



COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a new Hotel in Sydney, Australia

By Naci ARSLAN

Contact : <https://arслан0007.github.io/>

August 2020



Sydney, Australia



August 2020



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.

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.



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

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. (*List of Australian Capital Cities - Wikipedia*, n.d.) 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.

August 2020



DATA ACQUISITION AND CLEANING

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.



METHODOLOGY OF CAPSTONE

Firstly we need the data to get the list of neighborhoods in the city of Sydney. After that, we will do scraping the data using Python requests and beautifulsoup packages to get the geographical coordinates as a latitude and longitude due to we will use Foursquare API. In this process, the Geocoder package will allow us to convert an address into the geographical coordinates in the form of latitude and longitude.

METHODOLOGY OF CAPSTONE

After that, We will populate the data into a pandas DataFrame with the pandas' library then we will use the Folium package to visualize the neighborhoods on a map. This will allows us to perform a sanity check to make sure that the geographical coordinates data returned by geocoder are correctly plotted in the city of Sydney



METHODOLOGY OF CAPSTONE

Now time to use API to get the top of 100 venues that are within a radius of 2000 meters. Foursquare API will help us to get this data. For that, we need to register a Foursquare Developer Account due to obtaining the Foursquare ID and Foursquare Secret Key. Then we will be able to make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop.

METHODOLOGY OF CAPSTONE

Analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. So at the same time, we are also preparing the data for use in clustering. We will filter the "Hotel" as a venue category for neighborhoods.



METHODOLOGY OF CAPSTONE

Terminally, we will perform clustering on the data by using K-Means clustering. K-Means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is particularly suited to solve the problem for this project.

METHODOLOGY OF CAPSTONE

In this case, we will cluster the neighborhoods into 3 clusters based on their frequencies of occurrence for "Hotel". Then the results will allow us to identify which neighborhoods have a higher concentration of hotels while which neighborhoods have a fewer number of hotels. Based on this occurrence of hotels, it will help us to answer the question as to which neighborhoods are most suitable to open a new hotel.



RESULTS

The results of this data from the K-Means clustering show that we can categorize the neighborhoods into 3 clusters based on their frequency of occurrence for "Hotel". We can clarify these 3 clusters like as shown below:

RESULTS

Cluster 0: It is representing neighborhoods with a moderate number of hotels

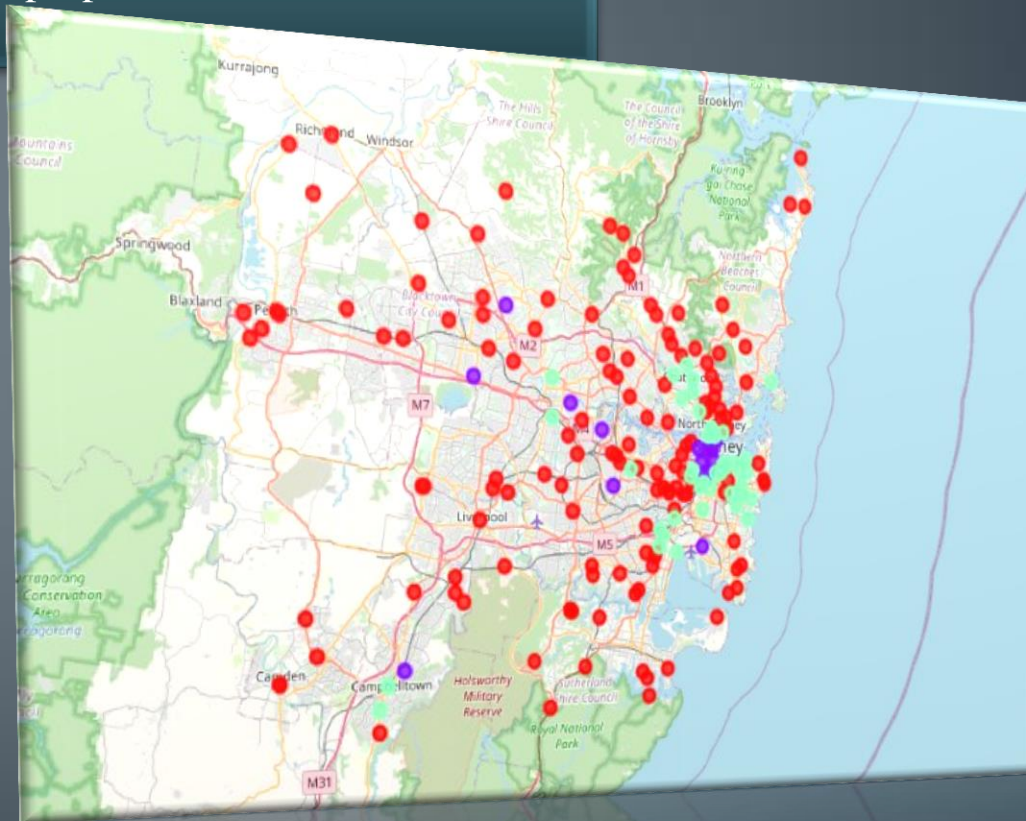
Cluster 1: It is representing of neighborhoods with a low number to no existence of hotels

Cluster 2: It is representing neighborhoods with a high concentration of hotels.



RESULTS

As we can see from the results of clustering are visualized in the map below with cluster 0 in red color, cluster 1 is in purple, and cluster 2 is in mint green color.



Sydney, Australia I: Clustering with Machine Learning

August 2020



DISCUSSION

As we can notice from the map in the Result section, we can see the highest number in cluster 2 is the most of hotels are concentrated in the central area of Sydney. If we look at other clusters such as cluster 1 has very low numbers to no hotels in the neighborhoods and moderate number in cluster 0.

DISCUSSION

Meanwhile, hotels in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of hotels, also in my research of that area for rent a room in a hotel. As I saw in forums; also central city has a parking problem, in this case, these can be effect badly to choose a hotel in this areas.



DISCUSSION

If we look another perspective, the result is also showing that the oversupply of hotels has mostly happened in the central area of the city, but as shown from the map also other areas have hotels but not as often in the central area. Hence, the project will recommend property developers to capitalize on these findings to open a new hotel in cluster 1 without petty competition.

DISCUSSION

Real estate agents with unique selling propositions to stand out from the competition can also open a new hotel in cluster 0 areas which are moderate competition. Eventually, real estate agents or investors are could be advised to avoid neighborhoods in cluster 2 which already have a high concentration of hotels.



CONCLUSION

First of all, we were through the process of identifying the business problem and afterward specifying the data as required, extracting as we need it, and also preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. real estate agent or investors regarding the best locations to open a new hotel.

CONCLUSION

As a suggestion for this situation; as shown from the map we seeing the neighborhoods in cluster 0 are the most preferred locations to open a new hotel. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new hotel.

August 2020



EXTRA RESEARCH AND COMBINE WITH THE CAPSTONE PROJECT

But if we follow this way and, also make extra searches such as which area is a good option for fun, and also good for parking in Sydney, etc. Then we can combine these areas with the result of this machine learning project. On that occasion, our decision would be more sensible to choose the location of the hotel area. And of course, we shouldn't forget the owner's decision, his decision could be change as depends on his taste. For that, we can separate areas such as Restaurants, Sightseeing, Families, Nightlife, etc

Circular Quay and Rocks, it shown in the map as Cluster 1 which means has very low numbers to no hotels in the neighborhoods and these represent high potential areas to open a new hotel.

King Cross on the map it is showing as a Cluster 2, and Cluster 2 is likely suffering from intense competition due to oversupply and high concentration of hotels.

Note:

And we can make this project more detail.

August 2020



LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this capstone, we only contemplate one factor which is the frequency of occurrence of hotels, also other factors exist like; parking areas, does the location is close to the airport? or shopping centers, an income of residents that could influence the location decision of a new hotel and population.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

For future research, the researcher can follow a methodology to estimate data to be used in the clustering algorithm to determine privileged locations to open a new hotel. Besides, this capstone letting us use free Sandbox Tier Account of Foursquare API which is have limitations as to the number of API calls, and with this way results are returned.

August 2020



REFERENCES

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

Foursquare Developer. (n.d.). Retrieved August 26, 2020, from <https://developer.foursquare.com/>

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

Where To Stay & Best Areas in Sydney – Updated for 2020. (n.d.). Retrieved August 26, 2020, from <https://santorinidave.com/best-places-sydney>



THANK YOU

By Naci ARSLAN

Graduated August 2020 from Istanbul University – School of Business

Contact : <https://arслан0007.github.io/>

August 2020