



# *Next Step for airbnb*

## *Business Development Analysis in Hong Kong*

### Abstract

This project aims to summarize Airbnb's current performance in HK and find out the special features of the local market. Various analysis techniques and machine learning models were applied. Two data-driven recommendations are presented to address the findings of our analysis and the company's goals for 2021.

*Isaac Lee, Nicole Tang*



## Contents

<i>Projects Introduction</i>	<i>P. 2 - 4</i>
<i>Summary</i>	<i>P. 5</i>
<i>Waterfall Schedule</i>	<i>P. 6</i>
<i>Data Dictionary</i>	<i>P. 7 - 14</i>
<i>Power BI Dashboard</i>	<i>P. 15 - 28</i>
<i>Progress Checklist</i>	<i>P. 29</i>
<i>Contacts</i>	<i>P. 30</i>



## **Next Step for Airbnb - Business Development Analysis in Hong Kong**

### **Keywords**

Business Development, Data Analysis, Machine Learning, Natural Language Processing, Sentiment Analysis, Time-Series Prediction, Business Intelligence, RFM analysis, Vacation rentals, Tourism, Marketing Targeting, Greater China, Hong Kong

### **Main Purpose**

**Business development** is crucial for every company.

By utilizing the existing company's data, we could identify the business positioning and the discovered advantages and even disadvantages. With the findings, the company is expected to have a clearer and foreseeable development plan which could be achieved in the future.

In the past year, pandemics have harmed the majority of industries in the economy, especially in the tourism and hospitality industry. Thus, we wonder how these industries would adjust and deploy their business plan to adapt to the new situation currently. As the young generation, when it comes to travel, Airbnb would be the first thing that pops up in our mind, and that's also the reason why we would choose Airbnb as our **analysis target**. Due to the better understanding of the regional situation, we have scaled down the targets to Greater China or even to Hong Kong.

Combining the data from October 2020 to July 2021, we conduct the **descriptive analysis** on the existing business data, along with the **simple predictive model for the sales forecasting**, **sentiment Analysis** on the customers' reviews, **RFM analysis**, and **classification model** for relationship maintenance and quality controls. Therefore, the management could have more accurate decision-making on the resource allocation and business development plan based on our **analysis and recommendations**. It would also benefit the prospective landlords to have more understanding of our business, become our business partners, and grow together.



## Business Values

- Grab the opportunities for possible business development;
- Find out our advantages and disadvantages;
- Generate insights that help for decision-making in marketing, resources allocation, customer experiences, and time-saving on report making;
- Provide improvement suggestions.

## Methodology

1. Get the related Data from Airbnb
2. **Data Analysis**
  - Operational Analysis
    - Understanding the current performance is the fundamental base of the project;
  - Host Analysis
    - The hosts are just like customers, we have to find out the valuable hosts and keep them with us, host retention is as important as host recruiting;
    - Dig into host quality and review the existing “Superhost” mechanism.
  - Customers Experience Analysis
    - Understanding the general feedback from customers for services improving;
3. Construct ***time-series models*** to predict the possible future developments in Hong Kong.
  - Develop various models for different locations and different KPIs, for example, estimated availability, services fees.
4. With the above analysis and prediction results, give out our ***recommendations and highlighted business insights***.
5. The final deliverables will be a **machine learning model prediction including classification & time-series prediction model**, a **dashboard** for management, and **reports** for further uses within the company.



## Data Source

### *Airbnb*

- Sources: Inside Airbnb.
- Time: October 2020 - Jun 2021
- Location: Greater China (Beijing, Shanghai, Hong Kong, Taiwan)
- Remark: From watchdog sites & self-scraped

## Deliverables

1. Dashboard with recommendations on marketing, business development.
2. Business Reports
3. Project Summary Leaflet

## Techniques Planned to Apply

- Python
- SQL
- Data Analysis
- Machine Learning (Time-series, Classification, NLP sentiment analysis)
- Dashboard – PowerBI
- Export report in PDF and CSV files for record

# BUSINESS DEVELOPMENT ANALYSIS PROJECT

## KEYWORDS

Business Development, Data Analysis, Machine Learning, Time-series Prediction, Classification, Business Intelligence, Customer Analysis, Sentiment Analysis, RFM Analysis, Group Segmentation

## INTRODUCTION

This project aims to summarize Airbnb's current performance in HK and find out the special features of the local market. Various analysis techniques and machine learning models were applied. Two data-driven recommendations are presented to address the findings of our analysis and the company's goals for 2021.

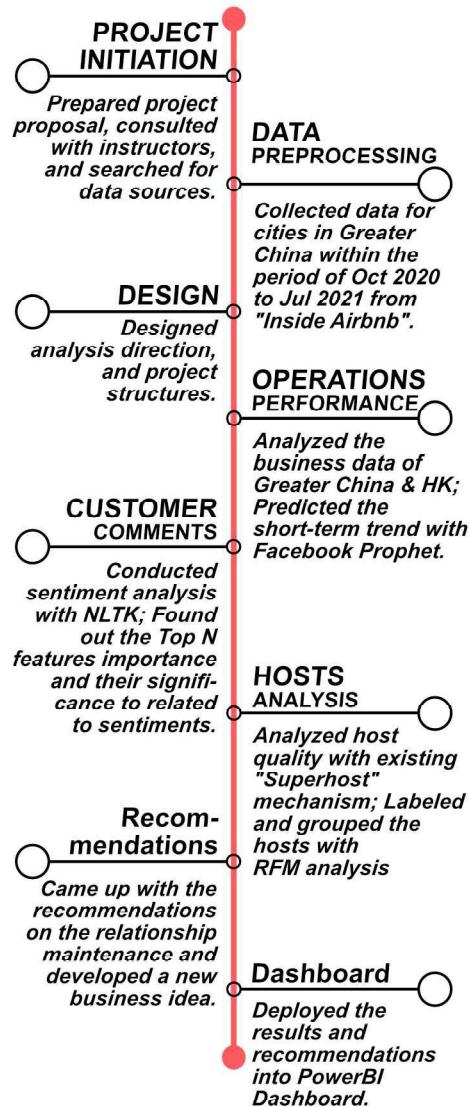
## DATA

SOURCES: Inside Airbnb

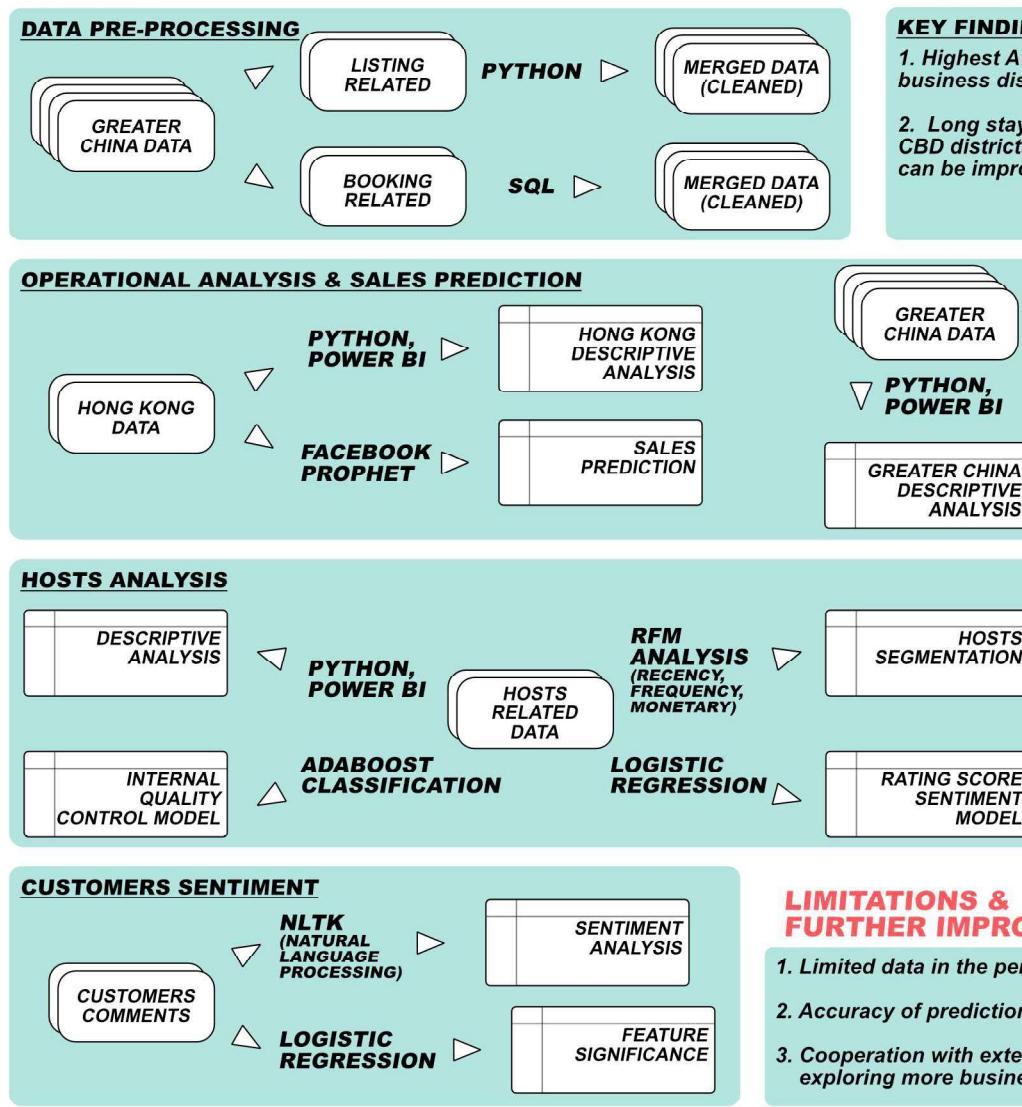
PERIOD: October 2020 - July 2021

LOCATION: Greater China (Beijing, Shanghai, Hong Kong, Taiwan)

## WORK FLOW



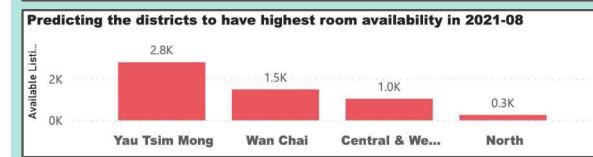
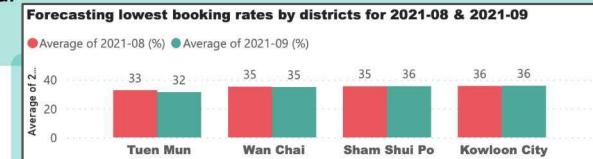
## ANALYSIS



## DATA-DRIVEN RESULTS & RECOMMENDATIONS

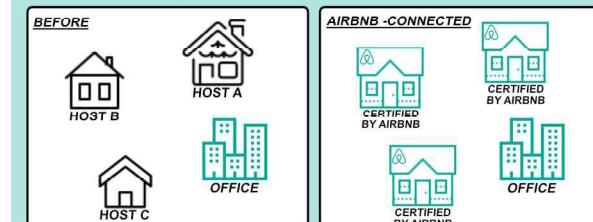
### KEY FINDINGS

- Highest Available Rooms in total and Low Booking Rate in central business district (CBD);
- Long stay guests account for large proportion of the demands in the CBD districts, especially in the Yau Tsim Mong district. Service quality can be improved.



### RECOMMENDATION

- Suggested to be the connector, lead in cooperating with the hosts and companies in CBD to provide standardized and discount long-term living services in which Airbnb would be directly supervised.



### LIMITATIONS & FURTHER IMPROVEMENT

- Limited data in the period from Oct-2020 to July-2021;
- Accuracy of prediction would be improved with completed data and more training;
- Cooperation with external rental platforms could be taken into consideration for exploring more business development opportunities.

Airbnb Business Development Analytics				Start Date	9/2/2021		Completed	In Progress	Phase Period																										
Project Members		Isaac Lee, Nicole Tang		Due Date	9/27/2021					week-1			week-2			week-3			week-4			final-week-5													
							2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
Task		Assigned To	Progress	Start	Duration	End	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W								
<b>Project Initiation Phase</b>			100%	9/2/2021	3	9/4/2021																													
Project Proposal			Isaac, Nicole	100%	9/2/2021	2	9/3/2021																												
Consultation with instructors			Isaac, Nicole	100%	9/2/2021	2	9/3/2021																												
Data source searching			Isaac, Nicole	100%	9/3/2021	2	9/4/2021																												
<b>Data handling Phase</b>			100%	9/5/2021	5	9/9/2021																													
Data Collection			Isaac, Nicole	100%	9/5/2021	1	9/5/2021																												
Data Study & Data Dictionary			Isaac, Nicole	100%	9/5/2021	2	9/6/2021																												
Table merging			Isaac, Nicole	100%	9/7/2021	2	9/8/2021																												
Data cleaning			Isaac, Nicole	100%	9/7/2021	2	9/8/2021																												
EDA - descriptive analysis			Isaac, Nicole	100%	9/9/2021	1	9/9/2021																												
<b>Design Phase</b>			100%	9/10/2021	5	9/15/2021																													
Design Analysis (Operation & Sales, Sentiment, Quality, & Recommendation)			Isaac, Nicole	100%	9/10/2021	2	9/11/2021																												
Methods choosing & testing (Facebook Prophet, Kats, VAR, Neural Prophet)			Nicole	100%	9/12/2021	2	9/13/2021																												
Dashboard comparing and choosing (Power BI/ Tableau)			Isaac, Nicole	100%	9/14/2021	1	9/14/2021																												
Merge and create new tables for analysis			Isaac, Nicole	100%	9/10/2021	5	9/15/2021																												
<b>Development Phase</b>			100%	9/16/2021	8	9/23/2021																													
Operation analysis - Performance Descriptive analysis (Greater China Region)			Nicole	100%	9/16/2021	4	9/19/2021																												
Operation analysis - Performance Descriptive analysis (Hong Kong)			Nicole	100%	9/16/2021	4	9/19/2021																												
Sales analysis - HK Sales descriptive analysis & prediction (FB Prophet Model)			Nicole	100%	9/16/2021	4	9/19/2021																												
Customer Sentiment analysis - Sentiment analysis on client reviews (NLTK)			Isaac	100%	9/16/2021	4	9/19/2021																												
Customer Sentiment analysis - Top N Features most affecting client sentiment			Isaac	100%	9/16/2021	4	9/19/2021																												
Host analysis - Quantitive - Existing Internal Standard -- "Superhost" descriptive and			Isaac	100%	9/16/2021	4	9/19/2021																												
Host analysis - Quantitive - Rating Score from guests - descriptive and predictive analysis			Isaac	100%	9/16/2021	4	9/19/2021																												
Host analysis - Quantile - Host grouping by valuable rankings of RFM Analysis			Nicole	100%	9/20/2021	2	9/21/2021																												
Recommendation - Relationship Maintenance & Superhost Mechanism			Nicole	100%	9/20/2021	2	9/21/2021																												
Recommendation - Potential Business Development - Booking Rate & Number of available			Isaac	100%	9/20/2021	2	9/21/2021																												
Overall Dashboard Summary			Isaac, Nicole	100%	9/20/2021	2	9/21/2021																												
Integration (combine Python & csv files in same directory) for deployment			Isaac, Nicole	100%	9/21/2021	3	9/23/2021																												
Apply model to all data and tables (all location)			Isaac, Nicole	100%	9/21/2021	3	9/23/2021																												
Deploy dashboard & testing			Isaac, Nicole	100%	9/21/2021	3	9/23/2021																												
Dashboard and server linkage			Isaac, Nicole	100%	9/21/2021	3	9/23/2021																												
<b>Testing &amp; Revision Phase</b>			100%	9/24/2021	3	9/26/2021																													
Final Testing			Isaac, Nicole	100%	9/24/2021	1	9/25/2021																												
Revision			Isaac, Nicole	100%	9/25/2021	2	9/26/2021																												
Presentation Materials, Handout, Resume Preparation			Isaac, Nicole	100%	9/25/2021	2	9/26/2021																												
			Total	100%																															

INTERNAL PRESENTATION

FINAL PRESENTATION & TALENT NIGHT

Airbnb Database			Initial Created: 3Sep2021	Last Updated On: 15Sep 2021	By: Isaac Lee, Nicole Tang
Table Names	Country	City	Definition & Description	Table Size & Info	Data Period
detail_listings	China	Shanghai	detail information about every apartment in different timestamp	entries: 236,688	26/10/2020 - 17/7/2021 (Without Mar, May, Jun 2021 data)
		Beijing		entries: 208,712	26/10/2020 - 17/7/2021 (Without May, Jun 2021 data)
		Taiwan		entries: 46,872	30/10/2020 - 23/7/2021 (without May 2021 data)
		Hong Kong		entries: 62,857	25/10/2020 - 13/7/2021 (without May 2021 data)
listings	China	Shanghai	basic information about every apartment with unique apartment	entries: 52,071	26/10/2020 - 17/7/2021 (Without Mar, May, Jun 2021 data)
		Beijing		entries: 39,602	26/10/2020 - 17/7/2021 (Without May, Jun 2021 data)
		Taiwan		entries: 6,936	30/10/2020 - 23/7/2021 (without May 2021 data)
		Hong Kong		entries: 9,302	25/10/2020 - 13/7/2021 (without May 2021 data)
detail_reviews	China	Shanghai	detail information of every comment with the apartment/ listing_id	entries: 495,571	26/10/2020 - 17/7/2021 (Without Mar, May, Jun 2021 data)
		Beijing		entries: 252,506	26/10/2020 - 17/7/2021 (Without May, Jun 2021 data)
		Taiwan		entries: 250,825	30/10/2020 - 23/7/2021 (without May 2021 data)
		Hong Kong		entries: 179,222	25/10/2020 - 13/7/2021 (without May 2021 data)
reviews	China	Shanghai	the date of every comment with the apartment/ listing_id	entries: 463,891	26/10/2020 - 17/7/2021 (Without Mar, May, Jun 2021 data)
		Beijing		entries: 238,323	26/10/2020 - 17/7/2021 (Without May, Jun 2021 data)
		Taiwan		entries: 221,383	30/10/2020 - 23/7/2021 (without May 2021 data)
		Hong Kong		entries: 160,568	25/10/2020 - 13/7/2021 (without May 2021 data)

Detail_ Listing Table		Initial Created: 3Sep2021	Last Updated On: 15Sep2021		By: Isaac Lee, Nicole Tang			
PK	Column	Variable_name	Label	Format-general	Format-Python	Codes/Ranges	Missing_value	Remark
*	1	<b>id</b>	<b>apartment id</b>	<b>Numeric</b>	<b>int64</b>	<b>xxxxxx-xxxxxxxx</b>	<b>Not_allowed</b>	<b>integer; unique</b>
	2	<b>listing_url</b>	apartment url	String	object	link	Not_allowed	unique
	3	<b>scrape_id</b>	id to label scrape batch	Numeric	int64	yyyymmddbbbbbb	Not_allowed	
	4	<b>last_scraped</b>	data scraped date	Date	datetime64[ns]	yyyy-mm-dd	Not_allowed	
	5	<b>name</b>	apartment name made by host	String	object	contains different languages	Not_allowed	encoding = utf-8-sig
	6	<b>description</b>	apartment description made by host	String	object	contains different languages	allowed	encoding = utf-8-sig
	7	<b>neighborhood_overview</b>	neighborhood description made by host	String	object	contains different languages	allowed	encoding = utf-8-sig
	8	<b>picture_url</b>	apartment url	String	object	link	allowed	unique
	9	<b>host_id</b>	host id	Numeric	int64	xxxxxx-xxxxxxxx	Not_allowed	
	10	<b>host_url</b>	host profile url	String	object	link	Not_allowed	
	11	<b>host_name</b>	host name	String	object	contains different languages	Not_allowed	encoding = utf-8-sig
	12	<b>host_since</b>	host registration date	Date	datetime64[ns]	yyyy-mm-dd	allowed	
	13	<b>host_location</b>	host registration location	String	object	contains different languages	allowed	encoding = utf-8-sig
	14	<b>host_about</b>	host self introduction	String	object	contains different languages	allowed	encoding = utf-8-sig
	15	<b>host_response_time</b>	calculated by system based on the host response record	String	object	"within an hour" or "within a day" or "within a few hours" or "a few days or more" or nan	allowed	
	16	<b>host_response_rate</b>	calculated by system based on the host response record	Numeric	object	0-100% or NaN	allowed	float
	17	<b>host_acceptance_rate</b>	calculated by system based on the host acceptance record	String	object	0-100% or NaN	allowed	float
	18	<b>host_is_superhost</b>	recognized by Airbnb for the best hospitality: 4.8+ overall rating, 10+stay, <1% cancellation rate, 90% response rate	Boolean	object	t/f	allowed	t = superhost; f = non-superhost
	19	<b>host_thumbnail_url</b>	host thumbnail url	String	object	link	allowed	
	20	<b>host_picture_url</b>	host picture url	String	object	link	allowed	
	21	<b>host_neighbourhood</b>	host registration location	String	object		allowed	
	22	<b>host_listings_count</b>	number of apartments from the host	Numeric	float64		allowed	

Detail_Listing Table		Initial Created: 3Sep2021	Last Updated On: 15Sep2021		By: Isaac Lee, Nicole Tang		
23	host_total_listings_count	total number of apartments from the host	Numeric	float64		allowed	
24	host_verifications	host's verified info	String	object	['xxxx','xxxx']	allowed	e.g. ['email', 'phone', 'reviews', 'sesame', 'sesame_offline']
25	host_has_profile_pic	host provides profile pic	Boolean	object	t/f	allowed	t = provided; f = didn't provide
26	host_identity_verified	host info verified	Boolean	object	t/f	allowed	t = verified; f = didn't verify
27	neighbourhood	host location provided by host	String	object	contains different languages	allowed	
28	neighbourhood_cleansed	host location revised by sys	String	object	contains different languages	Not_allowed	
29	neighbourhood_group_clea	host location district provided by host	Numeric	float64		allowed	
30	latitude	apartment's latitude	Numeric	float64	xxx.xxxxxx	Not_allowed	
31	longitude	apartment's longitude	Numeric	float64	xxx.xxxxxx	Not_allowed	
32	property_type	property type	String	object		Not_allowed	
33	room_type	type of apartment to offer	String	object	"Entire home/apt" or "Shared room" or "Private room"	Not_allowed	3 options to choose
34	accommodates	number of people can live in	Numeric	int64		Not_allowed	
35	bathrooms	number of bathrooms	Numeric	float64		allowed	many NaN
36	bathrooms_text	number of bathrooms	String	object		allowed	e.g. 1 bath ; 1.5 shared baths ; Shared half-bath
37	bedrooms	number of bedrooms	Numeric	float64		allowed	
38	beds	number of beds	Numeric	float64		allowed	
39	amenities	furniture and remarks from host	String	object	["xxxx", "xxx"]	Not_allowed	e.g. ["Essentials", "Washer", "TV", "Heating", "Wifi", "Cable TV", "Breakfast", "Air conditioning", "Kitchen", "Long term stays allowed"]
40	price	price per apartment per night	String	object	\$xxx.xx	Not_allowed	
41	minimum_nights	minimum number of night stay for the listing (calendar rules may be different)	Numeric	int64		Not_allowed	
42	maximum_nights	maximum number of night stay for the listing (calendar rules may be different)	Numeric	int64		Not_allowed	

Detail_Listing Table		Initial_Created: 3Sep2021	Last_Updated_On: 15Sep2021		By: Isaac Lee, Nicole Tang		
43	minimum_minimum_nights	the smallest minimum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
44	maximum_minimum_nights	the largest minimum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
45	minimum_maximum_nights	the smallest maximum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
46	maximum_maximum_nights	the largest maximum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
47	minimum_nights_avg_ntm	the average minimum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
48	maximum_nights_avg_ntm	the average maximum_night value from the calender (looking 365 nights in the future)	Numeric	float64		Not_allowed	
49	calendar_updated		Numeric	float64		allowed	many NaN
50	has_availability		Boolean	object	t/f	Not_allowed	
51	availability_30	number of days available in coming 30 from the date scrapping	Numeric	int64		Not_allowed	
52	availability_60	number of days available in coming 60 days from the date scrapping	Numeric	int64		Not_allowed	
53	availability_90	number of days available in coming 90 days from the date scrapping	Numeric	int64		Not_allowed	
54	availability_365	number of days available in coming 365 days from the date scrapping	Numeric	int64		Not_allowed	
55	calendar_last_scraped	scrapping date	Date	datetime64[ns]	yyyy-mm-dd	Not_allowed	
56	number_of_reviews	number of reviews	Numeric	int64		Not_allowed	
57	number_of_reviews_ltm	The number of reviews the listing has (in the last 12 months)	Numeric	int64		Not_allowed	

Detail_Listing Table		Initial Created: 3Sep2021	Last Updated On: 15Sep2021		By: Isaac Lee, Nicole Tang		
58	number_of_reviews_l30d	The number of reviews the listing has (in the last 30 days)	Numeric	int64		Not_allowed	
59	first_review	date of first review on the apartment	Date	datetime64[ns]	yyyy-mm-dd	allowed	
60	last_review	date of last review on the apartment	Date	datetime64[ns]	yyyy-mm-dd	allowed	
61	review_scores_rating	average rating on the apartment	Numeric	float64	0-100	allowed	0 = worse; 100 = best
62	review_scores_accuracy	average score for description accuracy	Numeric	float64	0-10	allowed	0 = worse; 10 = best
63	review_scores_cleanliness	average score for apartment cleanliness	Numeric	float64	0-10	allowed	0 = worse; 10 = best
64	review_scores_checkin	average score for check-in process	Numeric	float64	0-10	allowed	0 = worse; 10 = best
65	review_scores_communication	average score for host communication	Numeric	float64	0-10	allowed	0 = worse; 10 = best
66	review_scores_location	average score for apartment location	Numeric	float64	0-10	allowed	0 = worse; 10 = best
67	review_scores_value	average score for the price	Numeric	float64	0-10	allowed	0 = worse; 10 = best
68	license	The licence/permit/registration number	Numeric	float64	0-10	allowed	0 = worse; 10 = best
69	instant_bookable	book without waiting for host approval	Boolean	object	t/f	Not_allowed	t = can be booked currently; f = cannot be booked currently
70	calculated_host_listings_count	number of apartment from the same host	Numeric	object		Not_allowed	
71	calculated_host_listings_count	number of entire home type from the same host	Numeric	int64		Not_allowed	
72	calculated_host_listings_count	number of private room type from the same host	Numeric	int64		Not_allowed	
73	calculated_host_listings_count	number of shared room type from the same host	Numeric	int64		Not_allowed	
74	reviews_per_month	average number of reviews per month	Numeric	float64		allowed	

Listings table			Initial_Created: 3Sep2021	Last_Updated_On: 15Sep2021	By: Isaac Lee			
PK	Column	Variable_name	Label	Format-general	Format-Python	Codes/Ranges	Missing_value	Remark
*	1	<b>id</b>	<b>apartment id</b>	Numeric	Int64	00000000-99999999	Not_allowed	integer; unique
	2	name	apartment name made by host	String	Object	included different language	Not_allowed	encoding = utf-8-sig
	3	host_id	host id	Numeric	Int64	xxxxxx-xxxxxxxx	Not_allowed	
	4	host_name	host name	String	Object	contains different languages	allowed	encoding = utf-8-sig
	5	neighbourhood_group	district group (unknown)	String	Object	NaN	allowed	many NaN entries
	6	neighbourhood	district	String	Object	contains different languages	Not_allowed	encoding = utf-8-sig
	7	latitude	apartment's latitude	Numeric	Float64	xxx.xxxxxx	Not_allowed	
	8	longitude	apartment's longitude	Numeric	Float64	xxx.xxxxxx	Not_allowed	
	9	room_type	type of apartment to offer	String	Object	"Entire home/apt" or "Shared room" or "Private room"	Not_allowed	3 options to choose
	10	price	price per night	Numeric	Int64	0-999999	Not_allowed	integer
	11	minimum_nights	minimum night to rent	Numeric	Int64	0-999999	Not_allowed	integer
	12	number_of_reviews	number of guests' reviews	Numeric	Int64	0-999999	Not_allowed	integer
	13	last_review	date of last review	Date	datetime64[ns]	yyyy-mm-dd	allowed	e.g. 2014-01-20
	14	reviews_per_month	average monthly reviews since the apartment available online	Numeric	Float64	xxx.xx	allowed	float; number of reviews /total months since the listing(or apartment) available online
	15	calculated_host_listings_count	The number of listings the host has in the current scrape, in the city/region geography.	Numeric	Int64	xxx	Not_allowed	
	16	availability_365	number of days available in the future 365 days starting from the scrapping date	Numeric	Int64	xxx	Not_allowed	integer

detail_reviews table			Initial Created: 3Sep2021	Last Updated On: 15Sep2021		By: Isaac Lee		
PK	Column	Variable_name	Label	Format-general	Format-Python	Codes/Ranges	Missing_value	Remark
	1	listing_id	apartment id	Numeric	int64	00000000-99999999	Not_allowed	
*	2	<b>id</b>	<b>comment id</b>	<b>Numeric</b>	<b>int64</b>	<b>xxxxxx</b>	<b>Not_allowed</b>	<b>integer; unique</b>
	3	date	comment date	Date	datetime64[ns]	yyyy-mm-dd	Not_allowed	
	4	reviewer_id	reviewer id	Numeric	int64	xxxxxx	allowed	integer
	5	reviewer_name	reviewer name	String	object	contains different languages	allowed	encoding = utf-8-sig
	6	comments	comments on the apartment service	String	object	contains different languages	allowed	encoding = utf-8-sig

reviews_table			Initial Created: 3Sep2021	Last Updated On: 15Sep2021	By: Isaac Lee			
PK	Column	Variable_name	Label	Format-general	Format-SQL	Codes/Ranges	Missing_value	Remark
*	1	listing_id	apartment id	Numeric	INT NOT NULL	00000000-99999999	Not_allowed	unique when combine with date
*	2	date	comment date	Date	DATE NOT NULL	yyyy-mm-dd	Not_allowed	unique when combine with listing_id



# HONG KONG BUSINESS DEVELOPMENT ANAYLSIS

Isaac Lee | Nicole Tang

- PURPOSE
- GREATER CHINA
- HK PERFORMANCE
- HK PREDICTION
- HOST ANALYSIS
  - VALUES
  - SERVICES
- CUSTOMER ANALYSIS
- RECOMMENDATION
  - HOST
  - LISTING
- SUMMARY

**ANALYSING CURRENT  
PERFORMANCE & PREDICTING  
SHORT-TERM PERFORMANCE**

## Shareholder Letter Q2 2021



**RECOMMENDATIONS ON  
HOSTS RECRUITING & BETTER  
SERVICES QUALITY**

## Shareholder Letter

Q2 2021



### Progress on our 2021 Plan

TER

Our single priority for 2021 has been to prepare for the travel rebound. To do this, we've been perfecting our existing product. We've made significant improvements to the end-to-end experience of our core service for both Hosts and guests, and we continue to focus on the following areas:

- Educating the world about what makes Airbnb different—hosting
- Recruiting more Hosts and setting them up for success
- Simplifying the guest journey
- Delivering world-class service

In Q2 2021, we introduced more than 100 upgrades across every aspect of the Airbnb service, from our website and app to our community support and policies. We added features to give guests even more flexibility when planning their travel, and we made it simpler for anyone to become a Host. These upgrades were announced to our global community at our May 24 event.



Isaac Lee | Nicole Tang

Airbnb HK Business Development Analysis



## GREATER CHINA PERFORMANCE

**31009**  
Total Host

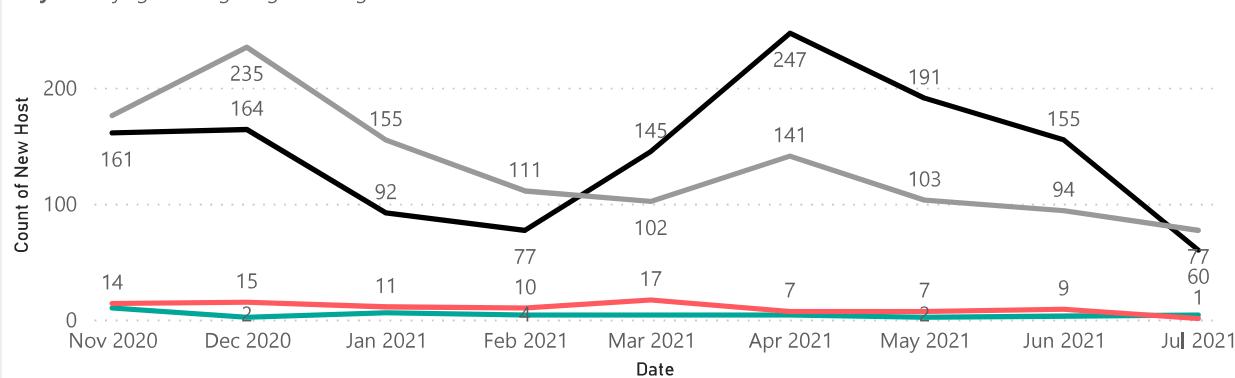
**1839**  
New Host (2021)

### Distribution



### New Host (monthly)

City

● Beijing
● HongKong
● Shanghai
● Taiwan


Host

All

Listing

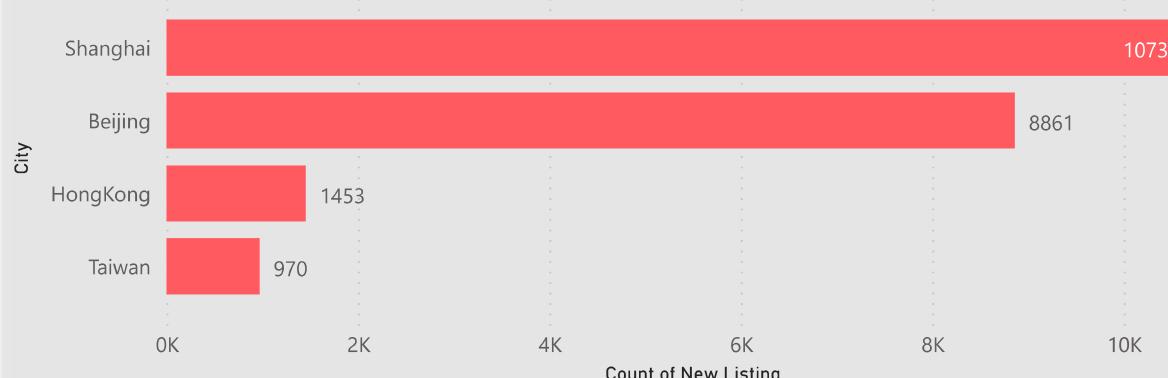
All

**59942**  
Total Listing

**22017**  
New Listing (2021)

### Distribution of New Listing 2021

City



**16023**

New Listing (2021 Existing)

### Remark

1. Greater China refers to Beijing, Shanghai, Hong Kong and Taiwan

2. The numbers are affected by significantly difference in landscape, population,etc.



# HONG KONG OPERATIONAL PERFORMANCE

6538

Total Listing

1453

New Listing 2021

5766

Listing Available in Future 30D

Future 30Day Hot

All

Future 90Day Hot

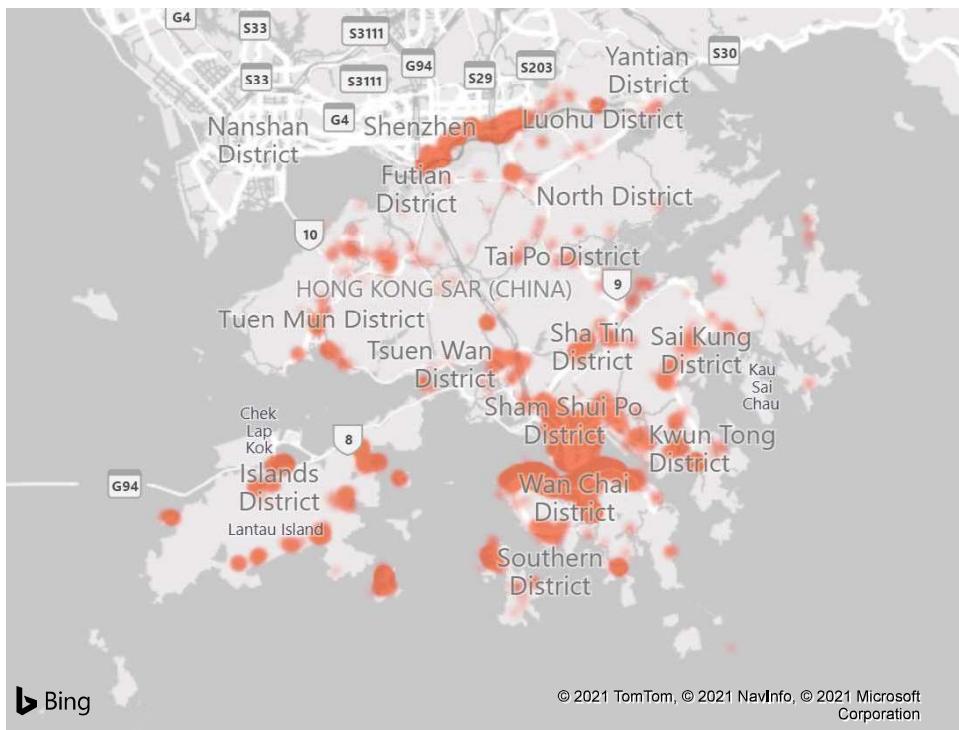
All

Future 365Day Hot

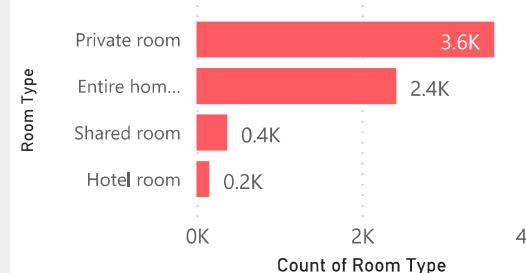
All

District

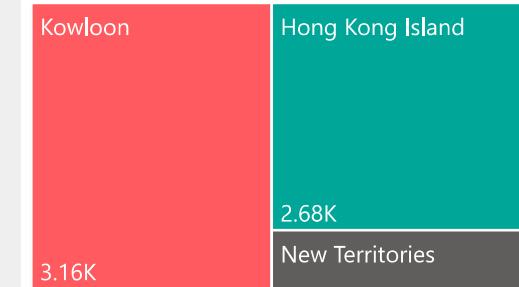
All



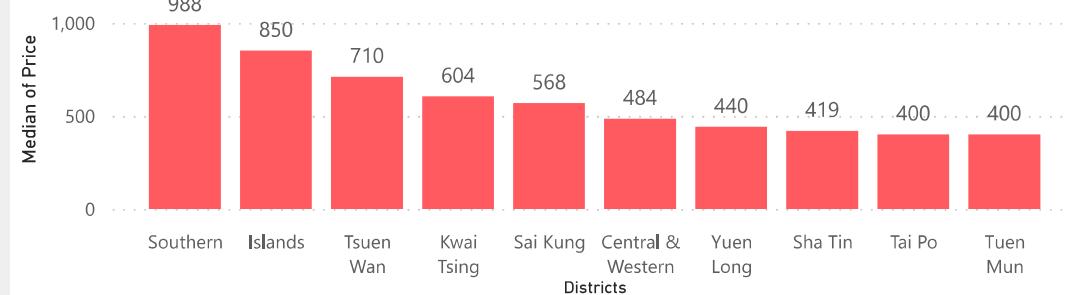
## Distribution of Room Type



## Distribution of Region



## Median of Price by Districts



Isaac Lee | Nicole Tang

Airbnb HK Business Development Analysis



# HONG KONG OPERATIONAL PERFORMANCE

**358**

Total Listing

**83**

New Listing 2021

**67**

Listing Available in Future 30D

Future 30Day Hot

All

Future 90Day Hot

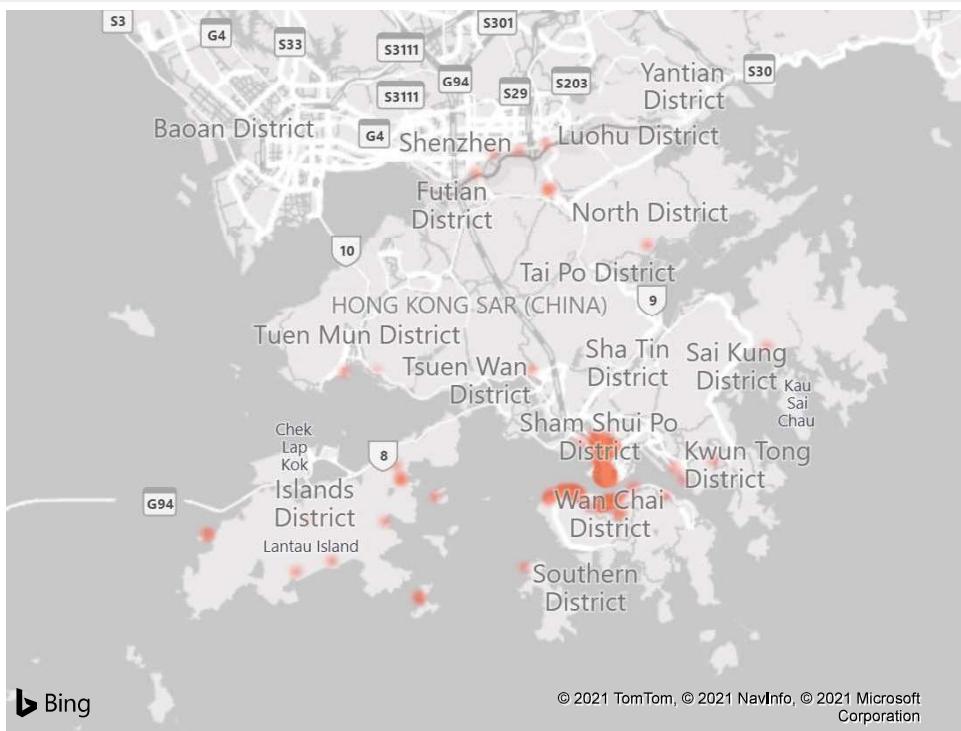
HOT (15>X)

Future 365Day Hot

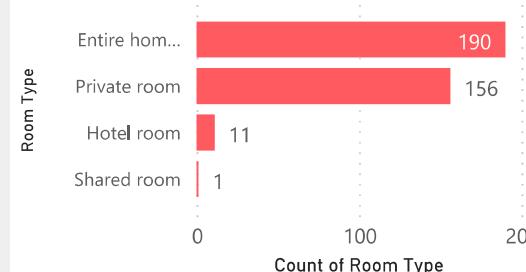
Multiple select... ▾

District

All ▾



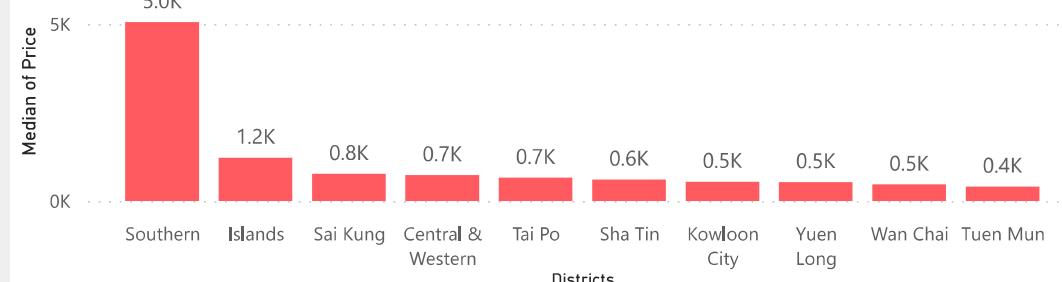
## Distribution of Room Type



## Distribution of Region



## Median of Price by Districts



Isaac Lee | Nicole Tang

Airbnb HK Business Development Analysis



# HONG KONG SALES PREDICTION

**5.78M**

21-08 Estimated Services Fees

**5.66M**

21-09 Estimated Services Fees

**9233**

Listings (for calculation)

**2715**

Hosts (for calculation)

Future 90Day Hot

All

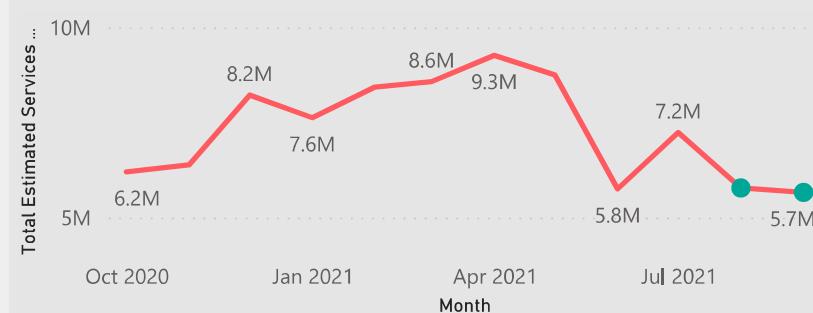
Future 365Day Hot

All

## 21-08, 21-09 Estimated Services Fees By Room Types



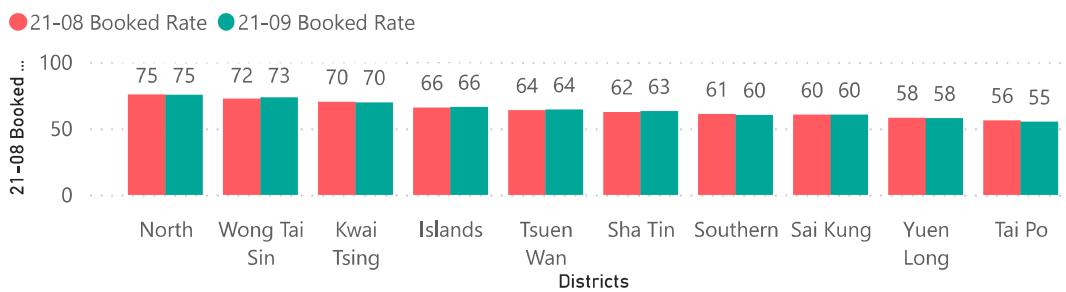
## Total Estimated Services Fees By Month



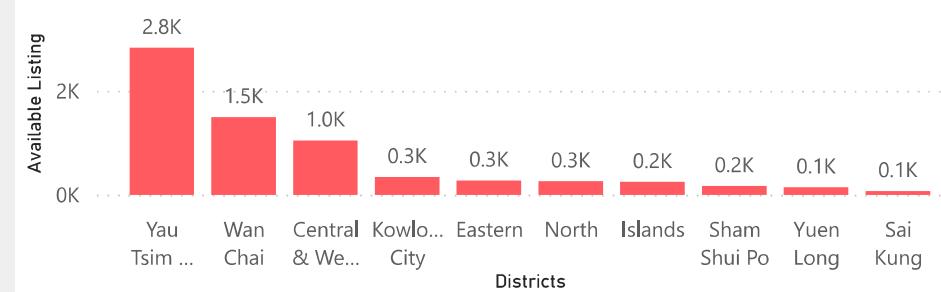
## Remark:

1. data of **Aug, Sept** are the predicted results.
2. Booked rates are predicted with **Time-series model, FaceBook Prophets** since the limitation of data.
3. The prediction results are mainly for finding the trend as insights.

## Top 10 Predicted Booked Rate by Districts



## Top 10 Predicted Available Listing Districts



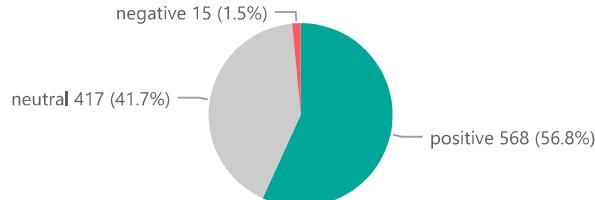
Isaac Lee | Nicole Tang

Airbnb HK Business Development Analysis

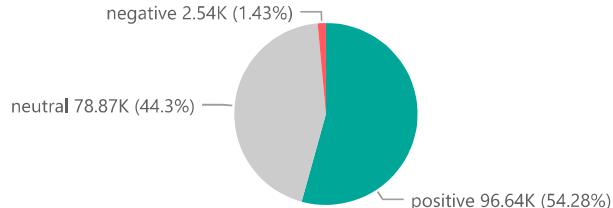


# CUSTOMER SENTIMENT ANALYSIS

## Count of Sentiment by Listing



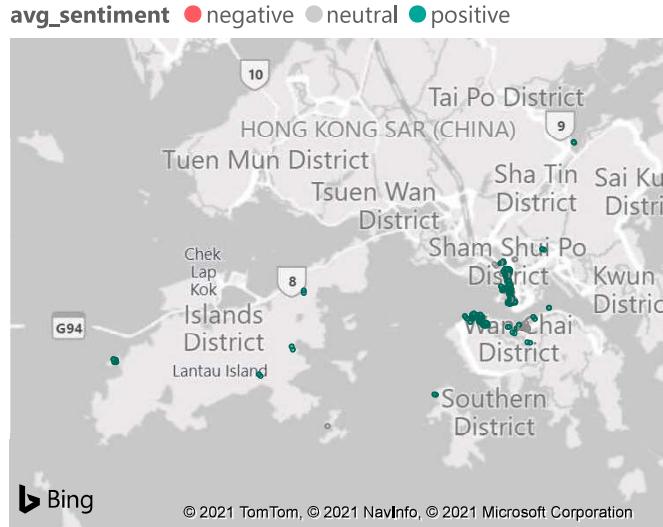
## Sentiment Percentage by Comment



## Key Word from Comments



## Sentiment Result By Location



Comment Sentiment

All

Districts

All

Feature Sentiment

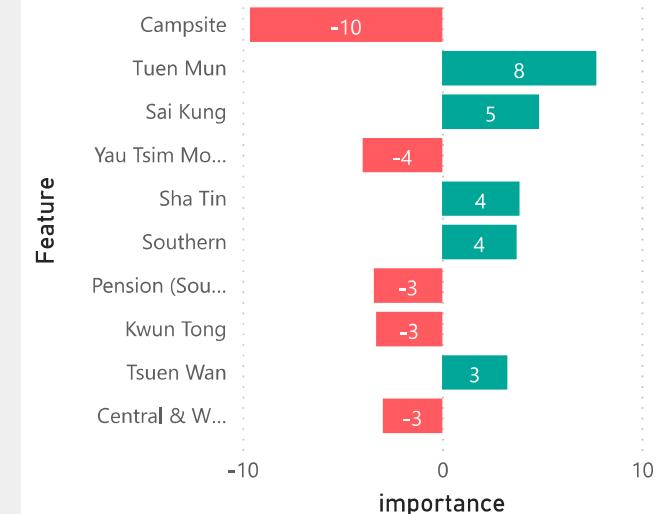
Select all

Bad Feature

Good Feature

Irrelevant Feature

## Feature Sentiment



Isaac Lee | Nicole Tang

Airbnb HK Business Development Analysis



# HOST QUALITY ANALYSIS

## Superhost Count by month in 2020-2021 and Jul-2021 Prediction

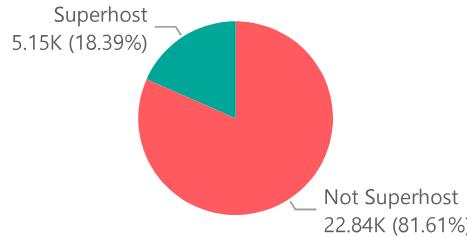
Superhost ● Not Superhost ● Superhost



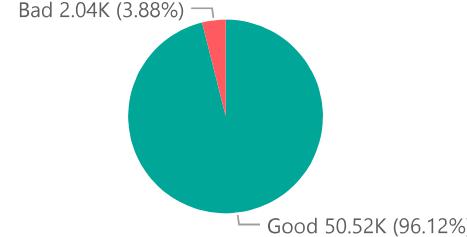
Year, Month

- Select all
- 2020
- 2021
- 1
- 2
- 3
- 4
- 6
- 7

## Overall Super Host Percentage (20.10 - 21.06)



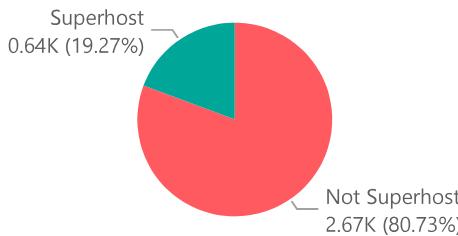
## Overall Good & Bad Score (20.10 - 21.06)



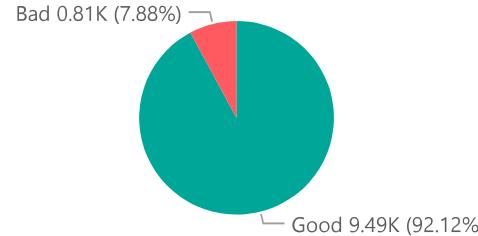
## Remark:

Data of **Jul** is the predicted results from Logistic Regression/ Adaboost Classification.

## 21.07 Super Host Prediction



## 21.07 Good & Bad Score Prediction



**Super-host** are experienced hosts who provide a shining example for other hosts, and extraordinary experiences for their guests.

It is our existing **standard to identify the quality host** at the same time.

## Requirement to become a Super-host :

1. Have hosted for at least 1 year
2. Completed at 100 nights
3. Response Rate at 90% or higher
4. Have a cancellation rate of 1% or lower



# HOSTS RFM ANALYSIS

**46.29**

Avg. Last Profitable Date

**73.90**

Avg. Review Score Rating

**25.66K**

Avg. Monetary Value

## RFM (tier 1 - best; tier 4 - Worst)

### Last Profitable (R)

The last time host making profit from the listing. (**Shorter the better**)

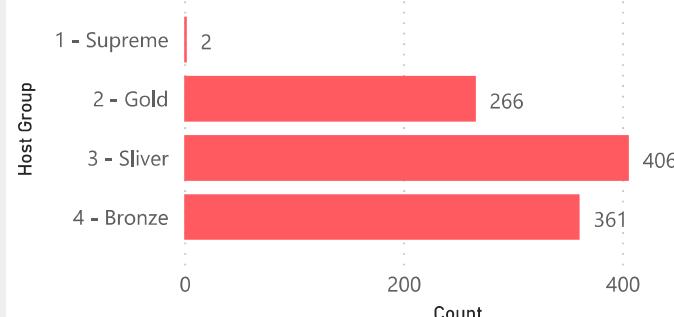
### Review Rating (F)

Performance rating of the host valued as the reviews score. (**Higher the better**)

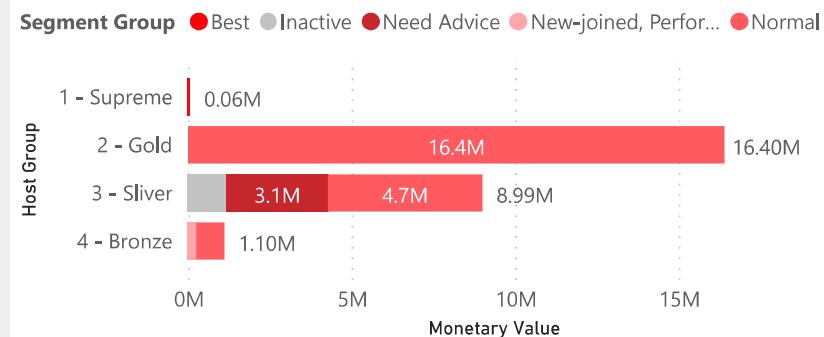
### Monetary Values (M)

Services fee we estimated earned from the host. (**Higher the better**)

## Distribution of Host Segmentation (Group)



## Monetary Value of Host Segmentation (RFM Score)



Host ID	Name	Superhost	Total Listing	Join Date	Last Profitable Day	Review Rating	Monetary Value	Segment
10056448	Annie	Superhost	4	Saturday, November 16, 2013		2	76.54	76,836.60 121
10086415	Alick	Superhost	2	Sunday, November 17, 2013		2	72.42	105,530.70 131
101171371	君曦	Superhost	4	Tuesday, October 25, 2016		83	80.00	3,195.75 323
101342921	BoBo	Not Superhost	2	Wednesday, October 26, 2016		2	76.13	98,400.00 131
101777143	Paul	Not Superhost	1	Sunday, October 30, 2016		83	97.00	1,277.25 314
101779290	Dorothy	Not Superhost	9	Sunday, October 30, 2016		2	80.55	27,601.20 121
101834949	M	Superhost	3	Sunday, October 30, 2016		3	79.46	1,659.00 223
102329302	Sophie	Not Superhost	1	Thursday, November 03, 2016		2	5.00	1,164.00 144



## RECOMMENDATION HOST

### RELATIONSHIP MAINTENANCE

#### Host Group That Need to Take Action

Host Group ● 1 - Supreme ● 3 - Sliver ● 4 - Bronze



**\*\*RFM Pattern = 141 142\*\***

#### Valuable Host Who Need Advice

##### Description

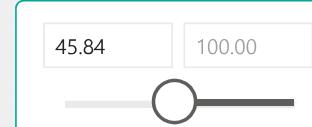
There are hosts who are **actively and as one of the largest services fee sources** but the quality of services provided to customers need to be improved.

##### Recommendations

It's good to provide them **follow-up advice in order to give them terrific incentives to enhance their services.**

### SUPER-HOST MECHANISM

#### REVIEW RATING

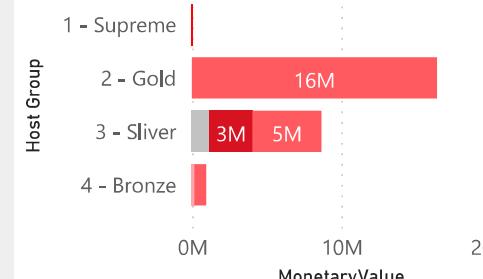


#### RFM SCORE



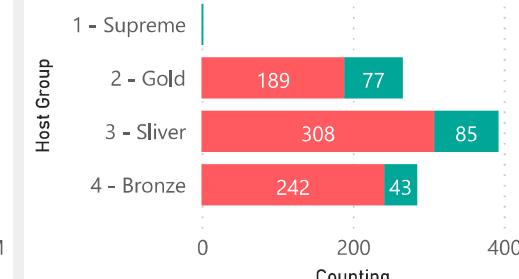
#### Superhost by Host Group

segment... ● Best ● Inactive ● Need Advice ➡



#### Superhost by Host Group

Superhost ● Not Superhost ● Superhost



#### Super-host Mechanism Review

- Effectively classify host in general as the first criteria, as it match the results with RFM analysis.

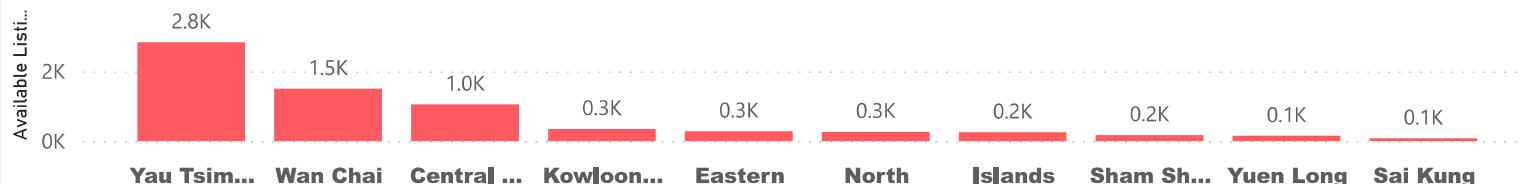
#### Remark:

Hosts are divided into group with its score:  
 Supreme = 3      Gold = 4-5  
 Silver = 6 -7      Bronze = 8 -12



## RECOMMENDATION LISTING

### Predicting the districts to have highest room availability in 2021-08



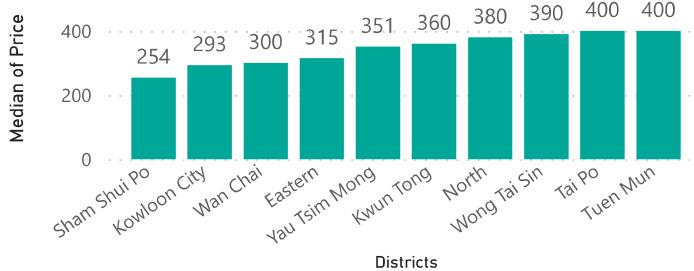
### Highlight

- Yau Tsim Mong has the most available rooms, but low booking rate (~40%)
- Supply is larger than Demand
- > Price comparably lower than the market
- Location or Service Level **below average**.

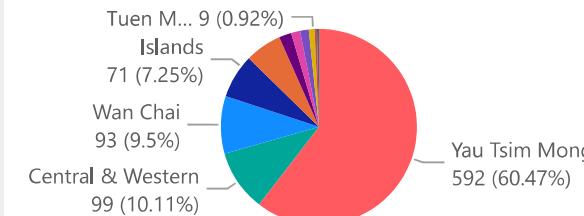
### Forecasting lowest booking rates by districts for 2021-08 & 2021-09



### Bottom 10 Overall Median Price by Districts



### Mid-Term & Long-Term Booking Distribution by Districts



### Potential Business Development

Work as **intermediary** to connect and cooperate with different hosts and companies to provide the staffs a standardized apartment service with a discount long-stay contract.

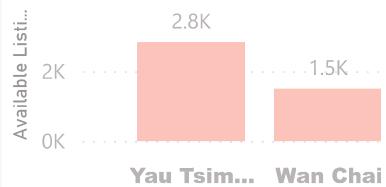
Case Example 1

Case Example 2

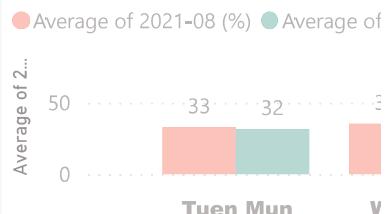


## RECOMMENDATION LISTING

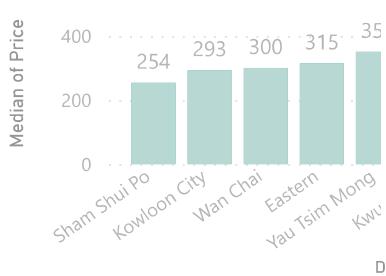
### Predicting the districts to have more available listings



### Forecasting lowest booking rate



### Bottom 10 Overall Median Price



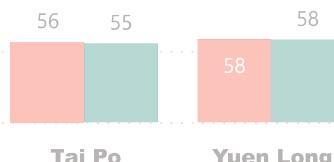
**Case 1**



- Commercial Building
- Residential Building

### Highlight

**Tsim Mong** has the most available ms, but low booking rate (~40%) Supply is larger than Demand Price comparably lower than the market location or Service Level **below average**.



### Business Development

Intermediary to connect and cooperate between hosts and companies to provide the standardized apartment service with a long-stay contract.

Case Example 1

Case Example 2



## OPERATION & SALES

- Maintain a satisfactory revenue in Hong Kong in 2021.
- As well as in the predicted revenue in future months by using Time-series models (Facebook Prophet).

## SERVICE QUALITY

- Super-host: The number of super-host would be maintaining and account for around 20% of total host base from the prediction result.
- The percentage of good rating given by the residents are maintained above 90% of total.

## HOSTS TARGETING (RFM ANALYSIS)

- Targeted 4 group of hosts, New-joined performed-badly, Need advice, Inactive (churned), Best performed.
- Suggested direction of maintaining relationship with different groups.

## POTENTIAL DEVELOPMENT

### 1) Relationship between booking rate and number of available rooms in CBD

- Location & Service Level lower than others, services improvements.

### 2) Potential Business Development

- Step-up as connector to link-up hosts and companies in that district.



Progress Checklist

Stage	Details	Progress	Remark
Preparation	Draft Proposal	✓	
	Consultation with Sat & Oillio (Lecturer)	✓	
Data Collection & Preprocessing	Data Collection and Merging	✓	
	Data management and queries - Data description	✓	
Data analysis	EDA - data checking, cleaning	✓	
	Decides KPI in use	✓	
	Machine Learning Preprocessing - Features Selection, prepare files for model training	✓	
	Host analysis - RFM analysis	✓	
	Customer analysis - Sentiment analysis	✓	
	Business Insights - Summary (Region level)	✓	
Machine Learning	Time-series model Building	✓	
	Models Selection and Start Training for Host analysis	✓	
	Combine result	✓	
Dashboard	Decide final outline of the materials	✓	
	Dashboard deployment	✓	
Testing	Final testing and confirm that deliverable is available to view	✓	