Coursera Capstone Project

IBM Applied Data Science Capstone

Opening up a new Cineplex in London, United Kingdom

Author: Niranjan Ganesan

August 2019

Introduction

Whenever our favourite movie releases we always wish to watch it in its first day of release. Cinema is not just an entertainment any more, it has become an integral part of our lives where we spend time with our loved ones. From 5-year-old kid to 90-year-old person, cinema has been the No.1 entertainment factor and not to forget that it has also become an inspiration to many. The demand for Cineplex in more crowded areas has always been high. Property developers can take this as an advantage to build new Cineplex theatres to attract new customers but choosing the location of building the new Cineplex has always been challenging. The new Cineplex must be built at more crowded areas in the city where other Cineplex's doesn't exist. Determining this location is very challenging and contributes to success of the new Cineplex being built.

Business Problem

The objective of this project is to analyse and select the best locations in the city of London, United Kingdom to open a new Cineplex. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In a multi-cultural city like London, if a property developer is looking to open a new Cineplex, where would you recommend, 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 Cineplex in the Capital city of England. General statistics shows that number of Cinemas sites in the UK has been increased steadily over a period of time. In 2017 UK has a total of 774 cineplex's in the country. The country's continued obsession with Cinemas will always welcome new Cineplex's in the city.

Data

To solve the problem, we will need the following data

- List of Neighbourhoods in London. This defines the scope of this project which is confined to the city of London.
- Latitude and Longitude coordinates of those neighbourhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to Cineplex. We will use this data to perform clustering on neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page(https://en.wikipedia.org/wiki/List_of_areas_of_London) contains a list of neighbourhoods in London. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of python requests and beautifulsoup packages. Then we will get the geographical coordinates using python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

Then, we will use Foursquare API to get the venue data for those neighbourhoods. 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 we are particularly interested in the Cineplex category in order to help us solve the business problem. This project will make use of many data science skills from web scraping, working with API, data cleaning, data wrangling, to machine learning and map visualization. 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.