Master of IT in Business
Capstone Project Poster

(Airbnb's impact on hotels and housing in Singapore)

(Clara Chua Kiah Hwii)

(Professor Kam Tin Seong)

SMU Singapore Management University

School of Information Systems

Motivation

Price is an important factor of why people choose Airbnb accommodation over the typical holiday accommodations. They may also value cultural exchanges with hosts, or urban / lifestyle tourism (e.g. food tourism, medical tourism, concerts). Thus, it helps to understand the characteristics of Airbnb that make it attractive to users.

As there are few spatial studies of Airbnb in Asia, we are also interested in looking at the spatial distribution of Airbnb listings in Singapore to discover if they are clustered in specific areas, and how this could inform hosts' pricing decisions, or impact areas hotel revenues, pricing of long-term rentals.

Scope of Work

- The project looks at geospatial analysis to explore and explain the data around Airbnb rentals in Singapore, namely to:
  - Understand the spatial distribution of Airbnb listings in Singapore
  - Develop a hedonic pricing model to explain factors affecting Airbnb pricing.
- The scope of work includes:
  - Review of spatial studies involving spatial distributions of Airbnb listings and its impact, geographically weighted regression (GWR) models for Airbnb pricing, and review of Asian based studies for Airbnb.
  - Exploratory data analysis (EDA) and spatial data analysis (ESDA) on Airbnb listings in Singapore, including Spatial Point Pattern Analysis.
  - Developing a hedonic (GWR) pricing model of Airbnb listings

Research Methods

Data

Inside Airbnb provides a snapshot of listings, scraped from the Airbnb website on a regular basis. We used data for Singapore, compiled on 22 June 2020.

R Packages: The project was done in Rmarkdown with the key R packages: tidyverse, GWmodel, spatstat, sf, purrr

Spatial Distribution:

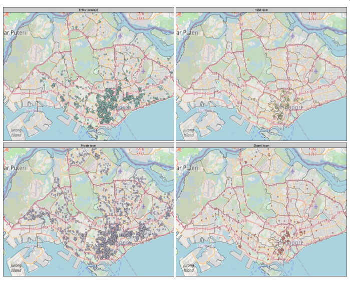
We used Spatial Point Pattern Analysis (SPPA) to discover if Airbnb listings are clustered in Singapore and where they occur. We use the K-function to test whether there is clustering or competition between listings, and the Kernel Density Estimation to ascertain where there are high intensities or densities of listings.

Hedonic Pricing Model:

We used Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR) to examine the price determinants of Airbnb listings.

Results: Spatial Distribution

The listings are clearly clustered around the Central Business District (CBD) and main shopping district (Orchard).

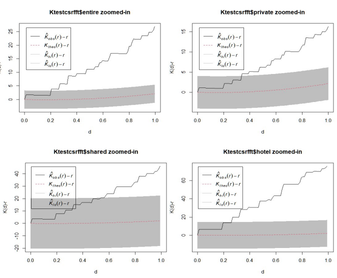


- Entire home/apartment listings are mainly in Kallang, Rochor and Novena area, close to the CBD and Orchard.
- Private room listings also show high concentration in the same neighbourhoods, and also in Outram, Bedok and Rochor, and more spread out within the neighbourhoods.
- Hotel room listings are mainly in city fringes of Outram, Kallang and Singapore River
- Shared room listings are concentrated in the Kallang and Rochor areas

Spatial Point Pattern Analysis

Using Ripley's K-function test, we determined that the distribution of Airbnb listings formed clusters at the radius shown below:

Entire home/apt	0.3km
Private rooms	0.5km
Shared rooms	0.6km
Hotel rooms	0.25km



Hedonic Pricing Model

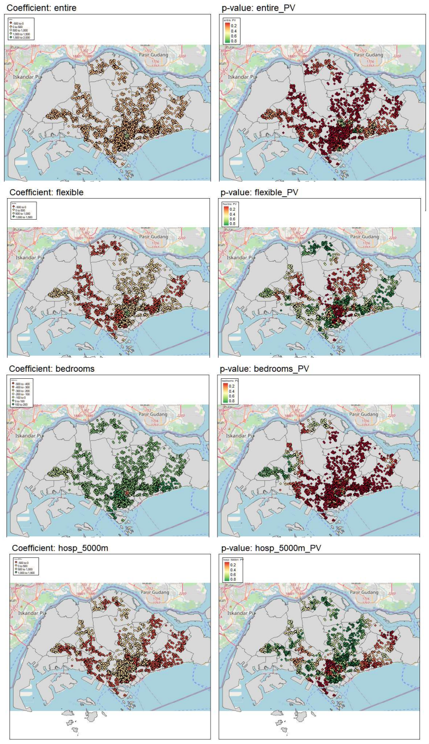
We conducted model selection using GWmodel for the gaussian and bisquare kernel (most commonly used) and both the AIC and CV approach when modelling the data. As expected, the GWR models outperform the generalised OLS model. We then selected the model with the best adjusted R^2 as the best GWR model and compared them with the OLS model:

Model performance

As expected, we see that the GWR outperforms the generalised OLS model. The table shows the results of the model:

Model	AIC	AICc	R2	Adjusted R2	RSS
OLS	109972.6	109972.6	0.04739	0.04621	1568875980
GWR (best)	108488.9	108763.7	0.2476165	0.2098307	1239122141

The following gives the results of the localised regression model with coefficients and p-values of the variables plotted:



Entire homes/apartments

Entire home listings is 'global' variable that raises the price of the listings, regardless of where it is located.

Flexible Cancellation Policy

Flexibility affects pricing positively in the central region (Kallang, Geylang, Downtown Core), and some outer regions (e.g. Serangoon, Hougang). Outer regions are more attractive with more flexibility, while central regions have more competition.

Number of bedrooms

No. of bedrooms is generally significant, corresponding to the number of guests a rental can accommodate, except a few small concentrations in the northern most tip and west regions (Jurong West, Bukit Batok).

Number of hospitals (5km)

Significant in the Tanglin, Newton neighbourhoods, corresponding to private hospitals (healthcare tourism) in those areas. This variable did not surface in the OLS model.

The other significant variables not shared here are hotels_250m, hotels_500m and malls_2000m

Project Challenges / Lessons Learnt

Key challenges & Lessons Learnt:

- *Analysis of a large datasets and large observation window*
  - We had to exclude areas that were not zoned for residential housing and dissolve boundaries to be able to run the simulations.
  - One solution was to learn how to parallelise operations using other tools as spatstat was not configured for parallelization. Parallelisation also had to be OS agnostic, which learning how to use packages like do.Parallel and foreach to get the simulations to run.
- *Lack of RAM memory for running regressions.* Lack of hardware made running regressions, especially the Multi-scale GWR impossible to run without spinning up a virtual machine (VM) with better hardware investments. We learned how to spin up and use Rstudio server on Azure to be able to run some of the GWR and model selections in a timely manner.
- Perseverance pays off: crucial to persevere to get the results. The best results were gotten when we went back to basics to understand the functions and issues, before building the code incrementally.

Acknowledgements

Special thanks to Prof Kam who has provided guidance along the way

References can be found in the project reports