hotel	Restaurant	Mediterranean Restaurant	Coffee Shop	French Restaurant
Algeria	Oran	35.7033	-0.649298	Hotel	Restaurant	Café	Ice Cream Shop	Boat or Ferry
Algeria	Bab Ezzouar	36.722	3.18567	Hotel	Mediterranean Restaurant	Coffee Shop	Restaurant	Lake
Angola	Luanda	-8.82727	13.244	Restaurant	Pizza Place	African Restaurant	Lounge	Hotel
Benin	Cotonou	6.3677	2.42525	Resort	Hotel	Restaurant	Shopping Mall	Hotel Bar

$$(2.3)$$

The k-means clustering algorithm was then run on city_venues_grouped (fig 2.1). It placed each city into a cluster between 0-10. These labels were incorporated into city_venues_final to create city_venues_merged:

Country	City	Type	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Algeria	Algiers	CC	28	2.99998	9	Hotel	Restaurant	Mediterranean Restaurant	Coffee Shop	French Restaurant
Algeria	Oran	MC	35.7033	-0.649298	0	Hotel	Restaurant	Café	Ice Cream Shop	Boat or Ferry
Algeria	Bab Ezzouar	MC	36.722	3.18567	9	Hotel	Mediterranean Restaurant	Coffee Shop	Restaurant	Lake
Angola	Luanda	CC	-8.82727	13.244	2	Restaurant	Pizza Place	African Restaurant	Lounge	Hotel
Benin	Cotonou	MC	6.3677	2.42525	7	Resort	Hotel	Restaurant	Shopping Mall	Hotel Bar

(2.4)

The next phase was to analyse the clusters.

Comparing and naming clusters

The next goal was to analyse the clusters and give them more descriptive names based on the insights from the analysis. Distinguishing the clusters is necessary for identification and marketing purposes. In order to compare them effectively, each original cluster like cluster1 (see figure 1.5) was morphed into a single column dataframe then put into one table with the others.

	Venues	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
0	Café	14	4	3	3	1	2	4	9	3	4
1	Hotel	14	5	5	3	1	2	5	12	1	6
2	Coffee Shop	13	NaN	3	NaN	1	1	2	4	1	4
3	Restaurant	12	5	5	3	1	2	4	11	3	6
4	Seafood Restaurant	12	5	NaN	NaN	1	NaN	NaN	NaN	2	NaN
5	Italian Restaurant	10	5	2	NaN	1	NaN	2	6	2	3
6	Fast Food Restaurant	8	2	5	3	1	NaN	5	8	3	3
7	Pizza Place	8	NaN	4	3	1	NaN	NaN	7	3	3
8	Bar	7	3	4	3	NaN	NaN	2	8	NaN	NaN
9	Burger Joint	7	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4
10	Indian Restaurant	7	3	2	NaN	NaN	NaN	NaN	NaN	3	3

(2.5)

There were 68 venues in total. NaN means that that venue does not appear amongst the top 20 venues in that cluster. Whereas a number means that it appears that many times. To see what each cluster has that the others do not, NaN was filtered out as follows:

```
clusters_mrgd_new[clusters_mrgd_new['cluster1']!='NaN']
```

	Venues	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
0	Café	14	4	3	3	1	2	4	9	3	4
1	Hotel	14	5	5	3	1	2	5	12	1	6
2	Coffee Shop	13	NaN	3	NaN	1	1	2	4	1	4
3	Restaurant	12	5	5	3	1	2	4	11	3	6
4	Seafood Restaurant	12	5	NaN	NaN	1	NaN	NaN	NaN	2	NaN
5	Italian Restaurant	10	5	2	NaN	1	NaN	2	6	2	3
6	Fast Food Restaurant	8	2	5	3	1	NaN	5	8	3	3
7	Pizza Place	8	NaN	4	3	1	NaN	NaN	7	3	3
8	Bar	7	3	4	3	NaN	NaN	2	8	NaN	NaN
9	Burger Joint	7	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4
10	Indian Restaurant	7	3	2	NaN	NaN	NaN	NaN	NaN	3	3
11	Ice Cream Shop	6	NaN	2	NaN	1	NaN	NaN	5	2	NaN
12	Lounge	6	2	NaN	3	NaN	2	2	9	3	6
13	Art Gallery	6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	French Restaurant	5	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4
15	Middle Eastern Restaurant	5	NaN	NaN	NaN	1	NaN	2	5	NaN	NaN
16	Historic Site	5	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN
17	Plaza	5	NaN	NaN	3	1	1	NaN	NaN	NaN	NaN
18	Shopping Mall	5	2	3	3	1	NaN	5	13	3	3
19	Diner	4	NaN	NaN	2	1	1	NaN	NaN	2	NaN

$$(2.6)$$

Section 3: Results

Using the above method, clusters were compared and named based on their unique offerings. The following are the findings:

Cluster 1
- (See figure 2.6 for table)
- Has 14 Cities
    - Search '**Out[247]**' in **Notebook** for specific cities

1) **General offerings** (A venue is a general offering if five or more clusters have it):

- Greatest quantity and variety of restaurants, cafes, and hotels.

2) **Moderately unique offerings** (A venue is a moderately unique offering if three or four clusters have it):

- Restaurants:
    + French
    + Middle Eastern
- Burger joints
- Ice cream shops
- Breakfast spots

3) **Highly unique offerings** (A venue is highly unique offering if one or two clusters have it):

- Historic sites
- Art Galleries

This cluster would be a cosmopolitan food lovers dream. It also has a lot to offer for people who consider themselves to be refined. It is the only cluster with a dense frequency of art galleries and one of two with a number of historic sites. Additionally, it has old countries like Egypt. Therefore, this will be classified as the **Ancient Africa** Cluster.

Cluster 2
- 5 Cities
  - '**Out[256]**'

Venues	Ancient Africa	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	3	1	2	4	9	3	4
Hotel	14	5	5	3	1	2	5	12	1	6
Restaurant	12	5	5	3	1	2	4	11	3	6
Seafood Restaurant	12	5	NaN	NaN	1	NaN	NaN	NaN	2	NaN
Italian Restaurant	10	5	2	NaN	1	NaN	2	6	2	3
Fast Food Restaurant	8	2	5	3	1	NaN	5	8	3	3
Bar	7	3	4	3	NaN	NaN	2	8	NaN	NaN
Indian Restaurant	7	3	2	NaN	NaN	NaN	NaN	NaN	3	3
Lounge	6	2	NaN	3	NaN	2	2	9	3	6
French Restaurant	5	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4
Shopping Mall	5	2	3	3	1	NaN	5	13	3	3
Resort	NaN	5	NaN	2	NaN	2	NaN	NaN	NaN	NaN
Beach	NaN	5	NaN	NaN	NaN	NaN	NaN	4	NaN	NaN
Cajun / Creole Restaurant	NaN	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Farmers Market	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Cocktail Bar	NaN	2	NaN	3	NaN	1	3	NaN	NaN	NaN
Pub	NaN	2	NaN	1	NaN	NaN	NaN	NaN	NaN	NaN
Nightclub	NaN	2	NaN	NaN	NaN	NaN	NaN	8	NaN	2
Mediterranean Restaurant	NaN	2	NaN	NaN	NaN	NaN	2	4	NaN	3
Bistro	NaN	2	NaN	NaN	NaN	1	3	NaN	NaN	NaN

(**3.0**)

1) **General offerings:**

   - Nothing of note

2) **Moderately unique offerings:**

   - Restaurants:
     + Mediterranean
     + French
   - Cocktail bars
   - Bistros
   - Resorts

3) **Highly unique offerings:**

- Steakhouses
- Nightclubs
- Spas
- Restaurants:
    + Portuguese
    + Cajun/Creole
- Beaches

This cluster has a number of unique offerings including the most beaches and resorts. Therefore, this will be classified as the **African Getaway** Cluster.

Cluster 3
- 5 Cities
    o '**Out[261]**'

Venues	Ancient Africa	African Getaway	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	3	1	2	4	9	3	4
Hotel	14	5	5	3	1	2	5	12	1	6
Coffee Shop	13	NaN	3	NaN	1	1	2	4	1	4
Restaurant	12	5	5	3	1	2	4	11	3	6
Italian Restaurant	10	5	2	NaN	1	NaN	2	6	2	3
Fast Food Restaurant	8	2	5	3	1	NaN	5	8	3	3
Pizza Place	8	NaN	4	3	1	NaN	NaN	7	3	3
Bar	7	3	4	3	NaN	NaN	2	8	NaN	NaN
Burger Joint	7	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4
Indian Restaurant	7	3	2	NaN	NaN	NaN	NaN	NaN	3	3
Ice Cream Shop	6	NaN	2	NaN	1	NaN	NaN	5	2	NaN
Shopping Mall	5	2	3	3	1	NaN	5	13	3	3
African Restaurant	NaN	NaN	5	NaN	NaN	NaN	NaN	10	NaN	3
Movie Theater	NaN	NaN	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Chinese Restaurant	NaN	NaN	3	NaN	NaN	NaN	NaN	NaN	3	NaN
Convenience Store	NaN	NaN	2	NaN	NaN	NaN	2	NaN	NaN	NaN
Fried Chicken Joint	NaN	NaN	2	NaN	1	NaN	NaN	NaN	NaN	NaN
Pharmacy	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Supermarket	NaN	NaN	2	NaN	1	NaN	2	NaN	NaN	NaN
Clothing Store	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN

(**3.1**)

1) **General offerings:**

   - Nothing of note

2) **Moderately unique offerings:**

   - Restaurants:
     - Chinese
     - African
   - Burger joints

3) **Highly unique offerings:**

   - Bus stations
   - Fried chicken joints
   - Clothing stores
   - Movie theaters

What stands out about this cluster is its high number of African restaurants, as well as the fact that it has movie theatres and bus stations. The stations imply that it would be fairly easy to get around even if you don't have a car. And the theatres make one think that this cluster is quite metropolitan. However, the presence of stations among the top venues also suggests that these places are not too westernized. Therefore, this cluster will be classified as the **African Experience** cluster

Cluster 4
- 3 Cities
  o '**Out[264]**'

Venues	Ancient Africa	African Getaway	African Experience	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	3	1	2	4	9	3	4
Hotel	14	5	5	3	1	2	5	12	1	6
Restaurant	12	5	5	3	1	2	4	11	3	6
Fast Food Restaurant	8	2	5	3	1	NaN	5	8	3	3
Pizza Place	8	NaN	4	3	1	NaN	NaN	7	3	3
Bar	7	3	4	3	NaN	NaN	2	8	NaN	NaN
Lounge	6	2	NaN	3	NaN	2	2	9	3	6
Plaza	5	NaN	NaN	3	1	1	NaN	NaN	NaN	NaN
Shopping Mall	5	2	3	3	1	NaN	5	13	3	3
Diner	4	NaN	NaN	2	1	1	NaN	NaN	2	NaN
Resort	NaN	5	NaN	2	NaN	2	NaN	NaN	NaN	NaN
Cocktail Bar	NaN	2	NaN	3	NaN	1	3	NaN	NaN	NaN
Pub	NaN	2	NaN	1	NaN	NaN	NaN	NaN	NaN	NaN
Airport	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN
Furniture / Home Store	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	1	NaN
Grocery Store	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	2	2
Pool	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN
Market	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN
Bakery	NaN	NaN	NaN	2	1	NaN	NaN	7	NaN	NaN
Comedy Club	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN	NaN

(**3.2**)

1) **General offerings:**

- Nothing of note

2) **Moderately unique offerings:**

- Nothing of note

3) **Highly unique offerings:**

- Nothing of note

All of its venues are general, including its unique offerings. Therefore, this cluster will be omitted, as it has nothing distinct to offer.

Cluster 5
- 1 City
  - '**Out[269]**'

Venues	Ancient Africa	African Getaway	African Experience	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	1	2	4	9	3	4
Hotel	14	5	5	1	2	5	12	1	6
Coffee Shop	13	NaN	3	1	1	2	4	1	4
Restaurant	12	5	5	1	2	4	11	3	6
Seafood Restaurant	12	5	NaN	1	NaN	NaN	NaN	2	NaN
Italian Restaurant	10	5	2	1	NaN	2	6	2	3
Fast Food Restaurant	8	2	5	1	NaN	5	8	3	3
Pizza Place	8	NaN	4	1	NaN	NaN	7	3	3
Ice Cream Shop	6	NaN	2	1	NaN	NaN	5	2	NaN
Middle Eastern Restaurant	5	NaN	NaN	1	NaN	2	5	NaN	NaN
Plaza	5	NaN	NaN	1	1	NaN	NaN	NaN	NaN
Shopping Mall	5	2	3	1	NaN	5	13	3	3
Diner	4	NaN	NaN	1	1	NaN	NaN	2	NaN
Fried Chicken Joint	NaN	NaN	2	1	NaN	NaN	NaN	NaN	NaN
Supermarket	NaN	NaN	2	1	NaN	2	NaN	NaN	NaN
Bakery	NaN	NaN	NaN	1	NaN	NaN	7	NaN	NaN
Soccer Field	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN
Tea Room	NaN	NaN	NaN	1	2	NaN	NaN	NaN	NaN
Latin American Restaurant	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN
Park	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN

(**3.3**)

1) **General offerings:**

- Restaurants:
  1) Seafood
  2) Italian
  3) Fast Food
  4) Pizza

2) **Moderately unique offerings:**

- Bakeries
- Diners
- Middle Eastern Restaurants
- Ice cream shops

3) **Highly unique offerings:**

  - Latin American Restaurants
  - Tearooms
  - Fried chicken joints

The most unique thing about this cluster is that it is the only one that has Latin American restaurants. This suggests that there is some Latin American influence on the culture. The fusion of Latin American and African culture could be quite an interesting thing to experience. This cluster will be classified as the **Latin Fusion** African Cluster.

Cluster 6
- 2 Cities
  - o '**Out[272]**'

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	cluster6	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	1	2	4	9	3	4
Hotel	14	5	5	1	2	5	12	1	6
Coffee Shop	13	NaN	3	1	1	2	4	1	4
Restaurant	12	5	5	1	2	4	11	3	6
Lounge	6	2	NaN	NaN	2	2	9	3	6
Historic Site	5	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN
Plaza	5	NaN	NaN	1	1	NaN	NaN	NaN	NaN
Diner	4	NaN	NaN	1	1	NaN	NaN	2	NaN
Resort	NaN	5	NaN	NaN	2	NaN	NaN	NaN	NaN
Cocktail Bar	NaN	2	NaN	NaN	1	3	NaN	NaN	NaN
Bistro	NaN	2	NaN	NaN	1	3	NaN	NaN	NaN
Tea Room	NaN	NaN	NaN	1	2	NaN	NaN	NaN	NaN
Spa	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN
Moroccan Restaurant	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN
Bed & Breakfast	NaN	NaN	NaN	NaN	2	3	NaN	NaN	NaN
Boutique	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN
Castle	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN
Garden	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN
History Museum	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN
Snack Place	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN

(**3.4**)

1) **General offerings:**
   - Nothing of note

2) **Moderately unique offerings:**

   - French restaurants
   - Resorts
   - Bistros
   - Diners

3) **Highly unique offerings:**

   - Historic sites
   - Spas
   - Tearooms
   - Bed & Breakfasts
   - Moroccan restaurants

- Museums
- Gardens
- Pubs
- Molecular gastronomy restaurants
- Sandwich places
- Castles

The most unique thing about this cluster is that it has castles, molecular gastronomy and Moroccan restaurants, and museums among other things. Given that it has the most unique offerings out of all of the clusters we have examined, it will be classified as the One-of-a-kind cluster.

Cluster 7
- 5 Cities
  - o '**Out[276]**'

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	One-of-a-kind	cluster7	cluster8	cluster9	cluster10
Café	14	4	3	1	2	4	9	3	4
Hotel	14	5	5	1	2	5	12	1	6
Coffee Shop	13	NaN	3	1	1	2	4	1	4
Restaurant	12	5	5	1	2	4	11	3	6
Italian Restaurant	10	5	2	1	NaN	2	6	2	3
Fast Food Restaurant	8	2	5	1	NaN	5	8	3	3
Bar	7	3	4	NaN	NaN	2	8	NaN	NaN
Lounge	6	2	NaN	NaN	2	2	9	3	6
Middle Eastern Restaurant	5	NaN	NaN	1	NaN	2	5	NaN	NaN
Shopping Mall	5	2	3	1	NaN	5	13	3	3
Cocktail Bar	NaN	2	NaN	NaN	1	3	NaN	NaN	NaN
Mediterranean Restaurant	NaN	2	NaN	NaN	NaN	2	4	NaN	3
Bistro	NaN	2	NaN	NaN	1	3	NaN	NaN	NaN
Convenience Store	NaN	NaN	2	NaN	NaN	2	NaN	NaN	NaN
Supermarket	NaN	NaN	2	1	NaN	2	NaN	NaN	NaN
Bed & Breakfast	NaN	NaN	NaN	NaN	2	3	NaN	NaN	NaN
Steakhouse	NaN	NaN	NaN	NaN	NaN	4	NaN	1	NaN
Food Court	NaN	NaN	NaN	NaN	NaN	2	4	2	NaN
Portuguese Restaurant	NaN	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN
Food	NaN	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN

(**3.5**)

1) **General offerings:**

- Variety of places to eat

2) **Moderately unique offerings:**

 - Restaurants:
   + Chinese
   + Mediterranean
   + Middle Eastern
   + Bistros
   + Cocktail bars

3) **Highly unique offerings:**

 - Bed & Breakfasts
 - Steakhouses
 - Portuguese restaurants

This cluster does not have any highly unique venues that set it apart. Therefore, it will be dropped.

Cluster 8
- 13 Cities
  - o '**Out[280]**'

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	One-of-a-kind	cluster8	cluster9	cluster10
Café	14	4	3	1	2	9	3	4
Hotel	14	5	5	1	2	12	1	6
Coffee Shop	13	NaN	3	1	1	4	1	4
Restaurant	12	5	5	1	2	11	3	6
Italian Restaurant	10	5	2	1	NaN	6	2	3
Fast Food Restaurant	8	2	5	1	NaN	8	3	3
Pizza Place	8	NaN	4	1	NaN	7	3	3
Bar	7	3	4	NaN	NaN	8	NaN	NaN
Ice Cream Shop	6	NaN	2	1	NaN	5	2	NaN
Lounge	6	2	NaN	NaN	2	9	3	6
Middle Eastern Restaurant	5	NaN	NaN	1	NaN	5	NaN	NaN
Shopping Mall	5	2	3	1	NaN	13	3	3
Beach	NaN	5	NaN	NaN	NaN	4	NaN	NaN
Nightclub	NaN	2	NaN	NaN	NaN	8	NaN	2
Mediterranean Restaurant	NaN	2	NaN	NaN	NaN	4	NaN	3
African Restaurant	NaN	NaN	5	NaN	NaN	10	NaN	3
Bakery	NaN	NaN	NaN	1	NaN	7	NaN	NaN
Food Court	NaN	NaN	NaN	NaN	NaN	4	2	NaN
Department Store	NaN	NaN	NaN	NaN	NaN	6	NaN	3
Thai Restaurant	NaN	NaN	NaN	NaN	NaN	4	NaN	NaN

(**3.6**)

1) **General offerings:**

- The highest occurrence of shopping malls by a significant margin
- The highest occurrence of lounges
- The second highest occurrence of cafe and coffee shops

2) **Moderately unique offerings:**

- Ice Cream Shops
- Restaurants:
  + Middle Eastern
  + African (most)
- Bakeries

3) **Highly unique offerings:**

- Breakfast spots
- Beaches
- Nightclubs
- Thai restaurants

This cluster has a lot to offer in terms of opportunities for shopping and nightlife. Not to mention things to do throughout the day since it has breakfast spots and coffee shops. What's also very unique about it is that it has the most African restaurants by a significant margin, while still having a fair amount of variety when it comes to places to eat. With a breadth of fun, luxurious, and culinary offerings, this cluster will be classified as the All-in-one cluster.

Cluster 9
- 3 Cities
  - ‘**Out[284]**’

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	One-of-a-kind	All-in-one	cluster9	cluster10
Café	14	4	3	1	2	9	3	4
Hotel	14	5	5	1	2	12	1	6
Coffee Shop	13	NaN	3	1	1	4	1	4
Restaurant	12	5	5	1	2	11	3	6
Seafood Restaurant	12	5	NaN	1	NaN	NaN	2	NaN
Italian Restaurant	10	5	2	1	NaN	6	2	3
Fast Food Restaurant	8	2	5	1	NaN	8	3	3
Pizza Place	8	NaN	4	1	NaN	7	3	3
Indian Restaurant	7	3	2	NaN	NaN	NaN	3	3
Ice Cream Shop	6	NaN	2	1	NaN	5	2	NaN
Lounge	6	2	NaN	NaN	2	9	3	6
Shopping Mall	5	2	3	1	NaN	13	3	3
Diner	4	NaN	NaN	1	1	NaN	2	NaN
Chinese Restaurant	NaN	NaN	3	NaN	NaN	NaN	3	NaN
Furniture / Home Store	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN
Grocery Store	NaN	NaN	NaN	NaN	NaN	NaN	2	2
Steakhouse	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN
Food Court	NaN	NaN	NaN	NaN	NaN	4	2	NaN
Gym / Fitness Center	NaN	NaN	NaN	NaN	NaN	NaN	2	NaN
Irish Pub	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN

(**3.7**)

1) **General offerings:**

   - Restaurants:
      + Seafood
      + Italian
      + Fast food
      + Indian
   - Shopping Malls

2) **Moderately unique offerings:**

   - Ice Cream Shops
   - Chinese restaurants
   - Diner

3) **Highly unique offerings:**

   - Movie theaters
   - Gardens

This cluster has no standout characteristics. Therefore it will be dropped.

Cluster 10
- 6 Cities
  - o '**Out[288]**'

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	One-of-a-kind	All-in-one	cluster10
Café	14	4	3	1	2	9	4
Hotel	14	5	5	1	2	12	6
Coffee Shop	13	NaN	3	1	NaN	4	4
Restaurant	12	5	5	1	2	11	6
Italian Restaurant	10	5	NaN	1	NaN	6	3
Fast Food Restaurant	8	2	5	1	1	8	3
Pizza Place	8	NaN	4	1	NaN	7	3
Burger Joint	7	NaN	NaN	NaN	NaN	NaN	4
Indian Restaurant	7	3	2	NaN	NaN	NaN	3
Lounge	6	2	2	NaN	2	9	6
French Restaurant	5	2	NaN	NaN	NaN	NaN	4
Plaza	5	NaN	NaN	1	1	NaN	2
Shopping Mall	5	NaN	3	1	NaN	13	3
Nightclub	NaN	2	NaN	NaN	NaN	8	2
African Restaurant	NaN	NaN	5	NaN	NaN	10	3
Department Store	NaN	NaN	NaN	NaN	NaN	6	3
Mediterranean Restaurant	NaN	NaN	NaN	NaN	NaN	4	3
Cupcake Shop	NaN	NaN	NaN	NaN	NaN	NaN	3
Salad Place	NaN	NaN	NaN	NaN	NaN	NaN	3
Golf Course	NaN	NaN	NaN	NaN	NaN	NaN	2

(**3.8**)

1) **General offerings:**

- Restaurants:
  + Italian
  + Fast food
  + Pizza
  + Indian
- Lounge

2) **Moderately unique offerings:**

- Restaurants
  + French
  + Mediterranean
  + African

3) **Highly unique offerings:**

- Salad place
- Cupcake shop

The most unique aspects of this cluster are that it appears to have restaurants that specifically serve salads, and cupcake shops. However, these aren't features that would be a big draw for tourists because they aren't substantial enough. For this reason, this cluster will be dropped.

Summary of Results

Based on the analysis, clusters were renamed and handled as follows:

Cluster 1 became **Ancient Africa**
- Standout characteristics:
    - Most historic sites, which are very rare
    - Only with art galleries
    - Most French restaurants, which are moderately rare

Cluster 2 became **African Getaway**
- Standout characteristics:
    - Most beaches, which are very rare
    - Most resorts, which are moderately rare

Cluster 3 became **African Experience**
- Standout characteristics:
    - Second most African restaurants, which are moderately rare

Cluster 4 was **dropped**
- Standout characteristics:
    - None

Cluster 5 became **Latin Fusion**
- Standout characteristics:
    - Only with Latin American restaurants

Cluster 6 became **One-of-a-kind**
- Standout characteristics:
    - Only with castles, gardens, and boutique shops.

Cluster 7 was **dropped**
- Standout characteristics:
    - None

Cluster 8 became **All-in-one**
- Standout characteristics:
    - Has beaches, which are moderately rare
    - Most nightclubs, which are moderately rare
    - Has almost every type of restaurant, including African
    - As well as the most shopping malls by far, which are a general venue

Cluster 9 was **dropped**
- Standout characteristics:
    - No distinct offerings

Cluster 10 was **dropped**
- Standout characteristics:
  - No distinct offerings

The final table consists of six clusters and its first 11 rows look as follows:

Venues	Ancient Africa	African Getaway	African Experience	Latin Fusion	One-of-a-kind	All-in-one
Café	14	4	3	1	2	9
Hotel	14	5	5	1	2	12
Coffee Shop	13	NaN	3	1	NaN	4
Seafood Restaurant	12	5	NaN	1	NaN	NaN
Restaurant	12	5	5	1	2	11
Italian Restaurant	10	5	NaN	1	NaN	6
Fast Food Restaurant	8	2	5	1	1	8
Pizza Place	8	NaN	4	1	NaN	7
Bar	7	3	4	NaN	NaN	8
Burger Joint	7	NaN	NaN	NaN	NaN	NaN
Indian Restaurant	7	3	2	NaN	NaN	NaN

(**3.9**)

Final map with the remaining six clusters:
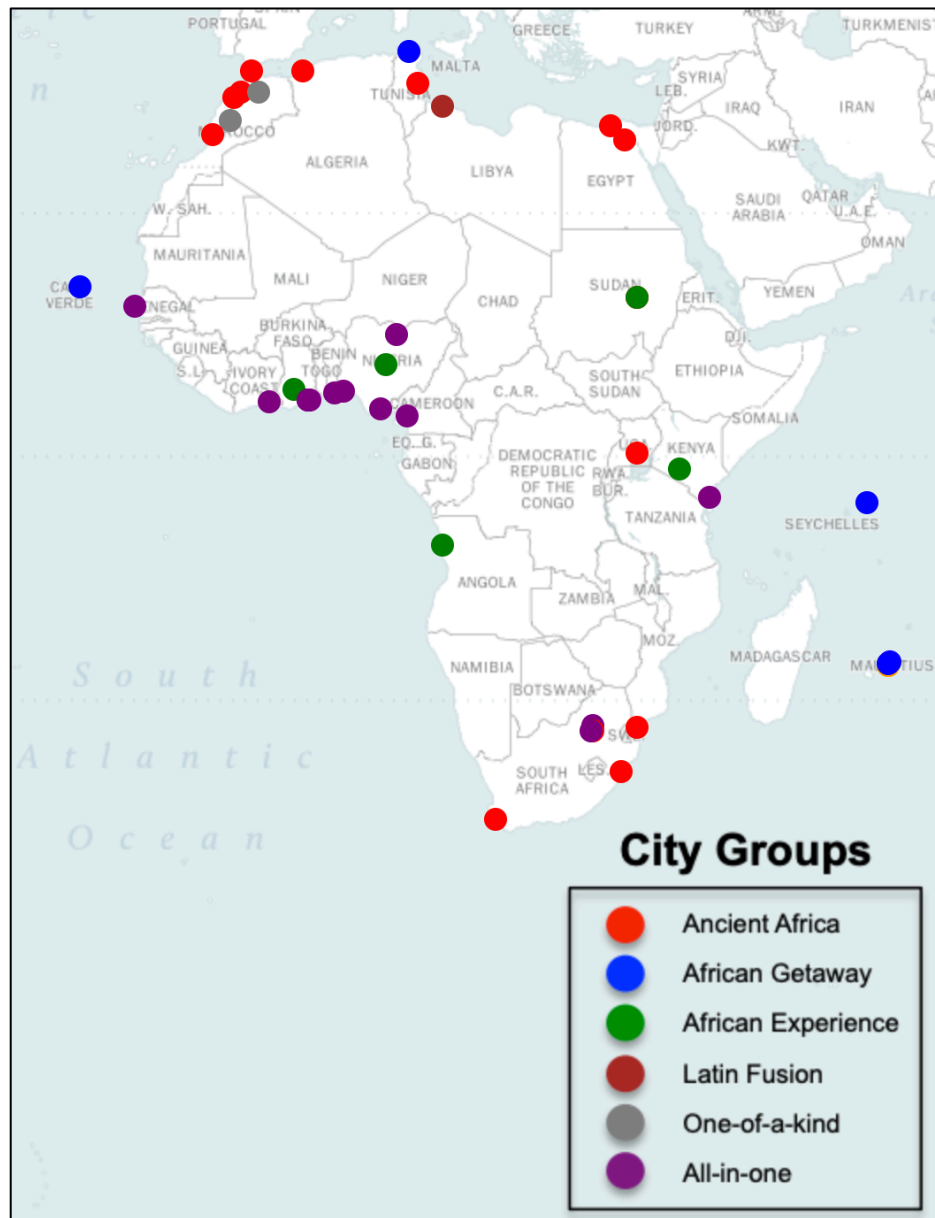
(**3.9.1**)

The html version (Search '**Out[307]**' in **Notebook**) is interactive and allows zooming as well as clicking for pop-up labels:

(3.9.2)
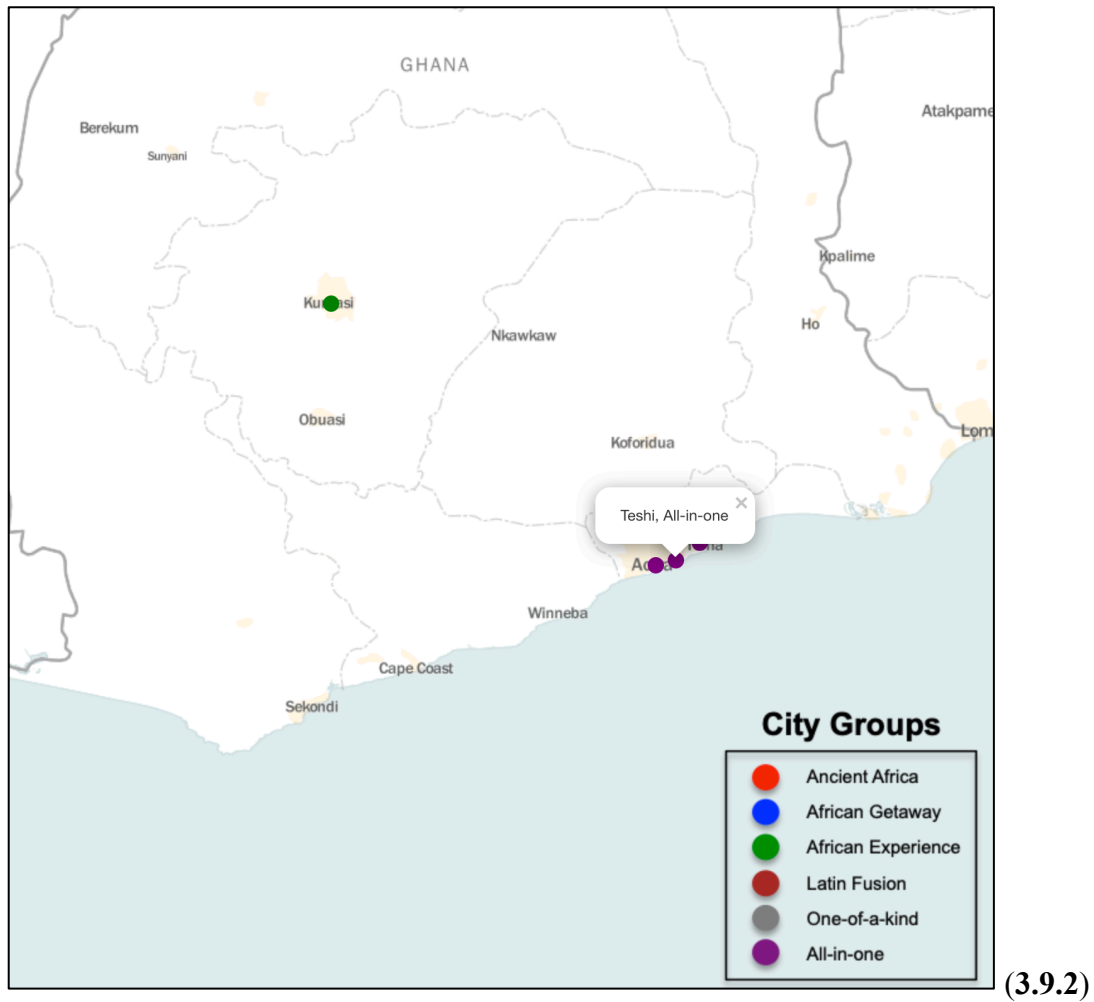
Section 4: Discussion & Conclusion

Discussion

Through this project, it has been discovered that 57 out of 269 major African cities have between 30 or more venues on Foursquare. Those 57 were split into 10 groups based on the similarity of their venues. Of these 10 groups, 6 were found to be distinct and thus offer a unique proposition to potential visitors. Therefore, they were named according to the venues that distinguished them and displayed on a map of Africa (figure 3.9.2). This easy-to-understand collation and synthesis of Foursquare data pertaining to African cities will help potential visitors plan their trips across the continent with greater ease and confidence. Consequently, there will be a boost in tourism within Africa—which could provide a potential revenue stream for Africa Tours.

In order to gain revenue from the findings, Africa Tours could sell the information through a subscription that allows potential visitors to access the cluster database. However, it is important to understand the limitations of the database and what it can and cannot be used for.

For instance, the unique venues for which a cluster is named do not necessarily exist in every city in that cluster. For example, Cluster 1 was named Ancient Africa because it contains the most historic sites, which are highly unique venues that tourists are likely to be interested in. However, only 5 of the 14 countries in cluster 1 have historic sites. Nonetheless, because this was the distinguishing characteristic of the cluster, it was named Ancient Africa. That is why it is recommended that customers use the map to see which cluster piques their interest, then delve into the actual database to see what each country offers before they make their decision. That being said, Cluster 1 is a bit of an exception out of the final clusters. The rest have a higher rare venue to number of countries ratio.

A way to increase this ratio would be to increase the number of clusters used in the K-means algorithm to perhaps 15. This would have the effect of casting the clustering net more narrowly, thus making the clusters smaller and more closely related. For example, it is likely that Ancient Africa would be split into multiple clusters, and one of them would have most of the historic sites. The only problem with this would be that having a lot of clusters becomes unwieldy. Hence why 10 were chosen.

Conclusion

The problem this study sought to solve is that there is not enough information about African cities for travellers compared to western cities. Making it difficult for Africans to plan a vacation to other African countries. In the data collection process, venue data was retrieved and tabulated for major African cities that had more than 30 venues listed on Foursquare—a website used by gallivants and tourists to rate venues they have been to. Thereafter, a K-means clustering machine learning algorithm was run on the venues to group the cities that had the most similar venues. From this operation came 10 groups. They were put side by side in a table and compared to see what stood out about each one that would interest a tourist. Four of them were deemed to be commonplace and dropped as a result. While the remaining ones were named according to their standout venues and displayed on a map of Africa for easy identification.

While valuable, the database alone would not be enough for customers to decide where to go because it does not take into consideration differing cultures, which would greatly affect one's experience of two places even if they have similar venues. Therefore, future research can focus on bringing context to the clusters by including descriptions of country and city cultures. Perhaps by incorporating pictures and the most detailed user comments on Foursquare into the database. These images and descriptions would give subscribers more of a feel for the culture of a place and whether or not it intrigues them. Thus enabling them to make a more informed decision.

Additionally, in the future, there should be a way of ranking the venues if they appear an equal amount of times. For example, in Cluster 9, one of the cities in Mauritius (Curepipe) had a volcano, but this very rare venue did not make it into the comparison table because some other venues like 'resort' got ranked above it when value_counts() was used and the top 20 venues taken. Therefore, venues might have to be looked over manually at the count stage to ensure that the most relevant ones are picked, with the goal of eventually developing additional criteria to automate the process.