

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Café in Kochi, Kerala, India

By: Job Thomas T

June 2020



Introduction

Kochi is becoming one of the busiest places in India. Kochi, known (a little tritely) as The Queen of the Arabian Sea, has much to offer the curious traveler, whatever their interests might be. While a lot of this has to do with its affinity for the arts, and embracing of the new and modern, its rich and colorful history cannot be denied. Kochi was one of the most important port cities in the international spice trade from the late 1400s, drawing traders from Europe, West Asia and China. Besides these influences, the city also had a small but thriving population of Jews and Anglo-Indians. The cuisine, architecture, and general culture features all of these influences including the traditions of the Portuguese, Dutch and English colonizers.

The city has always been quietly obsessed with food. However, in recent years—mostly thanks to the Kochi Muziris Biennale—the rest of the world has been let in on this open secret. And they just can't get enough.

As for an entrepreneur this is the right time to start a food joint such as a café in this vibrant city. Of course, as with any business decision, opening a new café requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the café is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Kochi, Kerala, India to open a new café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kochi, Kerala, India, if a property developer is looking to open a new café, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful for young entrepreneurs and investors looking to open or invest in café in Kochi. This project is timely as the city is becoming more popular with tourists and is one of the best place in Kerala with a night life.

Data

To solve the problem, we will need the following data:

- List of neighborhoods in Ernakulam district. This defines the scope of this project.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to cafe. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This Wikipedia page ("[https://en.wikipedia.org/wiki/Category:Suburbs_of Kochi](https://en.wikipedia.org/wiki/Category:Suburbs_of_Kochi)") contains a list of neighborhoods in Kochi, with a total of 44 neighborhoods. I myself have added one more which seemed relevant. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data, we are particularly interested in the Café category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.