Ames House Price Prediction

Group 5: Emma, Rohazeanti, Mingzi

Data Science Problem

Objective: Predict the price of housing in Ames, with higher accuracy and more features

We are a team of data scientists engaged by a real estate company to create a Machine Learning model that predicts the sale price of residential properties in Ames, Iowa with higher accuracy and more features as compared to available apps in the market and deploy it on an Application Programming Interface (API) for their agents.

Dataset Used for Prediction

80 explanatory variables

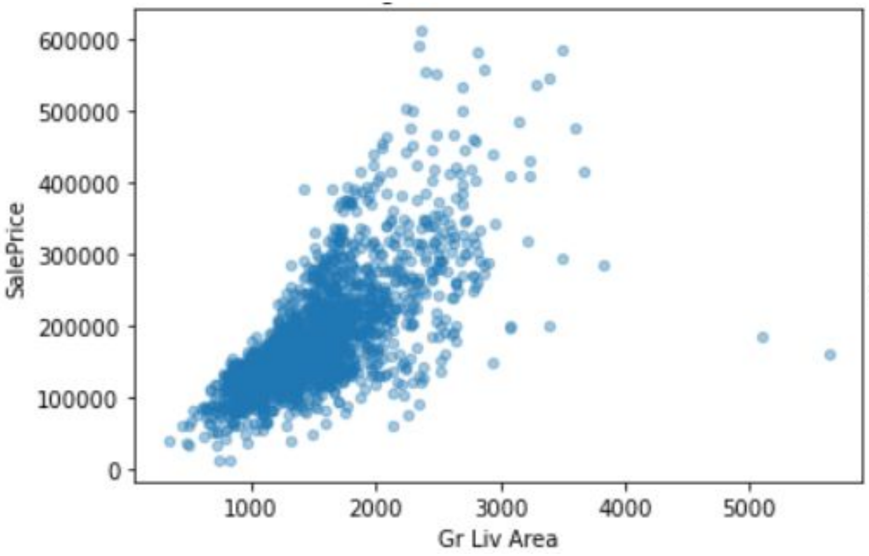
- Qualitative Features (eg quality and condition)
- Area specific Features (eg porches, bath, kitchen)
- Size Related Features (eg. total basement in sq ft, above grade (ground) living area, garage area)
- Location (eg. ms zoning and neighbourhood)
- Date-related (eg. year built, garage year built, year sold)

The features are also classified into the following data types:

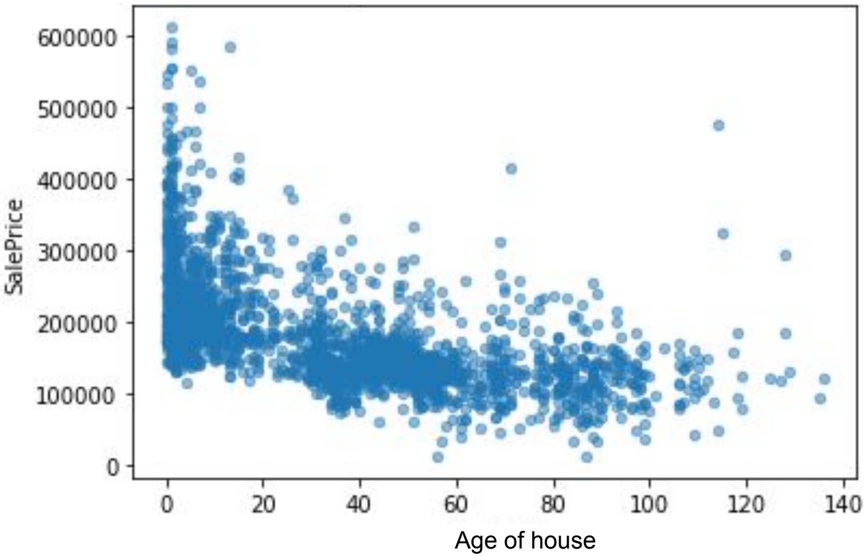
- Numerical (eg total basement square feet)
- Categorical (Ordinal) (eg overall condition)
- Categorical (Nominal) (eg. house style)

Overall Trend of Dataset Observations

Sale Price vs Gross Living Area

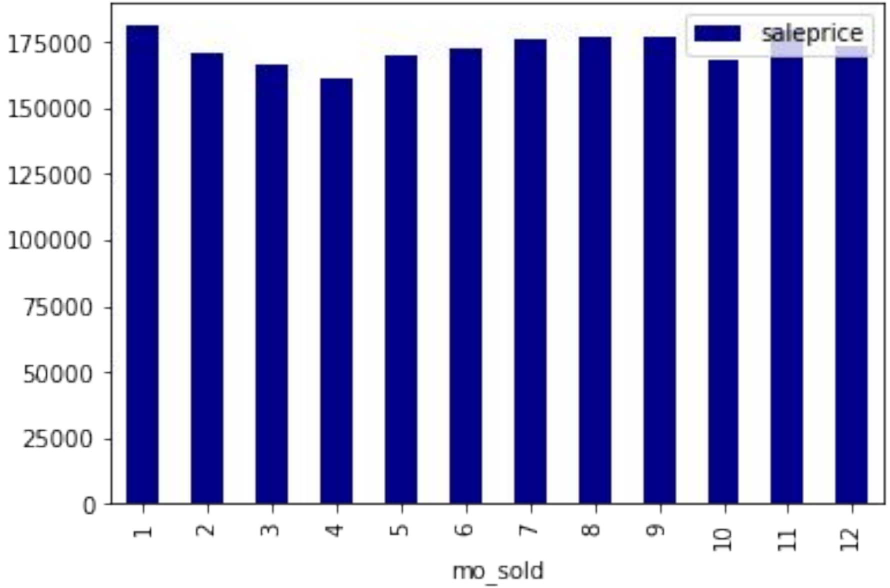


Sale Price vs House Age

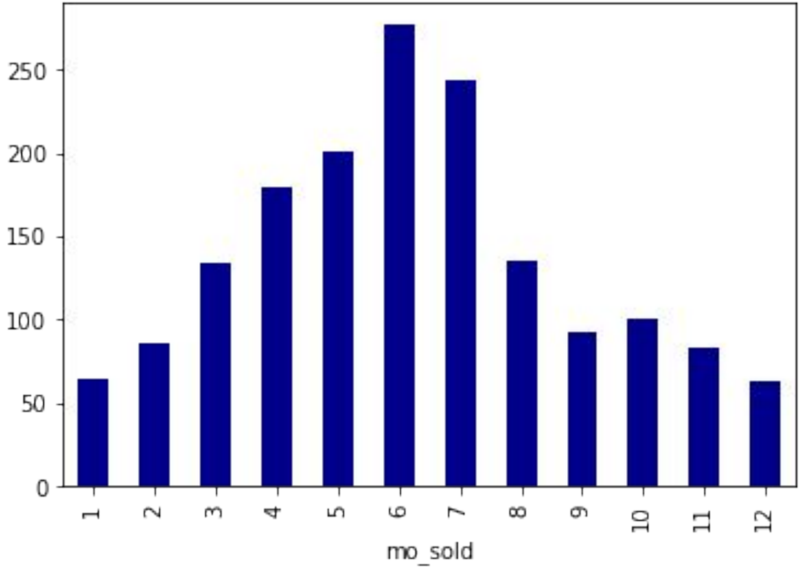


Overall Trend of Dataset Observations

Month sold vs Mean Sale Price

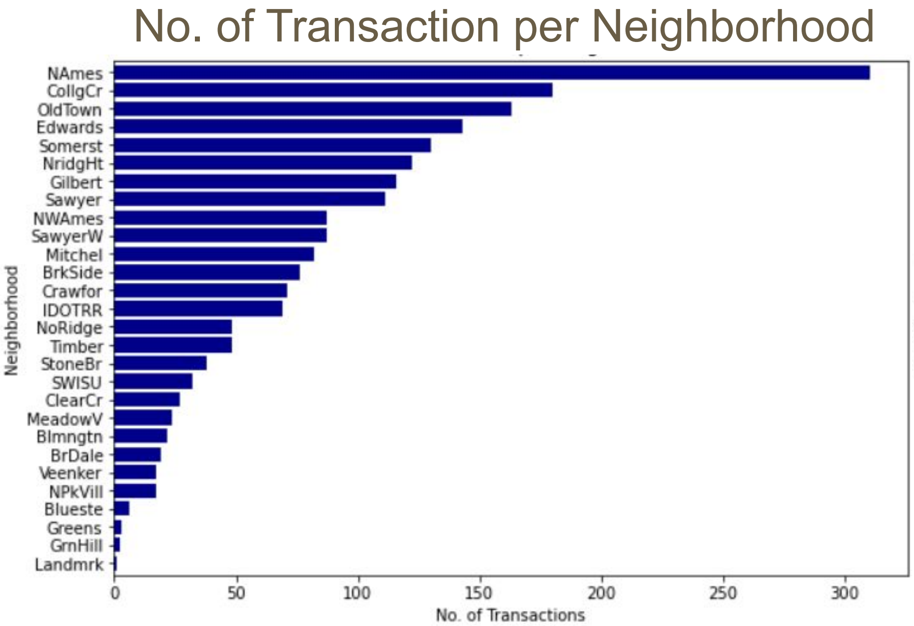


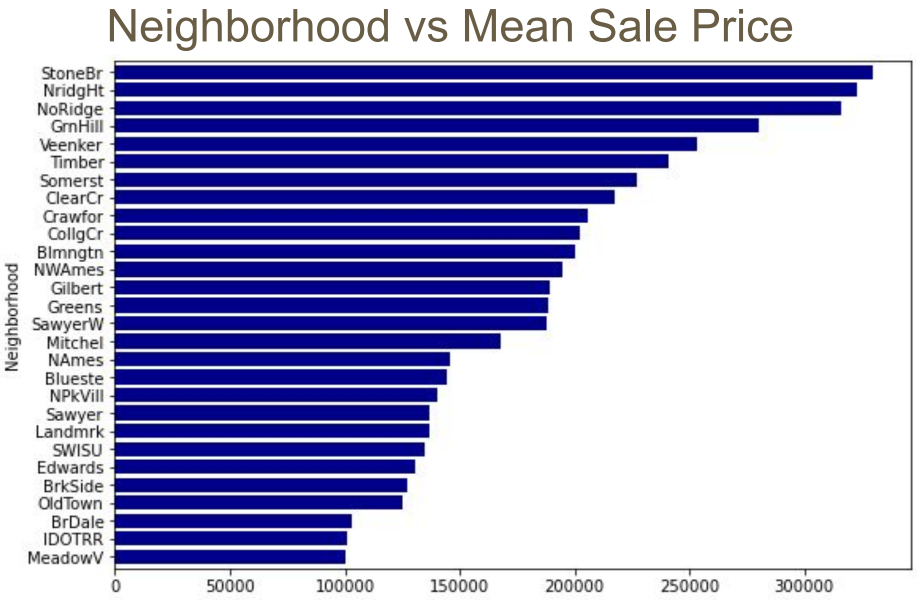
Month vs No. of Transaction



- Higher amount of transactions during summer
- Does not affect mean Sale Price

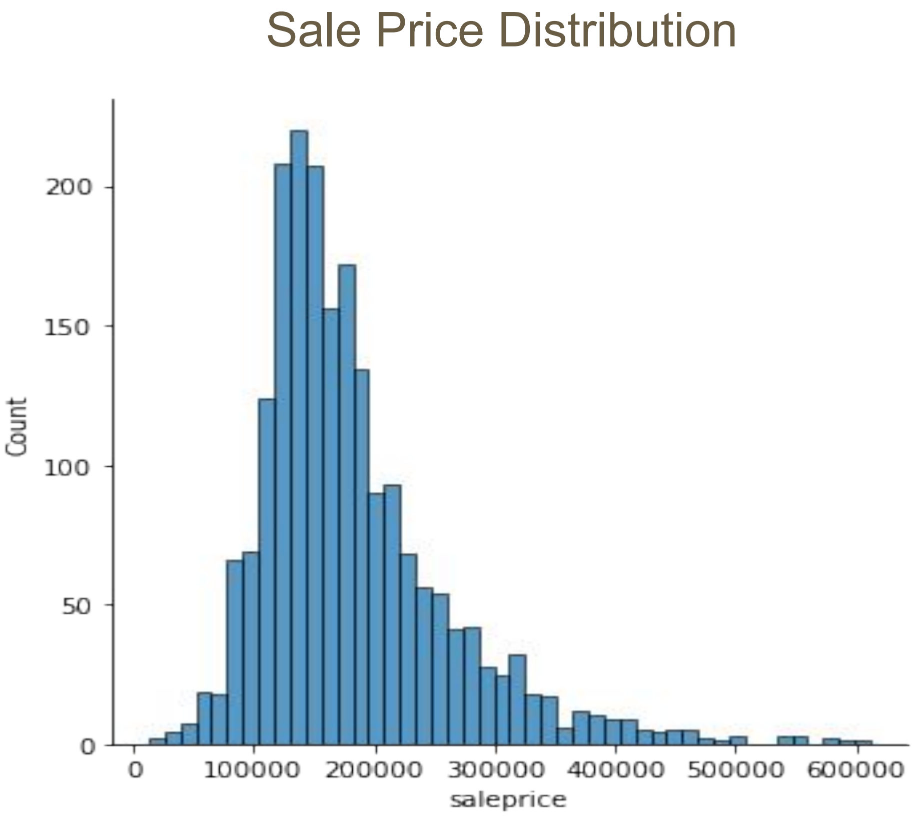
Overall Trend of Dataset Observations





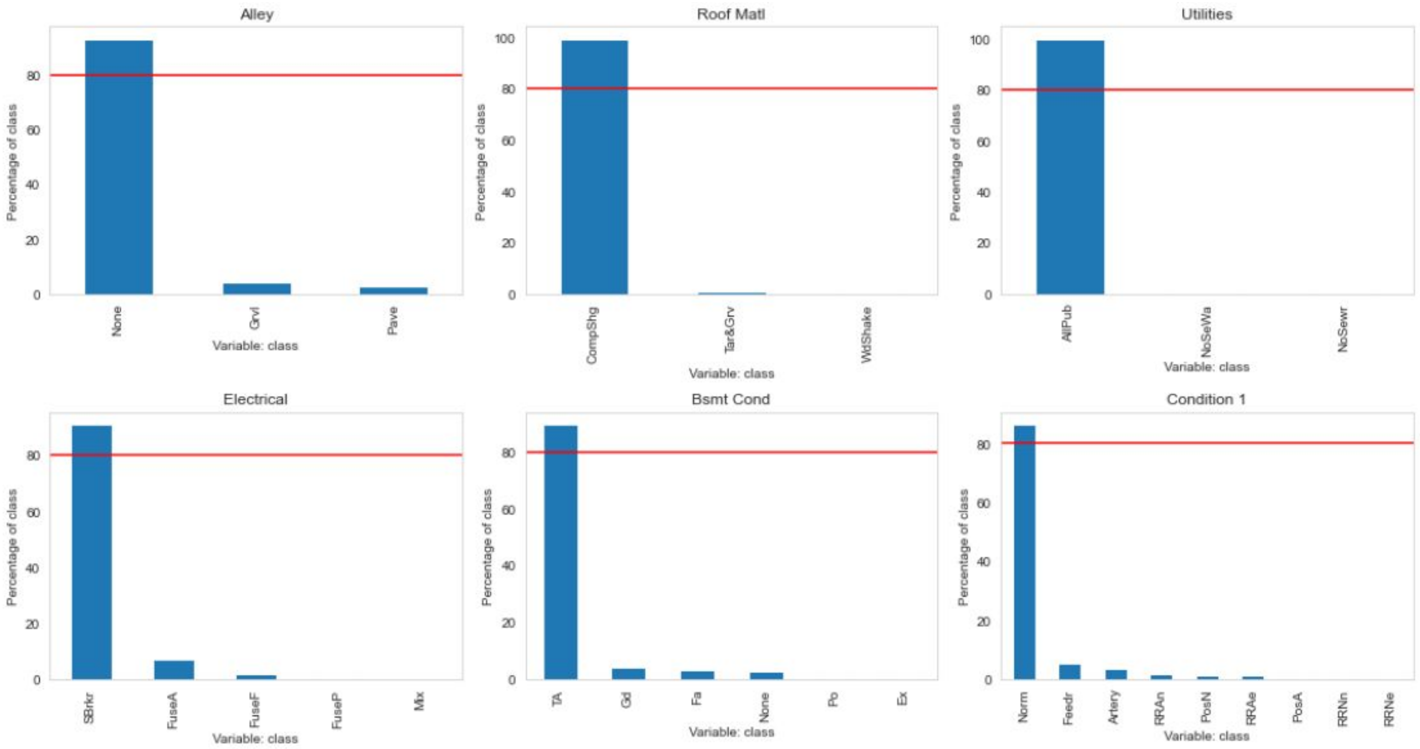
- Northwest Ames has the high amount of transactions
- Most of the sales happened for residential low and medium density area

Visualizing target variable - Sale Price



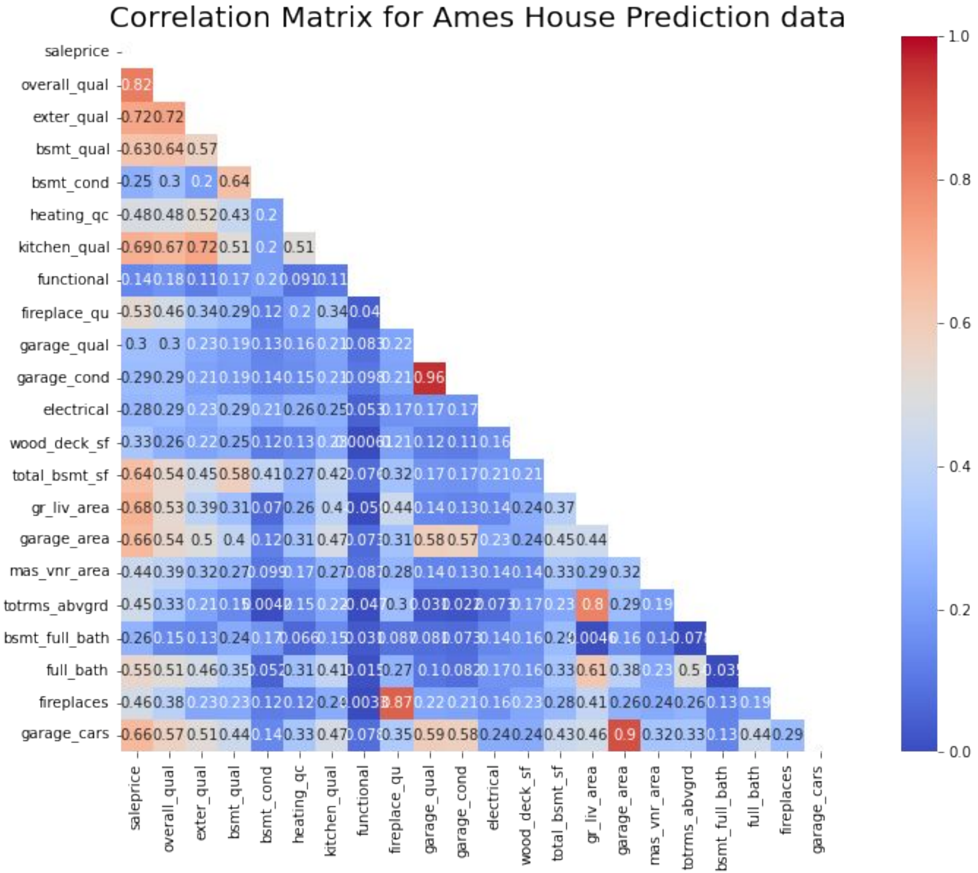
- Majority data capture between 100k to 330k
- Right-skewed, potential outliers

Feature Selection: High Frequency Class



Threshold set at 80%

Features Selection



- Selection based on top few correlated features (corr >0.50)

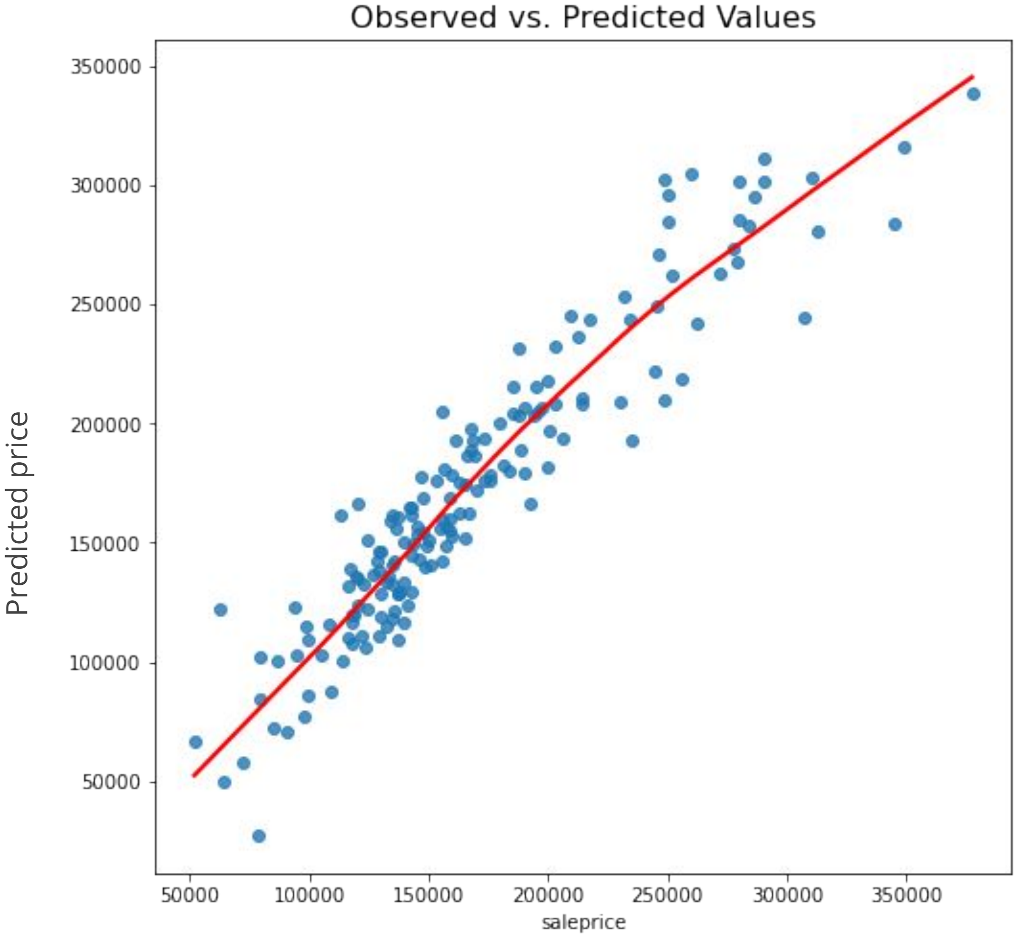
Features in the prediction model

Categorical (Ordinal)	- 'Overall_qual' - 'Fireplace_qu'
Categorical (Nominal)	- ‘Lot_config', 'Lot_shape', 'Land_slope', 'Land_contour' - 'Ms_zoning', 'Neighborhood', 'Bldg_type' - 'Street', 'House_style’, 'Roof_style', 'Exterior_1st’, 'Exterior_2nd', 'Mas_vnr_type', 'Foundation' - ‘Central_air', 'Paved_drive' - 'Sale_type'
Numerical	- 'Lot_frontage' - 'Lot_area', 'Garage_area', ‘Mas_vnr_area' - 'Total_bsmt_sf', 'Gr_liv_area' - 'Fireplaces', 'Full_bath' - 'Age'

Model Selection & Performance

	Linear Regression	Ridge (alpha = 1)	Lasso (alpha = 1)	Elastic Net (alpha = 0.01, l1_ratio = 0.4444)
MSE (Train)	453705780.33	455767752.73	453838558.22	470576542.68
MSE (Test)	461692464.30	429791086.29	452407612.99	421405745.57
Percentage difference	-1.76%	5.69%	0.32%	10.44%

Actual VS Predicted Sale Price



Deployment of ML Model

Web based deployment using **Streamlit**.

Process:

- Load and apply saved feature engineering
- Load and apply saved the best model
- Select each value of features using a slider window
- Make prediction button

Let’s try out our prediction webpage!

Streamlit

Demo Time

Limitations of the model

- Developed using data on houses sold between 2006 - 2010 (4 years)
  - Not enough to capture annual changes in Sale Price
- Housing prices at present (in 2022) may have also changed as a result of **inflation**
- Only specific to houses in Ames, model may not be applied to other city or country
- Prices do not take into consideration of social factors (i.e. crime rates, demographics in the neighborhood)
- Assumption of independence variables
- Good at predicting the price of properties within the range of $70,000 to $225,000

Conclusion

- Current application in the market: Zillow.com, Trulia
  - Provides estimates on the market value of property
- Our model is specifically build for **Ames, Iowa**
- More features as compared to others
- Complimentary tool on top of Comparative Market Analysis (CMA) reports
  - More confident in their proposed prices with **improved accuracy based on more features**
- Recommendation
  - Use better model such as XGBoost
- Future project consideration
  - Feature importance - from the model, see which coef are greater and identify those features.
  - Property survey portfolio: Property surveyor to evaluate the price of houses.
  - Extend to banks to establish the asset value of the houses