IBM Applied Data Science Capstone by Coursera
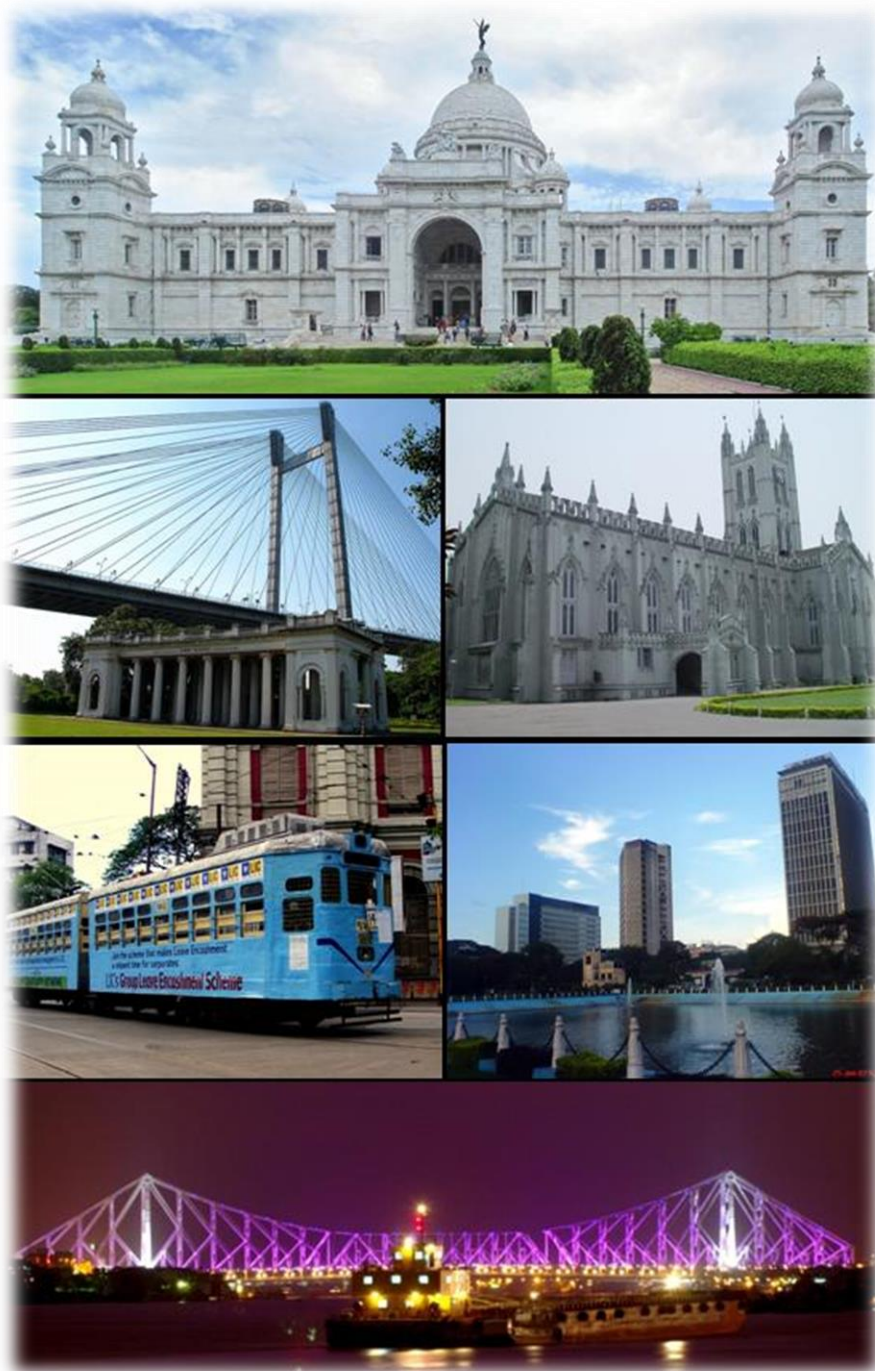
Opening a New Shopping Mall in Kolkata, India

By: **Barnendra Mohan Chaudhuri**

January 2020

Introduction

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. 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. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Kolkata and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, 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 analyse and select the best locations in the city of Kolkata, India 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: In the city of Kolkata, India, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful for property developers and investors looking to open or invest in new shopping malls respectively in the city of Kolkata, the easternmost and one of the four metropolitan cities in India. This project is timely as the city is currently suffering from a disproportionate distribution of shopping malls which leads over-congestion is some parts of the city while other parts of the city are eagerly awaiting shopping malls in their vicinity. This also leads to huge problems in traffic management by the Kolkata Traffic Police as people from all over the city and its suburbs move into the areas having a high density of shopping malls. The other related issue would be provisioning of adequate public transport by the Kolkata State Transport Corporation authorities for the movement of people in and out of the areas with high density of shopping malls which at times leads to taxis/cabs charging exorbitant prices especially during peak hours as well as late night hours.

Data

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

• List of neighbourhoods in Kolkata. This defines the scope of this project which is confined to the city and suburbs of Kolkata which is the easternmost and one of the 4 metropolitan cities of India

• Latitude and longitude coordinates of those neighbourhoods. 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 neighbourhoods.

Sources of data and methods to extract them

The Wikipedia page https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Kolkata contains a list of neighbourhoods in Kolkata and its suburb, with a total of 198 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of **Python requests** and **Beautiful Soup** packages. Then we will get the geographical coordinates of the neighbourhoods using **Open Cage Geocoder** package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use **Foursquare API** to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, 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.