Approach to address the Problem Statement

Sri Lanka is divided into 25 districts geographically. Each district is comprised of several Divisional Secretariats (DS Divisions) which are demarcated geographically. Colombo city lies in Colombo District itself. There is one more district very close to Colombo city, which is Gampaha district, as depicted in the below map. Therefore DS Divisions of these two districts (Colombo & Gampaha) will be considered to analyze.

Data for DS Divisions of two districts will be scraped from following Wikipedia sites:

https://en.wikipedia.org/wiki/Colombo_District

https://en.wikipedia.org/wiki/Gampaha_District

A data frame with features of DS Division, Area, Population, Population Density, and Latitude & Longitude will be created by scraping the data from above mention Wikipedia sites. Geo coordinates for each DS Division will be obtained using Nominatim geodecoding service.

Data visualization and Data Cleaning will be done according to fit into the requirements mentioned in the Introduction.

1. Extremely highly populated DS Divisions will be dropped if any.
2. DS Divisions which are more than 20km away from Colombo city will be dropped. (In Srilanka, it is quite common a person to travel 20km from his/her residence to work place daily basis.)

Foursquare API will be used to get the nearby venues for the DS Divisions and then DS Divisions will be clustered using K-Means clustering algorithm as per their nearby venues. Cluster with modern neighborhoods will be identified.

Finally DS Divisions with fairly modern neighborhoods with less populated and lies within reasonable distance to Colombo City will be recommended.