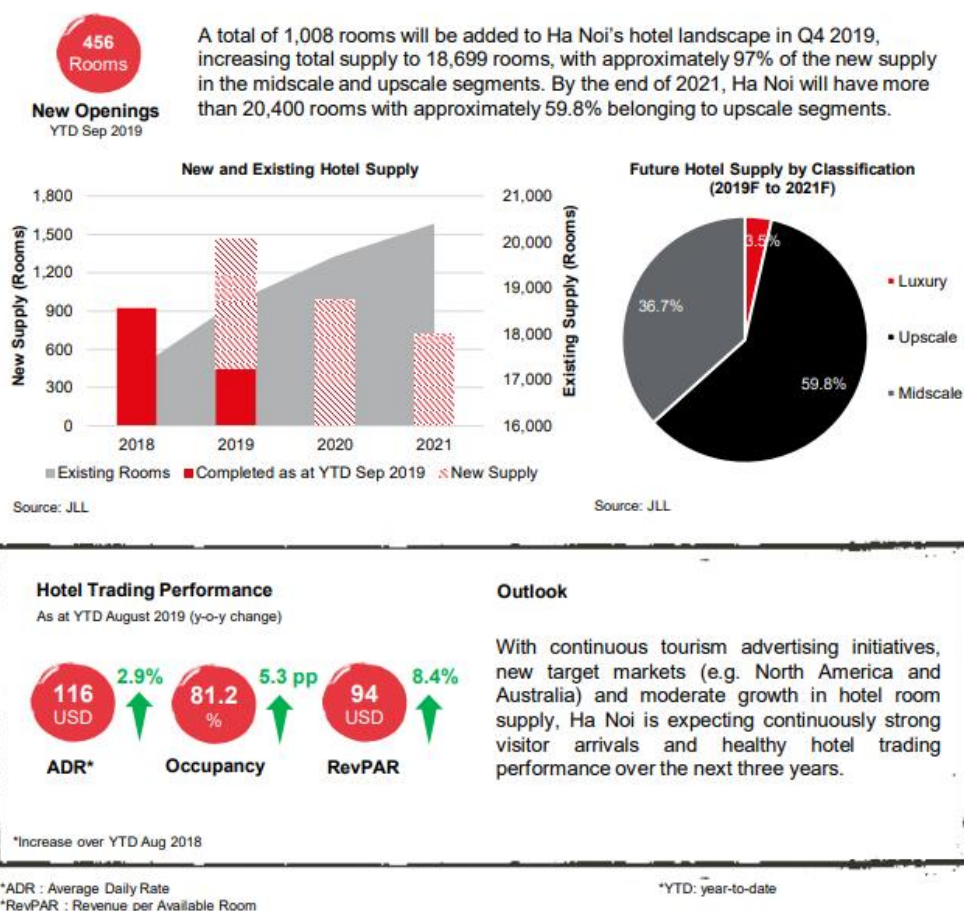


Coursera Capstone – Week 4

Opening a New Hotel in Ha Noi¹, Viet Nam

Problem

- The objective of this capstone project is to analyse and select the best locations in Ha Noi, Viet Nam to open a new hotel. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Ha Noi, if a property developer is looking to open a new hotel, where would you recommend that they open it?
- This project is particularly useful to property developers and investors looking to open or invest in hotels in Ha Noi. This project is timely as the city is currently suffering from oversupply of hotels.
 - o Data from Hotel news now²:
 - During the first two months of 2019, Hanoi reported double-digit increases in ADR (+10.1% to VND 2,856,659.12) and RevPAR (+14.0% to VND 2,200,763.03). Occupancy rose 3.5% to 77.0%, driven by a 4.1% jump in demand. The market experienced slight RevPAR growth in 2018 (+0.4%) once again pushed by ADR (+3.7%).
 - There are 224 hotels accounting for 17,615 rooms in Hanoi. The market continues to remain full of smaller hotels, with almost 75% of all hotels at 100 rooms or fewer and 60% of all hotels with fewer than 50 rooms. There are roughly 3,000 rooms in the development pipeline, with fewer than 1,000 rooms expected to open in 2019.
 - o Data from JLL Vietnam³:



¹ <https://en.wikipedia.org/wiki/Hanoi>

² <http://www.hotelnewsnow.com/Articles/294611/STR-Positive-outlook-for-Hanoi-hotel-industry>

³ <https://www.joneslanglasalle.com.vn/content/dam/jll-com/documents/pdf/research/apac/vietnam/jll-vn-hotel-market-snapshot-hn-dn-hcmc-sept-2019-en.pdf>

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

- List of neighbourhoods in Ha Noi. This defines the scope of this project which is confined to the city of Ha Noi.
- 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 hotels. 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:Districts_of_Hanoi) contains a list of neighbourhoods in Ha Noi, with a total of 30 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 API will provide many categories of the venue data, we are particularly interested in the hotel 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).