PRICE VS. QUALITY TRADEOFF IN LOS ANGELES AIRBNB MARKET

BY: ADAM MESSINA, SATHISH KOMIRE, NAGA RAJU NANNAPANENI, SIDDHARTH MAREDU & SAI VENKAT

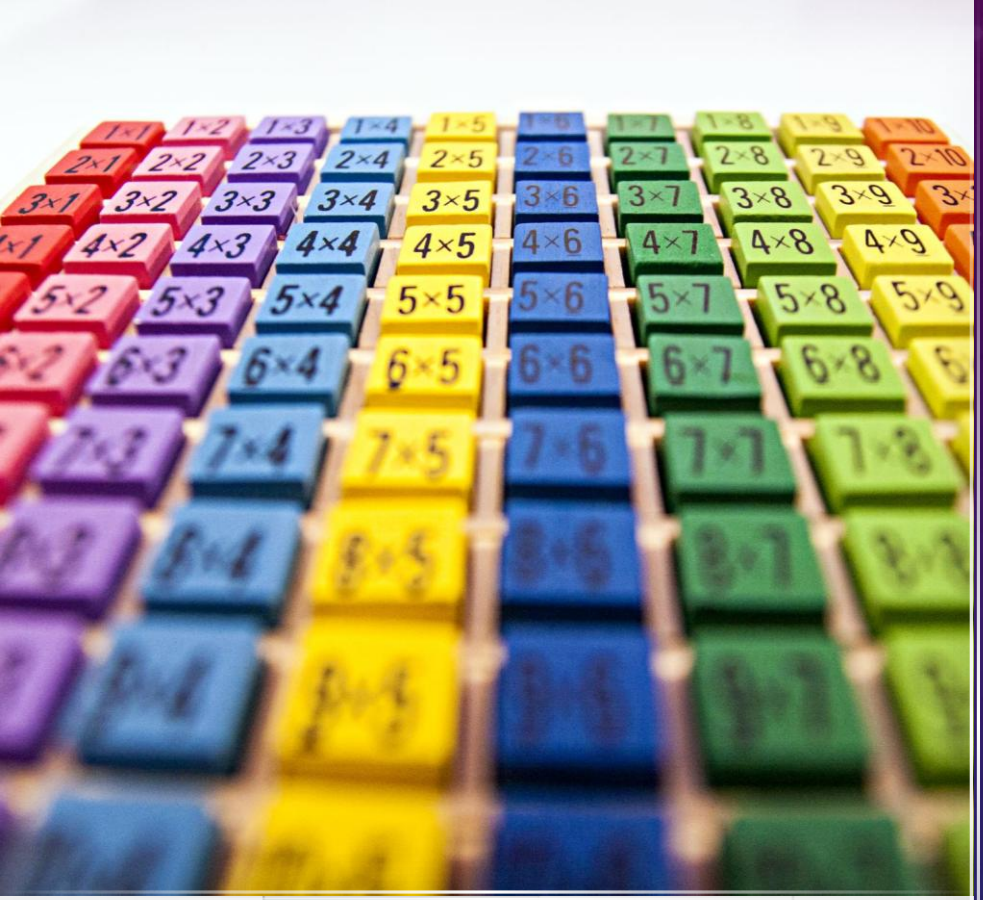
- Executive Summary
- Research Questions
- Methodology
  - General Predictive Model
  - LA Predictive Model
  - Causal Graph
- Results and Findings
  - Size
  - Price
  - Host Responsiveness
- Conclusion

RESEARCH QUESTIONS

- How is the high booking of Airbnb's effected by:
  - Size
    - Is size a significant variable?
    - What size Airbnb should our investor purchase?
  - Price
    - Is price a significant variable?
    - What is the best price to set for an Airbnb?
  - Host Responsiveness
    - Is host responsiveness a significant variable?
    - What expectations should we set to inform the investor on how often to respond to guests?

OVERVIEW OF METHODOLOGY

- EDA and Assumptions
  - Data Cleaning
  - Correlation Matrices
  - Checking For Multi-Collinearity
- Predictive Modeling
  - Achieve High AUC for all Airbnb Markets
  - Run the same model in LA
  - Get Importance Scores
- Create Causal Graph to Determine Relationships Between Variables
- Form Hypotheses Based On Causal Graph
  - Identify Confounders and Other Necessary Fixed Effects
  - Determine Best Size and Price of Airbnb's to Invest In
  - Set Management Expectations For Investor



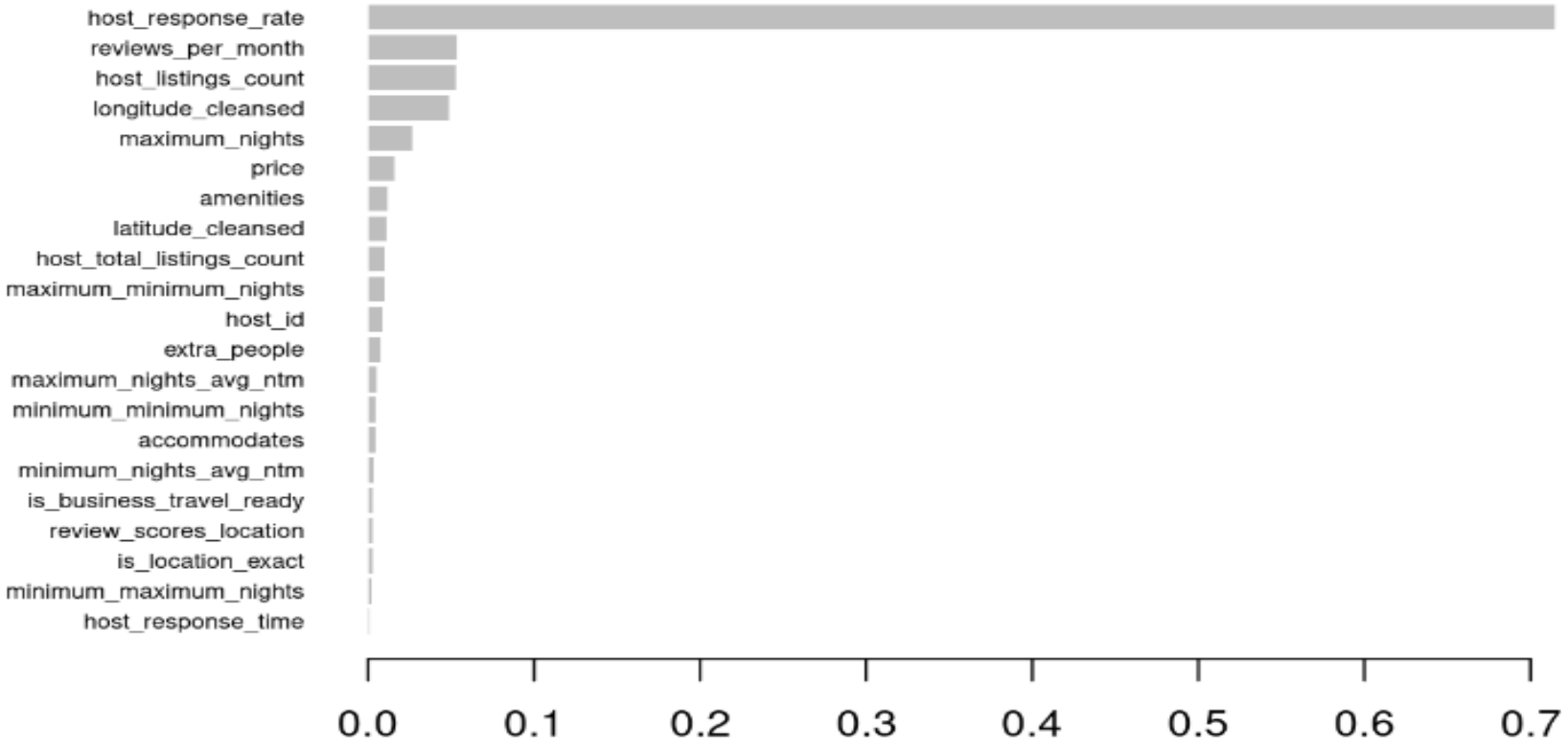
ost_response_time	host_response_rate	host_is_superhost	is_location_e
0.115780639	0.023304002	-0.055897265	-0.06115
1	0.62895691	0.139808309	-0.02933
0.62895691	1	0.131872087	0.02147
0.139808309	0.131872087	1	0.03510
-0.029334162	0.021474772	0.035101246	
0.080334067	0.044971569	0.085449692	0.01554
0.036924436	0.019796113	0.071161006	0.01669
0.070309406	0.043195757	0.07335277	0.02199
-0.013430397	-0.024235025	0.011354175	-0.00144
-0.011257998	-0.02239915	0.011922166	-0.00252
-0.00937366	-0.021110838	0.012064147	-0.00392
0.080746522	0.047155994	0.088848459	-0.04283

METHODOLOGY
EDA AND ASSUMPTIONS

- Basic Cleaning operation into one file used for modelling
  - Duplicates
  - Missing values handling - mean/median
  - Quantifying amenities
  - Special characters in values handling
  - Encoding categorical variables
- Logistic Regression
- Multicollinearity checks

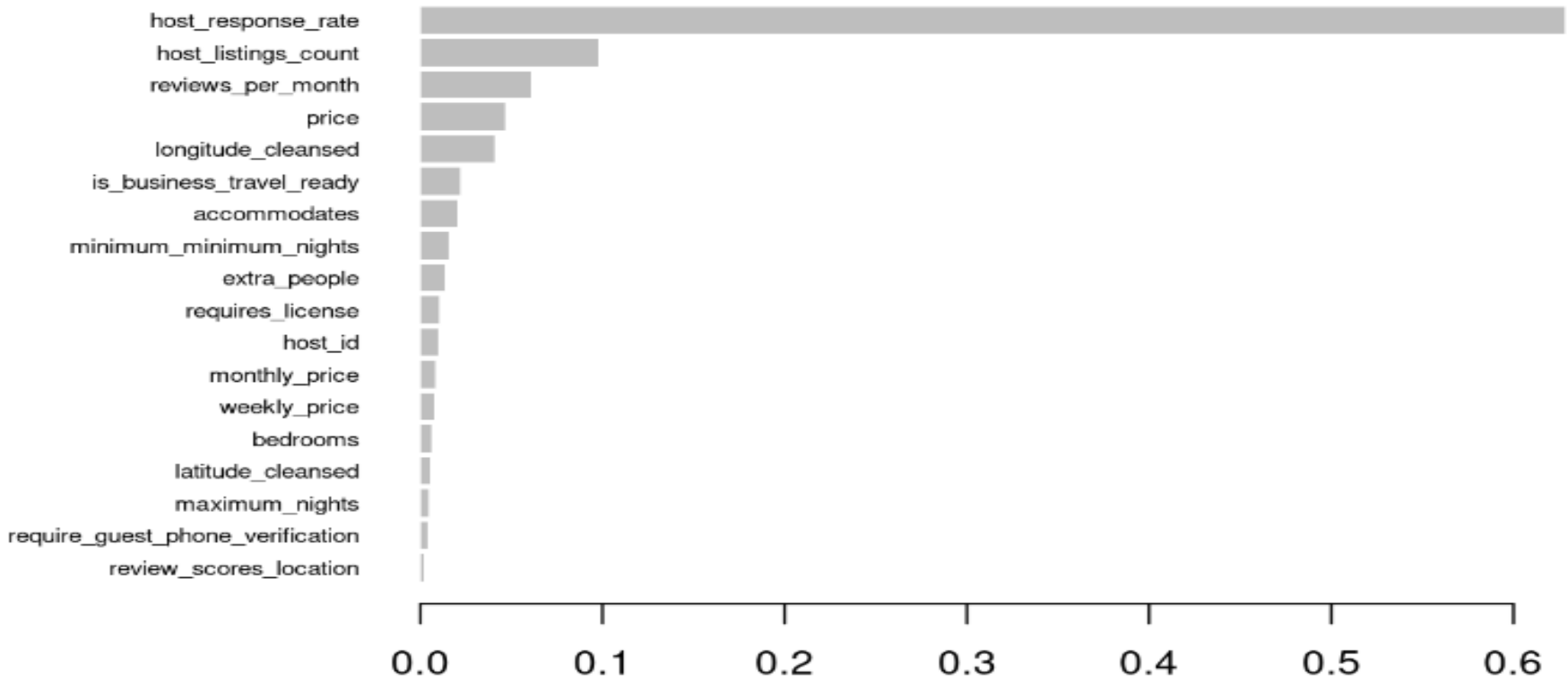
METHODOLOGY
GENERAL PREDICTIVE MODEL

```
## Area under the curve: 0.8401
```

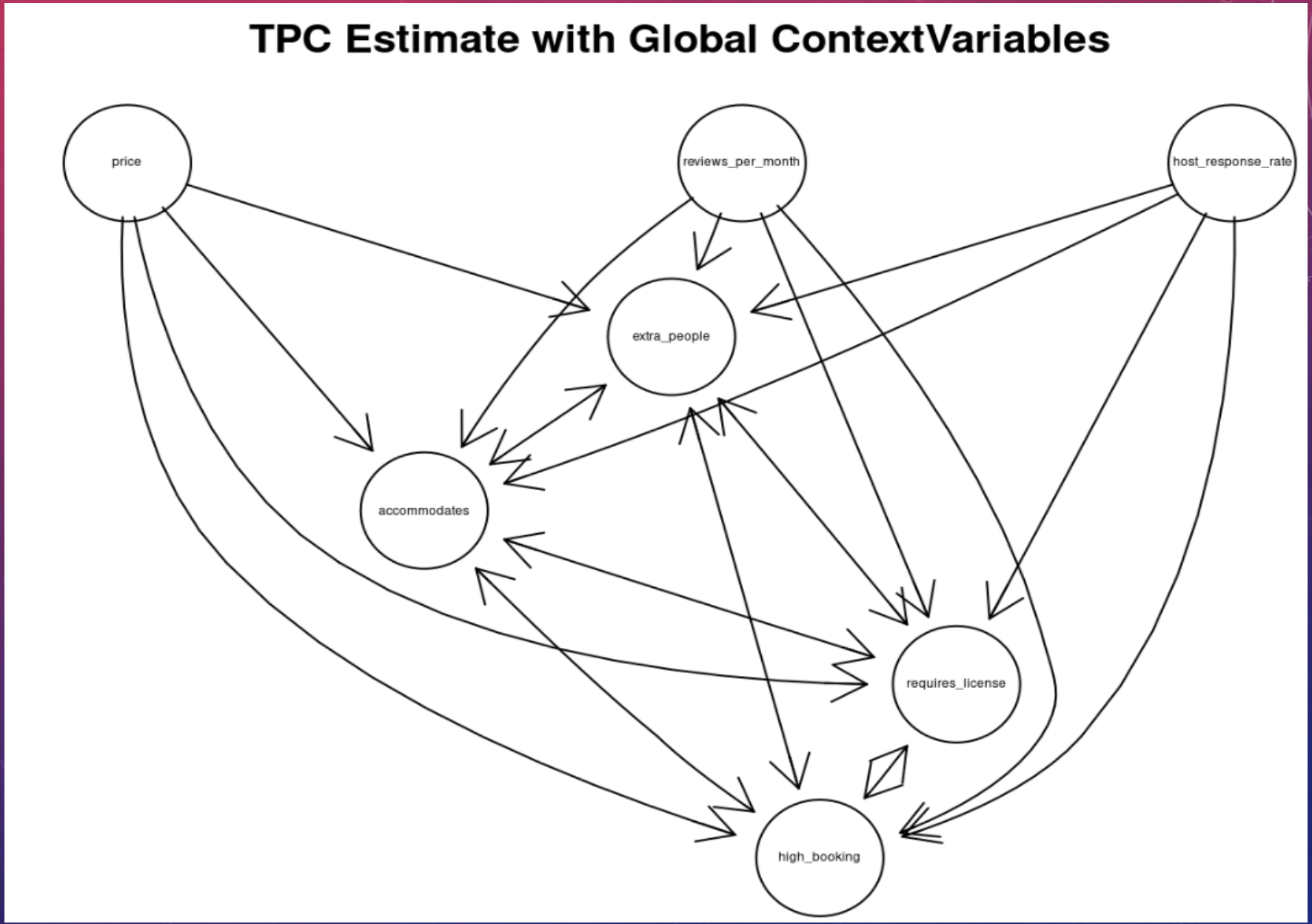


METHODOLOGY
LA PREDICTIVE MODEL

```
## Area under the curve: 0.849
```



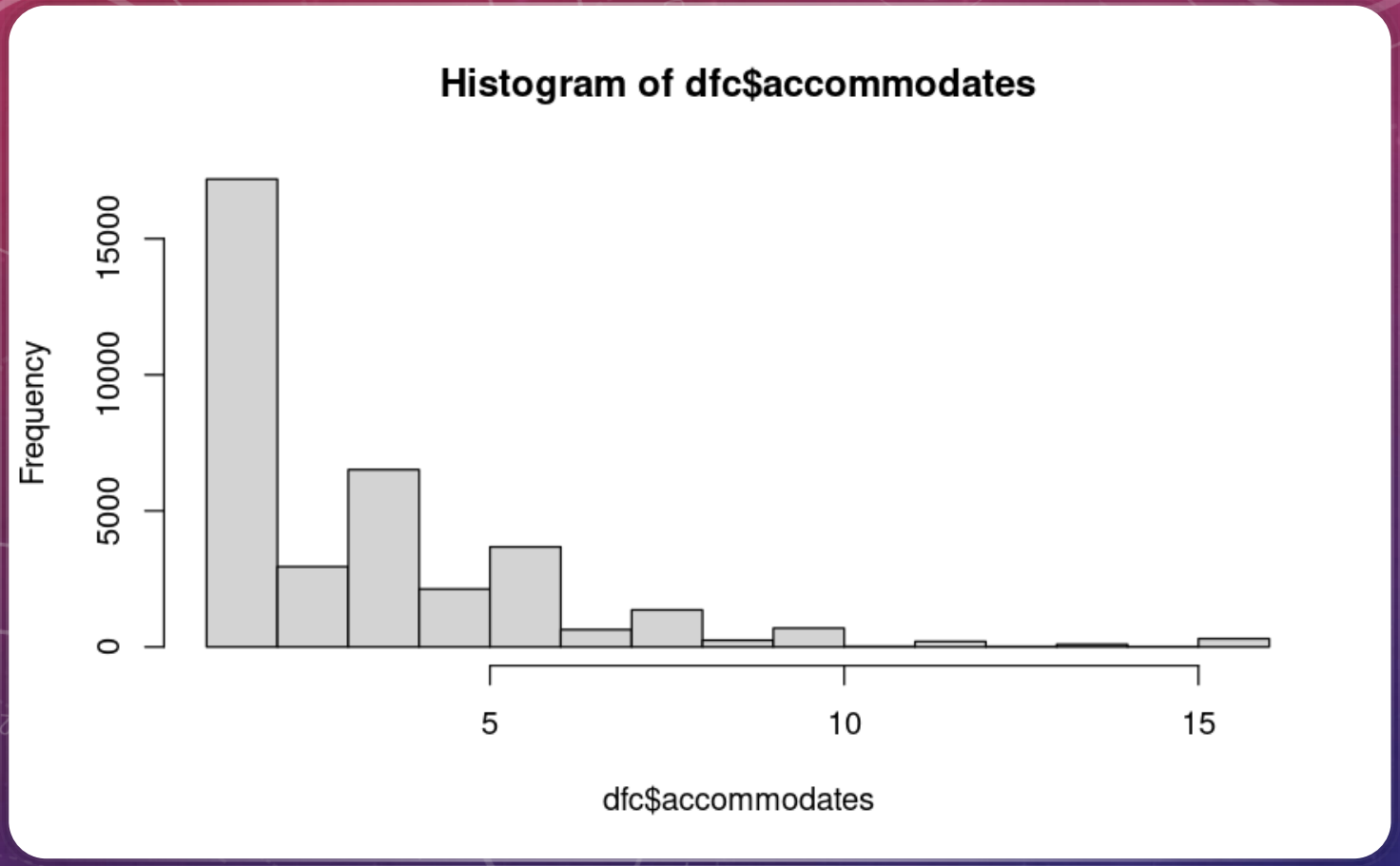
METHODOLOGY CAUSAL GRAPH



RESULTS AND FINDINGS

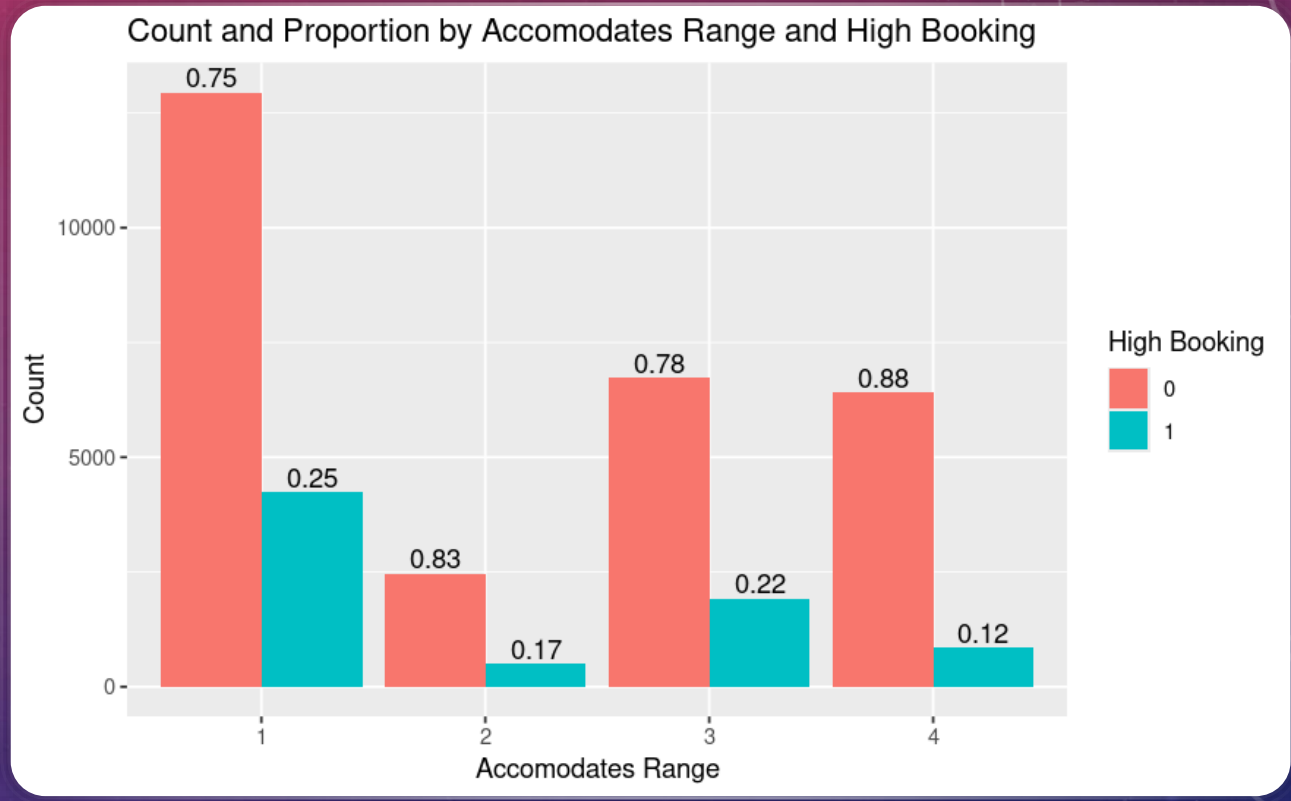
- Categorized Each Variable Based On Their Four Quantiles
- Used Each Quantile As Treatment and All Others As Control
  - Iterated This Process For All Quantiles of A Given Variable
- Controlled For Confounders and Modeled To Get ATE (Average Treatment Effect)

ACCOMMODATES HISTOGRAM

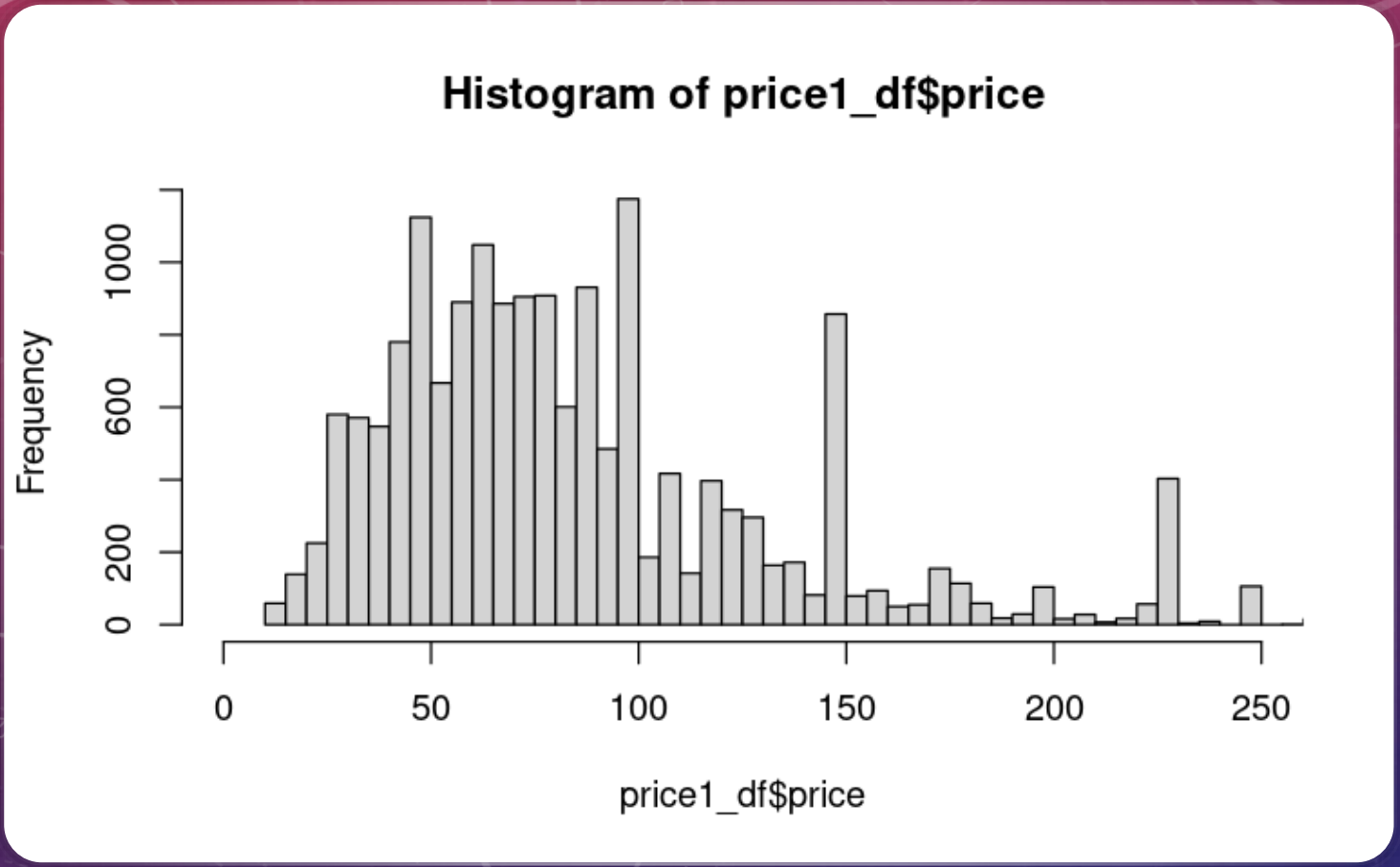


RESULTS AND FINDINGS : OVERALL SIZE

- 1st Quartile (Lowest 25%): accommodates at most 2 people
- 2nd Quartile (25% to 50%): accommodates at most 3 people
- 3rd Quartile (50% to 75%): accommodates at most 5 people
- 4th Quartile (Top 25%): accommodates at most 16 people

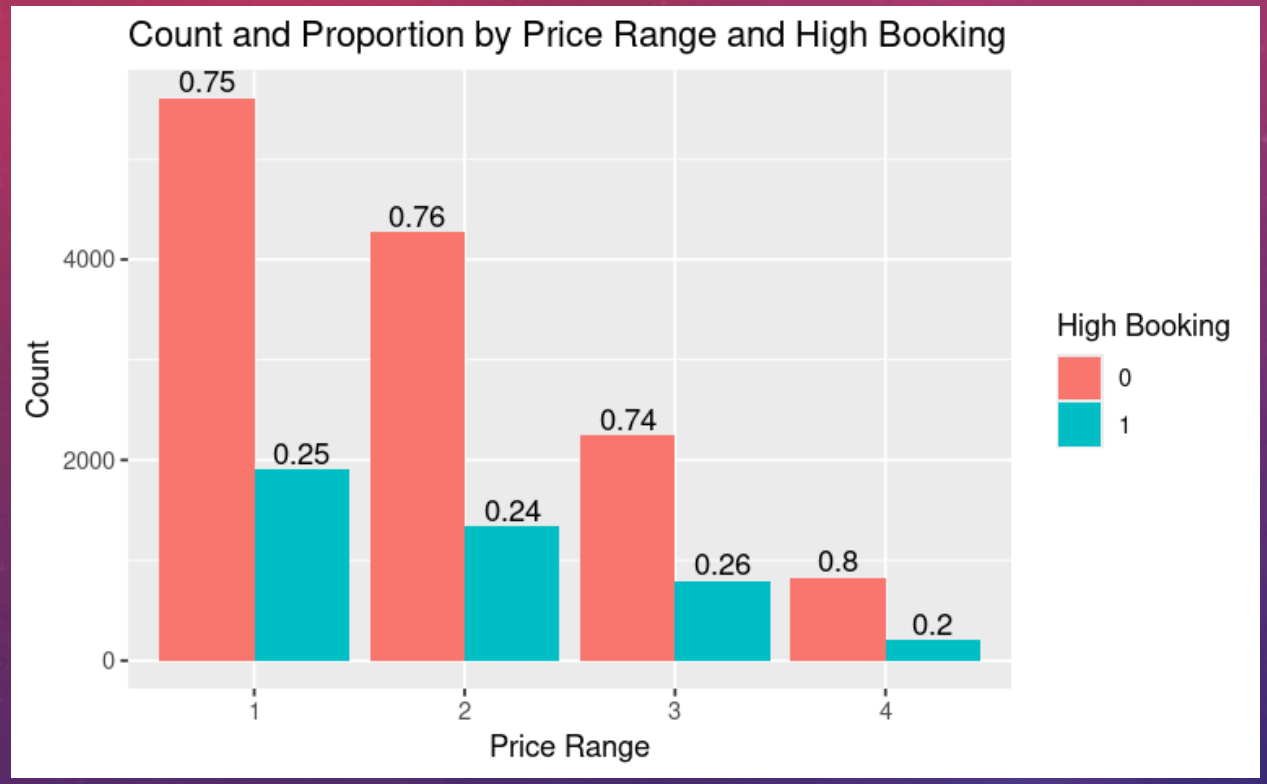


Higher chance of high booking for accommodating airbnb's accommodating 1-2 people or 4-5 people



HISTOGRAM OF AIRBNB'S PRICING

RESULTS AND FINDINGS: AIRBNB'S PRICING



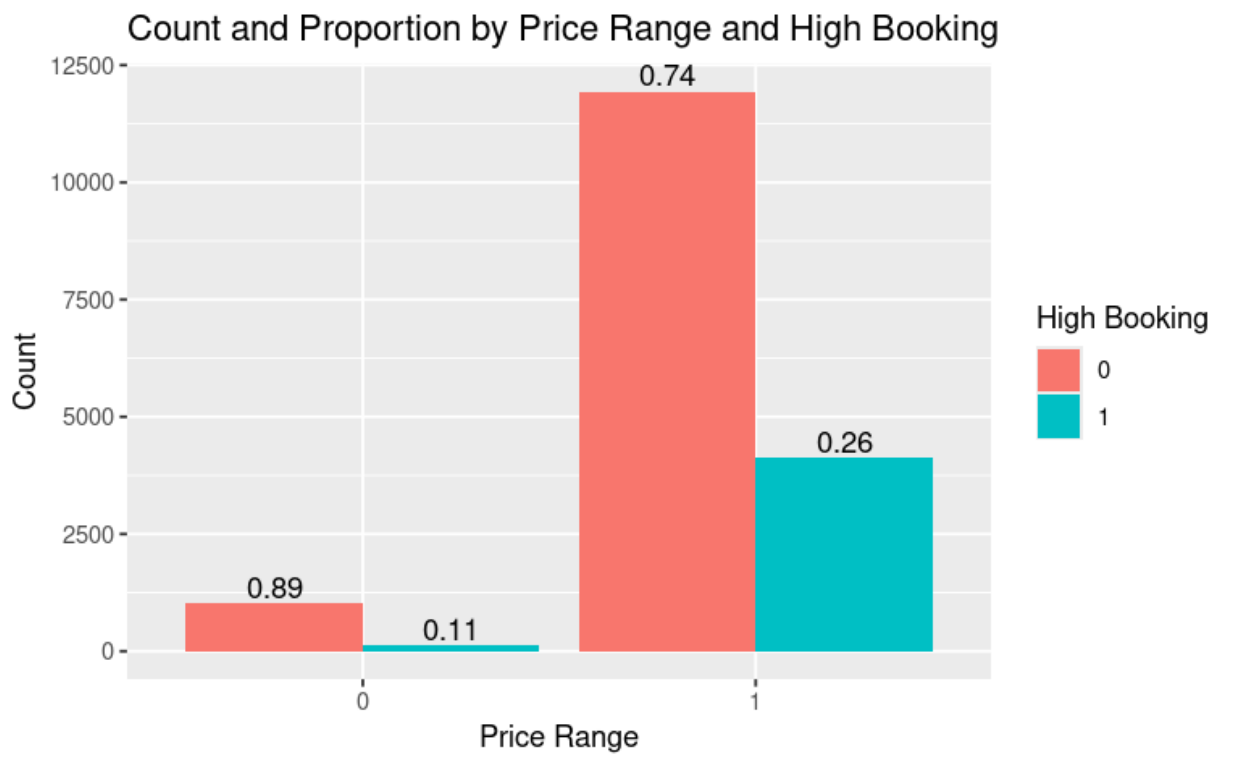

```r
quantile(price1_df$price)
```

```
  0%  25%  50%  75% 100%
  10   55   78  110 5000
```

Formulated count and proportion of price range by high booking

Advice to not charge Airbnb's over $110 for Airbnb's that accommodates 1-2 people

HOST RESPONSE RATE EFFECT ON HIGH BOOKING



Above 75% response rate improves high booking for Airbnb's that accommodates at most 2 people.

HOST RESPONSE RATE EFFECT ON 1-2 SIZED AIRBNB'S

- Airbnb's that accommodate at most 2 people, the host response rate should be at least 94%.

```{r}
quantile(host1_df$host_response_rate)
```

```
        0%        25%        50%        75%       100% 
 0.0000000 0.9463407 1.0000000 1.0000000 1.0000000 
```

CONCLUSION

01

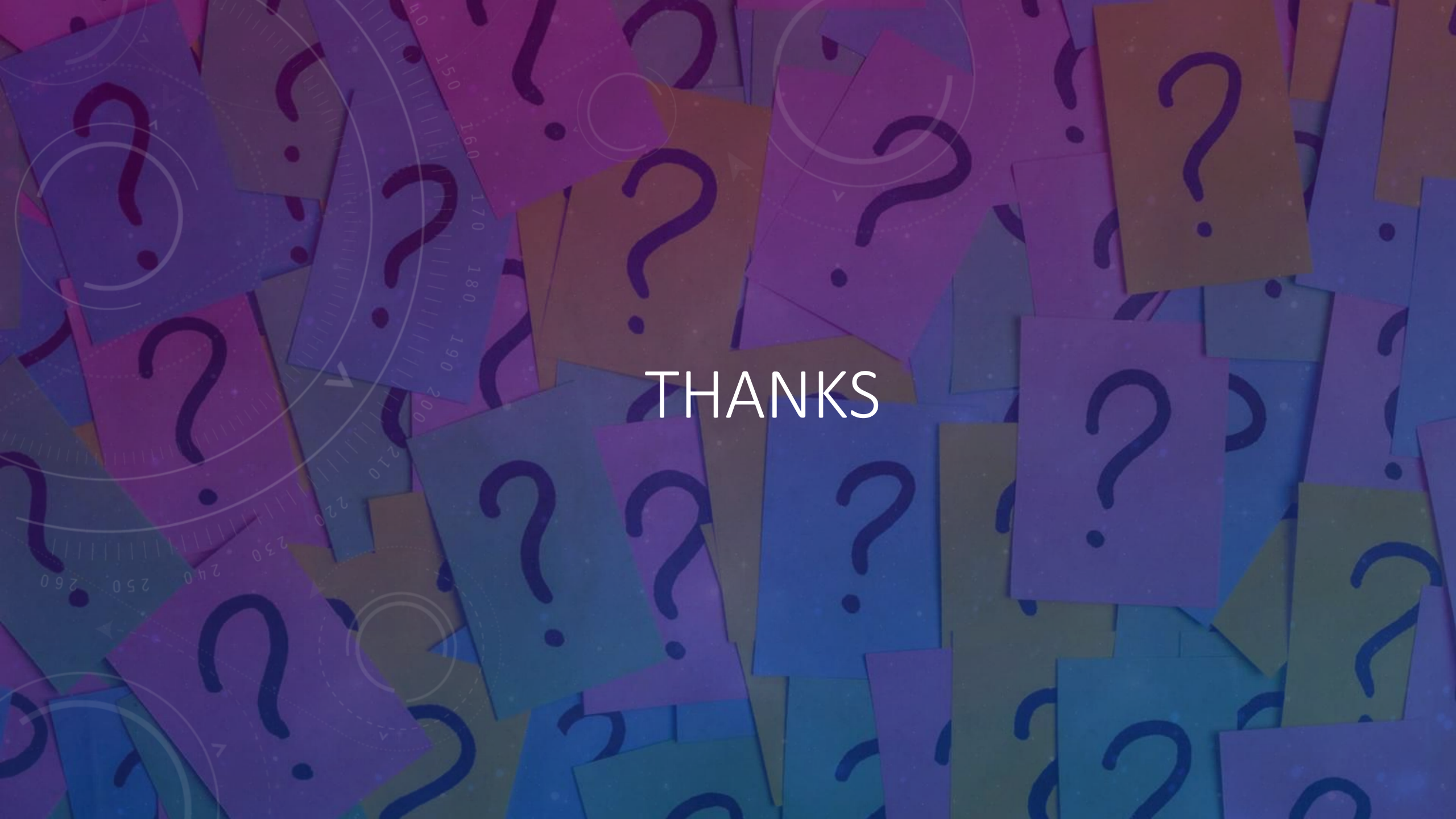
Invest In Airbnb's That Accommodates 2-3 People

02

Price These Airbnb's Less Than 100 or Greater Than 150

03

Achieve Host Response Rate 76.7% Or Higher


THANKS

Congratulations
CLASS OF
2024
Adobe Stock | #627095985