



# Austin Airbnb Market Analysis- Team 3

- Dharkar Gauri
- Dodamani Vishal
- Jethwa Shlok
- Joshi Jay
- Parekh Malav
- Shah Ami



# Contents

- Data Processing & Model used for Kaggle
- Why Austin Market
- Exploratory Analysis
- Business Case 1- Initial Acquisition
- Business Case 2 - Pricing and Management
- Business Case 3 - Upgrades and Renovations
- Summary for Business Cases
- Conclusion

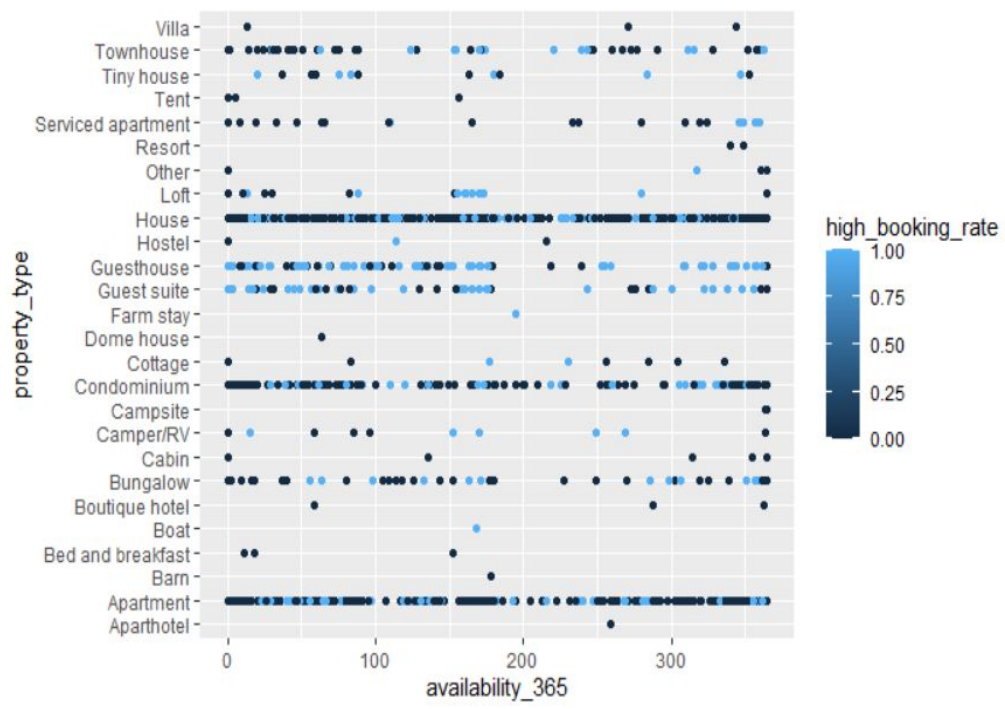
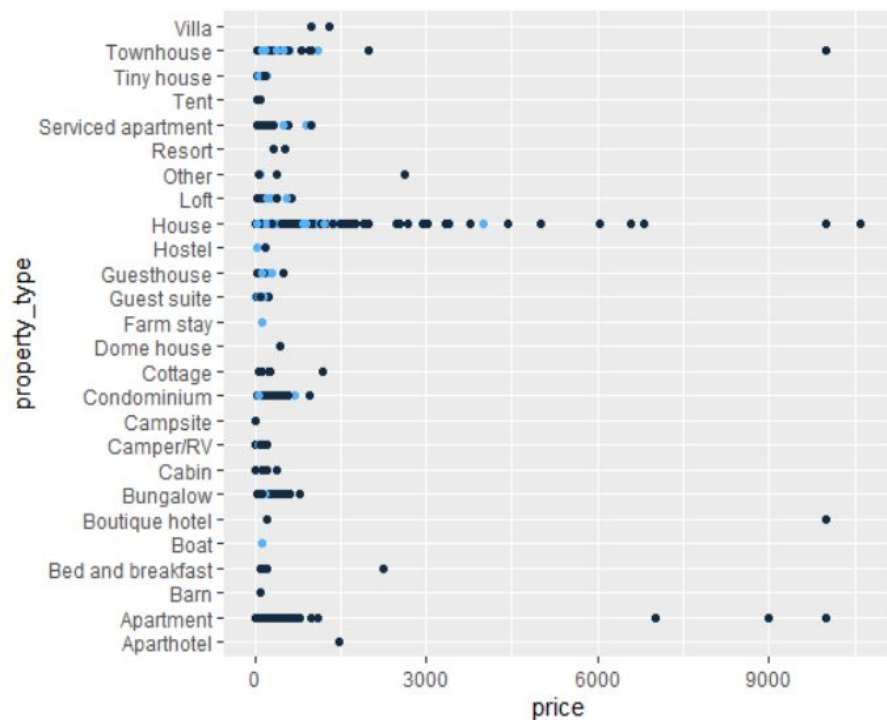
# Data Processing & Model used for Kaggle Competition

- Data Processing Steps - Cleaning the data and creating dummy variables, data transformation i.e. creating factors and levels out of normal data, and then we did attribute selection and removed some of the attributes.
- Final method used - XGBoost and the parameters were set to their default values. Except for one i.e. for eval\_metric we used AUC.
- Variables in Final Model - The variables such as host\_about, house\_rules, etc which were of less significance were removed.
- Performance(AUC) achieved - 0.803

# Why Austin Market?

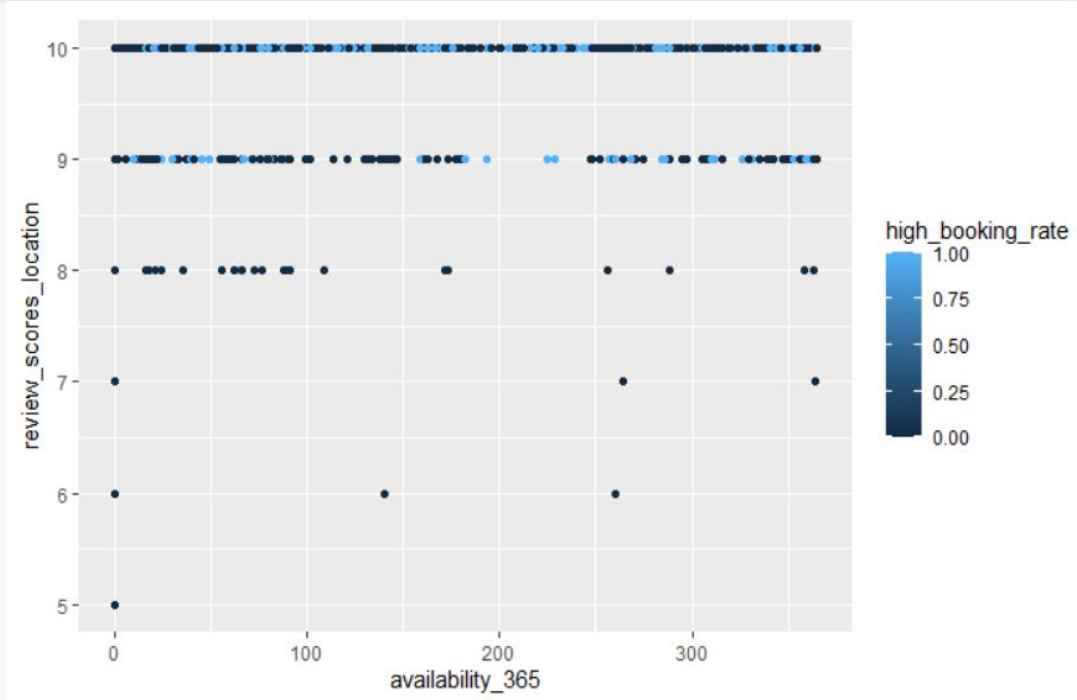
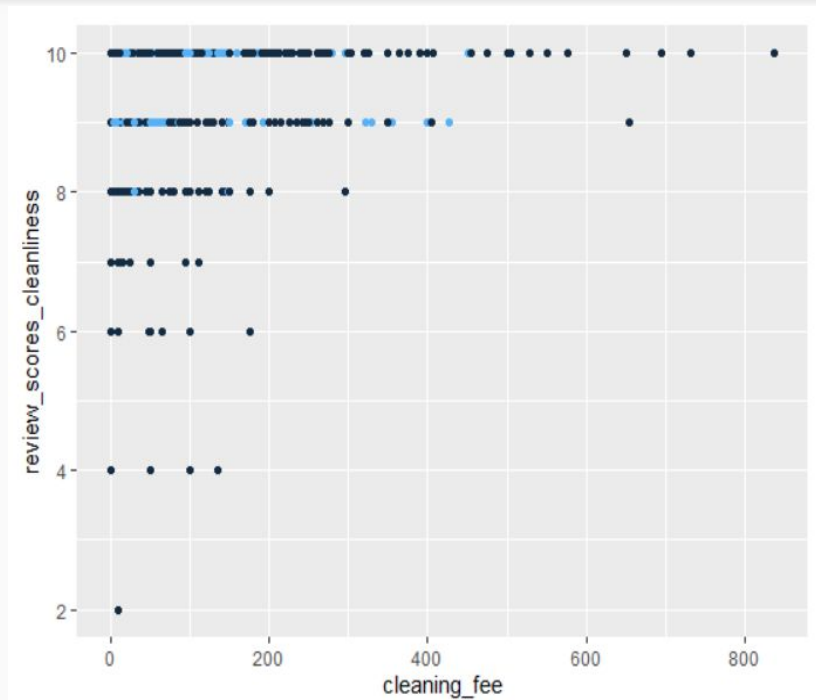
- **Cheaper to buy properties**
- **Quicker to get building permits**
- **Good ROI for Airbnb Austin** - There's clearly a demand for entire homes and so when renting it on Airbnb they can bring about \$2000 - \$4000
- **Growing future of Austin Market**- increasing jobs
- **Educational Hub** - UT Austin
- **Recreation and Activities for leisure** - Austin plays host to a large number of events such as ACL Festival and SXSW
- **Tourist Attractions** - Known for its sweeping views of the hill country, affluent communities, and an ideal location that is sandwiched between the popular Lake Travis area and downtown.

# Exploratory Data Analysis of Austin Data

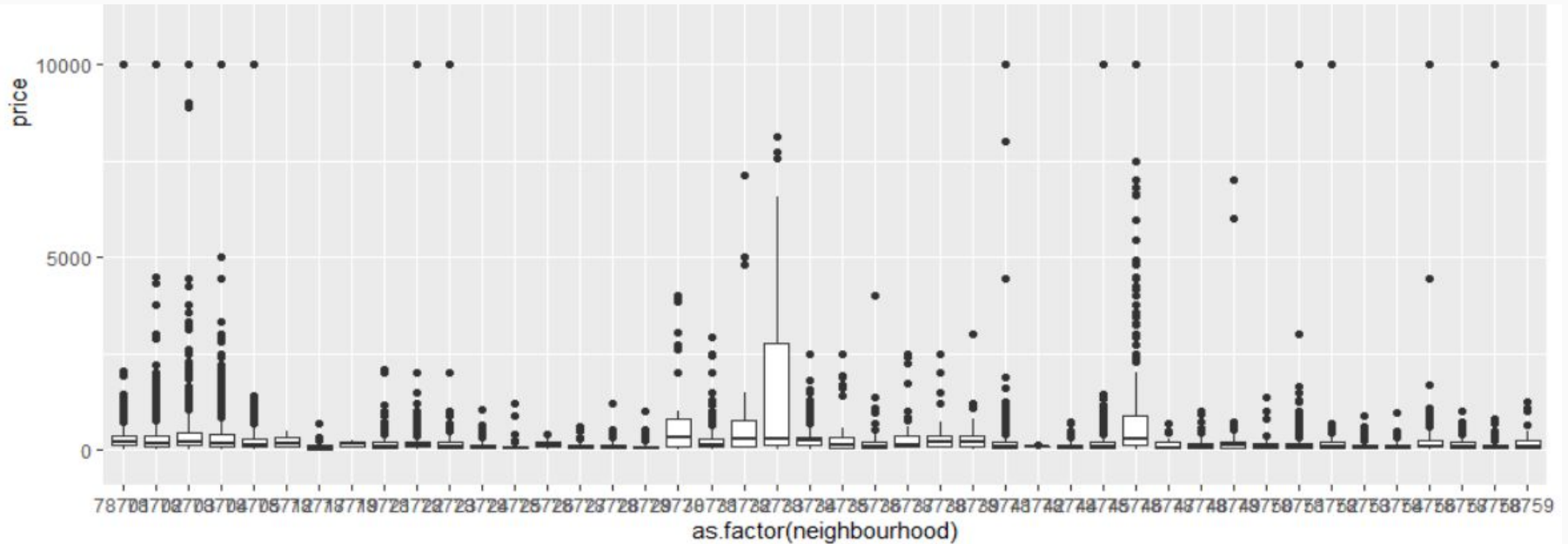




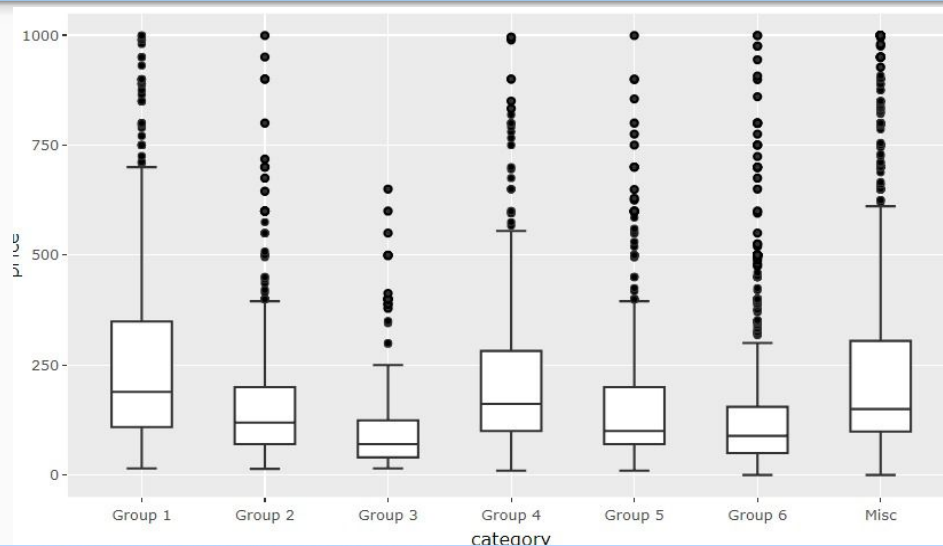
# Exploratory Data Analysis(Cont.)



# Box Plot for Neighbourhood v/s Price



# Box Plot for Prices according to different Zip Codes



```
Group1 <- c(78652,78701,78703,78705,78712)
Group2 <- c(78719,78721,78722,78723)
Group3 <- c(78717,78724,78725,78726,78727,78728,78729)
Group4 <- c(78730,78731,78732,78733,78734,78738,78739)
Group5 <- c(78735,78736,78737,78741,78749)
Group6 <- c(78742,78744,78745,78747,78748,78750,78751,78752,78753,78754,78756,78757,78758,78759)
Misc <- c(78613,78620,78650,78653,78660,78669,78746,78767,80211,78712)
```

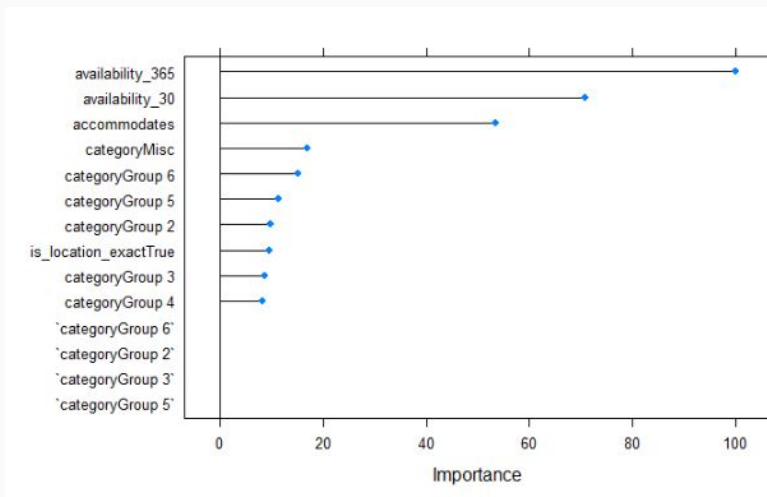


# Threshold Value for the models

- We have chosen the threshold for the models as 0.35,
- the main reason being that we should not miss any of the high booked property.
- Also, we wouldn't be at much loss even if we falsely predict not a high booking property as high booked property as the house will be an investment.

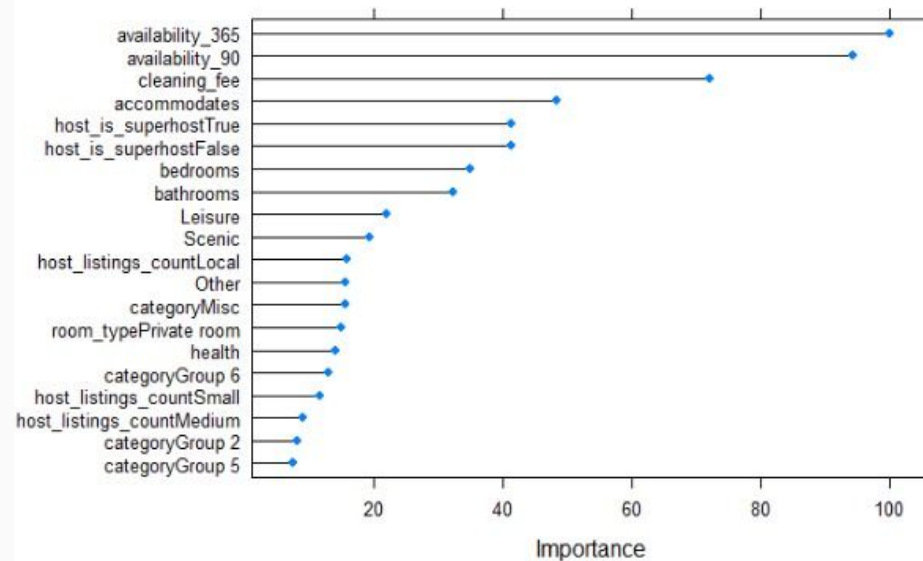
# Business Case 1- Initial Acquisition

- bagged trees ensemble methods
- Data preprocessing: clubbed zipcodes into geographical and logical groups
- Areas(zipcodes): Affordable vs luxurious
- We can see that some of the area groups have really great importance
- Used Available\_30(for weekend getaways) and Available\_365( for major event)



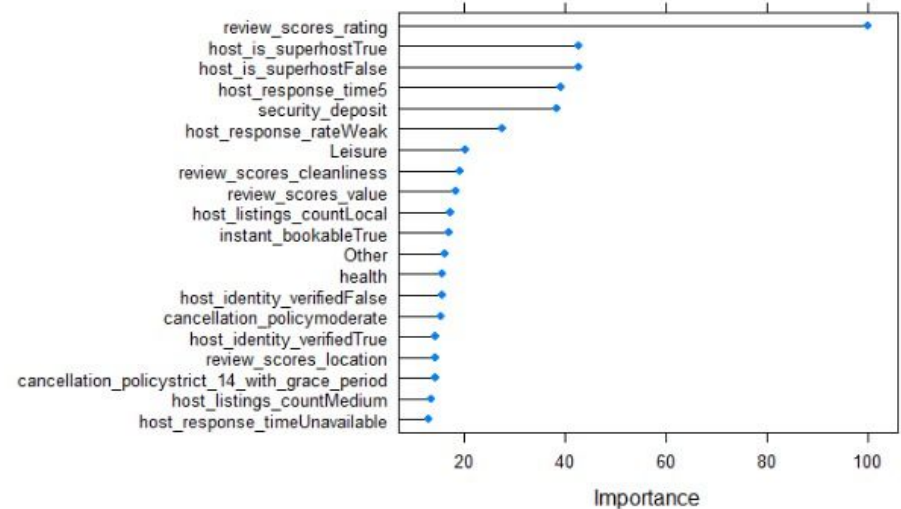
# Business Case 2(1)- Pricing

- Bagged tree ensemble method
- Graph shows variable importance for higher prediction accuracy
- Pricing most dependant on availability period, cleaning fee, accomodation, amenities and superhost or not



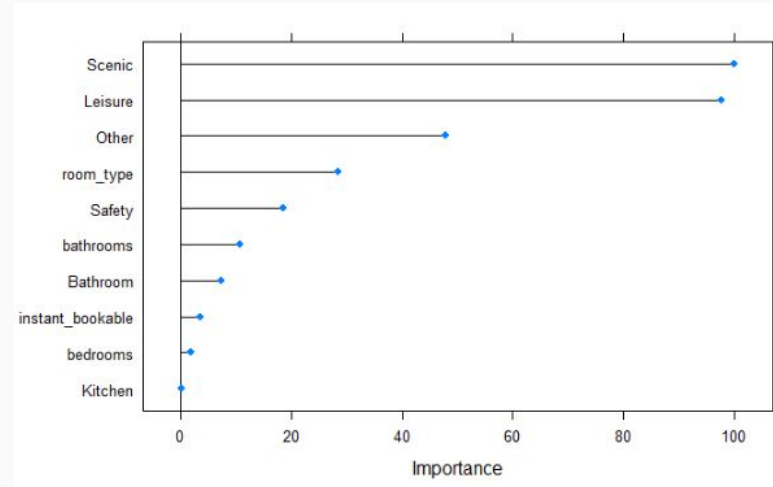
# Business Case 2(2)- Management

- Bagged tree ensemble method
- Graph shows variable importance for higher prediction accuracy
- Better review scores, host response time, host response rate help improve management
- We see that professional management is not that important



# Business Case 3- Upgrades & Renovations

- K Nearest Neighbors
- Preprocessing: Clubbed 188 unique into 11 desired categories
- People usually prefer Scenic views and leisure activities
- Upgrading to these amenities can be really worthful



# Summary for Business Cases

Business Case	Model	Accuracy	Sensitivity	Specificity
Business Case 1	Bagged Trees	72.77%	56.67%	77.79%
Business Case 2- Price	Bagged Trees	81.34%	78.33%	82.27%
Business Case 2- Management	Bagged Trees	81.45%	78.12%	82.34%
Business Case 3	KNN	72.38%	50.21%	79.29%

\*More focus on increasing sensitivity



# Conclusion

- Austin is a great place to buy properties at cheaper rate
- Along with accuracy, focus on increasing sensitivity as increased FN have more cost in this market
- Every business case demands variables of different choosing for higher performance
- Pricing depends on availability period, cleaning fee, accomodation, amenities and superhost or not
- Management depends upon review scores, host response time, host response rate
- Upgrades and Renovation depend on the property

**Thank You!**  
**Questions?**

