

DATA INVENTORY – DESCRIPTION - SOURCES

DATA INVENTORY AND DESCRIPTION

To achieve our goal thru this capstone assignment, we should have the following data:

- The geospatial coordinates of Paris (France)
- The order and the name of the 20 Paris Districts.
- The coordinates of a location in each district: we can take here the well-known coordinates of the Hall of the City in each district.
- The number of inhabitants in each district.
- The area of each district

In addition, we need to get the existing venues in each district. For each venue, we will need to know the category of the venue and their coordinates for our analysis. We will limit the number of venues to 150 and we will cover on 1,500 meters around of the location point.



After loading, cleaning and transforming the data, we will collect all the venues categories in each district and try to use a machine learning algorithm to group districts in some clusters presenting the same characteristics. From this result, we will try to respond to the business problem asked for the assignment.

DATA SOURCES

About the data related to the 20 Paris districts, Wikipedia provides on the name, area, number of inhabitants, density for each district [1]. Specific information on each district can be collected also from this page [2].

To get the data about the venues, categories and their coordinates, we will use the Foursquare API [3] which is the best place to collect the venues locations in any town over the world. We will use the standard free subscription to achieve this exercise.

[1] https://fr.wikipedia.org/wiki/Arrondissements_de_Paris#R%C3%A9partition_de_la_population

[2] <https://www.unjourdeplusaparis.com/paris-essentiel/paris-par-arrondissements>

[3] <https://enterprise.foursquare.com/products/places>