

Coursera Capstone
IBM Applied Data Science Capstone

Analysis of cafes within Colombo, Sri Lanka

By Dilshan Manathunge

May 2020



Introduction and Business Problem

Introduction

Sri Lanka is a relatively small island directly south of India, with a population of more than 21 Million residents (as of 2018) of which around 5.6 Million populate the capital of Colombo. Colombo is also the biggest city within the island and serves as the financial capital and a tourist destination. As such Colombo serves as a hunting ground for up and coming businesses such as cafes and restaurants. It is a can daunting for a person to search for the best place to eat/drink with the various options presented in this ever-growing city.

Business Problem

For this capstone I will be creating a guide utilizing the Foursquare API, to categorize cafes within Colombo, after which I will be clustering the cafes according the amount of likes they have received. My aim is to build a guide to traverse through the various coffee shops of Colombo and select the best ones.

Data

To solve this problem, I will be utilizing the following data,

- A list of the inter-city limits of Colombo.
- The latitudes and longitudes of those areas for visualization.
- Venue Data utilizing the foursquare API.

Data gathering

To obtain the list of inter-city limits I will be extracting the content from the Wikipedia Page (<https://en.wikipedia.org/wiki/Colombo>) which contains the inter-city limits. I will be using web scraping techniques to extract the data with the help of python libraries such as beautiful soup and requests. The Latitudes and Longitudes of the Inter-city Limits will be obtained using python's Geocoder package. After which I will be utilizing the Foursquare API to get venue Data of the relevant locations.

Methodology

When starting this project my thought process behind it was that the more likes they receive the better the coffee shop is. While this is an inaccurate measurement for measuring quality due to the restrictions of the free tier account from Foursquare, I was un-able to retrieve rating. This error has been somewhat rectified by binning the data into categorical variables.

I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4 clusters so that people looking for a coffee shop can easily see which coffee shops are the best and where in Hoboken they can look to eat.

Results

Running my clustering algorithm, I was able to generate four clusters of coffee shops. These are as follows:

Cluster 1

- Characteristics
 - Poor quality food

	Neighborhood	Latitude	Longitude	VenueId	VenueName	VenueLatitude	VenueLongitude	VenueCategory	total likes	total likes_cat	label
0	Slave Island	6.926520	79.850200	4d8a1106d85f37044a2ff5db	The Dining Room	6.929095	79.849568	Coffee Shop	22	poor	0
2	Slave Island	6.926520	79.850200	51dcff7e498e94b774bb1a0e	Pebbles cafe, Taj Samudra	6.923323	79.846273	Coffee Shop	11	poor	0
4	Slave Island	6.926520	79.850200	4c8b43d575ba9c749e81575c	Goodies	6.928913	79.849598	Coffee Shop	19	poor	0
5	Slave Island	6.926520	79.850200	5155636fe4b08e31444f9e39	Cafe Seventy Seven	6.920389	79.853801	Coffee Shop	5	poor	0
6	Slave Island	6.926520	79.850200	4de0e9d11f6ece647385c007	The Verandah	6.920024	79.845865	Coffee Shop	14	poor	0
8	Union Place	6.918760	79.861130	50cc7812e4b05e0e4e80e1e9	Odel Boulevard	6.916773	79.865148	Coffee Shop	14	poor	0
9	Union Place	6.918760	79.861130	590d92a2cad1b67f0cbf302e	Life Food	6.916240	79.853326	Coffee Shop	8	poor	0
11	Union Place	6.918760	79.861130	4e2959f952b1c37d35caca13	M3 Cafe	6.914540	79.857053	Coffee Shop	14	poor	0
13	Union Place	6.918760	79.861130	5155636fe4b08e31444f9e39	Cafe Seventy Seven	6.920389	79.853801	Coffee Shop	5	poor	0
14	Union Place	6.918760	79.861130	5a609723ee628b483065da85	Caramel Pumpkin	6.916296	79.869043	Coffee Shop	13	poor	0
15	Kollupitiya	6.912680	79.851350	553915b3498eb10d08ee7371	The Coffee Bean & Tea Leaf	6.911989	79.851029	Coffee Shop	24	poor	0
17	Kollupitiya	6.912680	79.851350	590d92a2cad1b67f0cbf302e	Life Food	6.916240	79.853326	Coffee Shop	8	poor	0
18	Kollupitiya	6.912680	79.851350	59da19494928143ec58f0c33	Tea Avenue	6.908325	79.849255	Coffee Shop	9	poor	0
22	Kollupitiya	6.912680	79.851350	4e2959f952b1c37d35caca13	M3 Cafe	6.914540	79.857053	Coffee Shop	14	poor	0
23	Kollupitiya	6.912680	79.851350	5489298a498e3141e039ed7d	Delifrance	6.906575	79.853310	Coffee Shop	6	poor	0
24	Bambalapitiya	6.904710	79.855070	55e5188a498e34b61c727637	The Coffee Bean & Tea Leaf	6.903622	79.852232	Coffee Shop	18	poor	0
25	Bambalapitiya	6.904710	79.855070	5aae6995e65d0c06df889c1d	The Embazzy	6.906572	79.858438	Coffee Shop	14	poor	0
30	Bambalapitiya	6.904710	79.855070	527b8ca211d2fbb60218d34c	Mitsi's Delicacies	6.902005	79.857477	Coffee Shop	21	poor	0
31	Bambalapitiya	6.904710	79.855070	59da19494928143ec58f0c33	Tea Avenue	6.908325	79.849255	Coffee Shop	9	poor	0
32	Bambalapitiya	6.904710	79.855070	5489298a498e3141e039ed7d	Delifrance	6.906575	79.853310	Coffee Shop	6	poor	0
33	Bambalapitiya	6.904710	79.855070	553915b3498eb10d08ee7371	The Coffee Bean & Tea Leaf	6.911989	79.851029	Coffee Shop	24	poor	0
34	Bambalapitiya	6.904710	79.855070	4fcf6db0e4b0f54f2917ec83	Cloud Cafe	6.897779	79.856392	Coffee Shop	15	poor	0
38	Havelock Town	6.887580	79.862560	59cba4328d0a536b6c9a39a8	Cafe Noir Blanc	6.892419	79.869752	Coffee Shop	11	poor	0

Cluster 2

- Characteristics
 - Good quality food

```
cafe_venues.loc[cafe_venues['label'] == 1]
```

	Neighborhood	Latitude	Longitude	VenueId	VenueName	VenueLatitude	VenueLongitude	VenueCategory	total likes	total likes_cat	label
3	Slave Island	6.926520	79.850200	4c7192ccf4d476b035bd5acf	Coffee Stop - Cinnamon Grand	6.917826	79.848547	Coffee Shop	133	great	1
7	Union Place	6.918760	79.861130	4c0771cb3cbcd13a7af40bc0	Coco Veranda	6.916424	79.868338	Coffee Shop	571	great	1
19	Kollupitiya	6.912680	79.851350	4c7192ccf4d476b035bd5acf	Coffee Stop - Cinnamon Grand	6.917826	79.848547	Coffee Shop	133	great	1
20	Kollupitiya	6.912680	79.851350	4c0b3fd5340720a191f28793	The Commons	6.908486	79.858051	Coffee Shop	286	great	1
26	Bambalapitiya	6.904710	79.855070	4b991f3df964a520ca6235e3	The Gallery Café	6.898672	79.854599	Coffee Shop	269	great	1
27	Bambalapitiya	6.904710	79.855070	4c0b3fd5340720a191f28793	The Commons	6.908486	79.858051	Coffee Shop	286	great	1
29	Bambalapitiya	6.904710	79.855070	4cb9f8f543ec6dc891aa131	The Barefoot Cafe	6.896644	79.853765	Coffee Shop	199	great	1
36	Havelock Town	6.887580	79.862560	4efaaa3df790731254d52c57	Java Lounge	6.892233	79.867115	Coffee Shop	318	great	1
39	Kirulapana	6.878604	79.873269	4c305b767cc0c9b69b40ed9a	Barista	6.877262	79.873039	Coffee Shop	207	great	1
45	Kirilapone North	6.881350	79.875880	4c305b767cc0c9b69b40ed9a	Barista	6.877262	79.873039	Coffee Shop	207	great	1
55	Pamankada	6.879540	79.869370	4c305b767cc0c9b69b40ed9a	Barista	6.877262	79.873039	Coffee Shop	207	great	1
61	Cinnamon Gardens	6.909510	79.867750	4b991ccbf964a5201e6235e3	The Coffee Bean & Tea Leaf	6.911452	79.866747	Coffee Shop	346	great	1
64	Cinnamon Gardens	6.909510	79.867750	4c0771cb3cbcd13a7af40bc0	Coco Veranda	6.916424	79.868338	Coffee Shop	571	great	1
71	Cinnamon Gardens	6.909510	79.867750	4f0d7c39e4b0281b8c4e48ff	Choco Luv	6.909549	79.874280	Coffee Shop	139	great	1
75	Borella	6.916690	79.877620	4f0d7c39e4b0281b8c4e48ff	Choco Luv	6.909549	79.874280	Coffee Shop	139	great	1
80	Maradana	6.924730	79.865700	4c0771cb3cbcd13a7af40bc0	Coco Veranda	6.916424	79.868338	Coffee Shop	571	great	1

Cluster 3

- Characteristics
 - Below average quality food

```
cafe_venues.loc[cafe_venues['label'] == 2]
```

	Neighborhood	Latitude	Longitude	VenueId	VenueName	VenueLatitude	VenueLongitude	VenueCategory	total likes	total likes_cat	label
1	Slave Island	6.92652	79.85020	4b6fa5cdf964a52006f92ce3	Barista	6.932601	79.843855	Coffee Shop	38	below avg	2
10	Union Place	6.91876	79.86113	56716368498ee66e925b8ffa	Black Cat Cafe	6.915373	79.869426	Coffee Shop	31	below avg	2
12	Union Place	6.91876	79.86113	521d962c498e314c4ac39d7a	Laksala Museum Gallery Café	6.910536	79.861873	Coffee Shop	29	below avg	2
37	Havelock Town	6.88758	79.86256	53c7cc97498e210e88986473	Java Lounge	6.882134	79.858934	Coffee Shop	42	below avg	2
51	Wellawatte	6.87924	79.85989	53c7cc97498e210e88986473	Java Lounge	6.882134	79.858934	Coffee Shop	42	below avg	2
53	Wellawatte	6.87924	79.85989	52e79d2211d21513c70f616a	That Cafe in Colombo	6.875635	79.861391	Coffee Shop	30	below avg	2
60	Pamankada	6.87954	79.86937	52e79d2211d21513c70f616a	That Cafe in Colombo	6.875635	79.861391	Coffee Shop	30	below avg	2
65	Cinnamon Gardens	6.90951	79.86775	562e0cc6498e85c97a0b9753	London House of Coffee	6.909475	79.866491	Coffee Shop	31	below avg	2
66	Cinnamon Gardens	6.90951	79.86775	56716368498ee66e925b8ffa	Black Cat Cafe	6.915373	79.869426	Coffee Shop	31	below avg	2
67	Cinnamon Gardens	6.90951	79.86775	4d3e8f813ec9a35dd2525b81	Aviraté Café	6.908334	79.865088	Coffee Shop	36	below avg	2
68	Cinnamon Gardens	6.90951	79.86775	521d962c498e314c4ac39d7a	Laksala Museum Gallery Café	6.910536	79.861873	Coffee Shop	29	below avg	2
70	Cinnamon Gardens	6.90951	79.86775	4e608ec4b993678bb574f058	Caffè Pascucci	6.910128	79.871694	Coffee Shop	41	below avg	2
72	Cinnamon Gardens	6.90951	79.86775	52cbe6e8498e966d6dea55b7	Tea Breeze	6.905846	79.863597	Coffee Shop	26	below avg	2
74	Borella	6.91669	79.87762	56716368498ee66e925b8ffa	Black Cat Cafe	6.915373	79.869426	Coffee Shop	31	below avg	2
76	Borella	6.91669	79.87762	4e608ec4b993678bb574f058	Caffè Pascucci	6.910128	79.871694	Coffee Shop	41	below avg	2

Cluster 4

- Characteristics
 - Average quality food

```
In [49]: cafe_venues.loc[cafe_venues['label'] == 3]
```

```
Out[49]:
```

	Neighborhood	Latitude	Longitude	VenueId	VenueName	VenueLatitude	VenueLongitude	VenueCategory	total likes	total likes_cat	label
16	Kollupitiya	6.91268	79.85135	5059f5cee4b0fd1a4976ab00	Coco Veranda	6.917147	79.848890	Coffee Shop	86	avg	3
21	Kollupitiya	6.91268	79.85135	52fe011d1d2f511eb909e2	Whight & Co	6.905171	79.850088	Coffee Shop	96	avg	3
28	Bambalapitiya	6.90471	79.85507	52fe011d1d2f511eb909e2	Whight & Co	6.905171	79.850088	Coffee Shop	96	avg	3
35	Bambalapitiya	6.90471	79.85507	4c835d57e63376b0fb47b93d	Barista	6.898225	79.860518	Coffee Shop	77	avg	3
62	Cinnamon Gardens	6.90951	79.86775	5111f973e4b0b4b884f62da3	Gerard Mendis Chocolatier	6.911246	79.870284	Coffee Shop	56	avg	3
63	Cinnamon Gardens	6.90951	79.86775	5300828811d2e07f25e5c5cc	The Paddington	6.911484	79.868723	Coffee Shop	56	avg	3

```
In [ ]:
```

Map of Clusters for Users

