

Description of the Data



In order to start our project, we are going to need information about the moroccan market (the market of cafes) we will need :

- Neighborhoods in some moroccan cities
- Latitude and longitude of each neighborhood
- Number of cafes per neighborhoud
- Habitation per neighborhoud
- Fousquare location data

Our problem here is to choose the appropriate place or neighborhood to invest in a cafe in order to guarantee a good profitability, so to settle this problem, we will start by :

1. getting all the neighborhoods in the following morrocan cities :

- casablanca

from : <https://fr.wikipedia.org/wiki/Casablanca#Quartiers>

- Rabat

from : <https://fr.wikipedia.org/wiki/Rabat#Quartiers>

- Marrakech

From : <https://fr.wikipedia.org/wiki/Marrakech#Quartiers>

- Tangier

From : <https://fr.wikipedia.org/wiki/Tanger#Quartiers>

2. get the coordinates (latitude and longitude) of each neighborhood (from <https://www.latlong.net/>)

we will get a result dataframe like the following :

[6]:

| | ville | quartier | Latitude | Longitude |
|---|------------|---------------|-----------|-----------|
| 0 | Casablanca | Aïn Chock | 33.532720 | -7.617050 |
| 1 | Casablanca | Aïn Sebaâ | 33.601560 | -7.545370 |
| 2 | Casablanca | Hay Mohammadi | 33.583580 | -7.561269 |
| 3 | Casablanca | Roches Noires | 33.599420 | -7.585870 |
| 4 | Casablanca | Anfa | 33.594150 | -7.646590 |
| 5 | Casablanca | Maârif | 33.580260 | -7.630220 |



3. Use the Foursquare location data to know the most visited venues and places per neighborhood and class them.

We will get the dataframe that look like :

[52]:

| | name | categories | lat | lng |
|---|--------------------|-------------|-----------|-----------|
| 0 | Les Quatre Cerises | Bakery | 33.534528 | -7.611007 |
| 1 | Idouar Ice | Coffee Shop | 33.533901 | -7.612866 |
| 2 | Café ANTIK | Coffee Shop | 33.532062 | -7.616460 |
| 3 | Fragola Gelato | Coffee Shop | 33.535425 | -7.609501 |
| 4 | West Village Café | Café | 33.533885 | -7.614602 |

4. use Clustering to divide these areas by clusters depending on their most common visited places, and choose the cluster or clusters that contain cafes in first positions.
5. Use the data of the number of cafes and habitation per neighborhood to calculate the correlation between these variables in order to focus on places where we can do our investment

From :

https://www.huffpostmaghreb.com/2017/05/29/casablanca-5e-ville-plus-densement-peuplee-monde_n_16865426.html

