**Coursera Capstone**

**IBM Applied Data Science Capstone**

***A New Shopping Mall in Cebu City, Philippines***

Johnry B. Morales

January 2019

**Introduction**

Philippines is country with a population of approximately 108 million, with much of the working population concentrated on major metropolitan cities and territories (the National Capital Region, metro Cebu and metro Davao). Filipinos are known for its sociable culture and demeanor with much of the people spending most of their free time on parks, open spaces and shopping malls. This is supported by the fact that four of the largest shopping malls in the world are located in the country; the SM Megamall, SM Mall of Asia, SM Seaside Cebu and SM City North. Visiting shopping malls is a great way to relax and enjoy during weekends and holidays. People can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities or even host events. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. In addition, the government aims to decentralize the already crowded National Capital Region (NCR) and urge the mall owners, developers and other investors to refocus development to other places. With this, property developers are also taking advantage of this trend to build more shopping malls to cater to the demand in the provinces. As a result, building of shopping malls has shifted from the NCR to other major and minor cities to provinces. With the current situation; COVID-19 pandemic, it is a challenge to pursue such project considering the weight of the effect on the economy the pandemic has brought both on the domestic and the international market. However, to be able to restart the domestic economy, the public and the private sector must encourage spending especially on the domestic level in order for the local economy to start-up fresh. With this, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. The location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

**Business Problem**

The objective of this capstone project is to examine, select and calculate the best locations in Cebu City, Philippines to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: considering the existence of large shopping malls like the Ayala Malls and the SM Malls, is there any location in Cebu City that would be suitable in opening a new shopping mall that could attract many people.

**Target Population**

This project is primarily useful to property developers and investors looking to open or invest in new shopping malls in the Philippines. With possible choice of cities in the country, developers and investors has a range of choice on which location, aside from the NCR, to establish their property. This project is timely as the government aims to decentralize both the public and private infrastructure investment to sprawl development in the provinces. Data from the Philippine Statistics Authority pegged the NCR’s density at approximately 27,000 persons/km2, making the NCR as one of the mostly densely populated cluster of cities in the world. Although the concentration of these financial and business establishments in the NCR, relatively, this kind of business is not affected that much considering the behavior of the populace making the prices of real estate not compromised. With the Philippine economy expected to further expand, consumerism in the country will still sprawl therefore the demand for further development such building of establishments.

**Data**

To solve the problem, we will need the following data:

* List of neighborhoods in the city of Cebu. This defines the scope and limitations of this project which is confined to the city of Cebu; a metropolitan city in the Visayas region.
* Latitude and longitude coordinates of these neighborhoods. This is required in order to plot the map and also to get the venue data.
* Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

**Sources of data and methods to extract them**

This Wikipedia page (https://en.wikipedia.org/wiki/Cebu\_City) contains a list of barangays in Cebu City, with a total of 80 barangays with a population of 922,611. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages or if not possible, we will build the data frame manually. Then we will retrieve the geographical coordinates of these barangays using Python Geocoder package which will give us the latitude and longitude coordinates. After that, we will use Foursquare API to get the venue data for those barangays. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, in which in this case, we are particularly interested in the Shopping Mall category in order 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.