



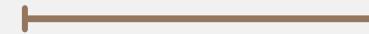
# Inside Airbnb Capstone



Analysis by: Kyle Reed



# Introductions



**Kyle Reed**

Data Analyst



**Rusty**

Potential Host



[www.linkedin.com/in/reed-kyle](https://www.linkedin.com/in/reed-kyle)



+850-428-6089

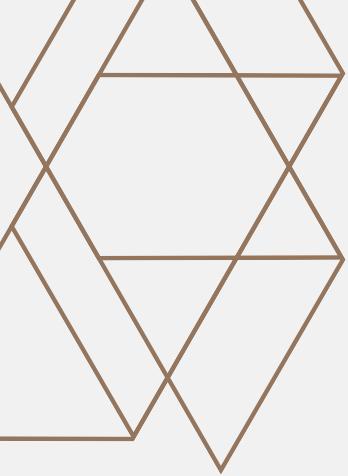


Denver, CO

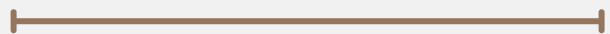


# Agenda

- 
- Problem Statement
  - Data
  - Outliers
  - Null Values
  - Modeling
  - Insights
  - Appendix

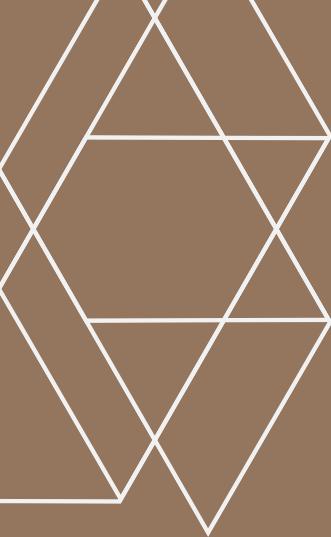


# Problem Statement



What features affect the listing  
price of an Airbnb?





# Data



**Inside Airbnb**  
Adding data to the debate





# Data Summary

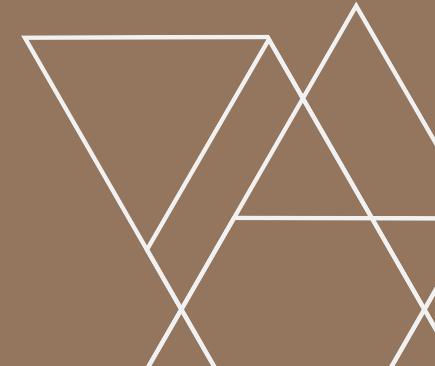
- Paris & London AirBnb's
- One listing per row

Before cleaning

- 477,605 rows, 75 columns
- 33 categorical, 42 numeric

After cleaning

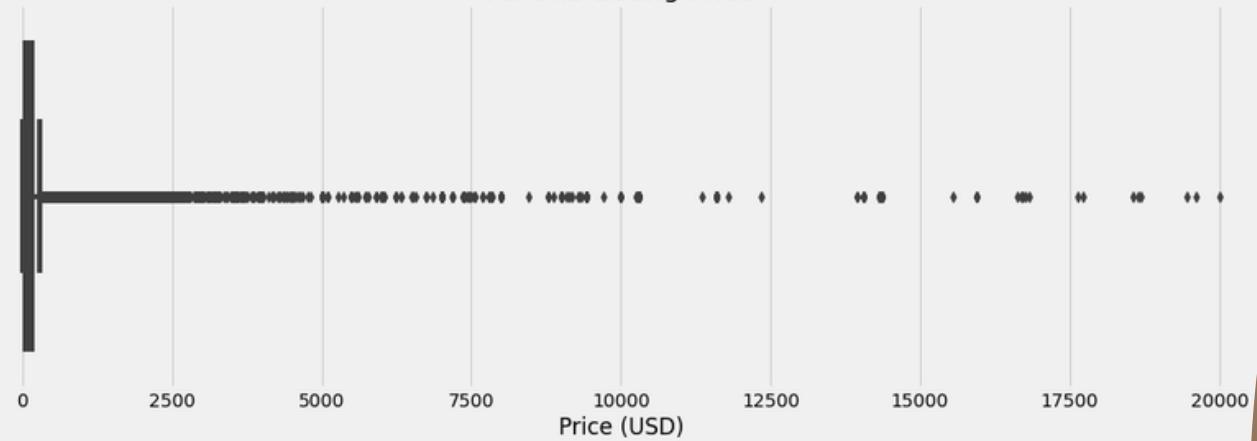
- 354,100 rows, 35 columns
- 2 categorical, 33 numeric





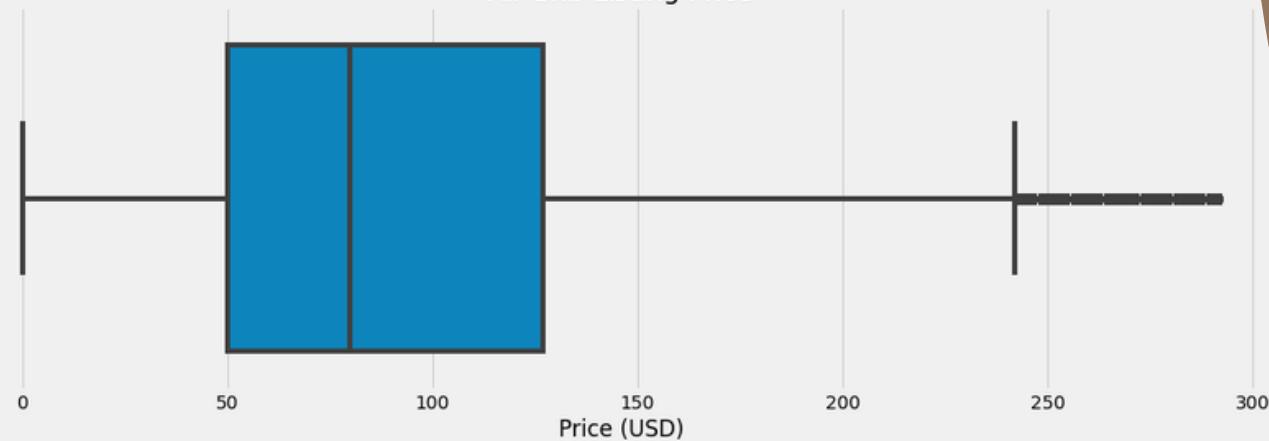
Before

Air Bnb Listing Price



After

Air Bnb Listing Price



# Outliers

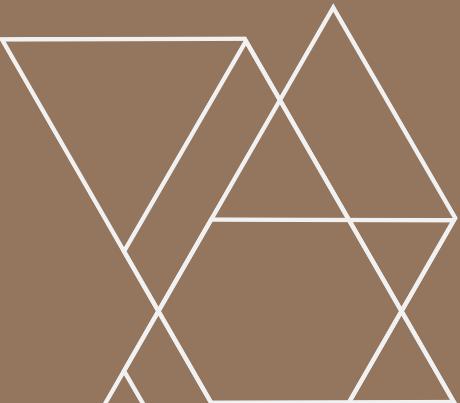
- Box-plot of features
  - Interquartile Range
    - Spread
    - Lower/Upper limit
- Target variable, price
  - Before: \$20,000 max
    - Mean ~ \$146, STD ~ \$330
  - After: ~ \$292 max
    - Mean ~ \$97, STD ~ \$59



# Null Values

→ → →

- 10% pricing data - 49,429 rows
- Paris
- Minimum maximum nights
  - 1125 - three years
  - Lease in France





Initial Results

	<b>train_cross_val</b>	<b>train_score</b>	<b>test_score</b>
<b>regr</b>	0.777921	0.971439	0.804992
<b>xgb</b>	0.651578	0.682195	0.654923
<b>tree</b>	0.571264	0.999028	0.617717
<b>ridge</b>	0.501256	0.501696	0.499925
<b>linear_regression</b>	0.501249	0.501691	0.499919
<b>lasso</b>	0.499651	0.500056	0.498817
<b>elastic</b>	0.376468	0.376546	0.374950

Top Performers

	<b>train_cross_val</b>	<b>train_score</b>	<b>test_score</b>
<b>xgb</b>	0.800774	0.995096	0.825650
<b>regr</b>	0.621917	0.669913	0.625128

# Modeling

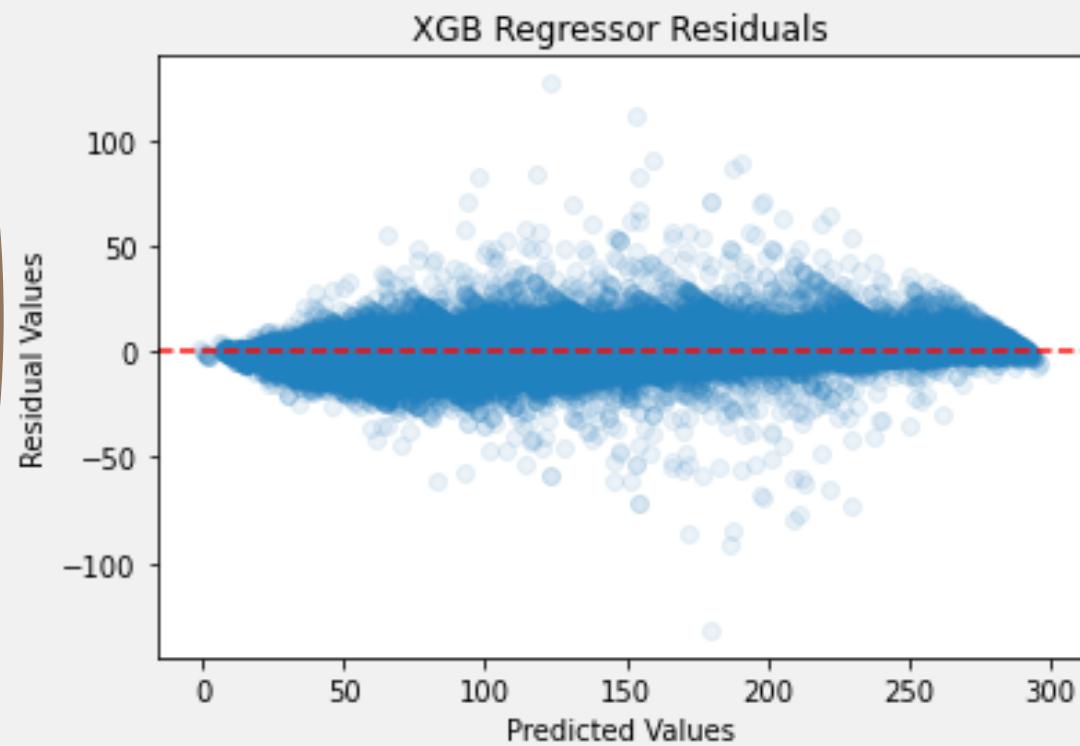


- Model selection
  - Feature importance
  - Performance
  - Work well with new data



# Error in Model

- Residual - error metric
  - Actual value - predicted value
- Mean Absolute Error of XGB
  - Predict within \$2.24 on average.



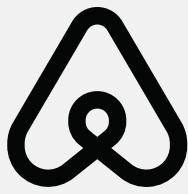


# Insights

---

- Room type
- Accommodate
- Neighborhood
- Bathroom





# Thank You



Questions?



# Appendix

## - Feature Correlation to Price

Before	price
accommodates	0.248932
beds	0.198975
baths	0.180582
calculated_host_listings_count	0.143309
calculated_host_listings_count_entire_homes	0.135159
availability_365	0.094506
calculated_host_listings_count_private_rooms	0.094431
availability_30	0.093767
availability_60	0.091845
availability_90	0.090988

After	price
accommodates	0.499798
beds	0.402875
availability_365	0.217887
availability_90	0.175311
availability_60	0.163915
baths	0.158293
availability_30	0.149311
host_id	0.116724
calculated_host_listings_count_entire_homes	0.115908
review_scores_location	0.091669

# Appendix

## - Feature Importance

XGBRegressor Results	<b>feature_importance</b>
<b>room type private room</b>	0.778062
<b>room type hotel room</b>	0.027723
<b>room type shared room</b>	0.021143
<b>accommodates</b>	0.011822
<b>calculated host listings count shared rooms</b>	0.010518
<b>neighborhood westminster</b>	0.007046
<b>neighborhood kensington and chelsea</b>	0.006991
<b>baths</b>	0.006382
<b>neighborhood city of london</b>	0.006371
<b>neighborhood croydon</b>	0.005017
<b>neighborhood richmond upon thames</b>	0.004275
<b>calculated host listings count entire homes</b>	0.003658
<b>calculated host listings count</b>	0.003276
<b>neighborhood luxembourg</b>	0.003012
<b>calculated host listings count private rooms</b>	0.003005

RandomForestRegressor Results	<b>feature_importance</b>
<b>room type private room</b>	0.318531
<b>accommodates</b>	0.157314
<b>availability 365</b>	0.092876
<b>latitude</b>	0.080296
<b>longitude</b>	0.052297
<b>calculated host listings count</b>	0.040882
<b>baths</b>	0.035855
<b>availability 30</b>	0.031471
<b>host id</b>	0.020934
<b>review scores location</b>	0.018161
<b>calculated host listings count entire homes</b>	0.012966
<b>calculated host listings count private rooms</b>	0.012465
<b>neighborhood westminster</b>	0.011414
<b>minimum minimum nights</b>	0.010979
<b>minimum nights avg ntm</b>	0.008281