IBM Applied Data Science Capstone

Opening a New Shopping Mall in Greater Manchester, UK.

By: Ahmed Hassan Metered

July 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 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 provide 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 demand. As a result, there are many shopping malls in the city of Greater Manchester and many more are being built. Opening shopping malls allow 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 analyze and select the best locations in the city of Manchester 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 Greater Manchester, 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 to property developers and investors looking to open or invest in new shopping malls in one of the biggest cities in the UK i.e. Greater Manchester. According to the India Times "Like other cities of the UK, it shouldn't come as a surprise that Manchester, too, is home to a pretty good shopping scene. After all, the city's long-running reign as the epicenter of business activities is well known. But for a neophyte, it could prove to be a tad hard to figure out exactly where to go for a worthy shopping experience."

Data

To solve the problem, the following data is required:

- List of Neighborhoods in Greater Manchester. This defines the scope of this project, which is confined to the city of Greater Manchester, located in UK at the Northwestern Europe.

- Latitude and Longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.

- Venue Data, particularly data related to shopping malls. This data will be used in clustering on the neighborhoods.

Sources of Data and Ways of Extraction:

This Wikipedia Page https://en.wikipedia.org/wiki/Category:Areas_of_Greater_Manchester contains a list of neighborhoods in Greater Manchester with a total of 81 Neighborhoods. Web scraping techniques are used to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then the geographical coordinates of the neighborhoods are received using Python Geocoder package which will provide us with the latitude and longitude coordinates of the neighborhoods. After that, Foursquare API is used to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API provides many categories of the venue data, we are particularly interested in the Shopping Mall category to help us 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.