



# **BUDT 758T - Hawaii**

## **Identifying Buying Targets for AirBnB Investment Property**

Team 2: Zhuxuan Xu, Jiangkun Xiong,  
Wittney Lee, Chunghao Lee, Jin Sun

# TABLE OF CONTENTS

**01**

**Kaggle Competition**

**02**

**Hawaii  
Market Research**

**03**

**Hawaii  
Model Analysis**

**04**

**Question and Finding**

**05**

**Conclusion**

A tropical-themed background featuring a large yellow sun partially obscured by a blue horizon line. The sky is light blue with two white clouds. Palm fronds are visible in the corners. The number '01' is prominently displayed in the center.

**01**

# **Kaggle Competition**

# Kaggle Competition

## Data processing steps:

- Convert variables into numeric or factor types & fill NA values.
- Construct four models: Logistic regression, Bagged Trees, Random forest and XGBoost.
- Compare AUC and choose the best model.

```
```{r}
# Function to convert to numeric
ConvertNumeric = function(dfCol, NAvalue){
  val <- (gsub("\\$", "", dfCol))
  val <- (gsub("\\%", "", val))
  val <- suppressWarnings(as.numeric(gsub("\\,", "", val)))
  val[is.na(val)] = NAvalue
  return(val)
}

# Function to convert to factor
ConvertFactor = function(dfCol, GoodVector, NAvalue){
  dfCol[!(dfCol %in% GoodVector)] = NAvalue
  dfCol = as.factor(dfCol)
  return(dfCol)
}
```

Model	AUC
Logistic	0.8765
Bagged Trees	0.9092
Random Forest	0.9318
XGBoost	0.9399

# Variables used

**Numeric Variables:** accommodates, availability\_30, availability\_60, availability\_90, availability\_365, bathrooms, bedrooms, beds, amenities, cleaning\_fee, extra\_people, guests\_included, host\_listings\_count, host\_response\_rate, host\_verifications, maximum\_nights, minimum\_nights, monthly\_price, price, review\_scores\_accuracy, review\_scores\_checkin, review\_scores\_cleanliness, review\_scores\_communication, review\_scores\_location, review\_scores\_rating, review\_scores\_value, security\_deposit, weekly\_price

**Factor Variables:** bed\_type, cancellation\_policy, host\_identity\_verified, host\_has\_profile\_pic, host\_is\_superhost, host\_response\_time, host\_since, instant\_bookable, is\_location\_exact, market, property\_type, require\_guest\_phone\_verification, require\_guest\_profile\_picture, requires\_license, room\_type

A tropical-themed background featuring a large yellow sun partially obscured by the number '02'. The sun is set against a light blue sky with two white clouds. Palm fronds are visible in the corners. The bottom of the image shows a blue body of water and a tan sandy beach.

**02**

# **Market Research**

# Quick Look on Hawaii Market



Most to Least Visited Places:

Oahu 49.6% > Maui 24.4% > Big Island 14.5% > Kauai 11.5%

# Quick Look on Hawaii Market

## Visiting Purpose

- Pleasure/Vacation
  - Honey Moon
  - Wedding
  - Conference



## Length of Stay

- First time visitors: 31%, 8.24 days
- Repeating visitors: 68%, 9 days

## Total Visitor Spending

- Increasing since 2003 (to 2019)
- Lodging, largest spending category



## Total Property

- 14247



# Kaggle Part (Team 2)



## Data Processing

- Convert selected variables into either numeric variables or factor variables
- Replace NAs for numeric variables with the average value or the highest frequency value
- Substitute NAs for factor variables with “Other” or the highest frequency value



## Final model

XGboost  
Parameters:  
Trees = 1000,  
`set_engine("xgboost")`,  
`set_mode("classification")`



## AUC

0.9399



## Variables

Numeric variables: `accommodates`, `availability_30`, `availability_60`, `availability_90`, `availability_365`, `bathrooms`, `bedrooms`, `beds`, `amenities`, `cleaning_fee`, `extra_people`, `guests_included`, `host_listings_count`, `host_response_rate`, `host_verifications`, `maximum_nights`, `minimum_nights`, `monthly_price`, `price`, `review_scores_accuracy`, `review_scores_checkin`, `review_scores_cleanliness`, `review_scores_communication`, `review_scores_location`, `review_scores_rating`, `review_scores_value`, `security_deposit`, `weekly_price`

Factor variables: `bed_type`, `cancellation_policy`, `host_identity_verified`, `host_has_profile_pic`, `host_is_superhost`, `host_response_time`, `host_since`, `instant_bookable`, `is_location_exact`, `market`, `property_type`, `require_guest_phone_verification`, `require_guest_profile_picture`, `requires_license`, `room_type`

A tropical-themed background featuring a large yellow sun partially obscured by a blue horizon line. The sky is light blue with two white clouds. Palm fronds are visible in the corners. The number '03' is prominently displayed in the center of the sun.

**03**

# **Model Analysis**

# Models - variables and performance

## Kaggle variables

```
recipe1 <-  
recipe(high_booking_rate ~ ., data = dfa_analysis) %>%  
  step_rm(id, access, city, description, host_about, host_acceptance_rate, host_location, host_neighbourhood, house_rules, interaction,  
  is_business_travel_ready, latitude, longitude, neighborhood_overview, neighbourhood, notes, space, state, transit, zipcode, square_feet,  
  '{randomControl}')
```

.metric <chr>	.estimator <chr>	.estimate <dbl>
roc_auc	binary	0.9319582

## Hawaii model with updated variables

.metric <chr>	.estimator <chr>	.estimate
roc_auc	binary	0.9374196

## Models: XGboost

## Variables Selection

- Weekly\_discount replace weekly\_price  
$$= 1 - \text{weekly\_price} / (\text{daily\_price} * 7)$$
- Add specific amenity columns  
(True/False): Family-kid friendly,  
parking, washer, dryer, hair dryer, Wifi

# Models - Cutoff and Cost Matrix

Costs at different cutoff

Cutoff	Cost
0.3	2318
0.4	1832
0.5	1463
0.6	1187
0.7	992

**Models: XGboost**

## Cut-off determination

This is a business advice for property purchasing, so we'd like lower False Positive. In cost matrix, we set cost of False positive 3 times higher than true positive. For the cost matrix, we found that cutoff = 0.7 has lowest cost.

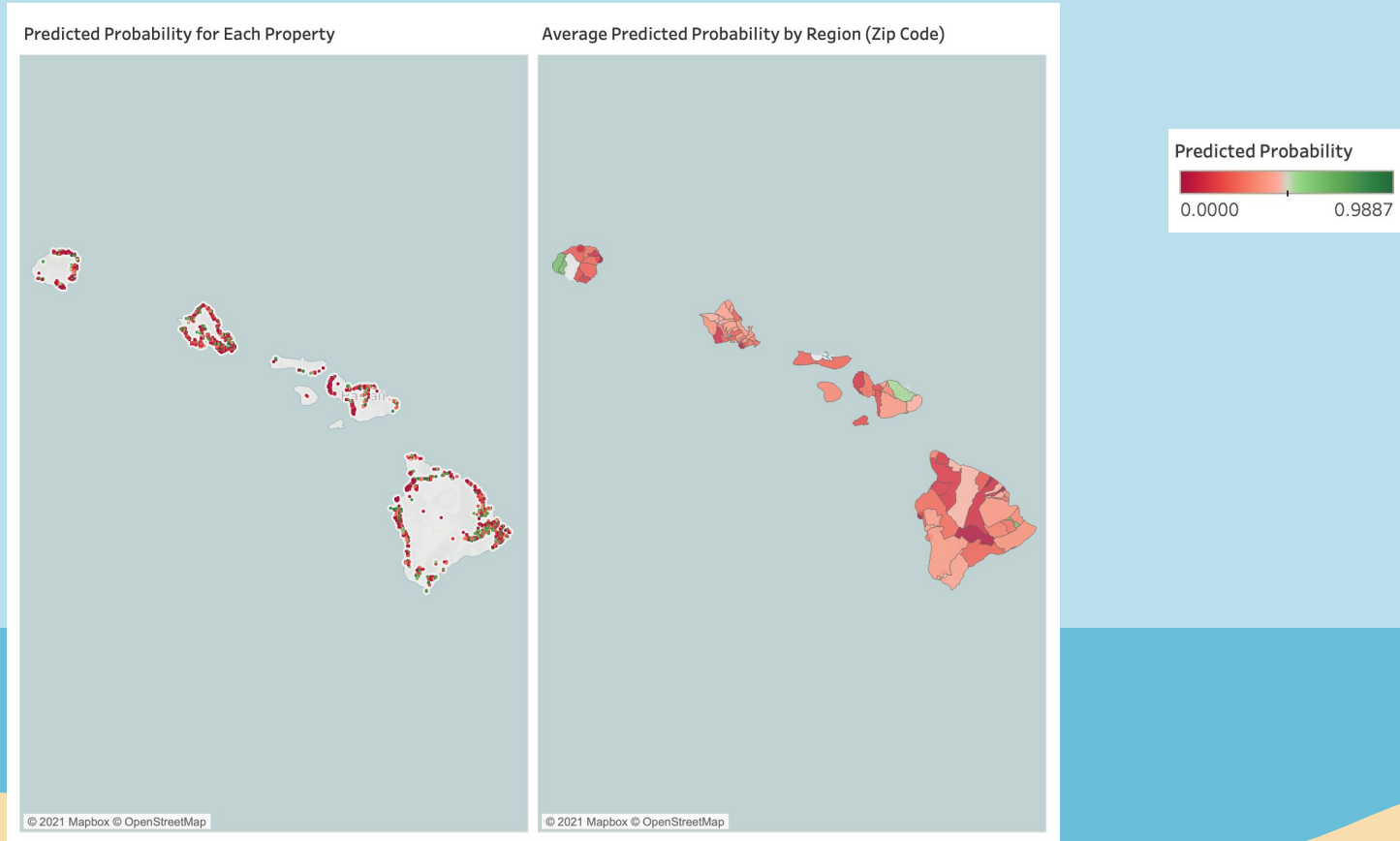
Cost Matrix		Actual	Class
		0	1
Predicted	0	0	1
Class	1	3	1

A tropical-themed background featuring a large yellow sun partially obscured by a blue horizon line. The sky is light blue with white clouds and brown palm fronds in the corners. The foreground is a solid blue band.

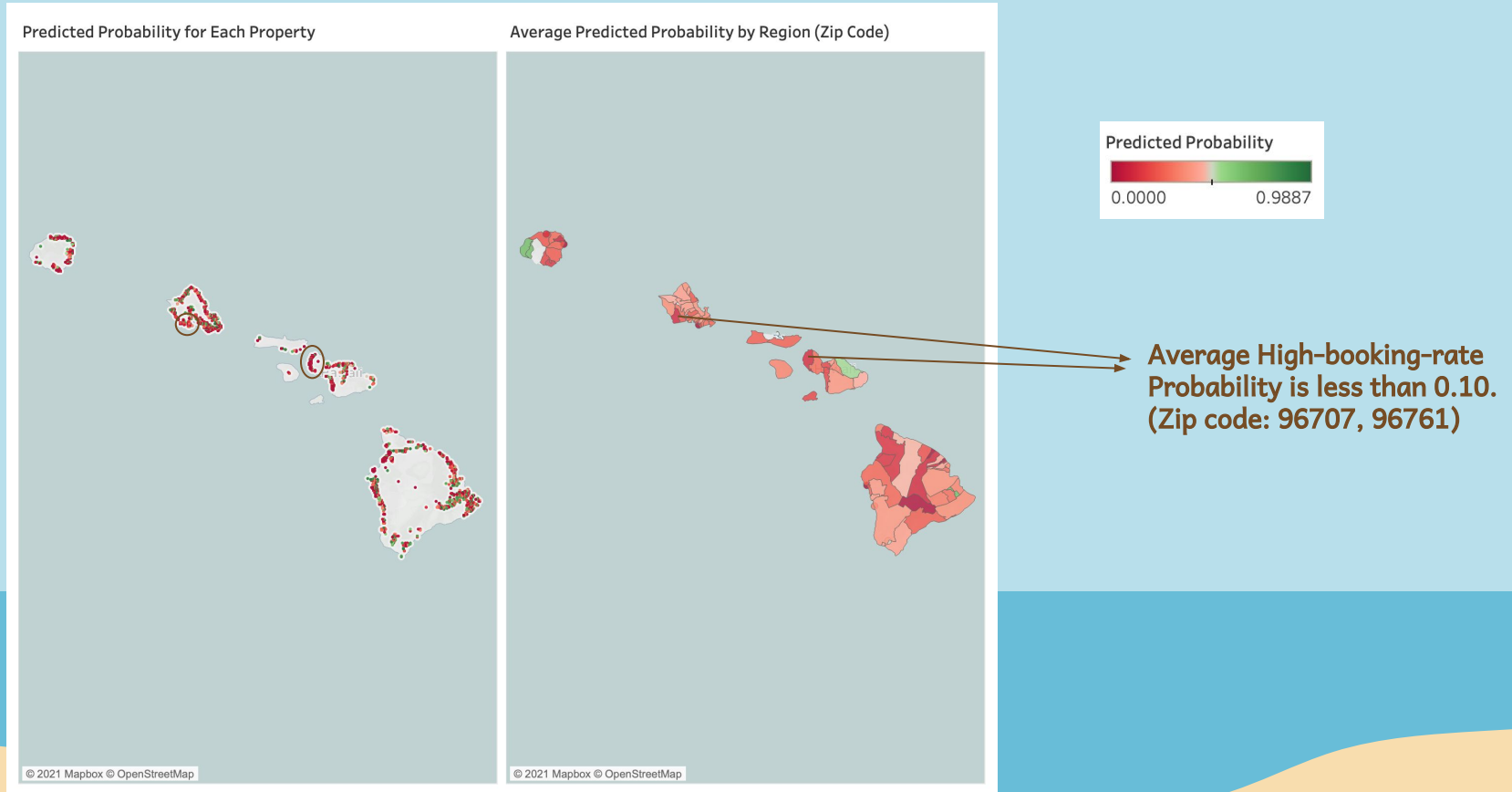
**04**

# **Research Questions**

# Q1: Which region is not good for investment?



# Q1: Which region is not good for investment?



## Q2: How does weekly price discount affect probability?

Weekly Discount Interval	Number	High-booking-rate Probability
$\geq 30\%$	92	0.4546
20% ~ 30%	127	0.6417
10% ~ 20%	197	0.6730
0% ~ 10%	183	0.6021
No discount	177	0.5967

High-booking-rate Probability for property cannot stay higher than a week = 0.2124

Two-sample t test for unequal sample size (unequal variance):

- Difference between 10% ~ 20% and no discount is significant (p-value = 0.01).
- Difference between 10% ~ 20% and 20% ~ 30% is not significant (p-value = 0.3623).
- Difference between 10% ~ 20% and 0 ~ 10% is significant (p-value = 0.02).

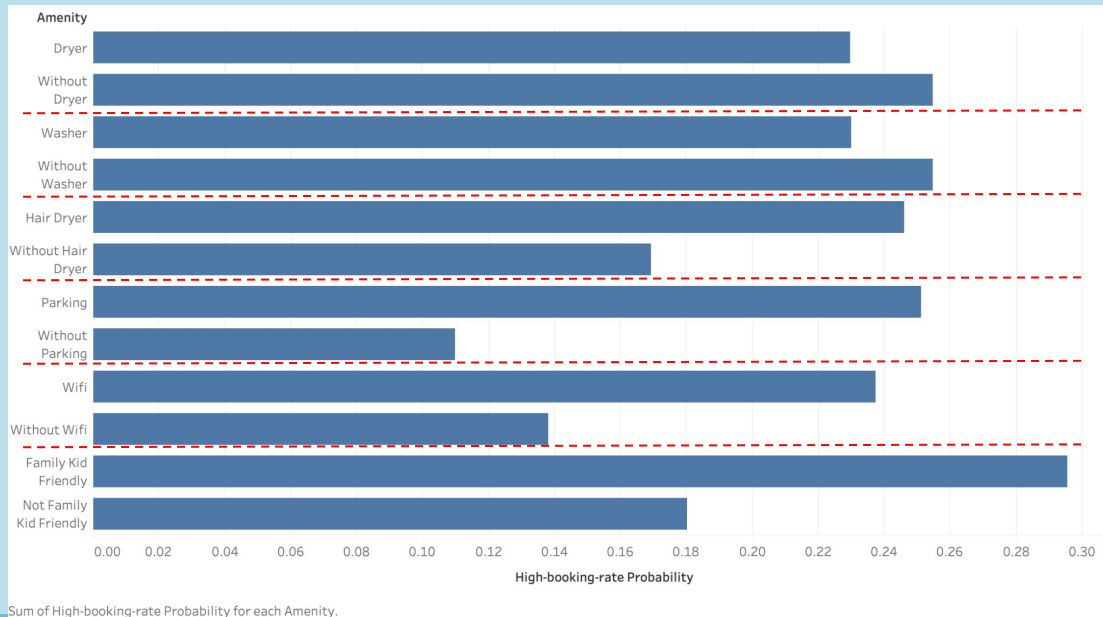
Management insights:

→ Allow customers to book more than a week (maximum\_nights  $\geq 7$ ).

→ Choose **10% to 30%** weekly discount as the price management strategy!



### Q3: How do provided amenities affect probability?



#### Management insights:

- Add Wifi and Hair Dryer equipments into the property.
- Try to add parking service.
- Washer & Dryer not important.
- Provide family-kid friendly service and equipments.

A tropical-themed background featuring a large yellow sun partially obscured by a blue horizon line. The sky is light blue with two white clouds. Palm fronds are visible in the corners. The number '05' is prominently displayed in the center.

**05**

# **Conclusion and Discussion**

# Conclusion

	DESCRIPTION
<b>Problem</b>	The airbnb/ short rental market is one of the most competitive and ever-changing markets for real estate investors.
<b>Goal</b>	Screen & Identify the right properties to invest
<b>Method</b>	A projection model to identify the properties with high booking rates and thus high income generating potential.
<b>AUC</b>	94%
<b>Limitations</b>	?

# Investor's Goal

High Booking Rate, Avg. Length of Stay

Operating Income

← Occupancy, Rental Pricing

Operating Cost

← Property Age, Property Price (& Mortgage)

Net Operating Income



Also consider :

- Investor's Budget
- Property Location



# References

Major Cities/Locations: Honolulu, Haleiwa, Kailua-Kona, Hawi, Captain Cook, Kihei, Paia, Princeville, and Kapaa (location)

<https://www.hawaii-guide.com/selecting-the-best-hawaiian-island>

<https://www.hawaii-guide.com/hawaii-tourism-statistics>

<https://blog.turnkeyvr.com/a-guide-to-hawaii-short-term-rental-regulations/>

<https://www.hawaii.house/articles/hawaii-investment-properties-airbnb-effect/>

<https://www.bizjournals.com/pacific/news/2017/02/22/top-airbnb-listings-in-honolulu-and-how-much.html>

<https://www.hawaiilife.com/blog/hawaii-vacation-rental-faqs/>  
recreational interests of visitors are changing: hiking > golfing

<https://www.civilbeat.org/2019/07/9-charts-that-show-how-hawaii-tourism-is-changing/>

Illegal vacation rentals problems

Airbnb horror stories analysis

<https://www.asherfergusson.com/airbnb/>

Buying Property for Airbnb (location, type, accessibility, neighborhood)

<https://www.fool.com/millionacres/real-estate-investing/rental-properties/buying-property-airbnb/>



# THANK YOU!

## Q&A

