Coursera Capstone

IBM Applied Data Science Capstone

Developing a commercial building in Canberra, Australia

By: Vijaya Bhusan

August 2019

Introduction

For many business, having buildings at commercial place is a great way to increase and scale their businesses by networking clisents, events and business partnership. They can travel for deals & meetings without hassle, manage any problem occured, host clients, and increase market share. Commercial space are like a one-stop destination for all types of business. For businesses, the central location and ease of transport provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more office spaces for booming startup culture to cater to the demand. As a result, there are many commercial spaces in Canberra and many more are being built. Opening commercial spaces allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new commercial spaces requiring to analyze location, demand and services to be provided. Particularly, the location is most important decisions that will determine whether it will be lucrative to business.

Business Problem

The objective of this capstone project is to analyse and select the best locations in Canberra, Australia to open. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: for location in Canberra.

Target Audience of this project

This project is particularly useful to start-ups, businesses, distributors and investors looking to open or plan to acquire and rent building spaces in . This project is timely as the city is currently suffering right palce and buildings for rental property. Data suggest that there will be modest increase in property and rental prices in 2019 and occupancy rate is also falling due to continued obsession with building more shopping space despite chronic oversupply.

Data

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

- List of neighbourhoods in Canberra. This defines the scope of this project which is confined to suburbs of Canberra, Australia in Oceania.
- 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 building and office. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Canberra) contains a list of neighbourhoods in Canberra, with a total of 124 neighbourhoods. 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 of the neighbourhoods using Python 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.