

AMES HOUSING

ANALYSIS & PREDICTIONS

Riche Ngo





TABLE OF CONTENTS

01

PROBLEM STATEMENT

The issue to be addressed

02

METHODOLOGY

The data science process

03

EXPLORATORY DATA ANALYSIS

Initial investigations on data
to discover patterns

04

TARGET ENGINEERING & FEATURE SELECTION

Transform target variable and
consider which features to
use

05

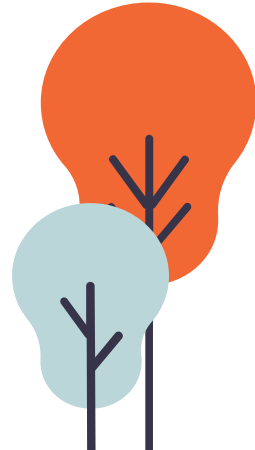
MODELLING

Construct a model using
training data to produce
predictions

06

CONCLUSIONS & RECOMMENDATIONS

Key takeaways and useful
insights for



PROBLEM STATEMENT

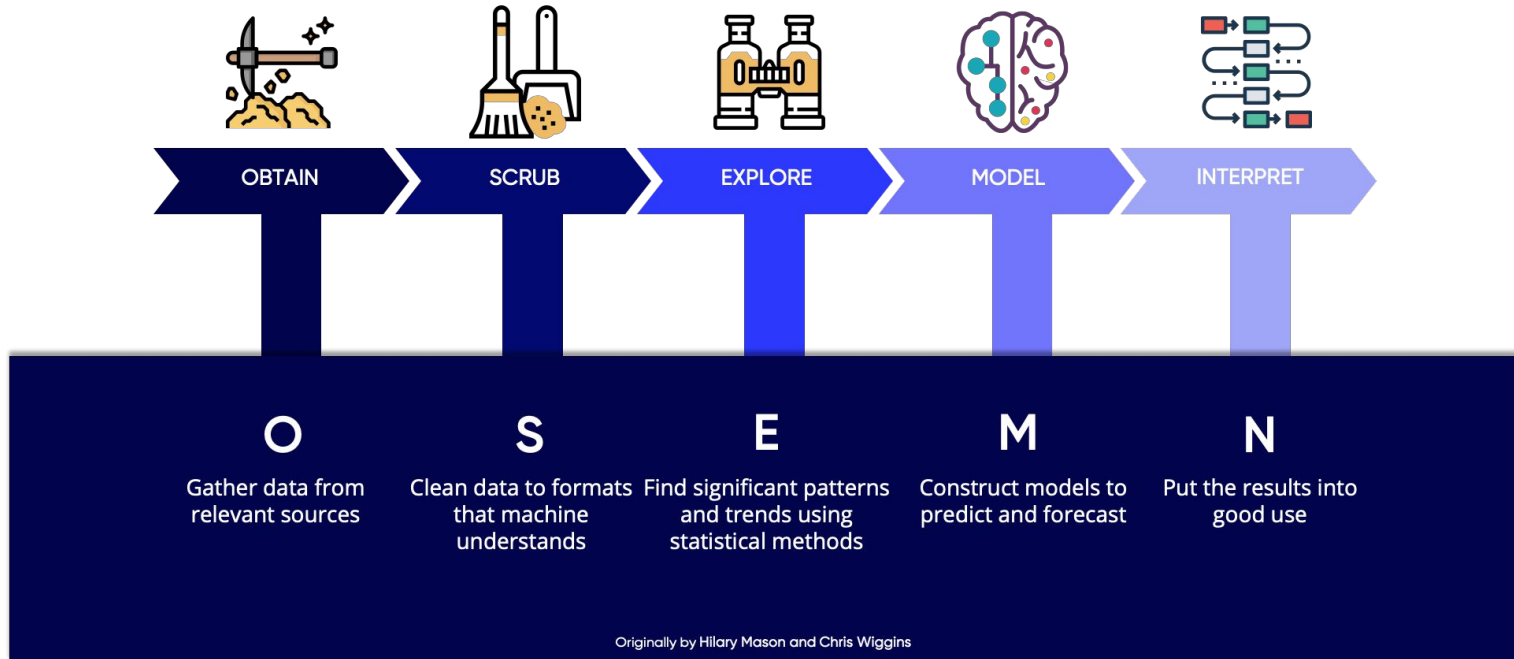
As part of a real estate agency in Ames, we are looking to boost annual sales. What is the best regression model to predict Ames housing prices? What expert advice could we give to the investors and homeowners of Ames, on which housing components and features would greatly influence housing prices?



FRIEDRICHE
IOWAREALTY

A MAKESHIFT AGENCY

Data Science Process



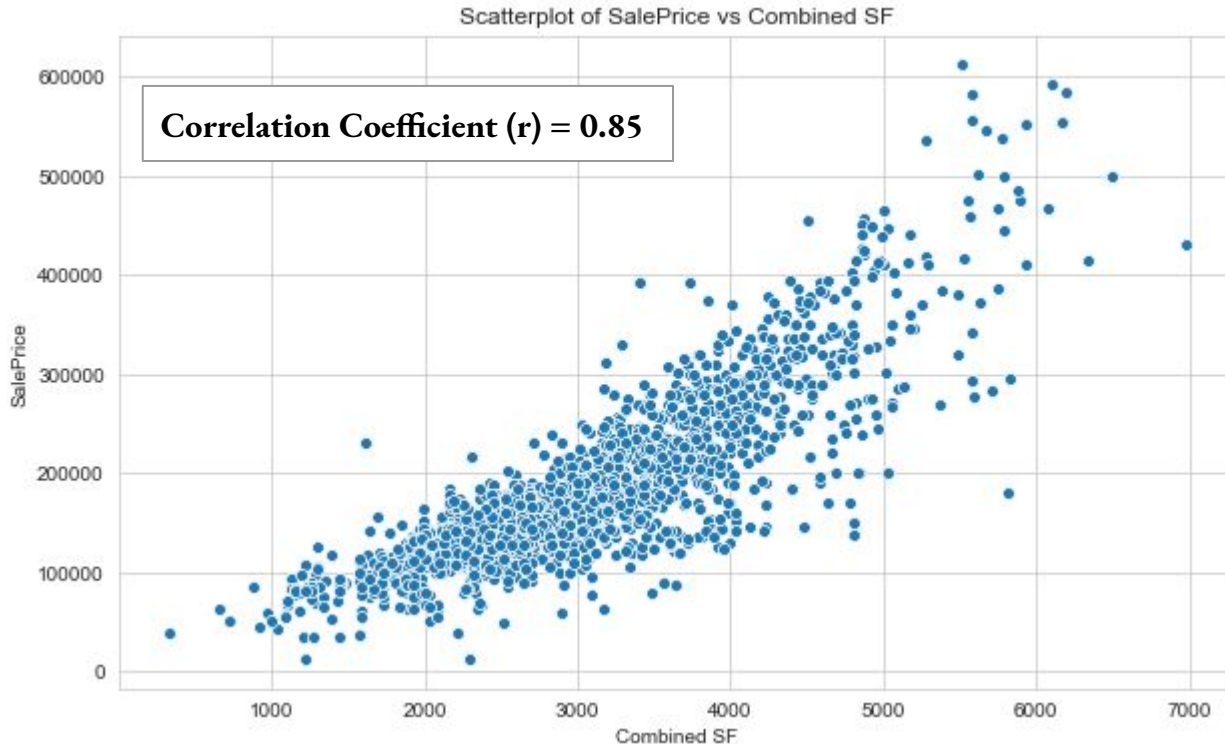
WHAT IS IN THE DATA?

The Ames Housing Dataset is an exceptionally detailed and robust dataset with more than 70 columns of different features relating to houses.

- Some of the features include the type of dwelling, the location of the house, ratings on different house aspects, etc.
- Information obtained from the Ames Assessor's Office, for computing assessed values for individual residential properties sold in Ames, Iowa from 2006 to 2010.



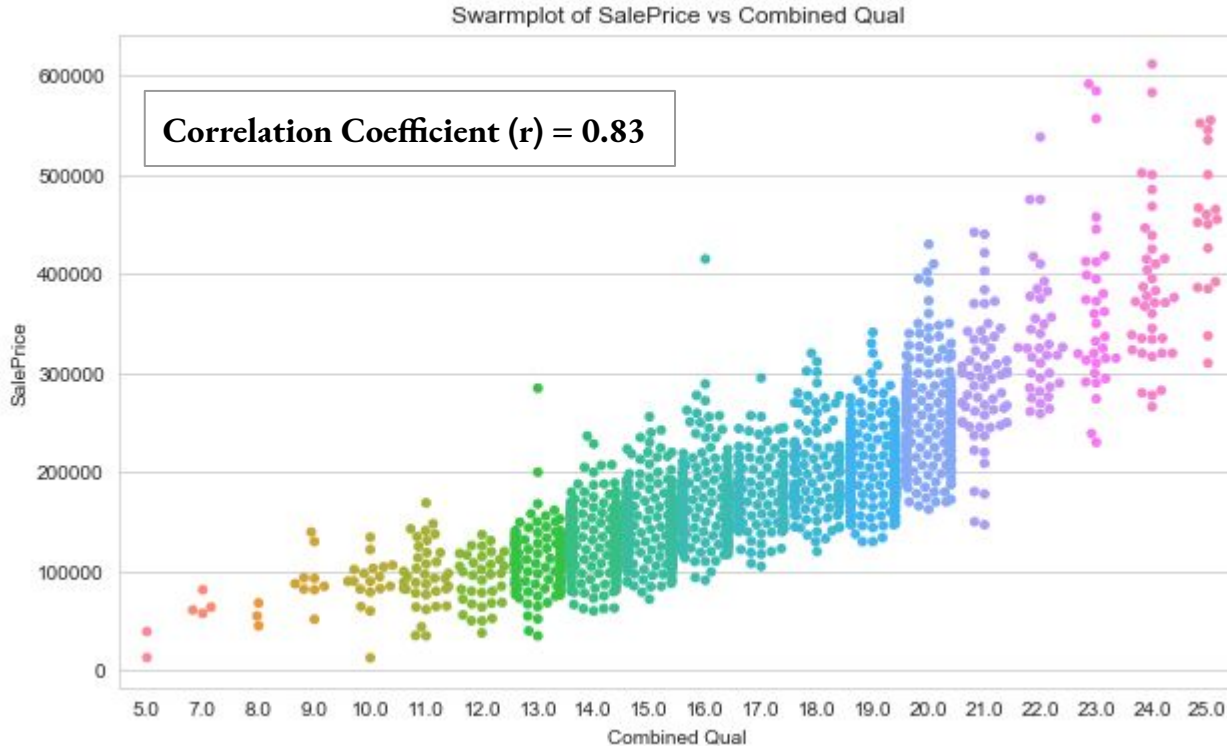
EXPLORATORY DATA ANALYSIS



Observations:

- Strong positive relationship between housing sale price and total area of the house, including the garage, basement, and living area above ground.
- High correlation coefficient.
- As total area of the house increases, sale price increases.
- Good indication of a linear relationship.

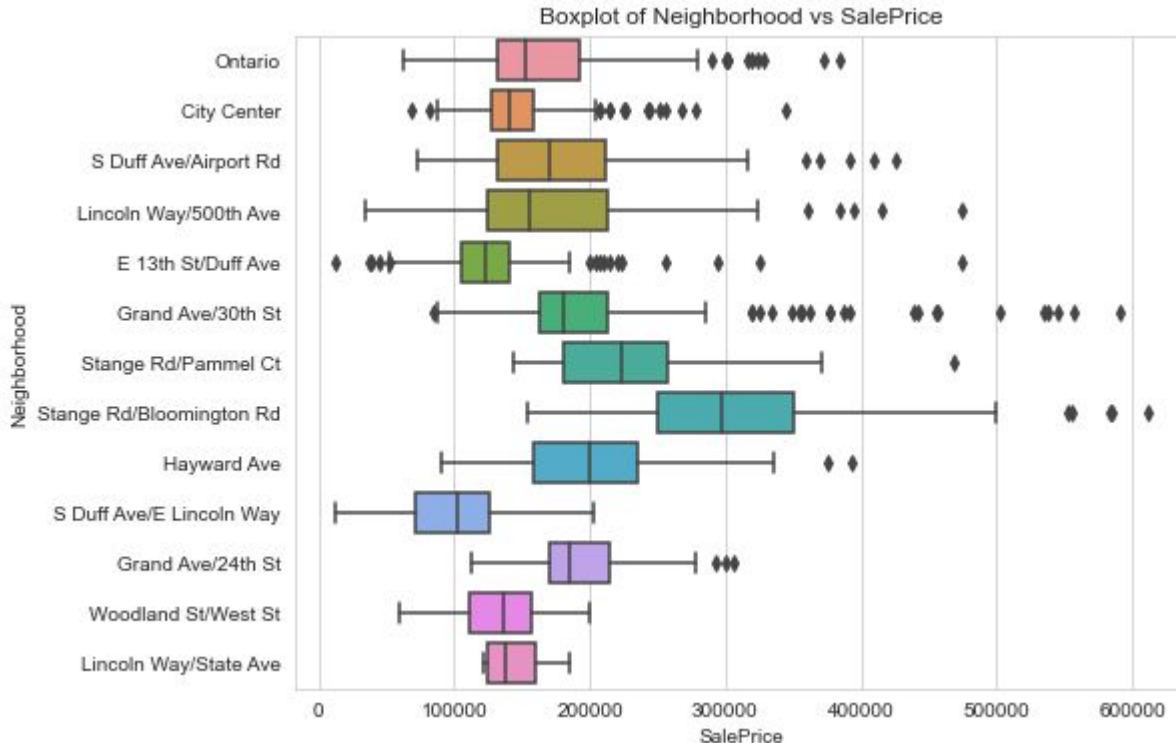
EXPLORATORY DATA ANALYSIS



Observations:

- Strong positive relationship between housing sale price and overall quality of the material, finishes, and exterior of the house.
- High correlation coefficient.
- As overall quality increases, sale price increases.
- Clear ordinal phenomenon.
- Largest number of houses within the overall quality range of 13-20.

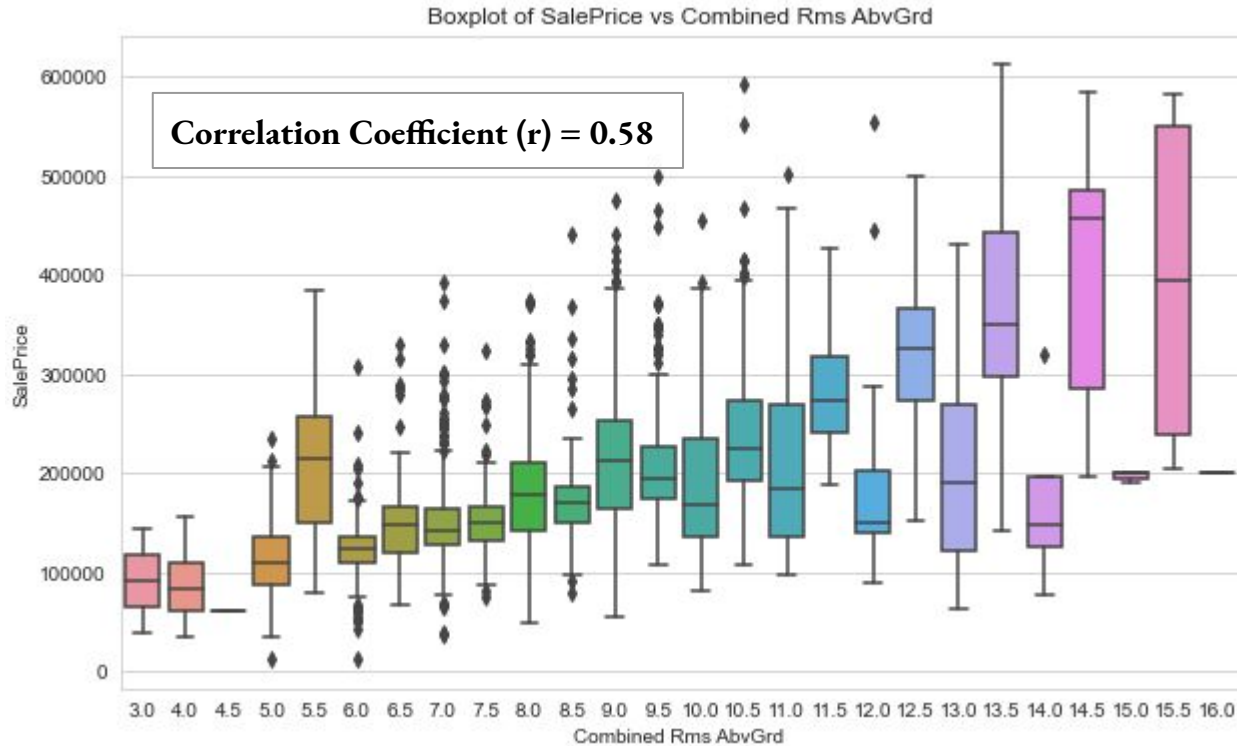
EXPLORATORY DATA ANALYSIS



Observations:

- Good variability in the data after grouping different neighborhoods into popular Ames regions.
- Highest median of sale price within the Stange Rd/Bloomington Rd region.
- Lowest median of sale price within the S Duff Ave/E Lincoln Way region.

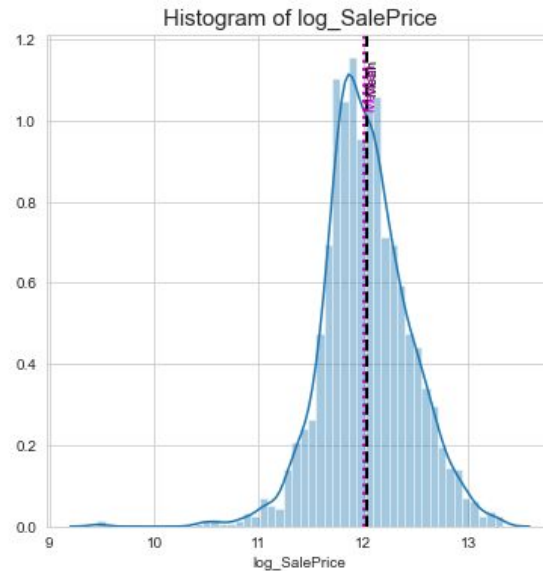
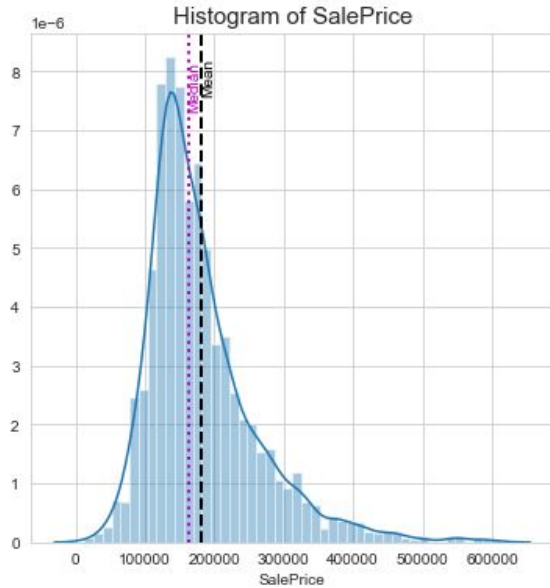
EXPLORATORY DATA ANALYSIS



Observations:

- Strong positive relationship between housing sale price and total number of rooms the house, including bedrooms and bathrooms.
- High correlation coefficient.
- As total number of rooms increases, sale price increases.
- Variability in data.

TARGET ENGINEERING



Improvement in normality

Observations:

- The distribution of target variable is right(positive)-skewed.
- Though not a major issue since we are not conducting inferences, correcting for the **violation in normality assumption** helps to improve predictions.
- Great improvement in normality after doing a **log transformation** on the target variable.
- Achieved a more homoscedastic model.

Lasso Regression used as a feature selection tool

STRENGTHS

- Fast and efficient method to trim down features.
- Automatic.
- Easy to interpret.

WEAKNESSES

- May produce a model that does not make sense.
- Arbitrary nature of selecting from a group of correlated features.

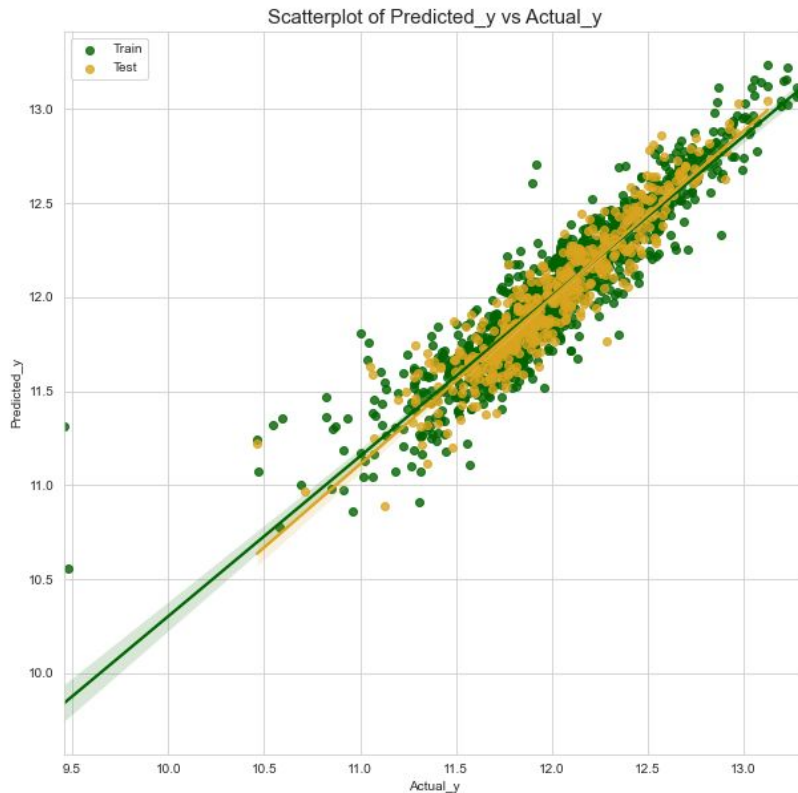
LASSO REGRESSION MODEL

This was the selected model after comparing prediction scores against other types.

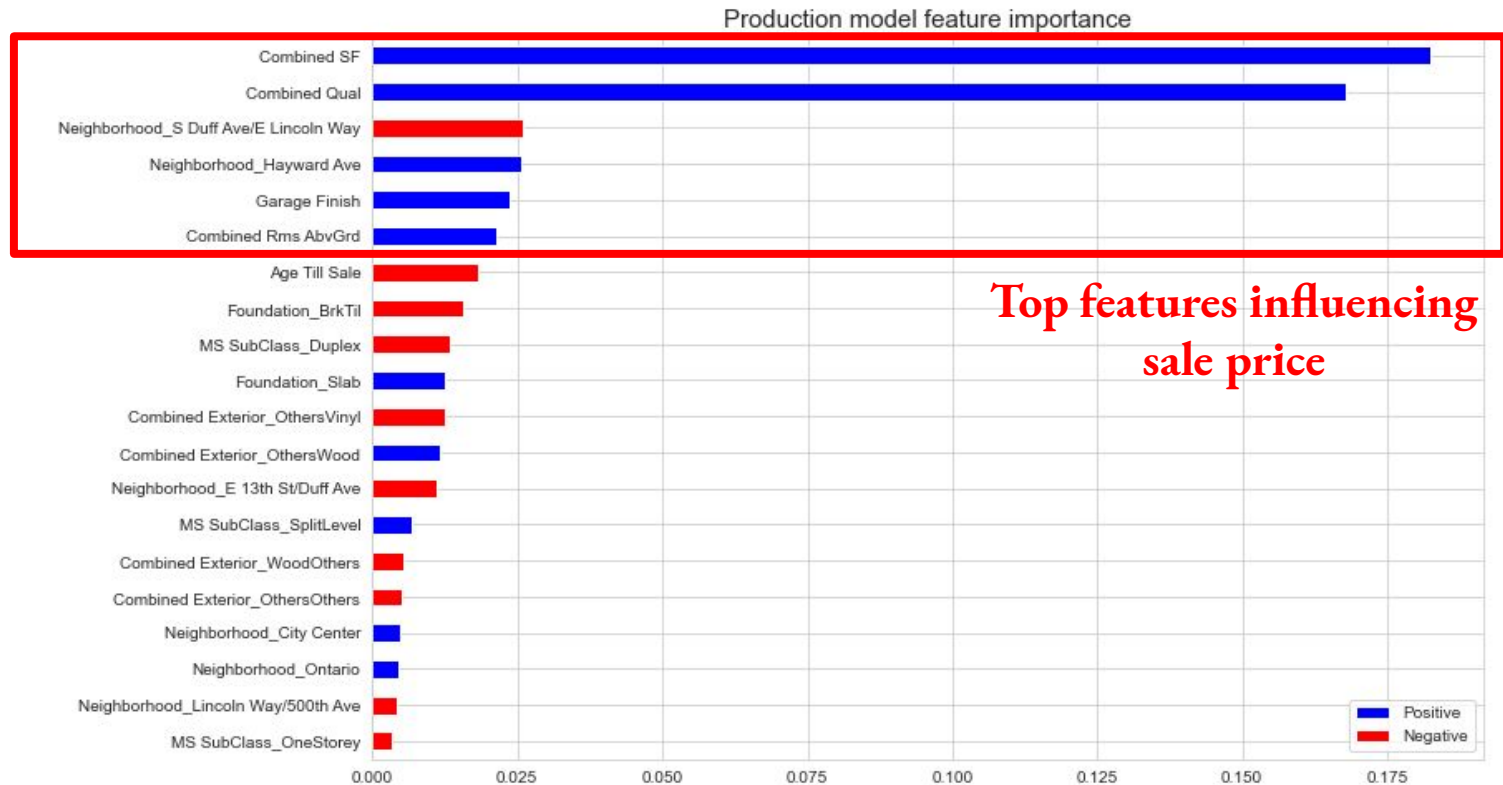
R^2 Score: **0.86**

RMSE Score: **22,841 (\$)**

- Strong linear relationship between the predicted sale price and actual sale price extracted from the train data.
- Minimal difference between the best-fit lines of the train data and test(hold-out) data.



MODELLING



CONCLUSIONS & RECOMMENDATIONS



GOAL

Boost annual sales for real estate agency



INVESTORS

Advise on what housing features to look out for



HOMEOWNERS

Advice on what housing features to upkeep

01

TOTAL AREA OF HOUSE

The larger the area, the higher the housing prices.

02

OVERALL QUALITY

The higher the rating, the higher the housing prices.

03

NEIGHBORHOOD REGION

Centrally located and easily accessible houses tend to fetch a higher price.

04

INTERIOR FINISH OF GARAGE

Make sure the garage interior is finished, first impressions matter.

05

TOTAL NUMBER OF ROOMS

More rooms can accommodate bigger families, higher demand means higher prices.

“Knowledge is power.”

—Francis Bacon





THANKS

Questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

