ANALYSIS OF OUTER SUBURBS AROUND COLOMBO

Based On Land Prices And Density of Social Places

By Yasiru Rangana

Introduction

- Colombo is the commercial capital of Sri Lanka.
- It is home to 5.6 million it is 28 percent of Sri Lanka’s population and accounts for about 45 percent of national GDP.
- It covers only about 6 percent of the country’s total land area.
- Therefore, it is extremely difficult to find a better suburb to buy land for a migrant

Objective

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.

Data

- A list of outer suburbs around Colombo was obtained from Wikipedia
- Geo-Coordinates for each suburb were retrieved from ArcGIS geocode library
- The most common places for a given suburb were acquired from the Foursquare API.
- A list of average land prices per perch for the above suburbs is available on lankapropertyweb.
- BeautifulSoup web scraping Python library and Pandas library was used to scrape data from the web.

Methodology

Data gathering and wrangling

1. list of outer suburbs around Colombo was scraped from Wikipedia page using BeautifulSoup web scraping python library.

```
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()
```

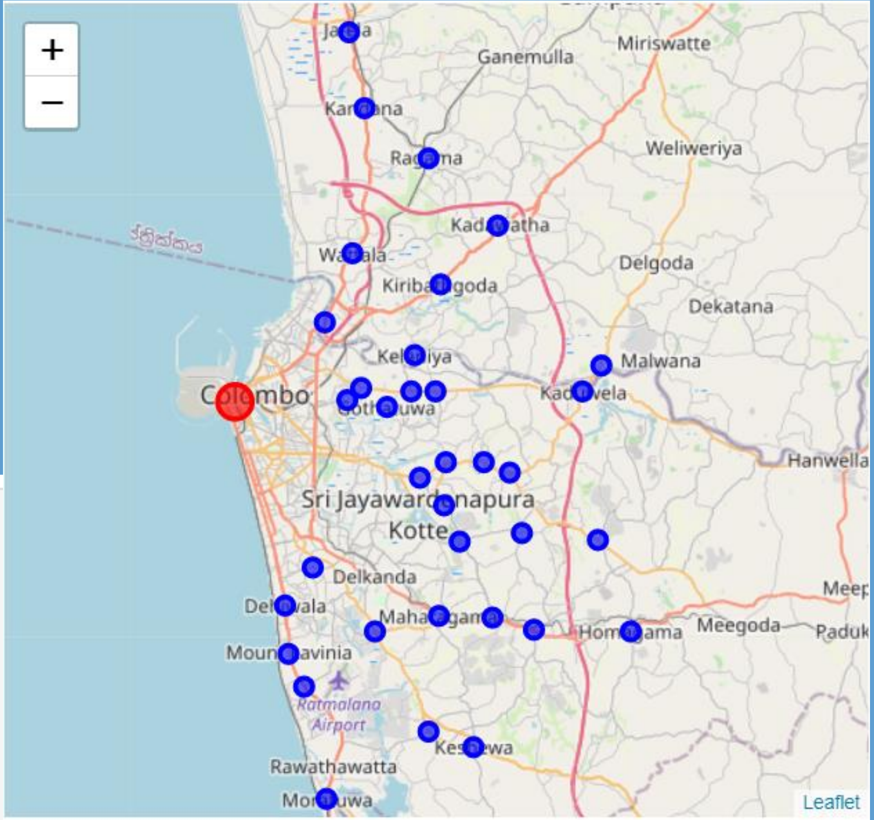
```
(35, 1)
```

	Name
0	Angoda
1	Athurugiriya
2	Battaramulla
3	Biyagama
4	Boralesgamuwa

Methodology

Data gathering and wrangling

2. 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))
```

Methodology

Data gathering and wrangling

3. The list of average land prices per perch was obtained

from Lanka property web.

```
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()
```

```
(49, 2)
```

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

Methodology

Data gathering and wrangling

4. The average land prices retrieved above were merged into the suburbs data frame.

	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

At this point, it was noticed that ‘KOSWATTE’ and ‘KOTIKAWATTA’ suburbs do not have land prices.

Methodology

Data gathering and wrangling

5. Prices of ‘KOSWATTE’ and ‘KOTIKAWATTA’ were imputed using ‘imputePrice’ function and added to the suburbs data frame.

```
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
```

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

Methodology

Data gathering and wrangling

6. Nearby common places were acquired using the Foursquare API.

For that, a free developer account was created on the foursquare developer portal.

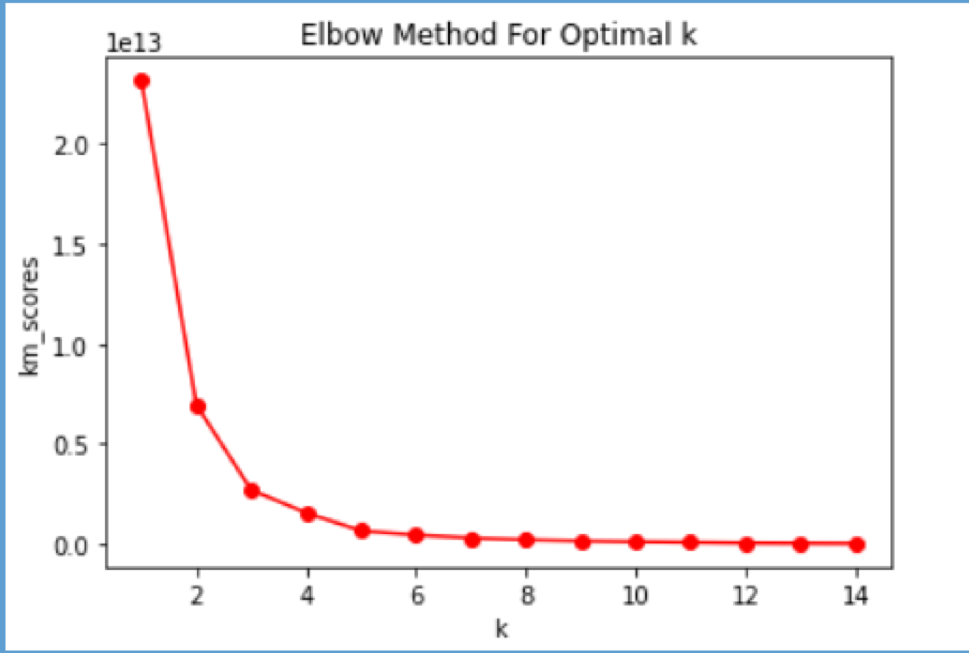
For the simplicity of this study nearest 20 places around 3km perimeter were queried.

Methodology

Analysis

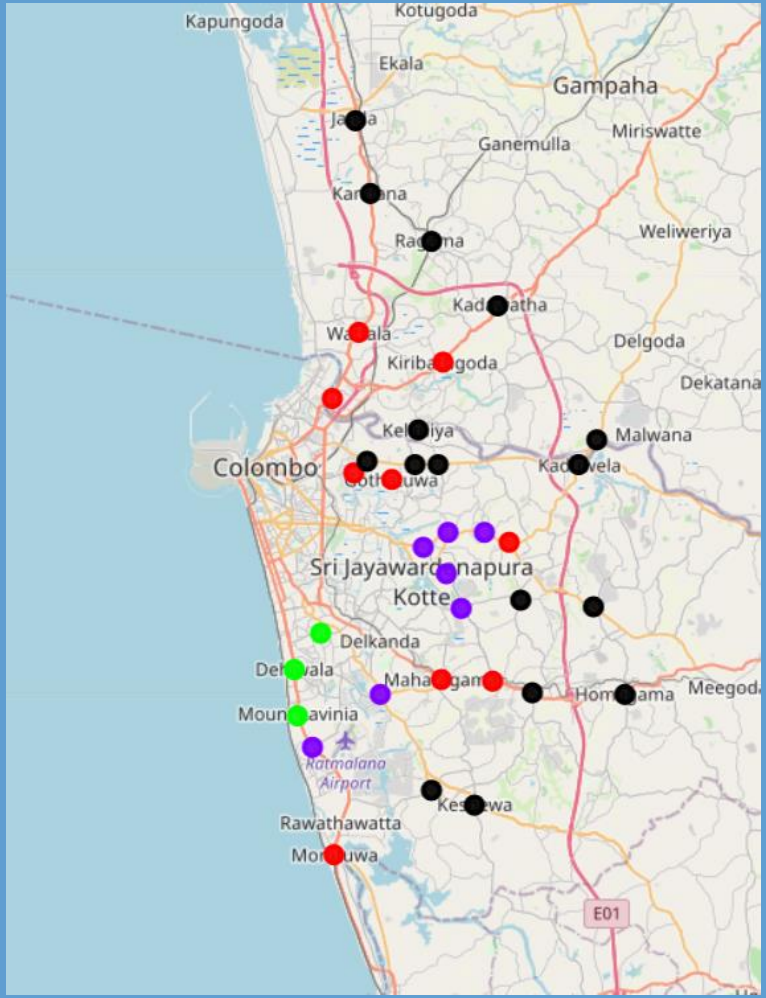
- K-means algorithm was used to cluster suburbs since it is the most promising unsupervised clustering technology suitable to this type of data set.

- Optimal K value was identified as 4 after doing the Elbow test.



Results

	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



Results

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

Conclusion & Future

- This study will be important to a migrant who wants to find similar suburbs based on land prices and surrounding venue types.
- 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.

Thank You !