Pillow Predictions

"Keep your money and sleep, let Pillow predict."

Sibel Tanoglu

Problem Statement

★ Create a regression model based on the Ames Housing Dataset in order to predict house prices for real estate market

Goal

★ Improve user satisfaction of Pillow app in Ames by providing more accurate predictions for the area

Data

➔ Data set contains information from the [Ames Assessor's Office](#) used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

➔ 2930 observations, 82 variables.

➔ Location, Area measurements, Lot / Land, Age, Appearance, Roof, Garage, Kitchen, Rooms / Bathrooms, Utilities, External (Deck, pool, porch etc.)

Workflow

Understanding data types, missing values, relationships, inconsistencies, correlations

Imputing missing variables, transforming and scoring, derived and interaction variables, feature engineering

Understanding the performance of the model, testing accuracy

Explore → **Visualize** → **Manipulate** → **Model** → **Evaluate**
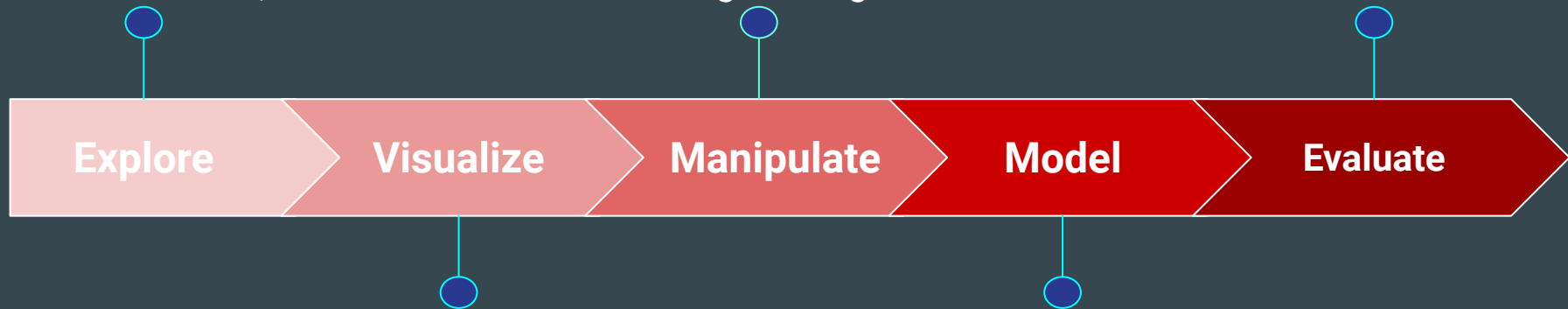
Histograms, Scatterplots, Box Plots, correlation matrix, heatmaps

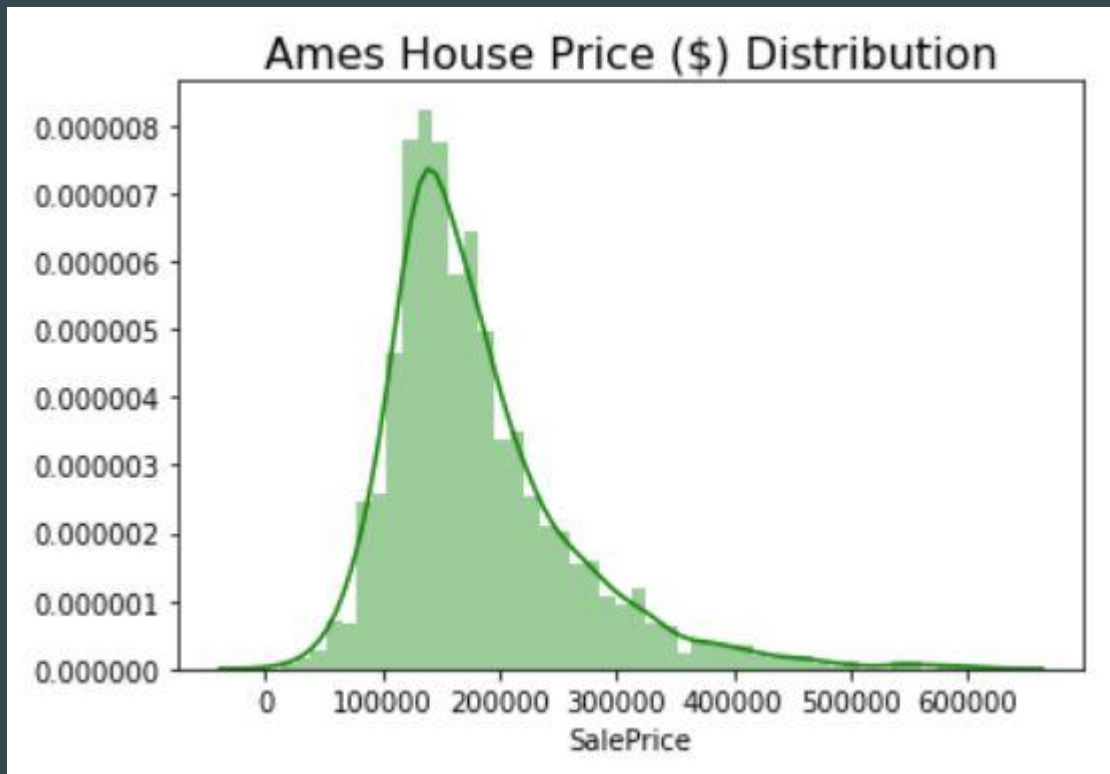**Multi-Linear Regression, Cross Validate Function, Scaling, RidgeCV, LassoCV**
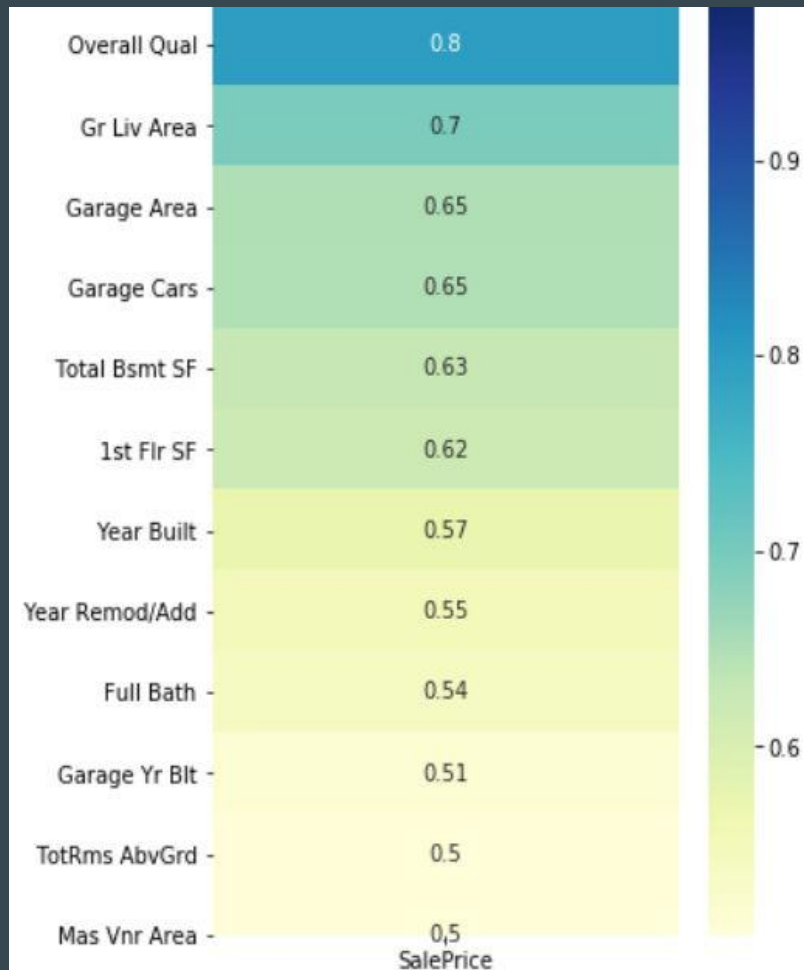
House Price Stats ($$)

Mean: $181,469

25%: $129,825

Median: $162,599

75%: $ 214,000

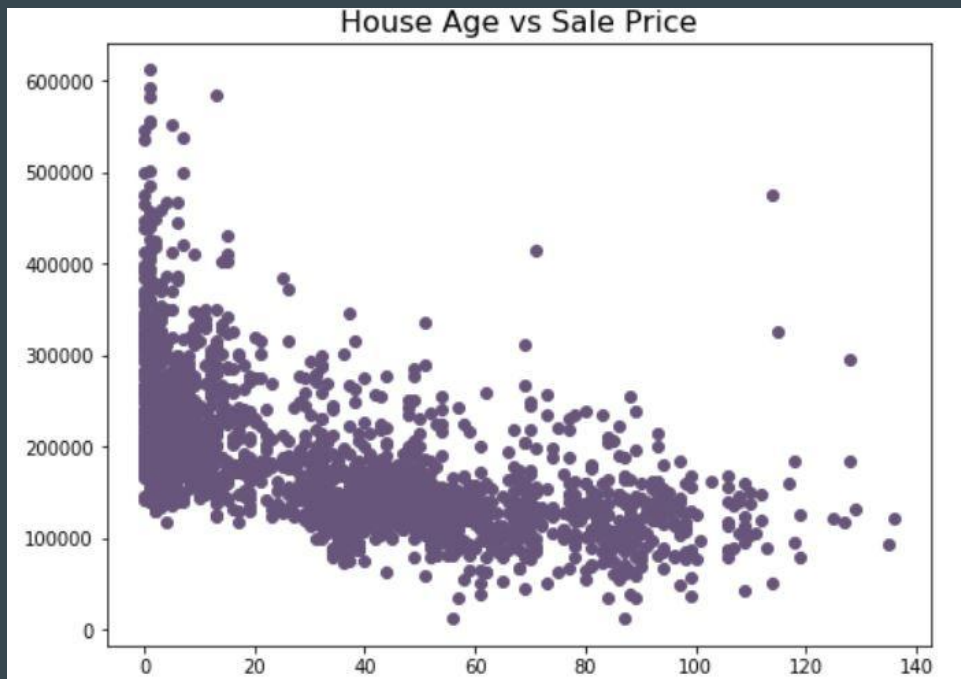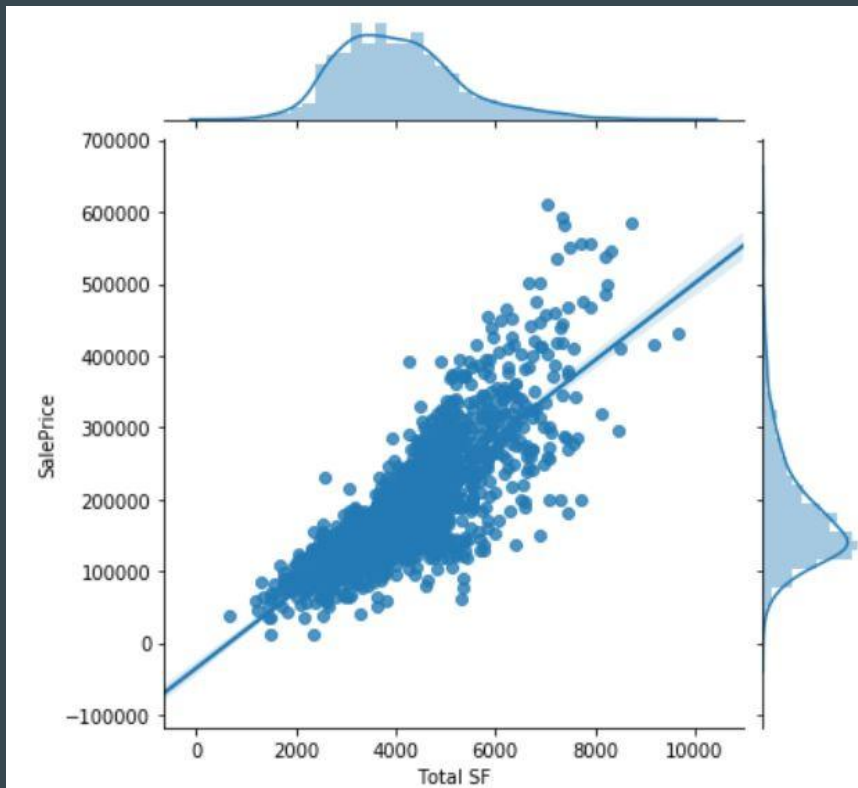Ames House Price ($) Distribution
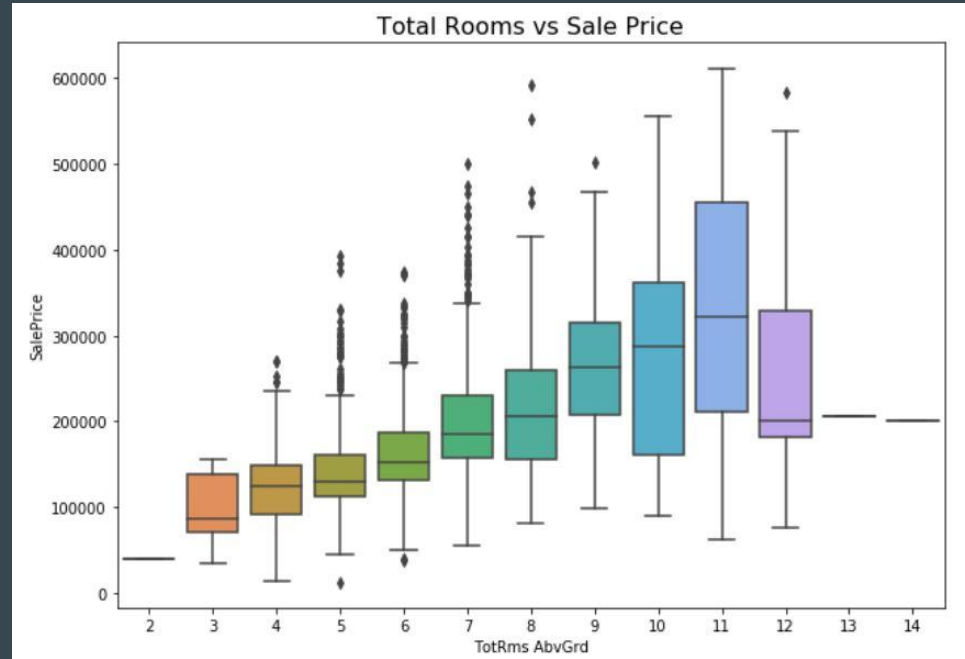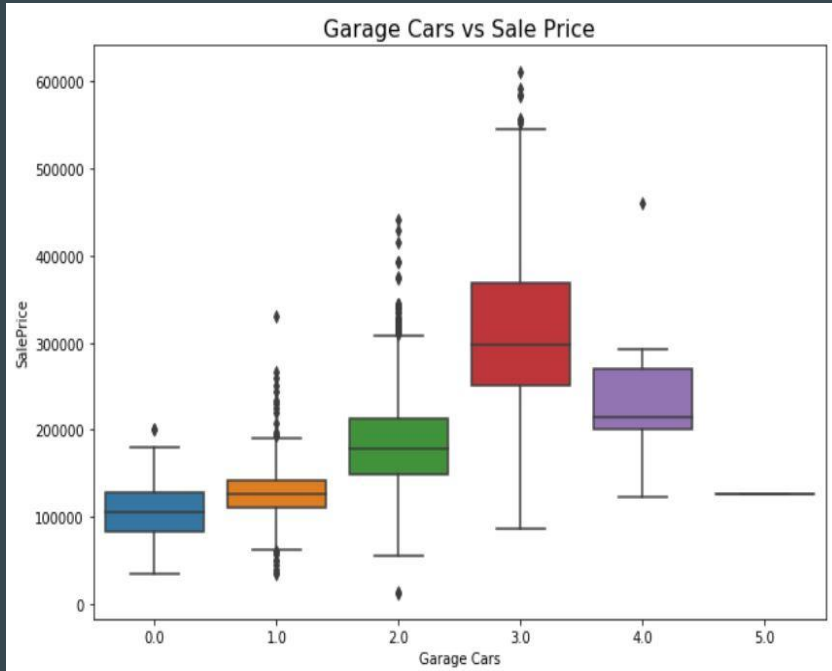
Correlations (%%)

Quantitative features correlated with Sale Price higher than 50% ratio

Living Area and House Age are most important

How many is too many?

QC: Quality and Condition are inseparable

Interactions between "Quality" and "Condition" variables increased our model's performance significantly.

- Garage QC
- Overall QC
- External QC
- Basement QC

Feature	Correlation Sale Price %
Garage Condition	0.2651
Garage Quality	0.2853
Garage QC	0.6672

Landscape-ability is not a thing in Ames

- Combined Lot Shape * Land Contour regularity and scored 1 - 16
- Every 1 point drops house value by $1,139.84

- Negative influence on price,

- Remove from Ames model and keep for other cities?

Model Predictors

Feature Coefficients (Scaled)
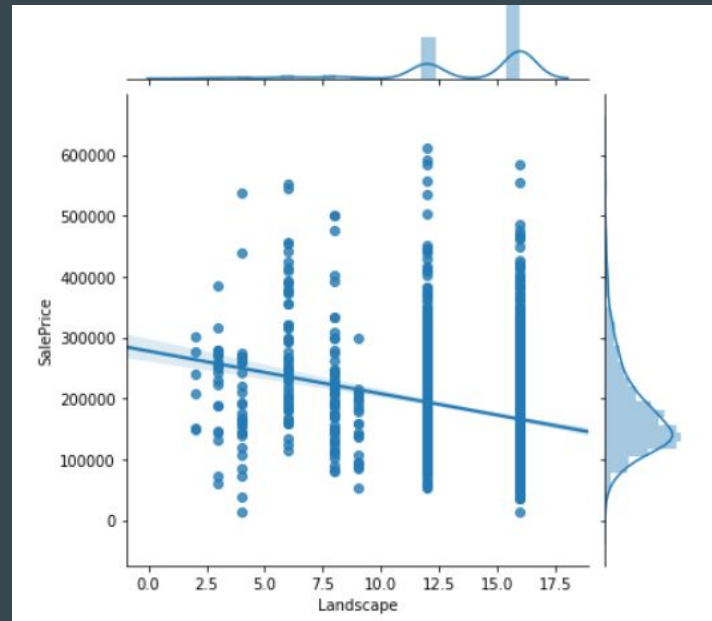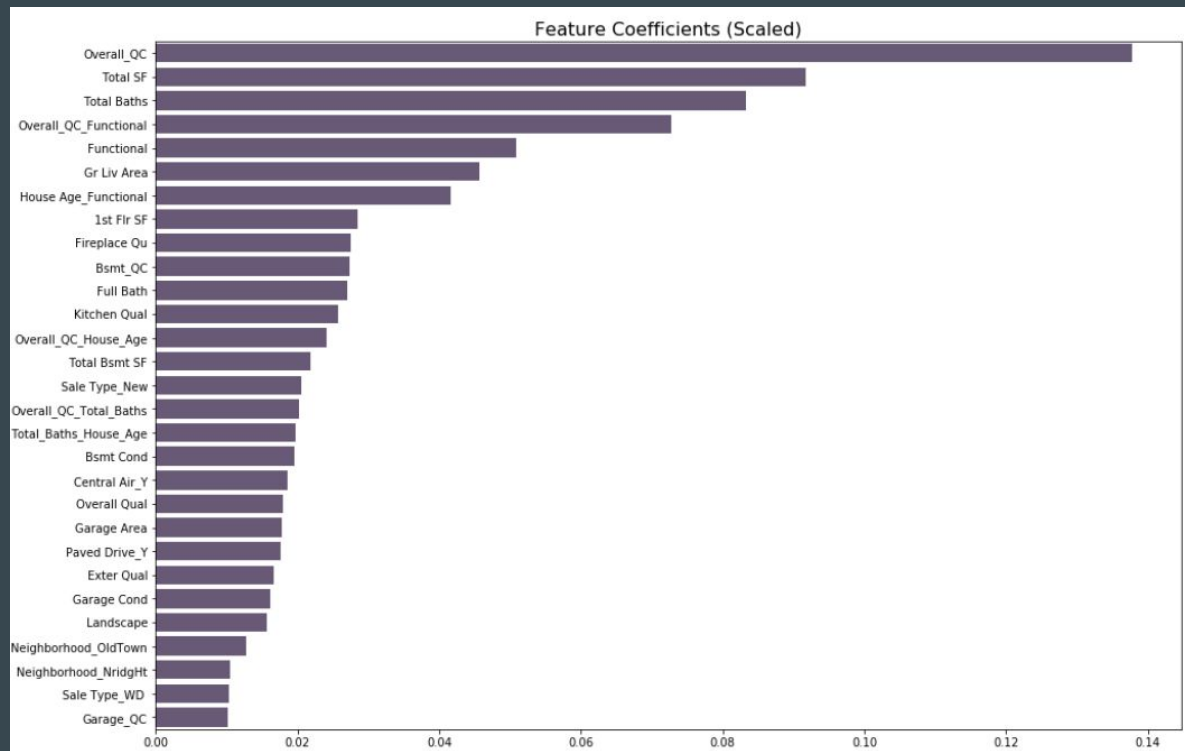
- ❏ Standard Scaler to regularize the features

- ❏ Lasso regression to zero out unrelated features

91.6% variance in house price can be explained by model

- ❏ R2 score: **0.916**

- ❏ R2 score on holdout: **0.898**

- ❏ RMSE on unseen data: **23,463**

Summary

- Regularize various different units (Sqft, Age, Number of rooms, etc.)

- Optimize feature selection

- More localized model for different cities (e.g landscape)
- More observations will improve learning, provide more accuracy
- Check dependency between predictors and remove unnecessary

Thank you...

Sources

❏ http://jse.amstat.org/v19n3/decock/DataDocumentation.txt

❏ https://pixabay.com

❏ https://homeguides.sfgate.com/landscaping-increase-property-value-7301.html

❏ https://nationwide-appraisal.com/what-determines-the-quality-condition-of-your-property/

❏ https://stackoverflow.com/