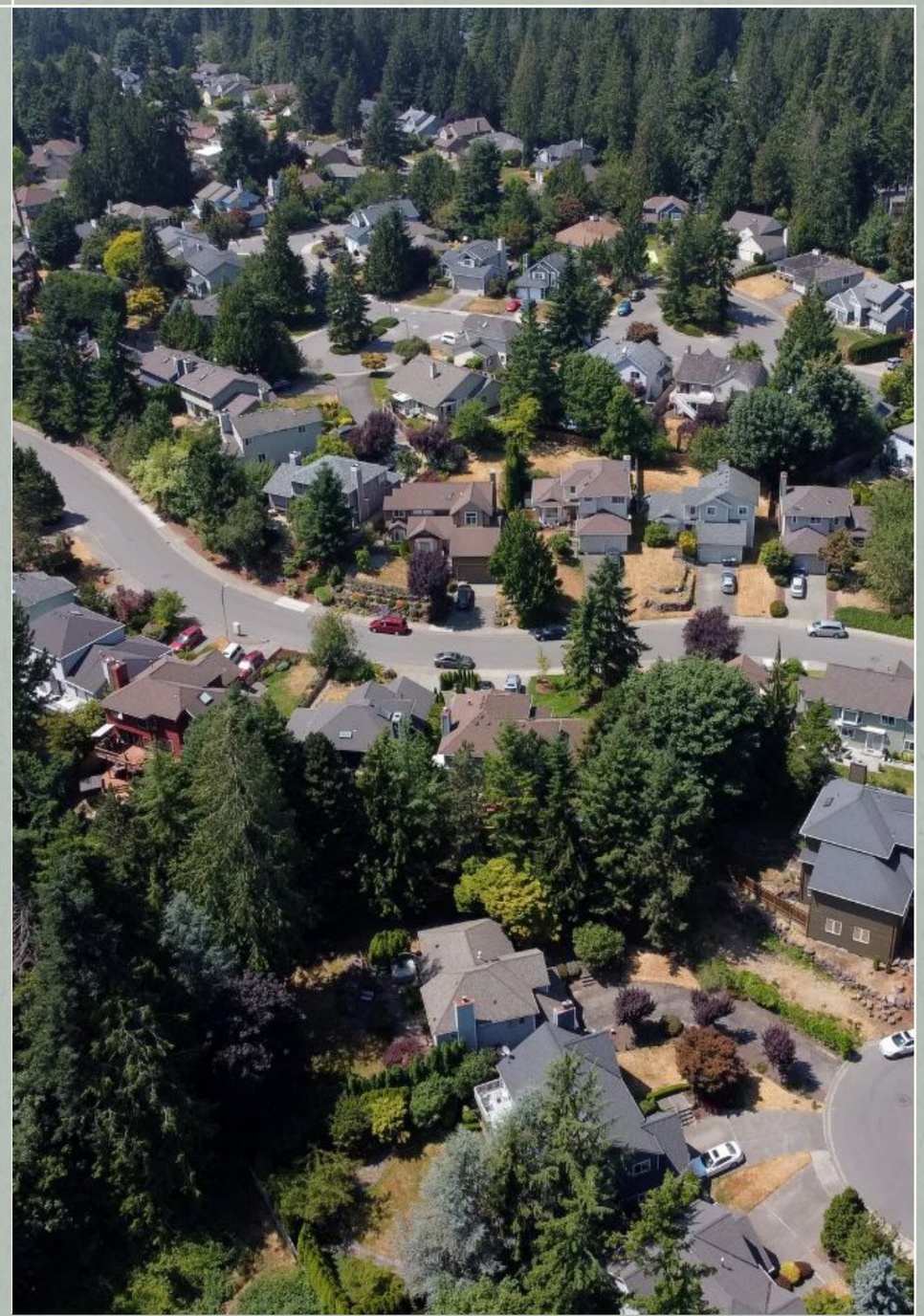
KING COUNTY REAL ESTATE

A Presentation by: Albane Colmenares

09/06/2023



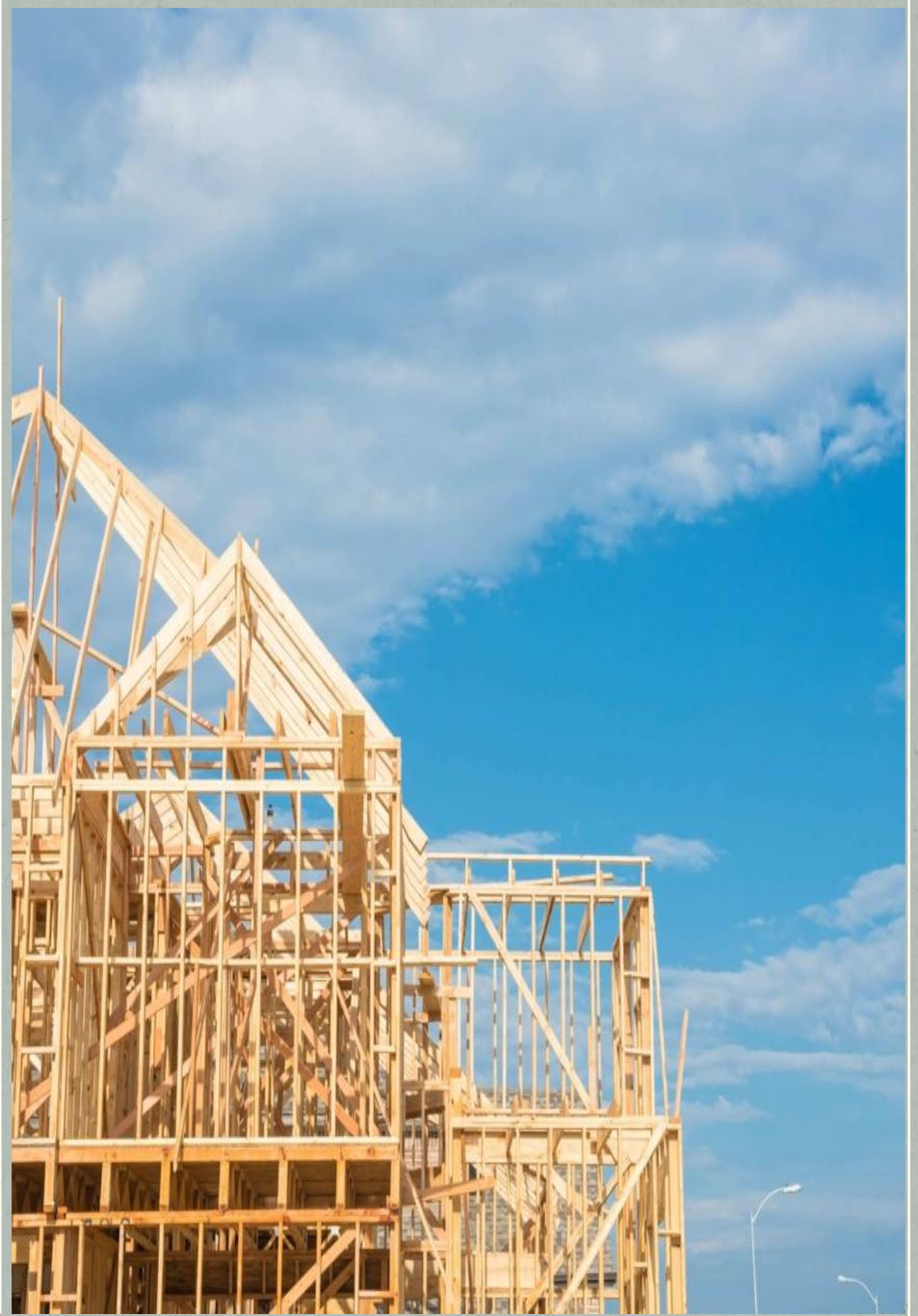
AGENDA

1 BUSINESS UNDERSTANDING

2 MODELING & REGRESSION RESULTS

3 RECOMMENDATIONS & NEXT STEPS

BUSINESS UNDERSTANDING

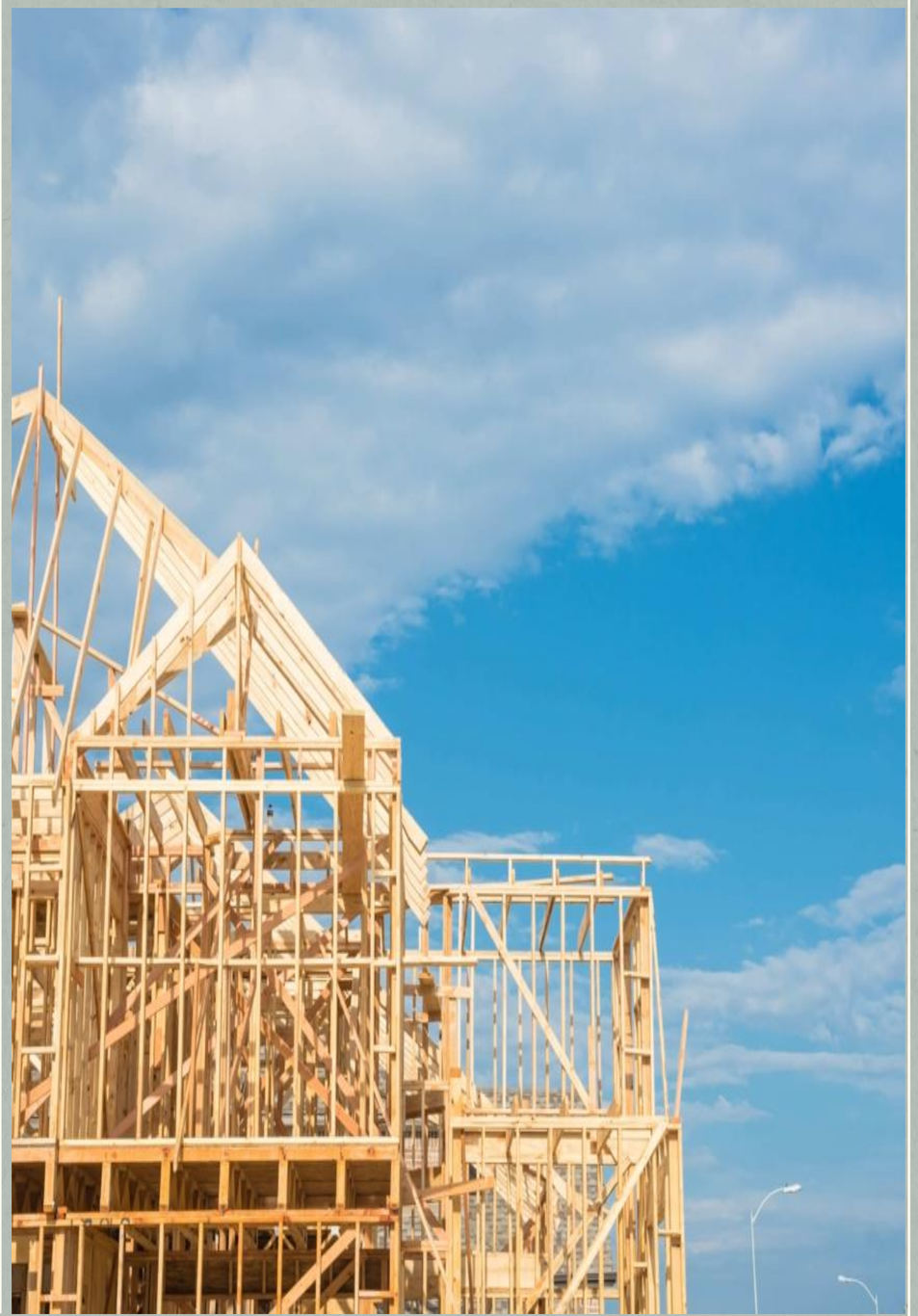


Home Owners	Selling Your House	King County, WA

RENOVATIONS:

HOW WILL THEY IMPACT YOUR SELLING PRICE?

MODELING



MODELING

1. Simple Linear Regression

A. Highest correlation with price
B. Initial simple linear regression
C. Exclusion of outliers
D. Improved linear regression

2. Multiple Linear Regression

A. Categorical feature
B. Condition & square footage of living area

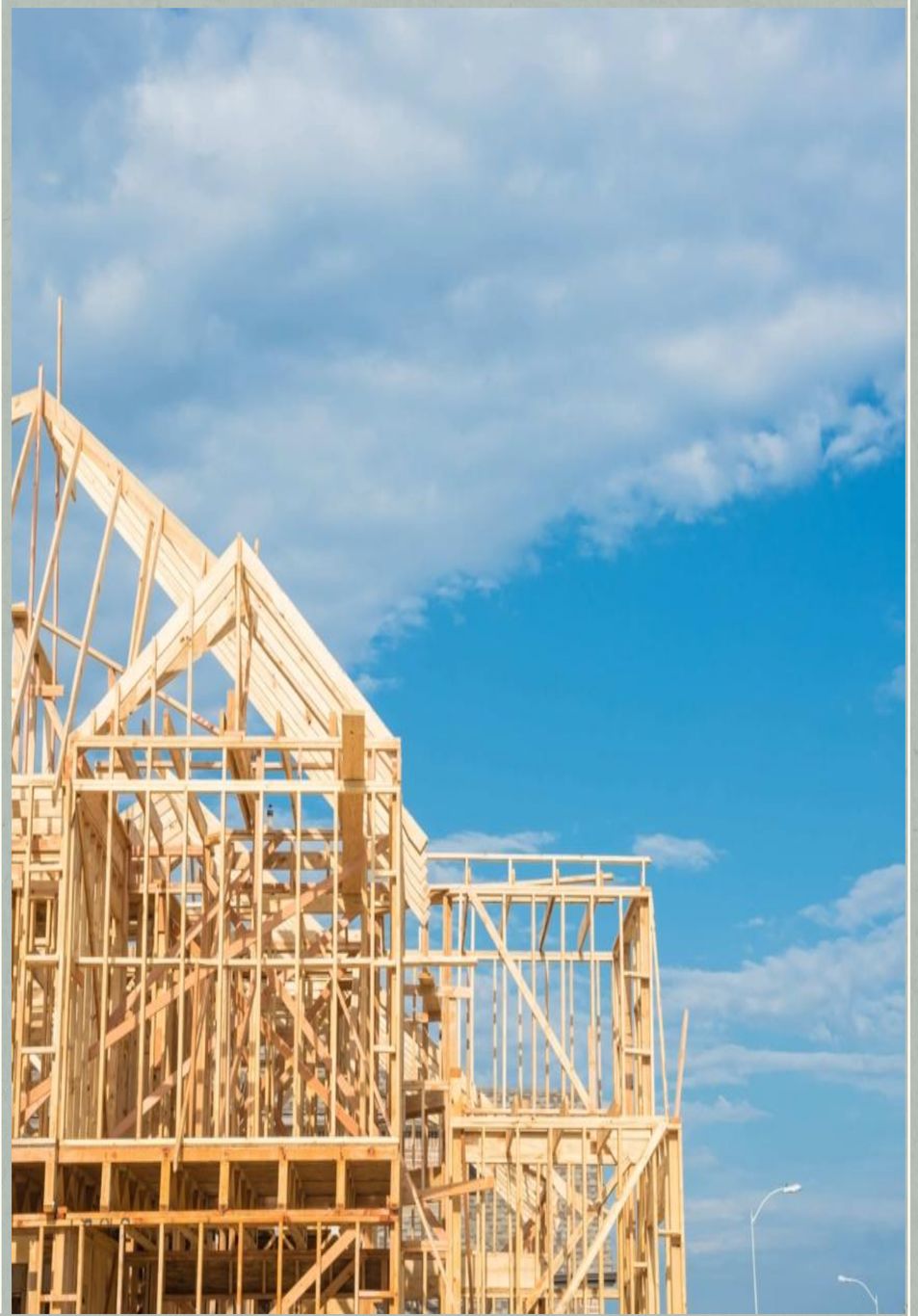
3. Multiple Linear Regression

A. Numeric features
B. Part 1: all variables

4. Multiple Linear Regression

A. Part 2: exclusion of lot square footage
B. Final model

REGRESSION RESULTS



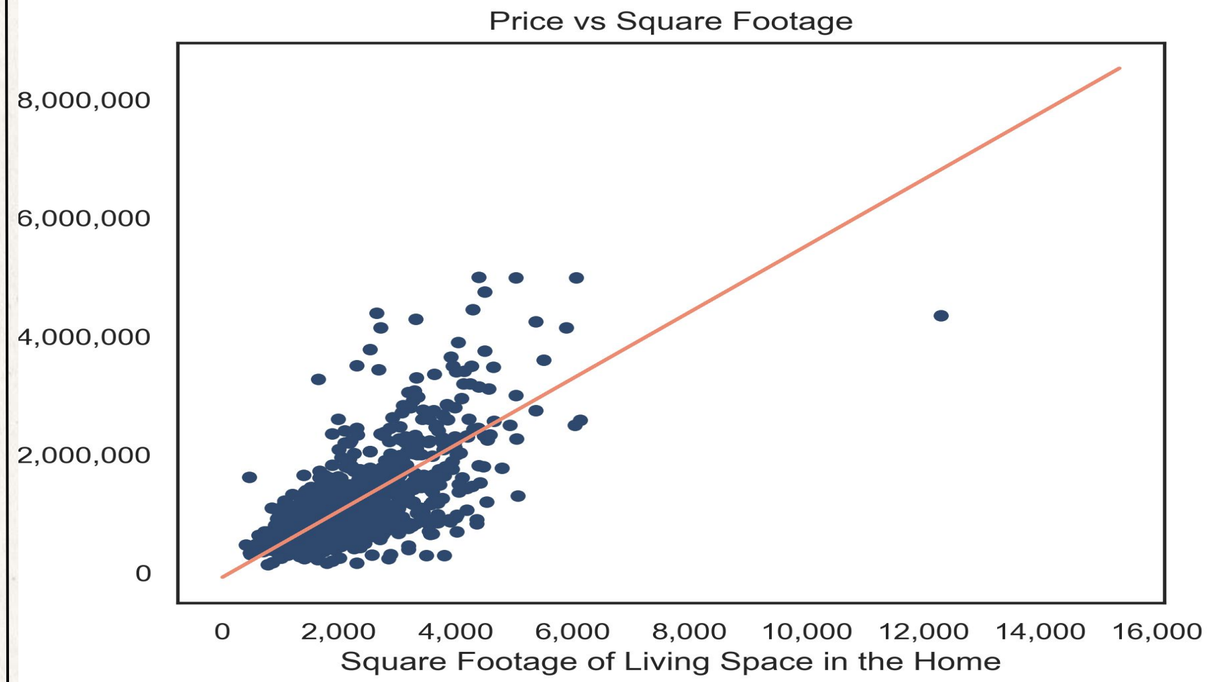
SIMPLE LINEAR REGRESSION

HIGHEST CORRELATION WITH PRICE

Square footage of living space

Results

1. Identification of predictor with highest correlation: **square footage of living space**
2. Ensuring accuracy by **excluding outliers**
3. Decent model explanatory power: **42.8%**
   Adjusted R-squared
4. Estimated price for 0 sq. ft house: **$77,330**
   *Constant coefficient or **intercept***



MULTIPLE LINEAR REGRESSION

1- CATEGORICAL FEATURE & 2- NUMERIC FEATURES

- Overall **condition** of the house
- Square footage of living space

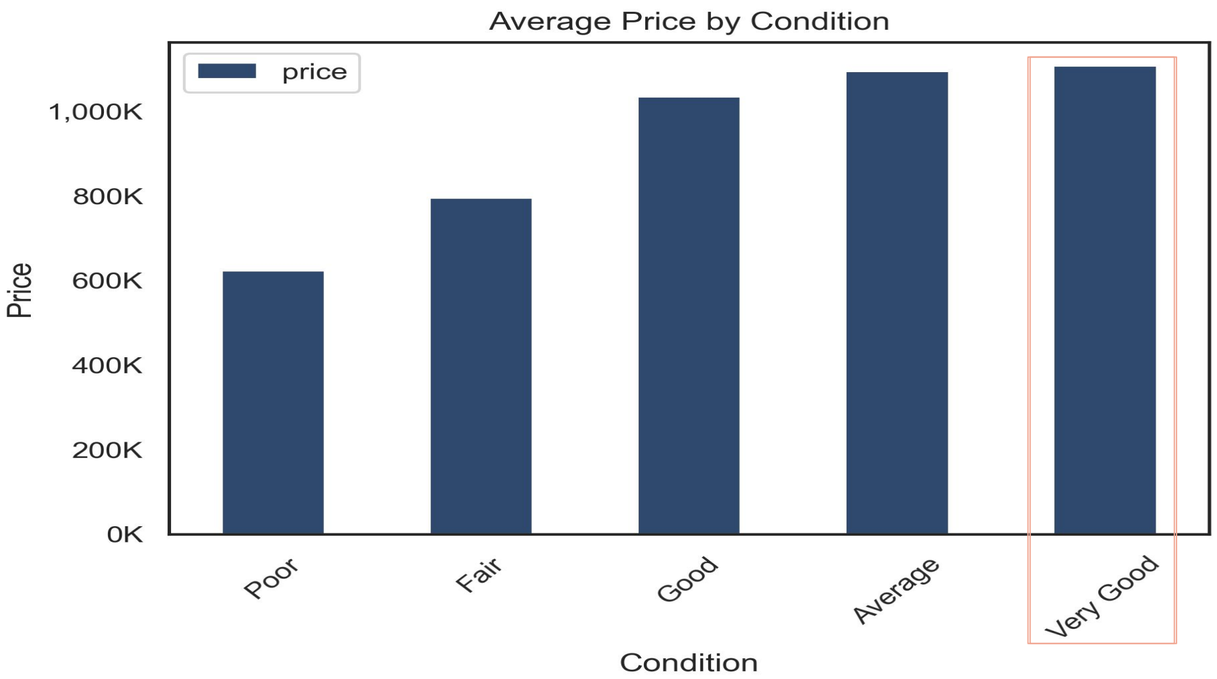
- **Square footage living area**
- Square footage patio
- Square footage above
- Square footage basement
- Square footage garage
- ~~Square footage lot~~
- Year built
- Floors
- **Bedrooms**
- **Bathrooms**

<u>Categorical</u>

1. **42.9%** of price's **variance** explained by categorical features
   Adjusted R-squared
2. "**Very good**" condition associated with **$135,700** increase on price + **statistically significant** impact on dependent variable
   Coefficient & p-values
3. Significant relationship with price
   F-statistic: 4478, Prob (F-statistic): 0

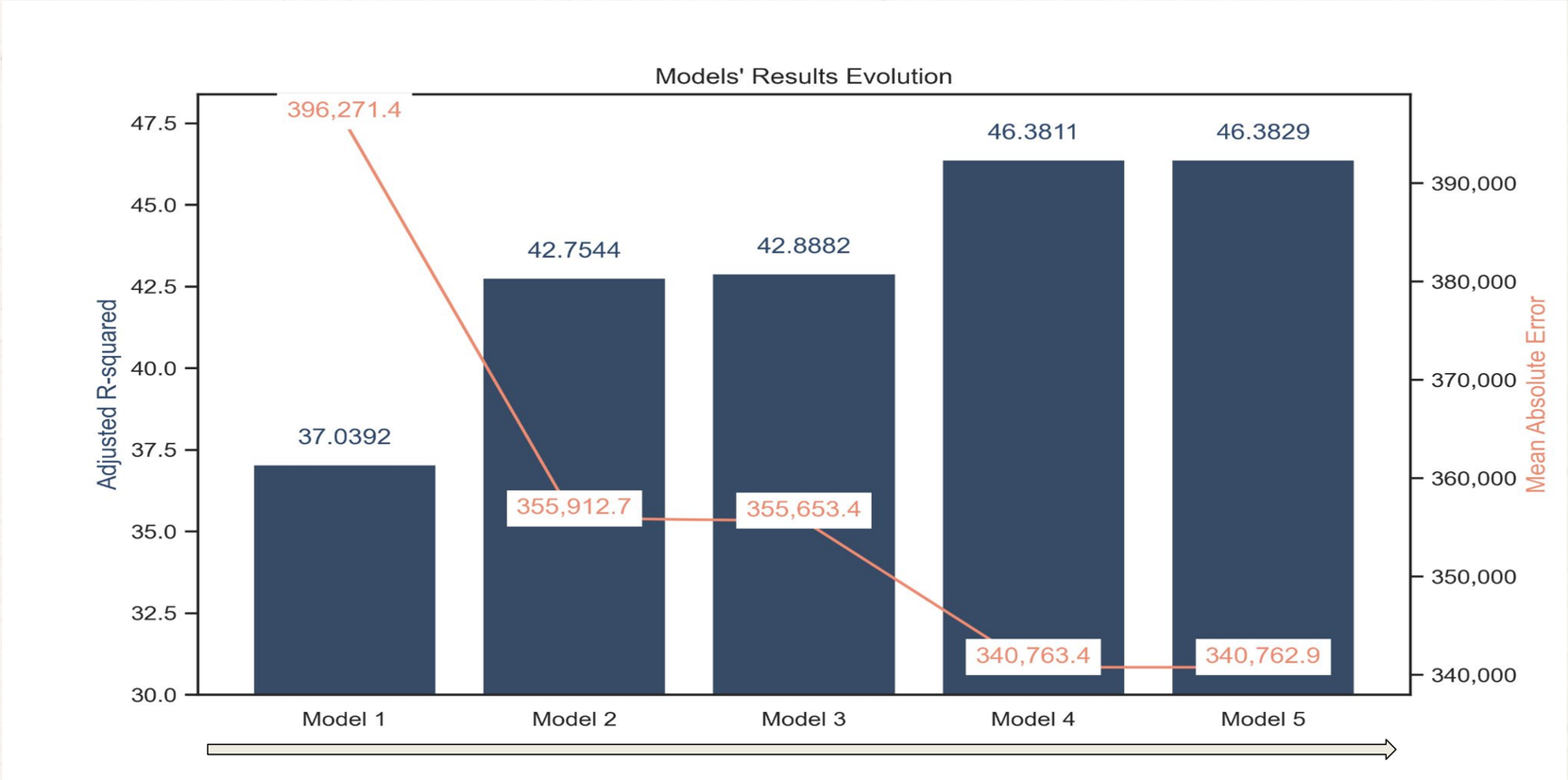
<u>Numeric</u>

1. **46.4%** of price's variance explained by **numeric features**
   Adjusted R-squared
2. Highest increase in price:
   a. **bathrooms**: **$116,200**
   b. **floors**: **$33,900**
   c. square footage **living area**: $303.3 → **$30,330** for 100 sq. feet
   Coefficient & p-values
3. Significant impact on price: F-statistic increased (2,866 vs 2,579 previously)



MODELS' RESULTS

- Two statistics measuring:
  - model's explanatory power → **Adjusted R-squared**
  - absolute differences between predictions & actual values → **Mean Absolute Error**
- Models' predictions more accurate:
  - Adj. R-squared **increasing from 37.0** (first model) **to 46.4** (last model)
  - **MAE decreasing** from 396,271.4 to 340,762.9
  - MAE seems like a high value, but acceptable for large price range ($131K to $5.6M)

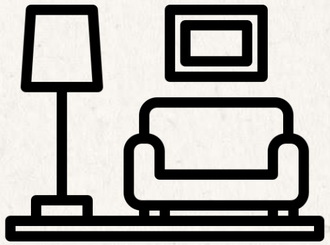


RECOMMENDATIONS

RECOMMENDATIONS

Very Good Condition

Aim for a Very Good Condition.
Associated with an increase of $135,700 in price

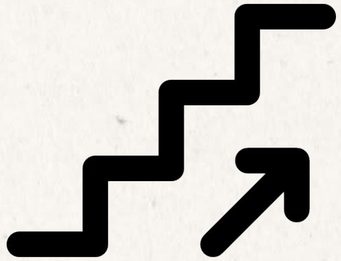


Living Area Square Footage

Increase the size of your living space.
Predicted to increase +$303.26 per square foot.

Bathroom

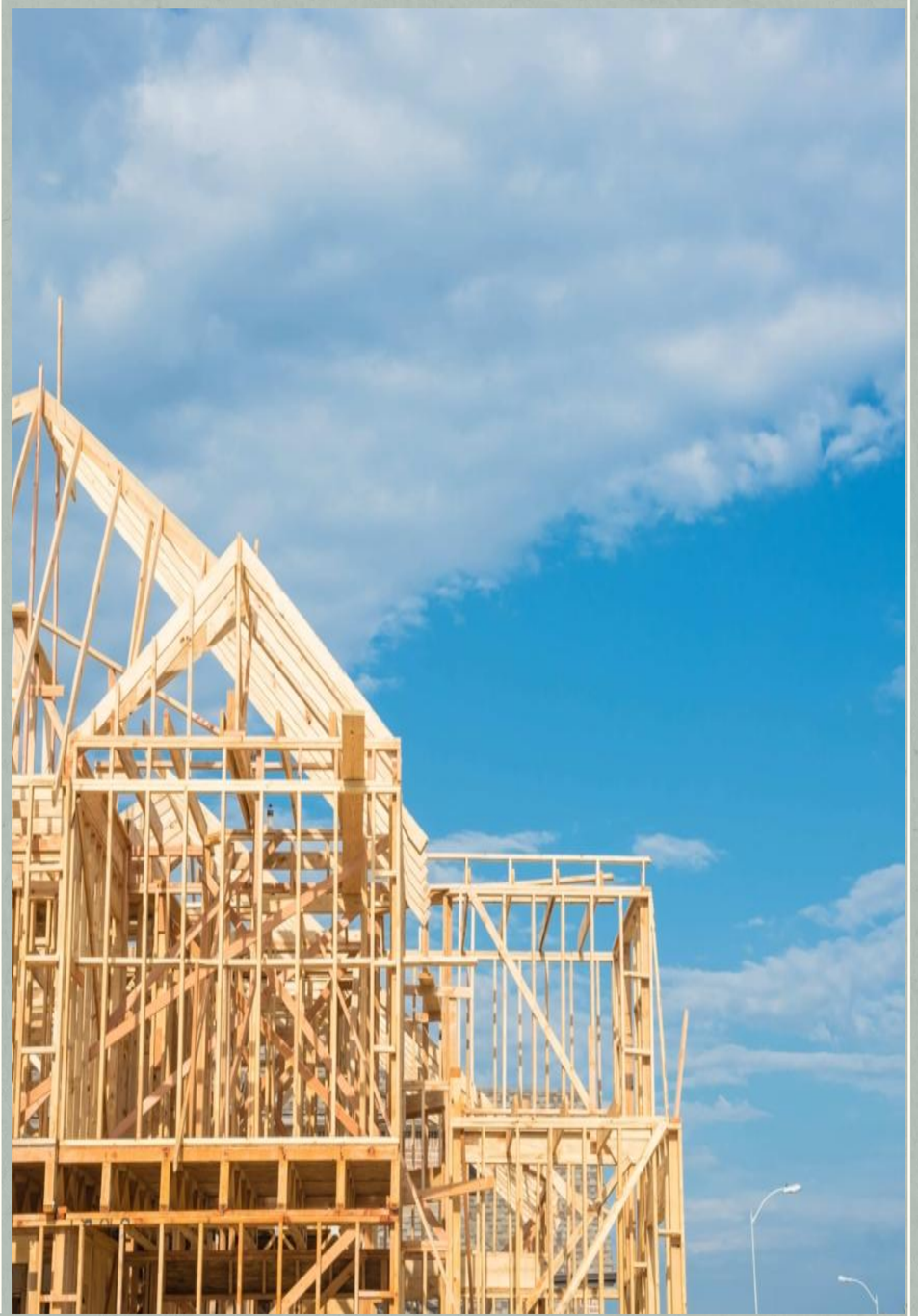
Add one bathroom if your house permits it.
Associated with $116,200 in price increase



Floor

Add one floor. Consider a mezzanine. Estimated increase in price: $33,900

LIMITS & NEXT STEPS



LIMITS
&
NEXT STEPS

LIMITS

Linear Regression Assumptions Threatened

- **Not a normal distribution**
- **Multicollinearity among independent variables**

NEXT STEPS

Linear Regression Assumptions Threatened

- **Normalizing distribution**
- **Scaling data to improve predictions and lower MAE**

CONTACT INFORMATION

For more details, contact:

albane.colmenares@gmail.com

THANK YOU