Coursera Capstone Project

IBM Applied Data Science Course

Opening a new Shopping Mall in Sydney

By: Akash M

Introduction

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and breaks. 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 terminus for all types of shoppers. For vendors, the central location and the large crowd at the shopping malls provides a great delivery 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 world now and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. As with any commercial decision, opening a new shopping mall requires serious thought and is a lot more complicated than it seems. Mainly, the location of the shopping mall is one of the most important decision that will determine whether the mall will be a success or a failure.

Business Problem

The aim of this capstone project is to analyse and choose the best places in the city of Sydney to open a new shopping mall. Using data science methodology and machine learning methods like clustering, this project aims to provide solutions to answer the business question: If a property developer is looking to open a new shopping mall in the busy city of Sydney, where would you recommend that they open it?

Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the city of Sydney.

Data

- List of neighbourhoods in Sydney. This defines the scope of this project which is confined to the city of Sydney, the state capital of New South Wales and the most populous city in Australia.
- 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

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_Sydney_suburbs)contains a list of neighbourhoods in city of Sydney. 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. The project makes 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).