

# Data Science Capstone Project

By Aayush Kamath

## Problem Statement in brief:

This project intends to analyze suburbs/neighborhoods in Mumbai, and come up with suburbs/neighborhoods which might be ideal for opening a café.

## Data Required to solve the business problem:

1. The list of suburbs/neighborhoods in Mumbai, India.
2. The geographical co-ordinates, that is, the latitude and longitude of the neighborhoods in Mumbai.
3. Data of venues near all neighborhoods, especially data of cafes.

## Data Sources:

- The list of suburbs/neighborhoods in Mumbai is scraped from the wikipedia page of suburbs in Mumbai.

(Link: [https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai))

The scraping of data is done using the BeautifulSoup library in Python.

- Once we have the list of all neighborhoods, we find the latitude and longitude for each of them using the Geocoder package in Python.
- The venue data for all neighborhoods are obtained by using the Foursquare API. By writing appropriate queries, we can access the necessary venue data through this API. Once the data is obtained, we sift through it to find the data that is necessary for arriving at a solution for our business problem.