# COURSERA - IBM APPLIED DATA SCIENCE

## Capstone Project – Week 5

## INTRODUCTION

**Pune** district is the cultural capital of Maharashtra state in India. With it's roots firmly anchored in the ancient Maratha culture, Pune can boast of a highly intellectual, cosmopolitan population with a homogeneous composition of people originating from all *corners of India*

Pune has been ranked as the eighth most populous city in India, with an estimated population of about 7.2 million as of 2019. It has been ranked as one of the most livable cities in India several times.

Along with its extended city limits Pimpri Chinchwad and the three cantonment towns of Pune, Khadki and Dehu Road, Pune forms the urban core of the eponymous Pune Metropolitan Region (PMR). According to the 2011 census, the urban area has a combined population of 5.05 million while the population of the metropolitan region is estimated at 7.27 million. Situated 560 metres (1,837 feet) above sea level on the Deccan plateau on the right bank of the Mutha river, Pune is also the administrative headquarters of its namesake district.

More popular as the IT Capital of India, Pune is also known as the "Oxford of the East" due to the presence of several well-known educational institutions. The city has emerged as a *major educational hub in recent decades, with nearly half of the total international* students in the country studying in Pune. Research institutes of information technology, education, management and training attract students and professionals from India and overseas.

## BUSINESS PROBLEM

Over the past decade, Pune city has seen a huge transformation in infrastructure due to the immense opportunities that it presents as one of the fastest developing cities of India. Hence, the platform for huge business opportunities and business friendly environment ends up attracting leading players into the market. This has thus resulted in a highly competitive market, and the costs of doing business is comparable to some of the highest in the country. This itself requires a thorough analysis of the business opportunities, and any investment must be strategically targeted so as to maximize returns,

### Background

The analysis presented in the Capstone project maybe reviewed by any investor, business shop or venture capitalist to arrive at an informed decision on what business opportunity should be explored with an eye on location parameters

### Business Problem

We expect that the following questions will have ready answers by the end of the analysis

- How is the entire city structured with regard to infrastructure, strategic locations, population?
- Which are the highly dense & well-developed areas that can be picked up for consideration
- Which are the areas in Pune that provide the highest business opportunities? And for what kind of business.

# DATA

## Data Sources

To make the analysis and observations more effective, it was decided to acquire data from multiple sources, following are the details:

From **Opendatasoft** API (https://public.opendatasoft.com/explore/?sort=modified&q=pune), Pune city data, along with neighbourhoods was retrieved. Following features were available to form the city map.

- **Country code:** Country information.
- **admin name 1:** State level details (Maharashtra, in this case).
- **admin name 2:** All cities included in Pune district.
- **admin name 3:** Defined neighbourhoods in Pune city.
- **place name:** The street, lane name or *peth* as locally called
- **postal code:** Zip code
- **Latitude:** The latitude value of the venue.
- **Longitude:** The longitude value of the venue.

Subsequently, **Foursquare's** explore API (https://api.foursquare.com/v2/venues) (which gives venues recommendations) was leveraged. With Pune as the center, and a radius of 500 m to scout around – each location/place was broken up into a series of venues and detailed information was procured to assist in identifying the tops spots within the city.

From Foursquare API (https://developers.zomato.com/api), following attributes for each venue were retrieved:
- **Name**: The name of the venue.
- **Category**: The category type as defined by the API.
- **Latitude**: The latitude value of the venue.
- **Longitude**: The longitude value of the venue.