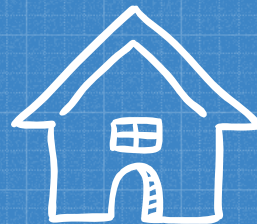


# AMES, IA

## HOUSE SALE PRICE ESTIMATION



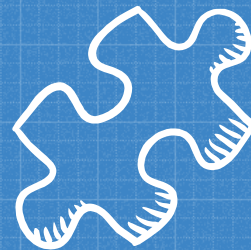
A LINEAR REGRESSION ESTIMATING HOUSE PRICES

BY JAMES SALISBURY



1

# THE PROBLEM STATEMENT





Home Flippers, LLC, a company focused on flipping homes in the Ames, IA region, contracted me to build a model to predict home values in the area. They were looking to use this model so they could calculate predicted prices for houses on the market in relation to historical sale prices of comparable houses. This information is of concern and value to them to enable them to identify opportunities with the greatest upside after making renovations or remodeling the house for sale.

Obviously the majority of the upside in the sale price from a home flip comes from the improvements made during the renovations, but by identifying undervalued homes in need of repair would allow them to calculate the potential upside compared to the amount of investment they put into the renovation.



2

# THE DATA





# \$181,469.70

Average house sale price

---

# 2,051

Data points

---

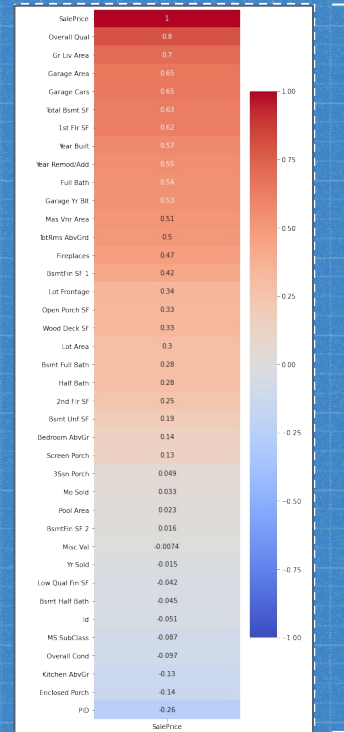
# 81

Features



# EXPLORING AND CLEANING THE DATA

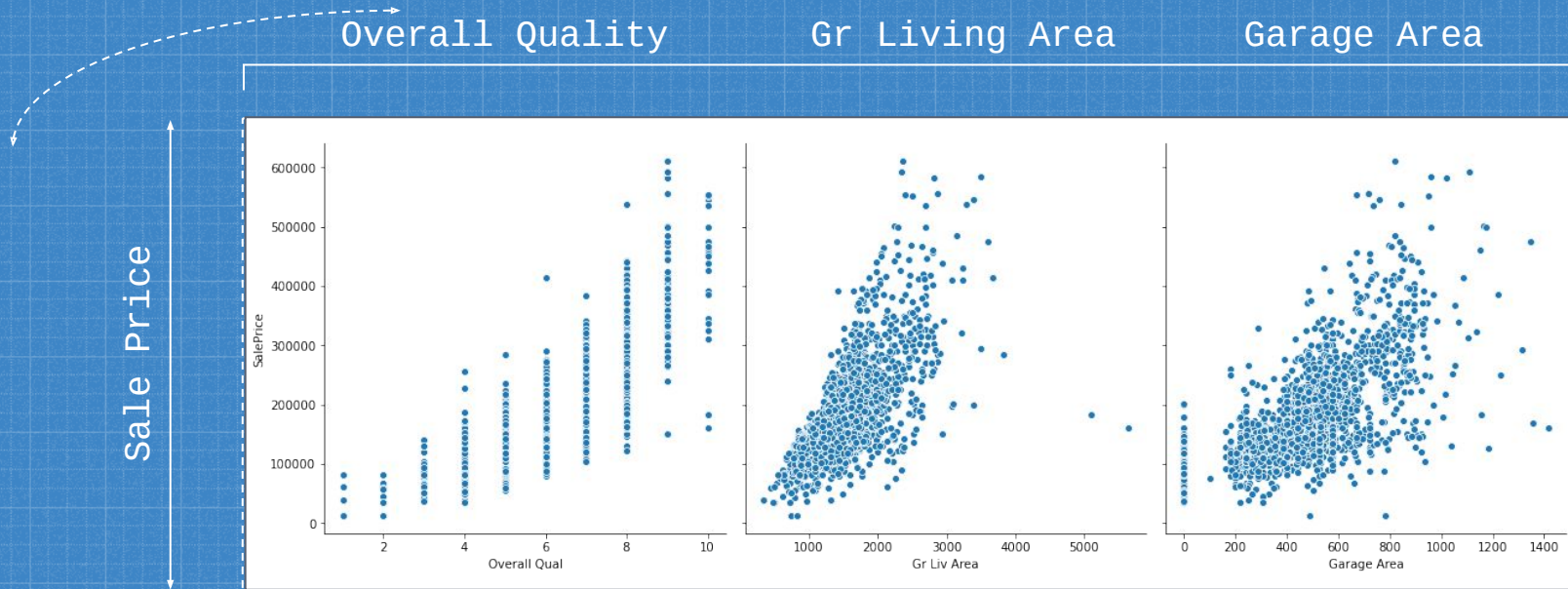
- Investigated correlations against sale price
  - Overall Quality
  - Above Grade Living Area
  - Garage Area
- Categorical features with small groupings
  - Neighborhood
  - Overall Quality
  - Foundation





# EXPLORING AND CLEANING THE DATA, CONT'D

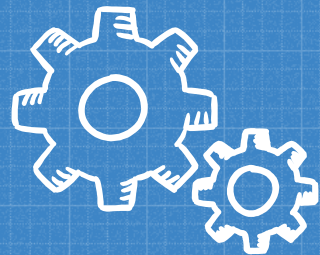
- Pairplot examining top three correlated features





3

# CLEANING & MAPPING FEATURE ENGINEERING





# CLEANING & MAPPING - Assumptions and Judgements

- Efficiency through functions!
- Filling in null values
  - Basement square footage
  - Garage type
- Overwriting condition values
- Mapping minority groups to their nearest neighbors
  - Neighborhoods
  - Quality and Condition ordinal rankings
  - Exterior covering
  - MS Zoning



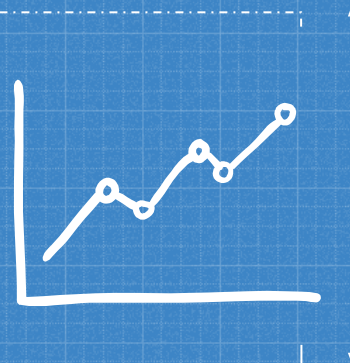
# FEATURE ENGINEERING

- Linear shift / MinMax scaling
- Interaction terms - synthesis of categories
  - Total rooms & Living area (above grade)
  - Full bathrooms & Living area (above grade)
  - Functional rating & Overall quality
- Logged features
  - Living area above grade
  - Lot area
  - Basement total square footage
  - Finished basement square footage



# 4

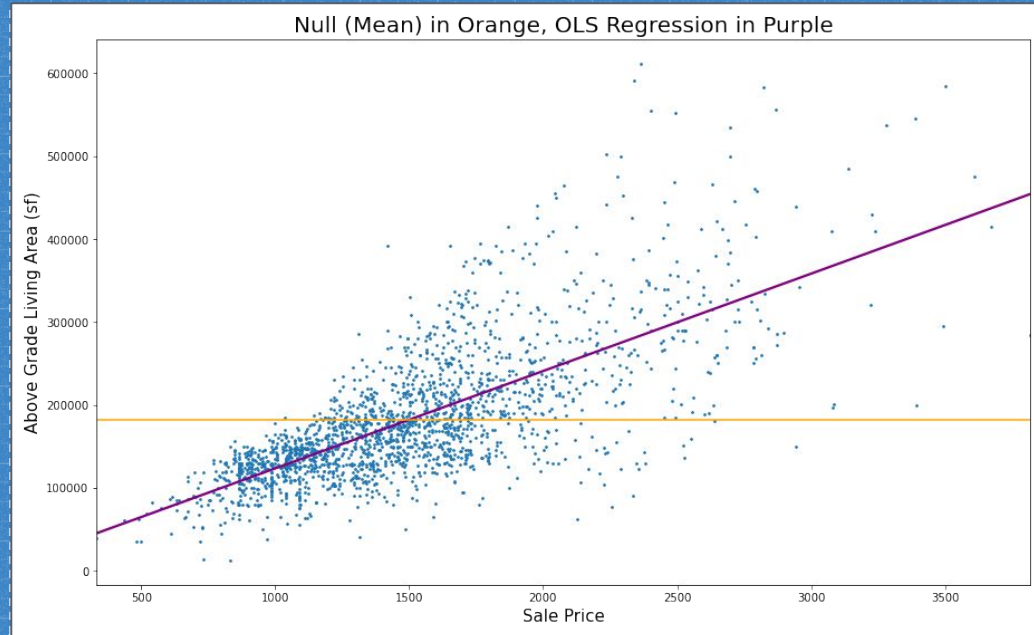
## THE MODEL





# MODELING & TUNING

- Establish baseline model
- Plot null hypothesis





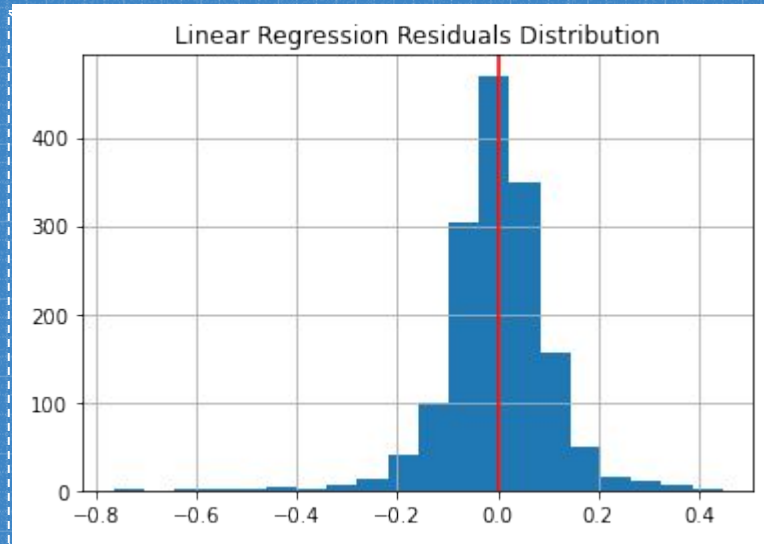
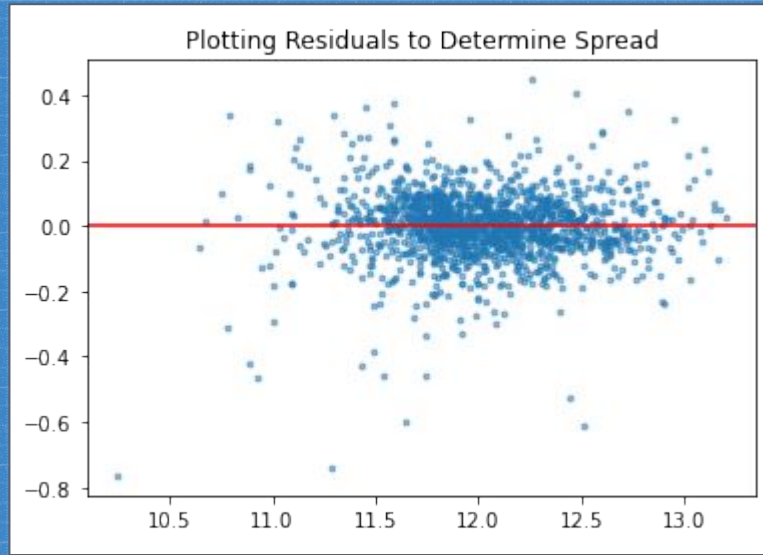
# MODELING & TUNING

- Feature selection
  - 20 features for dummification
  - 20 numeric columns (not including Id)
- Dum dum dum dum dum!
  - 113 columns total after `get_dummies()`
- Final model scores after many refinements

METRIC	SPLIT/MODEL	SCORE
RMSE (logged)	Training	0.1028
r2	Training	0.9375
RMSE (logged)	Validation	0.1280
r2	Validation	0.9035
RMSE	Baseline/Null	79276.56
RMSE	Linear Regression	19263.46



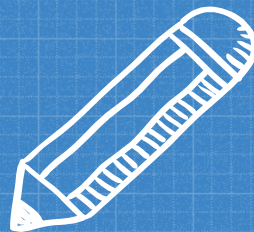
# MODELING & TUNING - Linear Regression





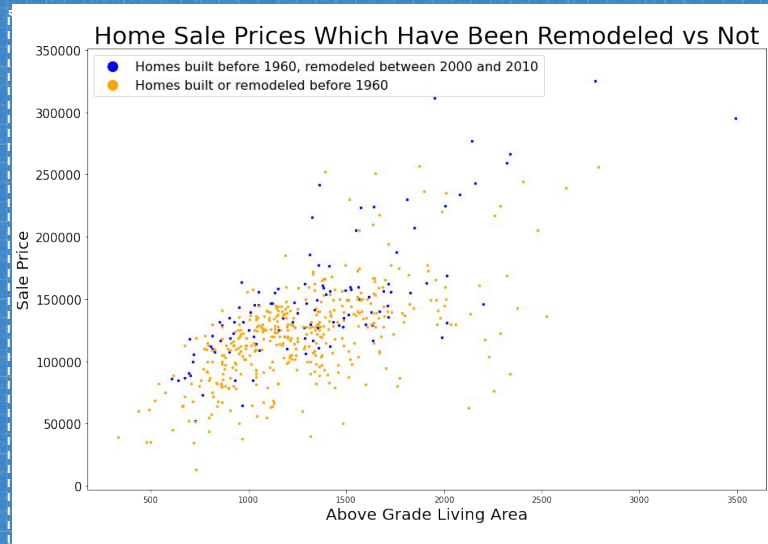
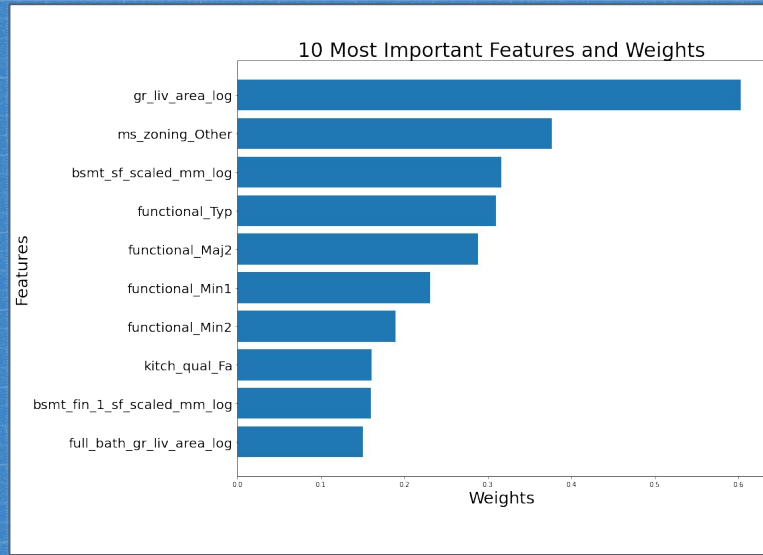
5

# THE FINDINGS





# FINDINGS





# POTENTIAL FLIPS

- Identified ten lowest priced houses in bins
- Highlighted opportunities in green
- Highlighted constraints in red

id	ms zoning	gr liv area	functional	bsmt sf	bsmt fin 1	bsmt fin 1 sf	exter 1	kitch qual	lot area	neighbor hood	cond_1	bldg type	style	overall cond	yr built	yr remodeled	roof style	exterior cond	foundation	bsmt ego	cent air	hdt_bath	hdt_bath	bedrooms	bst rooms	garage type	garage car size	pergola cond	second floor	bsmt full bath	bsmt half bath
1554	Other	733	Maj2	0	None	0	AsbShng	Fa	14584	DOFR	Norm	1fam	1Story	3	5	1952	1952	Gable	Fa	Other	N	1	0	2	4	Attchd	2	Fa	N	0	0
727	Other	720	Typ	720	Rec	495	Wd Sdng	TA	7879	DOFR	Norm	1fam	1Story	4	5	1920	1950	Gable	TA	CbClck	N	N	1	0	2	4	None	0	N	0	0
710	RM	968	Typ	690	Unf	0	VnylSd	TA	5925	DOFR	Norm	1fam	Fin	3	6	1910	1950	Gable	TA	BkTtl	N	N	1	0	2	6	None	0	N	0	0
738	Other	797	Typ	245	Unf	0	AsbShng	TA	1393	DOFR	Norm	1fam	1Story	4	4	1900	1950	Other	Fa	CbClck	N	N	1	0	2	5	None	0	N	0	0
709	RM	612	Typ	0	None	0	MetaISd	TA	5925	DOFR	Norm	1fam	1Story	3	4	1940	1950	Gable	Gd	BkTtl	N	N	1	0	1	4	Detchd	1	Fa	N	0
958	Other	952	Typ	756	Unf	0	Wd Sdng	TA	8712	DOFR	Norm	1fam	1Story	4	7	1896	1950	Hip	Fa	CbClck	N	Y	1	0	3	5	Detchd	1	TA	N	0
2880	Other	729	Mod	0	None	0	Wd Sdng	TA	12366	DOFR	RAA	1fam	1Story	3	5	1945	1950	Gable	TA	Other	N	N	1	0	2	5	None	0	N	0	0
2690	RM	693	Typ	693	Unf	0	Wd Sdng	Fa	4118	DOFR	RAA	1fam	1Story	4	4	1941	1950	Gable	TA	CbClck	N	N	1	0	2	4	None	0	N	0	0
1547	RM	796	Typ	796	Unf	0	AsbShng	TA	3636	DOFR	Norm	1fam	1Story	4	4	1922	1950	Gable	TA	BkTtl	N	N	1	0	2	5	None	0	N	0	0
946	Other	810	Maj1	0	None	0	Wd Sdng	TA	21280	DOFR	Norm	1fam	1Story	3	4	1910	1950	Gable	TA	CbClck	N	N	1	0	1	4	Detchd	1	TA	N	0
1556	Other	1217	Typ	649	Unf	0	Bedroom	TA	6500	DOFR	RAA	1fam	2Story	4	4	1905	1950	Other	Fa	BkTtl	N	N	1	0	3	6	Detchd	1	Fa	N	0
2044	RM	1484	Maj2	771	Unf	0	Wd Sdng	TA	10300	DOFR	Arney	1fam	Fin	4	3	1910	1950	Gable	Fa	CbClck	N	Y	2	0	1	6	Detchd	1	Fa	N	0
2017	RM	1092	Mod	0	None	0	VnylSd	Fa	5490	DOFR	Norm	1fam	2Story	4	5	1930	1950	Hip	TA	Other	N	N	2	0	3	7	None	0	N	0	
907	Other	1044	Typ	540	Unf	0	Wd Sdng	Fa	6712	DOFR	Norm	1fam	1Story	5	5	1952	1952	Hip	TA	CbClck	Ar	N	1	0	2	4	Baywood	2	TA	N	0
791	RL	1276	Mod	1276	Unf	0	Other	TA	9571	Edwards	Norm	1fam	1Story	5	3	1958	1958	Gable	Fa	CbClck	N	Y	1	0	3	5	Attchd	1	TA	Y	0
768	RL	1112	Typ	585	Unf	0	MetaISd	TA	1790	Edwards	Norm	1fam	Fin	5	3	1934	1950	Gable	TA	PCorc	N	N	2	0	4	6	Detchd	1	Fa	N	0
279	RL	1120	Typ	744	Unf	0	VnylSd	Fa	11792	Edwards	Norm	1fam	Fin	4	5	1948	1950	Gable	TA	CbClck	N	N	1	0	2	5	Detchd	2	Fa	F	0
2620	Other	1020	Typ	1020	Unf	0	Wd Sdng	Fa	4763	DOFR	Norm	1fam	Unfin	3	3	1918	1950	Gable	Fa	BkTtl	N	N	1	0	2	5	None	0	N	0	0
696	RM	1131	Typ	676	Unf	0	Wd Sdng	TA	5720	DOFR	Arney	1fam	Fin	5	6	1941	1950	Gable	TA	CbClck	N	Y	1	0	1	5	Detchd	1	TA	Y	0
2240	RM	1077	Maj2	961	Unf	0	MetaISd	TA	8967	DOFR	Norm	1fam	1Story	5	3	1920	1950	Gable	Fa	BkTtl	N	Y	1	0	2	6	Detchd	1	Fa	N	0
2603	RL	1760	Typ	1077	Unf	0	VnylSd	TA	18918	DOFR	Norm	1fam	Fin	4	4	1910	1950	Gable	TA	BkTtl	N	Y	1	1	4	7	Detchd	2	TA	Y	0
2984	Other	3400	Mod	0	None	0	Other	TA	33254	Majord	Arney	1fam	1Story	1	3	1951	1951	Gable	Fa	CbClck	N	Y	1	1	1	6	Attchd	1	TA	N	0
2699	RM	1774	Maj2	1095	Unf	0	AsbShng	TA	9600	DOFR	Norm	1fam	2Story	4	3	1900	1950	Gable	TA	BkTtl	N	N	2	0	4	8	27yaps	3	Fa	N	1
714	RM	1180	Typ	677	Unf	0	VnylSd	GF	3400	DOFR	Norm	1fam	2Story	4	4	1910	1950	Gable	Gd	BkTtl	N	N	2	0	3	7	Detchd	2	TA	N	0
701	RM	1701	Typ	690	Unf	0	AsbShng	Fa	7627	DOFR	Arney	1fam	2Story	4	6	1920	1950	Gable	TA	BkTtl	N	N	2	0	4	8	None	0	N	0	0
897	RL	1529	Mod	0	None	0	WdShng	TA	8405	Edwards	Norm	1fam	Fin	4	3	1945	1950	Gable	TA	Other	N	N	2	0	4	9	Detchd	1	TA	N	0
2613	RL	1022	Typ	1008	Unf	0	Wd Sdng	TA	6092	WdShg	Norm	1fam	Fin	5	5	1926	1950	Gable	Fa	BkTtl	N	Y	2	0	4	7	None	0	F	0	0
901	RL	1040	Maj1	1150	Rec	546	HdQuart	TA	11767	Edwards	Norm	1fam	1Story	5	5	1956	1956	Hip	TA	CbClck	N	Y	1	0	2	7	Attchd	1	TA	Y	0
2636	RL	1669	Maj2	1093	Unf	0	Wd Sdng	TA	8707	Edwards	RAA	1fam	Fin	4	5	1934	1950	Gable	TA	BkTtl	N	N	1	1	4	9	Attchd	1	TA	Y	0
799	RL	1130	Typ	795	Unf	0	MetaISd	TA	9884	Edwards	Arney	1fam	2Story	4	5	1940	1950	Gable	TA	CbClck	Maj	Y	1	0	4	7	Detchd	1	Fa	F	1
754	RM	2178	Maj2	1040	Unf	0	MetaISd	TA	3000	DOFR	Norm	1fam	2Story	5	3	1922	1950	Hip	TA	BkTtl	N	N	2	0	4	11	None	0	Y	0	0
187	RM	2756	Typ	840	Unf	0	MetaISd	TA	1645	DOFR	Norm	1fam	2Story	5	5	1910	1950	Gable	BkTtl	Maj	N	N	2	0	4	12	None	0	N	0	0
1594	RM	2377	Mod	662	Unf	0	Wd Sdng	TA	15000	DOFR	Norm	1fam	2Story	4	6	1885	1950	Gable	TA	PCorc	N	Y	2	0	4	11	Detchd	2	TA	N	0
753	RM	2729	Typ	795	Unf	0	Wd Sdng	TA	5020	DOFR	Norm	1fam	Fin	6	6	1912	1950	Gable	TA	CbClck	N	Y	1	0	5	8	None	0	Y	0	0
1596	RM	2130	Typ	1008	Unf	0	Wd Sdng	Fa	9900	DOFR	Norm	1fam	2Story	5	6	1880	1950	Gable	TA	BkTtl	Maj	Y	2	0	5	8	Detchd	1	TA	N	0
190	RM	2290	Typ	777	Unf	0	VnylSd	TA	11340	DOFR	Norm	1fam	2Story	4	6	1885	1950	Gable	TA	PCorc	N	Y	2	0	4	11	Detchd	2	TA	N	0
1114	RM	2058	Typ	735	Rec	319	AsbShng	TA	9600	DOFR	Norm	1fam	Unfin	6	5	1917	1950	Gable	TA	BkTtl	N	N	1	1	3	8	Detchd	2	Fa	F	0
2275	RL	1684	Maj1	0	None	0	MetaISd	TA	43000	Majord	Arney	1fam	1Story	1	3	1951	1951	Gable	TA	CbClck	N	Y	1	0	2	9	27yaps	3	TA	N	0
2562	RL	2136	Mod	2136	Unf	756	Wd Sdng	TA	10640	Majors	Norm	1fam	Fin	3	3	1951	1951	Hip	TA	CbClck	N	N	2	0	4	7	Detchd	2	TA	Y	0
661	RL	2377	Mod	1333	Unf	0	MetaISd	Gd	10410	DOFR	Norm	1fam	Fin	6	5	1915	1950	Gable	TA	BkTtl	N	Y	2	0	8	Detchd	2	TA	Y	0	0
ms zoning	gr liv area	functional	bsmt sf	bsmt fin 1	bsmt fin 1 sf	exter 1	kitch qual	lot area	cond_1	bldg type	style	overall cond	yr built	yr remodeled	exterior cond	foundation															
Other	733	Maj2	0	None	0	AsbShng	Fa	14584	Norm	1fam	1Story	5	1952	1952	Fa	Other															
Other	720	Typ	720	Rec	495	Wd Sdng	TA	7879	Norm	1fam	1Story	5	1920	1950	TA	CbClck															
RM	968	Typ	690	Unf	0	VnylSd	TA	5925	Norm	1fam	Fin	6	1910	1950	TA	BrkTtl															
Other	797	Typ	245	Unf	0	AsbShng	TA	13932	Norm	1fam	1Story	3	1900	1950	Fa	CbClck															
RM	612	Typ	0	None	0	MetaISd	TA	5925	Norm	1fam	1Story	4	1940	1950	Gd	BrkTtl															
Other	952	Typ	756	Unf	0	Wd Sdng	TA	8712	Norm	1fam	1Story	7	1896	1950	Fa	CbClck															
Other	729	Mod	0	None	0	Wd Sdng	TA	12366	RAA	1fam	1Story	5	1945	1950	TA	Other															
RM	693	Typ	693	Unf	0	Wd Sdng	Fa	4118	RAA	1fam	1Story	4	1941	1950	TA	CbClck															
RM	796	Typ	796	Unf	0	AsbShng	TA	3636	Norm	1fam	1Story	4	1922	1950	TA	BrkTtl															



6

# THE NEXT STEPS





## NEXT STEPS

- K-Nearest Neighbors in conjunction with my Linear Regression model
  - Identify comp set
  - Predict sale price after renovations based on improvements
- Extending the model to other towns, counties, or states
- Evaluate accuracy in the present day market



# Thanks !

