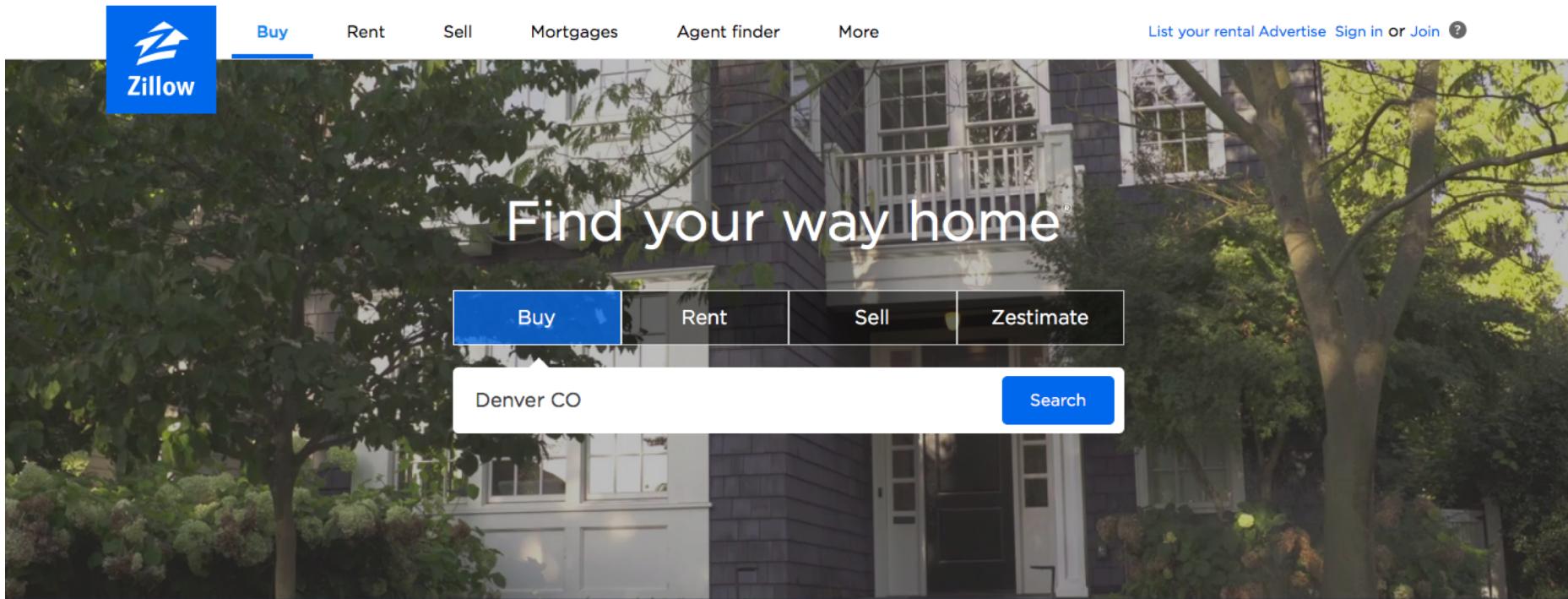


# Predicting house pricing in Denver, Colorado



Which features best  
explain the house  
prices in Denver?

# The data: 2912 houses



The image is a screenshot of the Zillow website homepage. The top navigation bar includes links for 'Buy', 'Rent', 'Sell', 'Mortgages', 'Agent finder', and 'More', with 'Buy' being the active tab. To the right are links for 'List your rental', 'Advertise', 'Sign in or Join', and a help icon. The main visual is a large, slightly blurred photograph of a two-story house with a porch, surrounded by trees and bushes. Overlaid on this image is the text 'Find your way home' in a large, white, sans-serif font. Below this, a search bar contains the text 'Denver CO' and a blue 'Search' button. The Zillow logo is in the top left corner of the page.

# Features

Categorical

Numerical

# Features

Numerical

- Size (sqft)
- Beds
- Baths

Year Built

School District  
Ranking  
GreatSchools ratings

Categorical

# Features

Numerical

- Size (sqft)
- Beds
- Baths

Year Built

School District  
Ranking  
GreatSchools ratings

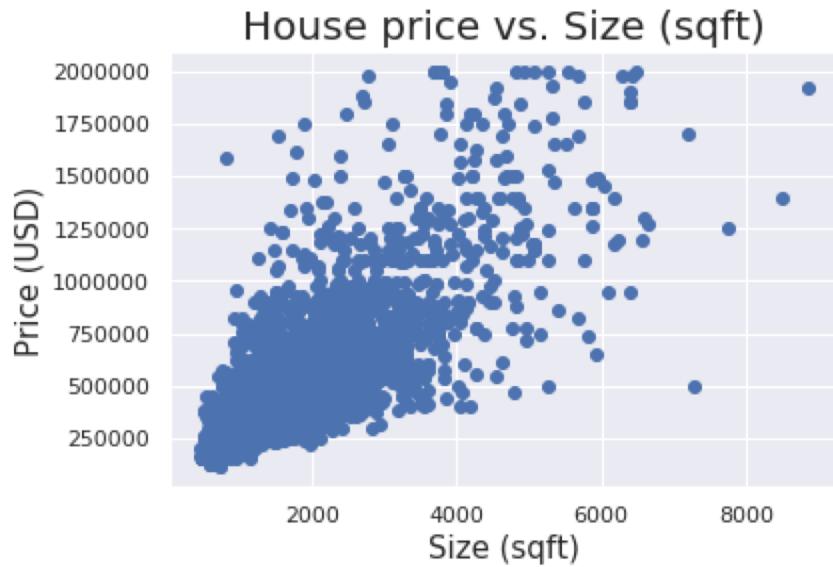
Categorical

Parking Spaces  
0 (on / off street), 1, 2, 3+

Type of house:  
Single family, Condo,  
Townhouse, other

Type of cooling:  
Central, non central

# Taking a look at the data

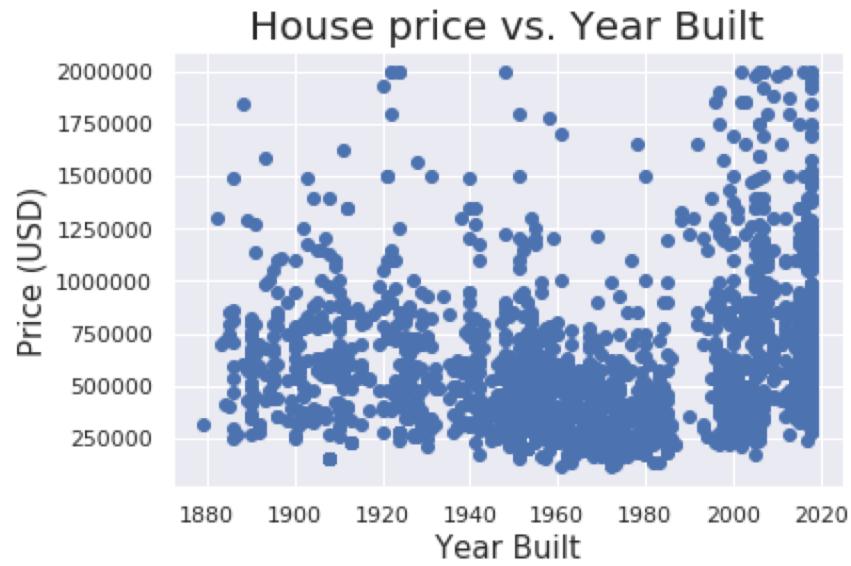


Linear relationship

# Taking a look at the data



Linear relationship



Polynomial relationship

# Modeling

## Linear

R2: 44% (cv=5)

- Lack of interactions
- Not capturing the polynomial relationships

# Modeling

Linear

Polynomial

R2: 44% (cv=5)

- Lack of interactions
- Not capturing the polynomial relationships

R2: 79% (cv=5)

- Too complex (>100 variables)
- Slightly over predicting
- Multicollinearity

# Modeling

Linear

R2: 44% (cv=5)

- Lack of interactions
- Not capturing the polynomial relationships

Polynomial

R2: 79% (cv=5)

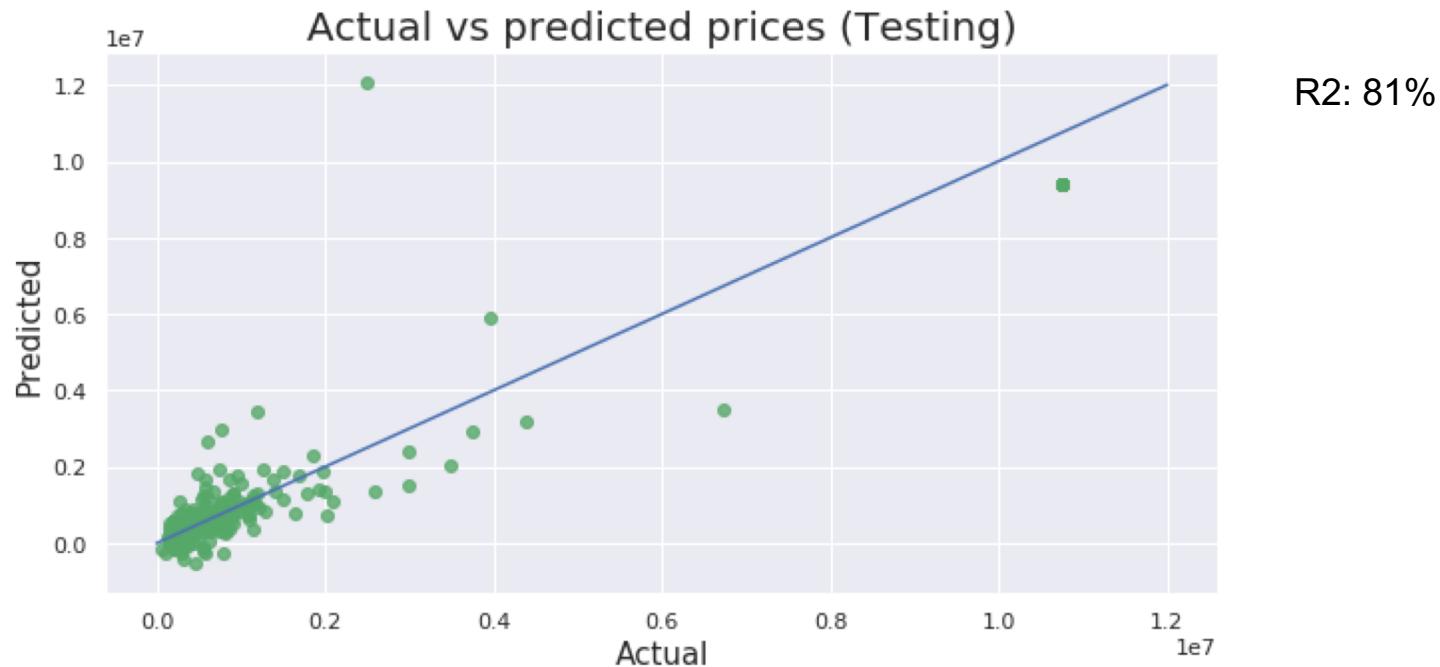
- Too complex (>100 variables)
- Slightly over predicting
- Multicollinearity

Lasso Regularization

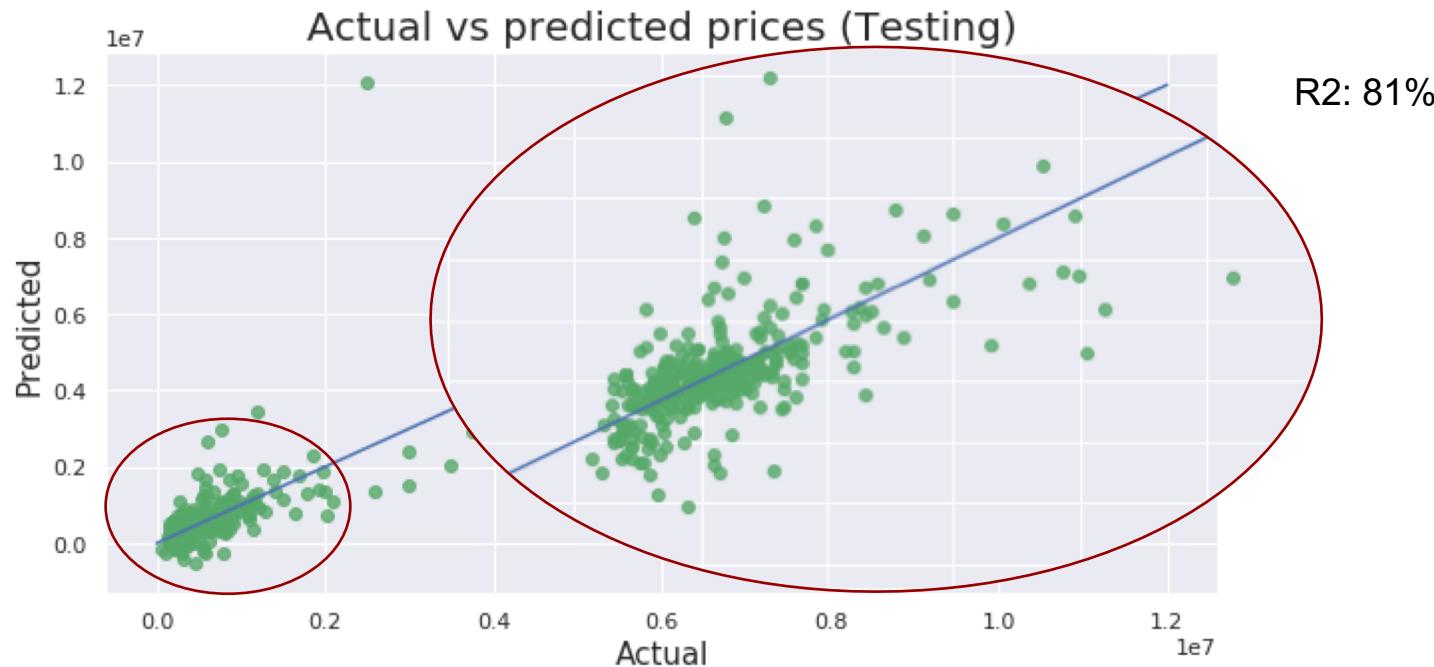
R2: 79% (cv=5)

- Simpler (46 significant coefficients)
- Still over predicting a little

# Performance



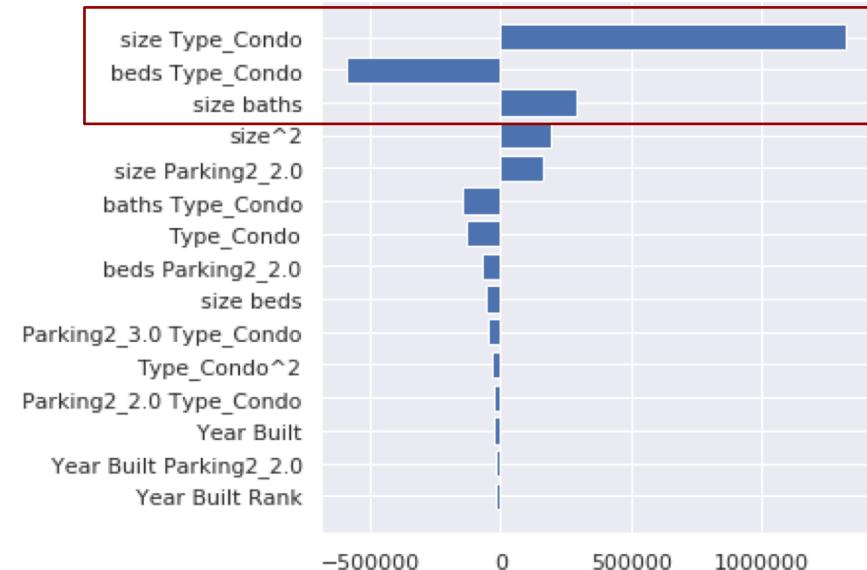
# Performance



# Conclusions

Interactions matter the most

Top 15 relevant variables

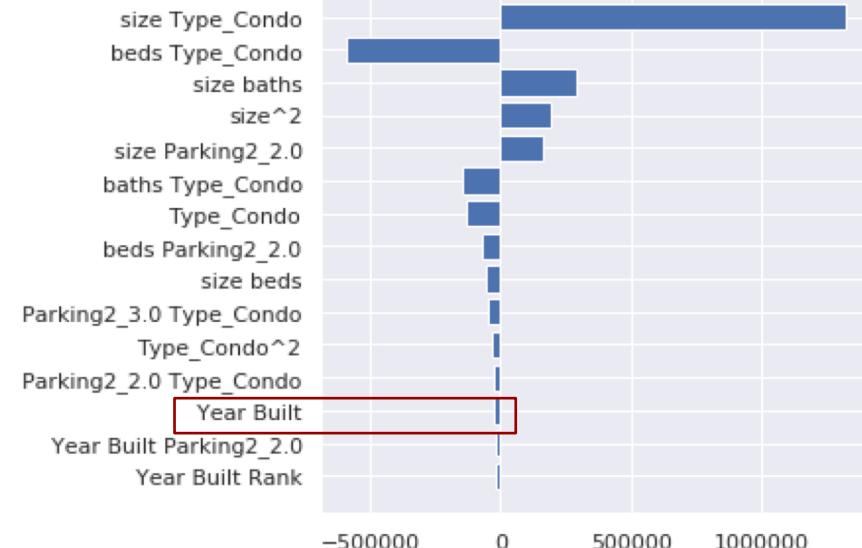


# Conclusions

Interactions matter the most

The newer the house the cheaper??

Top 15 relevant variables

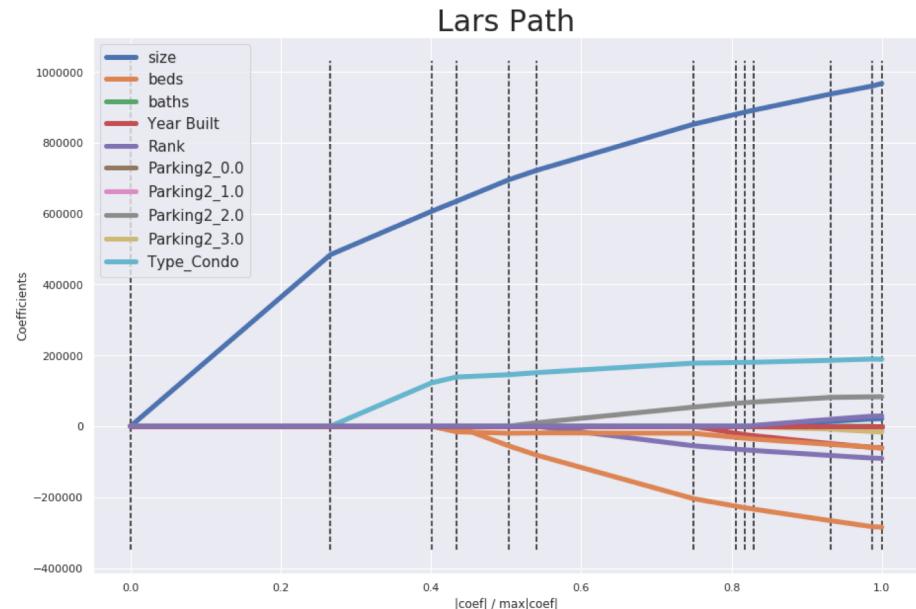


# Conclusions

Interactions matter the most

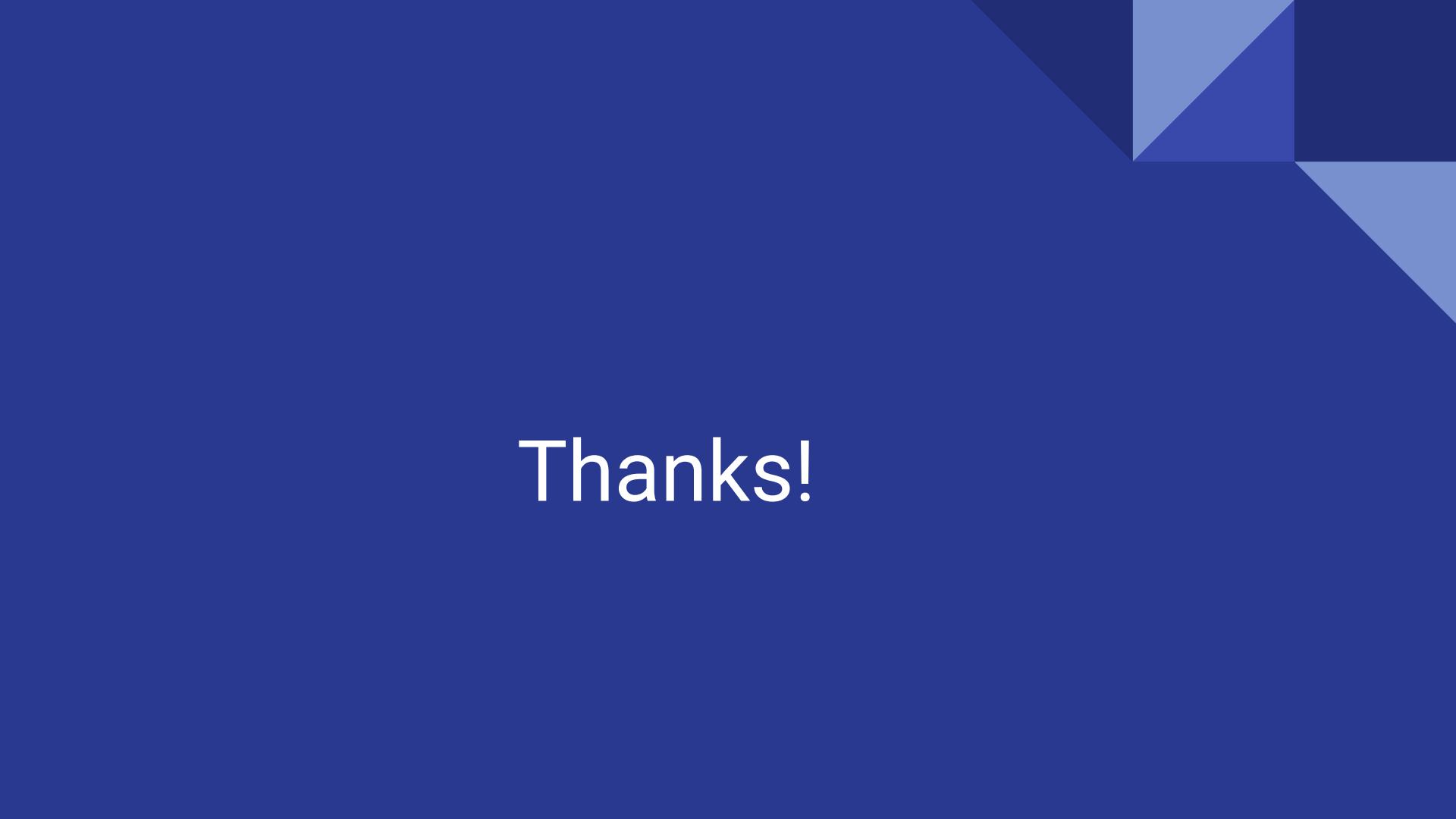
The newer the house the cheaper??

Size and Condo are the most relevant





# Questions?



# Thanks!