Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Cinema, London, United Kingdom

Bharath

Jun, 2020

Introduction

London is one of the main entertainment avenues in the United Kingdom and in western Europe. Millions of people from around the world visit London each year and they look for various entertainment outlets in London. Local population too is very vibrant and one of the main entertainment for this huge influx of people is visiting Cinemas. London too has a prolific movie industry and many Hollywood blockbusters use London for their movie premieres. Movie going is an experience in itself which many people experience week by week. Opening of a cinema hall is a huge investment from finding the ideal location and installing state of the art equipment. To make a profitable investment, finding an ideal location is crucial.

Business Requirement:

The objective of this capstone project is to analyse the councils and neighbourhoods within the Greater London area to open a new cinema hall in a council which has greater prospect of return on investment. Using data science methodology and machine learning techniques, the project will aim to answer the business requirement. Though there are various features that can be considered to solve the defined requirement such as looking into population in the council, population density in the council etc… the approach taken here is to identify existing councils with existing entertainment avenues and identify a neighborhood in the cluster.

Data:

Following data will be used in the analysis:

- List of councils/boroughs in Greater London
- Latitude and Longitude coordinates of the neighbourhoods within the councils
- Venue data of existing cinema halls

Wikipedia page (https://en.wikipedia.org/wiki/List_of_areas_of_London) will be used as the main data source.

Following actions will be performed on the data source:

- Python Requests library will be used to extract the data from the Wiki page and BeautifulSoup module will be used to scrap the data from the response.
- The geographical coordinates are retrieved using Python Geocoder package
- Foursquare APIs will be used to retrieve the venues in the neighbourhoods.
- With the retrieved data, pre-processing will be done and using K-Means Clustering algorithm clusters are defined