# AIRBNB - NEW YORK PRICING ANALYSIS



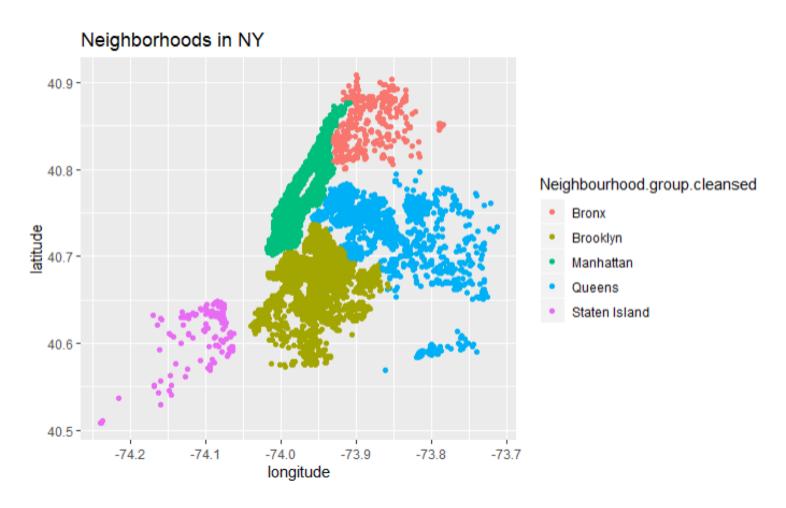
**RMSE** 

Aakanksha Baid, Isha Malte, Irene Sung

## **BUSINESS OBJECTIVE**

## **Predict Listing Price to:**

- Help Airbnb decide commission to charge on each property
- Help hosts in better understanding of price for their property based on multiple parameters
- Compete with nearby hotels by offering attractive price



## **ABOUT DATASET:**

# Airbnb NY city Listing data in 2017 Original dataset:

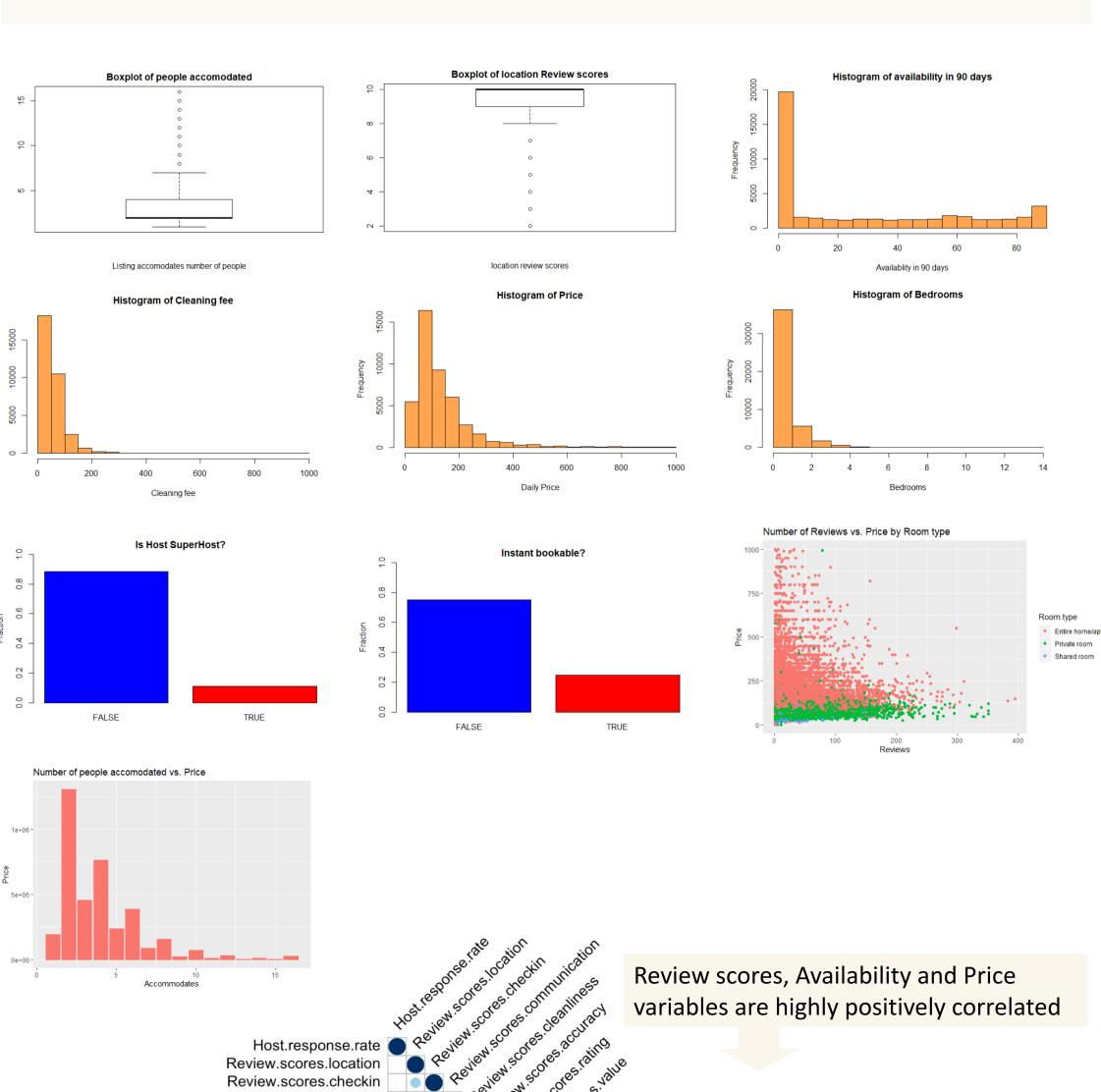
- > 44,308 observations
- > 94 variables

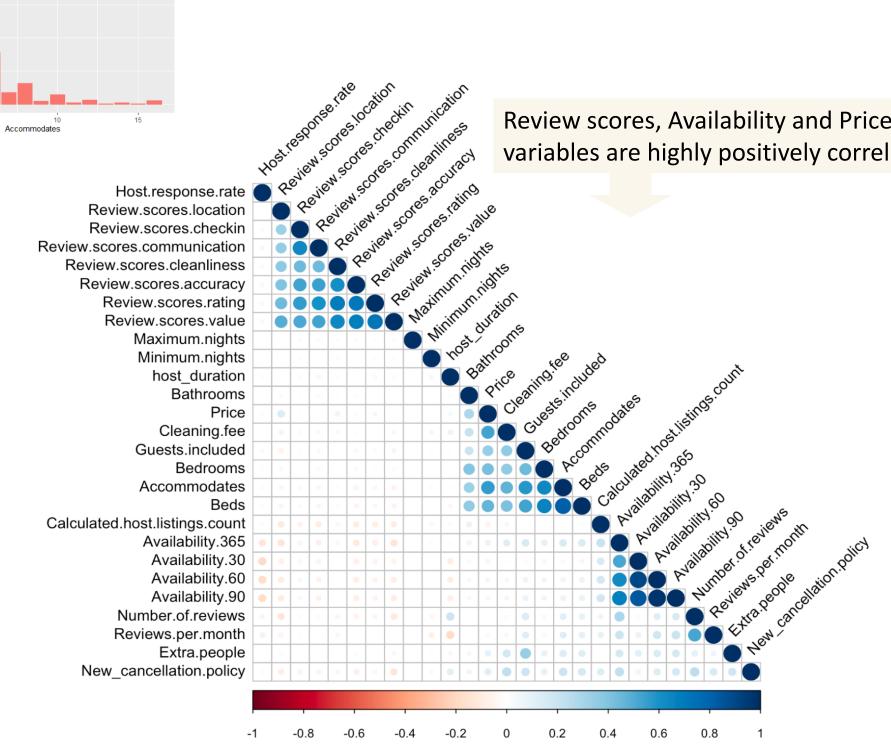
#### **Cleaned dataset:**

- > 43,760 observations
- > 180 variables (including dummy)

# **EXPLORATORY DATA ANALYSIS**

- > Variables selection: Reduced data to 44 variables having at least 60% non-missing data
- > Treatment: Replaced missing values by median of that variable
  - Cleaning fee- Replaced missing values by median fee of that neighborhood





## **MODELS**

Train: 30,649 (70%) Validate: 6,628 (15%) Test: 6,483 (15%)

**Prediction power comparison:** We used RMSE (Root Mean Square Error) to evaluate how well these models perform

### 1. Linear Regression

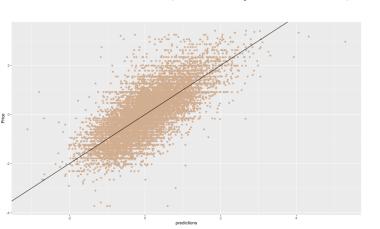
#### > Variables selection:

Removed independent variables that are highly correlated (VIF > 10) or statistically insignificant

#### > Pre-processing:

Log transformation + standardization

> RMSE: 0.66 (Actual price: 2.55)



term	estimate	std.error	statistic	p.value	term	estimate	std.error	statistic	p.valu
(Intercept)	0.002	0.004	0.581	0.561	Host greets you	-0.011	0.004	-2.822	0.
Host.response.rate	-0.007	0.004	-1.760	0.078	Hot tub	-0.011	0.004	-2.883	0.
Accommodates	0.370	0.005	69.434	0.000	Indoor fireplace	0.042	0.004	10.756	0.
Bathrooms	0.017	0.004	4.061	0.000	Internet	-0.024	0.005	-5.293	0.
Bedrooms	0.046	0.005	9.388	0.000	Iron	-0.015	0.005	-3.058	0.
Cleaning.fee	0.239	0.005	51.518	0.000	Keypad	-0.023	0.004	-5.791	0.
Extra.people	0.001	0.004	0.196	0.845	Kitchen	-0.007	0.004	-1.703	0.
Minimum.nights	-0.045	0.004	-11.103	0.000	Lake access	-0.008	0.004	-2.067	0.
Availability.30	0.060	0.004	14.436	0.000	Lock on bedroom door	-0.073	0.004	-17.427	0.
Number.of.reviews	-0.039	0.005	-8.621	0.000	Luggage dropoff allowed	0.008	0.004	1.831	0.
Review.scores.rating	0.019	0.006	3.014	0.003	Microwave	-0.022	0.005	-4.205	0.
Review.scores.cleanliness	0.045	0.006	7.921	0.000	Outlet covers	-0.012	0.004	-2.879	0.
Review.scores.location	0.156	0.005	34.320	0.000	Pack 'n Play/travel crib	0.010	0.004	2.388	0.
Review.scores.value	-0.100	0.006	-15.764	0.000	Pets allowed	-0.003	0.004	-0.880	0.
Calculated.host.listings.count	0.063	0.007	9.177	0.000	Pets live on this property	-0.052	0.004	-13.420	0.
Air conditioning	0.062	0.004	15.010	0.000	Pool	0.013	0.004	3.152	0.
Beachfront	-0.011	0.004	-2.697	0.007	Private bathroom	0.000	0.004	-0.095	0.
Buzzer/wireless intercom	0.022	0.004	5.217	0.000	Private entrance	0.007	0.004	1.604	0.
Cable TV	0.049	0.004	11.118	0.000	Private living room	-0.022	0.004	-5.501	0.
Crib	0.008	0.004	1.980	0.048	Safety card	0.017	0.004	4.089	0.
Dishwasher	0.030	0.005	6.262	0.000	Shampoo	0.040	0.004	9.226	0.
Doorman	0.068	0.005	15.151	0.000	Smoking allowed	-0.015	0.004	-3.871	0.
Dryer	0.013	0.004	2.994	0.003	Suitable for events	0.015	0.004	3.737	0.
Elevator	0.057	0.005	12.012	0.000	TV	0.076	0.004	17.542	0.
Essentials	-0.009	0.004	-2.059	0.040	Wheelchair accessible	-0.012	0.004	-2.787	0.
Family/kid friendly	0.034	0.004	7.673	0.000	Window guards	-0.021	0.004	-5.233	0.
First aid kit	-0.034	0.004	-7.890	0.000	Wireless Internet	-0.031	0.004	-7.527	0.
Free parking on premises	-0.059	0.004	-14.952	0.000	New_superhost_T	-0.001	0.004	-0.237	0.
Free parking on street	-0.011	0.004	-3.142	0.002	Room.type_SharedRroom	-0.098	0.004	-25.238	0.
Garden or backyard	-0.012	0.004	-2.831	0.005	New_cancellation.policy	0.003	0.004	0.701	0.
Gym	0.045	0.004	10.390	0.000	Instant.bookable_T	-0.017	0.004	-4.302	0.
Hair dryer	0.042	0.005	8.375	0.000	host_duration	0.054	0.004	12.746	0.
Hangers	-0.029	0.005	-5.998	0.000	Calculated.host.listings	-0.160	0.007	-22.081	0.

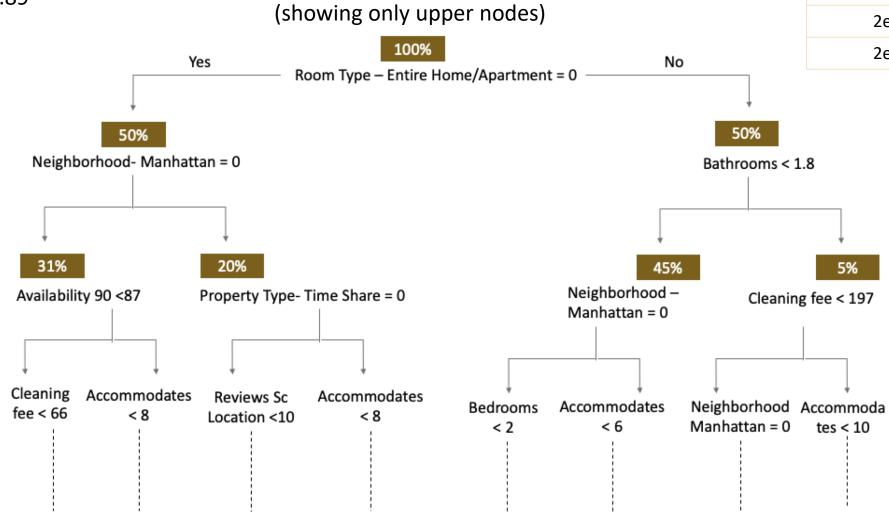
#### 2. CART

> Variables used in tree construction: Room type (Entire Home/apt), Neighborhood (Manhattan), Availability90, Accommodates, Property type (Timeshare), Review scores(location), Bathrooms.. +82 more

> **No. of splits:** 1800

> **RMSE:** 31.89

n type (Entire	3e-07	48.93035
Accommodates, Property type	2e-03	74.66704
nore	2e-04	60.59450
	5e-01	114.27472
	2e-05	50.43291
upper nodes)	2e-06	48.93901
00% No	2e-09	48.93035
ire Home/Apartment = 0		



#### 3. XgBoost

- > Objective: Linear regression
- > **RMSE:** 2.5

#### **RESULT**

- > RMSE: XgBoost 2.5, Linear regression 0.66 (actual price 2.55), CART 31.89
- > XgBoost has given the best predictability for price compared to linear regression model and CART
- > Room type, Bathroom and Neighborhood are the top parameters which influence the price