



Estate Developers Consultancy

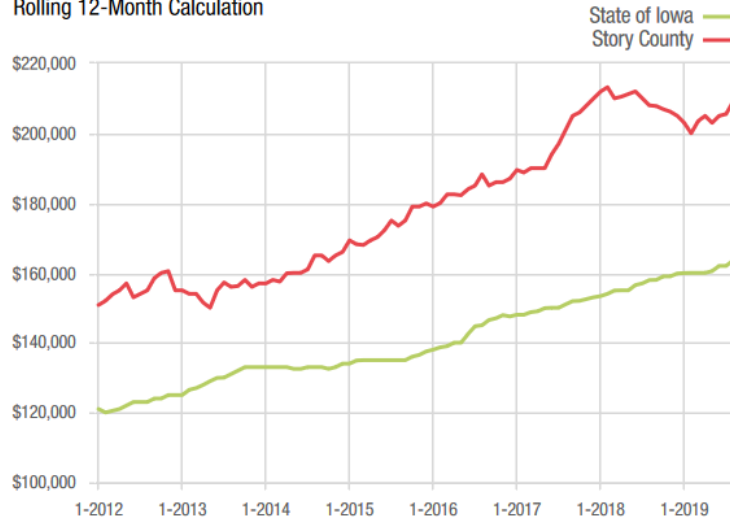
- 1. Identify which Estate for investment and development**
- 2. Proposed Property Appreciating Features**
- 3. Identify Property Depreciating Features**

**Estate Developers looking at 3 different
estate in Iowa for development**

North Ridge, College Creek, Greenhills

Median Sales Price - Single-Family Detached

Rolling 12-Month Calculation



Uptrend in sale prices of homes
in Iowa, Ames

Median : \$ 232,000

Average : \$ 245,328

Key Metrics	2018	2019	% Change	Thru 8-2018	Thru 8-2019	% Change
New Listings	83	99	+ 19.3%	894	1,025	+ 14.7%
Pending Sales	77	90	+ 16.9%	725	784	+ 8.1%
Closed Sales	119	121	+ 6.7%	705	762	+ 8.1%
Days on Market Until Sale	41	47	+ 14.6%	56	58	+ 3.6%
Median Sales Price*	\$202,000	\$231,984	+ 14.8%	\$210,000	\$212,250	+ 1.1%
Average Sales Price*	\$234,032	\$245,328	+ 4.8%	\$232,806	\$232,746	0.0%
Percent of List Price Received*	97.6%	98.2%	+ 0.6%	97.7%	97.7%	0.0%
Inventory of Homes for Sale	303	330	+ 8.9%	—	—	—
Months Supply of Inventory	3.7	3.8	+ 2.7%	—	—	—

Data Origin

- Published in Journal of Statistics Education, Volume 19, Number 3(2011)1
- www.amstat.org/publications/jse/v19n3/decock.pdf



Northridge

\$9034

\$2309

College Creek

Ames

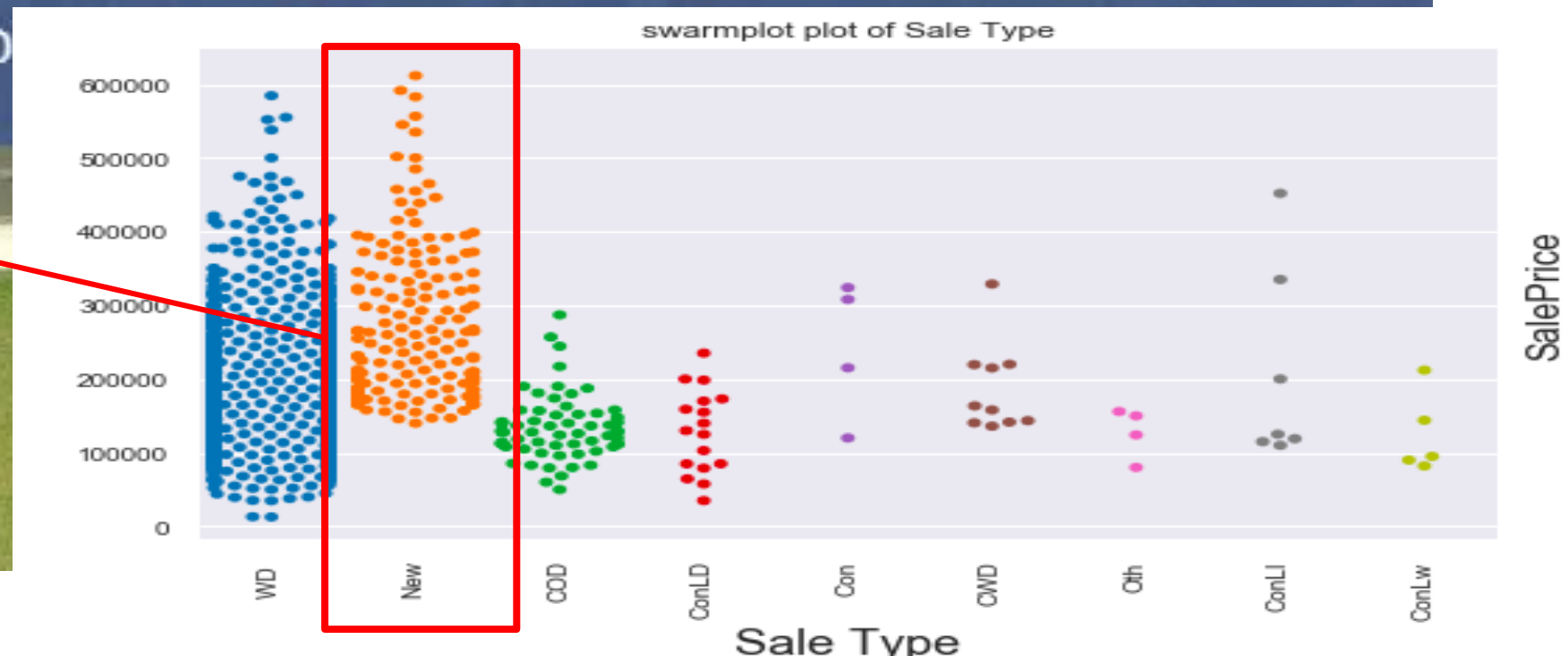
Green hills

\$5345

Property Appreciating Features

IOWA HOUSING MARKET

- Size of a home
- Sale type - New
- Overall Quality





FACTORS THAT INFLUENCE PROPERTY VALUE APPRECIATION



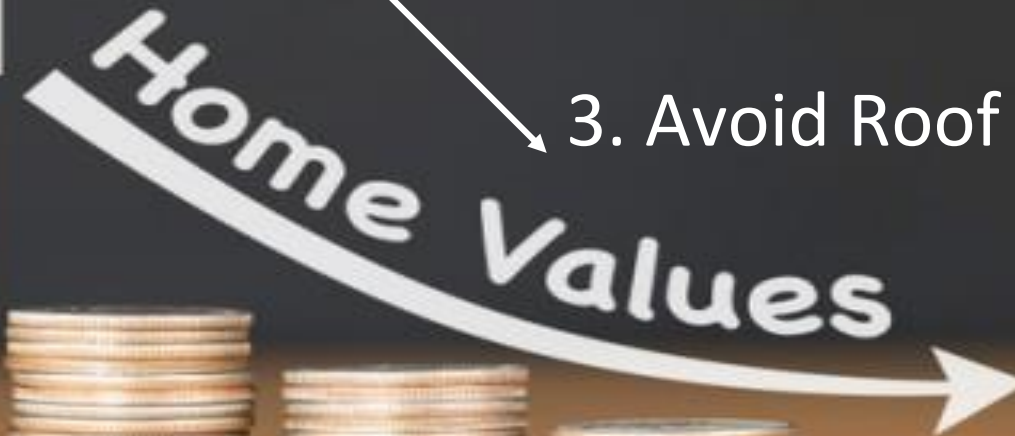
1. Number of Baths
+\$6275

2. Exterior Quality
+\$5335

3. Fireplaces
+\$4850

4. Kitchen Quality
+\$4810

Depreciating Property Features



1. Avoid Constructing Townhouse (-\$5881) and 2 storey (-\$2641)

2. Avoid Brick Face Masonry (-\$6780)

3. Avoid Roof Mansard (-\$2011)



Data cleaning

- Checking for null values and replacement with valid values if necessary
- Conversion of all ordinal and nominal data types into numerical format
- Dropping of outliers

df_train_alt: Replacing null values

Column 'Lot Frontage' has a mean 69.06 and median 68 being values close to each other.

This indicates that the distribution is almost Normal and hence the mean value is used for the null values

330 null values in 'Lot Frontage' were filled with mean value 69.055 ¶

```
: # df_train_alt: Replacing null values in 'Lot Frontage' with mean value
```

```
df_train_alt['Lot Frontage'] = df_train_alt['Lot Frontage'].fillna(df_train_alt['Lot Frontage'].mean())  
df_train_alt['Lot Frontage'].isnull().sum()
```

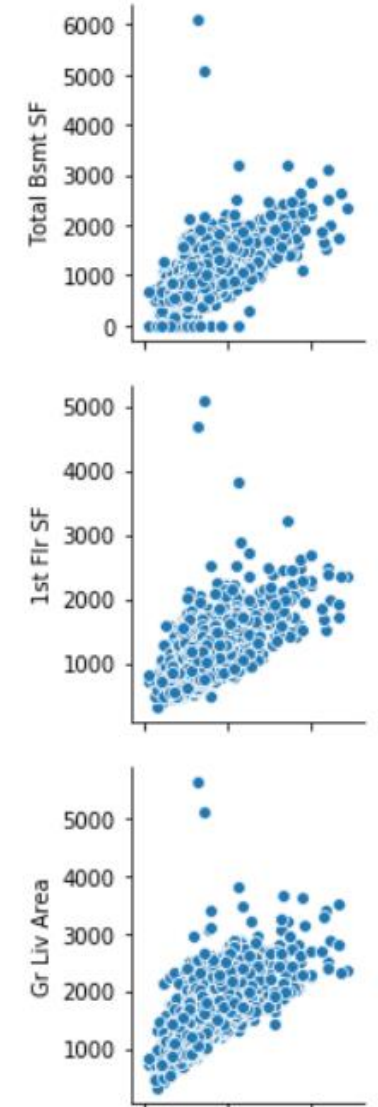
```
# changing ordinal ratings to rankings - 'Kitchen Qual'
```

```
mapping = {np.nan: 0, 'Po': 1, 'Fa': 2, 'TA': 3, 'Gd': 4, 'Ex': 5 }  
df_train_alt['Kit_Qual_rk'] = df_train_alt['Kitchen Qual'].apply(lambda x : mapping[x])
```


Features Identification

- Techniques used include:
 - **Heatmap**
 - Pearson Correlation coefficient
 - Filter Via Threshold of 50% correlation
 - **Reduce Collinearity** between features
 - **Eliminating poor Coefficient** from Lasso and Ridge modelling

	Id	PID	MS SubClass	Lot Frontage	Lot Area	Overall Qual	Overall Cond	Year Built	Year Remod/Add	Mas Vnr Area	BsmtFin SF 1	BsmtFin SF 2	Bsmt Unf SF	Total Bsmt SF	1st Flr SF	2nd Flr SF	Low Qual Fin SF	Gr Liv Area	Bsmt Full Bath	Bsmt Half Bath	Full Bath
Mo Sold	0.13	-0.033	0.013	-0.016	0.0032	0.019	-0.0031	-0.0071	0.012	-0.0039	-0.011	-0.014	0.02	0.0039	0.027	0.03	0.024	0.05	-0.0096	0.026	0.049
Yr Sold	-0.98	0.0085	-0.033	0.0084	-0.029	-0.012	0.048	-0.0036	0.043	-0.017	0.038	-1.3e-05	-0.044	-0.0043	-0.0074	-0.012	0.0013	-0.016	0.035	-0.01	0.0071
SalePrice	-0.051	-0.26	-0.087	0.33	0.3	0.8	-0.097	0.57	0.55	0.5	0.42	0.016	0.19	0.63	0.62	0.25	-0.042	0.7	0.28	-0.045	0.54



Numerical and Ordinal Features

SalePrice	1
Overall Qual	
Gr Liv Area	
Garage Area	
Garage Cars	0.65
Total Bsmt SF	0.63
1st Flr SF	0.62
Year Built	
Year Remod/Add	
Full Bath	
Garage Yr Blt	0.53
Max Yr Area	0.51
TotRms AbvGrd	0.5
Fireplaces	0.47
BsmtFin SF 1	0.42
Lot Frontage	0.34
Open Porch SF	0.33
Wood Deck SF	0.33
Lot Area	0.3
Bsmt Full Bath	0.28

Target Variable

Features Considerations

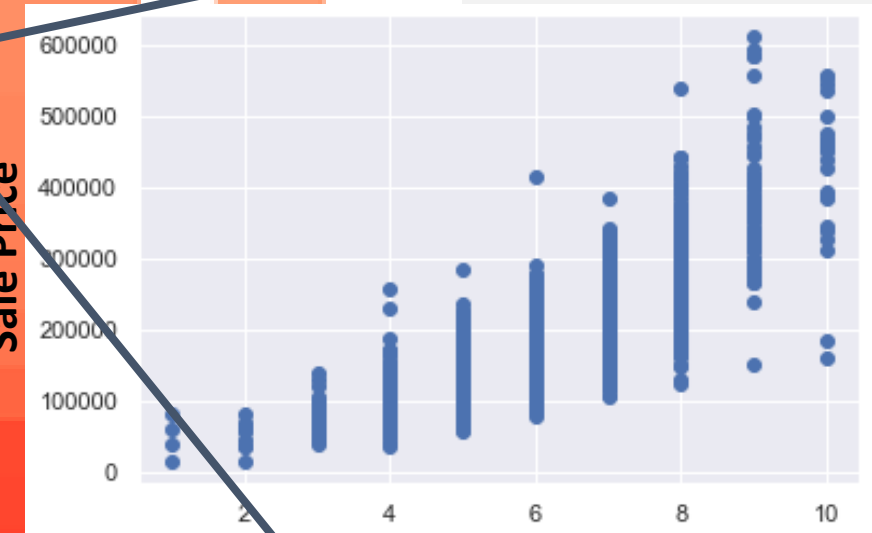
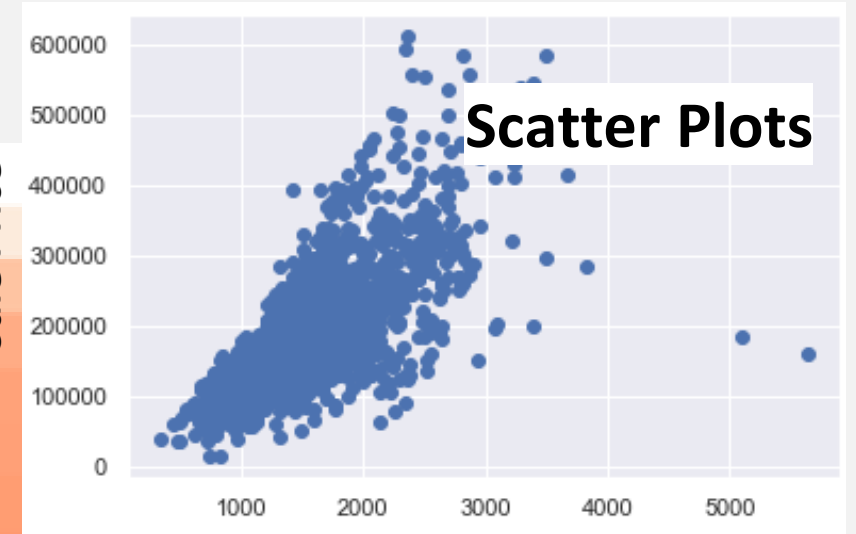
Sale Price

Sale Price

Scatter Plots

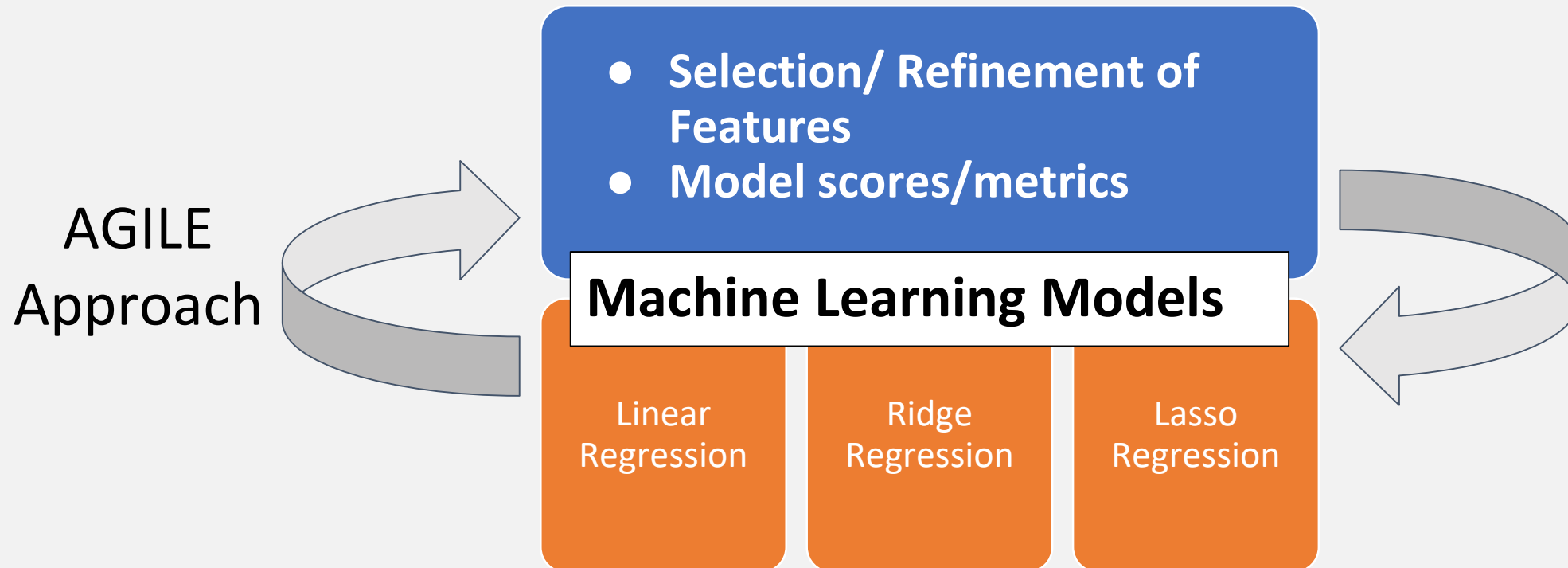
Gross Living Area

Overall Quality



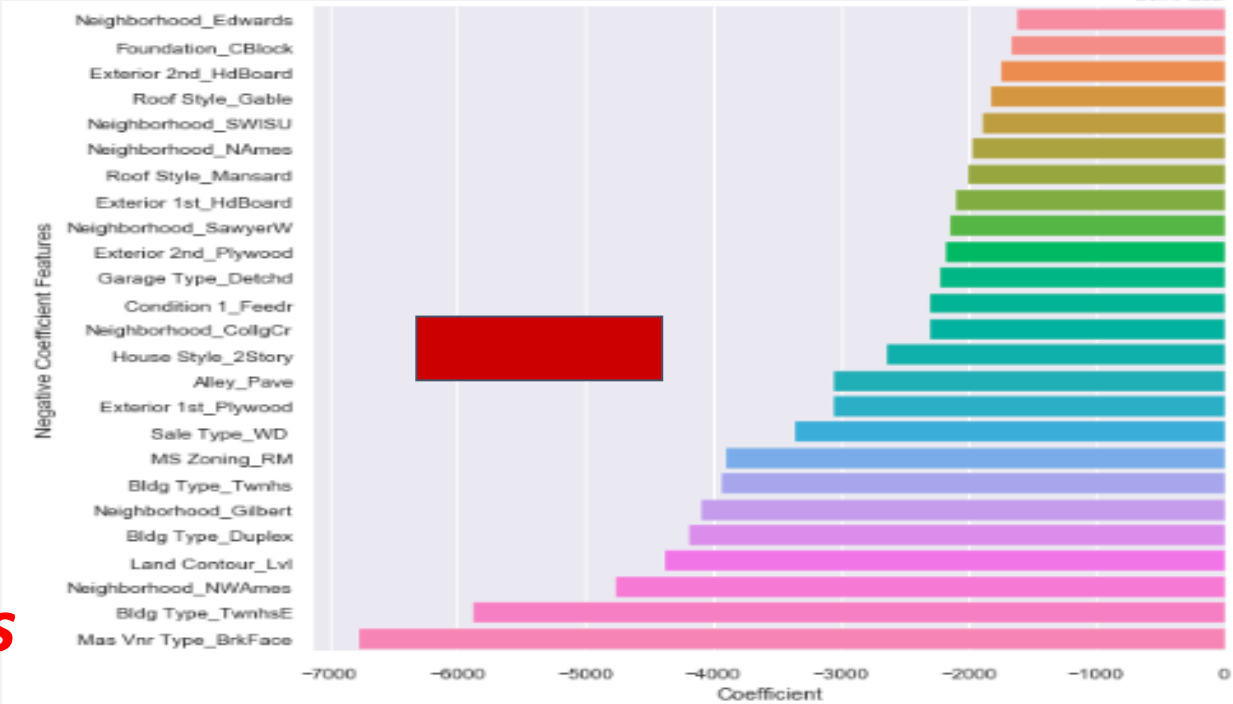
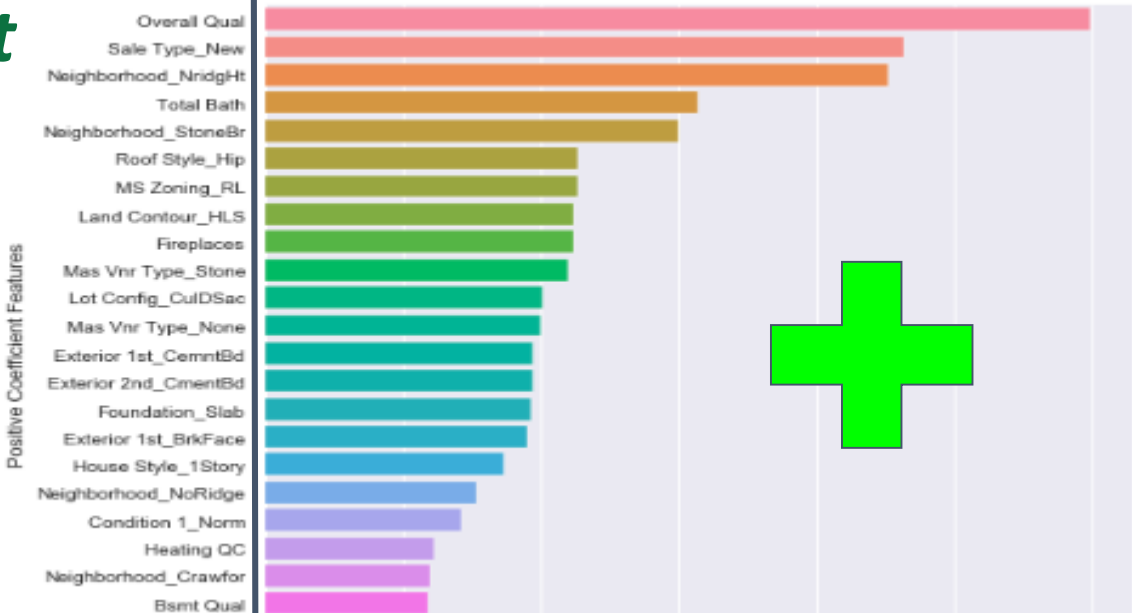
Development and validation of models

Machine learning model was trained iteratively on the data (> **2,000** homes) and further validated/tested on **about 900** homes.



Coefficient of Features

Best



Wors

t

Performance

Predicted Price



Predictions

Actuals

True Price

A close-up of Tony Stark (Robert Downey Jr.) from the Iron Man movies, looking upwards and to the side with a thoughtful expression. He is wearing a dark blue shirt and a dark jacket.

ANY QUESTIONS?