# Coursera Capstone

*IBM Applied Data Science*

## Opening a New Shopping Mall in Austin, Texas

By: Sahil Desai

June 2020

## Introduction

Shopping malls are a great way for people to spend their leisure time and socialize. Mall shopping, grocery shopping, activities including seeing movie theatres or attractions, and events are all present at malls and reap in a lot of profit. Malls bring different types of consumers or target markets due to the diverse number of things people can do there. Retailers want to seize the opportunity shopping malls provide for their business. It is a great area to advertise and launch new products, and a way to create a distribution channel. Property development companies want to take advantage by building malls to cater demand and bring in profit. Austin is listed as the fastest growing city in the United States. As a result, many malls and other venues are being built to adhere to the growing population. The goal of property developers is to make rental income of their tenants. Like any business decision, they need to choose the location of where they build very carefully. Shopping malls on average cost $25 million dollars to build, so they are no small investment!

### Business Problem

The objective of this capstone project is to find out where is the best location to build a shopping mall in Austin, Texas. Throughout the teachings of the courses, data science methodology and machine learning techniques like clustering will be implemented to make data driven decisions to clients. Based off the findings, where would be the best area to build a new mall in Austin as a property developer?

### Target Audience

The target audience of the project are property developers, architectural firms, subcontractors, or a lead contractor looking to invest or construct a mall in the Austin Texas area. With the rapidly growing population (majority being young professionals), Austin is constantly building new things weather its new offices/buildings for their tech companies, or attractions to keep their city lively and citizens invested financially to the city’s economy. Building a mall there could be a good investment due to the new population. It can be mutually beneficial to the citizens and the builders.

# Data

**To figure out this problem, we will need the following data:**

- List of neighborhoods in Austin, Texas. This defines the scope of this project, which is confined to this area, the capital city of the state of Texas in southwestern region of the U.S
- Latitude and longitude coordinates of the neighborhoods. This is required for plotting of the maps and to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

### Data Sources and method extraction

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Austin,_Texas) contains a list of neighborhoods in Austin, Texas. This consists of a total of 61 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of making API calls and beautiful soup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases and has venue data over 190 countries and is used by over 150,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the shopping mall category 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.