

## Table of Contents

<u>Capstone Proposal</u> .....	1
Data Used .....	1-4

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

# Capstone Proposal

## 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/](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 Istanbul	Turkish Restaurant
1	Algeria	Algiers	CC	Best Night Hotel	Hotel
2	Algeria	Algiers	CC	Salad box Algeria	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)

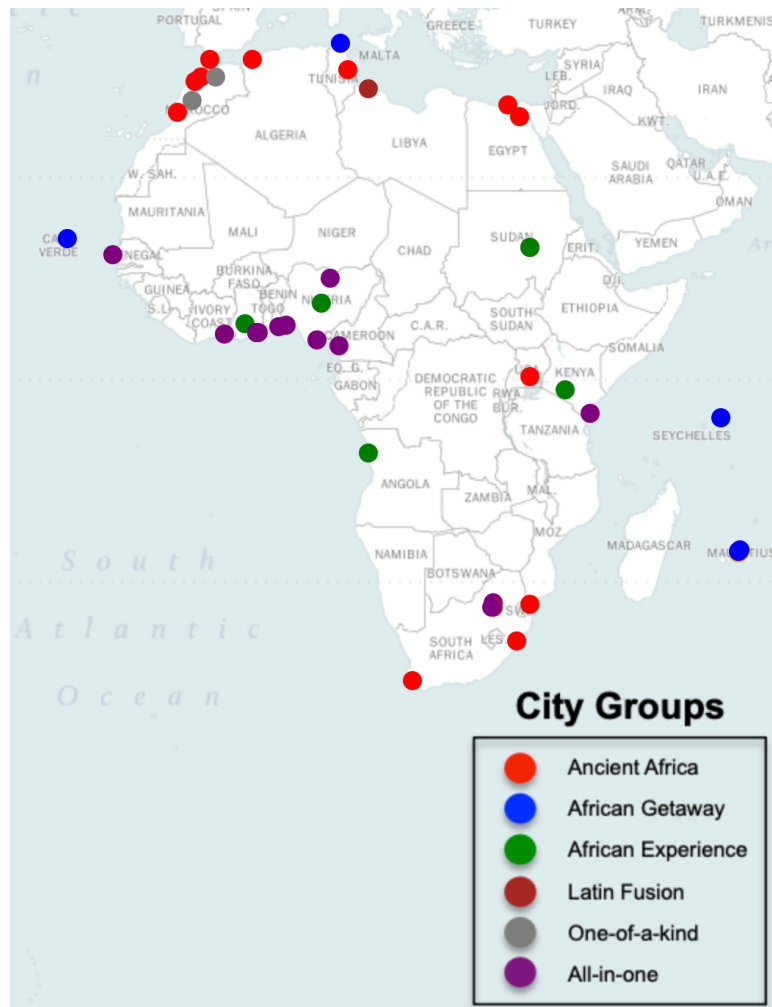
The dummied version of the above table will then be used to cluster the cities according to the venues they have in common. For example:

```
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.