

# Coursera Capstone

## IBM Applied Data Science Capstone

### *London, Ontario: A Shopping Mall venture*

Humble endeavor of: Jay S



## Introduction

With the advent of shopping malls as an internationalized idea of the European **piazza**, shopping became more socialized than ever, with people now being able to not only shop, like they used to when frequenting shopping streets or plazas, or dine at restaurants, but do all of that under the same roof, which served as the foundational vision for the creation of the first mall by the adept architect Victor Gruen, who was commissioned to build a shopping center in Edina, Minnesota. But, with the perpetually increasing consumer base and retail brands, it becomes an imperative for a retailer or a marketing strategist for brands to better assess the locale, and then decide which is the **right place to be in and what is the right time**. General baseline for all the retailers planning a market move, is usually, **central location** and **populace with ample buying capacity**. Furthermore, commercial undertakings like malls, also provides a steady rental income stream to the property developers, and with the city of London, Ontario being the epicenter of Canada's top earning shopping areas, according to the **Canadian Shopping Center study** of 2019, it is a jackpot of an opportunity for retailers and property developers alike, but before making the move the question of **WHERE** arises, and the main intent of this project is to tackle that, with the apt use of machine learning paradigms and data science.

## Business Problem

The capstone project intends to ease off the load, of manually analyzing the different neighborhoods in the city of London, ON, and undertake the tedious task of mining the data, to find informative pattern, by utilizing the various machine learning paradigms namely clustering algorithms and data science methodology, and to find answer to the question of “**Where is the right locality for a shopping mall venture**”.

## Target audience of the project

The capstone aims to cater the cohort of property developers, retailers and marketing strategists, by providing elucidating insights about the viability of a locality in sustaining the growth of the mall, furthermore, investors planning to invest in the sector, can also be greatly benefitted by the same. To bolster the relevance of the project undertaken, the city of London, Ontario, Canada, has been described by Canadian shopping center study, to be home to the 14<sup>th</sup> most profitable shopping complex per square foot, CF Maisonville, with the average earning grossing to \$986 per square foot, but at the same time according to the same study also hinting that there is a paucity of such wholesome shopping complexes in the city. The Retail Council of Canada (RCC) also points out that there has been a massive increment in the investment in the shopping complex sector in the years and the trend is likely to be followed , therefore the opportunity of growth of a shopping mall in an area which has a paucity of wholesome shopping complexes, together with the ever increasing investment into the sector, would definitely be an opportunity worth the toil for the said parties.

## Data

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

- List of neighborhoods in the city of London, Ontario, Canada, also renowned as the “forest city”
- Coordinate data for all the neighborhoods, constituting, the latitudes and the longitudes. It would be conducive to construct maps and get the venue data
- **Venue data**, pertaining to the shopping malls, using foursquare API, on which data clustering would be performed.

## Source and extraction methodology of data

The first source of data is the interactive map provided by [https://www.scribblemaps.com/maps/view/Neighbourhoods in London Ontario /Unki2xMmX6](https://www.scribblemaps.com/maps/view/Neighbourhoods%20in%20London%20Ontario/Unki2xMmX6), which essentially contains a mapped layout of the 46 neighborhoods in the city of London, furthermore, the population data for the neighborhoods is to be scraped from the official website of London city namely, <https://www.london.ca/About-London/community-statistics/neighbourhood-profiles>, to append the population and the per capita income data into the database, there will be a blend of strategies involved in the creation of a satisfactory dataset, which would include web scraping using the BeautifulSoup library, together with actually researching and appending population and per capita data, furthermore the coordinates of the same would be found out using the GeoPy library, and lastly the usage of RESTful APIs and Foursquare would yield the **results of the venue data pertaining to the shopping malls** for the neighborhoods which is required for further analysis, gauging the problem statement, one can easily deduce some crucial aspects which must be extracted from dataset, namely **the population** of the neighborhood which **would be crucial in understanding if there is a large enough consumer base to instate a mall for** and the **per capita income** to better assess if the selected demographics is, as the business jargon describe, **a high return target or a low return target**. The business problem at hand would utilize an appropriate mix of data wrangling, data cleansing, data analysis, web scraping, data visualization using Folium library, researching and machine learning methodologies (K-Means Clustering) to be successfully solved. In the following section, we would be describing the plan of action and the steps taken to facilitate all the processes mentioned above.