Exploring Airbnb Listings in Sydney

1. **Problem and background**

As aworld leader in accommodations of the “sharing economy”, Airbnb is a famous platform that allows travellers to find places to stay directly from individuals in thousands of cities around the world. Since I live in Sydney, which also happens to be a popular tourist destination, I am interested in finding out what the listings are like in this city. Specifically, I would like to find out similar listings by using the clustering algorithm learned from this course. This finding can then be used to make customized recommendations to travellers who are looking for a place to stay, based on their specific situation like budget, preferred room type, preferred nearby facilities, etc. On top of that, the finding will also provide such insights as what is a reasonable price for people who plan to list their properties on Airbnb.

1. **Description of the dataset**

The dataset was obtained from Kaggle’s Sydney Airbnb open data, originally sourced from publicly available information from the Airbnb site. The dataset consists of several csv documents, but the one of interest to me is the listings\_summary\_dec18 file, which contains all the essential information needed for this project, namely, name of listing, neighbourhood, latitude and longitude, room type, price, reviews per month, and availability.

As of December 2018, there have been more than 30,000 listings in the city of Sydney. To facilitate the analysis, I have decided to narrow down my research to only the CBD area, i.e. neighborhood with the name Sydney. Furthermore, only listings that have at least four reviews per month are considered. After applying these conditions, I am left with a dataset with only 847 listings, which is more feasible to explore.