MGMT 58500-002

TEAM 3

FINAL PROJECT

MANAGING AND DELIVERING ANALYTICS PROJECT

Jai Trivedi

Madhulika Chilla

Manoj Padmaraju

Olivia Ward

Thannir Malai Annamalai Kumar

Table of Contents

Plan Based Approach

For the plan-based approach, we simulated a 15-day project and created a plan considering the various roles and responsibilities associated in a professional setting. The costs and responsibilities assumed by the group members and the hourly costs associated with them are given in *[Exhibit 1]*.

The starting date for the project was 11/16/2022 and it was divided into 4 task categories with the following subtasks:

1. **Initiating tasks**: Reading the case, drafting project requirements and final review of objective
2. **Planning tasks**: Developing WBS, estimating duration, assigning resources, and determining inter-task relationships
3. **Executing tasks**: Data gathering, cleaning, EDA, and modeling
4. **Closing tasks**: Preparing project report, PowerPoint presentations, and recording presentation

The following table represents the WBS with the duration, predecessors, and resource allocation:

				Resource Allocation (%)				
S. No	**Task Name**	**Duration**	**Predecessors**	**PM**	**DS**	**AC**	**DA**	**DA**
1	Start Project	0 days						
2	Initiating Tasks *(Summary)*	1.5 days	1					
3	Read Case Individually	1 day	1	50%	50%	50%	50%	50%
4	Draft Project Requirements	0.5 days	3	10%	10%			
5	Review Project with Team	0.5 days	4SS	25%	25%	25%	25%	25%
6	Planning Tasks *(Summary)*	1.5 days						
7	Develop WBS	1 day	5	25%	25%	25%	25%	25%
8	Estimate Task Durations	0.5 days	7	10%	10%	10%	10%	10%
9	Assign Resources	0.5 days	7	10%				
10	Determine Task Dependencies	0.5 days	7	10%				
11	Executing Tasks *(Summary)*	9 days	10					
12	Data Gathering	2 days	10		25%	50%	100%	100%
13	Data Cleaning	1 day	12			25%	50%	50%
14-16	Status Updates	6 days	8			10%		
17	Exploratory Data Analysis	3 days	12		10%	25%	50%	50%
18	Preliminary Analysis from EDA	1 day	17		25%	50%		
19	Build Models	2 days	17,18		10%	25%	50%	50%
20	Review and Assess Results	1 day	19	10%	10%			
21	Closing Tasks *(Summary)*	3 days	20					
22	Prepare Project Report	2 days	20FS-10%	25%	25%	50%	50%	50%
23	Prepare Presentation	2 days	20FS-10%	25%	25%	50%	50%	50%
24	Record Presentation	1 day	22,23	50%	50%	50%	50%	50%
25	Final Quality Check and Submission	0.5 days	22,23	25%				
26	End Project	0 days	25					

As per the plan the total project was estimated to cost **$9,977** at the time of completion considering no delays and overtime work. The costs of each summary task are reported in *[Exhibit 2]*.

Agile Based Approach

For the agile based approach, we followed Kanban framework to effectively plan and track our project management activities. We visually tracked the tasks from backlog, through work-in-progress phases, and finally to completion. All the tasks are navigated by our team members on Kanban board *[exhibit 5]*, so we can get a quick overview of multiple tasks from a bird's-eye view. Additionally, we have used **pull system as a lean technique** to replace work that has been done.

Sample Kandan Board:

Backlog	Planning	In progress	Complete
Preparing the report	Discuss the future scope, recommendations, and key finding	Review the insights from EDA	Data Cleaning
Presentation	Perform Data visualization	Building the model based on EDA	Perform Exploratory Data Analysis

User Story Estimation:

Coming to user story estimation we have followed T-shirt sizing method, where we as team members had a session to discuss among ourselves to evaluate the efforts required by the user stories starting from XS to L. For instance, if the user story is given "M" size, it could mean the task requires less effort and time to finish. This helped us understand the user stories very quickly.

Prioritization:

MosCow method was used to prioritize user stories, where we placed our top priority requirements like "Data Extraction", "Data modelling" under must-have initiatives and secondary priority tasks like "Data cleaning" under should-have and less important tasks than must-have and should-have like "Preparing the report" under could-have and "Use Lean based approach to identify gaps in the current methodology" under wont'-have initiatives to get back later to work on these tasks.

Acceptance Criteria and Definition of Done:

We have evaluated each of our user stories against their respective acceptance criteria to be considered as done. For instance, for the user story "Data Cleaning" we defined our acceptance criteria as “Data should not have any null values or missing values”. Further, our product manager was responsible for writing the acceptance criteria of all our user stories which clarified the expected outcome of each user story.

Descriptive Analysis

To analyze the Airbnb listings and gather insights, the team narrowed down the cities of interest to San Francisco and Sydney. Exploratory Data Analysis was performed on Tableau for San Francisco and Sydney Airbnb listings for the year 2021. The key features/data points considered for analysis was involved categorical variables like Geography: City, Neighborhood; Listing features: Room Type; Host features: Host Verification, Superhost, instant bookablility. The numerical variables or metrics of interest considered were number of listings, price of listings, ratings, host response rate, host acceptance rate. The three key takeaways from initial EDA are:

1. Overview of Listings and Hosts:

- ***Sydney:*** The ***listings per host*** remained in the range of ***1-2*** across different neighborhoods, indicating ***location*** is ***not influencing*** the number of listings the hosts hold. *[Exhibit 6]*
- ***San Francisco:*** The ***listings per host*** varied between ***1-5*** across different neighborhoods, indicating ***location influences*** the number of listings the hosts hold. ***Luxurious neighborhoods*** (ex: Downtown Civic Center) had very high listings per host. *[Exhibit 7]*

2. Superhost Share: On average, ***San Francisco*** had ***higher % of superhosts*** compared to Sydney. This can be either because in San Francisco, the majority of the hosts provided great experience to their customers or the customers in general in were very happy with the experience irrespective of the service provided. *[Exhibit 8]*

3. Expansion by Room Type: Both San Francisco and Sydney had high number of Entire home/apt listings and private room listings and high customer reviews scores as well

- ***Sydney:*** In Sydney, even though hotel rooms have very high customer reviews scores (4.5), the number of listings were very less (141). In Sydney, Airbnb can plan its expansion for ***hotel room listings*** and reduce shared room listings as it is not doing well. *[Exhibit 9]*
- ***San Francisco:*** In San Francisco, even though shared room listings have very high customer reviews scores (4.6), the number of listings were less (133). So here, Airbnb can plan to expand ***shared room listings*** and reduce underperforming hotel rooms. *[Exhibit 10]*

Analysis and Modeling – Approach and Findings

The analysis was broken down into 2 parts – 1. COVID Impact assessment 2. Deploying a predictive pricing strategy based on statistically significant variables

1. **COVID Impact Assessment** – This section explores the shift in booking patterns caused by the spread of COVID across various phases of the pandemic in San Francisco and Sydney

a. **Total Yearly Reviews** – Considering review activity as a proxy for booking quantities, we can observe a dip in number of bookings by 86% in Sydney and 95% in SF with the onset of lockdown (April 2020) as compared to the previous year. With the ease of restrictions, Sydney recorded an increase in activity until July and further dropped by 30% until September before experiencing an increasing trend from October 2020 with the onset of phase-wise easing of restrictions across NSW. On the contrary, SF experienced a spike until August as the lockdown ended before plummeting to only about 1000 reviews towards the end of 2020. Overall, SF suffered a greater impact in terms of reduction in Airbnb activity

b. **Seasonality by Occupancy Rates** – On comparing the percent change in total monthly reviews between 2019 and 2020, we observed the least decrease in April 2020 for both cities as expected. However, the percent change flattened for Sydney from August while we observe an oscillating trend for SF. This can be attributed to the fact that August is a holiday month and Sydney is the most sought-after tourist destination which is why although the volumes remained low due to international travel, the percent decrease remained constant due to increased local activity. Overall, Sydney recorded increased rate of bookings post Covid compared to SF which recorded a nominal growth in activity as compared to 2019

c. **Price Distribution by Accommodation Type** – The number of active listings dropped by 84% and 91% in Sydney and SF respectively. Out of these listings that were active, hotel rooms and Entire homes are the most expensive on average followed by private rooms and shared rooms in both these cities The biggest drop in price from 2018 to 2020 happened for hotel rooms followed by entire homes whereas the price for shared rooms increased slightly by 9% and average price for private rooms decreased by 7% for Sydney where as the number of hotel rooms in SF fell by ~99% while recording an average price decrease of 18% in private rooms. In terms of neighborhood, the areas around downtown recorded the highest decrease in prices across all accommodation types.

2. **Pricing Modeling –** This explains the overall methodology employed to develop a localized listing level predictive pricing model and also /compare them with the existing calendar prices.

Methodology

a. Perform exploratory data analyses on the Airbnb calendar data for both cities to gauge the average prices per night and average availability

b. Segment all the listings into four different buckets in terms of prices and perform topic modeling to identify the major topics within each of these prices. These topics will serve as a pre requisite to understand the major drivers of price with in each of the buckets

c. Assess the feature importance of various variables individually through correlation test for both the cities to narrow down on the subset of features that contribute to a gain in information for predicting prices

d. Build machine learning models iteratively to reduce the error in prediction and fine tune the hyperparameters to obtain the most generalized model

e. Apply the learnt model to predict prices for Q4'22 and compare them with the calendar prices

Findings

a. Topics with key words like 'modern', 'parking', 'large' fall within the higher price buckets. We also observed a high degree of overlap in topics across buckets in both cities.

b. There are about 18 variables for Sydney and 15 variables for SF that contribute to information gain in prices. The three additional variables for Sydney are cleansed neighborhood flag, property type and length of association of listings

c. Gradient boosting model recorded a test accuracy of 74% for Sydney and 70% for SF which was further used to generate predictions at the listings level

d. The average predicted price for SF is higher than Sydney which is in accordance with the calendar prices. However, the pricing model for SF and Sydney suggested prices that are 21% and 19% higher respectively as compared to the calendar prices for Q4'22

Key Takeaway

The proposed solution considers the changing local market dynamics by building upon the most recent price information and provides more visibility into the price drivers which Airbnb can leverage to proactively monitor and update their pricing strategy in the post COVID era.

Conclusion and Future Scope

With the success of this project and using these approaches, future scope is an imminent and important discussion. First and foremost, this team would like to expand the scope of the project to investigate factors such as COVID's impact on the housing market. While the room-share market and the housing market saw different impacts initially during Covid through reporting by various media sources, this reporting slowed down, and a more in-depth analysis is required. Analysis on these trends internationally and in comparison, to the room-share market could assist in future trend prediction for these two markets, as they become highly influenced by each other. More often recently, people who are looking to enter the housing market to buy or sell have thoughts about renting out their space to make income, especially for second and third properties. These two industries are more intertwined than they lead on.

Another important relationship to investigate comes with the room-share and hospitality markets compared through regions that had incredibly different regulations when it comes to Covid. Regions in the southeast part of the US had very different regulations and regulatory period as compared to the west coast and northeast. These comparisons in addition to any international data that can be gathered and used for the same purpose could shed some light into the trends highlighted in this report. This expansion of scope, much like the housing expansion, relies on the gathering of sometimes hard to reach data. With greater quality and quantity of data, the team would also expand to further improve, validate, and solidify models and conclusions already made. Analysis upon sentiment of reviews and listings using natural language processing, correlations between amenities, bookings, and price, and solidifying the project management side with a lean based approach are all things crucial to success as the team moves forward with a greater scope. A focus on constant reassessment and evaluation is extremely important and continues to be at the forefront of any projects the team takes on.

A traditional analytics project such as this can be tricky to organize and manage, especially with a larger group of intelligent analysts who want to make a meaningful contribution. A plan that incorporates both agile and traditional project planning is useful in a situation such as this to ensure that everyone’s voices are heard, and the analysis is complete and useful. Through the analysis performed, strong conclusions were drawn addressing the impact COVID had on the same industry in two completely different places in the world with differing regulations and market characteristics. Additionally, a pricing model was able to be developed and would assist Airbnb in increasing their business intelligence in a given marketspace, something that could help in a plethora of decisions that need to be made. Projected values of revenue growth stand at approximately 21% for San Francisco and 19% for Sydney for the current listings as the world ventures into a relatively COVID-free period. Through the work already done and the future scope the team hopes to achieve, project management techniques have played and will play a very important role.

Appendix

Exhibit 1: Resource Allocation and Costs

Name	Role	Std. Rate
Thannir Malai Annamalai Kumar	Project Manager (PM)	60
Manoj Padmaraju	Data Scientist (DS)	50
Olivia Ward	Associate Consultant (AC)	45
Madhulika Chilla	Data Analyst (DA)	35
Jai Trivedi	Data Analyst (DA)	35

Exhibit 2: Summary Task Costs at Baseline

Task Name	Total Costs
Initiating Tasks	$1,169
Planning Tasks	$588
Executing Tasks	$4,540
Closing Tasks	$3,680

Exhibit 3: MS Project Workflow

Task Name	Duration	Start	Finish	Predecessors	Resource Names
Final Project	15 days	Wed 11/16/22	Tue 12/6/22		
1 Start Project	0 days	Wed 11/16/22	Wed 11/16/22		
2 Intiating tasks	1.5 days	Wed 11/16/22	Thu 11/17/22	1	
2.1 Reading Case Individually	1 day	Wed 11/16/22	Wed 11/16/22	1	Jai Trivedi[50%
2.2 Draft Project Requirements	0.5 days	Thu 11/17/22	Thu 11/17/22	3	Thannir Malai A
2.3 Review Project With Team	0.5 days	Thu 11/17/22	Thu 11/17/22	4SS	Thannir Malai A
3 Planning tasks	1.5 days	Thu 11/17/22	Fri 11/18/22		
3.1 Develop WBS	1 day	Thu 11/17/22	Fri 11/18/22	5	Thannir Malai A
3.2 Estimate Task Durations	0.5 days	Fri 11/18/22	Fri 11/18/22	7	Thannir Malai A
3.3 Assign Resources	0.5 days	Fri 11/18/22	Fri 11/18/22	7	Thannir Malai A
3.4 Determine Task Dependencies	0.5 days	Fri 11/18/22	Fri 11/18/22	7	Thannir Malai A
4 Executing tasks	9 days	Mon 11/21/22	Thu 12/1/22	10	
4.1 Data Gathering	2 days	Mon 11/21/22	Tue 11/22/22	10	Jai Trivedi,Mad
4.2 Data Cleaning	1 day	Wed 11/23/22	Wed 11/23/22	12	Jai Trivedi[50%
4.3 Status Updates	6 days	Wed 11/23/22	Wed 11/30/22	8	
4.4 Exploratory Data Analysis	3 days	Wed 11/23/22	Fri 11/25/22	12	Jai Trivedi[50%
4.5 Analysis from EDA	1 day	Mon 11/28/22	Mon 11/28/22	17	Olivia Ward[50
4.6 Build Models	2 days	Tue 11/29/22	Wed 11/30/22	17,18	Jai Trivedi[50%
4.7 Review and Assess Results	1 day	Thu 12/1/22	Thu 12/1/22	19	Manoj Padmar
5 Closing tasks	3 days	Fri 12/2/22	Tue 12/6/22	20	
5.1 Prepare Report	2 days	Fri 12/2/22	Mon 12/5/22	20FS-10%	Jai Trivedi[50%
5.2 Prepare Presentation	2 days	Fri 12/2/22	Mon 12/5/22	20FS-10%	Jai Trivedi[50%
5.3 Record Presentation	1 day	Tue 12/6/22	Tue 12/6/22	22,23	Jai Trivedi[50%
5.4 Final Quality Check and Submission	0.5 days	Tue 12/6/22	Tue 12/6/22	22,23	Thannir Malai A
6 End Project	0 days	Tue 12/6/22	Tue 12/6/22	25	

Exhibit 4: Workflow tracking using Smartsheet

Title	Description	Acceptance Criteria	Status	Done	Size	Priority	Assigned To
Reading the case	As a team member, I want to understand the case to further perform analysis on it.	Understand the project objectives	Complete	☑	XS	!	JT Jai Trivedi MP M
Requirements Gathering	As a business analyst, I want to list down all the requirements to establish a foundation for the project's vision.	Create user stories with the key requirements.	Complete	☑	S	–	JT Jai Trivedi MP M
Data Extraction	As a data engineer, I want to gather the data from external data sources for Airbnb and Covid 19 to identify relevant information to perform analysis.	1.Download the data for two cities. 2. Download the data for Covid-19	Complete	☑	M	!	JT Jai Trivedi MP M
Extract Data for COVID 19			Complete	☑			
Extract data for San Francisco			Complete	☑			
Extract data for Sydney			Complete	☑			
Data Cleaning	As a data engineer, I want to perform data cleansing to identify and fix the irrelevant information present in the raw data.	Data should not have any null values or missing values and all the irrelevalant data must be eliminated.	Complete	☑	S	!	JT Jai Trivedi M m
Eliminating all the missing and null values				☑			
Perform Exploratory Data Analysis	As a business analyst, I want to perform EDA to ensure the results produced are valid and are applicable to business outcomes	1. Identifying the outliers, skewedness and patterns in the data. 2. Generate hypothesis based on the trends.	Complete	☑	L	!	MP Manoj Padmaraju
Identify the import variables and tools				☑			

Exhibit 5: Kanban board view

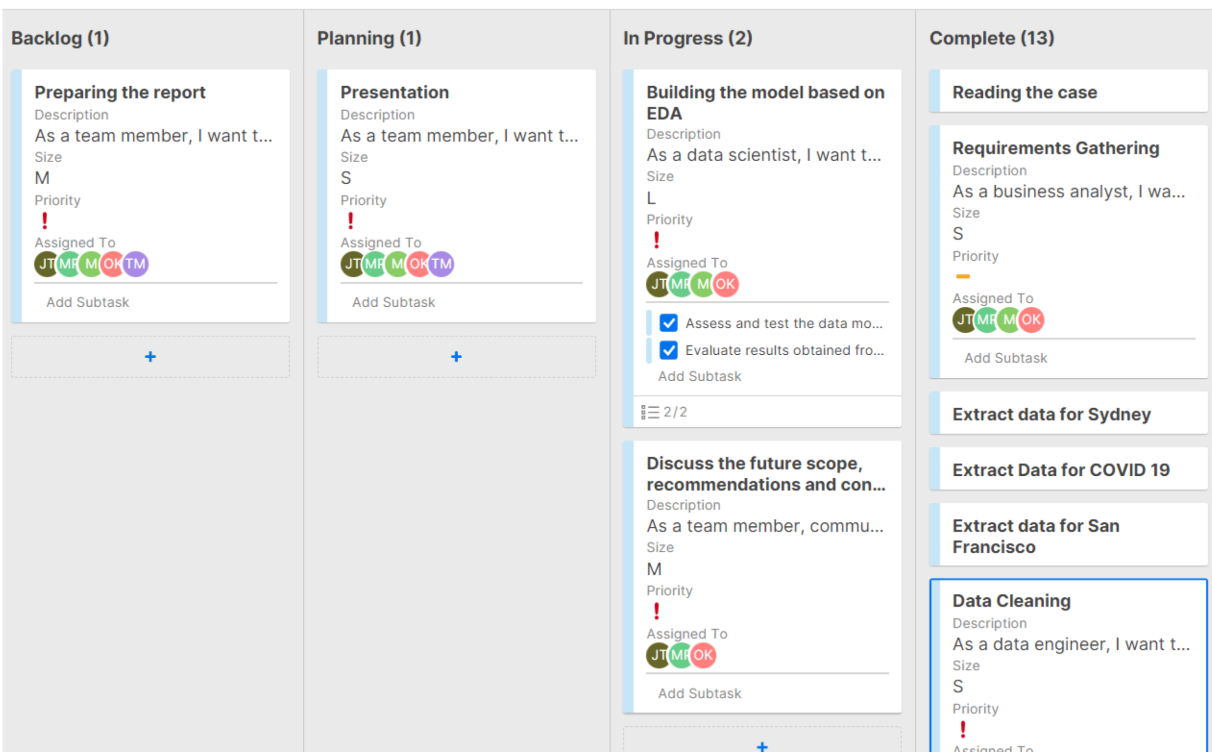

Exhibit 6: Sydney - Overview of Listings Hosts

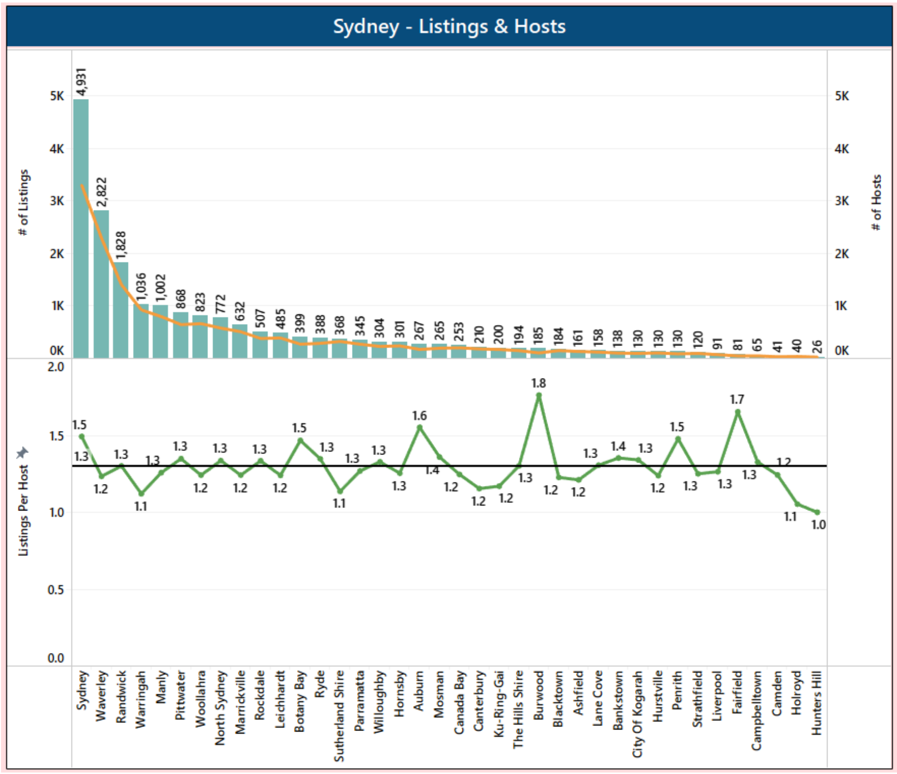

Exhibit 7: San Francisco Overview of Listings Hosts

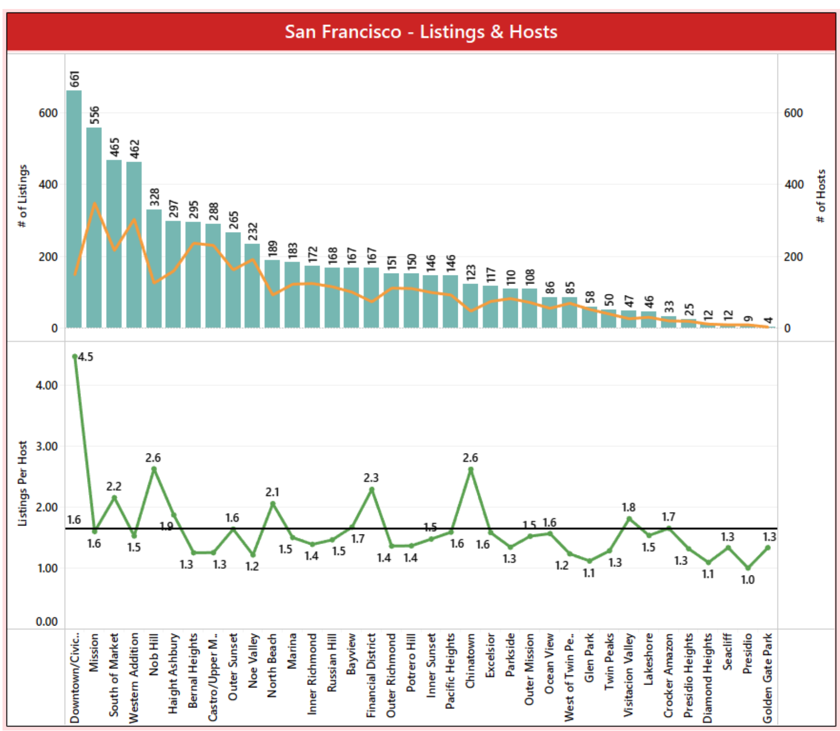

Exhibit 8: Neighborhood Drilldown

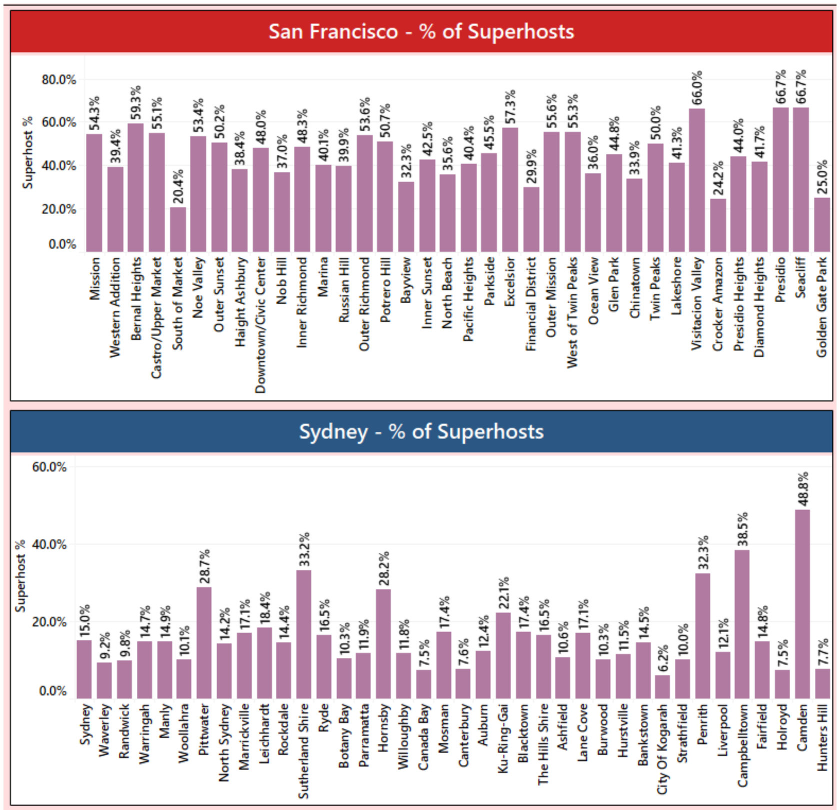

Exhibit 9: Sydney – Room Type Drilldown

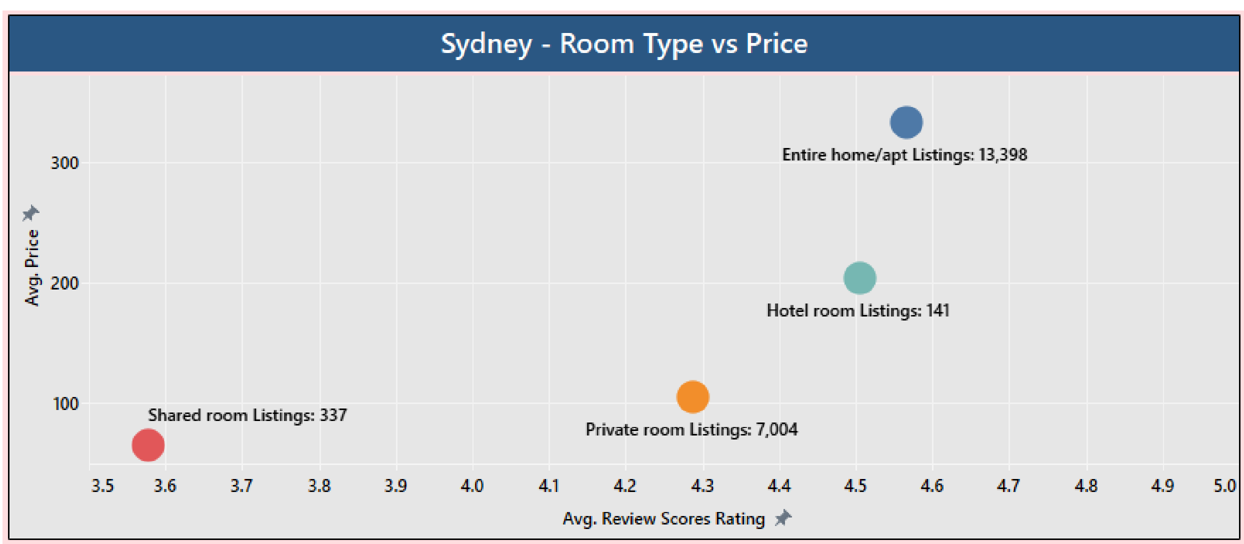

Exhibit 10: San Francisco – Room Type Drilldown