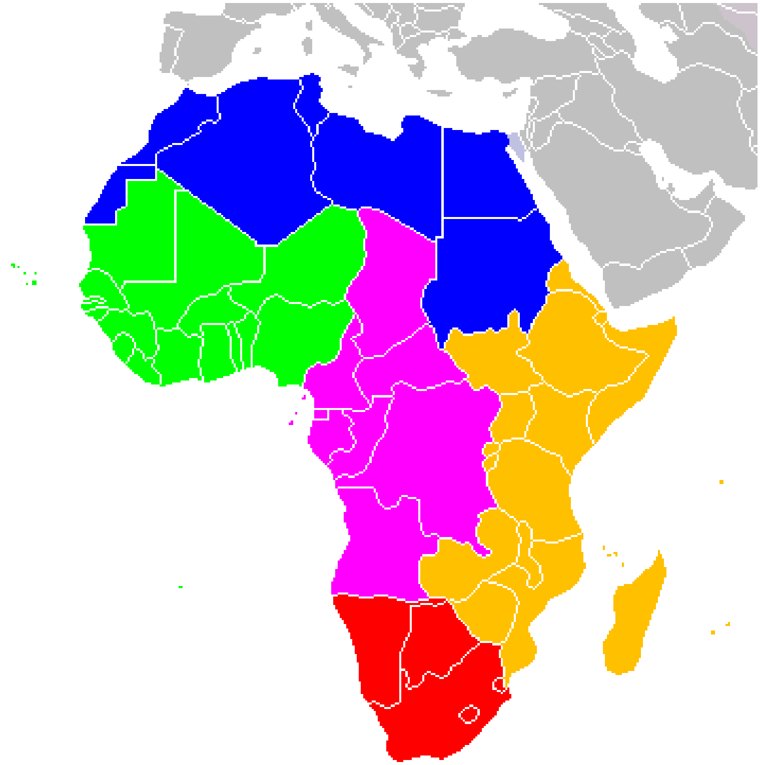
Clustering Similar African Cities using the Four-Square API.



Clustering Similar African Cities using the Four-Square API. Introduction and Data

Introduction

Africa is a diverse region as it is home to many cultures, histories and peoples. The diversity can also be observed in the standard of living and economic growth in the different parts. In this project we will use the location of major African cities and he four square API to seek and cluster similar African cities to cities within the region.

BY clustering similar cities it can be possible to enhance travel and tourism recommendations. This will allow tourists to better plan their trips and have a 'feeling' of what to expect when visitng Africa. Secondly, Africa is an emerging region with abundant natural resouorces and a young population this presents an opportunity for investors looking for higher returns and a strong market positons or monopolies. With the clustering possible in this project tourists, investors and etc, can better understand African Cities what they have to offer and the opportunites arising there.

Data

Our location Data will be retrieved from Simple Maps and Wikipedia. The data will provide the latitude and the longitude of most cities. The location will then be used to leverage Foursquare API to retrieve information on various venues around these locations. With the venue data and the location data we will the run a k-means algorithm to group cities with similar venues. With these clusters the Folium library will be used to plot similar cities.

The Clusters will be coloured and clearly labelled for peer-review. Secondly the Gross Domestic Product calculated for Purchasing Power parity and the populations will be used to produce GDP(PPP) per capita as a good proxy for the standard of living in these countries.

Thus the first data set from Simple Maps will have a coloumns for country, city, latitude, longitude. We shall the use a list of African countries retrieved from Wikipedia to filter our large list from Simple Maps. With this list We shall collect the locations for 150 venues in a 10km radius of the cities, this is possible with Four Squares API. Finally, the complete list will contian the city, country, latitude, longitude, 10 most common venues in a particulare city and the per capita purchaisng power parity.

Methodology

The goal of this project is to cluster similar african capital cities based on their per capita incomes and social life.

Fistly the locations of the african capitals are collect from simple maps, Foursquar and wikipedia. This information includes population, GDP corrceted for purchasing power parity and venue data from Four Square.

Secondly, the required rows are computed, which include GDP per capita and the 10 most popular venues in each city.

Lastly, similar african cities are clustered based on the popular venues and the per capita income.

Analysis

We begin by importing the necessary libraries to perform this analysis

In [143]:
```
import numpy as np # library to handle data in a vectorized manner
```

```python
import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

```
Libraries imported.
```

1.2 Data retrieval

We then use pandas to retrieve our information straight from wikipedia

The global cities database

In [241]:
```
Wcity = pd.read_csv('worldcities.csv')
City_coord = Wcity.iloc[:,[4,0,2,3,8]]
```

In [240]:
```
#Wcity.head()
Wcity = Wcity[Wcity['capital']=='primary']
```

Africa GDP(PPP) and Population

In [149]:
```
W_pop = pd.read_html('https://en.wikipedia.org/wiki/List_of_African_cou
ntries_by_GDP_(PPP)')
```

In [159]:
```
a_gdp_ppp_pc = W_pop[0]
```

In [161]:
```
C_names = a_gdp_ppp_pc.iloc[0]
```

In [162]:
```
C_names
```

Out[162]:
```
0                                           RegionRank
1                                              Country
2    Peak value of GDP (PPP) as of 2019Billions of ...
3                                            Peak Year
Name: 0, dtype: object
```

In [163]:
```
a_gdp_ppp_pc.columns = C_names
```

In [166]:
```
afr_income = a_gdp_ppp_pc.iloc[2:,]
```

In [355]:
```
afr_income.head(5)
```

Out[355]:

	RegionRank	Country	Peak value of GDP (PPP) as of 2019Billions of International dollars	Peak Year
2	1	Egypt	1391.734	2019

	RegionRank	Country	Peak value of GDP (PPP) as of 2019Billions of International dollars	Peak Year
3	2	Nigeria	1214.827	2019
4	3	South Africa	813.100	2019
5	4	Algeria	684.649	2019
6	5	Morocco	330.381	2019

In [174]:
```
population = pd.read_csv('population.csv', encoding = 'latin1')
```

In [176]:
```
popul = population.iloc[:,[0,5]]
```

In [187]:
```
popul.columns
```

Out[187]:
```
Index(['Region, subregion, country or area *', '2018'], dtype='object')
```

The list of African Countries

In [356]:
```
Df = City_coord.merge(popul, how = 'inner', left_on = 'country', right_
on = 'Region, subregion, country or area *')
Df.head(5)
```

Out[356]:

	country	city	lat	lng	capital	Region, subregion, country or area *	2018
0	South Africa	Pretoria	-25.7069	28.2294	primary	South Africa	57792.518
1	South Africa	Bloemfontein	-29.1200	26.2299	primary	South Africa	57792.518
2	South Africa	Cape Town	-33.9200	18.4350	primary	South Africa	57792.518
3	Zambia	Lusaka	-15.4166	28.2833	primary	Zambia	17351.708

	country	city	lat	lng	capital	Region, subregion, country or area *	2018
4	Zimbabwe	Harare	-17.8178	31.0447	primary	Zimbabwe	14438.802

Finally we merge the two datasets

In [243]:
```
df1 = Df.merge(afr_income, how = 'inner', left_on = 'country', right_on
 = 'Country')
df1.head(1)
```

Out[243]:

	country	city	lat	lng	capital	Region, subregion, country or area *	2018	RegionRank	Country	P (F 20 Inte
0	South Africa	Pretoria	-25.7069	28.2294	primary	South Africa	57792.518	3	South Africa	

In [196]:
```
df1.columns
```

Out[196]:
```
Index(['country', 'city', 'lat', 'lng', 'capital',
       'Region, subregion, country or area *', '2018', 'RegionRank', 'C
ountry',
       'Peak value of GDP (PPP) as of 2019Billions of International dol
lars',
       'Peak Year'],
      dtype='object')
```

In [244]:
```
df1['per_cap_inc']=df1['Peak value of GDP (PPP) as of 2019Billions of I
nternational dollars'].astype('float64')*1000000/df1['2018']
```

In [245]:
```
df1.head(1)
```

Out[245]:

	country	city	lat	lng	capital	Region, subregion, country or area *	2018	RegionRank	Country	P (F 20 Inte
0	South Africa	Pretoria	-25.7069	28.2294	primary	South Africa	57792.518	3	South Africa	

In [246]:
```
africa_data = df1.iloc[:,[1,2,3,4,8,11]]
```

In [247]:
```
africa_data.tail()
```

Out[247]:

	city	lat	lng	capital	Country	per_cap_inc
44	Conakry	9.5315	-13.6802	primary	Guinea	2306.776552
45	Malabo	3.7500	8.7833	primary	Equatorial Guinea	28792.757692
46	Bissau	11.8650	-15.5984	primary	Guinea-Bissau	1806.004685
47	Nairobi	-1.2833	36.8167	primary	Kenya	3715.907155
48	Moroni	-11.7042	43.2402	primary	Comoros	1662.817996

We now have the data in the preferred format!!!

2.0 Building our Map with the Folium Library and Foursquare API

In [ ]:
```
location
```

In [214]:
```
address = 'Africa'

geolocator = Nominatim(user_agent="Africa_explorer")
```

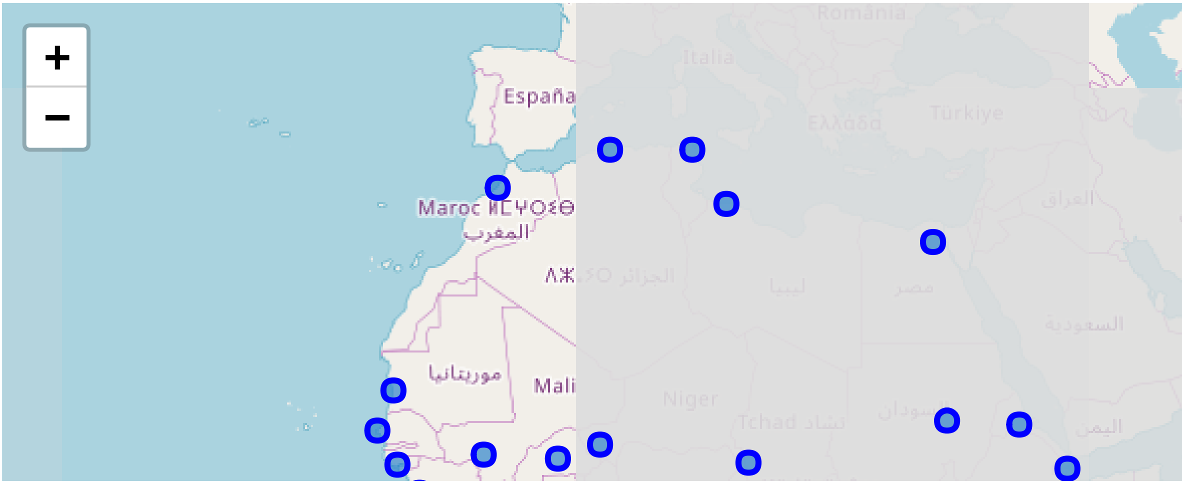
```
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinate of Africa are {}, {}.'.format(latitud
e, longitude))
```

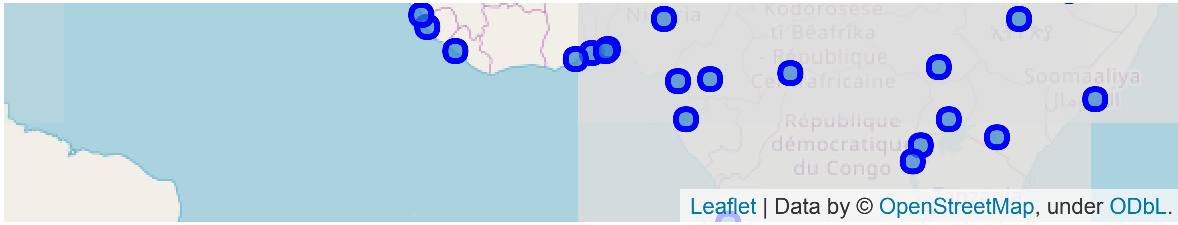
```
The geographical coordinate of Africa are 11.5024338, 17.7578122.
```

In [248]:

```
map_africa = folium.Map(location=[latitude, longitude], zoom_start=3)

# add markers to map
for lat, lng, label in zip(africa_data['lat'], africa_data['lng'], afri
ca_data['city']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_africa)

map_africa
```

Out[248]:





A distribution of our neighbourhoods

In [ ]:

```

```

In [254]:

```
CLIENT_ID = 'AYN2INSTLPV42E31M14BHJR2WFYDRCYTMU3RFIKJFX0KZMWI' #
CLIENT_SECRET = 'VZETUFMMNMOBMDXOYYKDJT5O4FNWPNE5BLA0HK5BKYILCAZT' #t
VERSION = '20180605' # Foursquare API version

print('My credentials are:')
print('CLIENT_ID: ' + 'Some random code')
print('CLIENT_SECRET: ' + 'A Secret')
```

```
My credentials are:
CLIENT_ID: Some random code
CLIENT_SECRET: A Secret
```

In [ ]:

```

```

A function to find nearby venues

In [283]:

```
# This Function will find venues around a selected neighbour h
def getNearbyVenues(names, latitudes, longitudes, radius=9000):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)
```

```
        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id=
{}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['it
ems']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list fo
r item in venue_list])
    nearby_venues.columns = ['city',
                  'city Latitude',
                  'city Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [279]:
```
LIMIT = 30
```

In [284]:
```
africa_venues = getNearbyVenues(names=africa_data['city'],
                                   latitudes=africa_data['lat'],
                                   longitudes=africa_data['lng']
                                  )
```

```
Pretoria
Bloemfontein
Cape Town
Lusaka
Harare
Monrovia
Maseru
Tripoli
Rabat
Antananarivo
Bamako
Nouakchott
Port Louis
Lilongwe
Maputo
Windhoek
Niamey
Abuja
Kigali
Victoria
Khartoum
Freetown
Dakar
Mogadishu
Juba
Ndjamena
Lomé
Tunis
Kampala
Luanda
Ouagadougou
Bujumbura
Porto-Novo
Cotonou
Gaborone
```

```
Bangui
Yaounde
Djibouti
Algiers
Cairo
Asmara
Addis Ababa
Libreville
Accra
Conakry
Malabo
Bissau
Nairobi
Moroni
```

In [285]:
```python
print(africa_venues.shape)
africa_venues.head()
```

```
(1064, 7)
```

Out[285]:

	city	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Pretoria	-25.7069	28.2294	Burger Bistro	-25.722152	28.227975	Burger Joint
1	Pretoria	-25.7069	28.2294	Brewers BBQ	-25.703939	28.240994	BBQ Joint
2	Pretoria	-25.7069	28.2294	Café 41	-25.744691	28.222439	Gastropub
3	Pretoria	-25.7069	28.2294	Fruit Stop	-25.718634	28.205505	Farmers Market
4	Pretoria	-25.7069	28.2294	Royal Danish Icecream	-25.742076	28.242174	Ice Cream Shop

In [286]:
```python
africa_venues.groupby('city').count()
```

Out[286]:

	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category

city	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
city						
Abuja	30	30	30	30	30	30
Accra	30	30	30	30	30	30
Addis Ababa	30	30	30	30	30	30
Algiers	30	30	30	30	30	30
Antananarivo	30	30	30	30	30	30
Asmara	4	4	4	4	4	4
Bamako	17	17	17	17	17	17
Bangui	6	6	6	6	6	6
Bissau	10	10	10	10	10	10
Bloemfontein	30	30	30	30	30	30
Bujumbura	13	13	13	13	13	13
Cairo	30	30	30	30	30	30
Cape Town	30	30	30	30	30	30
Conakry	16	16	16	16	16	16
Cotonou	4	4	4	4	4	4
Dakar	30	30	30	30	30	30
Djibouti	22	22	22	22	22	22
Freetown	10	10	10	10	10	10
Gaborone	30	30	30	30	30	30
Harare	30	30	30	30	30	30
Juba	6	6	6	6	6	6
Kampala	30	30	30	30	30	30

city	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Khartoum	30	30	30	30	30	30
Kigali	30	30	30	30	30	30
Libreville	18	18	18	18	18	18
Lilongwe	23	23	23	23	23	23
Lomé	18	18	18	18	18	18
Luanda	30	30	30	30	30	30
Lusaka	30	30	30	30	30	30
Malabo	12	12	12	12	12	12
Maputo	30	30	30	30	30	30
Maseru	6	6	6	6	6	6
Mogadishu	5	5	5	5	5	5
Monrovia	16	16	16	16	16	16
Moroni	6	6	6	6	6	6
Nairobi	30	30	30	30	30	30
Ndjamena	14	14	14	14	14	14
Niamey	17	17	17	17	17	17
Nouakchott	18	18	18	18	18	18
Ouagadougou	15	15	15	15	15	15
Port Louis	30	30	30	30	30	30
Porto-Novo	8	8	8	8	8	8
Pretoria	30	30	30	30	30	30
Rabat	30	30	30	30	30	30

city	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Tripoli	30	30	30	30	30	30
Tunis	30	30	30	30	30	30
Victoria	30	30	30	30	30	30
Windhoek	30	30	30	30	30	30
Yaounde	30	30	30	30	30	30

In [287]:

```
print('There are {} uniques categories.'.format(len(africa_venues['Venu
e Category'].unique())))
```

```
There are 160 uniques categories.
```

In [288]:

```
# one hot encoding
africa_onehot = pd.get_dummies(africa_venues[['Venue Category']], prefi
x="", prefix_sep="")

# add city column back to dataframe
africa_onehot['city'] = africa_venues['city']

# move Neighbourhood column to the first column
fixed_columns = [africa_onehot.columns[-1]] + list(africa_onehot.column
s[:-1])
africa_onehot = africa_onehot[fixed_columns]

africa_onehot.tail()
```

Out[288]:

	city	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Entertai
1059	Moroni	0	0	0	0	0	0	0	0	

1060	Moroni	0	0	0	0	0	0	0	0

	city	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Entertai
1061	Moroni	0	0	0	0	0	0	0	0	
1062	Moroni	0	0	0	0	0	0	0	0	
1063	Moroni	0	0	0	0	0	0	0	0	

In [289]:

```
africa_grouped = africa_onehot.groupby('city').mean().reset_index()
africa_grouped.head()
```

Out[289]:

	city	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Ent
0	Abuja	0.000000	0.0	0.0	0.033333	0.066667	0.0	0.000000	0.0	
1	Accra	0.100000	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	
2	Addis Ababa	0.000000	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	
3	Algiers	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.033333	0.0	
4	Antananarivo	0.066667	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	

In [264]:

```
africa_grouped[africa_grouped['city'] == 'Nairobi'].T.reset_index().ilo
c[1:]
```

Out[264]:

	index	20
1	African Restaurant	0.09375
2	Arts & Crafts Store	0
3	Athletics & Sports	0

	index	20
4	Bakery	0
5	Bar	0.03125
6	Bed & Breakfast	0
7	Bookstore	0
8	Breakfast Spot	0.03125
9	Brewery	0
10	Buffet	0
11	Burger Joint	0
12	Business Service	0
13	Café	0.03125
14	Casino	0
15	Chinese Restaurant	0
16	Clothing Store	0
17	Cocktail Bar	0.0625
18	Coffee Shop	0.28125
19	Comedy Club	0
20	Convenience Store	0
21	Cultural Center	0
22	Department Store	0
23	Dessert Shop	0.03125
24	Diner	0
25	Electronics Store	0
26	Ethiopian Restaurant	0.03125
27	Fast Food Restaurant	0
28	Flea Market	0

28	Flea Market	0

	index	20
29	Food	0
30	Fried Chicken Joint	0.03125
31	Furniture / Home Store	0
32	Grocery Store	0
33	Gym	0
34	Historic Site	0
35	History Museum	0
36	Hostel	0
37	Hotel	0.03125
38	Hotel Bar	0.03125
39	Ice Cream Shop	0.03125
40	Indian Chinese Restaurant	0
41	Indian Restaurant	0
42	Italian Restaurant	0
43	Juice Bar	0
44	Karaoke Bar	0
45	Light Rail Station	0
46	Lounge	0.0625
47	Middle Eastern Restaurant	0
48	Moroccan Restaurant	0
49	Movie Theater	0
50	Music Venue	0
51	Nightclub	0
52	Other Nightlife	0
53	Park	0

53	Park	0

	index	20
54	Performing Arts Venue	0.0625
55	Pizza Place	0.03125
56	Plaza	0
57	Pool	0
58	Pub	0
59	Restaurant	0.09375
60	Roof Deck	0
61	Seafood Restaurant	0
62	Shopping Mall	0
63	Soccer Field	0
64	Sports Club	0
65	Steakhouse	0
66	Supermarket	0
67	Tea Room	0.03125
68	Theater	0
69	Tunnel	0
70	Used Bookstore	0

In [290]:

```
num_top_venues = 5

for city in africa_grouped['city']:
    print("----"+city+"----")
    temp = africa_grouped[africa_grouped['city'] == city].T.reset_index
()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
```

```
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Abuja----
                venue  freq
0           BBQ Joint  0.07
1               Hotel  0.07
2              Arcade  0.07
3       Movie Theater  0.07
4  Chinese Restaurant  0.07


----Accra----
                venue  freq
0               Hotel  0.13
1  African Restaurant  0.10
2         Pizza Place  0.07
3              Lounge  0.03
4        Dessert Shop  0.03


----Addis Ababa----
                venue  freq
0               Hotel  0.27
1  Italian Restaurant  0.13
2                Café  0.10
3          Restaurant  0.07
4           Nightclub  0.07


----Algiers----
               venue  freq
0  French Restaurant  0.13
1              Hotel  0.10
2       Burger Joint  0.10
3             Lounge  0.07
4              Diner  0.07
```

```
----Antananarivo----
                venue  freq
0               Hotel  0.17
1          Restaurant  0.10
2  African Restaurant  0.07
3   French Restaurant  0.07
4       Grocery Store  0.03


----Asmara----
               venue  freq
0              Hotel  0.50
1            Airport  0.25
2      Movie Theater  0.25
3        Music Venue  0.00
4  Mobile Phone Shop  0.00


----Bamako----
              venue  freq
0             Hotel  0.29
1       Pizza Place  0.06
2               Bar  0.06
3     Grocery Store  0.06
4  Greek Restaurant  0.06


----Bangui----
                venue  freq
0  African Restaurant  0.17
1      Breakfast Spot  0.17
2               Hotel  0.17
3   French Restaurant  0.17
4          Restaurant  0.17


----Bissau----
                 venue  freq
0                Hotel   0.4
```

```
1   African Restaurant   0.1

2  Arts & Crafts Store   0.1
3          Pizza Place   0.1
4                 Port   0.1


----Bloemfontein----
                  venue  freq
0                 Hotel  0.10
1           Coffee Shop  0.10
2         Shopping Mall  0.07
3  Fast Food Restaurant  0.07
4    Seafood Restaurant  0.03


----Bujumbura----
               venue  freq
0              Hotel  0.31
1         Restaurant  0.23
2              Beach  0.08
3            Airport  0.08
4  Indian Restaurant  0.08


----Cairo----
           venue  freq
0  Historic Site  0.13
1         Lounge  0.10
2    Pastry Shop  0.10
3          Hotel  0.07
4           Café  0.07


----Cape Town----
                venue  freq
0         Coffee Shop  0.17
1               Hotel  0.17
2                Café  0.13
3             Theater  0.07
```

```
4  Italian Restaurant  0.03


----Conakry----
                  venue  freq
0                 Hotel  0.38
1     Mobile Phone Shop  0.12
2                  Pier  0.06
3            Restaurant  0.06
4  Gym / Fitness Center  0.06


----Cotonou----
    venue  freq
0  Bakery  0.25
1   Plaza  0.25
2   Beach  0.25
3    Pier  0.25
4    Park  0.00


----Dakar----
                venue  freq
0  African Restaurant  0.13
1      Ice Cream Shop  0.10
2       Shopping Mall  0.10
3               Hotel  0.07
4          Restaurant  0.07


----Djibouti----
                venue  freq
0               Hotel  0.14
1          Restaurant  0.14
2   Convenience Store  0.05
3  Seafood Restaurant  0.05
4       Shopping Mall  0.05
```

```
----Freetown----

                     venue  freq
0                    Hotel   0.2
1              Coffee Shop   0.1
2            Boat or Ferry   0.1
3  Mediterranean Restaurant   0.1
4                    Beach   0.1


----Gaborone----
               venue  freq
0       Shopping Mall  0.13
1               Hotel  0.13
2                Café  0.13
3          Restaurant  0.07
4  Mexican Restaurant  0.03


----Harare----
                  venue  freq
0         Shopping Mall  0.13
1            Restaurant  0.13
2  Performing Arts Venue  0.07
3                 Hotel  0.07
4                  Café  0.07


----Juba----
               venue  freq
0          Restaurant  0.33
1               Hotel  0.33
2       Grocery Store  0.17
3                Café  0.17
4  African Restaurant  0.00


----Kampala----
          venue  freq
0          Café  0.30
```

```
1    Coffee Shop  0.13

2          Hotel  0.10
3            Bar  0.07
4  Shopping Mall  0.03


----Khartoum----
          venue  freq
0          Café   0.1
1     Juice Bar   0.1
2         Hotel   0.1
3  Dessert Shop   0.1
4    Restaurant   0.1


----Kigali----
                venue  freq
0               Hotel  0.17
1         Coffee Shop  0.17
2                Café  0.10
3  African Restaurant  0.07
4          Sports Bar  0.03


----Libreville----
                venue  freq
0              Bakery  0.17
1  Italian Restaurant  0.17
2               Hotel  0.11
3          Restaurant  0.11
4   Convenience Store  0.06


----Lilongwe----
               venue  freq
0              Hotel  0.26
1      Shopping Mall  0.26
2               Café  0.09
3  Convenience Store  0.04
```

```
4     Soccer Stadium  0.04


----Lomé----
            venue  freq
0           Hotel  0.22
1      Restaurant  0.11
2           Beach  0.06
3  Ice Cream Shop  0.06
4         Airport  0.06


----Luanda----
            venue  freq
0      Restaurant  0.17
1           Hotel  0.10
2     Pizza Place  0.10
3     Coffee Shop  0.07
4  Ice Cream Shop  0.07


----Lusaka----
           venue  freq
0          Hotel  0.20
1  Shopping Mall  0.10
2           Café  0.07
3     Restaurant  0.07
4     Steakhouse  0.07


----Malabo----
        venue  freq
0       Hotel  0.25
1  Restaurant  0.17
2        Café  0.08
3     Airport  0.08
4   Hotel Bar  0.08
```

```
----Maputo----

                 venue  freq
0                 Café  0.10
1                Hotel  0.10
2          Pizza Place  0.07
3   Italian Restaurant  0.07
4         Burger Joint  0.03


----Maseru----
              venue  freq
0     Shopping Mall  0.33
1   Border Crossing  0.17
2             Hotel  0.17
3        Steakhouse  0.17
4       Gas Station  0.17


----Mogadishu----
                 venue  freq
0                Hotel   0.6
1                 Port   0.2
2                Beach   0.2
3   African Restaurant   0.0
4          Music Venue   0.0


----Monrovia----
                 venue  freq
0           Restaurant  0.19
1        Grocery Store  0.12
2                Hotel  0.12
3   African Restaurant  0.06
4                  Bar  0.06


----Moroni----
    venue  freq
0   Hotel  0.33
```

```
1   Plaza  0.17

2  Market  0.17
3    Port  0.17
4  Resort  0.17


----Nairobi----
                venue  freq
0               Hotel  0.17
1         Coffee Shop  0.13
2  African Restaurant  0.10
3      Ice Cream Shop  0.07
4              Lounge  0.03


----Ndjamena----
                venue  freq
0               Hotel  0.36
1          Hotel Pool  0.14
2   French Restaurant  0.14
3  African Restaurant  0.07
4              Resort  0.07


----Niamey----
                venue  freq
0               Hotel  0.12
1  Italian Restaurant  0.12
2   French Restaurant  0.12
3         Supermarket  0.06
4             Airport  0.06


----Nouakchott----
                      venue  freq
0                     Hotel  0.22
1                      Café  0.17
2                Restaurant  0.17
3                    Bakery  0.06
```

```
4  Mediterranean Restaurant  0.06



----Ouagadougou----
                      venue  freq
0                     Hotel  0.20
1        African Restaurant  0.07
2               Coffee Shop  0.07
3                      Food  0.07
4  Middle Eastern Restaurant  0.07


----Port Louis----
                venue  freq
0        Shopping Mall  0.10
1                Hotel  0.10
2  Chinese Restaurant  0.10
3                 Café  0.10
4          Pizza Place  0.07


----Porto-Novo----
                 venue  freq
0  Fast Food Restaurant  0.25
1        History Museum  0.12
2         Track Stadium  0.12
3         Shopping Mall  0.12
4                 Plaza  0.12


----Pretoria----
                venue  freq
0         Coffee Shop  0.17
1      Farmers Market  0.07
2          Restaurant  0.07
3  African Restaurant  0.03
4                 Gym  0.03
```

```
----Rabat----

                venue  freq
0        Historic Site  0.17
1  Moroccan Restaurant  0.13
2                 Café  0.10
3     Tapas Restaurant  0.07
4                Hotel  0.07


----Tripoli----
               venue  freq
0               Café  0.37
1  Italian Restaurant  0.13
2              Plaza  0.07
3      Ice Cream Shop  0.07
4         Coffee Shop  0.07


----Tunis----
                     venue  freq
0                     Café  0.13
1               Restaurant  0.10
2  Mediterranean Restaurant  0.10
3                    Plaza  0.07
4                    Hotel  0.07


----Victoria----
              venue  freq
0            Resort  0.27
1             Beach  0.10
2  French Restaurant  0.07
3               Bar  0.07
4             Hotel  0.07


----Windhoek----
               venue  freq
0              Hotel  0.13
```

```
1          Restaurant  0.10

2        Shopping Mall  0.10
3   Italian Restaurant  0.07
4                 Café  0.07


----Yaounde----
                 venue  freq
0               Bakery  0.13
1                Hotel  0.10
2           Restaurant  0.10
3   African Restaurant  0.07
4               Lounge  0.07
```

In [291]:
```python
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [292]:
```python
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['city']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicator
s[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
cities_venues_sorted = pd.DataFrame(columns=columns)
```

```python
cities_venues_sorted['city'] = africa_grouped['city']

for ind in np.arange(africa_grouped.shape[0]):
    cities_venues_sorted.iloc[ind, 1:] = return_most_common_venues(afri
ca_grouped.iloc[ind, :], num_top_venues)

cities_venues_sorted.head()
```

Out[292]:

	city	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 Mos Commo Venu
0	Abuja	Restaurant	Movie Theater	Arcade	Chinese Restaurant	BBQ Joint	Hotel	Frie Chicke Joi
1	Accra	Hotel	African Restaurant	Pizza Place	Pub	Music Venue	Modern European Restaurant	Loung
2	Addis Ababa	Hotel	Italian Restaurant	Café	Nightclub	Restaurant	Coffee Shop	Sp
3	Algiers	French Restaurant	Hotel	Burger Joint	Restaurant	Diner	Lounge	Steakhous
4	Antananarivo	Hotel	Restaurant	African Restaurant	French Restaurant	Hostel	Mediterranean Restaurant	Sandwic Plac

In [293]:

```python
africa_grouped.head()
```

Out[293]:

	city	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Ent
0	Abuja	0.000000	0.0	0.0	0.033333	0.066667	0.0	0.000000	0.0	
1	Accra	0.100000	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	
2	Addis Ababa	0.000000	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	

	city	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Ent
3	Algiers	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.033333	0.0	
4	Antananarivo	0.066667	0.0	0.0	0.033333	0.000000	0.0	0.000000	0.0	

In [303]:

```
africa_grouped_clustering1.head(1)
```

Out[303]:

	Per_capita_Income	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Art Craft Stor
0	14069.295268	0.0	0.0	0.0	0.033333	0.066667	0.0	0.0	0.

In [348]:

```
# set number of clusters
kclusters = 6

africa_grouped_clustering1 = africa_grouped.drop('city', 1)
africa_grouped_clustering1.insert(0, 'Per_capita_Income', africa_data[
'per_cap_inc'])
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(africa_groupe
d_clustering1)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[348]: array([0, 0, 0, 4, 1, 1, 4, 2, 3, 1])

In [349]:

```
cities_venues_sorted.head()
```

Out[349]:

	Cluster Labels	city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	1	Abuja	Restaurant	Movie Theater	Arcade	Chinese Restaurant	BBQ Joint	Hotel
1	1	Accra	Hotel	African Restaurant	Pizza Place	Pub	Music Venue	Modern European Restaurant
2	1	Addis Ababa	Hotel	Italian Restaurant	Café	Nightclub	Restaurant	Coffee Shop
3	5	Algiers	French Restaurant	Hotel	Burger Joint	Restaurant	Diner	Lounge
4	0	Antananarivo	Hotel	Restaurant	African Restaurant	French Restaurant	Hostel	Mediterranean Restaurant

In [350]:

```
africa_merged.head(1)
#['Cluster Labels']
```

Out[350]:

	city	lat	lng	capital	Country	per_cap_inc	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Com V
0	Pretoria	-25.7069	28.2294	primary	South Africa	14069.295268	4	Coffee Shop	Farmers Market	Resta

In [351]:

```
# add clustering labels
#cities_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
cities_venues_sorted['Cluster Labels']= kmeans.labels_
africa_merged = africa_data

# merge africa_grouped with africa_data to add latitude/longitude for e
ach city
```

```
africa_merged = africa_merged.join(cities_venues_sorted.set_index('cit
y'), on='city', how='inner')

africa_merged.head() # check the last columns!
```

Out[351]:

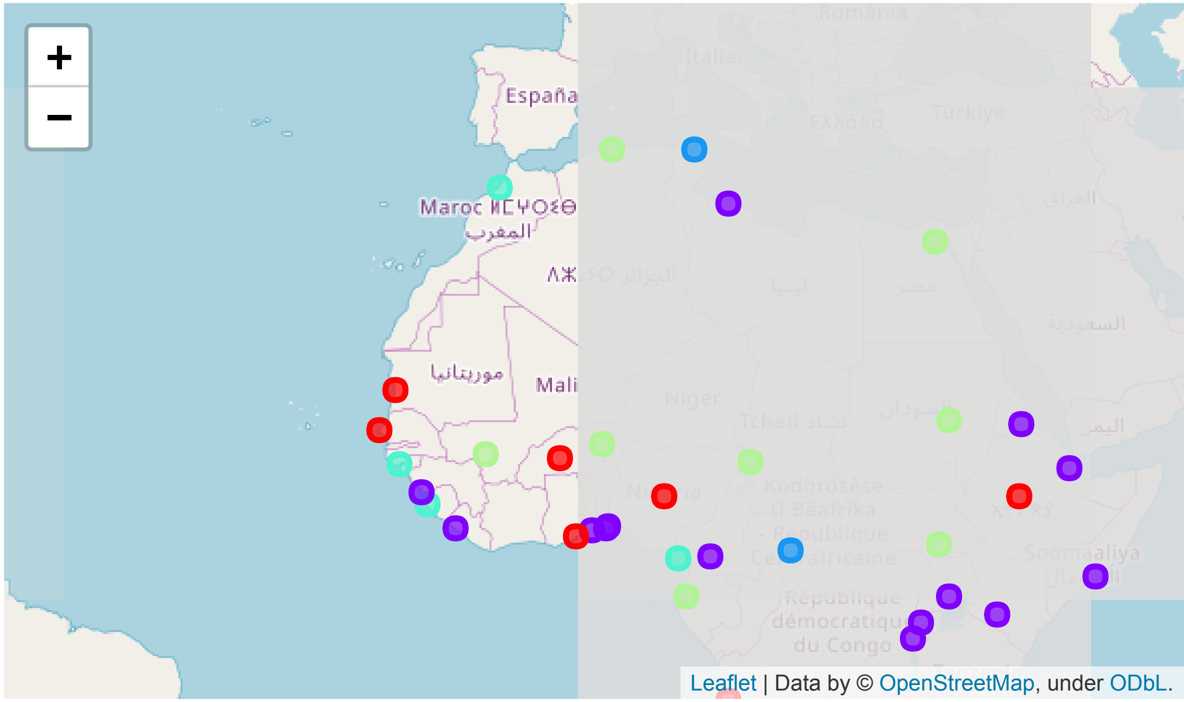
	city	lat	lng	capital	Country	per_cap_inc	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	Pretoria	-25.7069	28.2294	primary	South Africa	14069.295268	5	Coffee Shop	Farmers Market
1	Bloemfontein	-29.1200	26.2299	primary	South Africa	14069.295268	1	Coffee Shop	Hotel
2	Cape Town	-33.9200	18.4350	primary	South Africa	14069.295268	5	Coffee Shop	Hotel
3	Lusaka	-15.4166	28.2833	primary	Zambia	4409.940508	1	Hotel	Shopping Mall
4	Harare	-17.8178	31.0447	primary	Zimbabwe	2463.154492	2	Shopping Mall	Restaurant

In [352]:

```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=3)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(africa_merged['lat'], africa_merged[
'lng'], africa_merged['city'], africa_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_h
tml=True)
    folium.CircleMarker(
```

```
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[352]:



We also examine the average to gain and Idea of the clusters

In [353]:

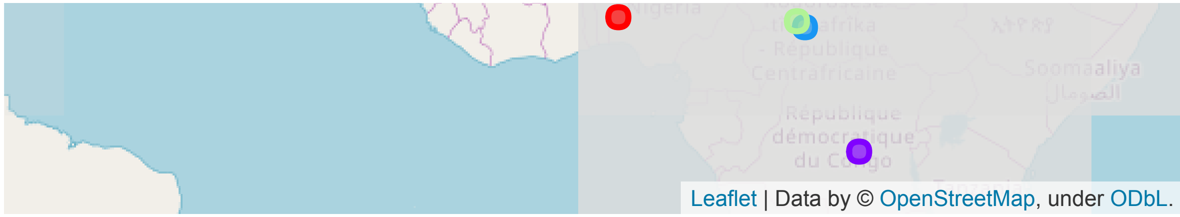
```
africa_c_groups = africa_merged.groupby('Cluster Labels').mean()
africa_c_groups = africa_c_groups.reset_index()
```

In [354]:

```
# create map
```

```python
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=3)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(africa_c_groups['lat'], africa_c_grou
ps['lng'], africa_c_groups['per_cap_inc'], africa_c_groups['Cluster Lab
els']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_h
tml=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[354]:



Results and Discussion

The k means algorithm created seven clusters, with analysis it appears the clusters can be describe as follows:

1. Spanish Atlantic Coastal Cities- Morocco, Malabo, Freetown and Bissau
2. Ancient African trading Civilasations- Ethiopia, Nigeria, Angola, Ghana, Senegal and Burkina Faso
3. Desert regions with high incomes and small populations- Namibia, Chad, Egypt, Sudan, South Sudan, Niger, Mali and Gabon
4. High Income modern societies- Pretoria, Cape Town and Mauritius
5. Middle Income States mostly in ex British Africa-
6. Crisis States- Tunisia, Zimbabwe and the Central African Republic

It is worth noting they are few anomalies in thes classifications:

1. Freetown was not a spanish colony, and it has lower incomes the its group members, but it has been clustered with these cities. It would be intresting to discover these lingkages.
2. Tunisia is a high-income democracy unlike CAR and Zimbabwe

For the other groups the classification was nearly perfect considering the history, geography and politics of the regions.

Conclusion

The purpose of this project was to classify similar african cities for the purpose of tourism and investment. Thhis has been achieved through K-means clustering although a few anomalies exist, most clusteer hold true to reality.

```
In [ ]:
```