Ames Housing Prices Analysis

Team: GGGG
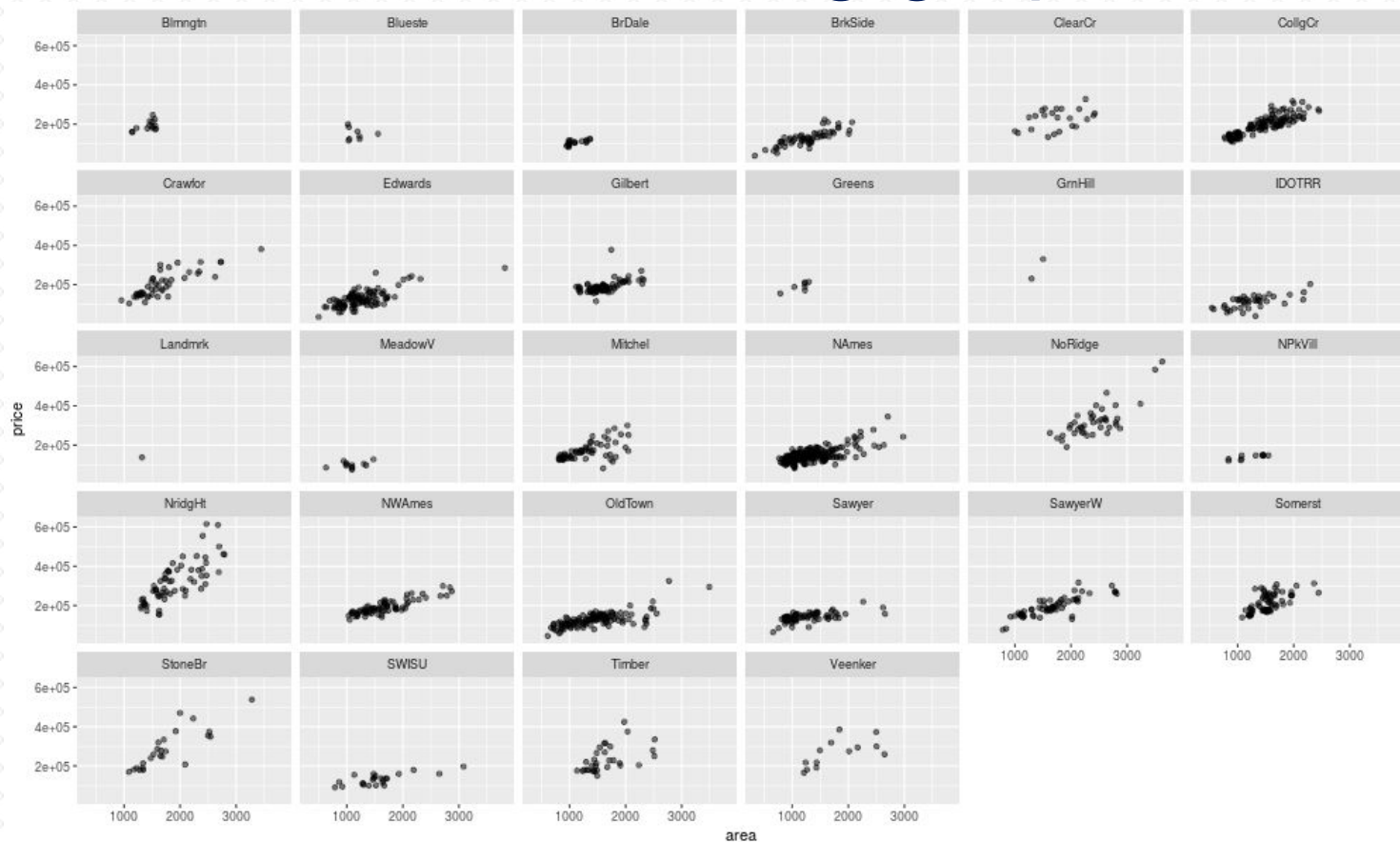
Yilin Gao   Mengrui Yin   Wenqi Cheng   Liyu Gong

Duke
UNIVERSITY

Most interesting graphic

Best Model

- Motivation: Need a Simple, Interpretable model for brokers
- Model: LASSO with interactions between area and house quality measures
- Reason:
    - Linear model (easy interpretation)
    - Similar performance on training and testing dataset (no overfitting)

Best Insights into predicting Sales Price

- Houses in the Green Hills neighborhood is 60% higher than the baseline
- Every 100 sq ft increase in area have different effect on price:
  - Brokers:
    - Avoid North Ames area
    - Sell 1 story or split foyer houses
  - Buyers:
    - Avoid Brookside, Stone Brook, Northridge Heights neighborhood
    - Buy houses with brick common masonry veneer type, mansard roof type

2 Best Houses to purchase

- Based on the difference between the **predicted price** and the **actual price** in the *testing set*, the 2 most **undervalued** houses are:
  - PID = 906402200, diff = 60760.297
  - PID = 533251110, diff = 47936.105
- Take other factors into consideration besides the model:
  - Hidden variables in determining house prices
  - Buyers' preferences