KING COUNTY REAL ESTATE

A Presentation by: Albane Colmenares   09/13/2023

AGENDA

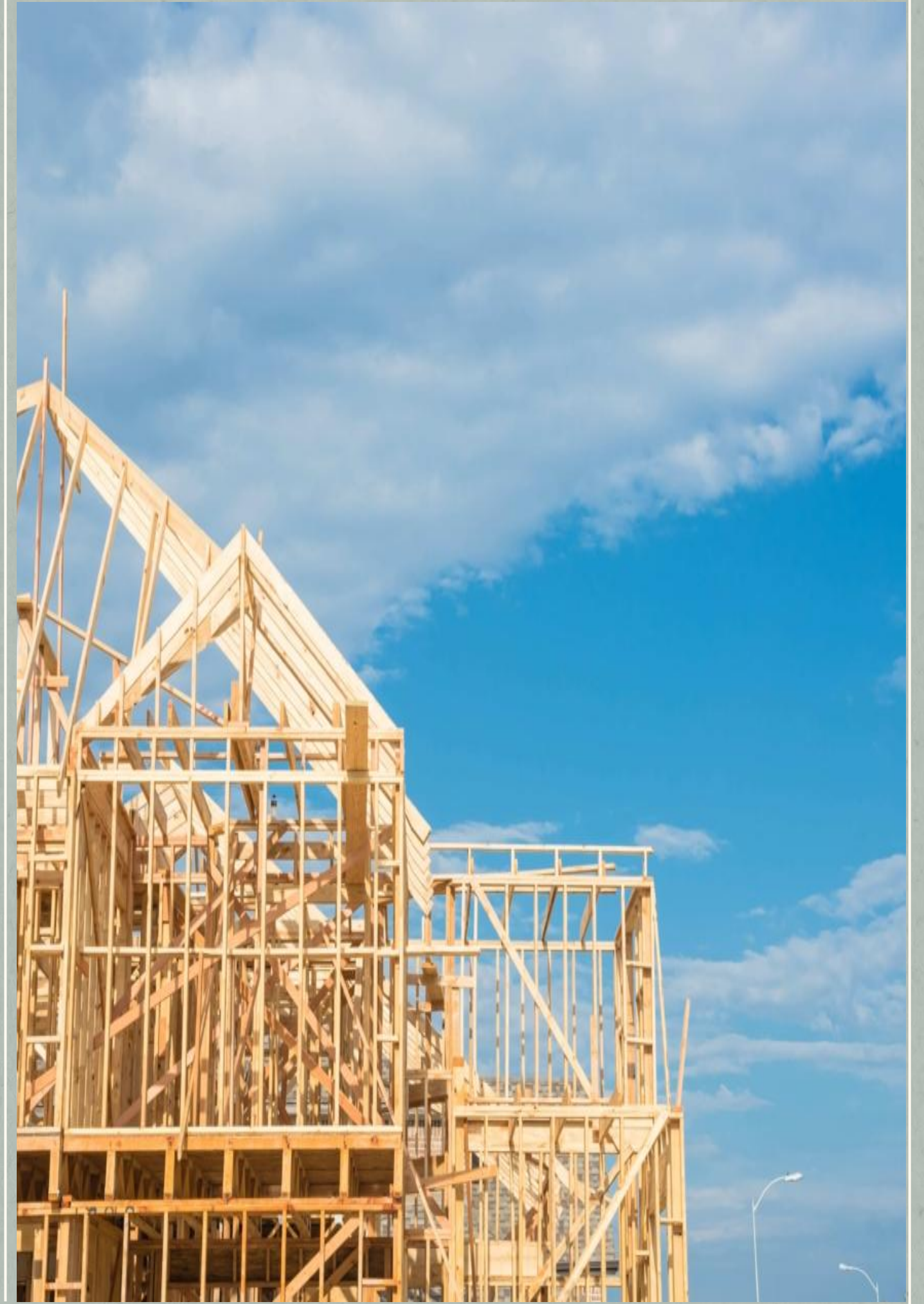**1** BUSINESS UNDERSTANDING

2 MODELING & REGRESSION RESULTS

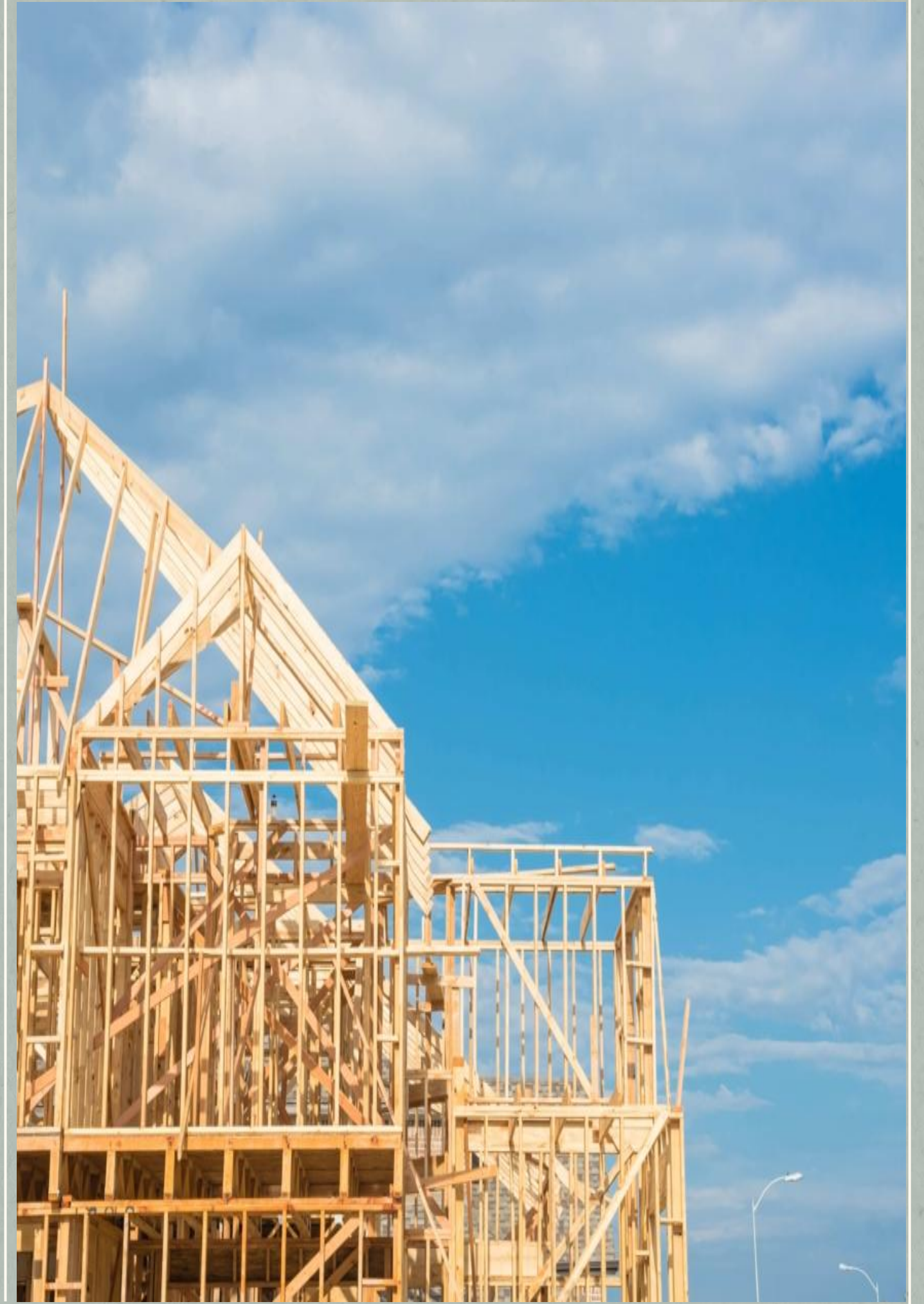**3** RECOMMENDATIONS & NEXT STEPS

BUSINESS UNDERSTANDING

HOW HOMEOWNERS CAN GAIN FROM RENOVATIONS:

PREDICTING THEIR IMPACT ON SALE PRICE

MODELING

MODELING

1. Simple Linear Regression

A. Highest <u>correlation</u> with price

B. Exclusion of <u>outliers</u>

2. Multiple Linear Regression

A. <u>Categorical</u> feature

B. Condition & square ft. of living area
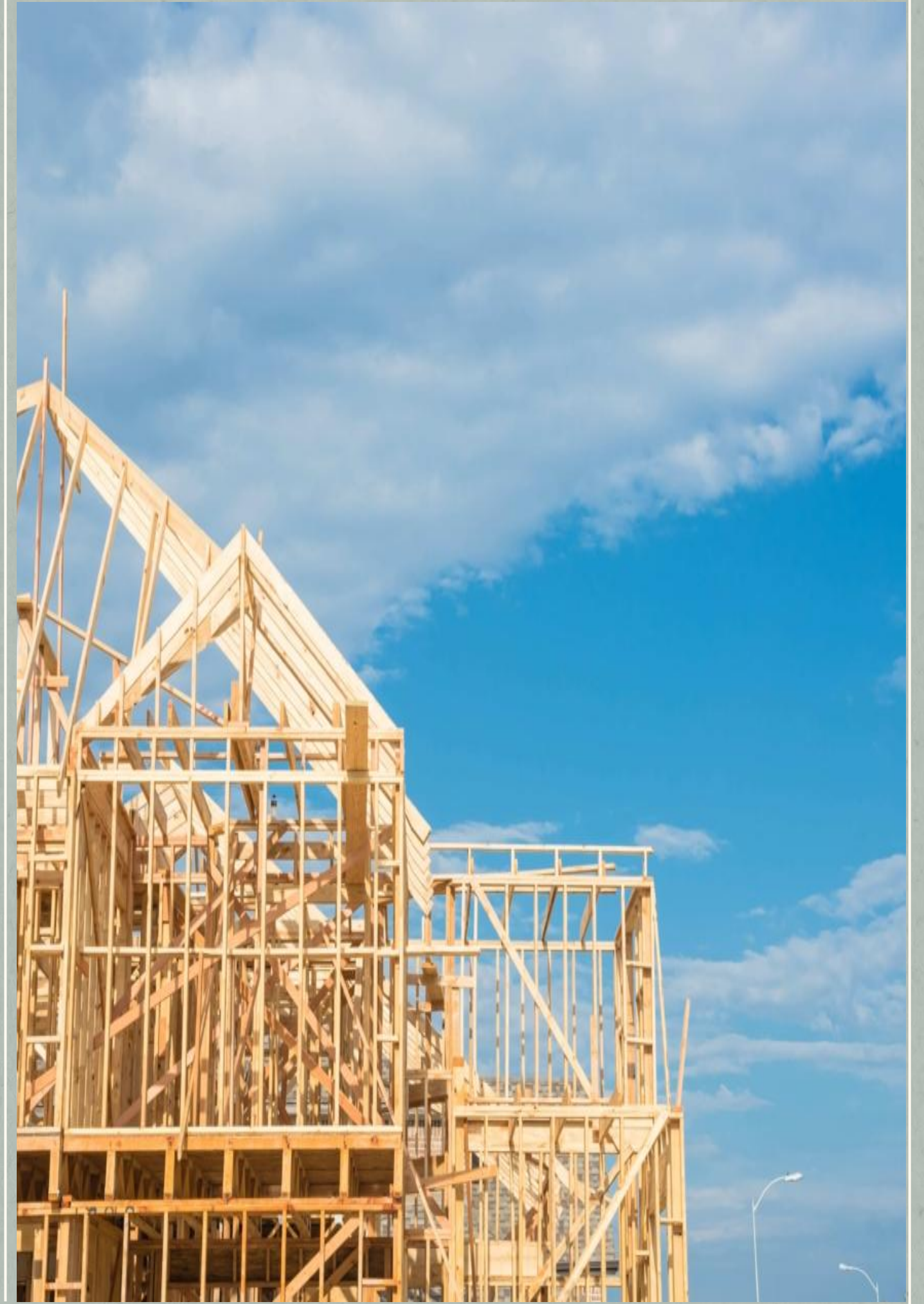
3. Multiple Linear Regression

A. Part 1: all variables

B. <u>Numeric</u> features

A. Part 2: exclusion of lot square ft.

B. Final model

REGRESSION RESULTS
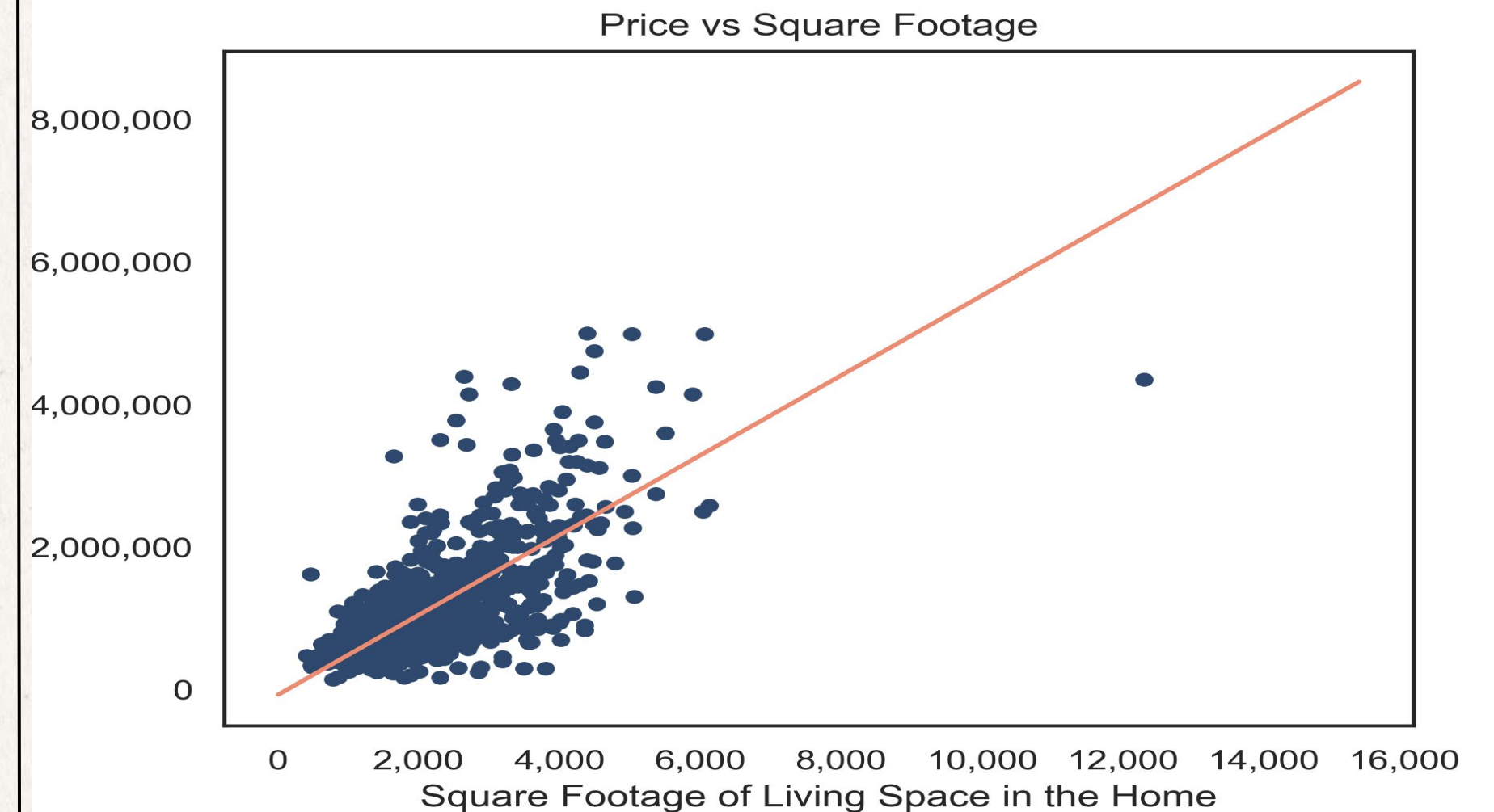
SIMPLE LINEAR REGRESSION

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

Price vs Square Footage

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
- **Bathrooms**
- Bedrooms
- **Floors**

Categorical

1. "**Very good**" condition associated with **$135,700** increase on price + **statistically significant** impact on dependent variable
Coefficient & p-values

2. **42.9%** of price's **variance** explained by categorical features
Adjusted R-squared

Numeric

1. Highest increase in price:
   a. **bathrooms**: **$116,200**
   b. **floors**: **$33,900**
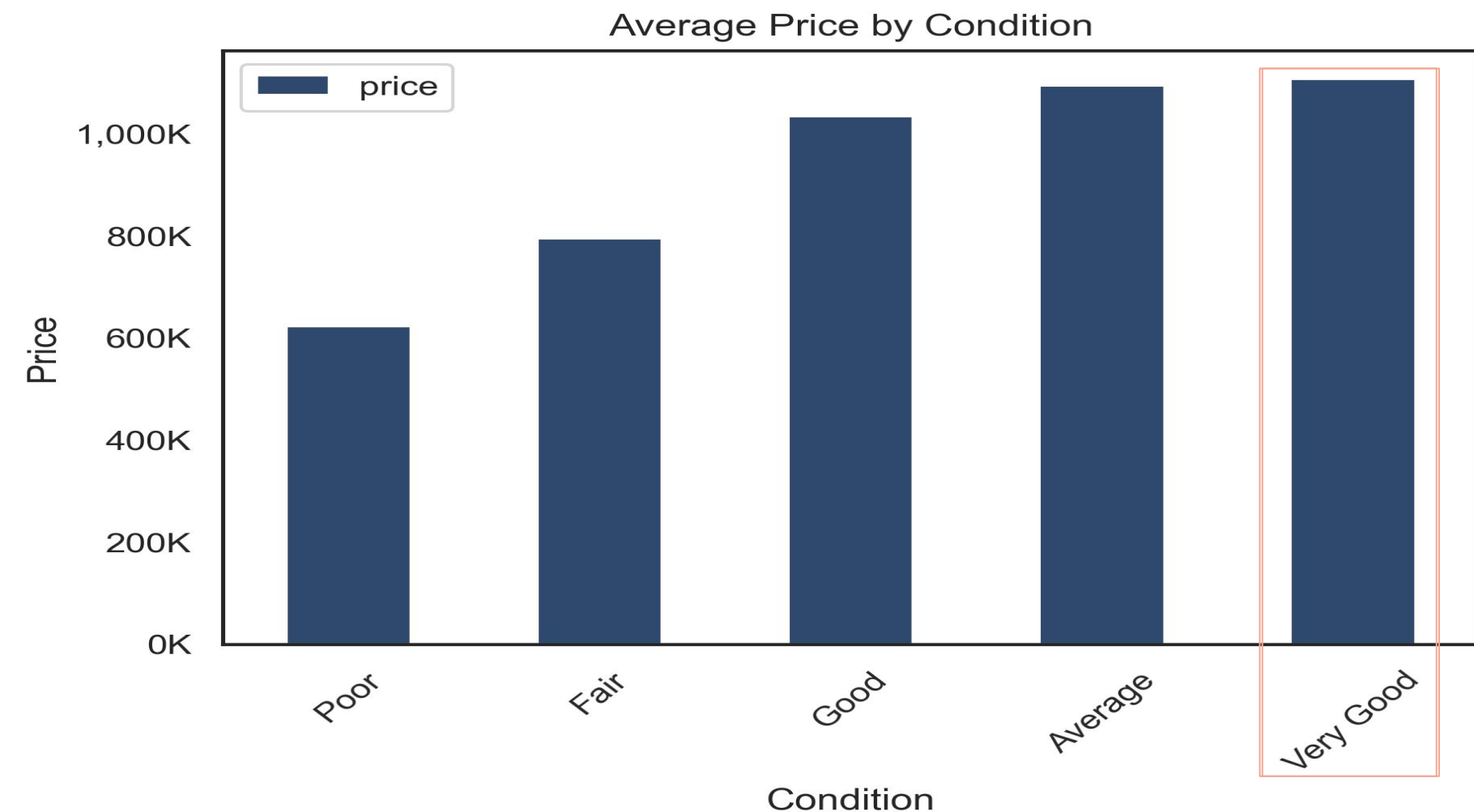   c. square footage **living area**: $303.3 → **$30,330** for 100 sq. feet
Coefficient & p-values

2. **46.4%** of price's variance explained by **numeric features**
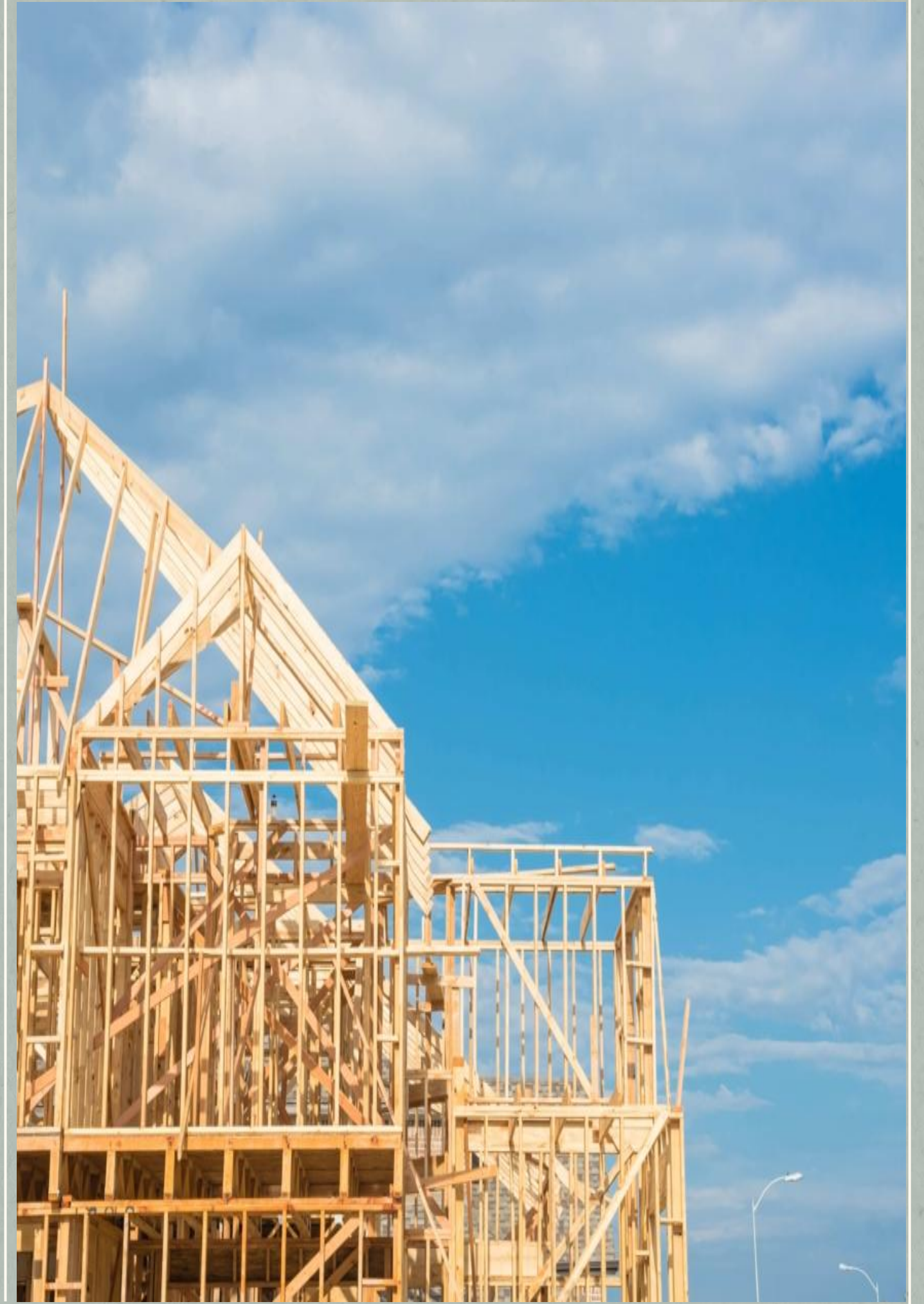*Adjusted R-squared*

Average Price by Condition

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
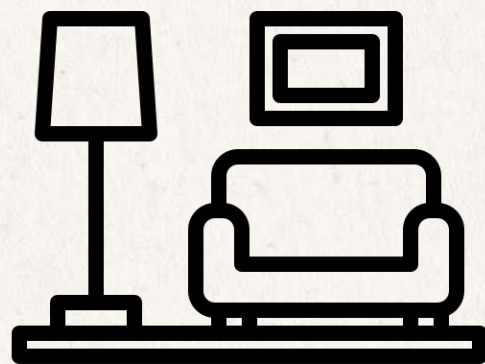   acceptable for large price range
    ($131K to $5.6M)

Models' Results Evolution
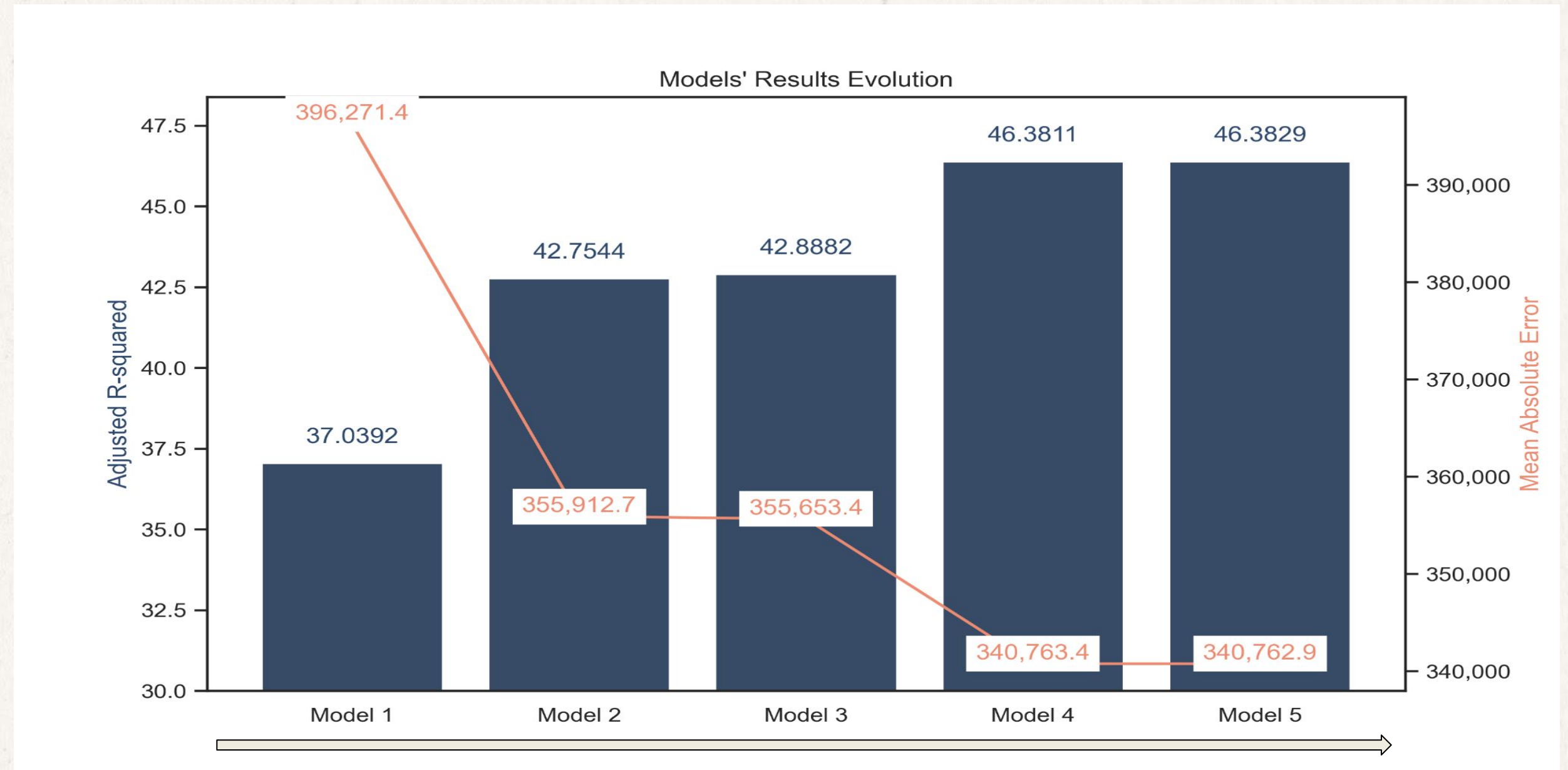
RECOMMENDATIONS

RECOMMENDATIONS

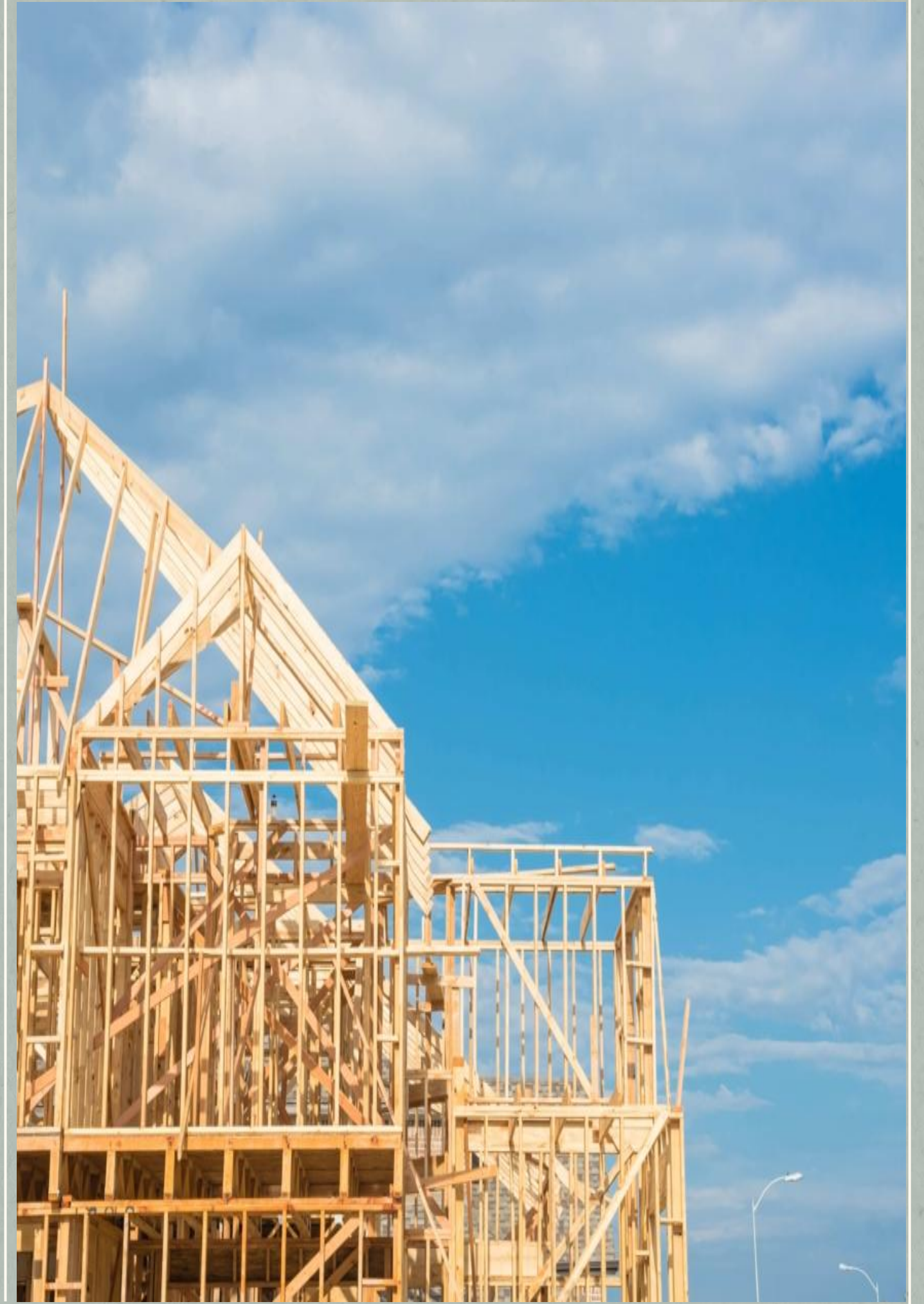| Very Good Condition	Living Area Square Footage	Bathroom	Floor
Aim for a "Very Good" condition	Increase the size of your living space	Add one bathroom if your house permits it.	Add one floor. Consider a mezzanine
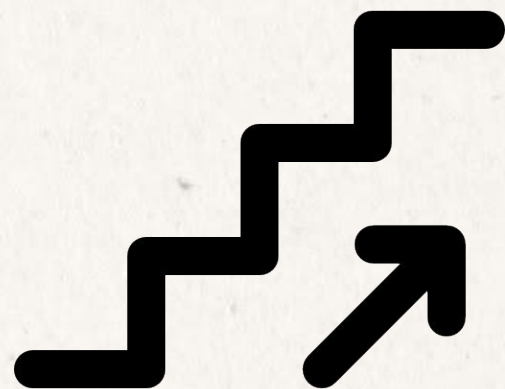
LIMITS &
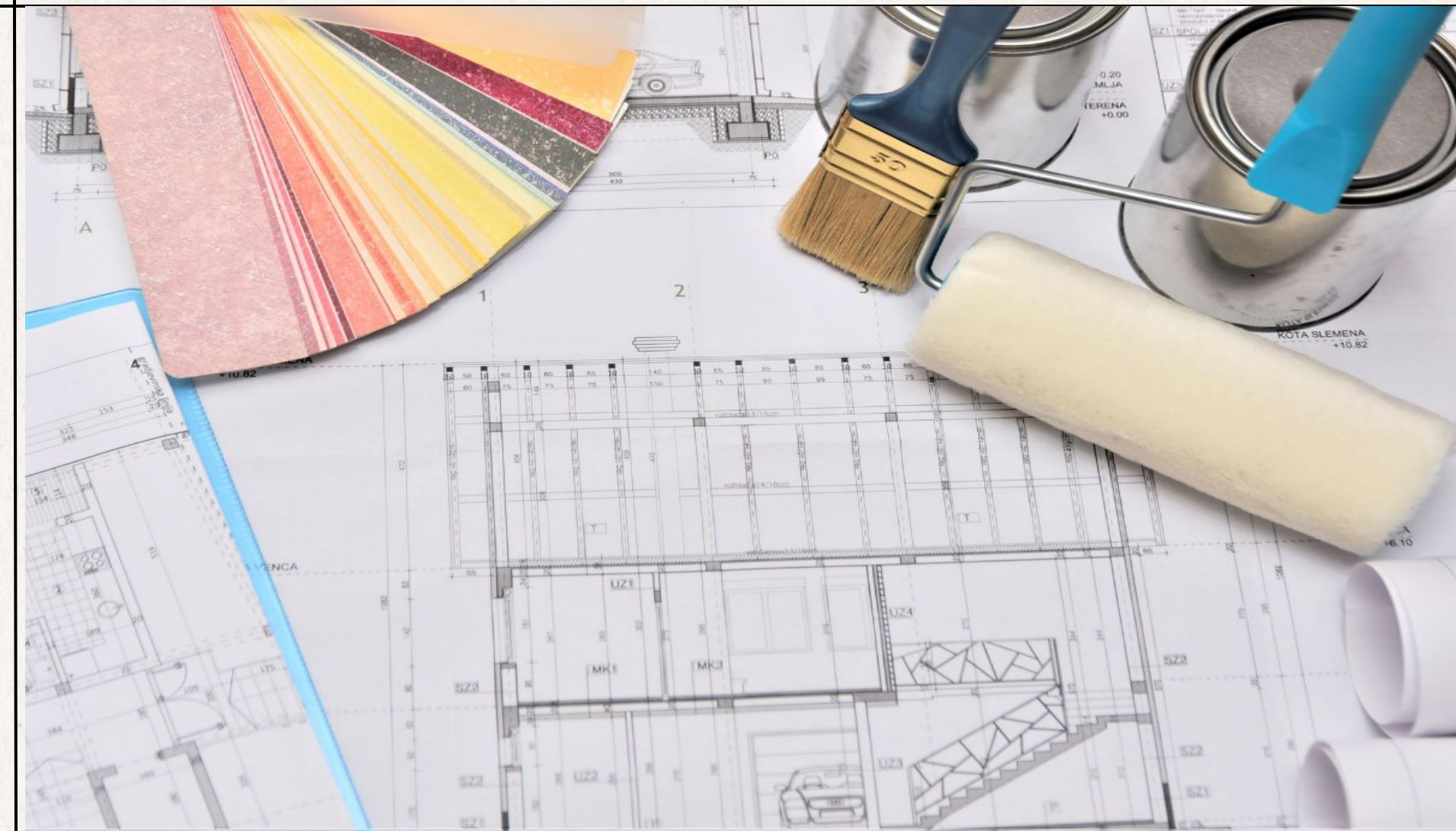# NEXT STEPS

LIMITS
&
NEXT STEPS

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