

An investigation Ames Housing SalePrices

Sujay Jangam, GA-DSI 27

Problem Statement.

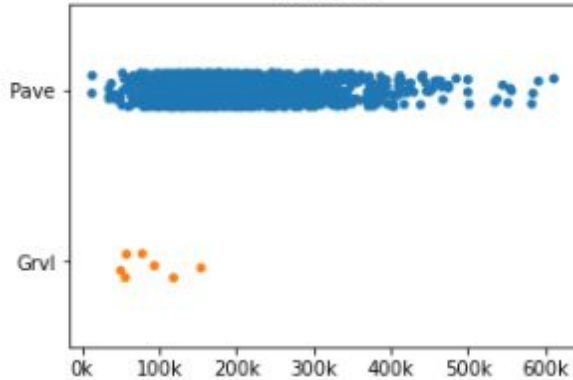
- A team of data scientists, which I am a part of, have been hired by an interior design and renovation firm.
- We hope to provide recommendations on how this firm improve property value.

Missing Data and Outliers.

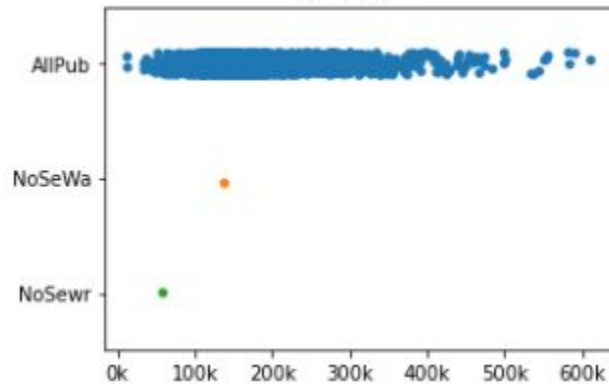
- There were many missing values in the dataset, a total of 9,822 values.
- Mostly managed.
- 'Lot Frontage' was imputed.
- 6 rows of Outlier data, dropped

Feature Selection – drop.

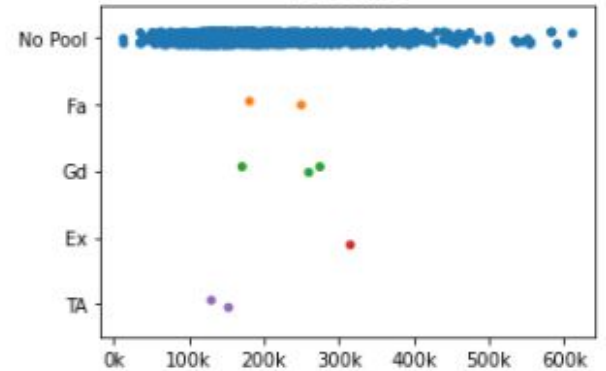
Street



Utilities

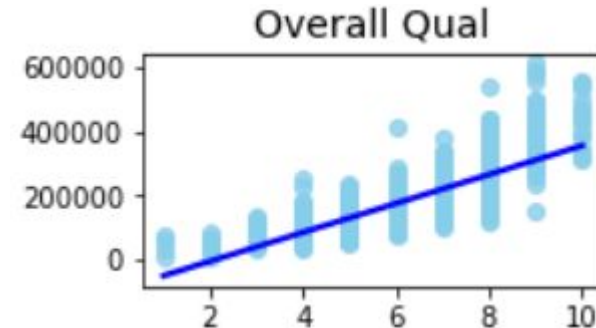
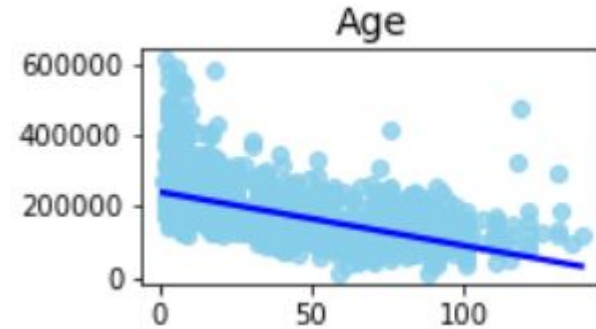
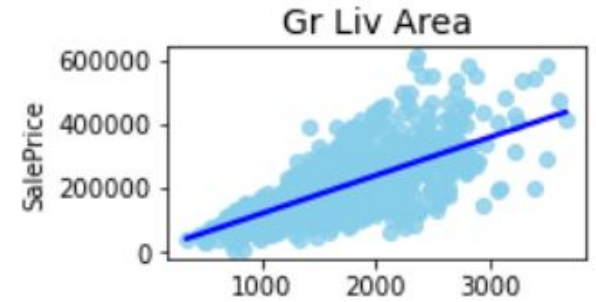


Pool QC



Feature Selection.

- Based on visualisation and correlation scores, we selected groups of feature such as:
 - Square Footage
 - Quality
 - Location
 - Others: e.g. Age

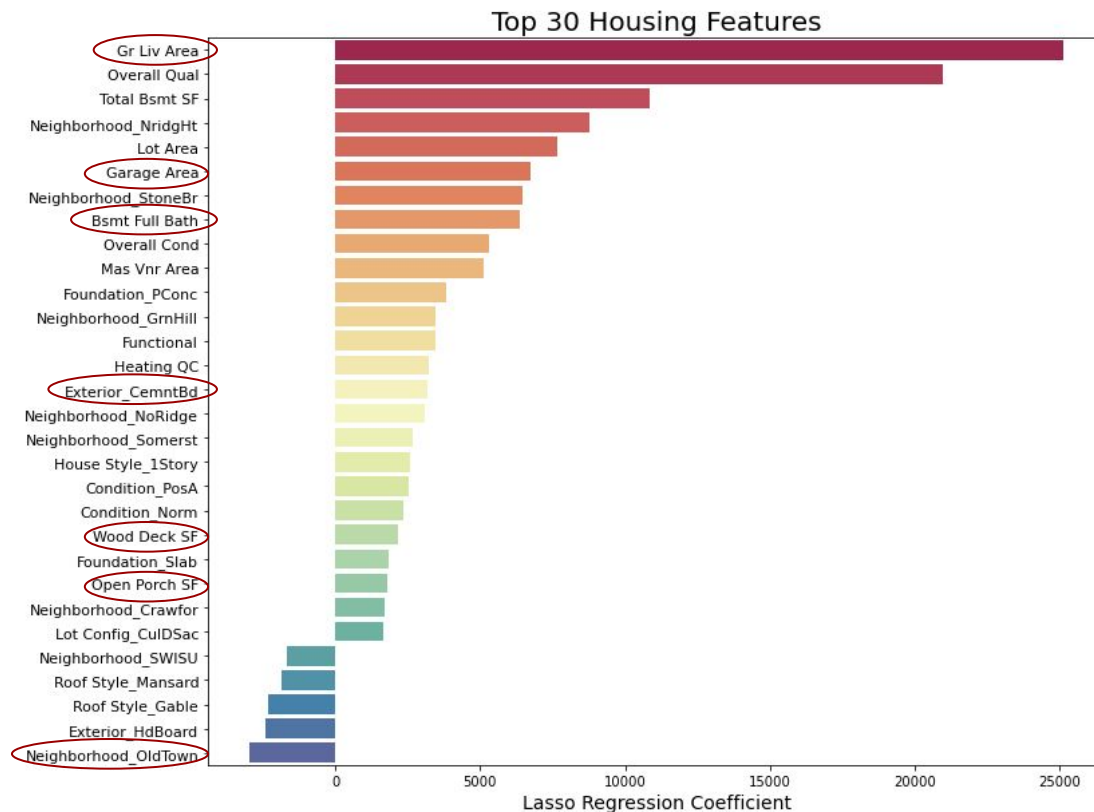


Model Accuracy.

- Best fit line passes through most of the points
- Outliers



Some Encouraging Figures



Conclusion and Recommendations:

Recommendation:

- Overall and exterior quality of home
- Heating Quality and Heating type
- Switch to a cement, brick or vinyl siding, wood shingles exterior. Avoid hardboard, wood siding, or plywood.
- Garage size.
- House style
- Basement Square Footage
- Bathrooms in the basement

Conclusion and Recommendations:

Recommendation:

- Home expansion (1 Storey)
- Overall and exterior quality of home
- Heating Quality and Heating type
- Switch to a cement, brick or vinyl siding, wood shingles exterior. Avoid hardboard, wood siding, or plywood.
- Garage size.
- House style
- Basement Square Footage
- Bathrooms in the basement

Model Shortcomings and Improvements.

Shortcomings

- Generalizes well to the city of Ames
- Inflation not taken into account

Improvements

- All data filled in (missing values)
- Use of Algorithms
- Interactivity between factors