

IBM Applied Data Science Capstone Project

Kalaivanan.S

12-February-2020

Part-1: Problem Description

Objective:

The objective of this project is to analyze a suitable location to open up a coffee shop in Chennai, India.

Introduction:

Chennai is a metropolitan, capital city located in the southern state of Tamilnadu, India. Being a cosmopolitan city, people from all over India stay and work in this city, the major employers being Information Technology industry and Automobile sector. In addition to the people living there, foreign tourists flock this place all through the year.

This has led to a spur in the food and hospitality industry. People throng to the various cuisines located in the suburbs of Chennai on a daily basis. The most common place visited by people of all cultures (both domestic and foreign) would be a nice coffee shop.

Business Problem:

This project analyzes the neighborhoods present in Chennai, and the distribution of coffee shops in these neighborhoods. The final objective would be to identify a suitable place to open up a new coffee shop in the suburbs of Chennai. This project clusters the neighborhoods according to the number of coffee shops present in each neighborhood. Web scraping and machine learning techniques are utilized to arrive at a conclusion.

Part-2 Data Description

Data required:

The below mentioned Data would be required to solve the objective

- The list of neighborhoods in the city of Chennai.
- The geographical coordinated of the neighborhoods
- The list of venues, particularly coffee shops located in these neighborhoods

Data Sources:

The neighborhoods data of Chennai is scraped from the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai). The list contains a total of 61 neighborhoods located in Chennai. BeautifulSoup package is used to scarp the data from Wikipedia page and store the neighborhoods in the dataframe.

Geocoder from geopy package is used to extract the geographical coordinates of the stored neighborhoods. The latitudes and longitudes are append to the neighborhoods dataframe.

Foursquare API is used to get the details of trending venues for the corresponding neighborhoods in the dataframe. Among the trending venues, the coffee shop venues are segregated and clustering algorithms are applied to find out a suitable place for opening up a new coffee shop. The methodology and inference are detailed in the next part of the documentation.