ANALYSIS OF OUTER SUBURBS AROUND COLOMBO CONCERNING LAND PRICES AND DENSITY OF SOCIAL PLACES

Yasiru Rangana

Contents

Introduction

Colombo is the commercial capital of Sri Lanka. It is also the administrative capital of the Western Province and the district capital of Colombo District [1]. Colombo is the financial center of the island and major hub for transport, healthcare, and education. It is located on the west coast of the island in between Kalutara and Gampaha districts.

When compared with the rest of Sri Lanka, Colombo metropolitan area is home to 5.6 million it is 28 percent of Sri Lanka's population and accounts for about 45 percent of national GDP. Moreover, it dominates 80 percent of industrial value in Sri Lanka [2]. It is the country's major urban agglomeration and is growing faster than any other area in Sri Lanka. Because of Colombo is the main urban area in Sri Lanka, hundreds of thousands of people migrate to Colombo and nearby areas each year for several reasons. According to the department of senses and statistics, more than 31 percent of the migrant population in Sri Lanka, migrate to Colombo and Gampaha which is an adjacent district of Colombo [Figure:1] [3]. It is more than five times higher than the people who migrate to Kandy which is the second major city in Sri Lanka. However, the Colombo metropolitan area covers only about 6 percent of the country's total land area [2]. Because of higher demands, real estate prices are skyrocketing each year. Therefore, it is extremely difficult to find a better suburb to buy land for a migrant. They want to find areas where real estate values are lower. at the same time, they want to find areas that have a higher density of social places according to their interests. The objective of this study is to cluster outer suburbs around Colombo concerning land prices and density of social places which will enable migrants to identify suburbs with similar price bands and socially popular places.

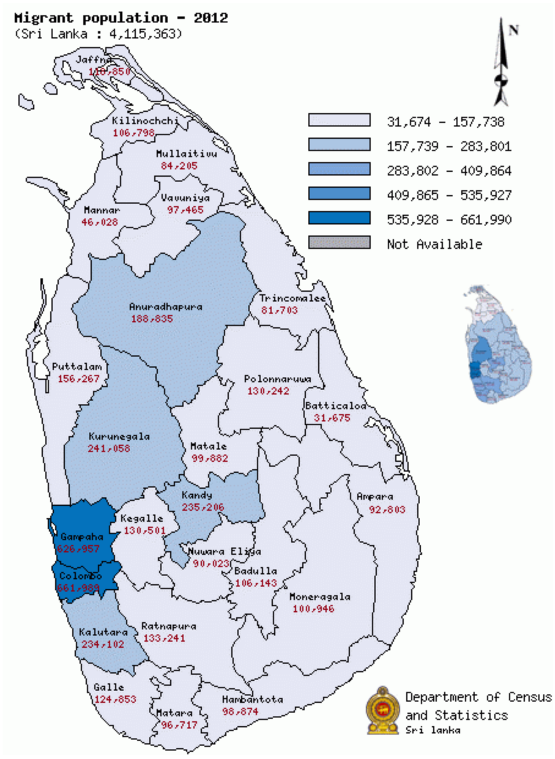

Figure 1: Migrant Population of Sri Lanka

Data Section

- A list of outer suburbs around Colombo was obtained from Wikipedia article: Colombo [4]. For this study, inner-city suburbs and capital zone suburbs in Colombo are not used since they are more suitable for business purposes rather than residential purposes.

- Geo-Coordinates for each suburb were retrieved from ArcGIS geocode library for Python [5].

- The most common places for a given suburb were acquired from the Foursquare API [6].

- A list of average land prices per perch for the above suburbs is available on lankapropertyweb [7].

- BeautifulSoup web scraping Python library and Pandas library was used to scrape data from the above web pages [8].

Methodology

Data gathering and wrangling

1. Suburbs

First, the list of outer suburbs around Colombo was scraped from Wikipedia page using BeautifulSoup web scraping python library. For this study, inner-city suburbs and capital zone suburbs in Colombo are not used since they are more suitable for business purposes rather than residential purposes.

```
url='https://en.m.wikipedia.org/wiki/Colombo'
page = requests.get(url)

soup = BeautifulSoup(page.content, 'html.parser')

listSuburbs = []
div = soup.find_all('div', class_='div-col')
for ul in div:
    for li in ul.findAll('li'):
        for a in li:
            title = a.string
            listSuburbs.append(title)

suburbs = pd.DataFrame(listSuburbs,columns=['Name'])

print(suburbs.shape)

suburbs.head()
```

```
(35, 1)
```

	Name
0	Angoda
1	Athurugiriya
2	Battaramulla
3	Biyagama
4	Boralesgamuwa

Figure 2: scrapping wikipedia page

There are thirty-five popular outer suburbs around Colombo according to this data.

2. Geo-Coordinates

ArcGIS geocode library for Python was used to obtain geo-coordinates for the above suburbs.

```
def get_geocordination(city):
    location = geocoder.arcgis('{},  COLOMBO, LK'.format(city))
    lat_lng_coords = location.latlng
    latitude = lat_lng_coords[0]
    longitude = lat_lng_coords[1]
    print('{},{},{}'.format(city,latitude,longitude))
    return latitude,longitude

suburbs['Latitude'],suburbs['Longitude'] = zip(*suburbs['Name'].apply(get_geocordination))
```

Figure 3: use of ArcGIS to get Geo-coordinates

After retrieving latitude and longitude for centroids in the above suburbs, the data frame of suburbs was shown as follows. Only the first few rows are displayed.

	Name	Latitude	Longitude
0	Angoda	6.935934	79.925654
1	Athurugiriya	6.877397	79.989996
2	Battaramulla	6.902228	79.919574
3	Biyagama	6.946234	79.991284
4	Boralesgamuwa	6.841013	79.901718
5	Dehiwala	6.851325	79.865995
6	Gothatuwa	6.929570	79.906005
7	Hokandara	6.880341	79.959809
8	Ja-Ela	7.077637	79.891086
9	Homagama	6.841217	80.003118
10	Kadawatha	7.001477	79.950466

Figure 4: head of suburbs data frame

Using the folium library above suburbs can be visualized on the Sri Lankan map as displayed in figure 5.

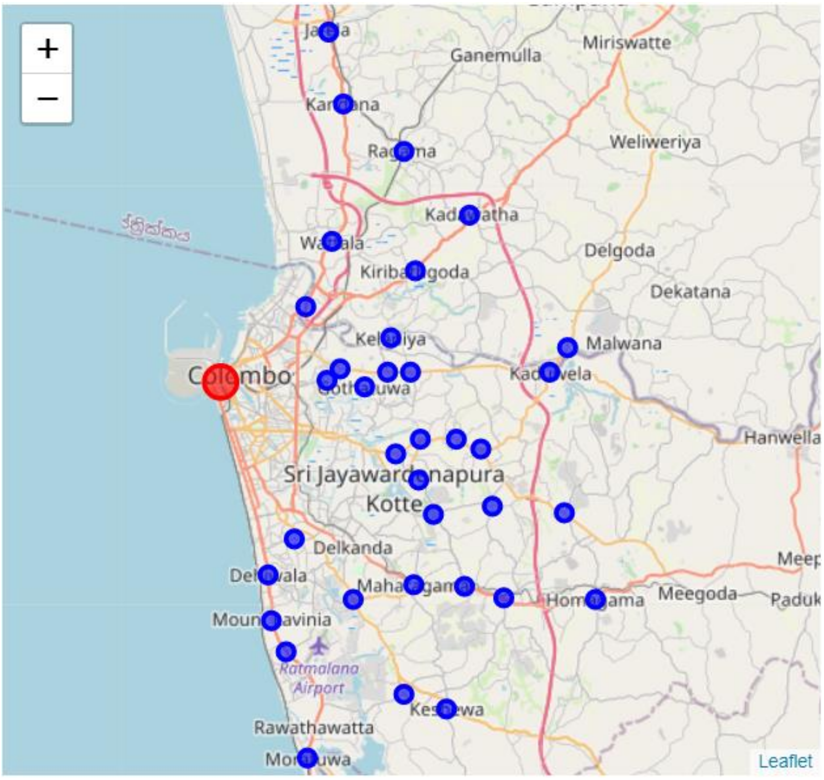

Figure 5: map of selected outer-Colombo suburbs

3. Land Prices

The list of average land prices per perch was obtained from Lanka property web and columns were renamed to be meaningful.

```
landPricesWEB = pd.read_html('https://www.lankapropertyweb.com/house_prices.php')
landPrices = landPricesWEB[4].rename(columns={'Unnamed: 0': 'Name',
                                              'Average Price (Per Perch)': 'Price'})
landPrices['Price'] = landPrices['Price'].str.replace(r"\(.*\)","").str.replace(",","")
print(landPrices.shape)
landPrices.head()
```

Figure 6: scrapping web page to get land prices

There are 49 suburbs with average land prices in Sri Lankan rupees (LKR). Our list of suburbs had only 35 suburbs.

```
(49, 2)
```

	Name	Price
0	Angoda	625972
1	Athurugiriya	484436
2	Bandaragama	197576
3	Battaramulla	1880478
4	Biyagama	521190

Figure 7: shape & head of landPrices data frame

4. Merging

The average land prices retrieved above were merged into the suburbs data frame. At this point, it was noticed that 'KOSWATTE' and 'KOTIKAWATTA' suburbs do not have land prices. Even though it was easy to drop those two suburbs, it was decided to impute those two suburbs with an average price of the nearest three suburbs.

	Name	Latitude	Longitude	Price
15	Kesbewa	6.795586	79.940721	490000
16	Kiribathgoda	6.978221	79.927310	918511
17	Kolonnawa	6.932621	79.890414	945000
18	Koswatte	6.908071	79.929538	NaN
19	Kotikawatta	6.936024	79.915914	NaN
20	Kottawa	6.841609	79.964486	783554
21	Maharagama	6.847323	79.926728	1296439
22	Malabe	6.903965	79.955072	976897
23	Moratuwa	6.775091	79.882487	961883
24	Mount Lavinia	6.832674	79.867385	2635636

Figure 8: suburbs data frame with average prices

5. Imputing missing values

To impute the average price of a given suburb using the average price of the nearest k number of suburbs, python function was implemented.

```
def imputePrice(Name,k):

    a = distance.cdist(suburbs.loc[suburbs['Name'] == Name][['Longitude', 'Latitude']],
                       suburbs[['Longitude', 'Latitude']], 'cityblock')
    ind=([])

    for i in range(k):
      ind.append(np.argpartition(a, i+1)[0][i+1])

    print('Nearest '+ str(k) +' Suburbs for ' + Name + ' are ')
    print(suburbs.iloc[ind])

    average = suburbs.iloc[ind]['Price'].astype(int).mean().astype(int)

    print('imputed price : ' + str(average))

    return average
```

Figure 9: python function to impute price averaging k number of nearest suburbs

Prices of 'KOSWATTE' and 'KOTIKAWATTA' were calculated using 'imputePrice' function and added to the suburbs data frame.

	Name	Latitude	Longitude	Price
18	Koswatte	6.908071	79.929538	1905159
19	Kotikawatta	6.936024	79.915914	733935

Figure 10: imputed prices for 'KOSWATTE' & 'KOTIKAWATTA'

6. Get nearby locations

Nearby common places were acquired using the Foursquare API. For that, a free developer account was created on the foursquare developer portal. For the simplicity of this study nearest 20 places around 3km perimeter were queried. It was noticed that some locations were categorized into different category names even though they were the same. This may happen as Foursquare collects data from the public. As an example, some train stations were labeled as light rail stations and some as metro stations. After merging those locations manually there were 91 unique location types across all suburbs.

Analysis

After acquiring categories of most common places using foursquare API, they were listed against each suburb using one-hot encoding. Moreover, the average land price of each suburb also listed against each suburb. In order to cluster suburbs using the above data, the whole data set was normalized.

K-means algorithm was used to cluster suburbs since it is the most promising unsupervised clustering technology suitable to this type of data set. Optimal K value was identified as 4 after doing the Elbow test.

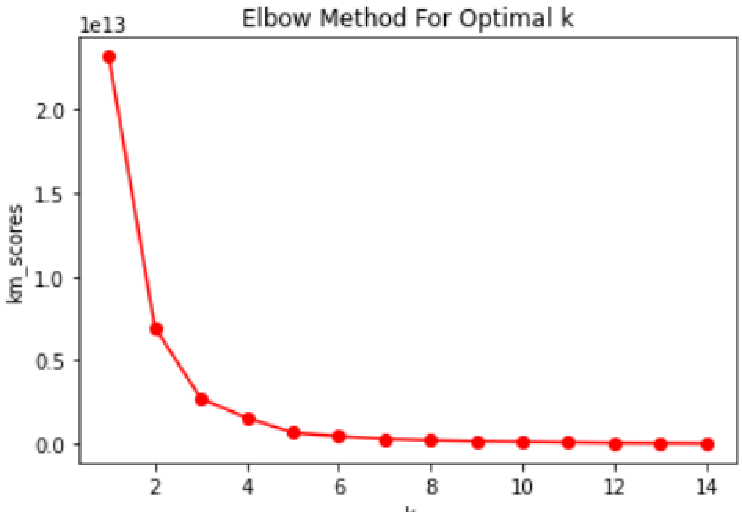

Figure 11: Plot of Elbow Method for Optimal K

Figure 12 shows suburbs merged with cluster labels after K-means clustering. Only the first seven suburbs with the two most common venues are shown in the figure.

	Name	Latitude	Longitude	Price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	Angoda	6.935934	79.925654	625972	1	Bus Stop	Shopping Mall
1	Athurugiriya	6.877397	79.989996	484436	1	Shopping Mall	Gym
2	Battaramulla	6.902228	79.919574	1880478	0	Bakery	Gym
3	Biyagama	6.946234	79.991284	521190	1	Gym	Shopping Mall
4	Boralesgamuwa	6.841013	79.901718	1554242	0	Department Store	Shopping Mall
5	Dehiwala	6.851325	79.865995	4008277	2	Cosmetics Shop	Pizza Place
6	Gothatuwa	6.929570	79.906005	945000	3	Shopping Mall	Italian Restaurant

Figure 12: suburbs with cluster labels

Results

Clusters of suburbs based on land prices and most common venues can be visualized as the following figure.

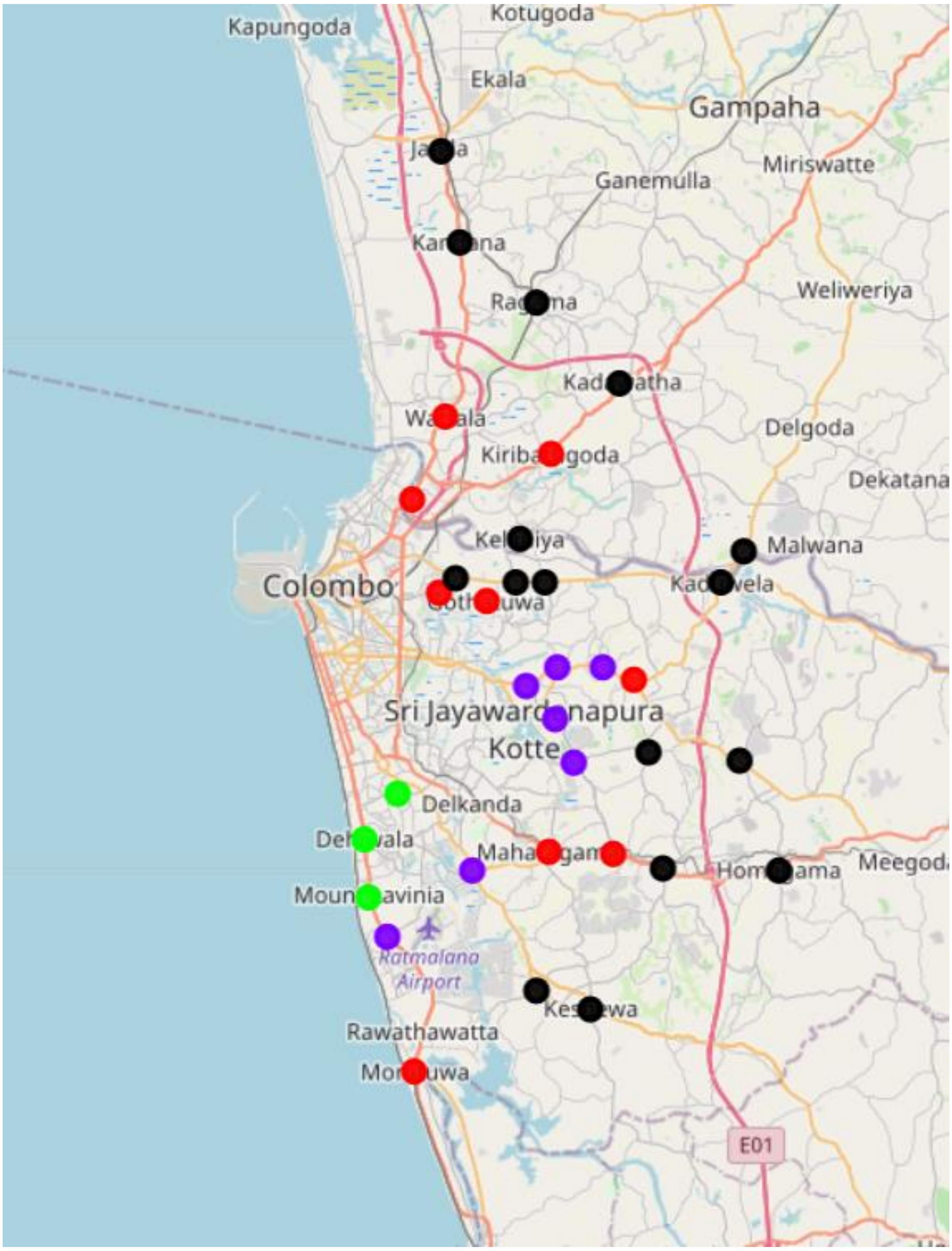

Figure 13: Visualised map with clusters

The summary of the result of this study was shown in a tabular form as follows.

CLUSTER	SUBURBS	AVERAGE LAND PRICE (LKR)
1	Angoda, Athurugiriya, Biyagama, Hokandara, Ja-Ela, Homagama, Kadawatha, Kaduwela, Kandana, Kelaniya, Kesbewa, Kotikawatta, Kottawa, Piliyandala, Ragama, Wellampitiya	579,970.81
2	Gothatuwa, Kiribathgoda, Kolonnawa, Maharagama, Malabe, Moratuwa, Pannipitiya, Peliyagoda, Wattala	1,065,355.33
3	Battaramulla, Boralesgamuwa, Koswatte, Pelawatte, Rathmalana, Thalahena, Thalawathugoda	1,845,639.43
4	Dehiwala, Kalubowila, Mount Lavinia	3,235,008.00

Discussion

In this study, I clustered outer suburbs around Colombo which is the commercial capital of Sri Lanka. I considered land prices and density of social places to identify suburbs with similar price bands and socially popular places. This study will be useful to a migrant who wants to find similar suburbs based on the above measures.

I could identify 4 clusters of suburbs based on their average land price and common venues. According to the results of this study, we can clearly see that cluster one which is colored as black has the lowest land prices and it has suburbs far away from the Colombo city and 'Sri Jayawardhanapura Kotte' which is the capital of Sri Lanka.

At the same time 'Dehiwala', 'Kalubowila', and 'Mount Lavinia' suburbs are clustered together as cluster 4 with the highest land prices. Office locations, beaches, pubs, restaurants, and departmental stores were the most common venues in those suburbs.

Cluster 2 and cluster 3 have low-mid level and high-mid level land prices respectively. Moreover, cluster 2 has common places like clothing stores, shopping malls, and fast-food restaurants. Cluster 3 is popular for gyms, cocktail bars, juice bars, bakeries, coffee shops, and book stores.

Conclusion and Future Studies

As an emerging city in south Asia, Colombo is becoming a challenging location to buy a property. Since the lack of location-based data and availability of land prices, these types of studies had never been done based on any city in Sri Lanka. As a start-up, this study was focused on a few available suburbs. This study will be important to a migrant who wants to find similar suburbs based on land prices and surrounding venue types.

As an example, a person with a lower budget can focus on lands around Piliyandala, Kesbewa, Kottawa, Homagama, Kaduwela, Kadawatha, Kelaniya, Ragama and expect the same level of density of popular venues. At the same time, a person with a lower-middle budget can get the same level of density of popular venues in Moratuwa, Maharagama, Malabe, Kiribathgoda, and Wattala.

This study was done only based on average land prices and most common venues derived via foursquare API. But further studies can be done using other factors like ethnicity, crime density, air pollution, etc…. Moreover, web application or mobile application can be developed to assist migrants to find similar properties based on their interests.

References

1. Colombo
2. Overview of Sri Lanka by World Bank
3. Department of Census and Statistics
4. Outer Suburbs of Colombo
5. ArcGIS Geocoding
6. Foursquare API for Developers
7. Average Land perch prices in Western Province
8. Beautiful Soup Documentation