# Analyzing the Townships in Algiers City

**Samia BELHADDAD**
**IBM Data Science Professional Certificate Applied project**

**May 21, 2020**

## 1. Introduction

The final project of the IBM DS Professional Certificate is an applied capstone that allows students to apply the skills they learnt in 8 earlier courses.

The main purpose of this project is to apply Data Science Methodology ,Python programming skills ,and Data Analysis to solve real world problems. IBM only provides a guideline to execute the project. Learners are responsible for finding a business problem and creating an analytics approach to solve that problem and get insights from data.

In this project, Learners are required to leverage Foursquare location data to explore cities of their choice .

## 2. Business and Problem Understanding:

Algiers city is the economic and political capital of Algeria, It is a large city with 57 districts. This project aims To boost economic development in Algiers.Business owners need to figure out the best place to locate their businesses based on competition and business needs in an area of the city. The project aims to answer the following questions:

- What is the best place to locate a business in Algiers?
- What are the most common categories of venues in Algiers?
- Which area in Algiers city is the best for launching a new business?

## 2.Methodology

### a. Data understanding and collection:

The needed data is:

Location data of the townships of the city of Algiers: Scraping Wikipedia page of Algiers province to get the townships of Algiers using BeuatifulSoup API

Link: https://en.wikipedia.org/wiki/Algiers_Province

Latitude and longitude data of each township: Using geopy API to get the latitude and longitude if each township in Algiers.

Venues data : Getting the venues related to location coordinates using Foursquare API

Link: https://fr.foursquare.com/developers/apps

**b. Data preprocessing:**

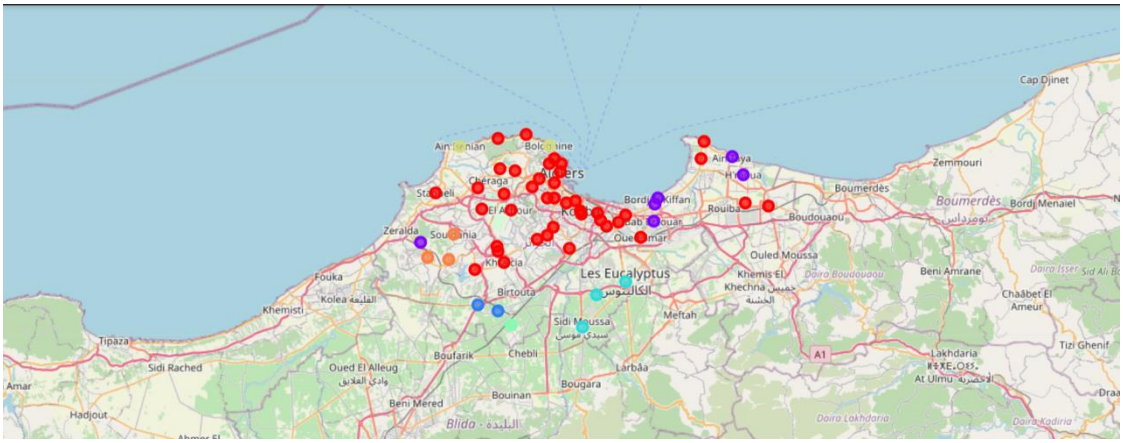
Some of the townships has to be modified manually because of the confusion they create when trying to get the location coordinates(names similarity with other irrelevant locations)

The data that comes from the different sources needs to be put in one dataframe . so the dataframes were merged to one.

In order to apply the Kmeans Algorithm, we use onehot encoding for algiers venues categories to convert them to numerical data.

**c.** The use of Kmeans algorithm for segmentation and clustering of the city based on venues in each neiborhood

## 3. Result :




Visualizing the final result of clusters using folium marked map that shows 7 different clusters based on the most common venues in the city,
The main venues in Algiers are basic residential and commercial venues such as hotels, restaurants and shops .
The townships differ by some few venues .

## 4. Recommendation

The townships differ by some few venues and thus, every township is a potential business location as soon as the business is not a restaurant, a hotel or a grocery shop!

## 5. Conclusion

After analyzing the clusters of the city of Algiers:

-The clusters were not able to be disjoint as the venues are mainly similar to eachother in almost every neiborhood.

-Algiers City is a residential city more than being a commercial one or a touristic one

-Only the basic services are available in every neiborhood .

-the most common venues are hotels,restaurants, grocery stores,