House Price Prediction Model & Urban Development

by Suwicha Saeling
& Pichaya Charoonpongsakdi

Our Role:

Data Scientists from the Housing Division, City of Ames

Suwicha Saeling (Bird)
Lead Data Scientist
Housing Division, City of Ames

Pichaya Charoonpongsakdi (Anik)
Data Scientist
Housing Division, City of Ames

1 **Problem Statement**
What and Why?

2 **Model Development**
How?
- Data Wrangling
- Linear Regression Model
- Feature Selection
- Evaluation

3 **Conclusion and Q&A**

Problem Statement

Department of Housing,
City of Ames

"Developing the best actionable & explainable price prediction model based on current available Ames housing dataset 2010"

Model Development Process

Part I

Clean

drop – impute – cast type

EDA

get insight – clean outliers

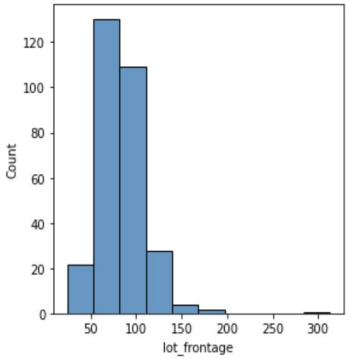
Encode

dummy

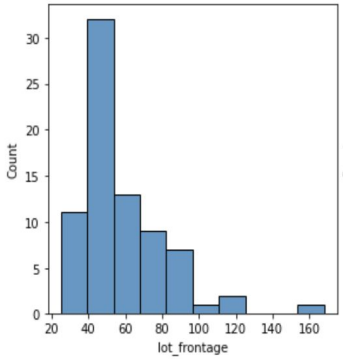
Part II

Data Cleaning – Fill null values

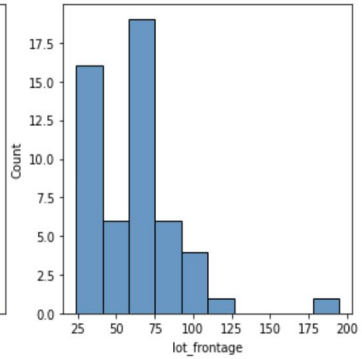
- Fill null values in the columns ‘lot frontage’ with the **median** values of this column when grouped by type of ‘lot configuration’



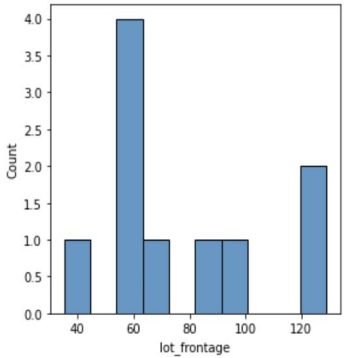
Corner Lot



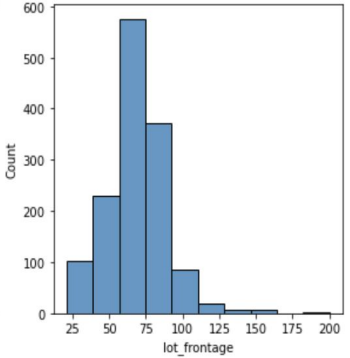
Cul-de-Sac



Frontage on
2 sides of property



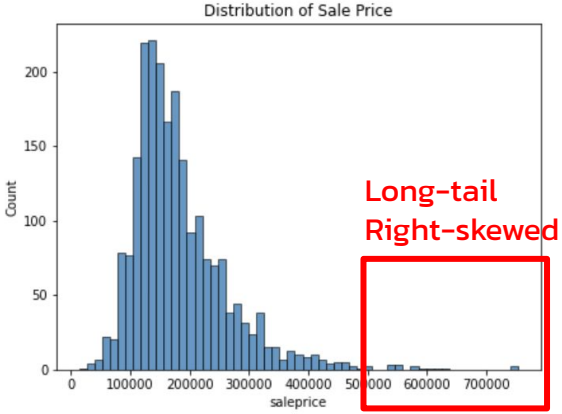
Frontage on
3 sides of property



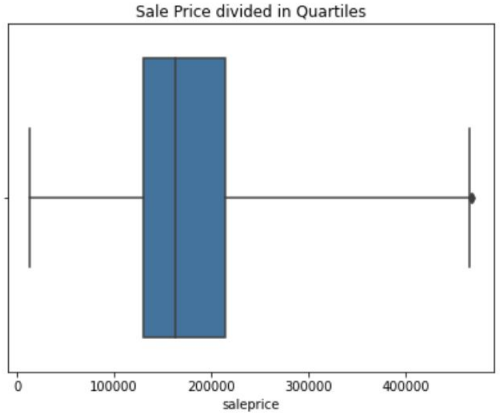
Inside Lot

Data Cleaning – Outliers

Before removing outliers



After removing outliers



Model Development Process

Part I

Clean

drop – impute – cast type

EDA

get insight – clean outliers

Encode

dummy

Part II

82 features

feature selection > correlation

Lasso and Ridge

LINEM assumption

Feature Selection

Location

zoning

surrounding

neighborhood

House Condition

overall condition

year built

number of rooms

garage

House size & Land size

land configuration

total square feet

total usable area

house type

Other factors

pool

porch

sale type

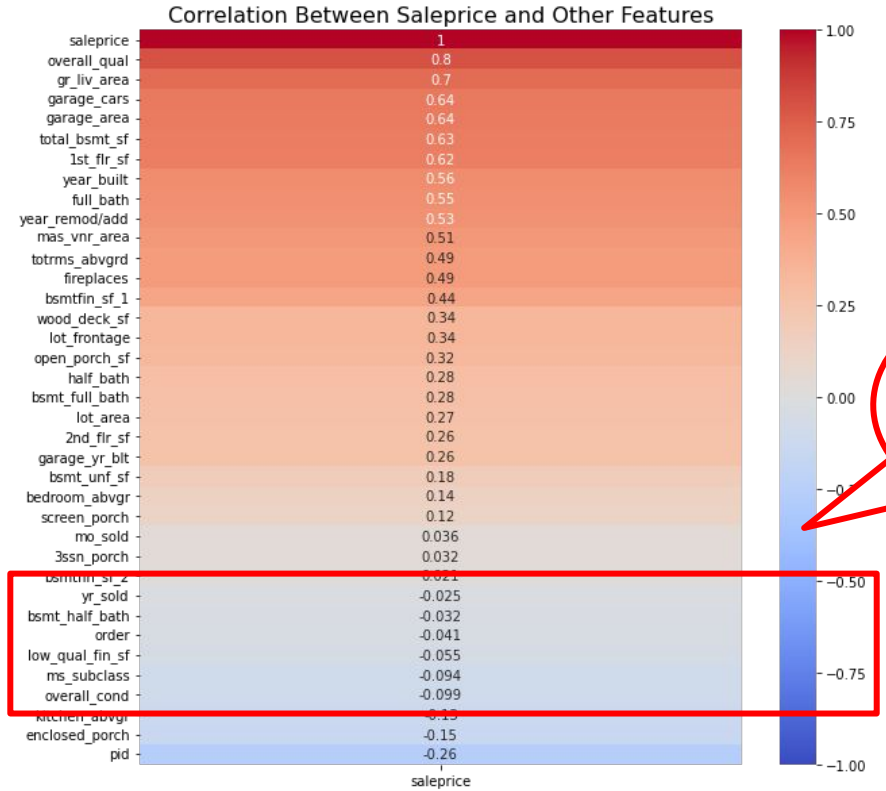
scores

30,000+

Feature Selection – Correlation Scores

score improvement
1,000+

scores
29,000+



Remove factors with correlation lower than 0.10 and higher than -0.1

LINEM assumption

L linearity

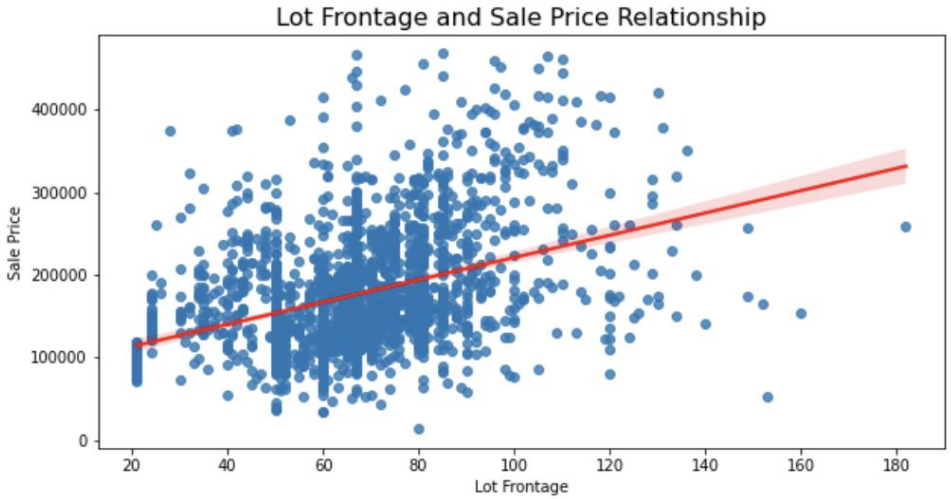
I independence

N normally distributed errors

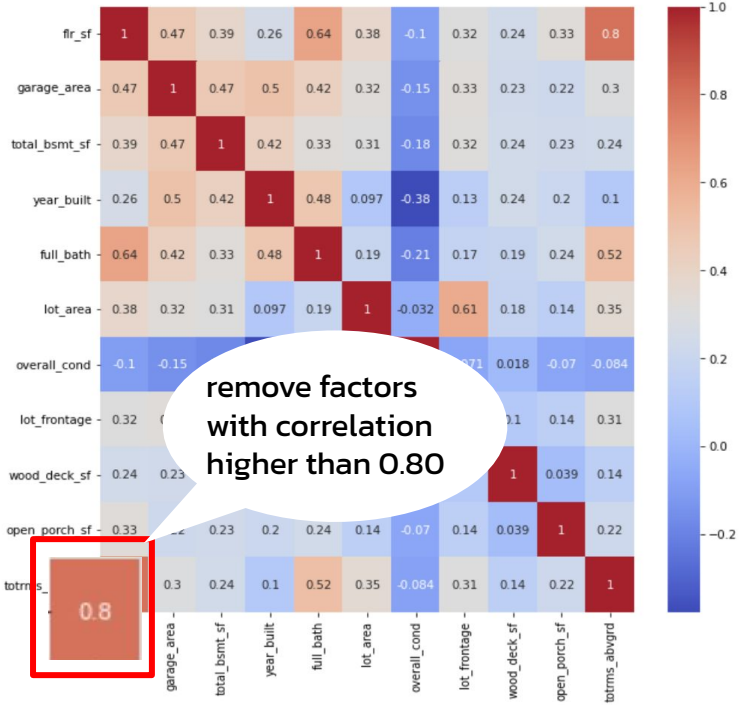
E equal variance

M multicollinearity

LINEM Assumption

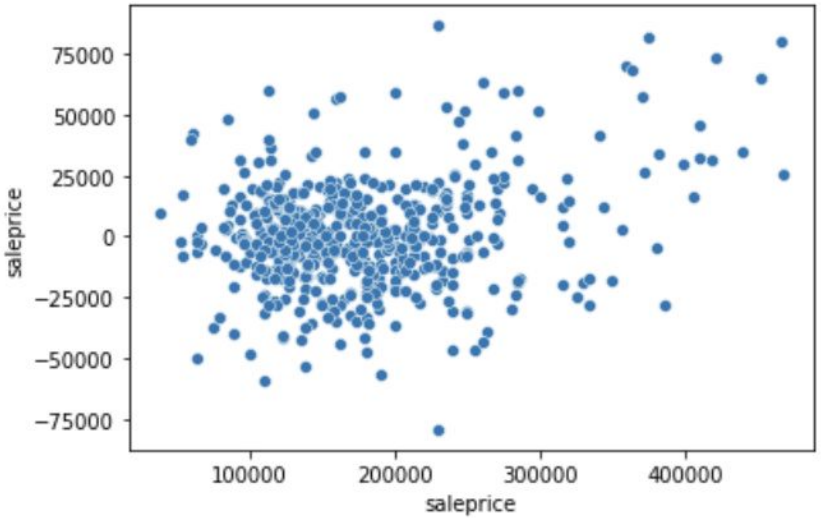


Linearity

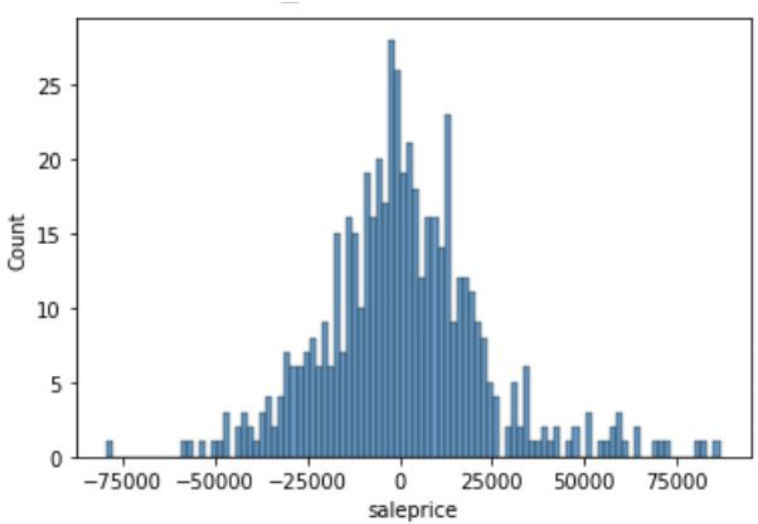


Multicollinearity

LINEM Assumption



Equal Variance of Errors



Normally Distributed Errors

Price Prediction – Final Model

Features

'flr_sf'	'ms_zoning'	'overall_cond'
'garage_area'	'ms_subclass'	'heating_qc',
'total_bsmt_sf'	'neighborhood'	'house_style'
'year_built'	'central_air'	'kitchen_qual'
'full_bath'	'condition_1'	'roof_matl'
'lot_area'	'bsmt_qual'	'functional'
'lot_frontage'	'paved_drive'	
'wood_deck_sf'	'exter_qual'	
'open_porch_sf',	'foundation'	

R Square

91 %

RMSE

28,099

Summary and revert back to problem statement

"Developing the best actionable & explainable price prediction model based on current available Ames housing dataset 2010"

- Equal & Affordable Housing
- Urban Planning
- Appraisal Price

Limitations of Current Model & Outlook on Future Models

Limitations

- Based on physical features
- Lack of contextual analysis/features e.g. proximity to schools, supermarkets, current supply and demand, etc.
- Data is out-of-date
- Less predictive power as compared to other supervised machine learning models

Outlook on Future Models

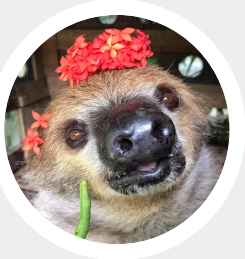
- More contextual features e.g. schools, hospitals, supermarkets, etc.
- Collect new data to keep the model up-to-date
- Price Prediction with other supervised learning models

Thank you!

Q & A

Meet the team!

Suwicha Saeling (Bird)
Seasoned Python Auditor
- Watch me out!



Pichaya Charoonpongsakdi (Anik)
The Russian Spy
- Don't mess with me!

Nong Lumyai the Great
Head of Emotional Support
- Love everyone to the moon and back! Kiss kiss <3

Let’s make a better life for LUMYAI !