Airbnb investment in Austin, TX

market analysis and income modeling

DSI course

Feb 2024

By: Masoud Alfi



Problem statement

- How is the current Airbnb market in Austin?
- How to maximize income in this market?
- What are investment opportunities for short term rentals in Austin?

- Main objectives?
- Who benefits?
- Why is it important?



Our data

 Quarterly scraped Airbnb data (insideairbnb.com)

Home price information (Zillow.com\data)

- How did we deal with missing/unavailable info?
 - Data extracted from details, descriptions, and texts
 - Looked at previously scraped data

Our approach

Objective:

Estimate yearly (or quarterly) income

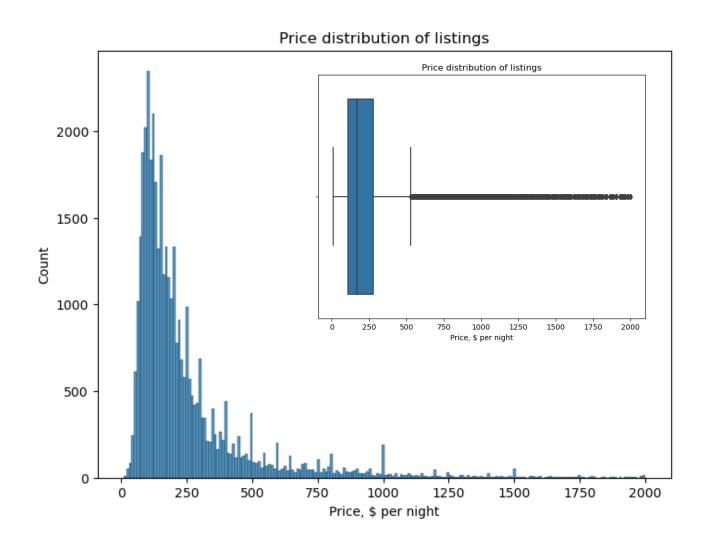
How?

Income = price × occupancy rate × days

- Target variables:
 - Listing price
 - Occupancy rate



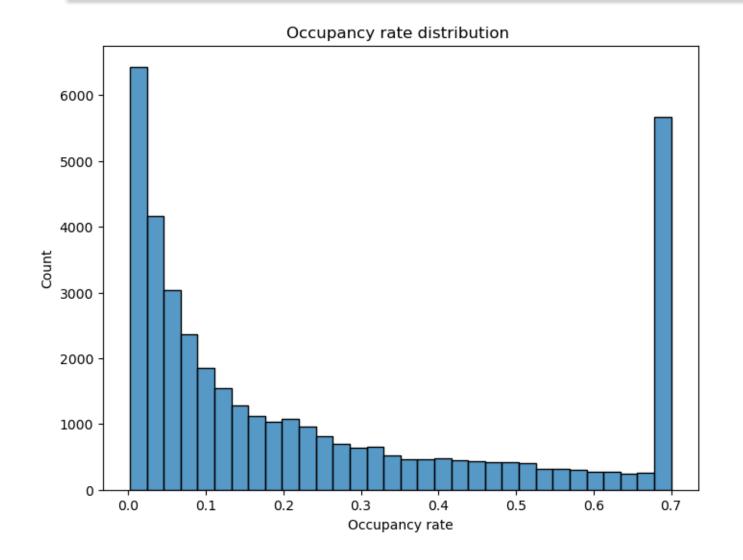
Listing price



- Average = \$280/night
- Min = 1
- Max = \$65k
- Cut-off value: \$2000/night

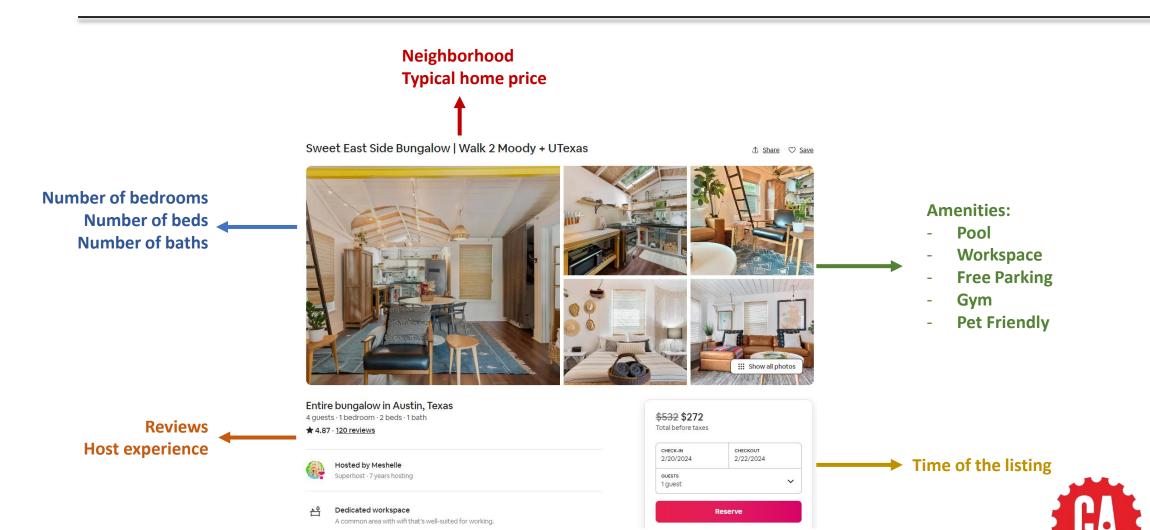


Occupancy rate

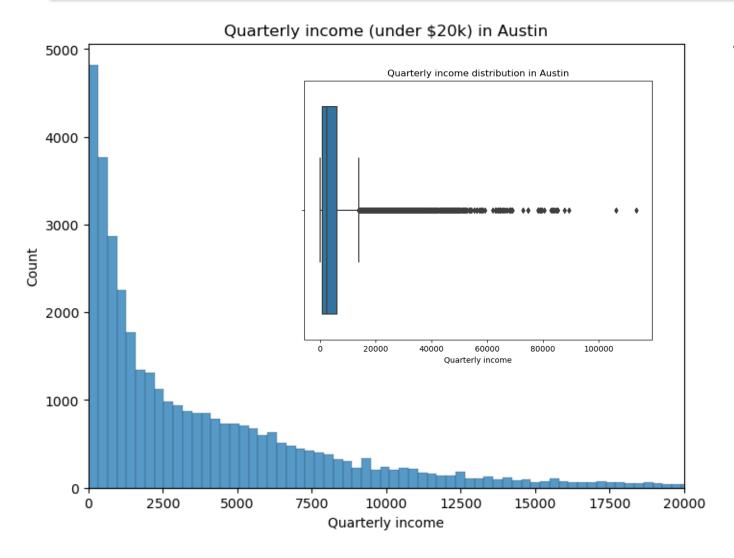


- We estimated it (how?)
- Max set to be 0.7 (why?)
- Average = 0.25
- Challenges:
 - Hard to verify
 - Hard to model (hardly correlated with our features)

What features to use?



Quarterly income



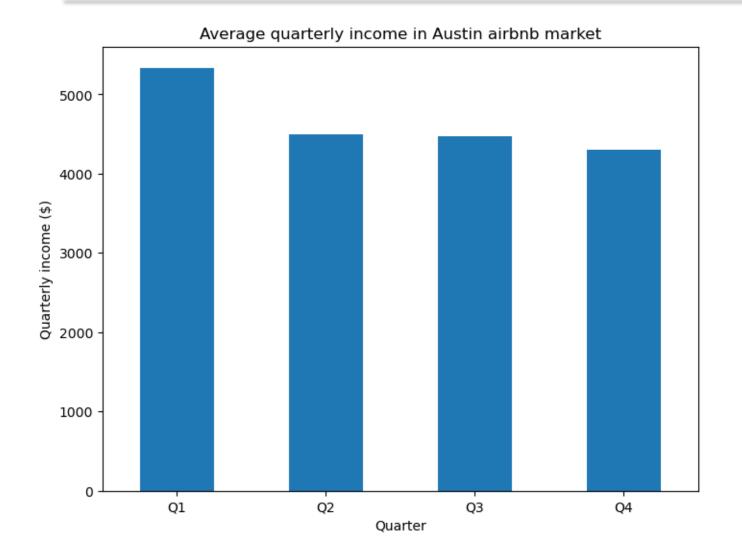
Very right-skewed

• Mean: \$4600/quarter

• Median: \$2400/quarter

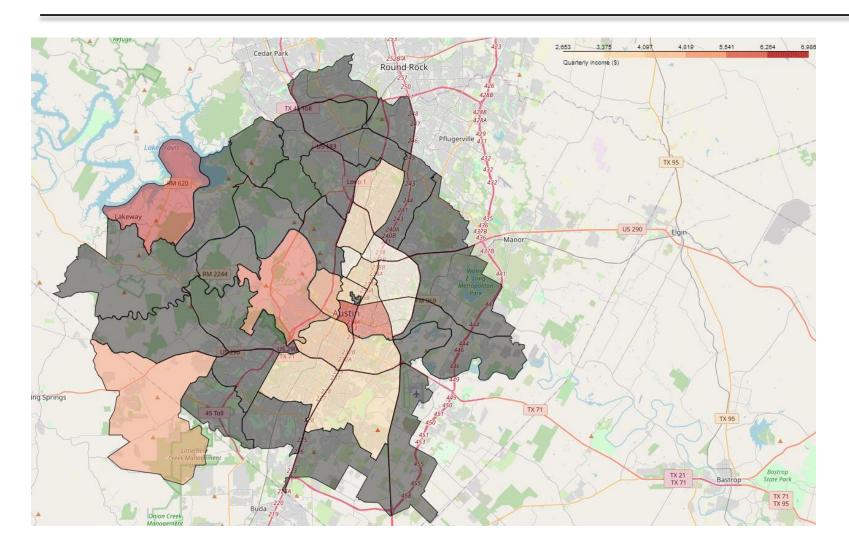


Quarterly income



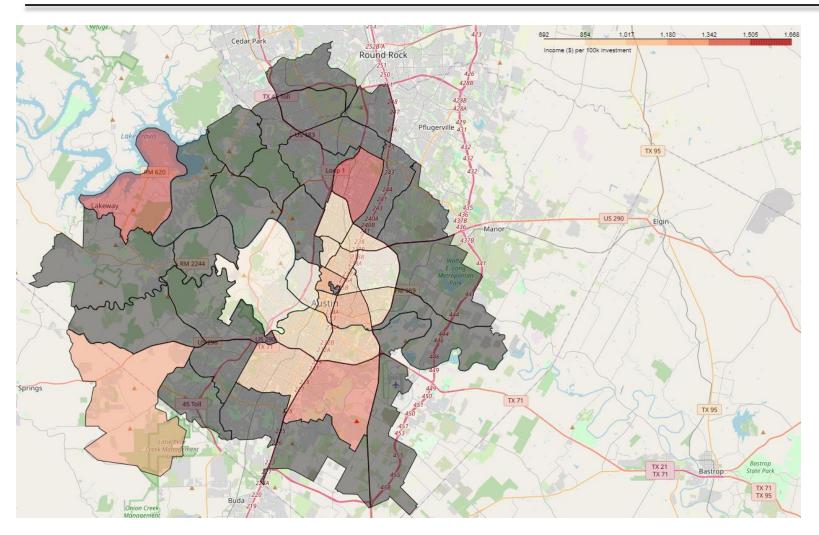
- Highest income in Q1
 - Higher occupancy rate
 - Higher prices
- Holiday time!

Highest income regions



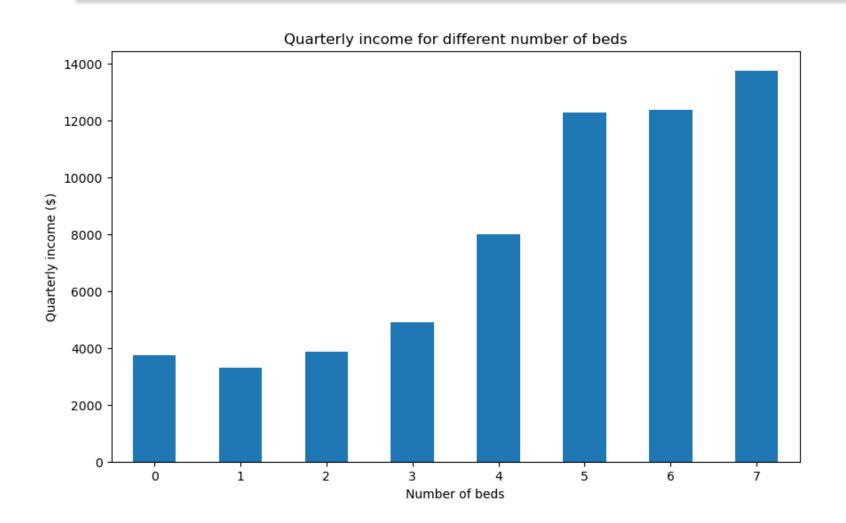
- Neighborhoods with 125+ homes
- Highest income:
 - Downtown area
 - Lake Travis area

Best investment regions



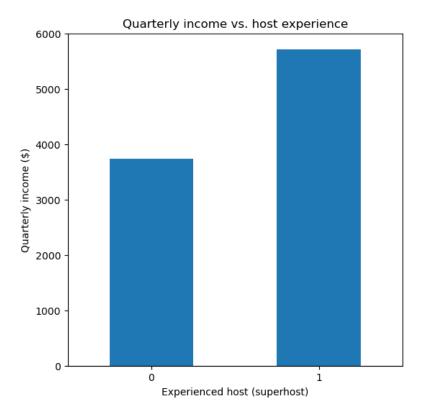
- Income per 100k investment
- Best for new investors
- Some regions are very expensive
- Away from the center!

Income vs. number of bedrooms

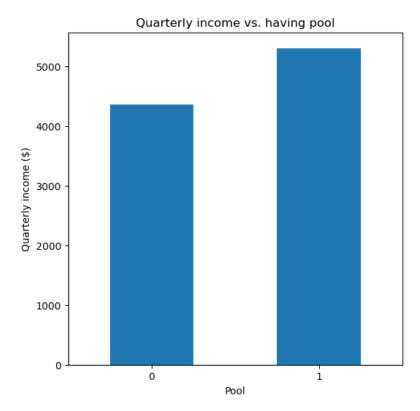


 Income jumps for 4+ bedrooms

Other important features



• Host's experience impacts income



 Having pool bumps up the income (for 2+ bedroom houses)

'Listing price' modeling

• Linear regression, Neural Network, and XGBoost

Regressor	Specifics	R2 score (train data)	MAE (train data)	R2 score (test data)	MAE (test data)
<u>Linear regression</u>	Polynomial features, Ridge	0.56	\$93	0.52	<u>\$97</u>
Neural Network	4 hidden layers (32- 256 nodes)	0.71	\$73	<u>0.58</u>	<u>\$88</u>
<u>XGBoost</u>	BayesSearchCV for hyperparameters	0.90	\$42	<u>0.71</u>	<u>\$72</u>

'Occupancy rate' modeling

- Very challenging modeling and poor performance
- Linear regression and XGBoost

Regressor	Specifics	R2 score (train data)	MAE (train data)	R2 score (test data)	MAE (test data)
Linear regression	Polynomial features, Ridge	0.17	0.18	0.13	<u>0.19</u>
<u>XGBoost</u>	BayesSearchCV for hyperparameters	0.66	0.11	<u>0.45</u>	<u>0.14</u>

Demo Time!



Conclusions

- The average quarterly income from airbnb in the city of Austin is \$4600.
- First quarter of the year usualy generates more income than other times.
- Houses with 4+ bedrooms show a jump in the income, compared to smaller properties.
- Being an experienced host (a.k.a. superhost) significantly impacts the yearly income of listings.

Future work?