

# Coursera Capstone

IBM Applied Data Science Capstone

## *Opening a New Coffee Shop in New Delhi, India*

By: Harsh Sharma

October, 2019



[This Photo](#) by Unknown Author is licensed under [CC BY](#)

# Introduction

For many people either it be working professionals or just hanging out buddies a coffee shop provides a soothing place to hang-out or to chill along with your friends. Thus a vast majority of people belonging to age groups 18-35 enjoys coffee breaks from their monotonous and hectic schedule to go for a coffee break. Hence the coffee shops have become an integral part of their lives.

## Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of New Delhi, India to open a new Coffee Shop. 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 Delhi, India, if a property developer is looking to open a new Coffee Shop, where would you recommend that they open it?

## Target Audience of this project

This project is particularly useful to shopkeepers who want to introduce new ventures such as coffee shops under their trade names. Thus due to vast area of this city a normal businessman or a person willing to open a coffee shop would rely on the other's opinion regarding the best place to open a shop, but it would be time consuming and cumbersome. Thus This project would help them to identify the best places to set up their shops

# Data

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

- List of headquarters in the delhi city. This defines the scope of this project which is confined to the city of New Delhi, the capital city of the country of India in South East Asia.
- Latitude and longitude coordinates of those headquarters. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to coffee shops. We will use this data to perform clustering on the headquarters.

## **Sources of data and methods to extract them**

This Wikipedia page ([https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Delhi](https://en.wikipedia.org/wiki/List_of_districts_of_Delhi)) contains a list of neighbourhoods in Delhi, with a total of 11 headquarters. 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 headquarters 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.