Visualisation to Depict if AirBnB is worth Property Investment

Dhruvisha Gosai (JC810547)

MA5830 – Data Visualisation, James Cook University

Abstract

In recent years, the hotel industry has seen a consistent decline in demand with a rapidly growing Airbnb community with competitive rates, experiences and assisting small businesses and localities a way to earn extra money. Accommodating around 124.5 thousand guests in Hobart and regional towns alone, Airbnb guests have become a significant driver in contributing to Tasmania's Gross State Product (GSP) by $55 million in 2015-16 and supporting 600 jobs in addition to the hosting activities. (O'Mahony, Simes, Hill, Liu, & Alvaro, 2017). However, in 2018, Tasmania faced an unexpected challenge with low supply for long-term tenancy where the government had to offer landlords almost $13,000 to entice them to opt for long-term rental although Airbnb had a much better return via hosting than leasing their unit. (Bleby, 2018) With this newfound investment stream poses a unique question to the investors – is Airbnb's short-term rental a better property investment than long-term tenancy and if so, is there still demand short-term rentals?

Introduction

Assuming the property investor owns multiple properties, they are potentially looking to invest in Tasmania. Although not having formal statistical knowledge, with part of their day consumed by looking at various visualisations to assess the risk and potential growth opportunity, they still have a good grasp on forecasting and economic principles for future growth.

To provide the property investor with sufficient depth and overview of the key performance indicators (KPIs) for Airbnb listings in Tasmania, a dashboard featuring informative visualisations has been created.

This information will enable the investor to make an informed decision to discover and verify their hypothesis whether to invest in hosting activities on Airbnb and what sort of investment would ideally yield low vacancy rates and high return.

Key emphasis is placed on the following enabling the investor to decode the required information –

- **Rental yield**: Understanding the average listing price by area/locality. This is achieved by creating a map of Tasmania and a Z-score bar graph that depicts the suburbs with listing price much greater or lower than the *mean* listing price across the whole state.
- **Vacancy Rate**: Using average number of reviews per month plotted on boxplots to establish the vacancy rates, assuming the majority of the Airbnb users leave a review for the host after staying over.
- **Type of property** to invest in: Comparing the average price of a room type in the property against the vacancy rate of that room type would assist in narrowing down the most ideal property type to list on Airbnb.

Data

For this report, provided Airbnb data is used comprising of the files calendar.csv, listings.csv, listings_summary.csv, neighbourhoods.geojson (spatial data) and reviews.csv. These are observational datasets detailing the particulars of Airbnb listings, their reviews, availability, and location details in Tasmania over a period. A total of 4 csv files were joined in Tableau to create 2 datasets – listings.csv + neighbourhoods.geojson and calendar.csv + listings_summary.csv; to be able to use the variables interchangeably for both tables for various visualisations. There are 22 variables in total with 12 derived variables, used for visualisation purposes. A list of used variables is given below with vis worksheet number on Tableau for reference –

Field Name	Derived	Dataset	Description	Format	Work - sheet
Longitude (generated)		Neighbourhood	Auto-generated coordinates on Tableau from GeoJSON file	Integer	1
Latitude (generated)		Neighbourhood	Auto-generated coordinates on Tableau from GeoJSON file	Integer	1
Area Price	Derived	Listings	Average price by locality	Numeric	1
Neighbourhood Price	Derived	Listings	Average price by suburb	Numeric	1
Geometry		Neighbourhood	Defined suburb boarders provided in the dataset	Character	1
Entire Place Rate	Derived	Listings	Price derived only for room_type – “Entire home/apt”	Numeric	1
Hotel Room Rate	Derived	Listings	Price derived only for room_type – “Hotel room”	Numeric	1
Private Room Rate	Derived	Listings	Price derived only for room_type – “Private room”	Numeric	1
Shared Room Rate	Derived	Listings	Price derived only for room_type – “Shared room”	Numeric	1
Difference from Region	Derived	Listings	Difference between locality average price and suburb price	Numeric	1
Above or Below?	Derived		Determine if “Difference from Region” is above or below suburb price	Character	1
Average Price	Derived	Listings	Average of total Area price	Integer	2
STDEVP Price	Derived	Listings	Standard deviation of total Area price	Integer	2
Z-score	Derived		(total area price – average area price)/ “STDEVP Price”	Integer	2
Neighbourhood Cleansed		Listings	Suburb Name	Character	2
Room Type		Listings	Type of room listed – e.g., Entire home, hotel room, etc.	Character	4
Property_Type_Banding	Derived	Listings	Summarised or rolled up level of “Property Type”	Character	4
Price		Listings	Price of listing	Numeric	1,2,4
Room Type		Listings_Summary	Type of room listed – e.g., Entire home, hotel room, etc.	Character	3
Number of Reviews		Listings_Summary	Number of reviews for each listing	Numeric	3
Neighbourhood		Listings_Summary	Suburb Name	Character	3
Date		Calendar	Date of the listing availability	Date	3

Methods & Results

All the data manipulation for creation of new variables was performed in Tableau version 2020.4.0. Relationships were created for 4 imported datafiles to create 2 final datasets - listings.csv + neighbourhoods.geojson and calendar.csv + listings_summary.csv; by joining on the unique key. This enabled the viewing of the data together but to use the variables interchangeably from both datafiles, they had to be logically *joined*. Once the joins were created for both datasets, listings.csv + neighbourhoods were used to create visualisation for worksheets 1, 2 and 3, and calendar.csv + listings_summary.csv dataset was used to create visualisation for worksheet 4.

1. Geomap (Worksheet 1): It is of interest to understand what regions of Tasmania have higher or lower average listing price and compare it to other suburbs. A geomap is created using the neighbourhood spatial data to show the locality price compared to the suburban price of the Airbnb listings. This will allow the investor (who is a consumer) to survey what localities will provide what sort of rental yields.
   Latitude and longitude coordinates generated from the spatial data is used as row and column respectively and average *Area Price* is used as measure with channel: colour (hue) to create a geomap with high saturation colour for either extremes of higher or lower price. Colour scale of "Red-Blue-White Diverging" is used to assist in decoding of the information by the investor (who is a consumer) and a filter of average price range is added to assist the investor quickly filter out the suburbs yielding lower rental income. A **tooltip** is added specifically to provide a breakdown of what is the average area/locality price and compare it to the neighbourhood/suburban average of the listings to assist in informing whether the average listing price has increased or reduced compared to the suburb. Additional details of rates for room type are provided as a quick breakdown of what sort of room type is causing the price of entire locality to go higher or lower.

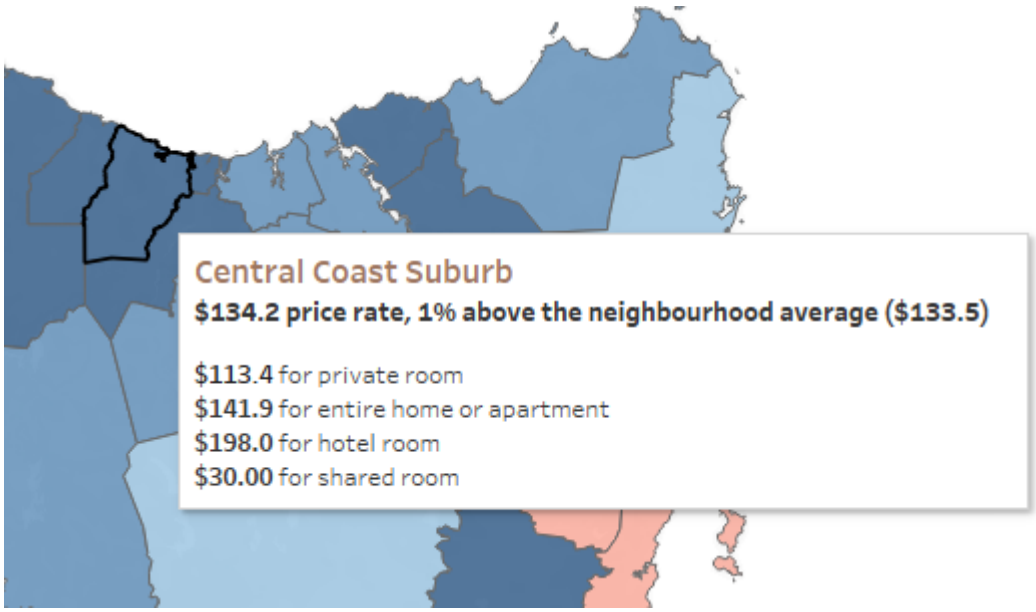

Figure 1: Tooltip overview for visual 1

2. Horizontal bar graph (Worksheet 2): To further deep dive into what suburbs have higher or lower listing price than the *mean* of total listing prices in Tasmania, a Z-score bar graph is created. We have a distribution of z-scores broken up by individual suburbs. This is intended to help the investor shortlist the suburbs that fall between the z-score of -2 and +2. Anything beyond that can be expected to have significantly higher or lower average listing price compared to the whole state. In this case, Kingsborough and Huon Valley have significantly higher average listing price than any other suburb. A potential speculation could be that there are not many listings in that region, or the value of these properties is higher.
   This graph is created using the derived variable *z-score* by the variable *Neighbourhood Cleansed*. Colour scheme is kept constant with the colours used in the prior geomap as this visual is intended to support the analysis of suburb profile. When creating the dashboard "*Suburb Profile*", a common **highlighting** rule is created for the map and bar graph to assist in decoding the information for individual suburb at the same time and reducing cognitive overload of the information.

Figure 2: Action highlight for dashboard "Suburb Profile"

3. Boxplot (Worksheet 3): Once introduced through the suburb profile, it is of interest to establish what sort of room type has high reviews on average to understand the vacancy rate. The most common way to overcome the lack of data for guests checked in and checked out is to use reviews. The Airbnb structure is set up in a way that only once the check-out is confirmed can people leave reviews, meaning the number of reviews on a listing is a meaningful way to determine the vacancy rate on room types.

   Using the variable number of reviews by room type for the neighbourhood for available rooms only and a **filter** on calendar year set to 2020, it provides an effective overview of the distribution of the vacancy across the room types. With the inner box indicating the middle 50% of the data and whiskers indicating the amount of spread within 1.5 times (Tableau, 2020), this visual is indicative of both optimistic and pessimistic vacancy number for the average listing under room type.
   A filter for the month of the year is provided in case the investor is interested in checking out the non-peak months of the year to see if the spread of the reviews is consistent with the yearly average or not.

4. Area graph (Worksheet 4): To further dig into what sort of property type will yield high returns, an area graph for average price is provided by property type. Additional breakdown of what sort of room a particular property type belongs to is provided to understand the inclination of the potential Airbnb lodger. Average form **Tableau**

analytics is used to add an average line to provide an overview on which type of room type in the property listing is not widely preferred.
Dashboard "*Property Preference*" is created with **action highlighting** to allow user to only select a particular room type and compare both area graph and boxplot.

Two dashboards and a story are created for final presentation of the visuals.

- Dashboard 1, "*Suburb Profile*" is created with an intent to assist investors infer a suburb profile with a price range to determine the rental yield. Combining a geomap and bar graph along with functionality of action highlighting makes it an easy view of the suburb profile. A filter of Area price range on the side helps the investor change the range of average listing price by their needs. Colour schemes are constant between the geomap and bar graph with use of "Red-Blue-White Diverging" scale to tie back to Airbnb colour palette and visually differentiate the opposite end of the spectrum – average price range with geomap and z-score with bar graph.
- Dashboard 2, "*Property Preference*" is created to equip investor with knowledge of what sort of property is widely in-demand and how their vacancy rates compare to different types of room types. This dashboard is combination of worksheet 3 and 4 which help by having details of average price of room type displayed right next to the number of reviews used to determine the vacancy rates. Colour scheme of red is kept consistent with Airbnb colour scheme and dashboard 1 to blend and be easy on the eyes.

Discussion

Overall story created in Tableau is with an intent to assist investors make an informed decision on whether it may be worthwhile for them to invest in Airbnb hosting.
Detailed below is how each visualisation in the dashboard is intended to be used.

Visual 1: Geomap

A geomap is an intuitive way to demonstrate the suburb profile to which an investor can get themselves familiarised to. To reduce the mental processing, gentle colour saturation is used to provide the eyes visual cues to help the eyes and brain process the targeted potential points of interest more quickly. (Knaflic, 2015). Average price range by latitude and longitude to create this map with scale range between lowest rate ($106.5) and highest rate ($355.7) of the listing with visual makes it easier for the potential investor to process the details faster than plain numbers with no visual cues. There is a legend provided for the map to understand what colour depicts what range spectrum.
There is a significant degree of interactivity ranging from tooltips to help decode the price rates and rates split by room types, to action highlighting of the suburbs against the bar graph. With very minimal cognitive interference, this visualisation has two colour scales but the human perceptual system is fundamentally based on relative judgements, not absolute ones (Munzner & Maguire, 2015) which is why a scale with two colours that doesn't blend too considerably makes it easier to visually gauge which suburbs have high and low listing rates. The Gestalt principles used in this map include **similarity** - where two different colours at different ends of the spectrum to denote the higher and low listing rate by suburb, as well as **continuity** - where the same colour scale is used for visual 2 (bar graph) in addition to the map in dashboard 1. Channel of spatial region and colour is used as attribute here.

Visual 2: Bar graph

A horizontal bar graph is presented next to geomap as them assist in decoding similar information that can be used together. This is created to inform investors of the suburb's details when comparing the suburb listing price to the mean of overall listing prices across Tasmania. No legend is provided as the numbers on the bars are self-intuitive with positives in blue and negatives in red. This reduces the load of information provided to investor to digest. Interactivity with this graph is at par with action highlighting shared between bar graph and geomap to assist with suburb exploration. Gestalt principles in use are **proximity** (with use of white space to separate chart elements), **similarity** (with use of blue and red to identify the groups of positive z-score and negative z-score), and **continuity** (where bar continues flawlessly separating both the groups into two parts – higher or lower than average total listing price).

There is no cognitive interference or distractors in this visualisation with only 2 groups to visualise and the choice of colour schemes remain consistent and mild on the eyes. Representation of z-score split with suburb average listing price helps put the results into perspective in terms of the favourable suburb for high rental yield for investors and effectively devote time into further researching on selected suburbs.

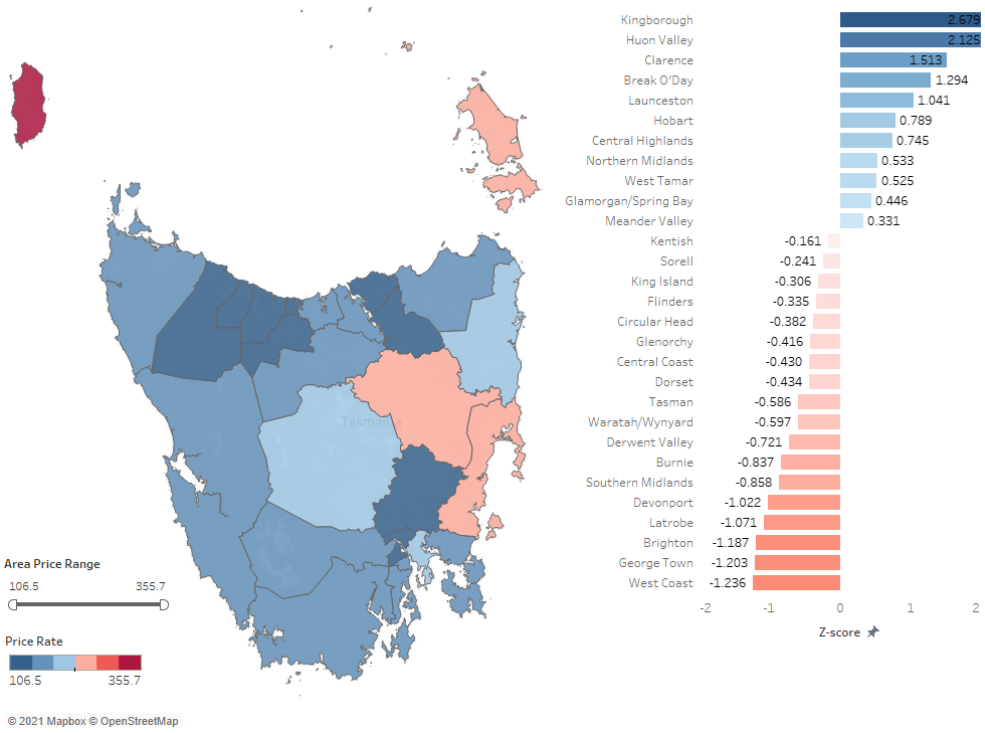

Figure 3: Dashboard 1 "Suburb Profile" with visual 1 (geomap) and visual 2 (bar graph)

Visual 3: Boxplot

The average number of reviews are used to determine the vacancy rate of a listing by month where investor will be able to determine what room type is more in demand worth investing in if considering hosting with Airbnb.

No legend is provided with box plot, but it is self-explanatory with darker shade of the box plot denoting the spread lower than mean number of reviews and lighter shade denotes the spread higher than the mean number of reviews. This colour scheme is consistent with the area graph for visual 4 with use of Airbnb colours.

There is some interactivity with this graph where a filter for month is present to filter out the reviews to get ones of interest for months. No cognitive load or distractors are present for this visualisation with only 2 groups, shape as marker and colour as channel. No background contract is present which lessens the cognitive load and delivers required detail with ease. Gestalt principles in use are **proximity** with use of white space used to separate "room type" groups and **similarity** with use of two colours to denote different groups at different spectrum of the mean.

Visual 4: Area graph

The use of an Area graph is employed here to help indicate what type of property is more appealing to the tenants using Airbnb and an average listing price they are willing to pay for. This graph provides greater context when placed beside the prior boxplot as it can be noted that 50% of the data for average number of reviews is the highest for "Entire home/apt" which means that vacancy rate for that room type is potentially low. To understand what sort of prices the average individual is willing to pay is as high as $233 for "Natural/Earthy/remote" property types.

Use of Tableau analytics to signify the average price of the data spread assists to understand what room types are worth shortlisting and which ones may not be worth the investment – for instance, shared room has the lowest number of reviews and lowest rental yield. Action highlighting for the room type is used between bar graph and area graph to select same group when wanted.

With no legends provided for this graph, the Gestalt principle of **continuity** helps guide the eyes through different groups from highest average price for listing to the lowest. Gestalt principle of **similarity** is used with different shade of the same colour palette to distinguish different groups. There is minimal cognitive interference with no distractors and more white space to differentiate between the groups and graphs.

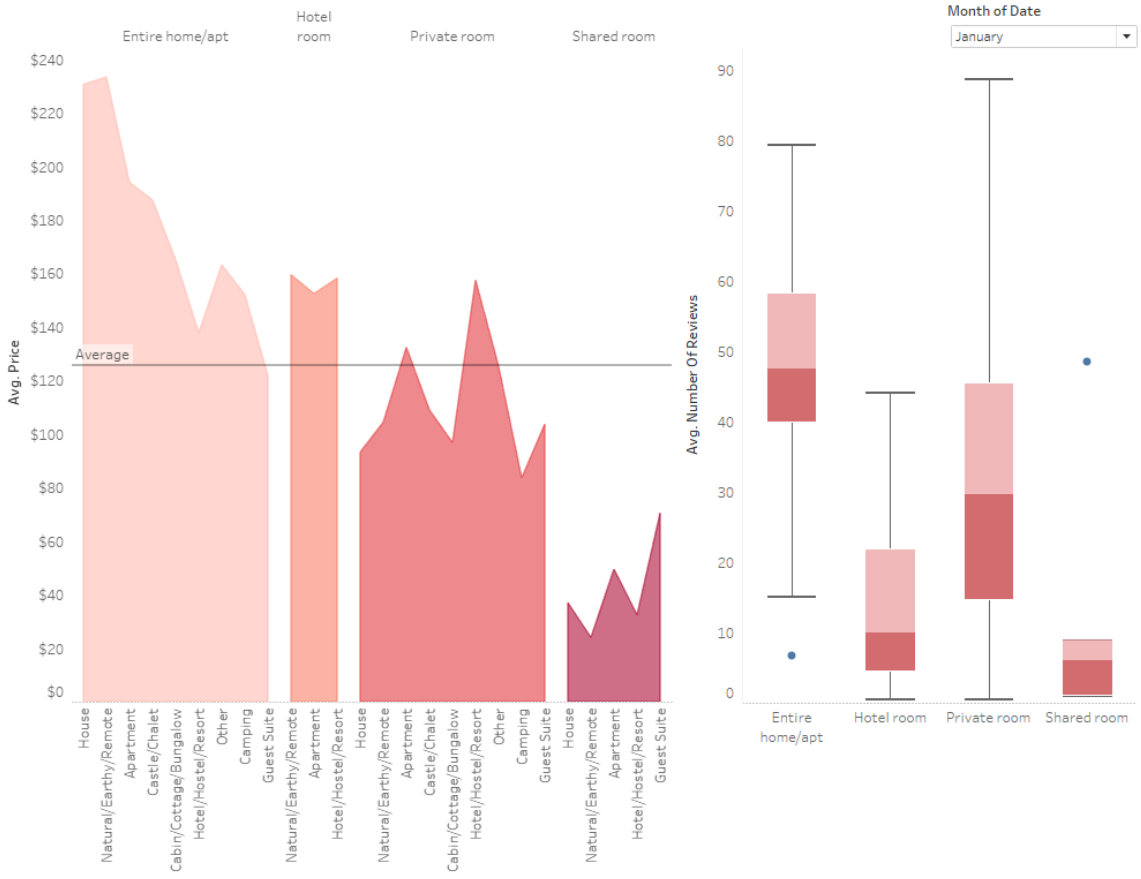

Figure 4: Dashboard 2 "Property Preference" with visual 3 (boxplot) and visual 4 (area graph)

Conclusion

As observed from these results, there is a growing potential for investment in Tasmania with such a huge tourism industry boost thanks due to the growth of Airbnb with high rental yields and low vacancy rates. This dashboard layout and presentation paves the way for future analysis and underpins an opportunity to expand on comparisons of the Airbnb rental yield against the long-term tenancy yields, population % and tourism volume by each suburb.

In terms of limitations faced, it is critical to observe potential downstream effects in the data validity as the majority of 2020 faced ongoing consequences due to the COVID19 epidemic, resulting in an expected decline across the tourism industry and shared room spaces as no international traveller or backpackers were travelling to Australia. This would presumably result in skewed results and potentially adjusting the results against the average of prior years would benefit the investor in making more informed decision.

It is also worthwhile to note that the data available for this dashboard does not include the total number of visits for each property or property type, meaning that inferences have been drawn to assume the popularity of a property type based on the number of reviews instead. This however is an imperfect solution as the high number of reviews observed for the property type “Entire home/apt” may not necessarily be an indication of popularity, but because the majority of homes available on Airbnb fit this description.

From the dashboard, addition of narratives to build on story point would have helped the investor more than just viewing visual representation and leaving the interpretation to individual’s perception and has room for future development and further refinement.

References

Bleby, M. (2018, April 13). *Get out of Airbnb: Tasmania offers landlords $13,000 for long-term rentals*. Retrieved from Australian Financial Review: https://www.afr.com/property/get-out-of-airbnb-tasmania-offers-landlords-13000-for-longterm-rentals-20180412-h0yozy

Knaflic, C. N. (2015). *storytelling with data.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Munzner, T., & Maguire, E. (2015). *Visualization analysis & design .* CRC Press LLC.

O'Mahony, J., Simes, R., Hill, A., Liu, Y., & Alvaro, E. (2017). *Economic effects of Airbnb in Australia, Tasmania.* Deloitte Access Economics.

Tableau. (2020). *Build a Box Plot*. Retrieved from tableau.com: https://help.tableau.com/current/pro/desktop/en-us/buildexamples_boxplot.htm