



$$\text{PEOPLE} + \text{PLACES} + \text{LOVE} + \text{AIRBNB} = \text{Airbnb}$$

Airbnb in Boston

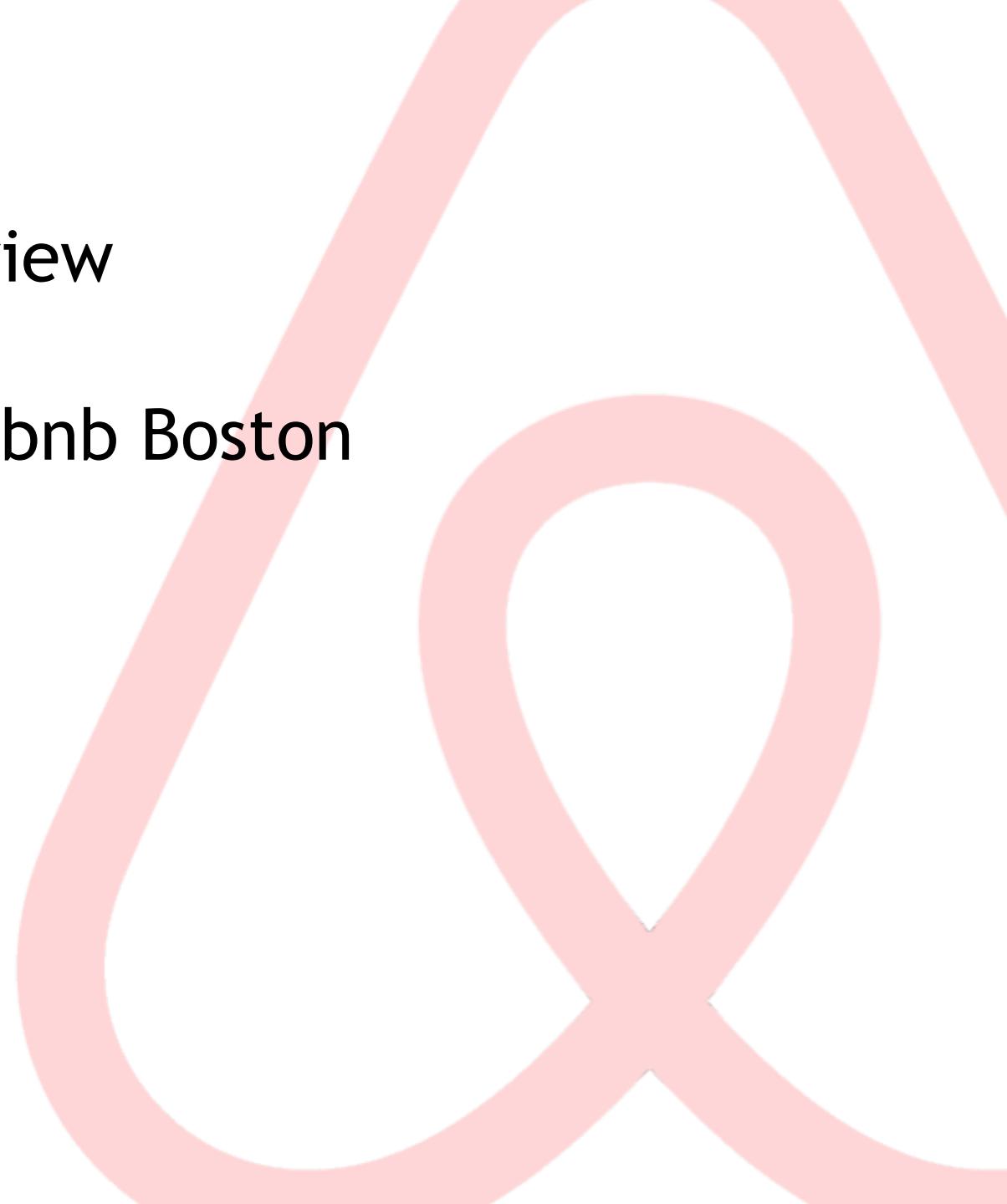
An Exploratory and Predictive Analytics

By Team SXNB

Xiang (Allen)

Shixuan(Alex)

Jiawei(Mike)



1

Industry Overview

2

Glimpse on Airbnb Boston

3

Data Mining

4

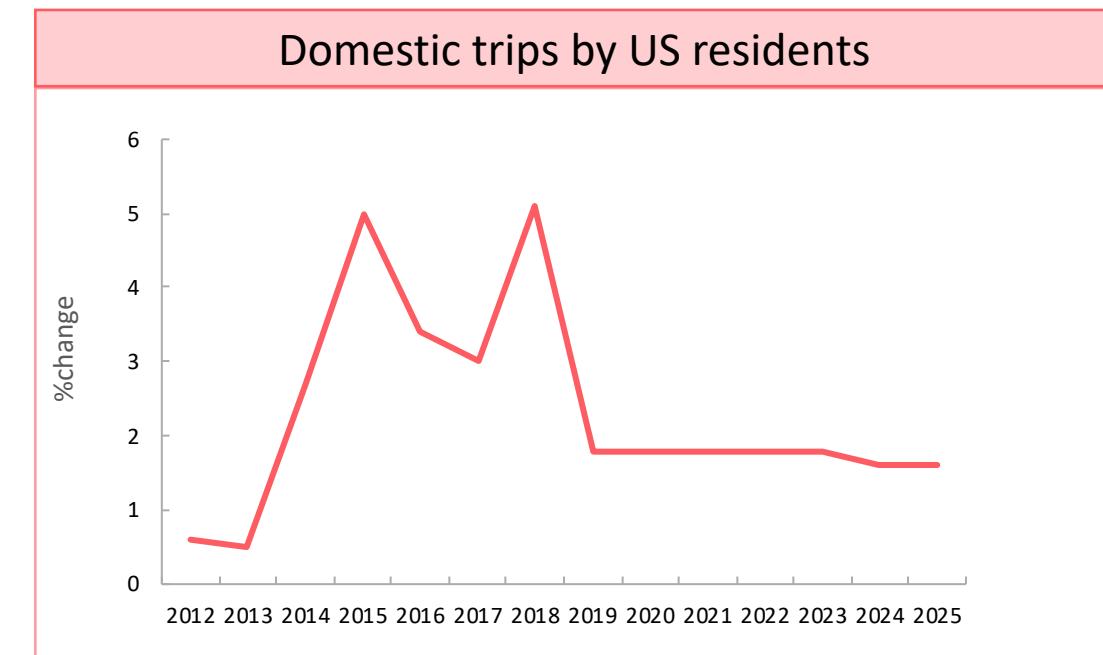
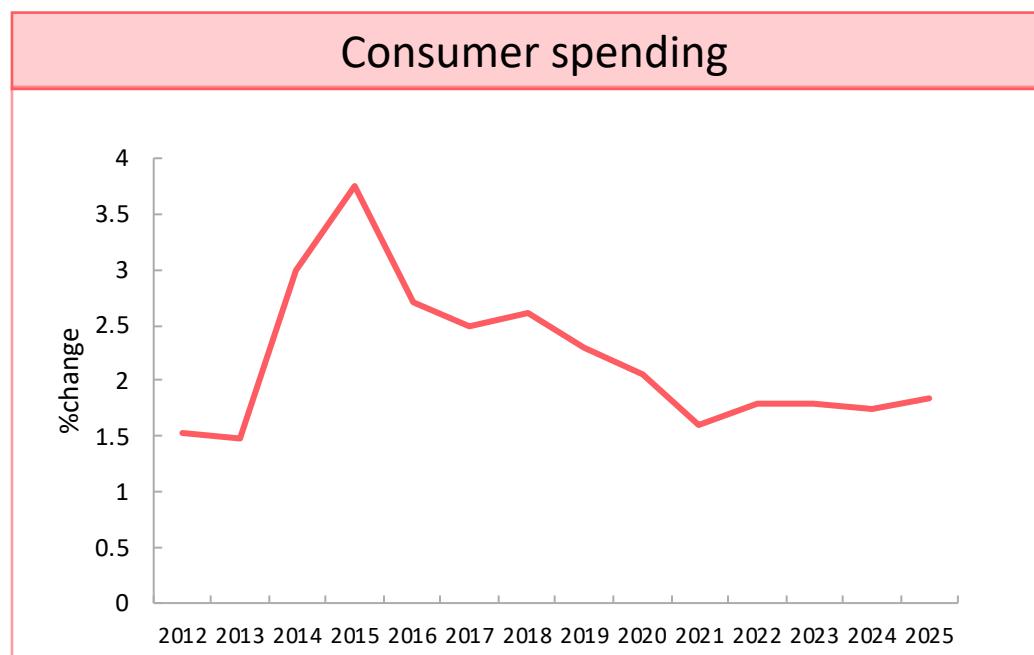
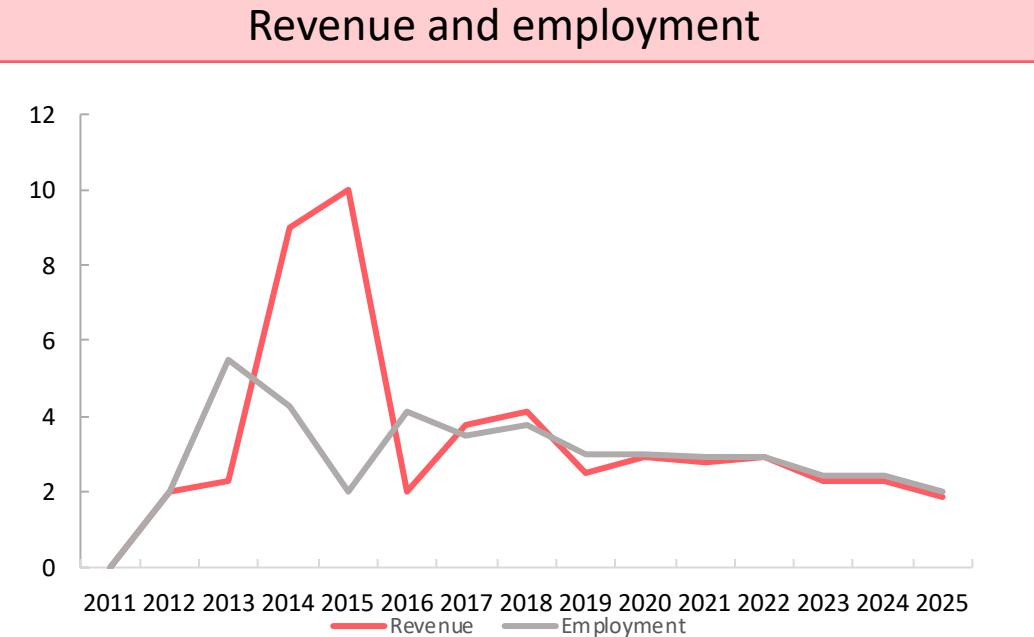
Conclusion

Industry overview

Hostel Accommodations Industry

Revenue Profit annual growth 14-19

\$3.1bn **\$208mn** **4.4%**



1

Industry Overview

2

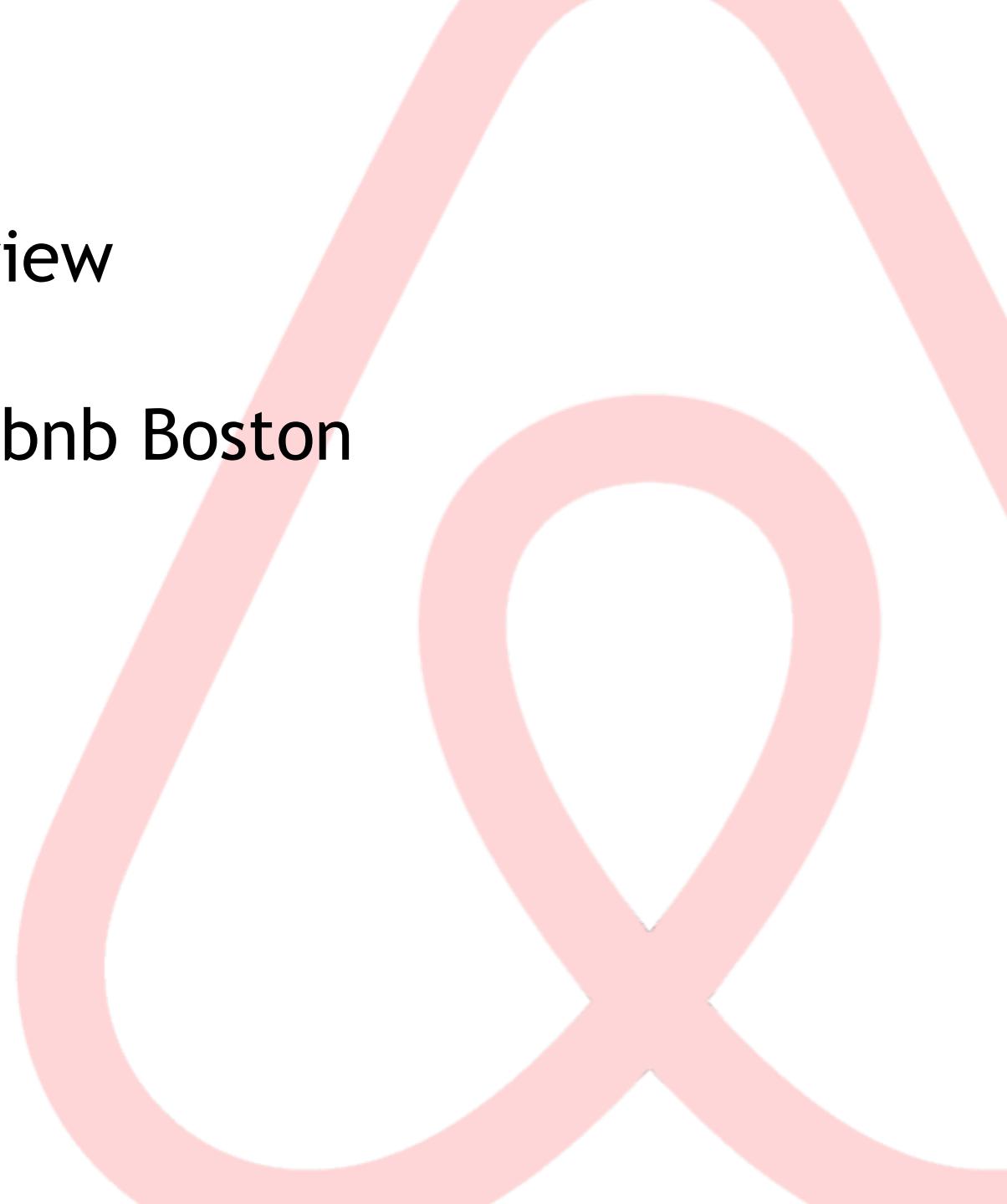
Glimpse on Airbnb Boston

3

Data Mining

4

Conclusion



Data & descriptive stats

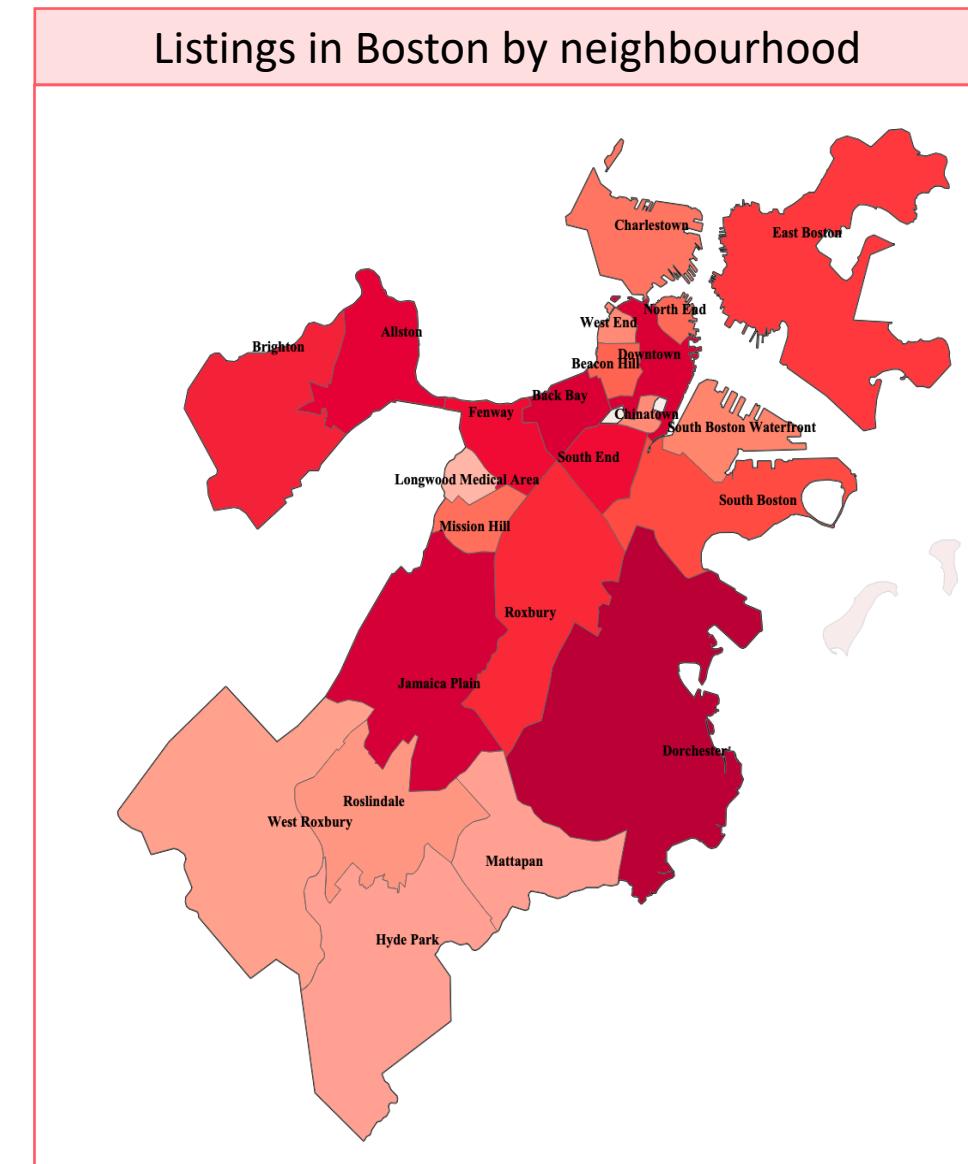
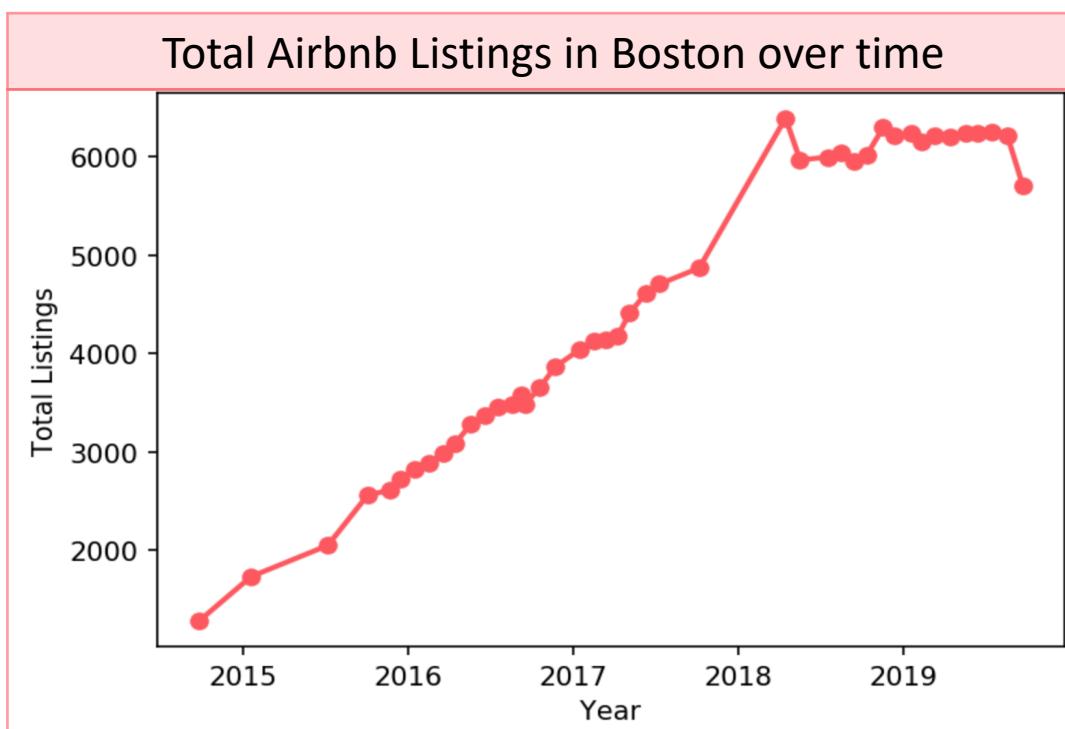


| Data source | | |
|------------------|---|-------------|
| File | Description | Time period |
| listings_details | Detailed Listings data for Boston | 2015 – 2019 |
| calendar | Detailed Calendar Data for listings in Boston | 2019 – 2020 |
| reviews_details | Detailed Review Data for listings in Boston | 2009 – 2019 |

| Descriptive Statistics | | | | | | |
|------------------------|------|--------|--------|-------|--------|------|
| | min | mean | median | max | s.d. | n |
| price | 0 | 196.2 | 149 | 10000 | 324.6 | 8806 |
| minimum nights | 1 | 5.78 | 2 | 1000 | 20.22 | 8806 |
| number of reviews | 0 | 26.99 | 4 | 650 | 53.83 | 8806 |
| reviews per month | 0.01 | 1.886 | 1.08 | 13.05 | 1.996 | 6311 |
| host listing count | 1 | 30.32 | 4 | 309 | 63.04 | 8806 |
| availability 365 | 0 | 138.33 | 89 | 365 | 128.84 | 8806 |

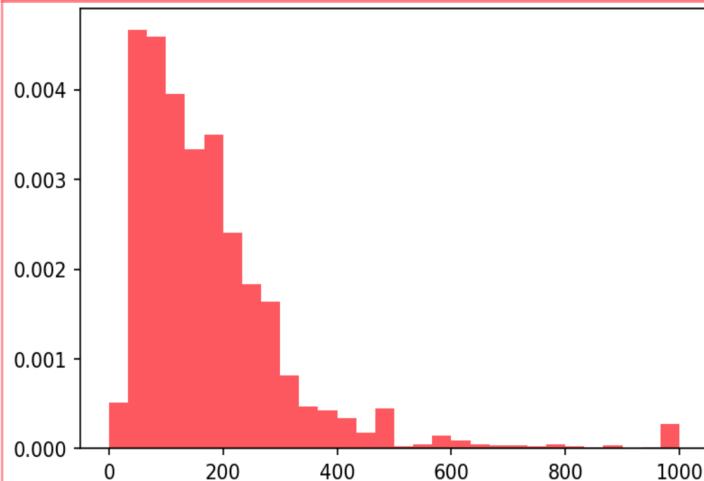
Listings

- Number of listings is increasing steadily from 2015 – 2018 and stays at a relatively stable level recently.
- Large areas generally have high listings (Jamaica Plain, Dorchester, East Boston, Roxbury, Allston). Listing density of Back Bay and downtown area is large

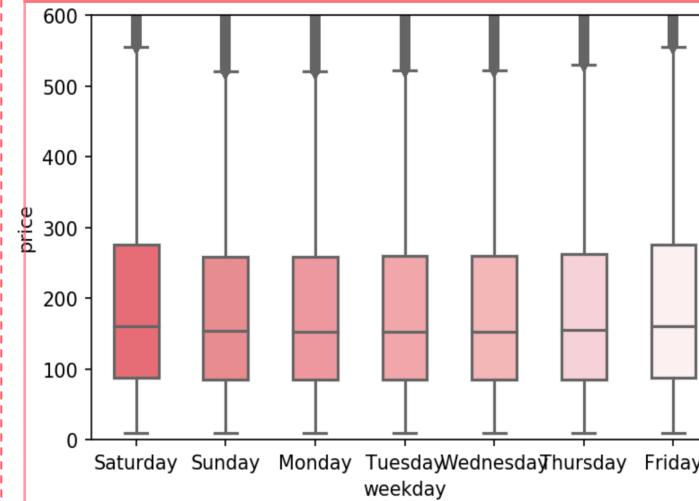


Price

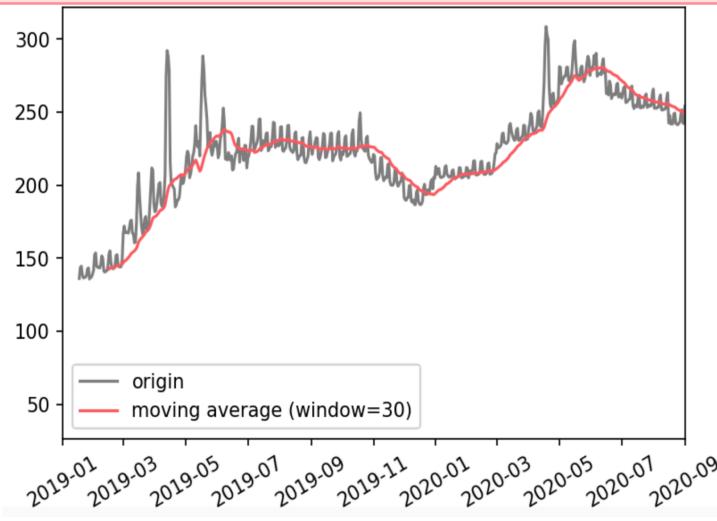
Price Hist -- right skewed



Higher price on Fri and Sat



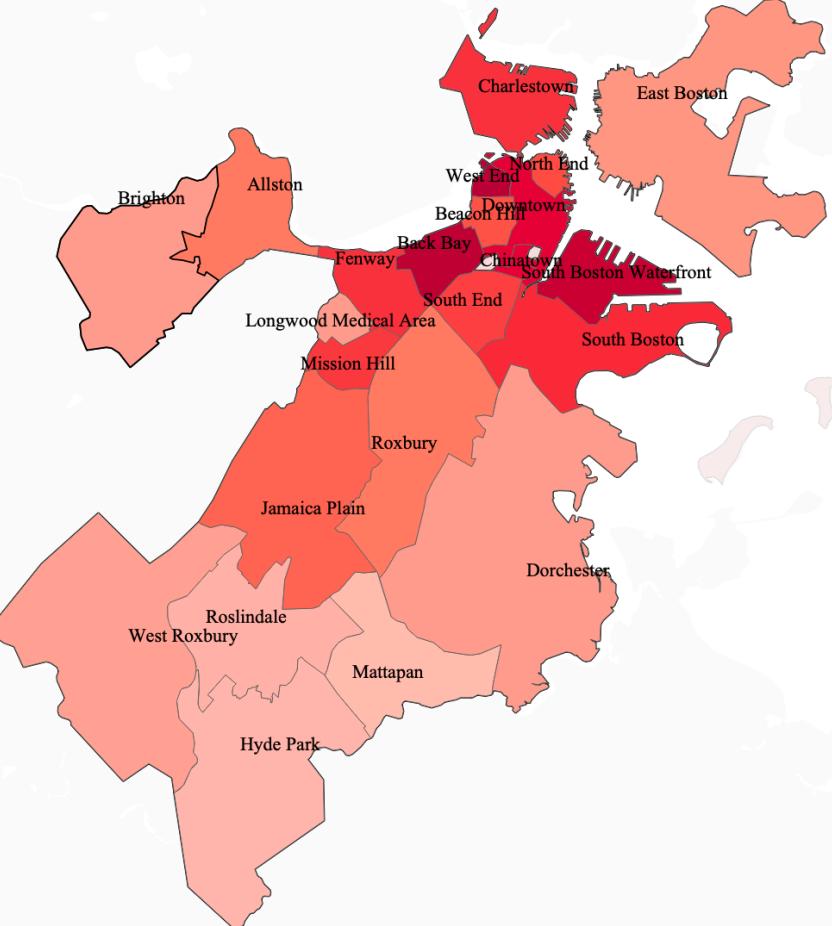
Seasonality! higher from Jun to Nov



Weekday Price

| Weekday | Price |
|---------|--------|
| Mon | 219.3 |
| Tue | 219.1 |
| Wed | 219.78 |
| Thu | 221.53 |
| Fri | 230.24 |
| Sat | 231.26 |
| Sun | 220.29 |

Inner harbor area has higher price (Back Bay, West End, Downtown, Chinatown, South Boston).



Minimum Occupancy rate

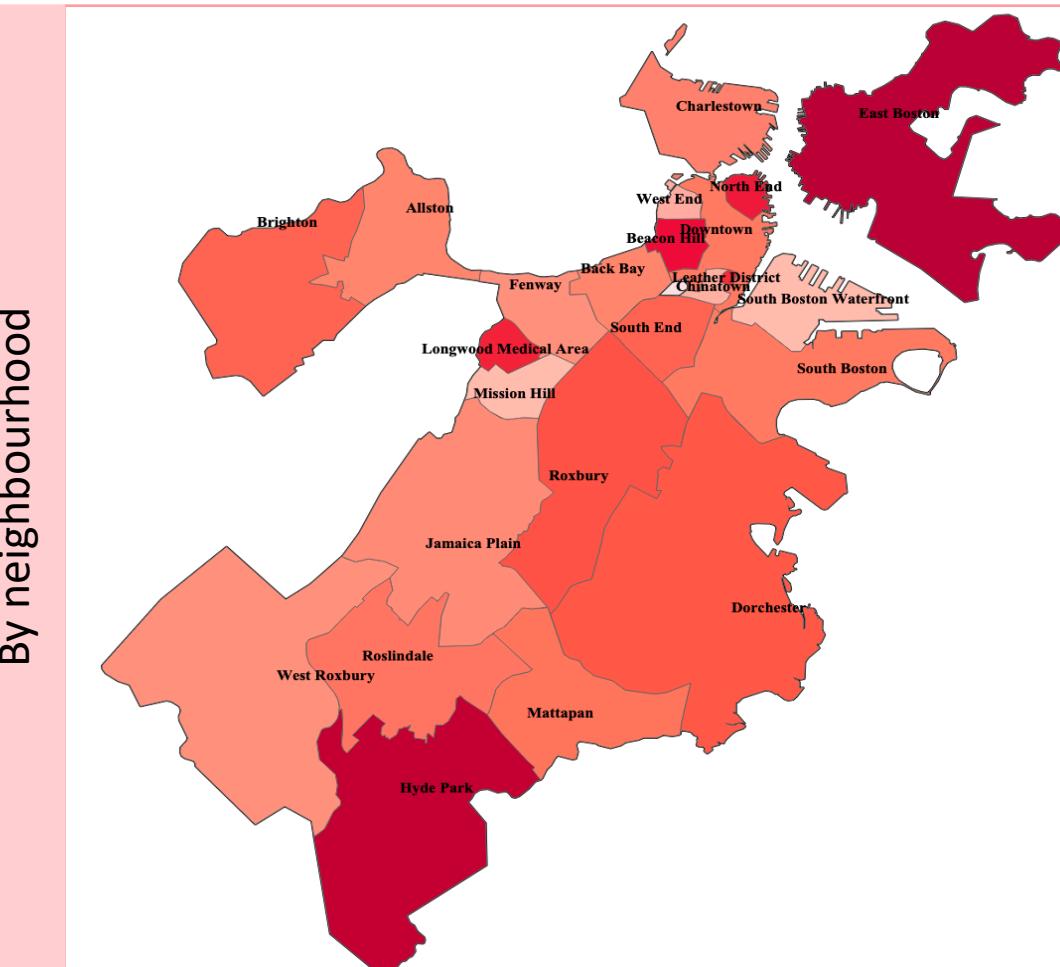
Min occupancy rate estimated by

$$\text{Reviews per month} \times \max(\text{avg stay in Boston}, \text{minimum night})$$

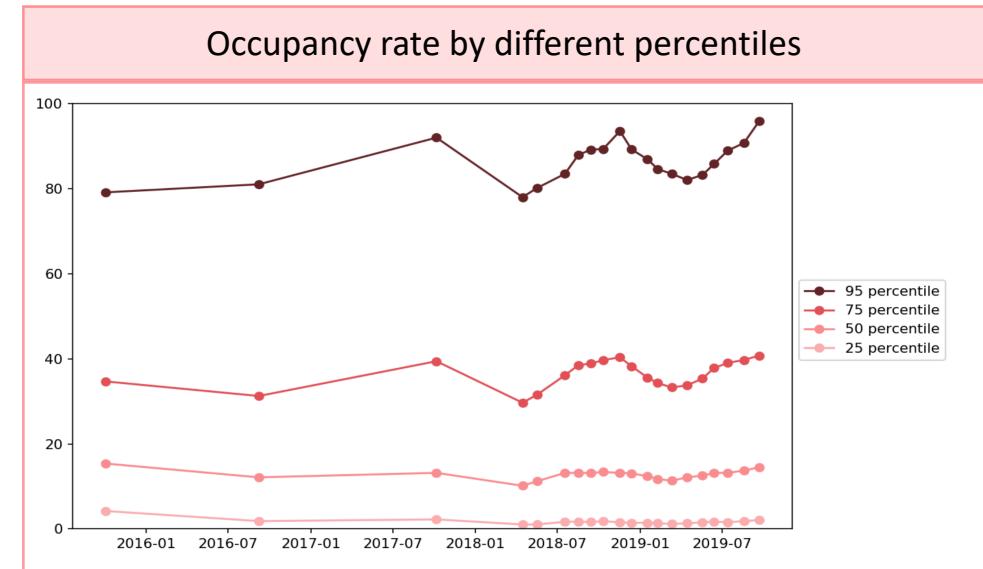
review rate

1. Avg stay: domestic (81.5%) stay 2.2 nights and overseas (18.5%) stay 8 nights
2. Review rate = 80%

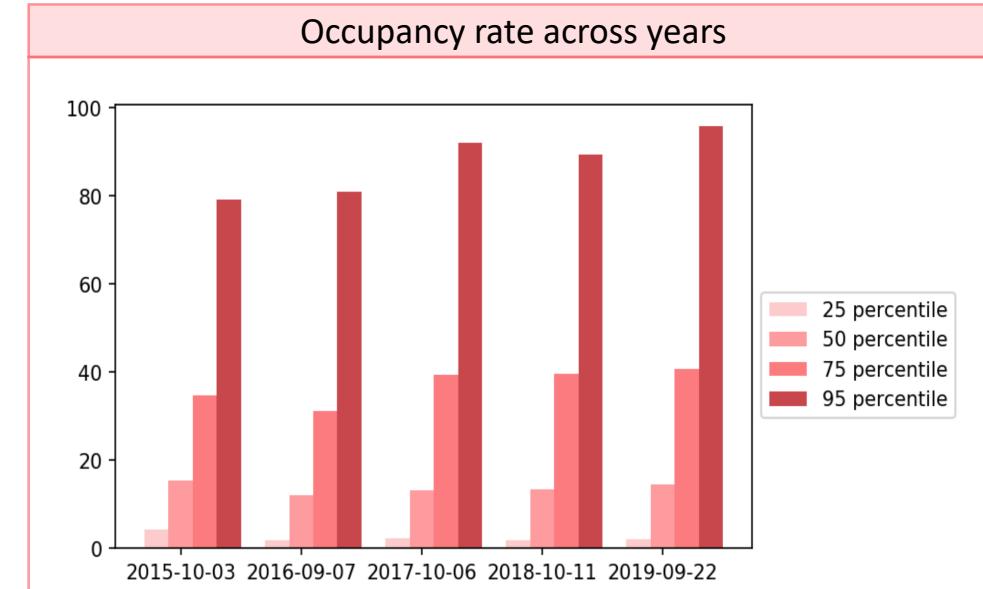
By neighbourhood



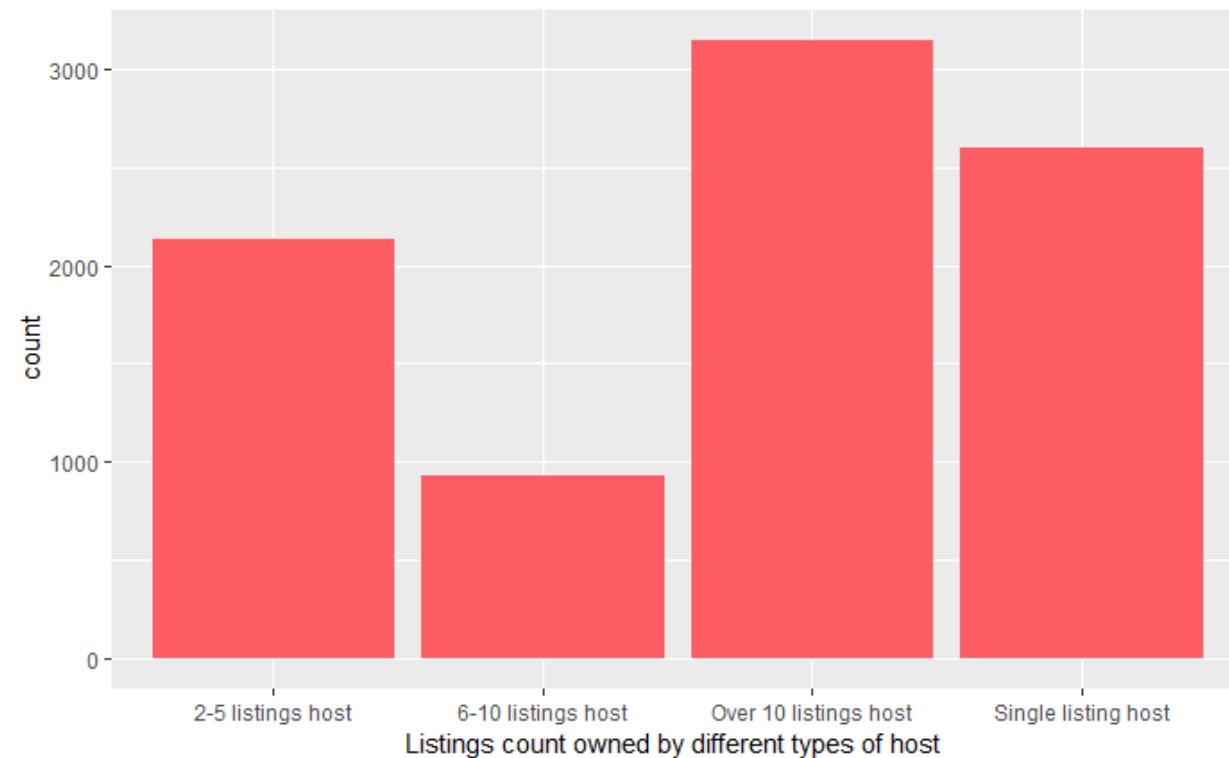
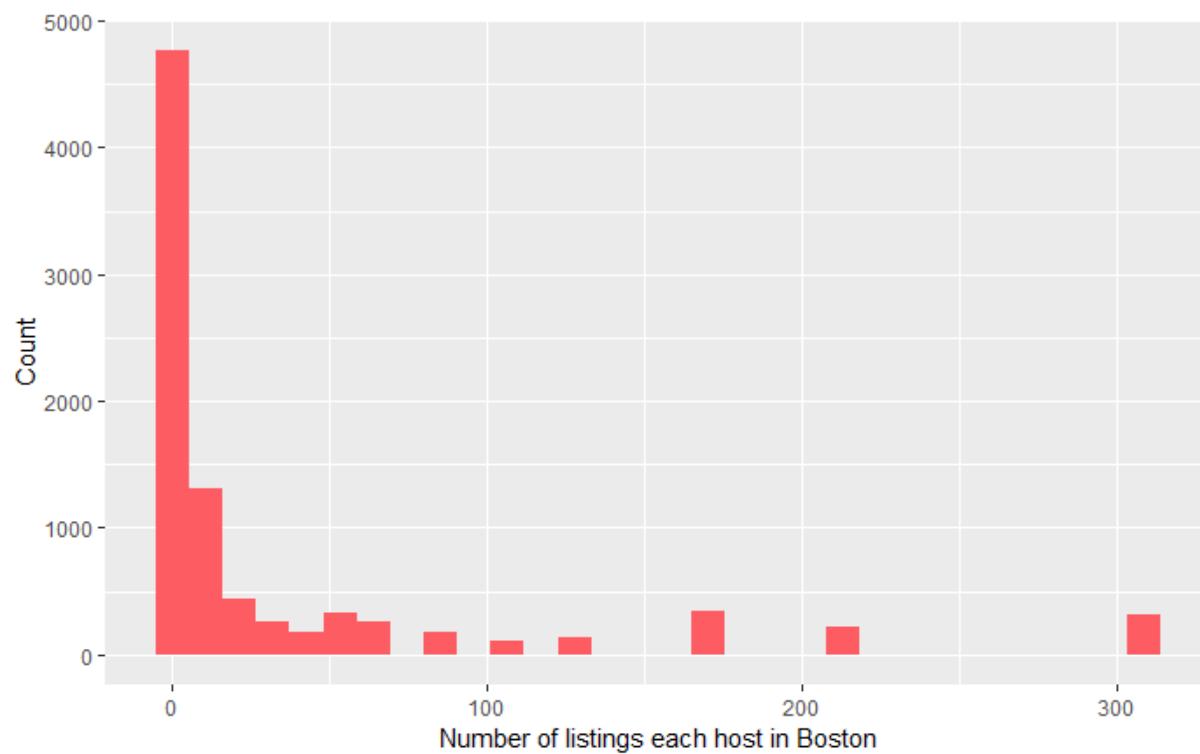
Top 5% listings are full all year while 75% listings occupancy rate are below 40%



OR didn't see significant increase across years.



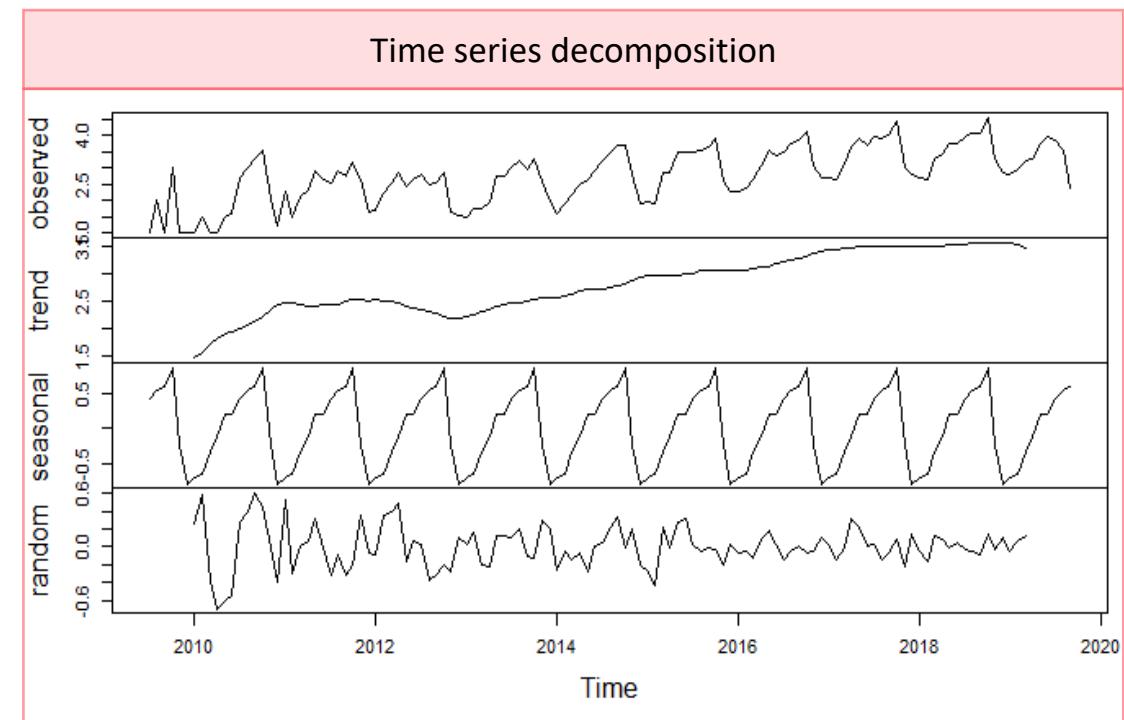
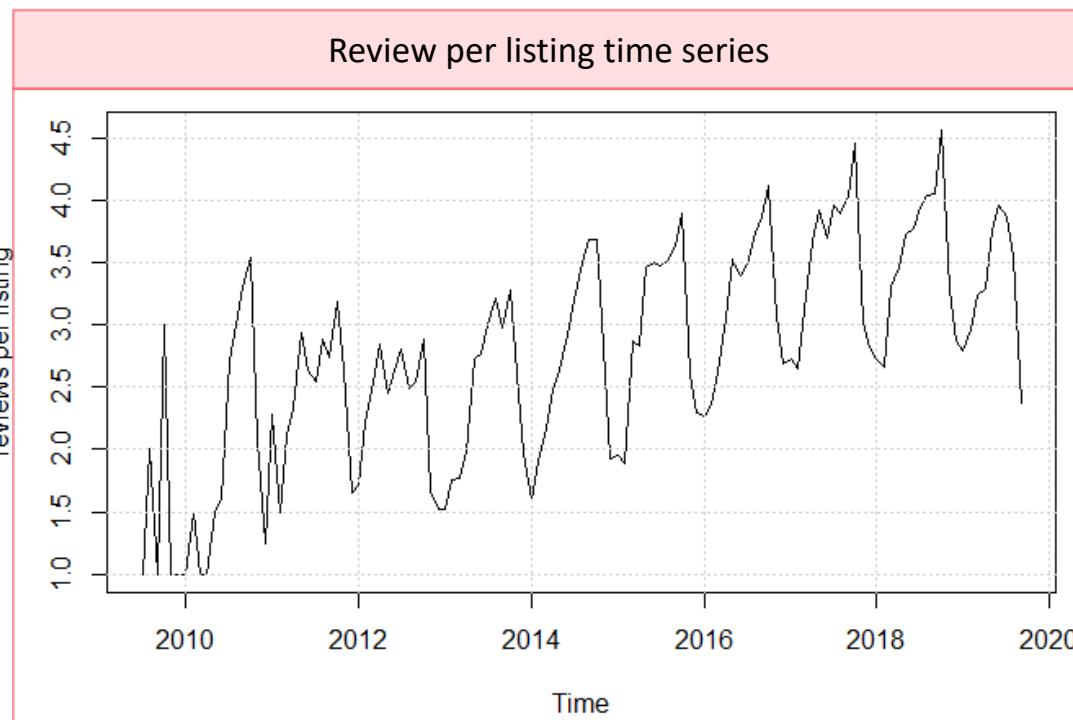
- Most listings are owned by hosts with multiple listings.
- Implying the Airbnb market in Boston are mostly controlled by professional landlords.



Reviews



- There is a clear seasonal pattern.
- The rising trend is stagnating.



1

Industry Overview

2

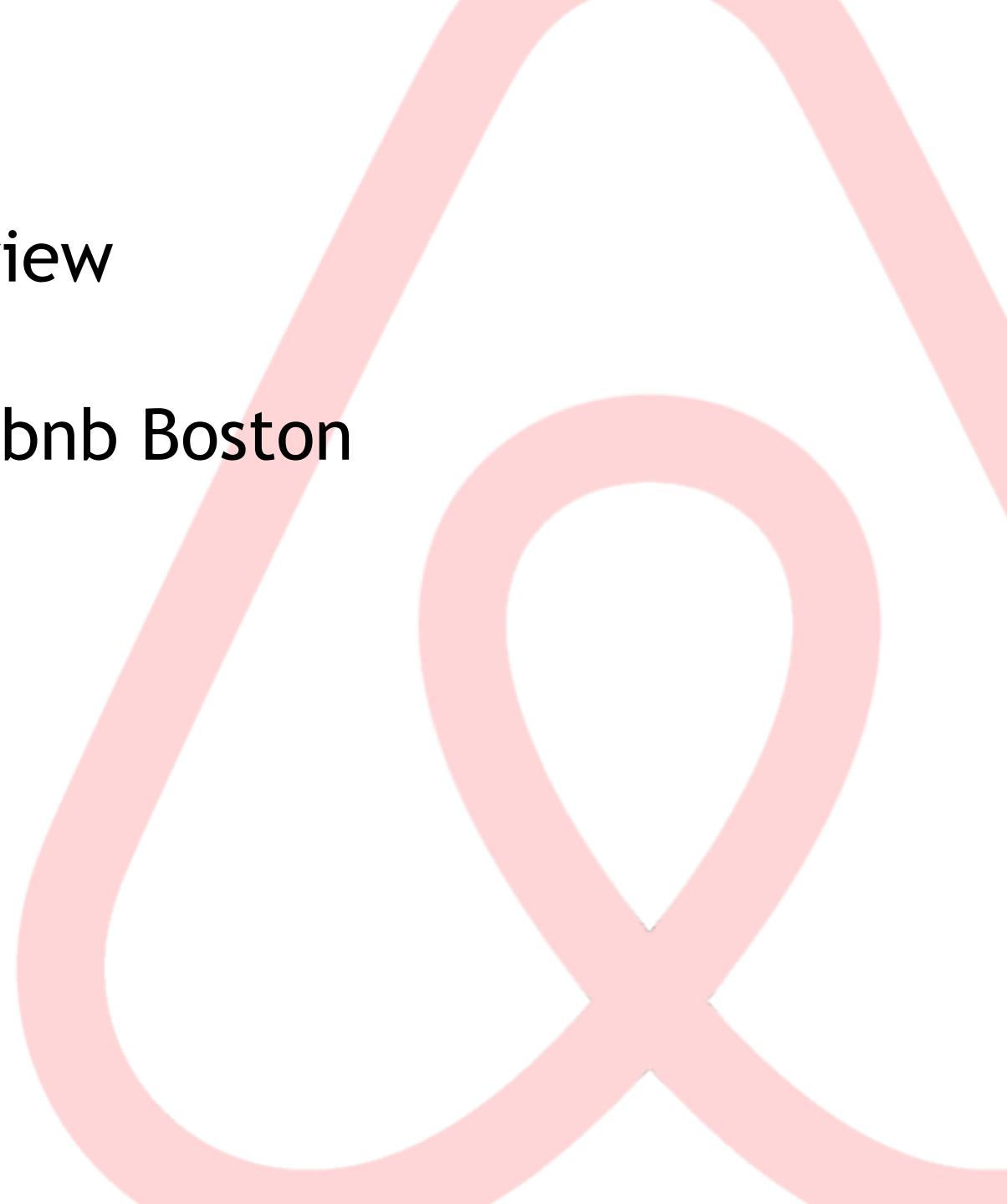
Glimpse on Airbnb Boston

3

Data Mining

4

Conclusion



Linear Regression



- Use stepwise regression (forward) to select the best model.
 - 38 vars, 0.73 adjusted R²

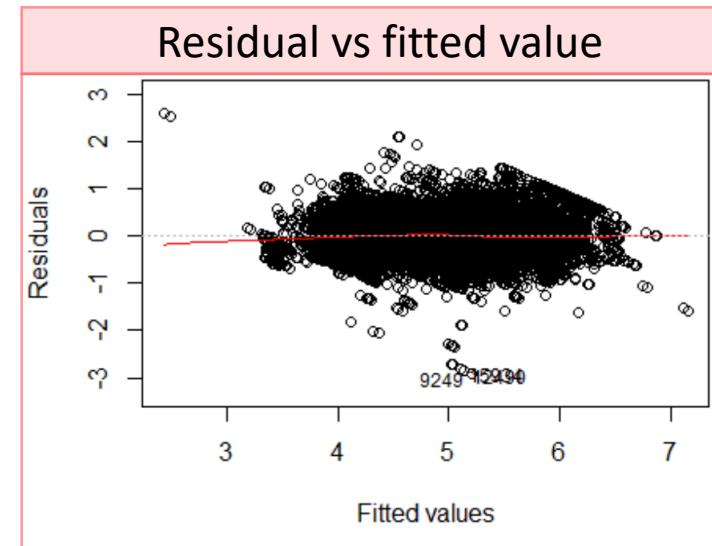
Coefficients:

Model diagnose

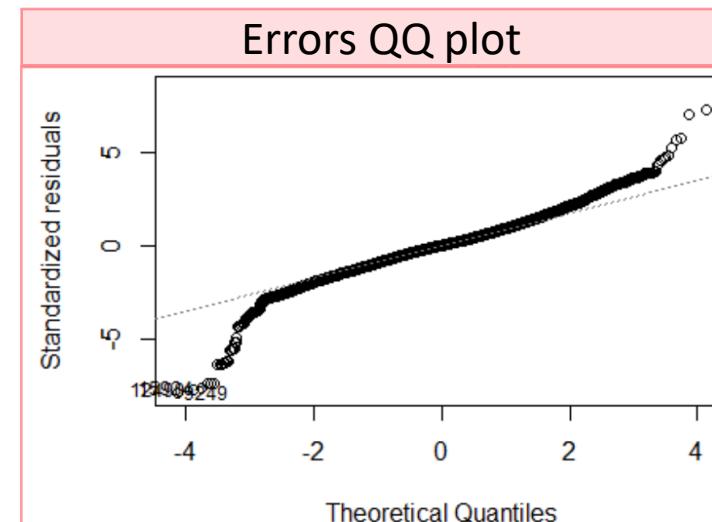
No severe collinearity

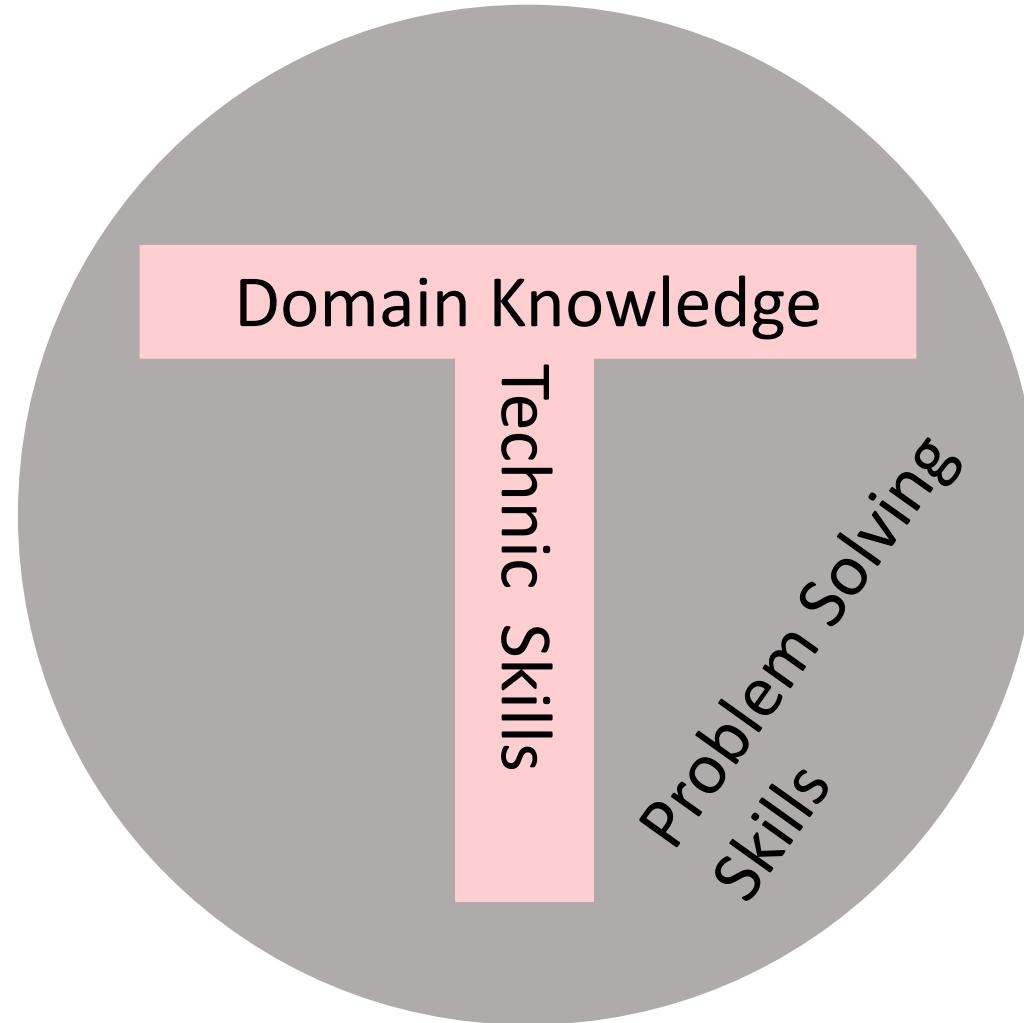
| Vif between vars | |
|----------------------------------|------|
| Features | GVIF |
| host_response_rate | 2.96 |
| host_listing_count | 1.94 |
| accommodates | 7.15 |
| bathrooms | 1.57 |
| bedrooms | 3.58 |
| beds | 5.34 |
| security_deposit | 1.19 |
| guests_included | 1.85 |
| extra_people | 1.24 |
| minimum_nights | 1.81 |
| minimum_minimum_nights | 1.69 |
| availability_90 | 1.6 |
| availability_365 | 1.7 |
| number_of_reviews | 2.64 |
| number_of_reviews_ltm | 5.41 |
| reviews_scores_rating | 3.62 |
| review_scores_cleanliness | 1.93 |
| review_scores_checkin | 1.79 |
| review_scores_communication | 2.26 |
| review_scores_location | 1.67 |
| review_score_value | 2.59 |
| reviews_per_month | 4.08 |
| host_is_superhost | 1.46 |
| host_has_profile_pic | 1.02 |
| instant_bookable | 1.41 |
| require_guest_profile_picture | 5.03 |
| require_guest_phone_verification | 3.96 |
| host_response_time | 3.94 |
| neighbourhood | 7.09 |
| property_type | 3.54 |
| room_type | 2.91 |
| bed_type | 1.14 |

Errors distributed randomly align the y axis

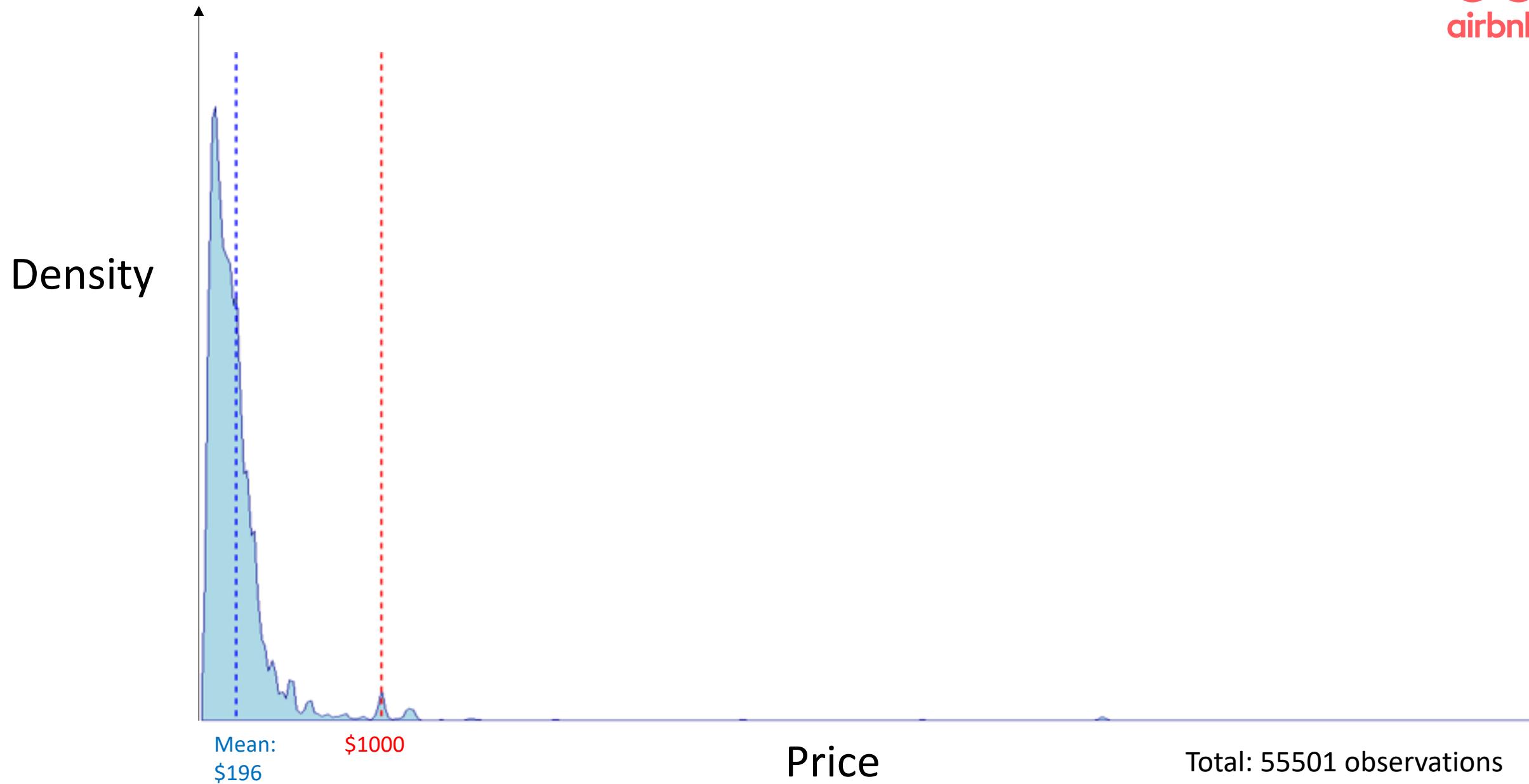


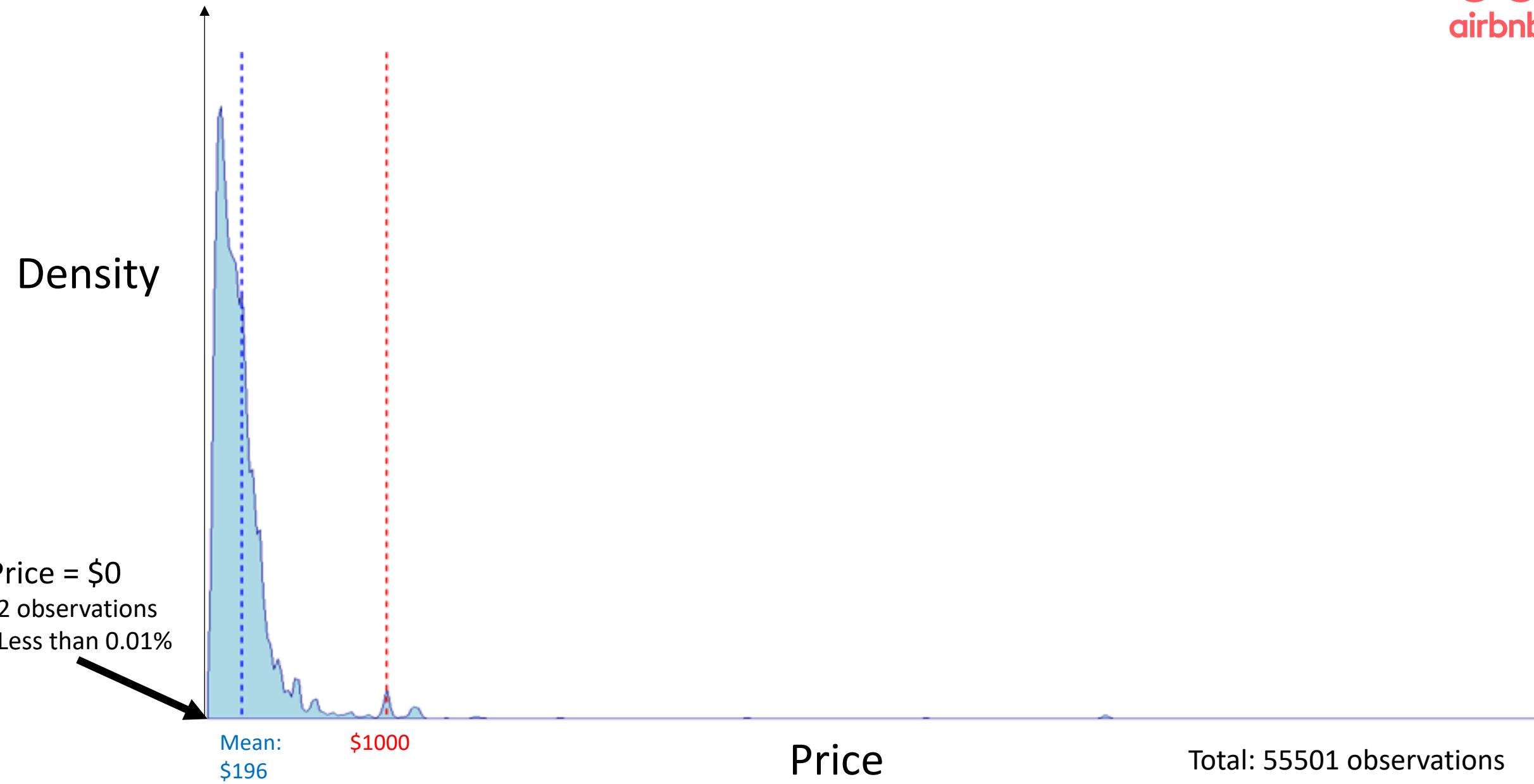
Most of errors follows normal distribution



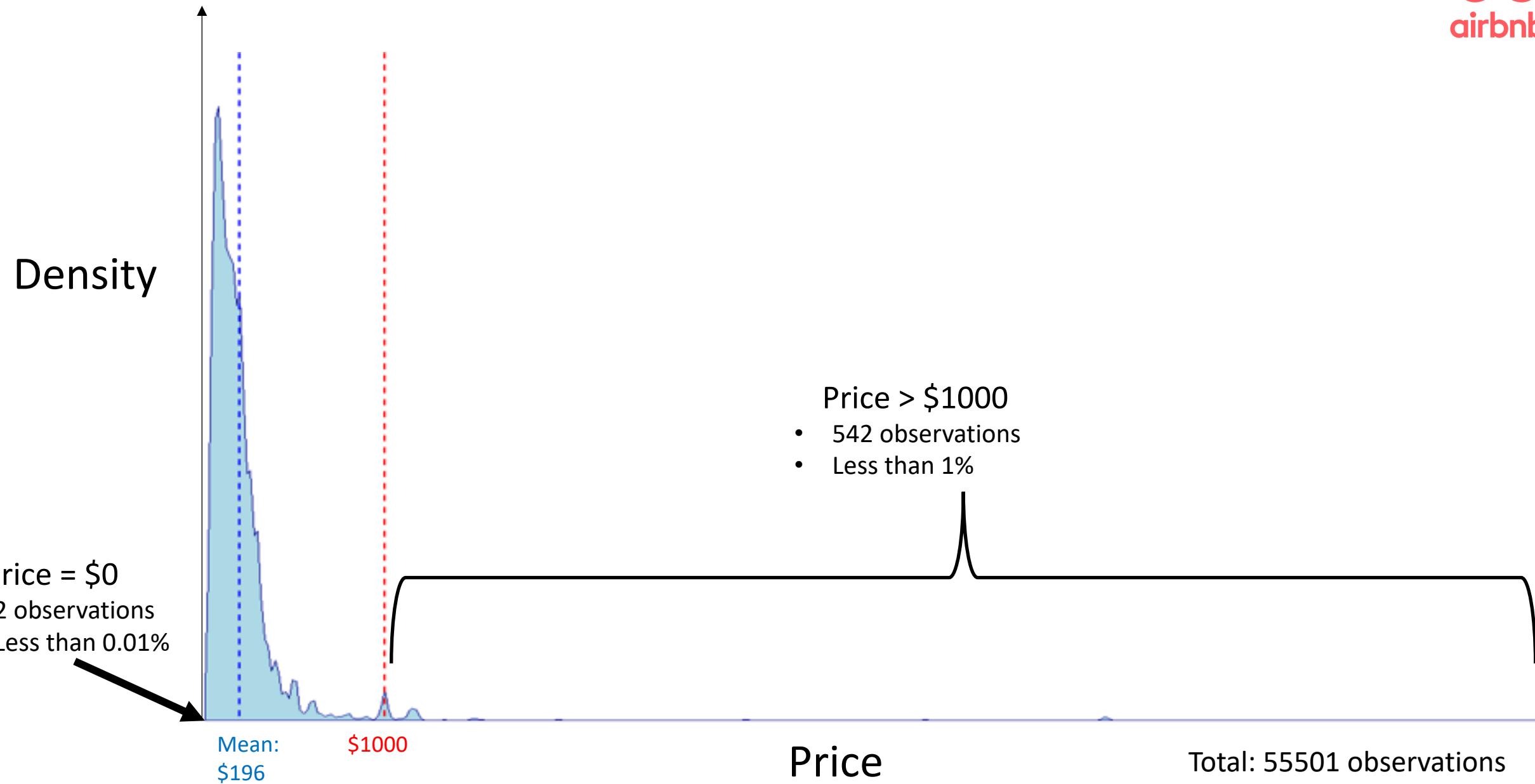


Deal with outlier



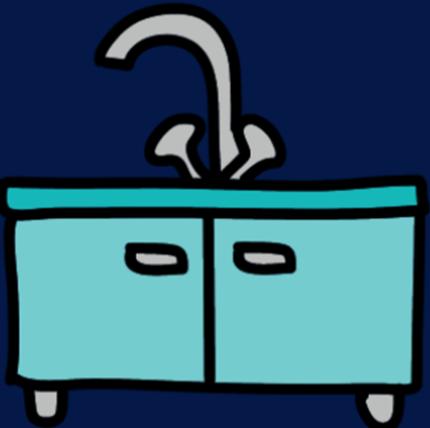


Deal with outlier



- Regression assumptions no longer matter
- Kitchen sink approach

Everything but
the Kitchen Sink



Natural Language Processing – word cloud

Key Words:

- location/neighborhood
- cleanliness
- room type
- beds/bedroom



Word cloud of reviews



Feature selection



Feature Pool

Numeric Features

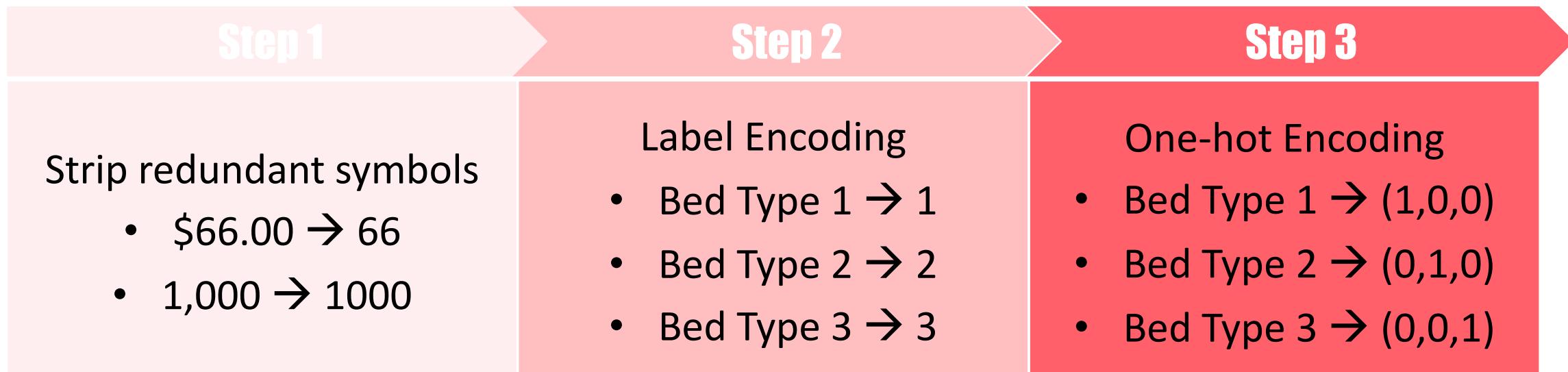
host_response_rate, host_listings_count, accommodates, bathrooms, bedrooms, beds, security_deposit, guests_included, extra_people, minimum_nights, maximum_nights, minimum_minimum_nights, maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights, minimum_nights_avg_ntm, maximum_nights_avg_ntm, availability_30, availability_60, availability_90, availability_365, number_of_reviews, number_of_reviews_ltm, review_scores_rating, review_scores_accuracy, review_scores_cleanliness, review_scores_checkin, review_scores_communication, review_scores_location, review_scores_value, calculated_host_listings_count, calculated_host_listings_count_entire_homes, calculated_host_listings_count_private_rooms, calculated_host_listings_count_shared_rooms, reviews_per_month

Binary Features

host_is_superhost, host_has_profile_pic, host_identity_verified, is_location_exact, requires_license, instant_bookable, is_business_travel_ready, require_guest_profile_picture, require_guest_phone_verification

Categorical Features

host_response_time, neighbourhood, property_type, room_type, bed_type, cancellation_policy



Model Evaluation



Mean (price) = 196
s.d. (price) = 144

| Learning Method | Running Time | RMSE (Train) | RMSE (Test) |
|---|--------------|--------------|-------------|
| Linear Regression | 1s | 86.07 | 88.09 |
| Ridge Regression | 1s | 86.16 | 87.73 |
| Lasso Regression | 1s | 87.44 | 86.65 |
| K-Neighbor Regression | 30s | 71 | 64.61 |
| Random Forest Regression with Grid Search | 5min | 77.07 | 78.82 |
| Neural Nets Regression | 2min | 63.7 | 73.83 |

Model Evaluation



Mean (price) = 196
 s.d. (price) = 144

| Learning Method | Running Time | RMSE (Train) | RMSE (Test) |
|---|--------------|--------------|-------------|
| Linear Regression | 1s | 86.07 | 88.09 |
| Ridge Regression | 1s | 86.16 | 87.73 |
| Lasso Regression | 1s | 87.44 | 86.65 |
| K-Neighbor Regression | 30s | 71 | 64.61 |
| Random Forest Regression with Grid Search | 5min | 77.07 | 78.82 |
| Neural Nets Regression | 2min | 63.7 | 73.83 |
| XGBoost Regression with Grid Search | 814min 😞 | 0.72 | 27.96 |
| LightGBM Regression with Grid Search | 10min | 10.79 | 33.79 |
| CatBoost Regression with Grid Search | 5min | 17.05 | 33.11 |

Feature importance



| LightGBM | |
|------------------------------|------------|
| Features | Importance |
| reviews_per_month | 7418 |
| availability_365 | 6582 |
| number_of_reviews | 6156 |
| availability_90 | 5286 |
| number_of_reviews_ltm | 4980 |
| host_listint_count | 4486 |
| availability_60 | 4141 |
| reviews_scores_rating | 3503 |
| availability_30 | 3391 |
| cal_host_listings_count | 2719 |
| accommodates | 2640 |
| cal_host_listings_entirehome | 2514 |
| extra_people | 2507 |
| security_deposit | 2127 |
| maximum_nights | 2108 |
| beds | 1787 |
| host_response_rate | 1758 |
| guests_included | 1651 |
| minimum_maximum_nights | 1565 |

| CatBoost | |
|------------------------------|------------|
| Features | Importance |
| host_listing_count | 14.02 |
| room_type | 9.57 |
| bathrooms | 9.12 |
| neighbourhood | 8.31 |
| bedrooms | 7.95 |
| security_deposit | 6.04 |
| extra_people | 4.78 |
| accommodates | 4.15 |
| cal_host_listings_count | 4.05 |
| reviews_per_month | 3.31 |
| cal_host_listings_entirehome | 3.22 |
| guests_included | 2.72 |
| availability_365 | 2.2 |
| property_type | 1.96 |
| minimum_nights_avg_ntm | 1.88 |
| number_of_reviews | 1.65 |
| beds | 1.27 |
| reviews_scores_rating | 1.27 |
| cancellation_policy | 1.19 |
| number_of_reviews_ltm | 0.9 |

Feature importance



| LightGBM | |
|------------------------------|------------|
| Features | Importance |
| reviews_per_month | 7418 |
| availability_365 | 6582 |
| number_of_reviews | 6156 |
| availability_90 | 5286 |
| number_of_reviews_ltm | 4980 |
| host_listint_count | 4486 |
| availability_60 | 4141 |
| reviews_scores_rating | 3503 |
| availability_30 | 3391 |
| cal_host_listings_count | 2719 |
| accommodates | 2640 |
| cal_host_listings_entirehome | 2514 |
| extra_people | 2507 |
| security_deposit | 2127 |
| maximum_nights | 2108 |
| beds | 1787 |
| host_response_rate | 1758 |
| guests_included | 1651 |
| minimum_maximum_nights | 1565 |

| CatBoost | |
|------------------------------|------------|
| Features | Importance |
| host_listing_count | 14.02 |
| room_type | 9.57 |
| bathrooms | 9.12 |
| neighbourhood | 8.31 |
| bedrooms | 7.95 |
| security_deposit | 6.04 |
| extra_people | 4.78 |
| accommodates | 4.15 |
| cal_host_listings_count | 4.05 |
| reviews_per_month | 3.31 |
| cal_host_listings_entirehome | 3.22 |
| guests_included | 2.72 |
| availability_365 | 2.2 |
| property_type | 1.96 |
| minimum_nights_avg_ntm | 1.88 |
| number_of_reviews | 1.65 |
| beds | 1.27 |
| reviews_scores_rating | 1.27 |
| cancellation_policy | 1.19 |
| number_of_reviews_ltm | 0.9 |

Rank the listing (some extra work)



- Use CatBoost algorithm to predict review scores to rank unreviewed listings, RMSE 4.5
- It's also useful for the Airbnb to display searching results.

| Id | Predict score |
|----------|---------------|
| 21470225 | 103.94 |
| 20088992 | 103.74 |
| 21711766 | 102.74 |
| 21711766 | 102.63 |
| 21470225 | 102.27 |
| 21010620 | 102.09 |
| 21423860 | 102.01 |
| 20088992 | 101.93 |
| 20088992 | 101.91 |
| 23287239 | 101.86 |
| 24323074 | 101.81 |
| 21470225 | 101.75 |
| 23287239 | 101.65 |
| 23287239 | 101.65 |
| 24473093 | 101.62 |
| 11470818 | 101.53 |
| 14924831 | 101.52 |
| 15528667 | 101.52 |
| 20088992 | 101.49 |



Lux 1BR Apt in Bulfinch Triangle w/gym & WiFi

Boston

3 guests | 1 bedroom | 2 beds | 1 bath

Mike

\$235 per night

Dates
Check-in → Checkout

[View Photos](#)



Luxury 2BR apt located minutes from Fenway Park!

Boston

5 guests | 2 bedrooms | 3 beds | 2 baths

Kara

\$258 per night
★4.80 (5 reviews)

Dates
Check-in → Checkout

[View Photos](#)

1

Industry Overview

2

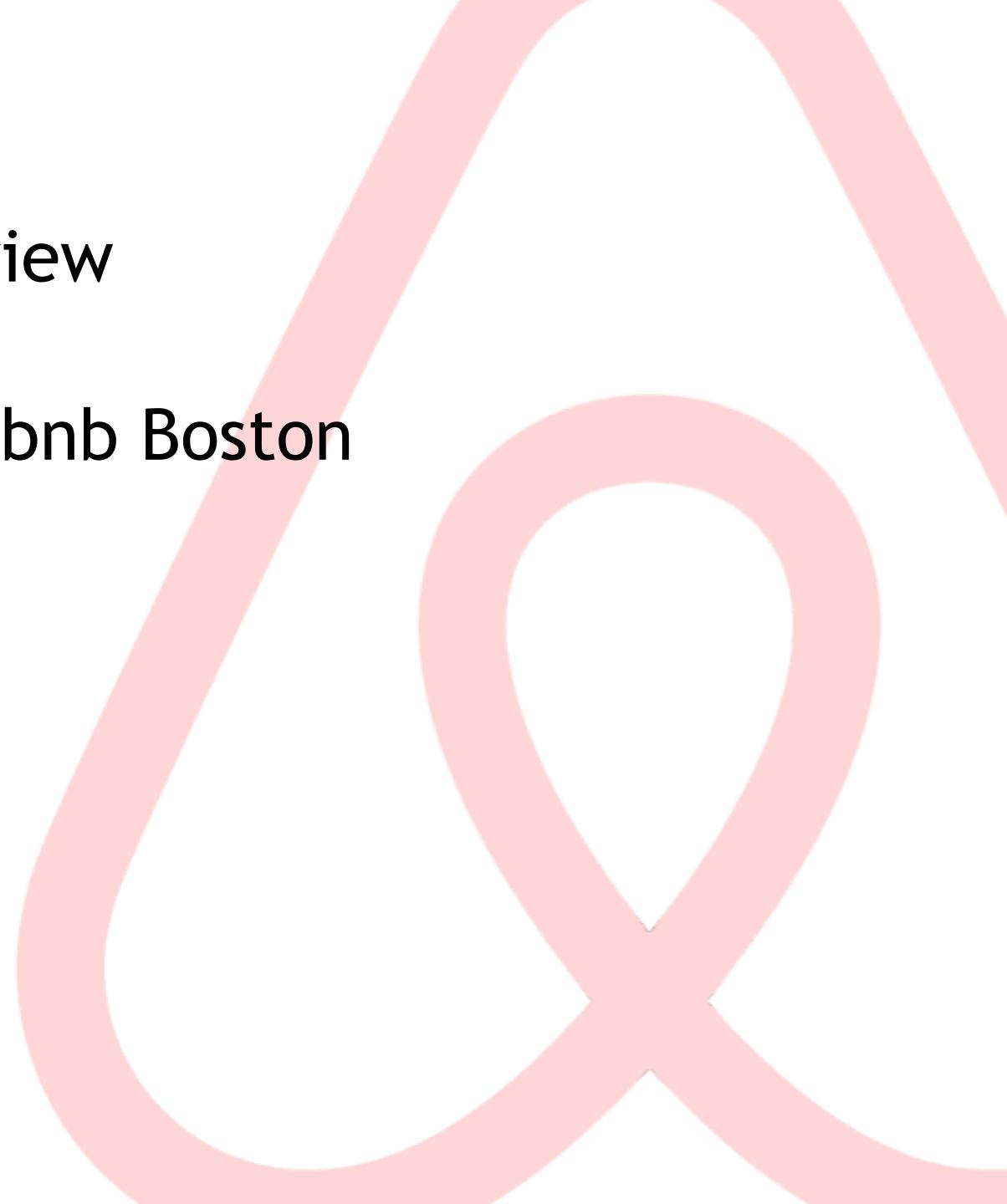
Glimpse on Airbnb Boston

3

Data Mining

4

Conclusion



Suggestions



For
host

- Take extra attention on the neighborhood of their rooms
- Make sure the rooms are clean and well prepared.
- Try to get along well with renters

For
Airbnb

- Use a “price calculator” to give real-time price recommendations to hosts
- Rank the unreviewed listings.
- Adopt policies that encourages customers to write reviews and give review scores.
- Provide a playbook about how to prepare a room to hosts.
- Find solutions of legal problems.



Thanks for listening!