Capstone Project – Brief Report

Introduction:

If you are travelling to a city such as Sydney and looking to book an Airbnb residential property or similar, where should you choose?

It is actually people's common-sense that choosing somewhere either close to the CBD or tourist destinations would be a good option, if not considering the costs, however, are there might be other spots that you could get both affordable accommodation and convenience (for both travel and general shopping). Under this particular project, only the surroundings are considered. (limited to data sources and effort spent in this project)

I believe that many people have experienced the difficulty in finding an ideal place that he/she can be renting, especially when the person is either travelling or studying or just started working in a different city.

Usually, people would consider their valuation matrix that includes: location(including the commute to destination/work/study, nearby shops, etc.), surrounding environment(noise, crime rate, etc.), internal design, rent costs, etc. (Considering the fact that this is just a small project, the plan is not to involve too many variables(e.g., CNN can be used to analyse the internal that help user to filter better designs, collecting the crime data, or accurate estimation of commute time all require a significant workload and limitation on data availability). This project will be mainly utilising available data from Foursquare and some other public data.)

Data Source:

Data Source 1: Transport Management Center (Sydney, NSW): to find out where the spots for public transportation are. The dataset is actually the data during even the New Year's Eve that those public transport are still running, in other words, they are the "key transportation hub/ interchange" for areas close to Sydney central area. Data Source 2: Foursquare: given any particular latitude + longitude, find the nearby facilities that would help "make life more convenient". Using the nearby facilities that can be found with FourSquare API, we can evaluate whether the nearby surroundings close to the rental property is actually convenient or not. Data Source 3: Airbnb price and location data (from Kaggle) With the dataset from Airbnb, we could have an estimation of the general price point for the areas that we are investigating.

Methodology:

Using Airbnb price data as a benchmark, we could obtain an estimated price-level and supply condition of residential properties in Sydney Suburbs.

Using the public transport dataset, we could get location data of "some benchmark points" where the surrounding facilities can be searched in Foursquare.

Based on the nearby facilities, we could then classify the locations into different groups, with which the difference in nearby facility nature/ groups can be classified.

The data from Airbnb can be borrowed here to advise the average price per suburb in Sydney, so as to assist decision for choosing areas to rent in Sydney.

Results:

Contrary to the anticipation that CBD area would be much more costly, these areas are actually quite affordable compared to "rich nearby suburbs". (note that the difference in property type is not considered here, also less space/ area for CBD accommodations.)

North Sydney appears to be the only area where both convenience level and affordability can be allocated into the same group as the Sydney CBD area.

Discussion:

The results indicate that both CBD area and North Sydney are areas where both cost and convenience requirements are fulfilled. (This is contrary to the fact that these areas would have higher property value.) (And one potential explanation is that the higher occupancy rate and competition in these areas are making keeping the rate at a relative-affordable level.)

If using the location data of these Airbnb rental properties into the location data in the first place, the result would then be changed into a classification of Airbnb properties by their convenience level.

Conclusion:

Based on the price level and nearby facilities, we could recommend locations that are both affordable and convenient for tourists.

If adding other user-level data and user-preference data, we could potentially build a recommender system for travellers/ international or interstate students and other rental property searchers.

There are many other research aspects(such as combining the comments from users to better quantify the merits and demerits of each location) that we can investigate into, and many classifications (different property type, different season influence, etc.) that can be done to make the project more "meaningful".

Based on the venue_id selected, there can be further analysis of the venues such as the number of likes, ratings, price-point, etc. I am postponing here to work on other projects.