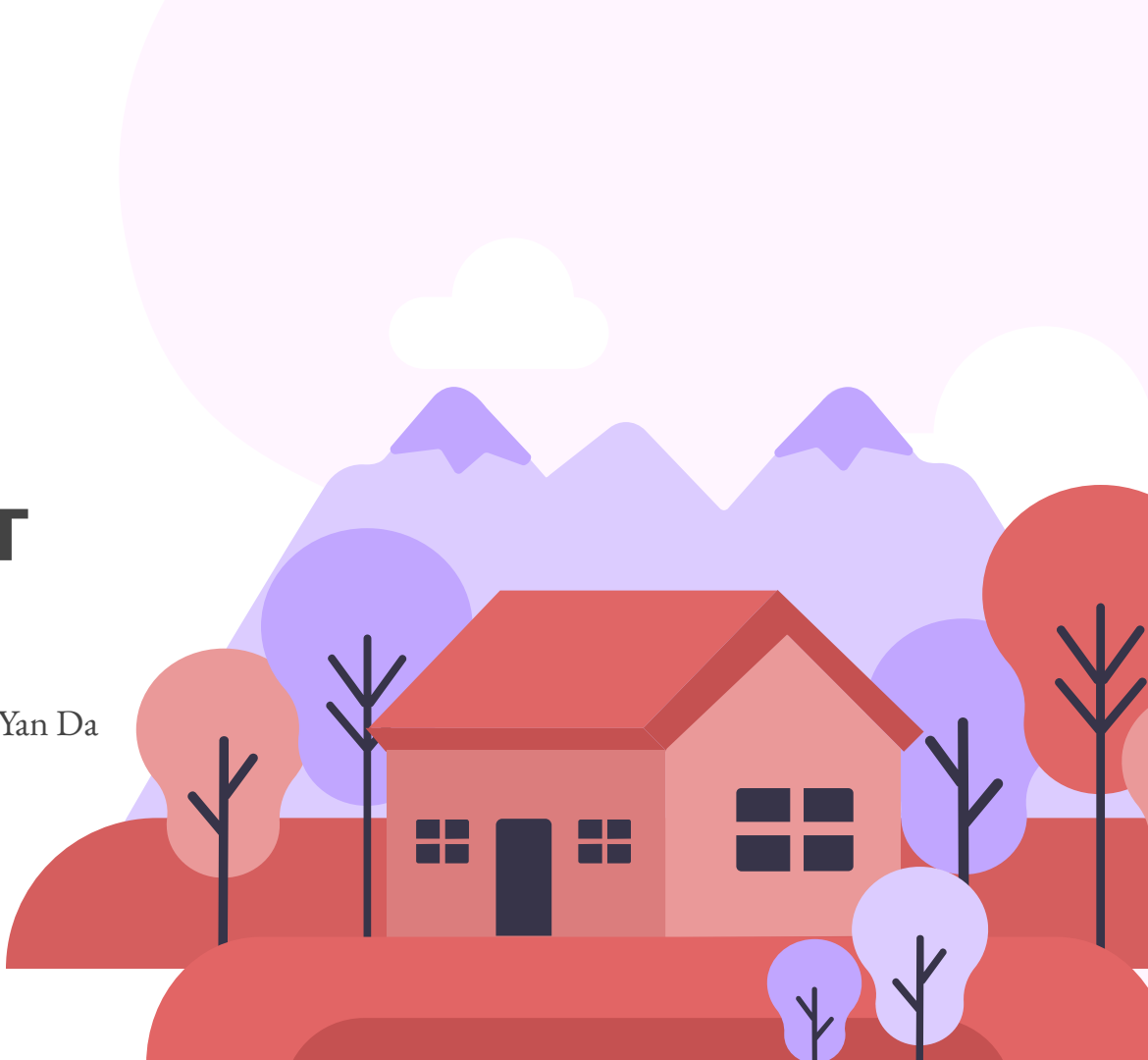


AMES HOUSING DEVELOPMENT

DSI 33

Mary-Anne | Wei Zhe | Daniel | Jimmy | Yan Da



CONTENTS

- Background
- Problem Statement
- Workflow Process
- Conclusion & Recommendations



BACKGROUND

Real Sky Estate Development has been developing residential areas around Ames and is looking to **procure and develop new housing** in the area.

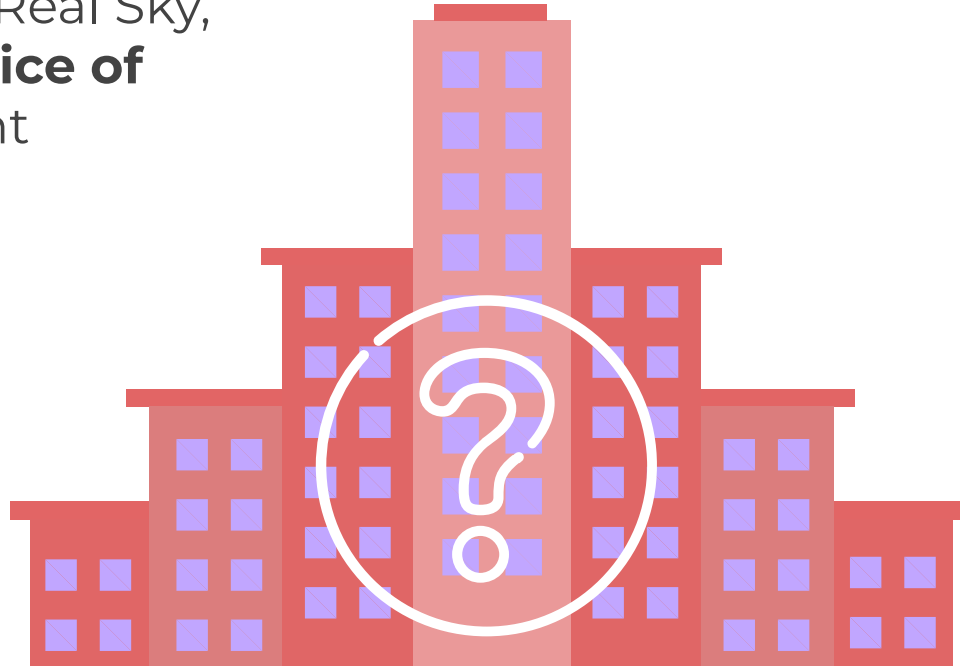
Previous housing developments **were not returning favourable profits** due to the recent pandemic and political tensions. It also caused a rise in the costs of living and building materials over the years.

Facing a forecasted recession in the upcoming year, Real Sky senior management team has reached out to our data science team to **pinpoint factors** that will direct towards **revamping the company's focus structure to improve the attractiveness and sales price** of new housing developments.



PROBLEM STATEMENT

What **core features** should we, Real Sky, focus on to **increase the sale price of homes** for our next development project?



WORKFLOW PROCESS



AMES HOUSING DATASET OVERALL

DATA

80 Features

2930 Entries

Home sales from
Years 2006-2010

3 TYPES OF DATA

Numerical

Square foot | No. of

Ordinal

Excellent, poor, average

Nominal

Material types | Names

MISSING VALUES

Replaced with best
fit values

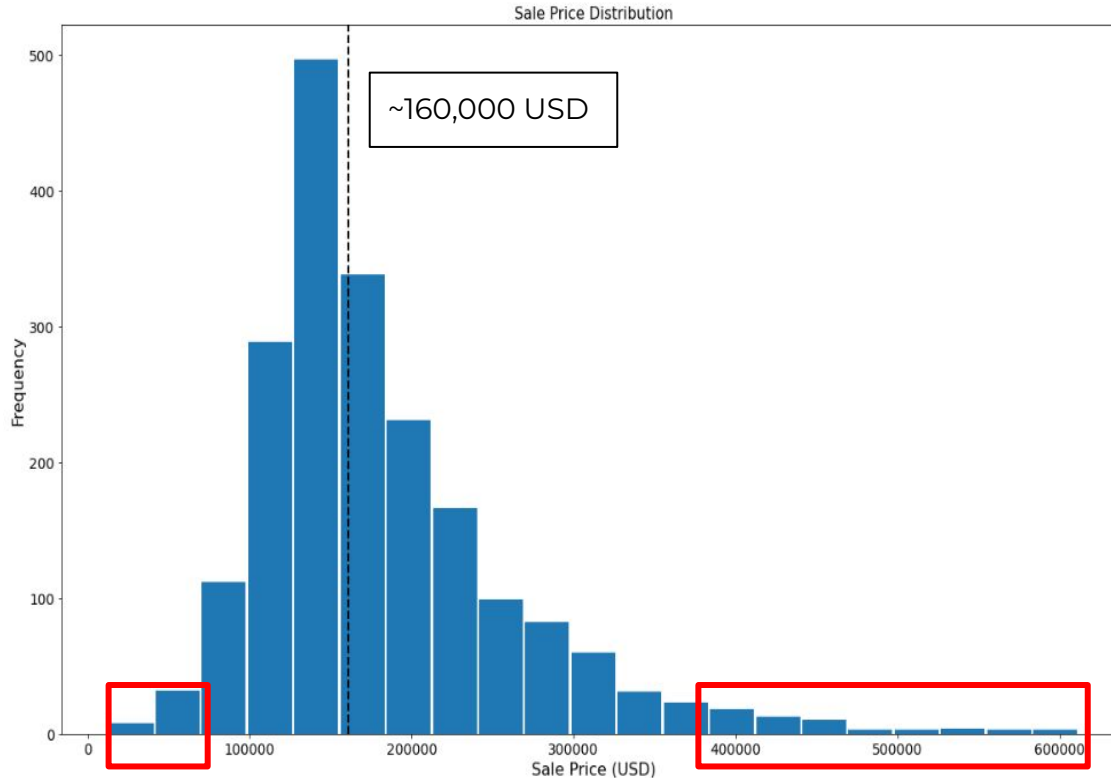
Median Values

Removed

Misc Features : 97% Missing
Values

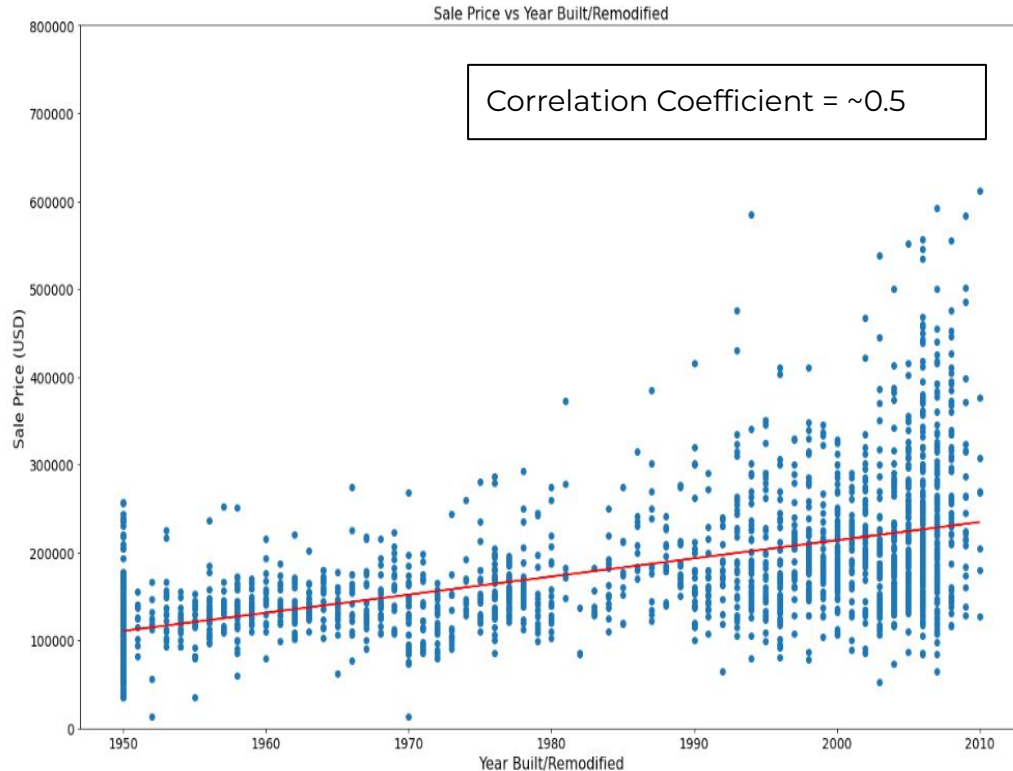


EDA: DISTRIBUTION OF HOUSING SALE PRICE WITHIN 100K TO 300K USD



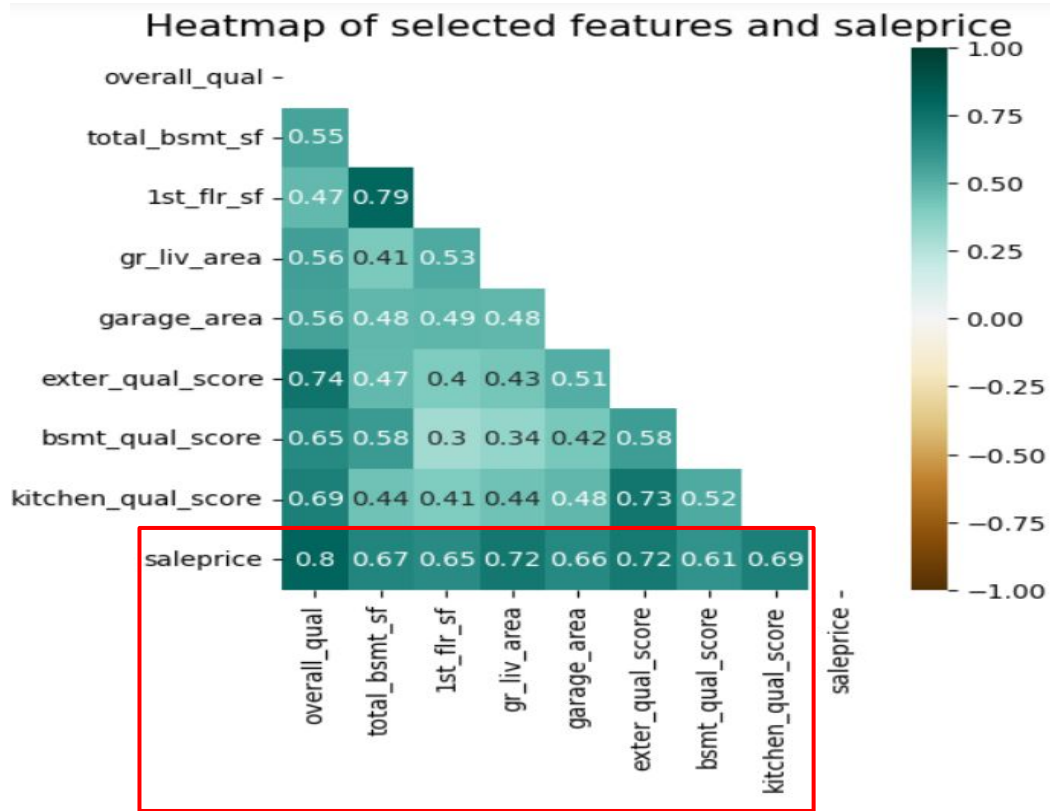
- Main distribution of sale prices within 100-300K USD range.
- Median sale price at ~160,000
- Limited data for sale prices < 50K and > 400K.

EDA: MINIMAL IMPACT OF YEAR BUILT/REMODIFIED ON SALE PRICES



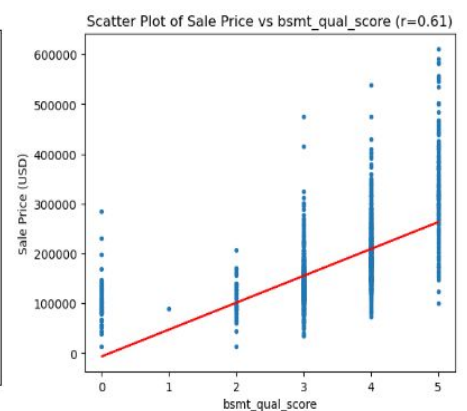
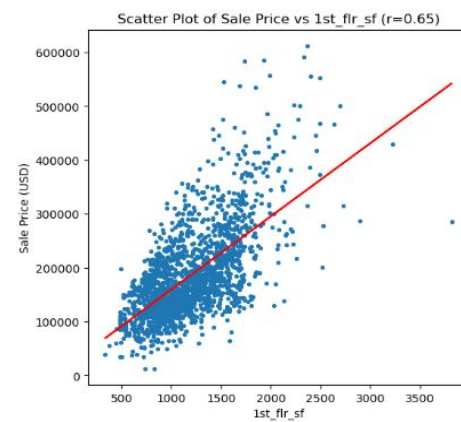
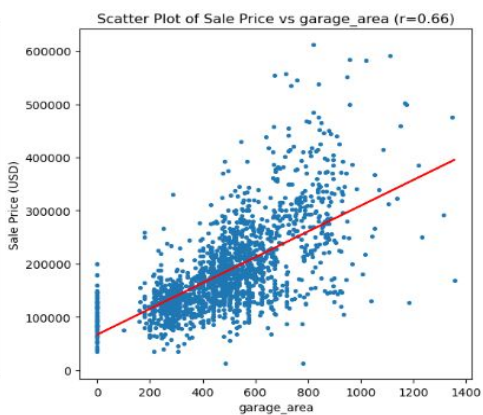
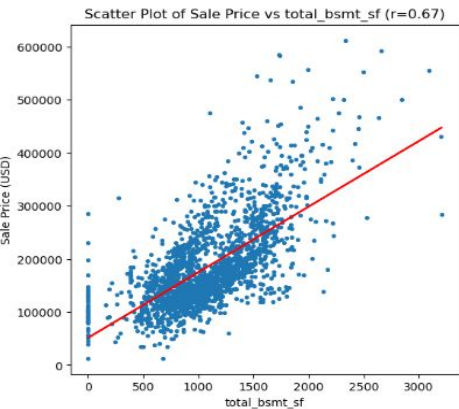
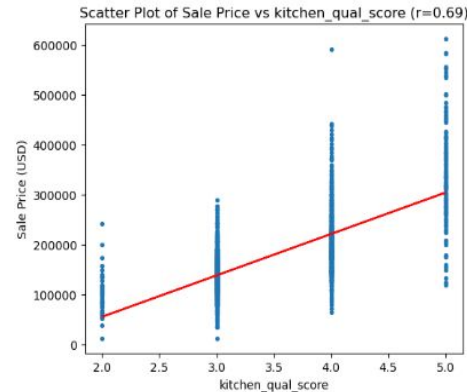
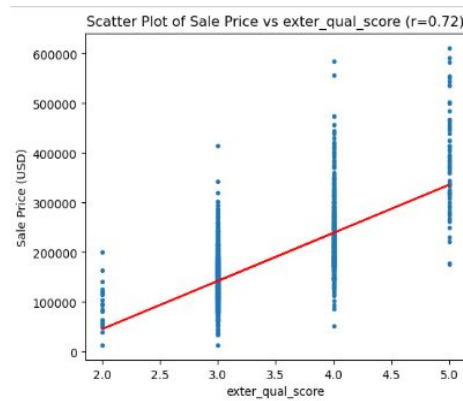
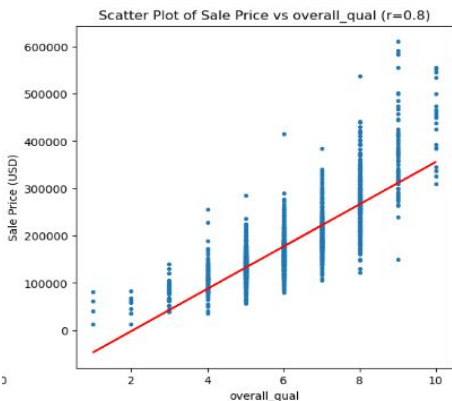
- Model creation for sale price prediction of new housing developments.
- Weak correlation coefficient value of ~ 0.5 .
- Poor linearity between Sale Pricing and Year Built/Remodified.
- Not a significant factor that will affect sales price.

FEATURES SELECTION : FEATURES WITH > 0.6 CORRELATION COEFFICIENT WITH SALE PRICE

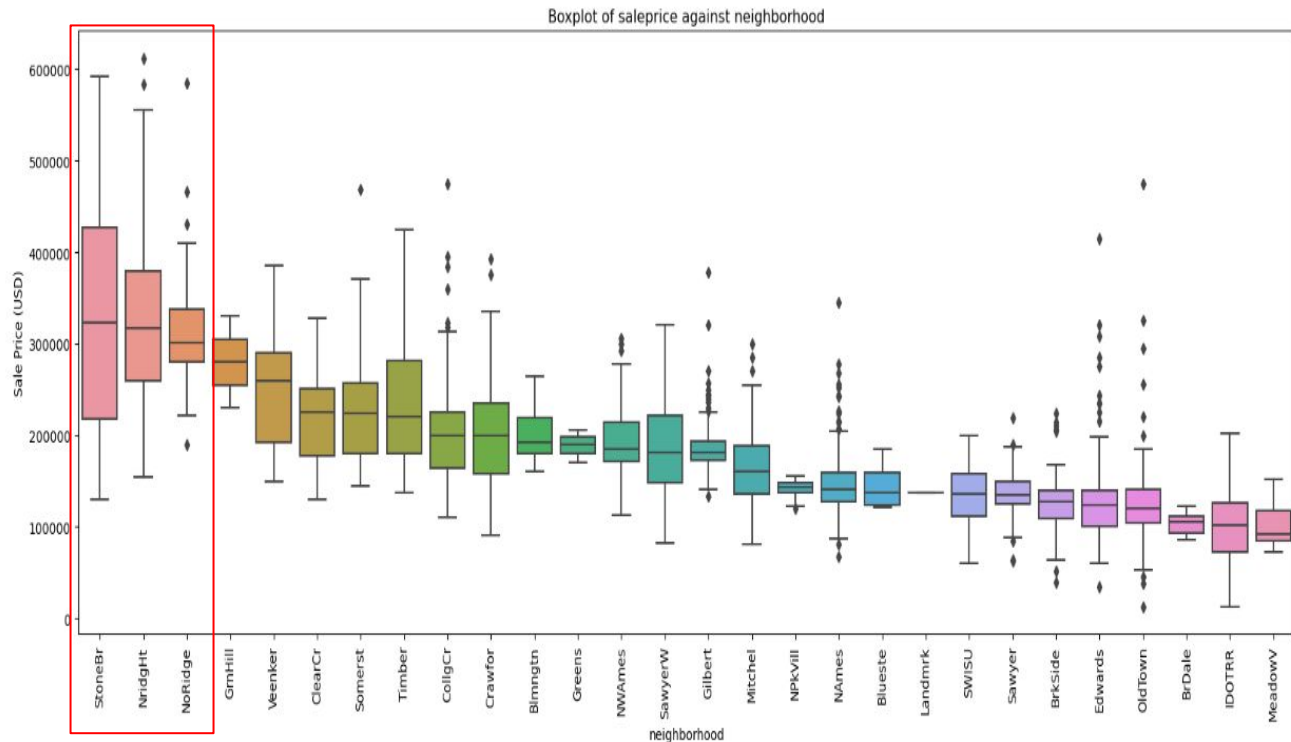


Selected Feature	Description	Corr. Coeff.
overall_qual	Overall material and finish of the house	0.80
gr_liv_area	Above ground living area	0.72
exter_qual_score	Exterior material quality	0.72
kitchen_qual_score	Kitchen quality	0.69
total_bsmt_sf	Total basement area (Square Feet)	0.67
garage_area	Size of garage (Square Feet)	0.66
1st_flr_sf	Area of 1st floor (Square Feet)	0.65
bsmt_qual_score	Basement height	0.61

FEATURES SELECTION : POSITIVE CORRELATION BETWEEN SELECTED FEATURES AND SALE PRICE



FEATURE SELECTION : SELECTION OF “Neighborhood” AS MODELLING FEATURE



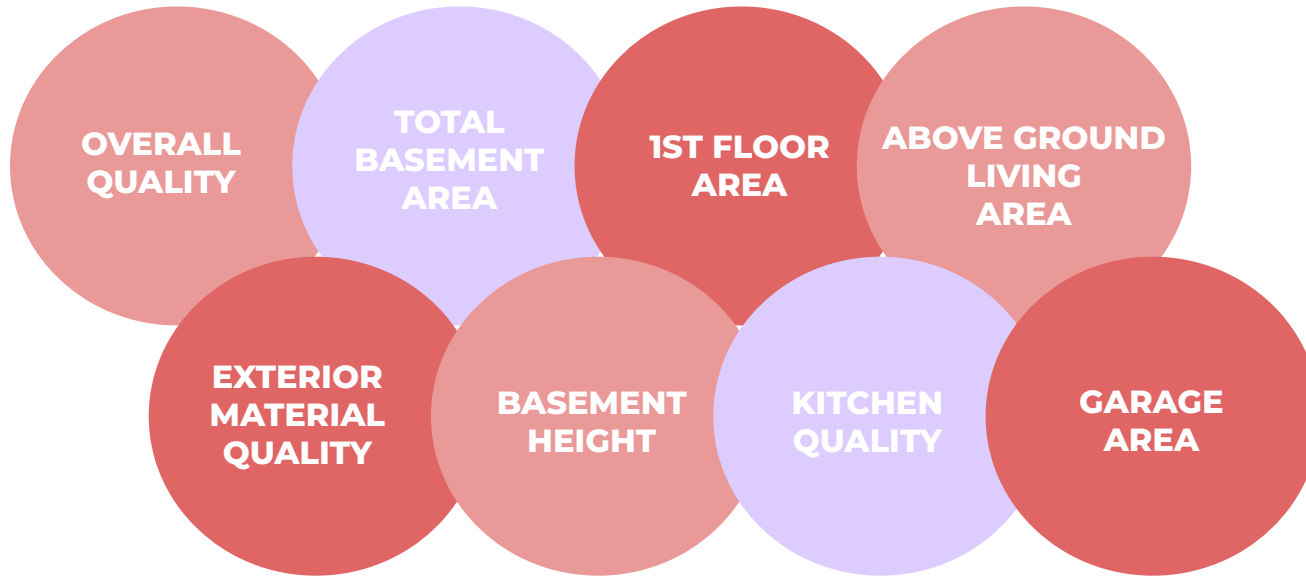
- Good location as one of the top 5 most desired home features from Forbes*.
- Top 3 locations with better sales price:
 - Stone Brook
 - Northridge Heights
 - North Ridge

Reference links:

<https://www.forbes.com/sites/forbesrealestatecouncil/2020/04/27/15-most-desirable-home-features-to-todays-buyers/?sh=7997a5677a4e>

CORE FEATURES: OVERVIEW OF SELECTED FEATURES

Numerical/Ordinal Features



Categorical Features



MODEL SELECTION - BEST PREDICTION MODEL: RIDGE REGRESSION

Model	Train R2 score	Train RMSE	Test R2 score	Test RMSE
Baseline	-	-	0.0	79277
Linear Regression	0.88816	26256	0.86899	29765
Lasso Regression	0.88816	26256	0.86898	29766
Ridge Regression	0.88816	26256	0.86901	29763

R2 score: measures how much the variability in sale price can be explained by the selected features in our model
RMSE (Root Mean Square Error) : measures the average difference of the predicted value from actual sale price



Higher R2 score



Better Model Performance

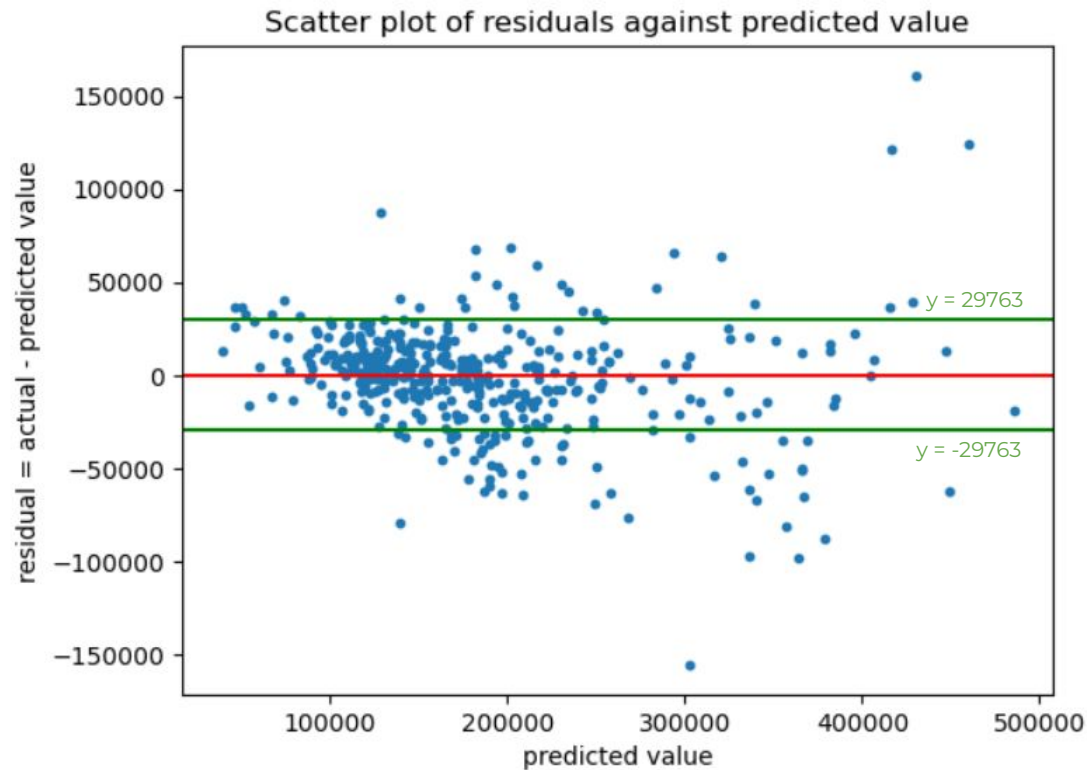


Lower RMSE

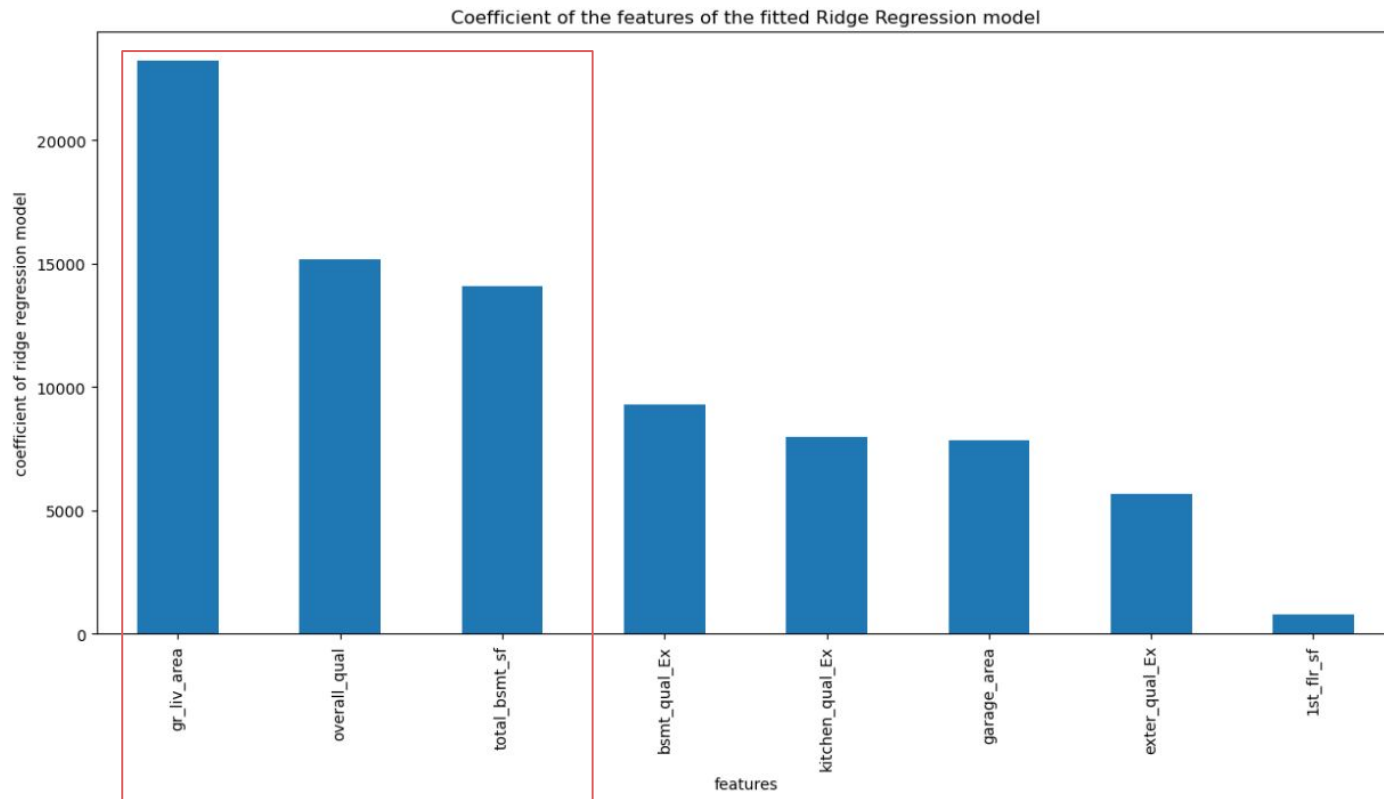


Better Model Performance

RMSE for Ridge Regression = 29763

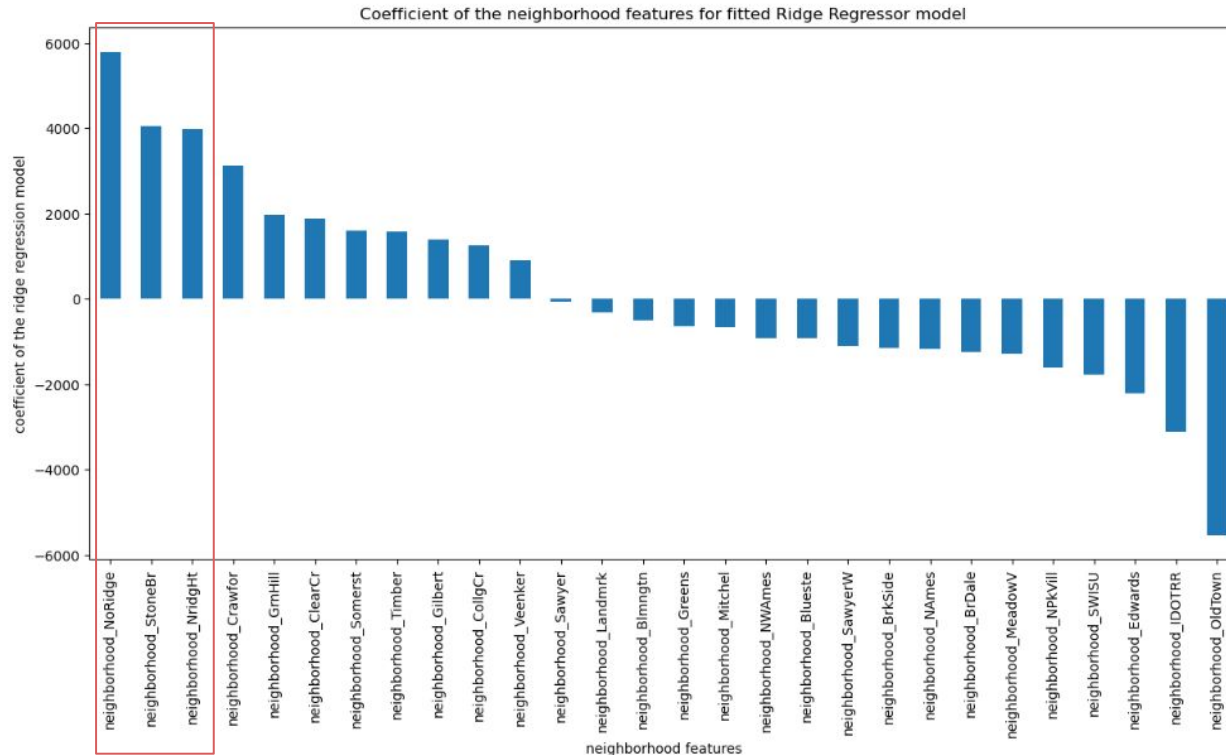


Top 3 features are ground living area, overall quality and total basement area



	Median
Ground Living Area	1444 sq. ft.
Overall Quality	Score 6
Total Basement Area	994 sq. ft.

DESIRABLE LOCATIONS



Top 3 neighborhoods:

- **Northridge, Stone Brook and Northridge Heights**
- Located North-west of Ames with lower crime rate and better education ranking*

*www.neighborhoodscout.com/ia/ames

LIMITATIONS

INSUFFICIENT DATA

External factors
skewing sale prices
(eg. recession, natural
disasters)

NEW DATA

More recent year data

TIME CONSTRAINT

Different models



MOVING FORWARD

MONITORING NEW DATA

- Change in trends
- Change in consumer preference

CLARITY IN FEATURES VALUES

Garage Qual (Ordinal): Garage quality

Ex **Excellent**

Gd **Good**

TA **Typical/Average**

Fa **Fair**

Po **Poor**

NA **No Garage**

RECOMMENDATIONS

Potential Development Area

North-West of Ames



Budget Allocation

Prioritization of
Core Features



Marketing

Core Features

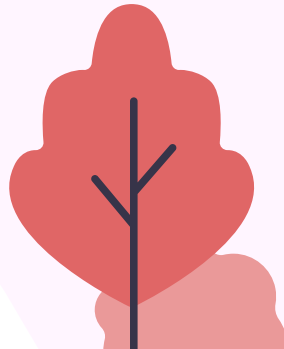


CONCLUSION

PROBLEM STATEMENT: What **core features** should we, Real Sky, focus on to **increase the sale price of homes** for our next development project?



- **Top 3 Core features**
 - GROUND LIVING AREA , OVERALL QUALITY, TOTAL BASEMENT AREA
- **Model used : Ridge Regression (89% R2 Score)**
- **Top 3 Neighborhoods**
 - Northridge, Stone Brook and Northridge Heights

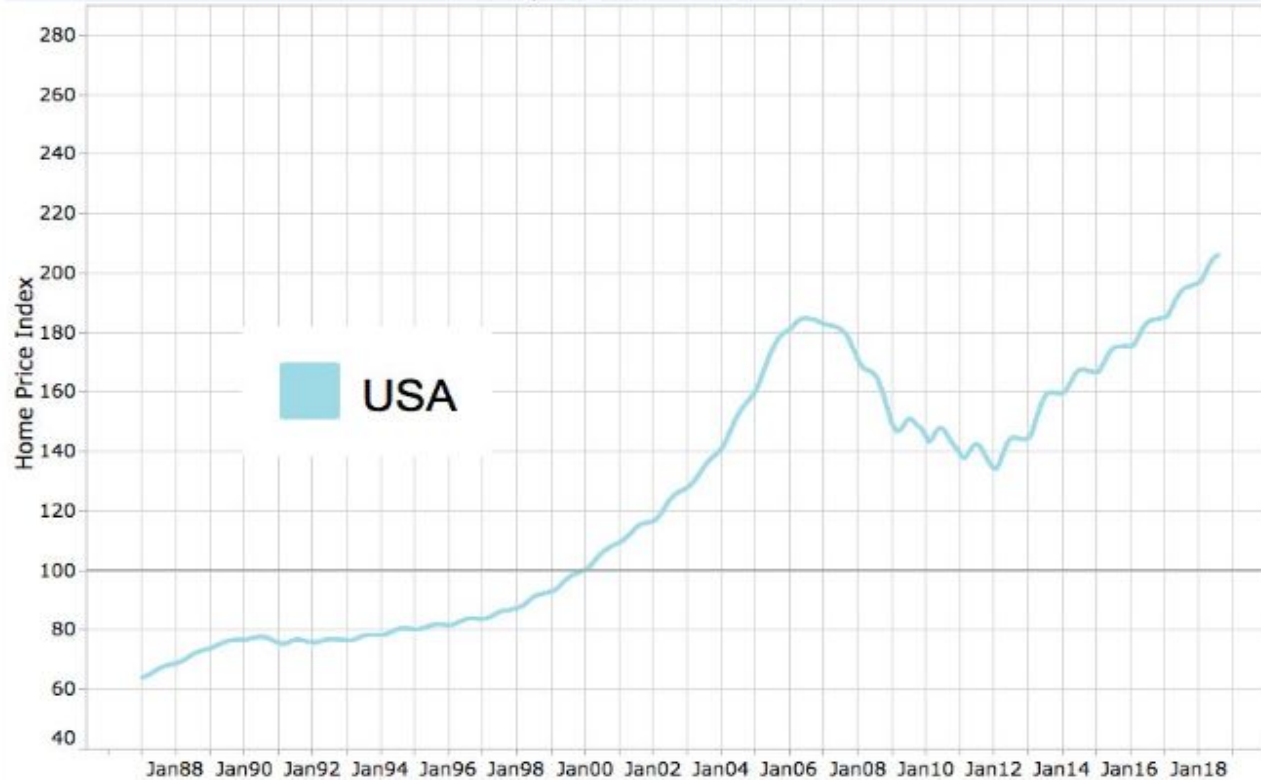


THANKS



Case-Shiller Home Price Index

January 2000 Home Prices = 100



Reference link:

<https://www.forbes.com/sites/johnwake/2018/11/02/the-next-housing-bust/?sh=353094318b79>