

Coursera Capstone
IBM Applied Data Science Capstone

Analysis of cafes/restaurants 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/restaurants according to the price/location and relevant cuisine. I will then cluster these restaurants based on their similarities so that a user can easily determine what type of cuisine they are looking for and in what price range.

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 BeautifulSoup 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.