



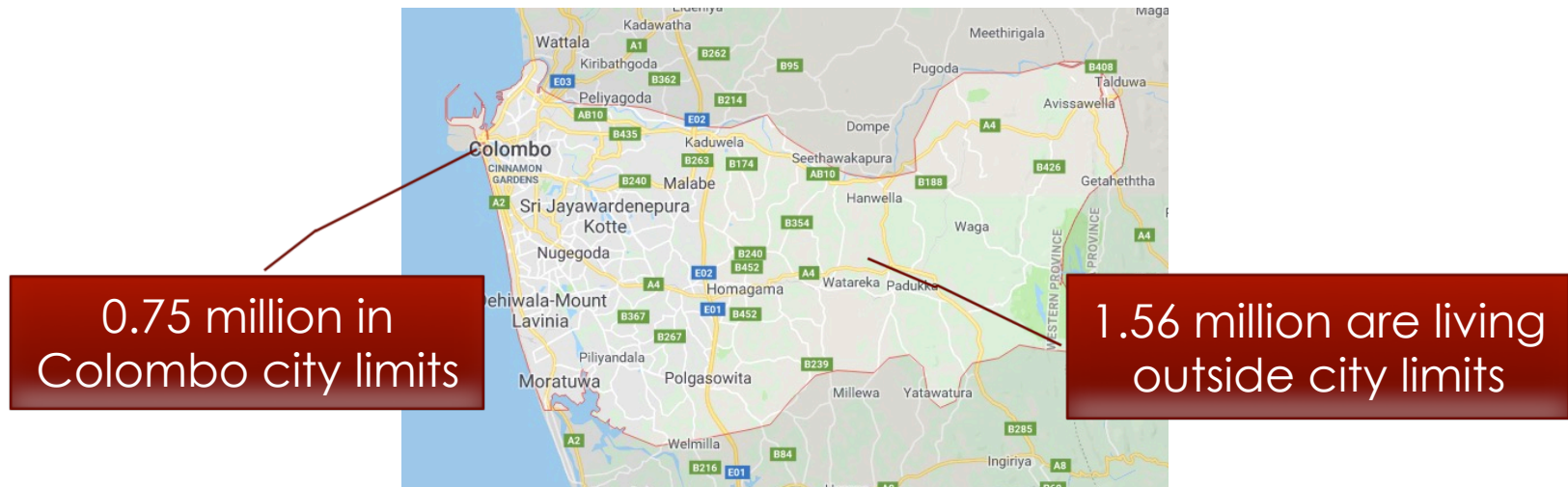
Exploring the Suburbs of Colombo to Open a Bakery

Modisha Jayaratne

Introduction

Background

- Majority of the commercial venues in Colombo district in Sri Lanka are restricted to Colombo city limits
- There is a demand as well as potential for businesses to thrive in suburban areas



Introduction to the Analysis

- Main objective: to analyze the Colombo suburban areas and to identify the most favorable location to open a new bakery
- Target audience: bakery owners who are interested in opening their next outlet in Colombo suburban area, but unsure of the ideal location

Data Extraction & Preprocessing

Data Sources and Extraction

- Colombo district is divided into 13 Divisional Secretary's divisions
- Division names and population data were scraped from the [wikipedia page on Colombo district](#)
- Real estate prices were scraped from [Lanka Property web page](#)
- [Foursquare API](#) to retrieve details of the popular venues in real-time
- [Nominatim](#) Geocoder in GeoPy library to retrieve coordinates of each location

DS Division	Main Town	Divisional Secretary	GN Divisions [6]	Area (km ²) [7]	Population (2012 Census) [8]				C
					Sinhalese	Sri Lankan Moors	Sri Lankan Tamil	Indian Tamil	
Colombo	Colombo	K. G. Dharmathilaka	35	18	79,468	126,345	97,690	8,635	
Dehiwala	Dehiwala-Mount Lavinia	K. Champa N. Perera	15	8	53,182	17,870	12,917	1,240	
Homagama	Homagama		81	121	231,878	508	1,907	460	

Average Land perch prices in Western Province (Q1 2018)

	Average Price (Per Perch)
Aluthgama	278,286
Angoda	536,255
Athurugiriya	344,404

Data Preprocessing

- Removed Colombo and Thimbirigasyaya divisions from the dataset as they fall within Colombo city limits
- Renamed “Sri Jayawardenapura Kotte” as “Kotte” for the purpose of obtaining Foursquare data
- When the division name is different from that of the main city (i.e. Kotte and Seethawaka), the land price of the main city (Ethul Kotte and Avissawella respectively) was assigned explicitly
- Renamed “Rathmalana” as “Ratmalana” in property data, as the table join is affected by the slight difference in spelling

Preprocessed Data

	DivSec	Population	Land Price	Latitude	Longitude
0	Dehiwala	87834	3131549	6.851279	79.865977
1	Homagama	236179	271067	6.841273	80.003058
2	Kaduwela	252057	350032	6.935703	79.984331
3	Kesbewa	244062	353686	6.795740	79.940848
4	Kolonnawa	190817	825000	6.932625	79.890314
5	Maharagama	195355	1030166	6.847278	79.926608
6	Moratuwa	167160	809972	6.774682	79.882610
7	Padukka	65167	95720	6.841538	80.091647
8	Ratmalana	95162	1112115	6.815259	79.866778
9	Seethawaka	113477	164454	6.952948	80.218633
10	Kotte	107508	2010233	6.888322	79.918741

Methodology

Exploratory Data Analysis

- An area within 5km radius from the main city of each division was considered
- Maximum of 100 venues were retrieved per division
- This returned a total of 637 venues across all 11 divisions

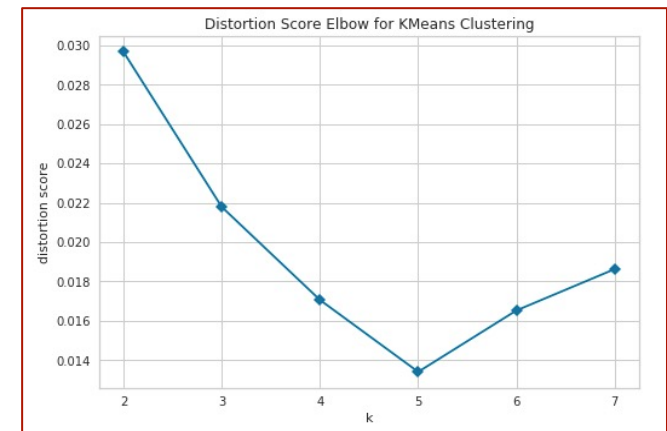
	DivSec	Venue
0	Dehiwala	100
1	Homagama	38
2	Kaduwela	29
3	Kesbewa	25
4	Kolonnawa	100
5	Kotte	100
6	Maharagama	100
7	Moratuwa	31
8	Padukka	7
9	Ratmalana	100
10	Seethawaka	7

Top Ten Most Common Venues

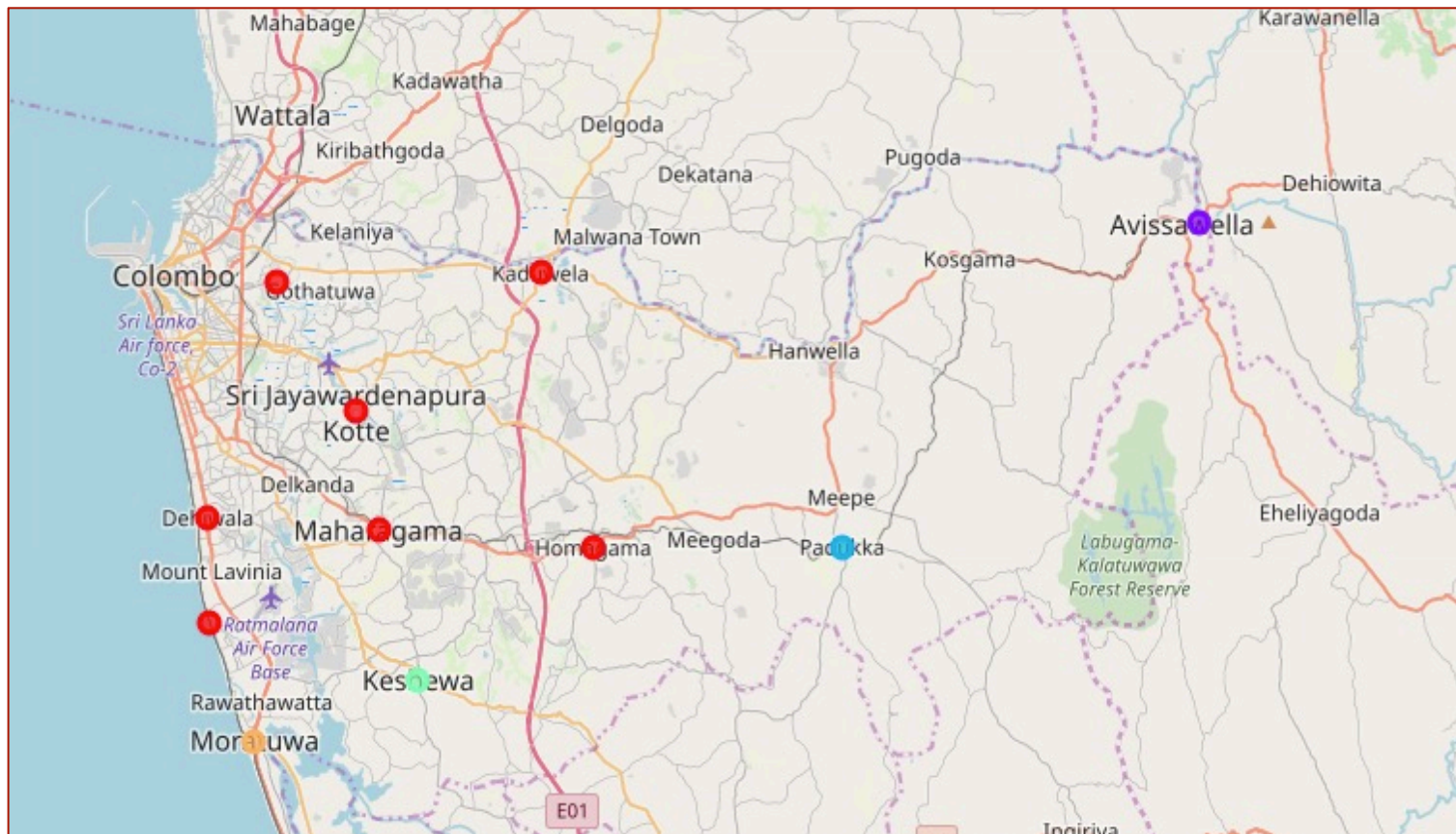
	DivSec	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Dehiwala	Bakery	Restaurant	Clothing Store	Asian Restaurant	Café	Pizza Place	Coffee Shop	Women's Store	Cosmetics Shop	Beach
1	Homagama	Supermarket	Convenience Store	Chinese Restaurant	Bakery	Gym / Fitness Center	Shopping Mall	Pizza Place	Restaurant	Bus Station	Asian Restaurant
2	Kaduwela	Pizza Place	Restaurant	Asian Restaurant	Supermarket	Gym	Fast Food Restaurant	Diner	Movie Theater	College Cafeteria	Snack Place
3	Kesbewa	Bus Station	Grocery Store	Gym / Fitness Center	Department Store	Chinese Restaurant	Shopping Mall	Snack Place	Flea Market	Clothing Store	Pizza Place
4	Kolonnawa	Restaurant	Dessert Shop	Pub	Bakery	Café	Italian Restaurant	Hotel	Sri Lankan Restaurant	Seafood Restaurant	IT Services
5	Kotte	Bakery	Gym	Restaurant	Convenience Store	Café	Supermarket	Clothing Store	Asian Restaurant	Coffee Shop	Fast Food Restaurant
6	Maharagama	Supermarket	Bakery	Gym	Convenience Store	Chinese Restaurant	Pizza Place	Asian Restaurant	Restaurant	Café	Bookstore
7	Moratuwa	Clothing Store	Restaurant	Train Station	Chinese Restaurant	Pizza Place	Resort	Food Court	Supermarket	Juice Bar	Fast Food Restaurant
8	Padukka	Bakery	Resort	Train Station	Bus Station	Tea Room	Shopping Mall	Women's Store	Fair	College Cafeteria	Comfort Food Restaurant
9	Ratmalana	Restaurant	Pizza Place	Clothing Store	Shopping Mall	Asian Restaurant	Bakery	Chinese Restaurant	Fast Food Restaurant	Department Store	Convenience Store
10	Seethawaka	Resort	Restaurant	Pizza Place	Comfort Food Restaurant	Café	Convenience Store	Bus Station	Cocktail Bar	Coffee Shop	College Cafeteria

Clustering Divisions

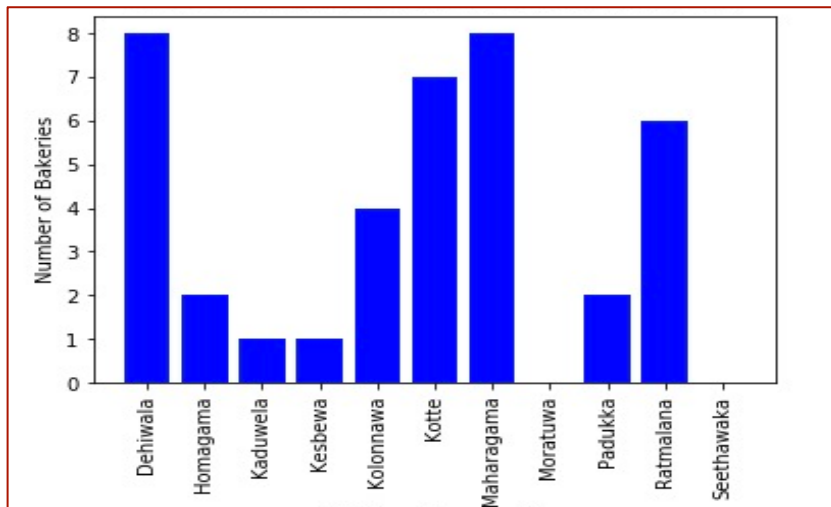
- K-means clustering is an unsupervised learning algorithm that helps to identify structure in the data
- This algorithm was used with the intention of uncovering the hidden similarities between different divisions and grouping them appropriately
- The “elbow method” was used to identify the optimal k value and it returned k=5



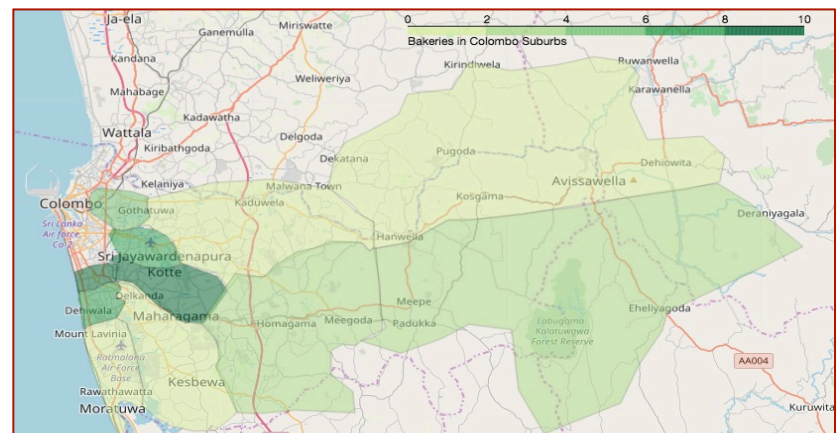
Division Clusters



Exploratory Analysis on Bakeries



Number of Bakeries	DivSec
0-1	Kaduwela, Kesbewa, Moratuwa, Seethawaka
2-3	Homagama, Padukka
4-5	Kolonnawa
6-7	Kotte, Ratmalana
8-9	Dehiwala, Maharagama



Top five preferable locations to Open a Bakery

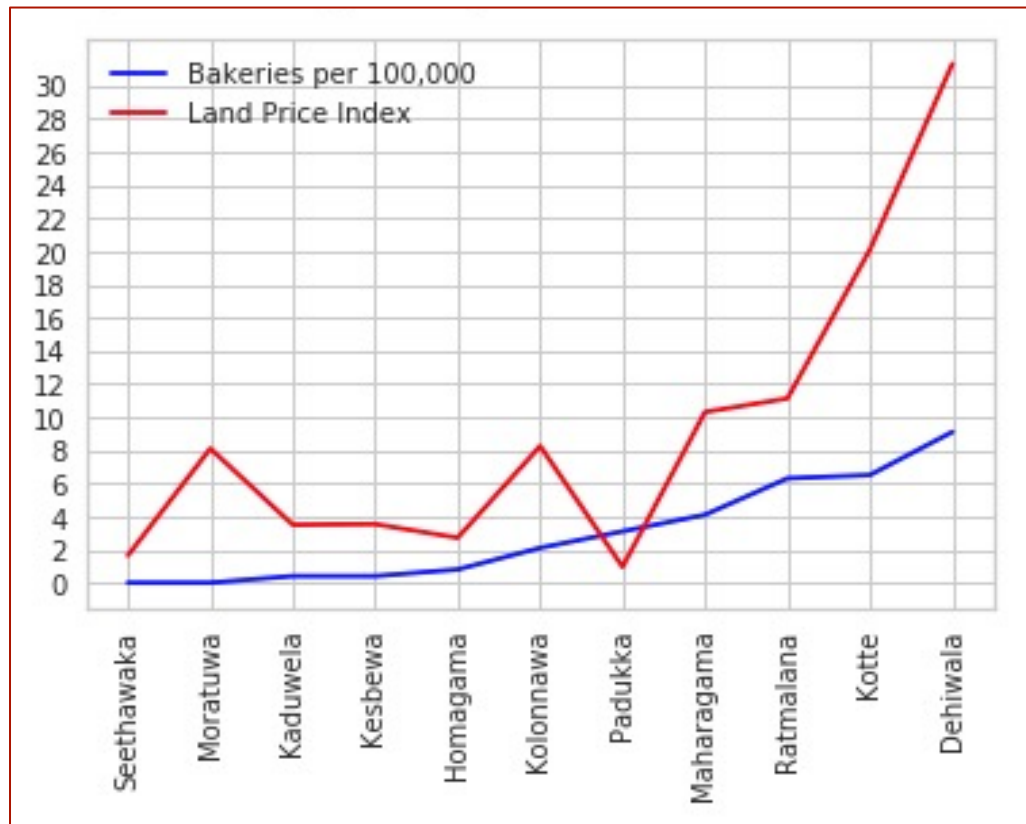
- Criteria for deriving the sequence of preference:
 - Number of bakeries for 100,000 people in each division
 - Property price (tiebreaking criteria when two divisions are similar in terms of the per capita bakeries)

	DivSec	No of Bakeries	Population	Bakeries per 100,000	Land Price
0	Seethawaka	0	113477	0.0	164454
1	Moratuwa	0	167160	0.0	809972
2	Kaduwela	1	252057	0.4	350032
3	Kesbewa	1	244062	0.4	353686
4	Homagama	2	236179	0.8	271067

1

2

Visualizing the Results



Results

Summary of Results

- Bakeries alone do not follow the same clustering pattern when all the venues are considered together
- Seethawaka, Moratuwa, Kaduwela, Kesbewa and Homagama are the top five divisions to consider
- Seethawaka and Moratuwa are the most favorable divisions to open a bakery
- When the cost of real estate is added as a second criterion, Seethawaka is clearly on the lead
- Considering the investment in real estate, Kaduwela, Kesbewa and Homagama are preferred to Moratuwa

Discussion

Discussion (1/2)

- The analysis was solely based on the data acquired from external sources
- The reliability of the analysis is directly dependent on the reliability of the data
- There could be many venues that are not captured by Foursquare
- The population statistics are from 2012 and may have changed significantly
- The land prices are from the first quarter of 2018 and may also be outdated

Discussion (2/2)

- Only the popular venues, population and the land price of each division were considered
- There may be many other aspects to consider before selecting a location to open a bakery
- In order to arrive at a concrete conclusion, a comprehensive analysis should be conducted with more data pertaining to these other dynamics

Conclusion

Summary

- Colombo suburban area was studied to identify the optimal division to open a new bakery
- The main factors were popular venues and the population
- Land prices were used as a supplementary factor in a tiebreaking scenario
- With these three factors, Seethawaka is the most favorable division to open a bakery
- This conclusion is subject to limitations such as the reliability of data and the limited factors considered

Thank you