



# PROPERTY PRICE PREDICTION (REGRESSION MODELS)

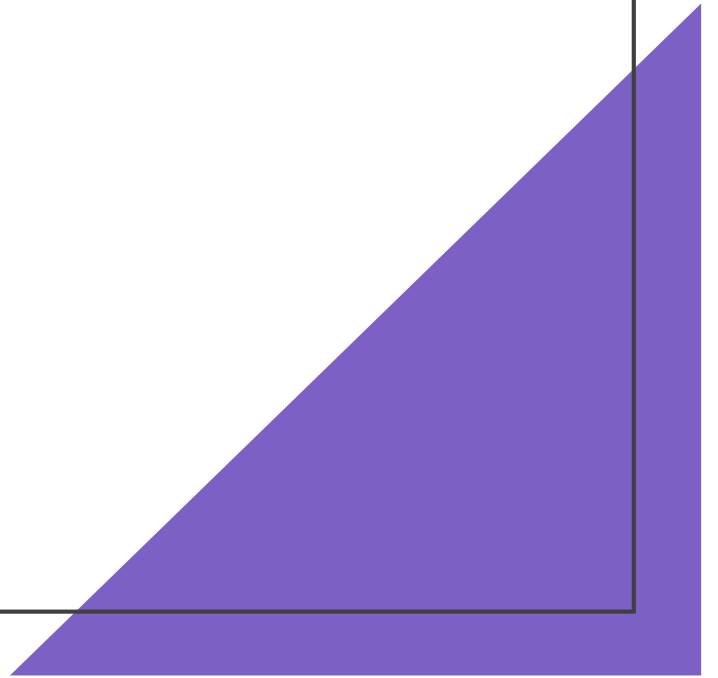


GROUP VII



# OUTLINE

- Project Overview
- Business Problem
- Data Understanding
- Data Cleaning and preparation
- Data Analysis
- Conclusion
- Recommendation



# PROJECT OVERVIEW

The project aims to support our stakeholder: **The National Association of Realtors** (NAR) with appropriate and comprehensive insights and information they can utilize to advice their clients: proprietors and homeowners about how different factors affect home sale prices in the county.





# BUSINESS PROBLEM

1. To determine Property Valuation by considering the impact of various property attributes
2. To identify the most influential features in determining property prices.
3. To evaluate potential real estate investment opportunities thus assessing profitability and potential ROI

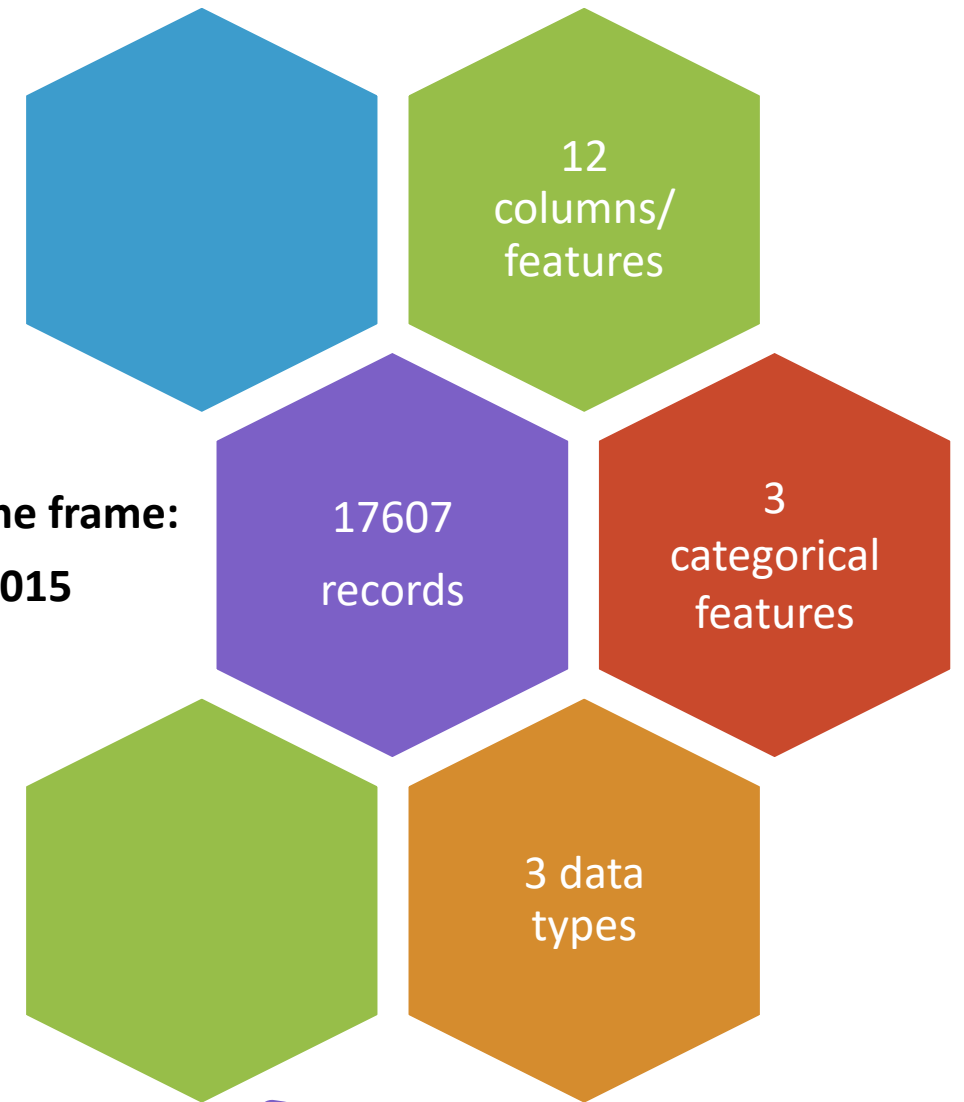
# DATA UNDERSTANDING

This project uses the north western county dataset.

It includes the below features:

- price
- bedrooms
- bathrooms
- sqft\_living
- zipcode
- yr\_built

**Data time frame:**  
**1900 - 2015**

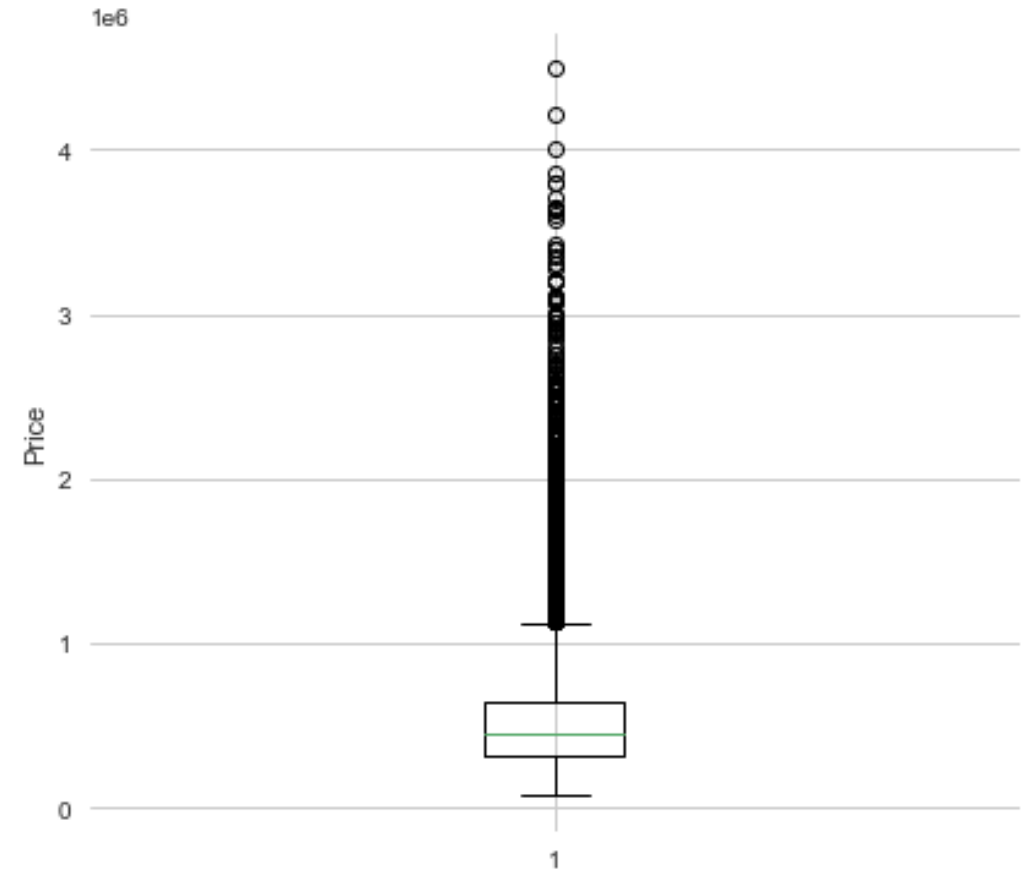
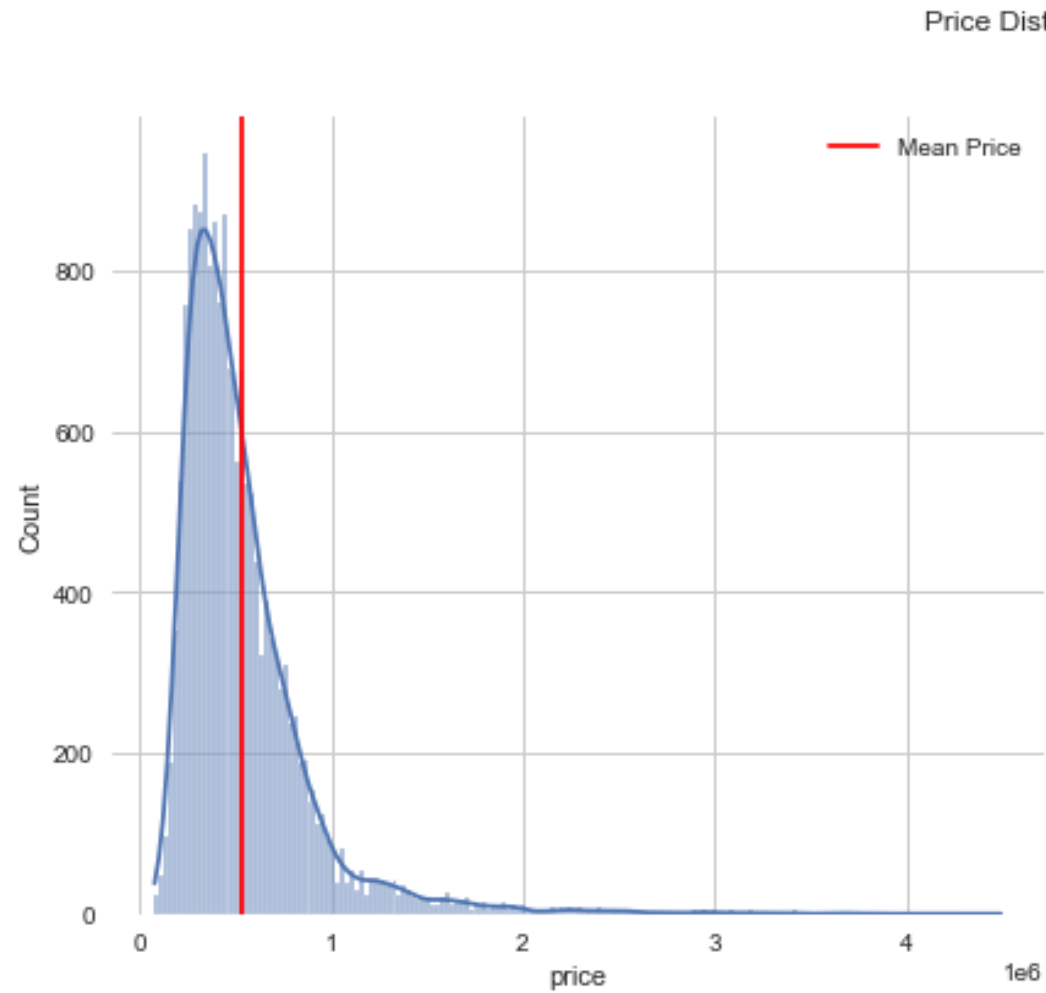


# DATA CLEANING & PREPARATION

- ☐ Check and drop duplicates in the 'id' column
- ☐ Identify and handle missing values
- ☐ Check for place holders in 'price' column
- ☐ Convert data date types if necessary
- ☐ Identify outliers
- ☐ Feature Engineering by creating new columns i.e 'is\_renovated'
- ☐ Dropping irrelevant columns

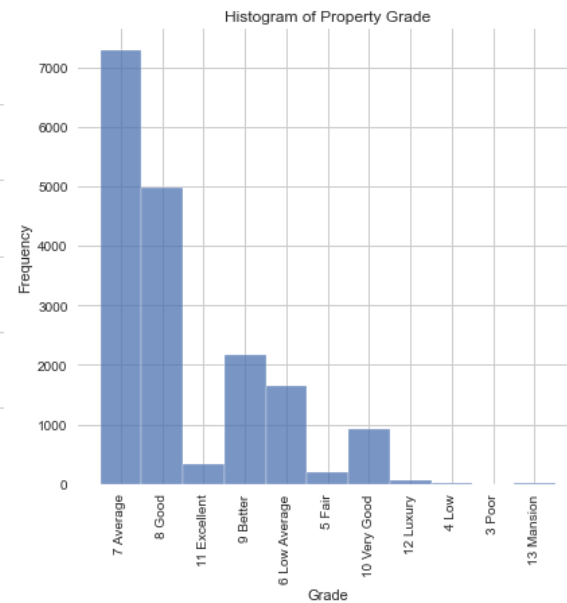
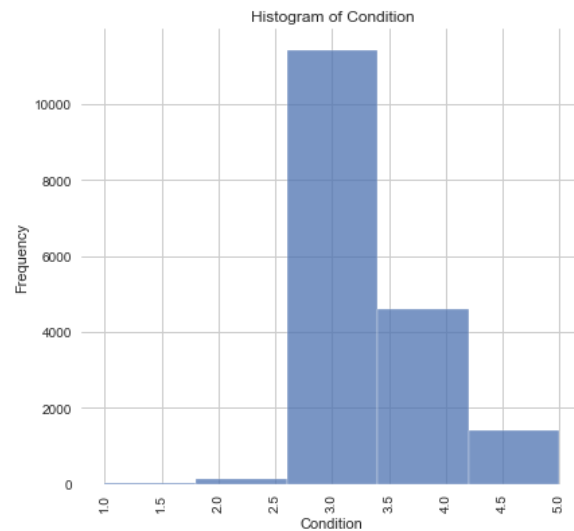
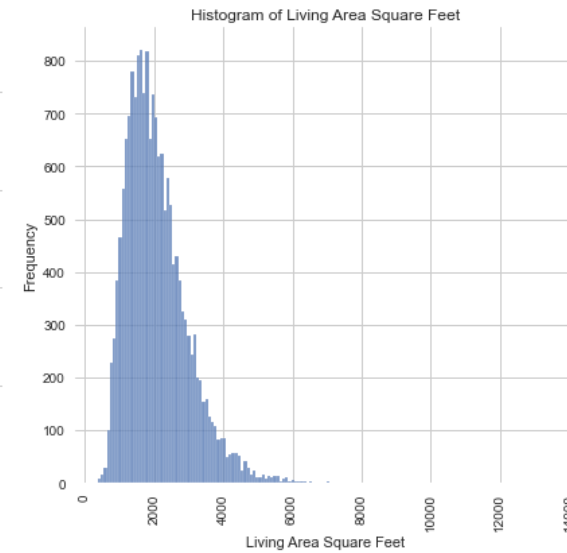
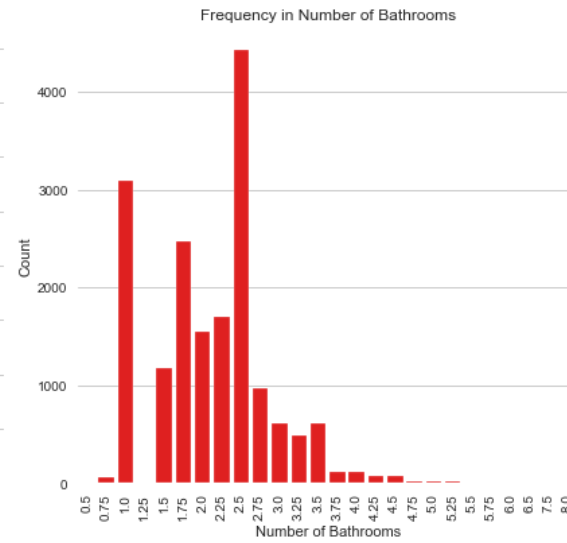
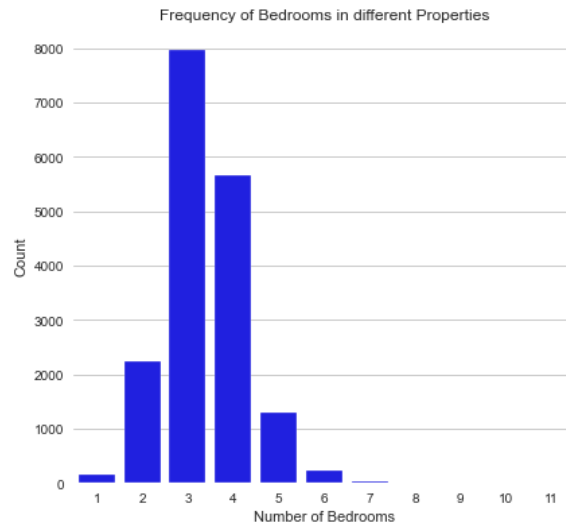


# PRICE DISTRIBUTION

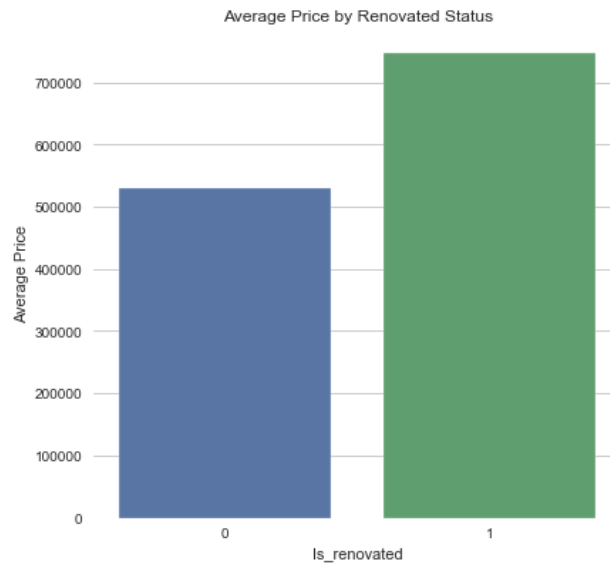
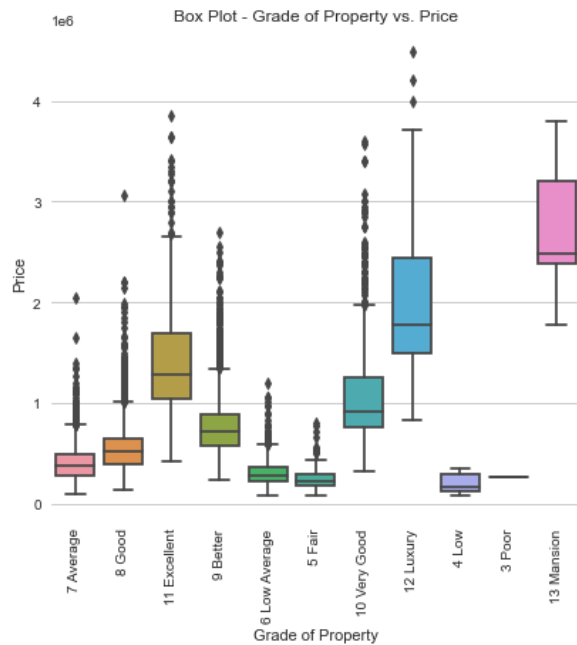
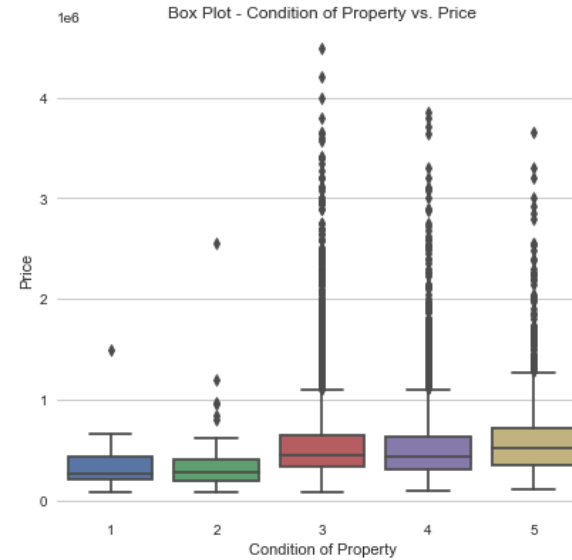
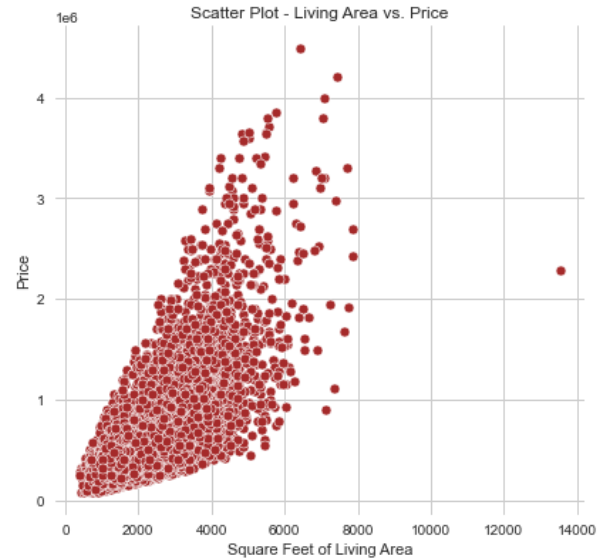
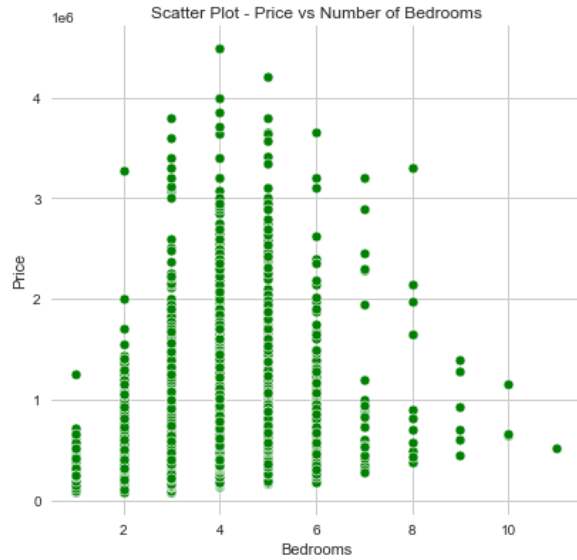


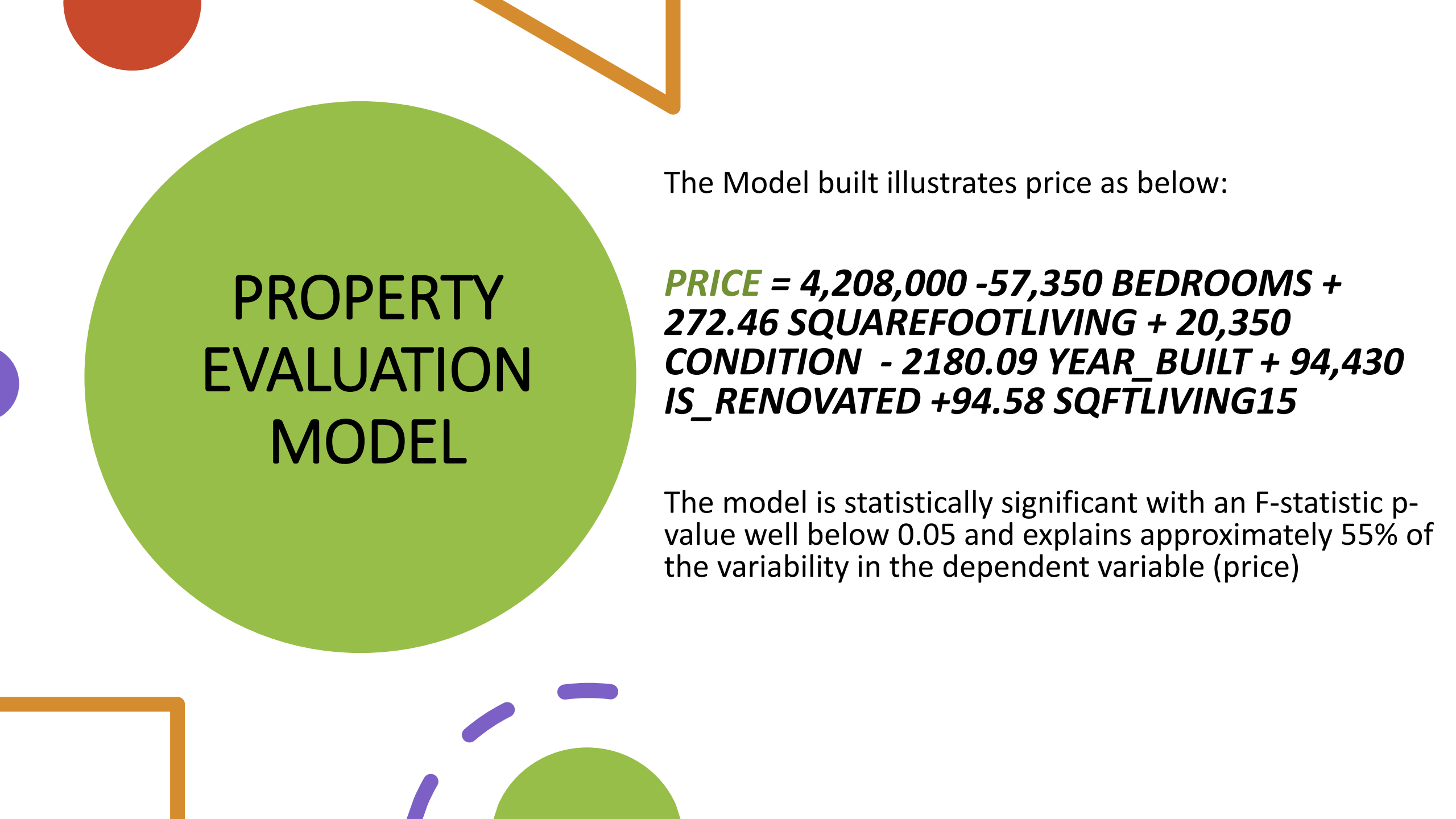


# UNIVARIATE ANALYSIS



# BIVARIATE ANALYSIS





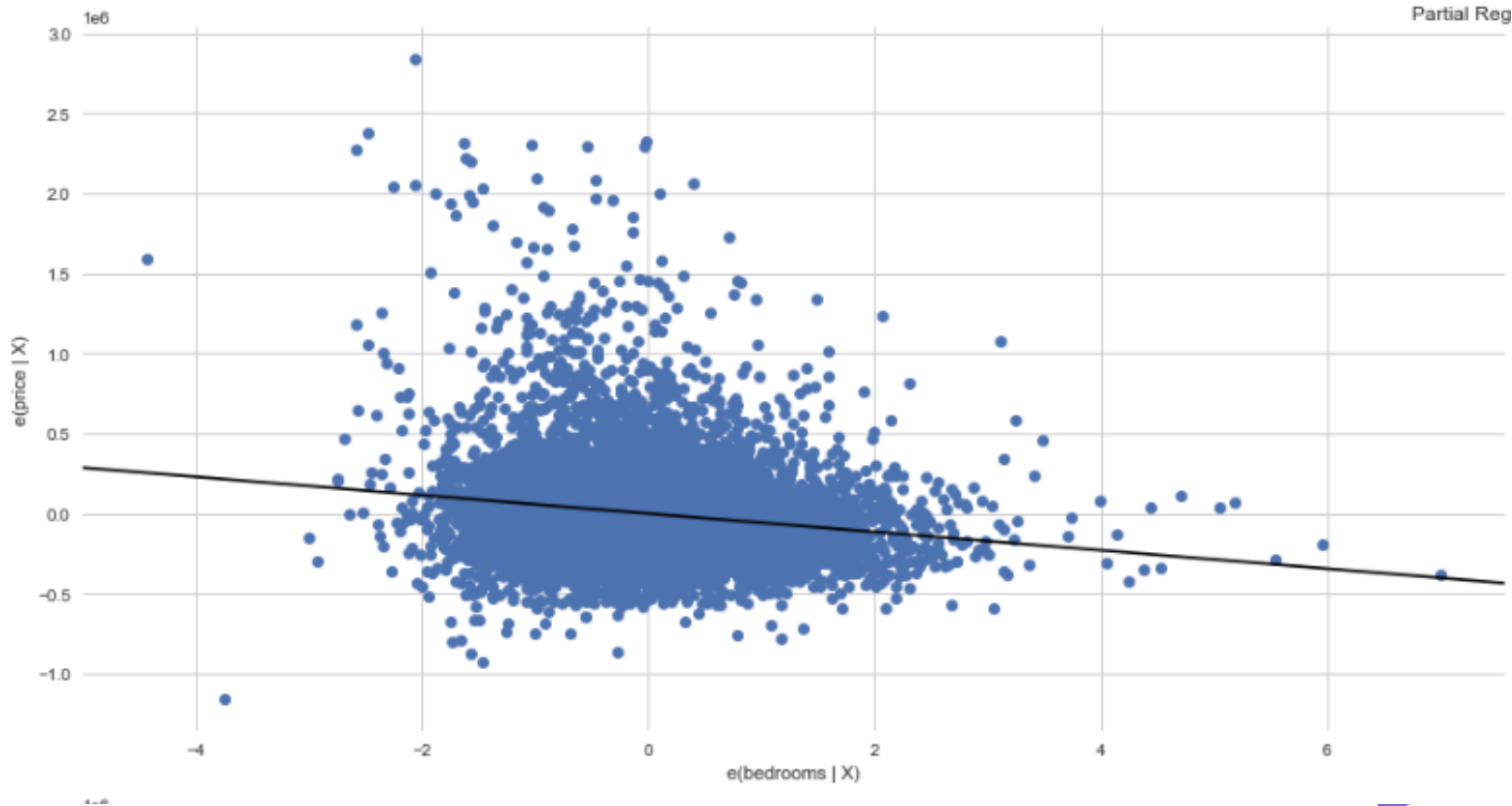
# PROPERTY EVALUATION MODEL

The Model built illustrates price as below:

$$\text{PRICE} = 4,208,000 - 57,350 \text{ BEDROOMS} + 272.46 \text{ SQUAREFOOTLIVING} + 20,350 \text{ CONDITION} - 2180.09 \text{ YEAR\_BUILT} + 94,430 \text{ IS\_RENOVATED} + 94.58 \text{ SQFTLIVING15}$$

The model is statistically significant with an F-statistic p-value well below 0.05 and explains approximately 55% of the variability in the dependent variable (price)

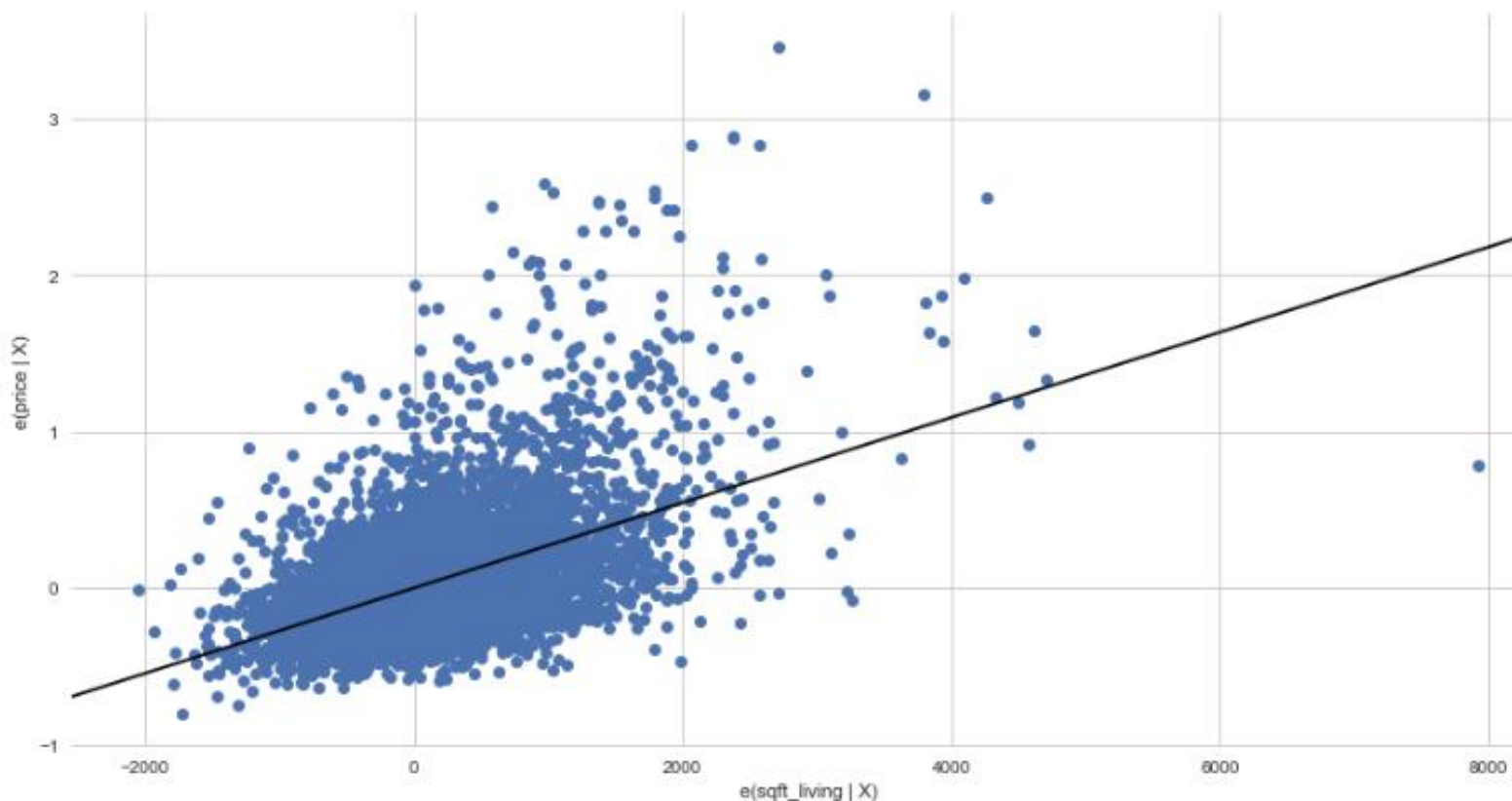
# BEDROOM FEATURE



On average, each additional bedroom is associated with a decrease of approximately USD 57,350 in the price.

The model coefficient for bedrooms is statistically significant, with t-statistic p-value well below 0.05.

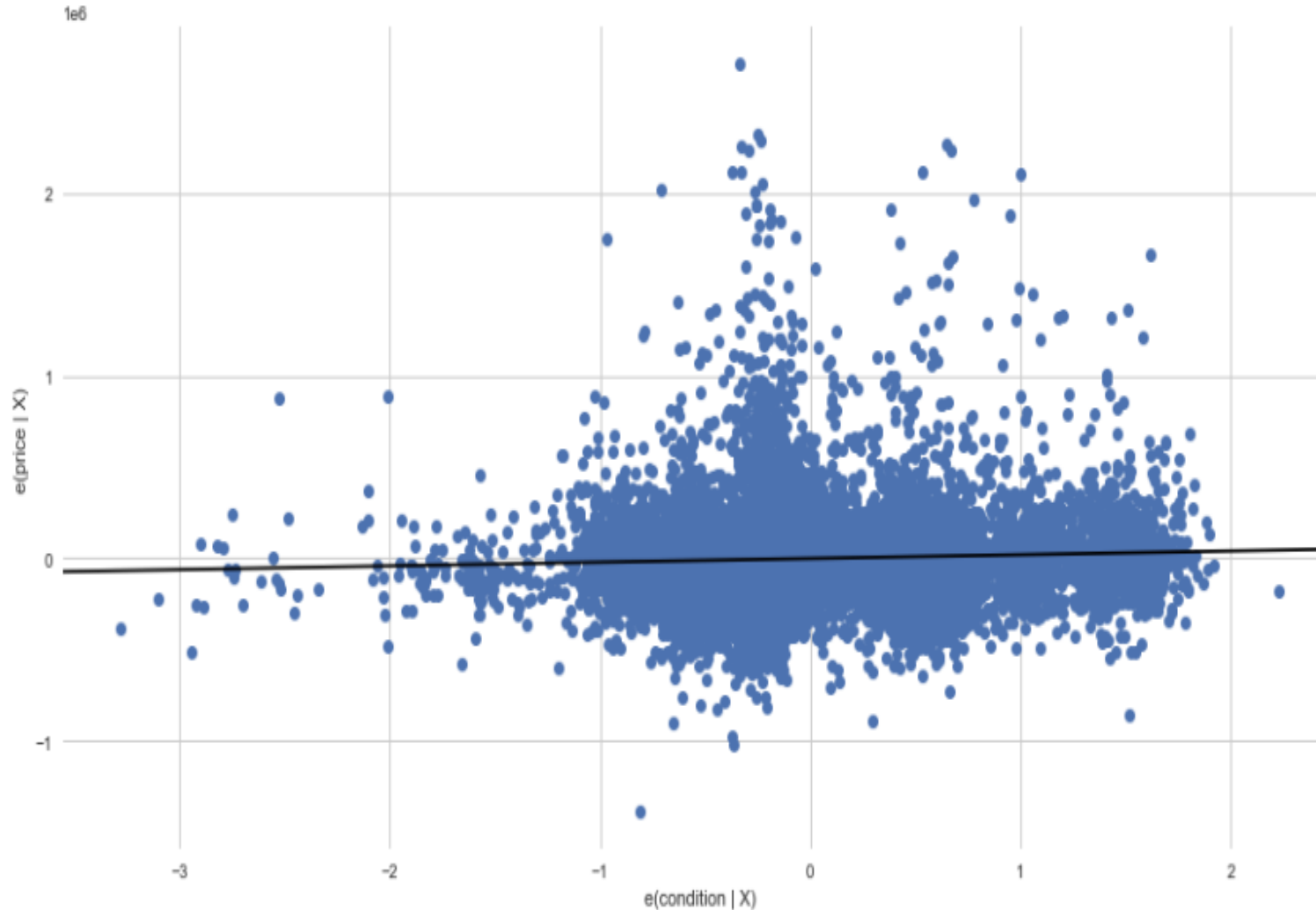
# SQUARE FOOT LIVING FEATURE



The model coefficient `sqft_living` is statistically significant, with t-statistic p-value well below 0.05.

For each additional square foot of living area is associated with an increase of approximately USD272.46 in the price.

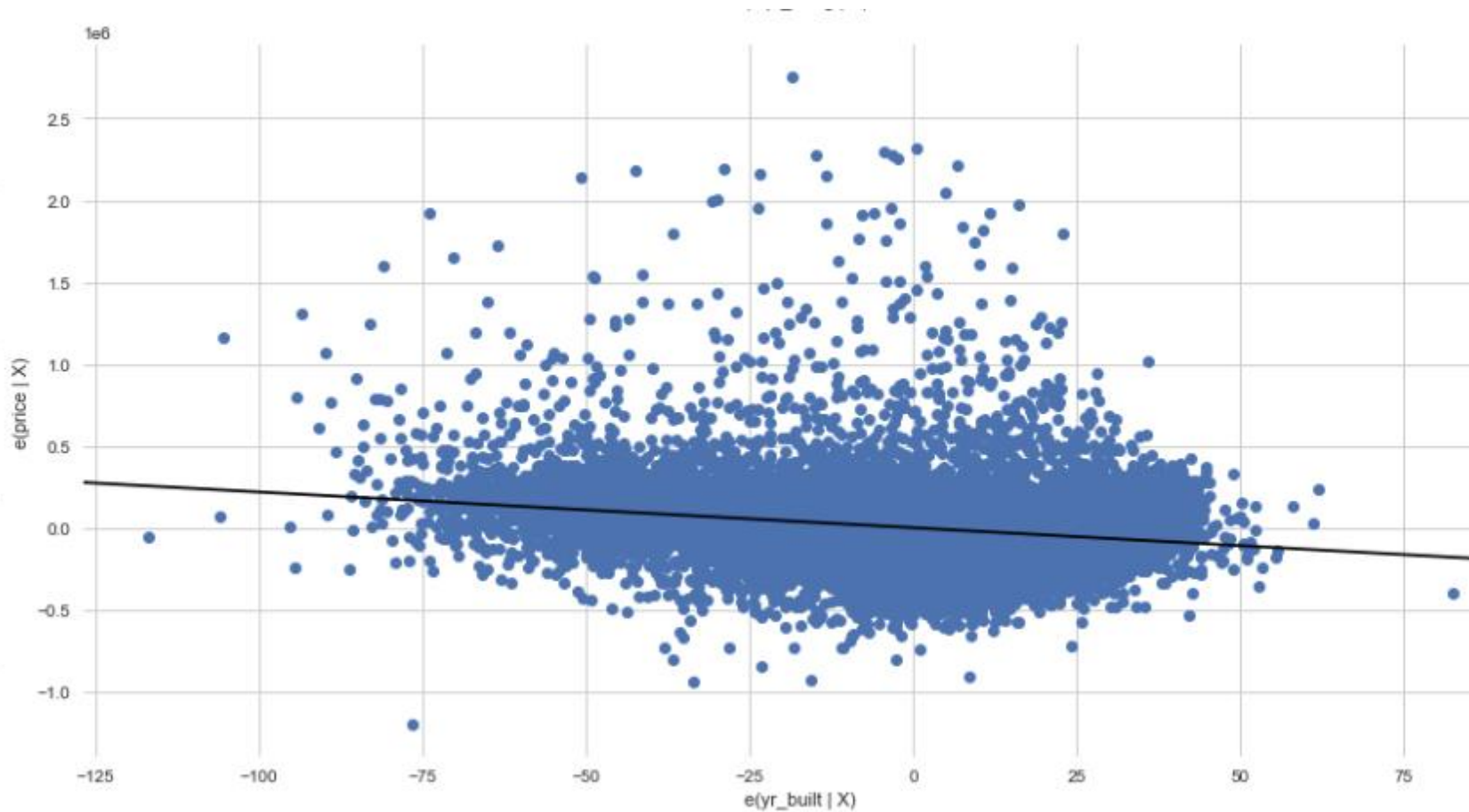
# CONDITION FEATURE



The model coefficient condition is statistically significant, with t-statistic p-value well below 0.05.

On average, each unit increase in condition is associated with an increase of approximately USD20,350 in the price.

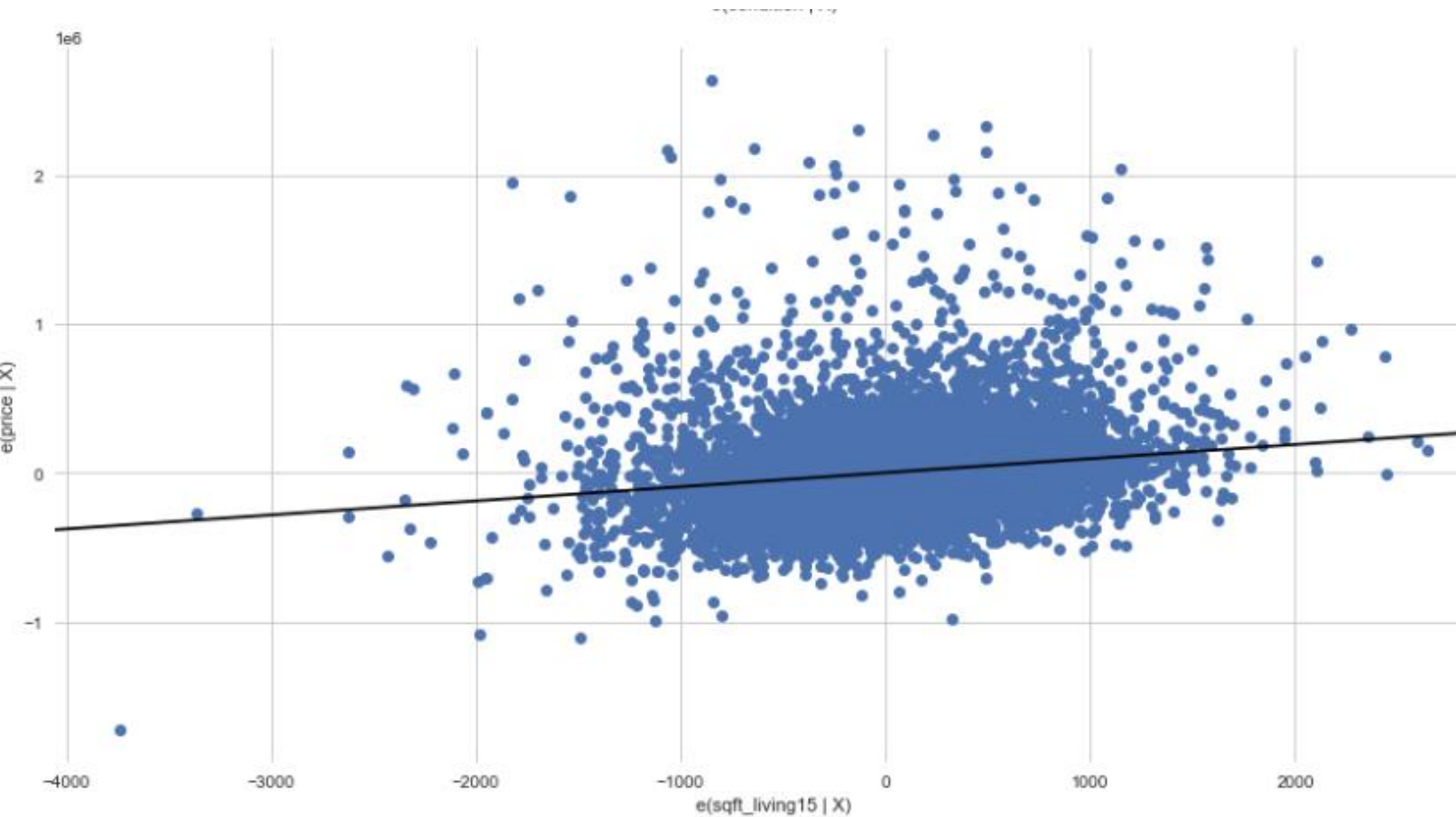
# YEAR BUILT FEATURE



The model coefficient `yr_built` is statistically significant, with t-statistic p-value well below 0.05.

The `yr_built` on the other hand has an associated decrease in price the older the house is by approximately USD 2184

# SQUARE FOOT LIVING FOR THE NEAREST 15 NEIGHBOURS FEATURE

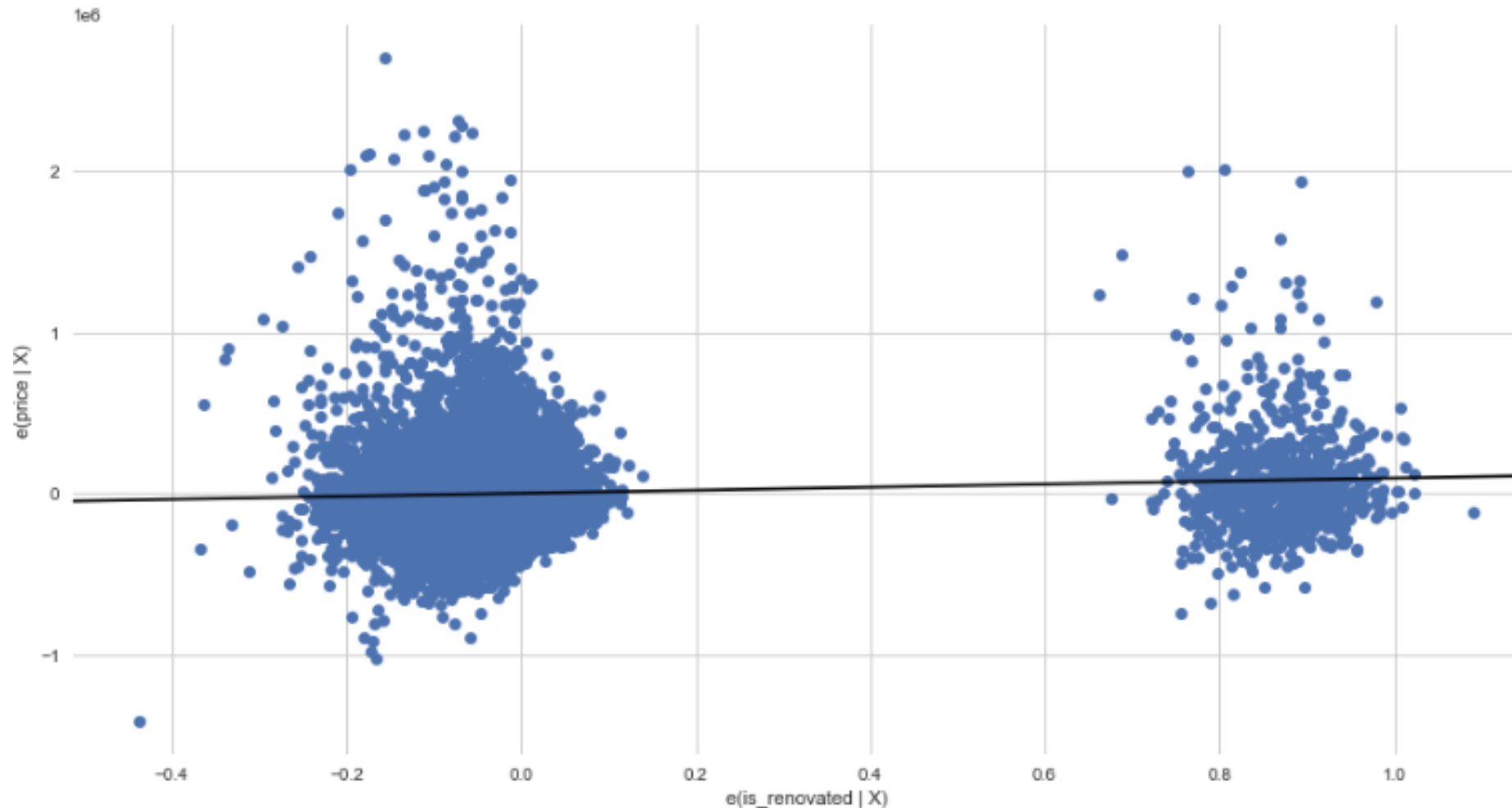


The model coefficient `sqft_living15` is statistically significant, with t-statistic p-value well below 0.05.

On average, each additional square foot of the neighboring properties' living area is associated with an increase of approximately USD 94.59 in the price.



# RENOVATION STATUS FEATURE



The model coefficient `is_renovated` is statistically significant, with t-statistic p-value well below 0.05. A renovated property increases the price by USD 94,300

# CONCLUSION

1. The model shows a moderate level of predictive power . The model can explain approximately 55% of the variability in home prices.
2. Significant predictors of price as per the model are ;
  - the number of bedrooms
  - square footage of living area
  - condition of house
  - year built
  - whether the property has been renovated
  - the square footage of neighboring properties
3. Normality assumption: The Q-Q plots of the model's residuals suggest that they approximately follow a normal distribution.





# RECOMMENDATIONS TO HOMEOWNERS

Consider the influence of neighboring properties

Renovations can add value

Pay attention to the year built

Consider the number of bedrooms

Focus on the square footage

Maintain the condition of the property



# RECOMMENDATIONS TO MEMBERS NAR



Stay updated on market trends

Educate clients on the impact of features

Stay informed about regulations and policies

Collaborate with appraisers

Conduct thorough market analyses

Provide renovation recommendations

Any  
Questions....

---



Thank you!

---

