Predicting the Price of AirBnb in New York

Balram Prasad, Rafael Silva, Claire Zhu, Matt Campbell

Overview

Company Background:

Airbnb is a community-based, two-sided online platform that facilitates the process of booking private living spaces for travelers.

Website is a competitive marketplace, and the onus of pricing is left to the lister

Client Problem:

- Our client is AirBnb. Our goal is to help AirBnb give their customers recommendation on **pricing**.

Our Role:

- Our team built a model using data from the AirBnb website to predict the price of listings

Data

Our dataset had 17 columns not including price. We believe that there are three broad categories that would drive the price of AirBnb listings
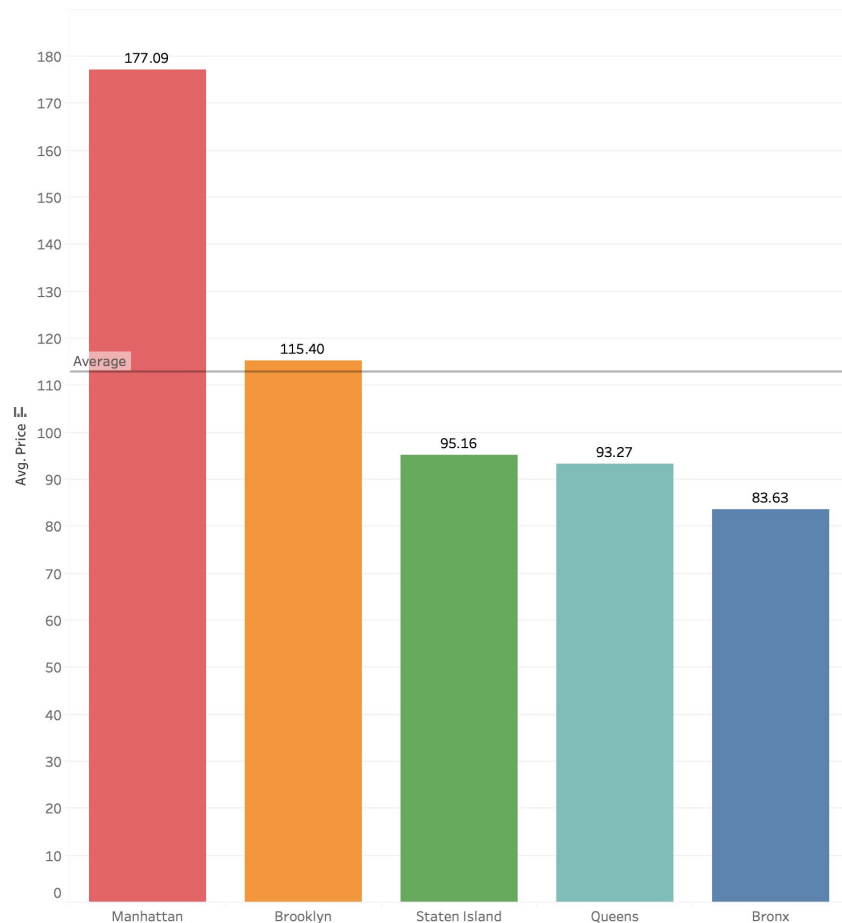
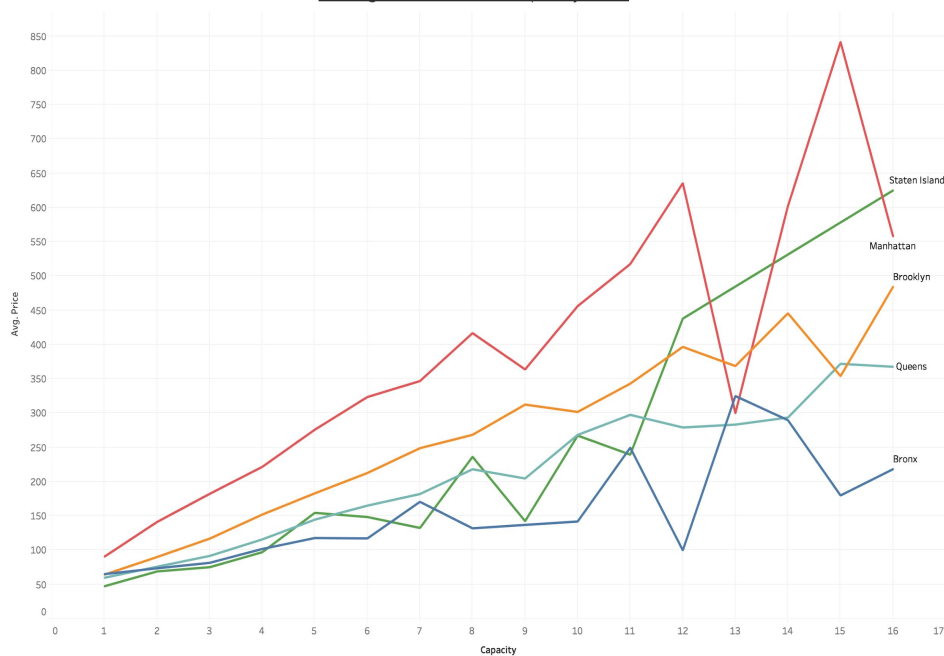|	Type	Potential Variables
Location	Qualitative	Boroughs Neighborhoods
Type of Listing	Qualitative	Property Type Room Type
Capacity	Quantitative	Beds Bedrooms Accomodations
Bathrooms	Quantitative	Bathrooms

Important Variables

1. **Boroughs and Price**

Average Price of Listings along NYC Boroughs

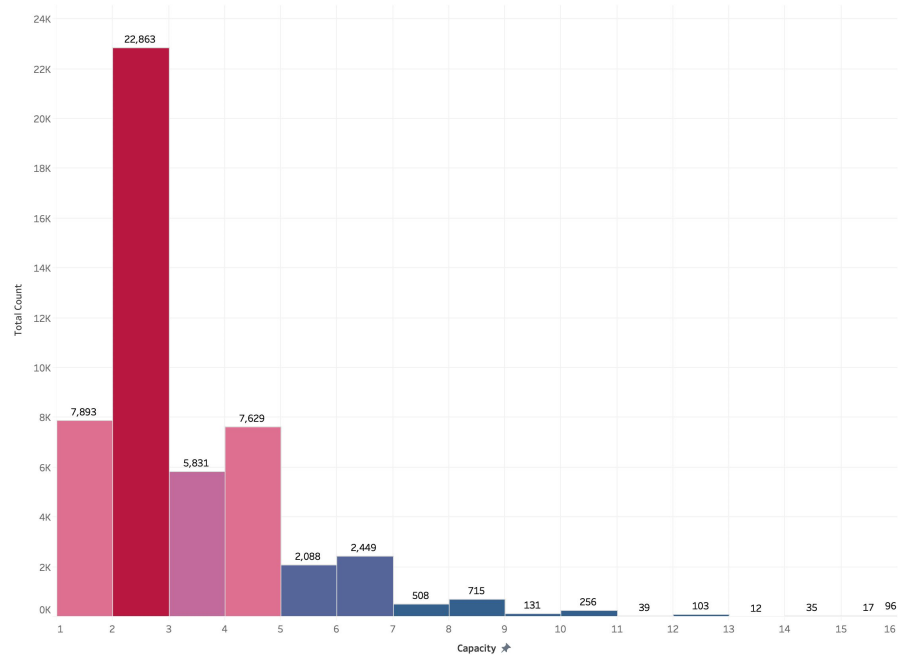| Borough	Avg. Price
Manhattan	177.09
Brooklyn	115.40
Staten Island	95.16
Queens	93.27
Bronx	83.63

Average (line at ~112)

Important Variables

Average Price for Each Capacity Level

Total Observations per Capacity Level

2. Capacity and Price

Important Variables

3. Capacity and Price

Entire Home/Apartment
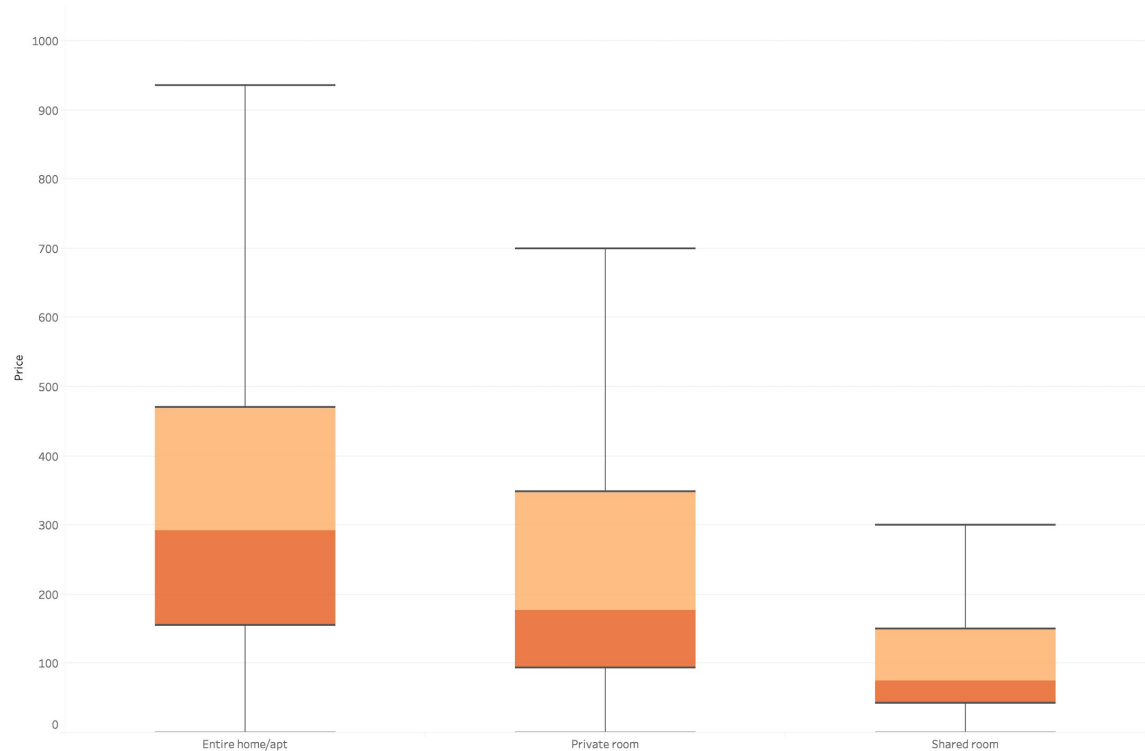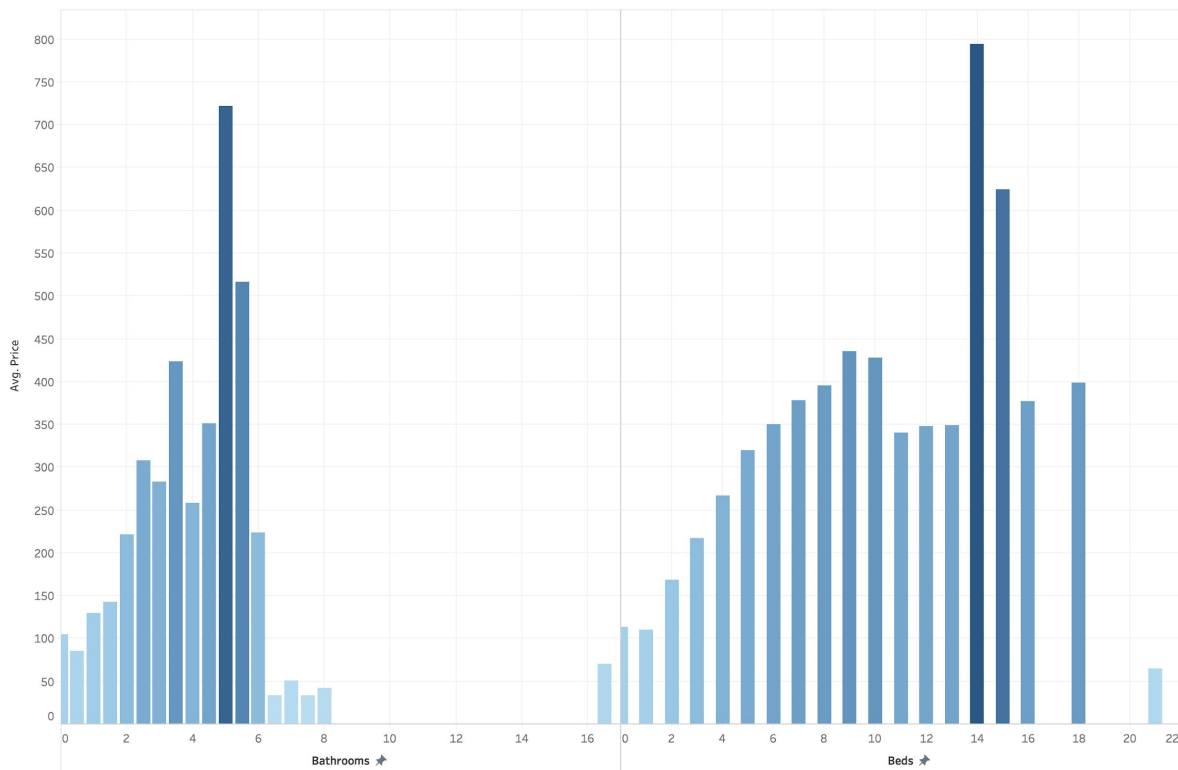
⬇

Private Room

⬇

Shared Room

Room Type vs. Price

Important Variables

Average Price for amount of Bathrooms and Beds

4. *Bathrooms*

Mostly positive correlation

Did not include Beds because of multicollinearity

Model – QDA

	Predictions			
	$0 - $69	$69 - $106	$106 - $175	$175 +
Actual observations $0 - $69	803	1538	109	72
$69 - $106	349	1542	333	345
$106 - $175	81	685	537	1145
$175 +	38	219	239	2101

Table 1. Confusion matrix

QDA Evaluation

	$0 - $69	$69 - $106	$106 - $175	$175 +
sensitivity	0.3183981	0.6002336	0.2193627	0.8090104
specificity	0.5489887	0.4547377	0.5783039	0.3822788
accuracy	0.491614			

Table 2. Evaluation based on sensitivity, specificity, and accuracy.

Model – Linear Regression

Model Formula: Price ~ Accommodates + Bathrooms + Room_Type + Boroughs

Visualization
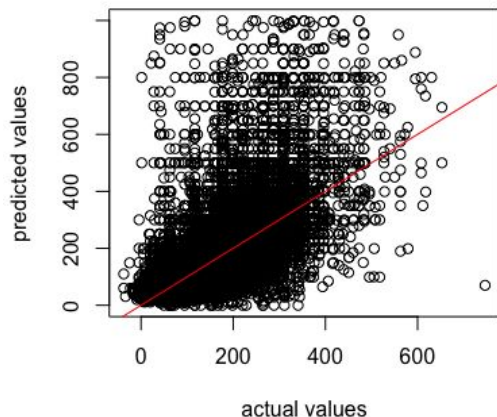
Model Statistics

```
Residuals:
    Min      1Q  Median      3Q     Max
-676.46  -43.23   -9.75   23.28  933.38

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                15.2704     2.9473   5.181 2.21e-07 ***
accommodates                               21.8186     0.2520  86.595  < 2e-16 ***
bathrooms                                  44.9478     0.9597  46.833  < 2e-16 ***
neighbourhood_group_cleansedBrooklyn       18.7340     2.7170   6.895 5.45e-12 ***
neighbourhood_group_cleansedManhattan      72.8619     2.7167  26.820  < 2e-16 ***
neighbourhood_group_cleansedQueens          6.0630     2.8810   2.104   0.0353 *
neighbourhood_group_cleansedStaten Island -10.5061     5.3256  -1.973   0.0485 *
room_typePrivate room                     -60.1517     0.8959 -67.138  < 2e-16 ***
room_typeShared room                      -80.5925     2.5859 -31.166  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 83.88 on 50656 degrees of freedom
Multiple R-squared:  0.4495,    Adjusted R-squared:  0.4494
F-statistic:  5170 on 8 and 50656 DF,  p-value: < 2.2e-16
```

Final Product

For an AirBnB listing, the price should be calculated following this <u>formula:</u>

- *Start* at base price of **$15**
- *Add* **$22** for every person you can accommodate
- *Add* **$45** for every bathroom in the listing
- *Add* **$70** if the listing is located in Manhattan
- *Add* **$20** if the listing is located in Brooklyn
- *Add* **$6** if the listing is located in Queens
- *Discount* **$10** if the listing is located in Staten Island
- If the listing is a private room (not an entire home), discount **$60**
- If the listing is a shared room (not an entire home), discount **$80**

Validating the Model

AirBnB Pricing

Rules:	Base Price:	$15.00	
Per Person	$22.00	Queens	$6.00
Per Bathroom	$45.00	Staten Island	-$10.00
Manhattan	$70.00	Private Room	-$60.00
Brooklyn	$20.00	Shared Room	-$80.00

Property Name	Location	Room Type	Capacity	Bathroom	Model Prediction	Actual Value	Percentage Error
1	Manhattan	Entire Home	4	1	$218.00	$225.00	3.11%
2	Bronx	Private Room	2	1	$44.00	$42.00	-4.76%
3	Staten Island	Entire Home	8	2	$271.00	$200.00	-35.50%
4	Queens	Entire Home	1	1	$88.00	$75.00	-17.33%
5	Brooklyn	Entire Home	8	1	$256.00	$260.00	1.54%
6	Brooklyn	Private Room	2	1	$64.00	$79.00	18.99%
7	Manhattan	Entire Home	2	1	$174.00	$185.00	5.95%
8	Bronx	Entire Home	3	1	$126.00	$125.00	-0.80%
9	Queens	Entire Home	3	1	$132.00	$100.00	-32.00%
10	Staten Island	Entire Home	5	1	$160.00	$221.00	27.60%

Recommendation

- We **would recommend** that AirBnb adopt our model to recommend estate pricing for their users

- Very low R2 - only ~50 of variation in price is explained by our model.

- Validation method showed that most predictions are within 30%
  - Not concerning given context

- Value to consumers would outweigh the error

Limitations

- Can only use the model for New York
- Income in the area or proximity to tourist destinations - major confounding variable, especially in New York
- Seasonality, customer sentiment, related costs, etc.

Next steps

1. Evaluate the model in the field

2. Consider and gather new variables (regional income, parking)

3. Evaluate seasonality in the model

4. Evaluate data progression through time for more accurate results if there is knowledge of Time Series

5. Extrapolate the model to other cities

Questions?