COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a new Hotel in Sydney, Australia

By Naci ARSLAN

Graduated 2020 from Istanbul University – School of Business

August 2020

Sydney, Australia

Contents

1.INTRODUCTION

1.1 BACKGROUND OF THIS CAPSTONE

Hotel is the place where we can stay with the best amenities without any worries. Generally, most people would like to go to hotels to relax and take a load of their busy lives. Such establishments offer luxury & comfort, along with other benefits that can make a trip enjoyable for everyone that is involved in the proceeding. And people would like to rest during a weekend away in this case hotel could be a good option. If the hotel is close to the airport, some people may use the hotels to avoid missing their flight.

Staying in a hotel is always an exciting occasion, especially for children as they tend to compare it to an adventure. Also Due to the repetitiveness of everyday life, we want a better silently time. In this case, people could like to go hotel. Because in the hotels due to cleanness, peacefulness, changeable day they can feel very nice.

1.2 BUSINESS PROBLEM

The objective of this capstone project is to analyze and select the best locations in the city of Sydney, Australia to open a new hotel. For finding a solution we will use a data science methodology and machine learning techniques like clustering, and this project aims to provide solutions to answer the business question

1.3 TARGET AUDIENCE

This project is useful to the investors and who are interested in looking at a nice location to open or invest in a new hotel in the capital city of the New South Wales Territory, Australia i.e. Sydney."https://en.wikipedia.org/wiki/List_of_Australian_capital_cities" The best area for first-time visitors to Sydney: Darling Harbour in the Central Business District with its many waterside cafés, bars, and restaurants could be the nice option.

2. DATA ACQUISITION AND CLEANING

For solve this problem, we need the do the following steps as shown below:

We need to define the scope of this project which is confined to the city of Sydney, the capital of New South Wales Territory, Australia.(*List of Australian Capital Cities - Wikipedia*, n.d.)

For that, we need a list of neighborhoods in Sydney. (*Category:Suburbs of Sydney - Wikipedia*, n.d.)

Data scraped from sources. Now we need latitude and longitude coordinates of neighborhoods. This is required to get the venue data and in order to plot the map. Venue data, specifically data related to hotels. We will use this data to perform clustering on neighborhoods.

2.1 SOURCES OF THE DATA AND METHODS TO EXTRACT THEM

This data is we scraped from Wikipedia page. contains a list of neighborhoods in Sydney, with a total of 200 neighborhoods. We will use a web scraping technique to extract the data from the Wikipedia page, also we will use the help of Python requests and beautifulsoup packages. Afterward, we will get the geographical coordinates of the neighborhoods using Python geocoder package which will help us to latitude and longitude coordinates of the neighborhoods.

Subsequently, we need to get venue data of those neighborhoods and for that, we can use Foursquare API which has one of the largest databases of 105+ million places and is used by over 125.000 developers. It will provide many categories of the venue data, and in these data, we will focus on the Hotel category, due to help us to solve this business problem put forward.

This project mostly will help us with scraping the data, working with Foursquare API, data cleaning - wrangling, and also we will use machine learning techniques like K-Means clustering and or map visualization we'll get benefits from the Folium. In this next step, we will discuss the Methodology.

3. REFERENCES

Category:Suburbs of Sydney - Wikipedia. (n.d.). Retrieved August 26, 2020, from
https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney

List of Australian capital cities - Wikipedia. (n.d.). Retrieved August 26, 2020, from
https://en.wikipedia.org/wiki/List_of_Australian_capital_cities