

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Ice-Cream shop in Bangalore, India

By: Nikhil Agarwal

April 9, 2020

Introduction

Ice-Cream has always been favourite choice for everyone when it comes to food during entertainment, trips or on day outings. And so do business, franchises and eateries shops. Now a days there are ice-cream franchises, local shops made on a big scale. As the trend of fancy shops have increased so has the investments in it and with big investments come big risks. So, before opening a shop one has to be aware about its surrounding like competition it is going to get. So, Location becomes very important for any business. Because audience is everything.

Business Problem

The objective of this capstone project is to analyse and select all neighbourhood areas in the city of Bangalore, India to open a new Ice-Cream parlour. 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 Bangalore, India, if a property developer is looking to open a new ice-cream parlour, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new ice-cream parlour in the capital city of Bangalore, India. Bangalore is huge IT hub and people here lives for the weekend and lot of people here loves to spend life eating and celebrating. Now, to make more profit we need to find an area which is famous or good but has less number of ice-cream parlour so that we can get minimum competition. So, before an investment results of this project can be really helpful.

Data

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

- List of neighbourhoods in Bangalore: This defines the scope of this project which is confined to the city of Bangalore, the metropolitan city of the country of India in South Asia.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to existing Ice-cream shops. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore) contains a list of neighbourhoods in Kuala Lumpur, with a total of 65 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and pandas packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. 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 Shopping Mall 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.