Chathusha Karunathilake

Finding lands to construct apartment buildings in colombo

# Introduction

## Background

Colombo is the commercial capital of Sri Lanka. The city is packed with office and commercial buildings and road traffic is a common scene. People, who commute daily to attend to their jobs are the ones who are mainly affected by this traffic.

Mainly due to this traffic issue, there is a high market for apartments in the Colombo suburbs among people with high income. Building developers have identified this market and currently, in Colombo, there are more than 5 high-rise apartment constructions that are ongoing and more at the planning stage. The market for apartments seems to be booming.

Further, building developers are searching for lands with marketable infrastructures at proximity to construct apartments in the future. By marketable infrastructure I mean, schools, roads, public transport facilities, supermarkets, shopping centers, etc.

## Problem

With current market for apartment building, building developers are looking for lands to purchase near marketable infrastructures to construct apartment buildings in near future. This project aims to provide some insights on the locations preferable to construct the apartment complexes based on nearby venue data.

## Interest

Mainly, building developers in Sri Lanka would be very interested in finding proper locations to construct apartment complexes. Also, buyers would be interested too.

# Data Acquisition and Cleaning

## Data Sources

* 1. Census data of Colombo district from Department of Census and Statistics, Sri Lanka.
  2. Venue details using Foursquare API.
  3. Latitude and longitude data of each neighborhood from Google Maps API.

## Data Cleaning

Census data of Colombo district from Department of Census and Statistics, Sri Lanka was used in the analysis. This data was used to identify the neighborhoods (we call, grama niladhari division) of Colombo area. However, the census data came as .pdf file. So, Tabula app was used to convert .pdf tables to .csv files. Some manual modifications were done to the .csv file to make it analyze friendly.

Then, after acquiring location data of each neighborhood, distance from the Colombo city central was calculated. For that I’ve used the geopy package. Then the locations further than the 15km from city center was removed from the analysis because people in those location will take more than 1 hour to reach to the Colombo city. Same as before step, I’ve also removed the locations closer than 5km to the city center because the land prices will be too much high.

## Data Acquisition

Latitudes and longitude data of each neighborhood had not included in the cleaned census data. So, I had to use Google Maps API to get the latitude and longitude of each location. For that I’ve passed the neighborhood name to Maps API and cleaned up the results. Then the latitude and longitude data were saved to a .csv file with neighborhood name for future analysis.

Foursquare API was used to get the nearby venue data. Nearby venue data was collected under following categories.

1. Road
2. Intersection
3. Bus station
4. Bus stop
5. Train station
6. Residence
7. Schools
8. Office
9. Medical center
10. Factory
11. Recreation

Residence, Office, and Factory venue data were mainly collected to get an understanding about the neighborhood, whether it is industrial, residential or city. Other venue categories were observed, because as I think those are the main features that a customer looking for near an apartment building.

# Modeling

I’ve used the K-Means clustering method to cluster each location based on nearby venues.

## K-Means Clustering

Firstly, I’ve used the clustering with K values from 1 to 15 to do the modeling. Then, I’ve plotted each K value against sum of squared error. From that I’ve found the best value to do the modeling is 3.

Chart, line chart

Description automatically generated

Then, the actual modeling is done with K=3, and cluster each neighborhood to 3 categories.

# Conclusion

From this exercise, I’ve identified 80 preferable locations to construct the apartment buildings out of 577 neighborhoods evaluated.

Those preferable locations are listed below,

|  |  |  |  |
| --- | --- | --- | --- |
| Sammanthranapura | Hokandara North | Sri Saranankara | Attidiya South |
| Halmulla | Hokandara East | Vilawala | Piriwena |
| Ambathale | Kumaragewatta | Dutugemunu | Wedikanda |
| Dahampura | Wickramasinghapura | Kalubovila | Vihara |
| Dahampura | Madiwela | Hathbodhiya | Kandawala |
| Malgama | Thalawathugoda West | Galwala | Angulana North |
| Maligagodella | Thalawathugoda East | Dehiwala West | Kaldemulla |
| Rajasinghagama | Liyanagoda | Dehiwala East | Thelawala South |
| Udumulla South | Kottawa North | Udyanaya | Lakshapathiya North |
| Elhena | Polwatta | Nedimala | Lakshapathiya Central |
| Dodamgahahena | Pamunuwa | Malwatta | Angulana South |
| Kaduwela | Thalapathpitiya | Jayathilaka | Katubedda |
| Kothalawala | Pragathipura | Karagampitiya | Divulpitiya West |
| Thalahena North | Pannipitiya North | Kawdana East | Bellanvila |
| Malabe North | Wijerama | Mount Lavinia | Boralesgamuwa West A |
| Pore | Kirulapone | Kawdana West | Boralesgamuwa West C |
| Malabe East | Wellawatta North | Watarappala | Rattanapitiya |
| Thalahena South | Wellawatta South | Wathumulla | Egodawatta |
| Muttettugoda | Pamankada West | Katukurunduwatta | Boralesgamuwa East A |
| Kalapaluwawa | Pamankada East | Attidiya North | Boralesgamuwa East B |