# *Coursera capstone project: Applied data science*

Saurabh Chaudhari
Saurabhwvu@gmail.com

---

# 1.Introduction

Mumbai is the capital city of Maharashtra, with a population of around 20 millions. It is second most populated city in India and 7$^{th}$ most populated city in the world. The city of Mumbai consists of two distinct regions: Mumbai City district and Mumbai Suburban district, which form two separate revenue districts of Maharashtra. The western part of the Mumbai Suburban district forms the Western Suburbs and the eastern portion forms the Eastern Suburbs. The suburbs of Chembur, Govandi, Mankhurd and Trombay lie to the south-east of the Eastern Suburbs. These suburbs are generally not considered as part of the Eastern Suburbs and are sometimes referred to as the "Harbour Suburbs".

# 2.Business Problem

Most of the population in the Mumbai is working population and it need to eat breakfast, lunch, dinner and snacks, while travelling through different modes of transportation such as bus, local trains and taxies. Foods that attract busy people on the go include egg sandwiches, fries, pizza, burgers, microwaveable or cold prepared meals. Beverages such as coffee, tea, wraps, bottled water, soda and juice also sell well.
Thus, the main objective of the project will be to find ideal spots in the city where fast food retail chains can be put up, aiming at the above demographic, thereby helping the owners of the outlets to extract maximum profits out of them.

# 3.Data

The data for this project has been retrieved and processed using through multiple sources.

## 3.1 Neighborhood

The data of the neighborhoods in Mumbai can be extracted out by web scraping using BeautifulSoup library for python. The neighborhood data is scrapped from a Wikipedia webpage. This data is converted into a pandas Data frame. The good thing about it data is that it provides the latitude and longitude of the neighborhoods too and these values are further stored into initial dataframe.

## 3.2 Venue data

From the location data obtained after Web Scraping and Geocoding, the venue data is found out by passing in the required parameters to the FourSquare API, and creating another DataFrame to contain all the venue details along with the respective neighborhoods.