

Data Requirements

Data requirements and Sources:

- A list of Suburbs in Colombo Sri Lanka from https://en.wikipedia.org/wiki/Category:Suburbs_of_Colombo.
- Latitude and longitude values for the suburbs scraped from the website. These will be obtained using the geopy library to geocode the scraped locations.
- Venue data will be obtained in order to cluster the locations (Suburbs) of Colombo based on venue categories.

Data Use:

- Colombo is a relatively small city therefore it important to capture as many smaller subunits as possible and understand their venue compositions thoroughly. Therefore, we scrape this page.
- The data is geocoded since foursquare needs to be provided a latitude and longitude along with a radius in order to look for an 'n' number of locations in said locality.
- The four square API's explore endpoint will be used to obtain as much venue data as possible in these vicinities in order to do a satisfactory clustering and understand clusters of areas that are busy with people who are looking to grab a bite, but also where the market is already not over-saturated with fast casual and casual dining restaurants making it difficult to succeed after opening planned pizzeria location.