

Description of the data

Data Sources

To consider the problem we can list the data as below:

- Kampala Neighborhoods

Name, latitude and longitude coordinates were scrapped from Wikipedia

- Venues of every neighborhood

Foursquare API was used to identify Chinese restaurants

As a database, I used GitHub repository in my study. My master data, which has the main components *Restaurants*, *Latitude* and *Longitude* information of the city.

Data Cleaning and Wrangling

First, I scrapped Wikipedia to extract the name of every neighborhood. I got a data frame and I deleted those columns that were not important.

Then it was necessary to collect all the links from the same web of every neighborhood in order to obtain the coordinates (latitude and longitude).

Finally, I created a dataset (CSV file) with the population of every neighborhood from a website and then I imported this set and performed a new data frame. Then I merged this data frame with the main data frame (Previous data frame).

I used the Foursquare API to obtain the venues, check which neighborhood has a Chinese restaurant and which does not. I also used a folium library to obtain a map that marks the selected neighborhoods and a prominent mark in the neighborhoods with a Chinese restaurant.