Hello!

I am Katie Sylvia

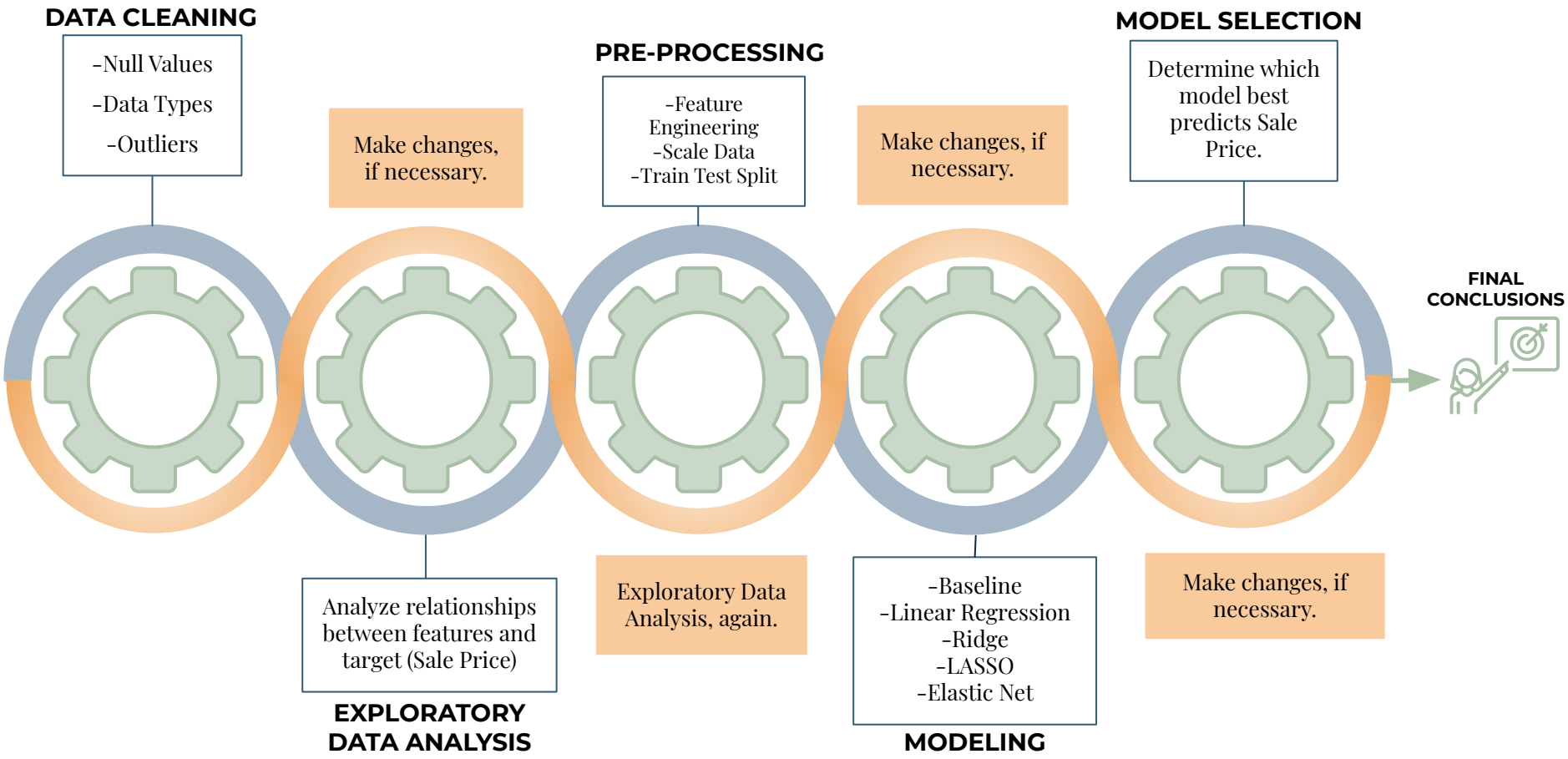
I have been hired by Ames Realty Co. to build a model that best determines sale prices for properties in Ames, Iowa.

The Ames Housing Dataset:

PREDICTING HOME PRICES WITH LINEAR REGRESSION

Workflow Process

The Journey to Linear Regression is Not So Linear


DATA CLEANING
-Null Values
-Data Types
-Outliers
Make changes, if necessary.
EXPLORATORY DATA ANALYSIS
Analyze relationships between features and target (Sale Price)
PRE-PROCESSING
-Feature Engineering
-Scale Data
-Train Test Split
Exploratory Data Analysis, again.
Make changes, if necessary.
MODELING
-Baseline
-Linear Regression
-Ridge
-LASSO
-Elastic Net
MODEL SELECTION
Determine which model best predicts Sale Price.
Make changes, if necessary.
FINAL CONCLUSIONS

EDA Discoveries

Top Numerical Features that Increase Sale Price

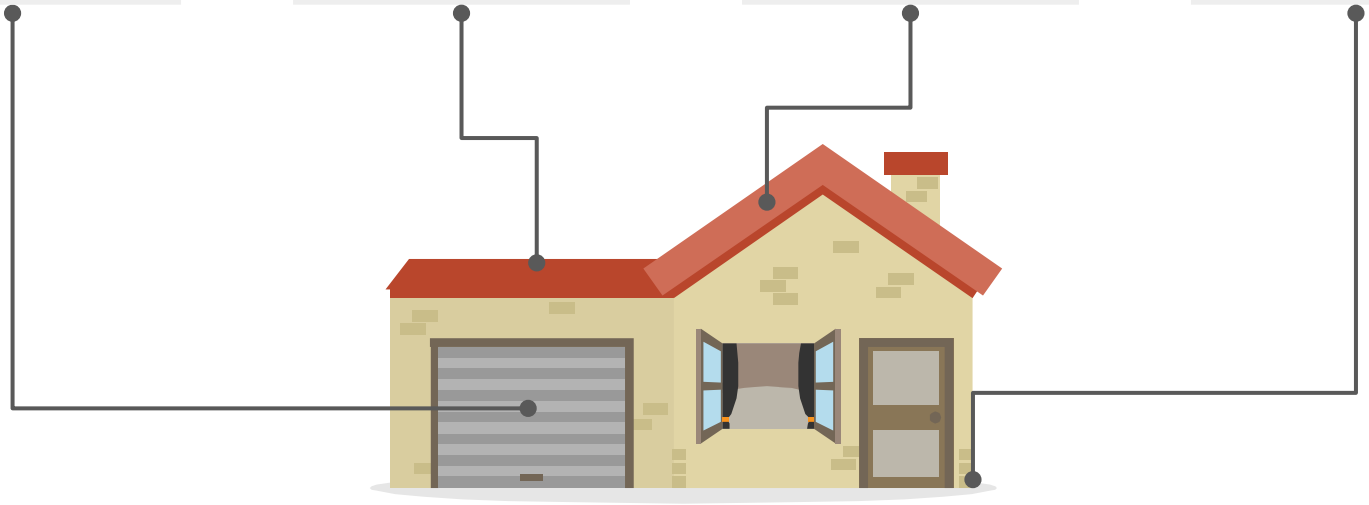
01	**Overall Qual**	Overall material and finish quality
02	**Gr Liv Area**	Above ground living area (square feet)
03	**Exter Qual**	Exterior material quality
04	**Kitchen Qual**	Kitchen Quality

Top Categorical Features that Increase Sale Price

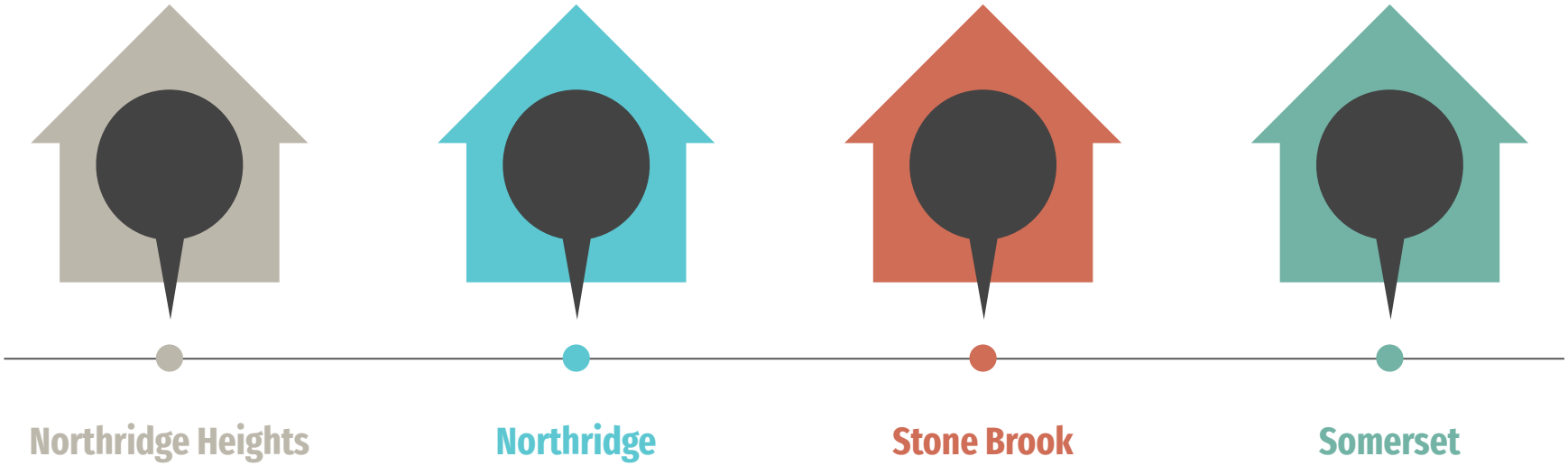
01	**Foundation**	Poured Concrete
02	**Garage**	Finished Garage
03	**Building Class**	2-STORY 1946 & NEWER
04	**Exterior Covering**	Vinyl Siding

Top Features that Decrease Sale Price

Garage Type	Condition 1	Roof Material	Land Contour
Unfinished	Within 200' of East-West Railroad	Gravel & Tar	Flatness of Property - Low Depression



Neighborhoods with High Correlation with Sale Price



Northridge Heights

Northridge

Stone Brook

Somerset

Modeling and Selection

Model Results

Model Type	Training Score	Testing Score
Baseline	0.0	0.0
Linear Regression	0.936465	0.916183
Ridge	0.93199	0.921822
LASSO	0.936446	0.916870
Elastic Net	0.932014	0.921822


Comparison of Actual Price to Predicted Price
Predicted Price ($)
Actual Price ($)
600000
500000
400000
300000
200000
100000
0
0
100000
200000
300000
400000
500000
600000

For each prediction made with the **baseline model** (predicting all sale prices as the median sale price), the price is likely off by:

$68,946.92

For each prediction made with the **my model** (using Ridge Regression), the price is likely off by:

$19,933.64

71%
Improvement

Conclusions

Using the final model, I am able to account for approximately 92.2% of the variation in Sale Price. With a 71% improvement from the base model, my model proves to be an efficient method of determining sale prices of homes in Ames, Iowa.

With more time, an adjustment to the model would be made to increase the accuracy of predicting the sale prices in more expensive homes.

I hope you, Ames Realty Co., are satisfied with my model and I look forward to seeing its implementation in the future.

Thanks!

Any questions?