A.2. Data Description

To solve the problem, we will need the following data:

• **List of districts in Oran**. This defines the scope of this project.

• **Latitude and longitude coordinates** of those districts. This is required in order to plot the map and also to get the venue data.

• **Venue data**, particularly data related to restaurants. We will use this data to perform clustering on the districts.

Source of the data and method to extract them:

This website page (https://journals.openedition.org/droitcultures/28) contains a list of districts name in Oran, I've chosen this page because it contains the old and new denominations for those districts to avoid duplicate with other commune name when searching the geographical coordinates. We will use web scraping techniques to extract the data from the website page, with the help of Python requests and beautifulsoup packages.

Then we will get the geographical coordinates of the districts using Google Geocoder API which will give us the latitude and longitude coordinates.

After that we will use Foursquare API to get the venue data for the districts in Oran, I personally will have used Google places API instead of Foursquare API, because of the number of users in Algeria on the Google platform is more interesting to get the maximum of venues, but I'll follow the request of the stakeholder and use foursquare API. Foursquare API will provide many categories of the venue data, we are particularly interested Restaurants category in order to help us to solve our business problem.