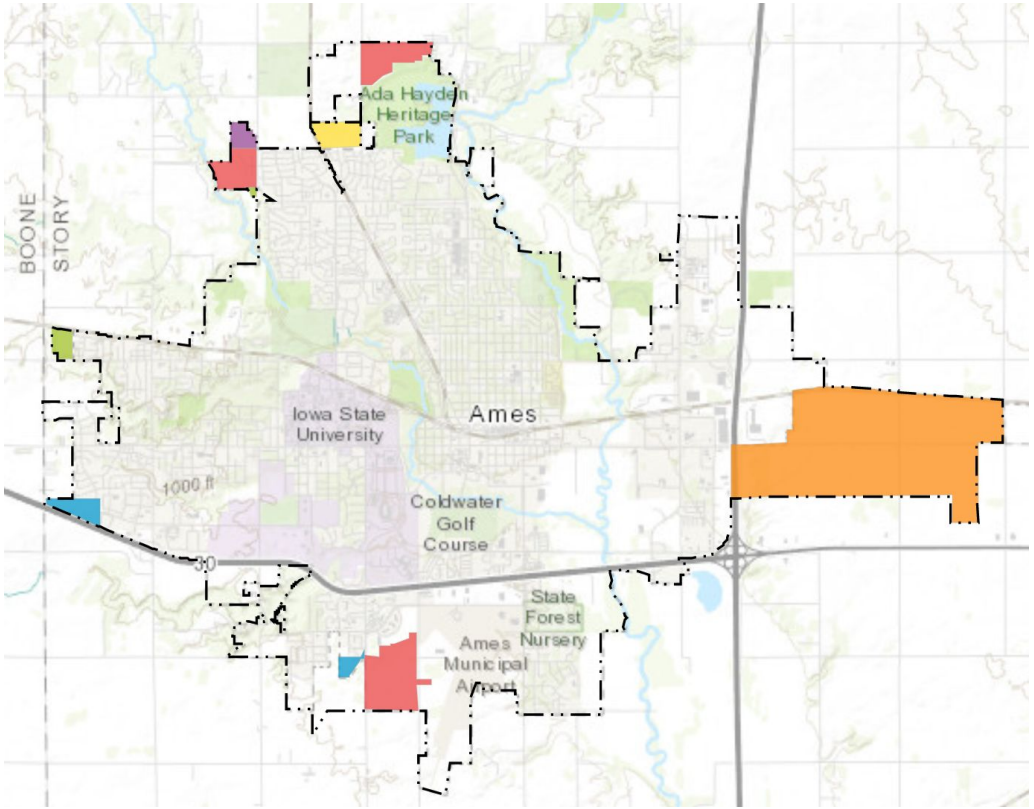
Project 2: Housing Prices in Ames, Iowa

Ganesh Jing Chuan Radhe Yu Zheng

About The Business

- **Company Profile**: Property listing and pricing platform, connecting home buyers and sellers
- **Business Context**: Planning expansion of operations into the city of Ames, Iowa
- **Business Problem**: Current platform's automatic suggested price for new property listings does not work well in a new city.
- **Objective**: Estimate the most appropriate price based on selected characteristics of the property (to reduce the amount of information that we need to collect).

City of Ames

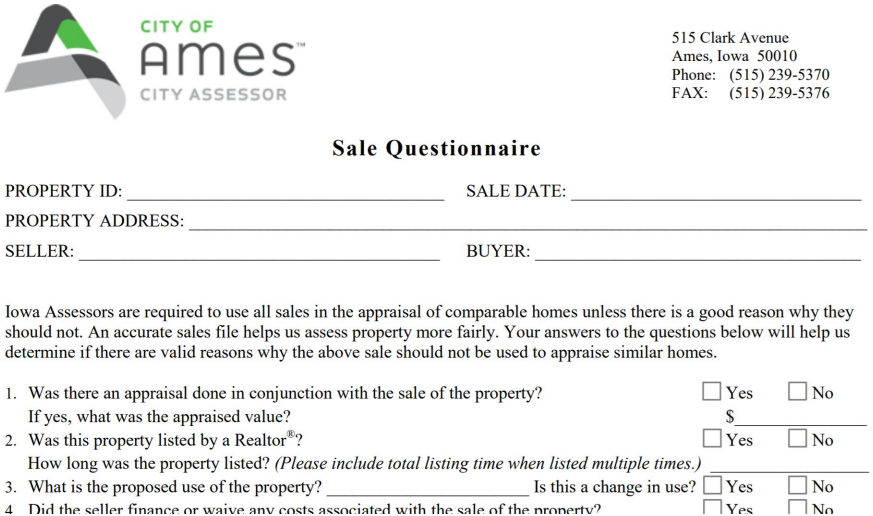


- Population largely comprised of the Iowa State University population. University is also the largest employer.
- The areas annexed by the city has been growing slowly over the years.

Source: City of Ames

Data Source

- Data from the Ames City Assessor
- Collated from surveys filled in by home buyers
- Obtained by Dean De Cock, of Truman State University
- Accessed on Kaggle



CITY OF ames CITY ASSESSOR

515 Clark Avenue
Ames, Iowa 50010
Phone: (515) 239-5370
FAX: (515) 239-5376

Sale Questionnaire

PROPERTY ID: _____ SALE DATE: _____

PROPERTY ADDRESS: _____

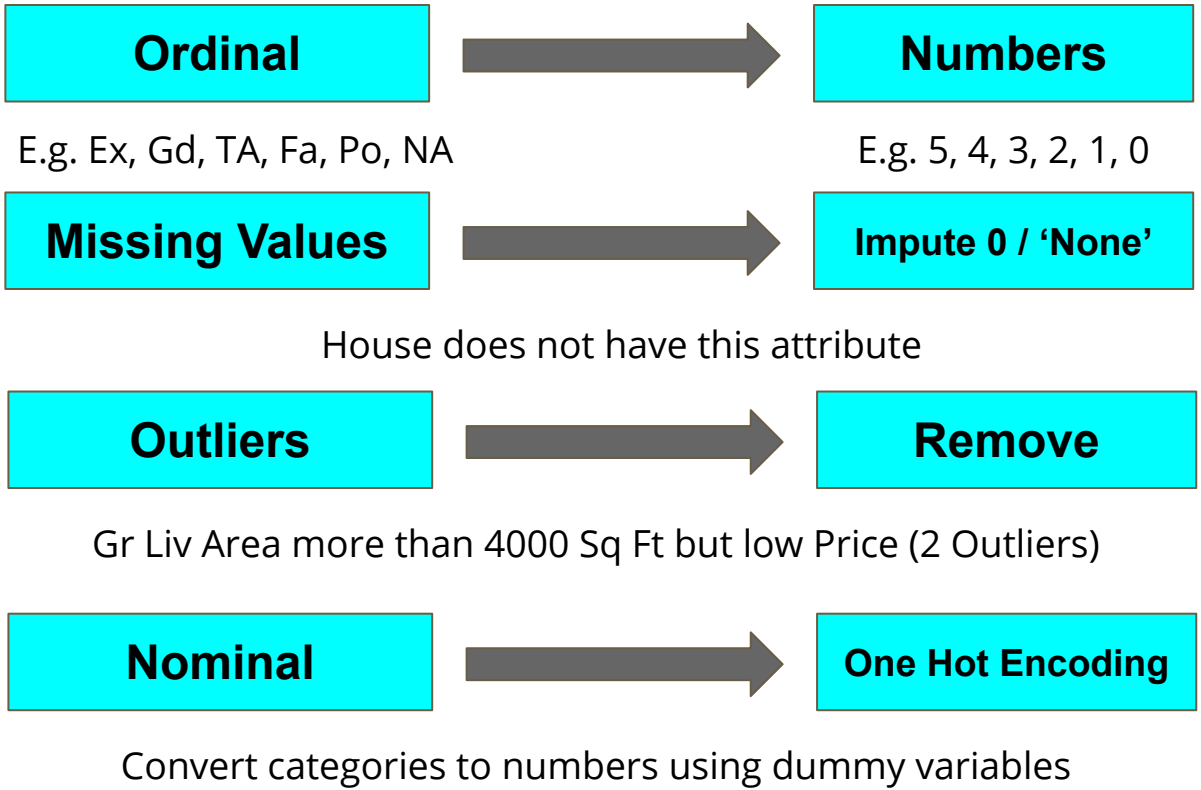
SELLER: _____ BUYER: _____

Iowa Assessors are required to use all sales in the appraisal of comparable homes unless there is a good reason why they should not. An accurate sales file helps us assess property more fairly. Your answers to the questions below will help us determine if there are valid reasons why the above sale should not be used to appraise similar homes.

1. Was there an appraisal done in conjunction with the sale of the property? ☐ Yes ☐ No
   If yes, what was the appraised value? $_____
2. Was this property listed by a Realtor®? ☐ Yes ☐ No
   How long was the property listed? *(Please include total listing time when listed multiple times.)* _____
3. What is the proposed use of the property? _____ Is this a change in use? ☐ Yes ☐ No
4. Did the seller finance or waive any costs associated with the sale of the property? ☐ Yes ☐ No

Source: City of Ames

Data Cleaning

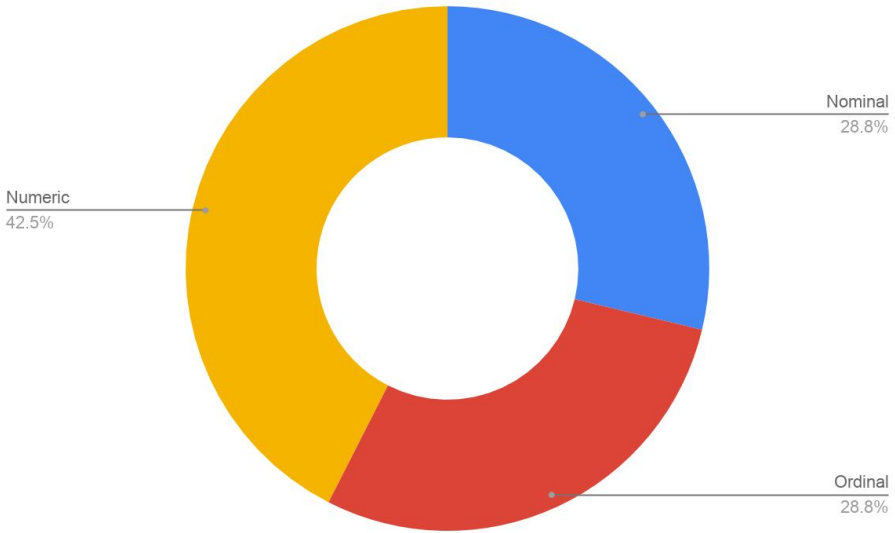


All Variables

Numeric Variables

Ordinal Variables

Nominal Variables



Numeric

34 Variables

	Feature	Description
2	Lot Frontage	Linear feet of street connected to property
3	Lot Area	Lot size in square feet
18	Year Built	Original construction date
19	Year Remod/Add	Remodel date (same as construction date if no remodeling or additions)
25	Mas Vnr Area	Masonry veneer area in square feet
33	BsmtFin SF 1	Type 1 finished square feet
35	BsmtFin SF 2	Type 2 finished square feet
36	Bsmt Unf SF	Unfinished square feet of basement area
37	Total Bsmt SF	Total square feet of basement area
42	1st Flr SF	First Floor square feet
43	2nd Flr SF	Second floor square feet
44	Low Qual Fin SF	Low quality finished square feet (all floors)
45	Gr Liv Area	Above grade (ground) living area square feet
46	Bsmt Full Bath	Basement full bathrooms
47	Bsmt Half Bath	Basement half bathrooms
48	Full Bath	Basement full bathrooms
49	Half Bath	Basement half bathrooms

	Feature	Description
50	Bedroom AbvGr	Bedrooms above grade (does NOT include basement bedrooms)
51	Kitchen AbvGr	Kitchens above grade
53	TotRms AbvGrd	Total rooms above grade (does not include bathrooms)
55	Fireplaces	Number of fireplaces
58	Garage Yr Blt	Year garage was built
60	Garage Cars	Size of garage in car capacity
61	Garage Area	Size of garage in square feet
65	Wood Deck SF	Wood deck area in square feet
66	Open Porch SF	Open porch area in square feet
67	Enclosed Porch	Enclosed porch area in square feet
68	3Ssn Porch	Three season porch area in square feet
69	Screen Porch	Screen porch area in square feet
70	Pool Area	Pool area in square feet
74	Misc Val	$Value of miscellaneous feature
75	Mo Sold	Month Sold (MM)
76	Yr Sold	Year Sold (YYYY)
78	SalePrice	Sale price $$

Ordinal

23 Variables

	Feature	Description
6	Lot Shape	General shape of property
8	Utilities	Type of utilities available
10	Land Slope	Slope of property
16	Overall Qual	Rates the overall material and finish of the house
17	Overall Cond	Rates the overall condition of the house
26	Exter Qual	Evaluates the quality of the material on the exterior
27	Exter Cond	Evaluates the present condition of the material on the exterior
29	Bsmt Qual	Evaluates the height of the basement
30	Bsmt Cond	Evaluates the general condition of the basement
31	Bsmt Exposure	Refers to walkout or garden level walls
32	BsmtFin Type 1	Rating of basement finished area

	Feature	Description
34	BsmtFin Type 2	Rating of basement finished area (if multiple types)
39	Heating QC	Heating quality and condition
41	Electrical	Electrical system
52	Kitchen Qual	Kitchen quality
54	Functional	Home functionality (Assume typical unless deductions are warranted)
56	Fireplace Qu	Fireplace quality
59	Garage Finish	Interior finish of the garage
62	Garage Qual	Garage quality
63	Garage Cond	Garage condition
64	Paved Drive	Paved driveway
71	Pool QC	Pool quality
72	Fence	Fence quality

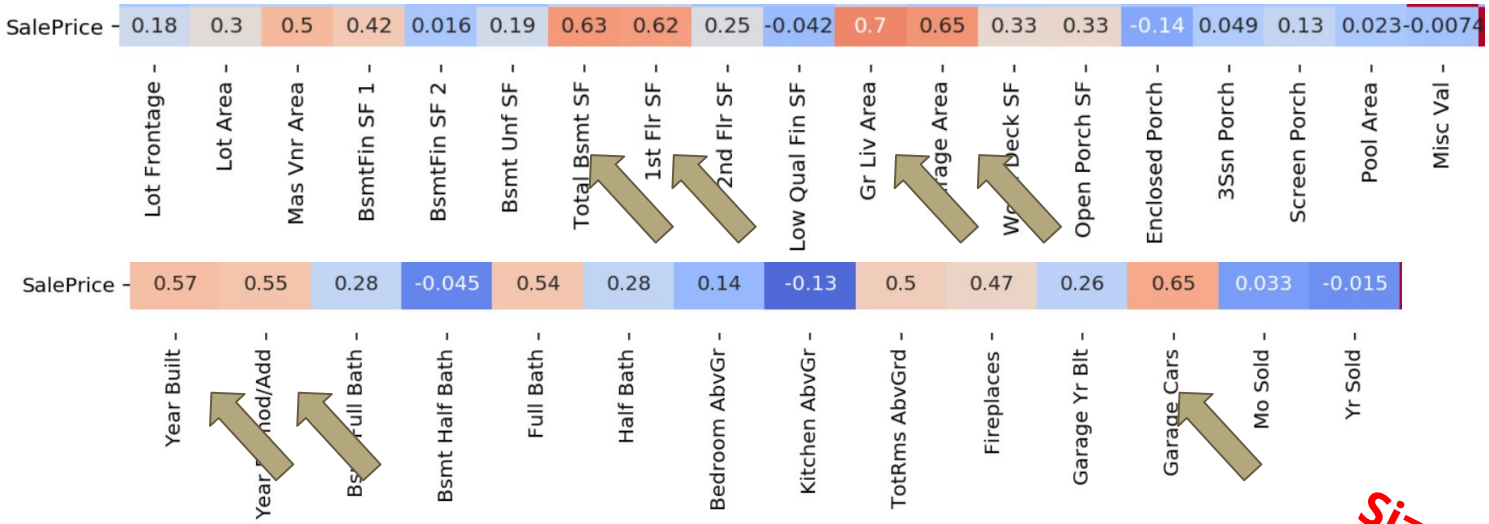
Nominal

22 Variables

	Feature	Description
0	MS SubClass	Identifies the type of dwelling involved in the sale.\t
1	MS Zoning	Identifies the general zoning classification of the sale.
4	Street	Type of road access to property
5	Alley	Type of alley access to property
7	Land Contour	Flatness of the property
9	Lot Config	Lot configuration
11	Neighborhood	Physical locations within Ames city limits (map available)
12	Condition 1	Proximity to various conditions
13	Condition 2	Proximity to various conditions (if more than one is present)
14	Bldg Type	Type of dwelling
15	House Style	Style of dwelling

	Feature	Description
20	Roof Style	Type of roof
21	Roof Matl	Roof material
22	Exterior 1st	Exterior covering on house
23	Exterior 2nd	Exterior covering on house (if more than one material)
24	Mas Vnr Type	Masonry veneer type
28	Foundation	Type of foundation
38	Heating	Type of heating
40	Central Air	Central air conditioning
57	Garage Type	Garage location
73	Misc Feature	Miscellaneous feature not covered in other categories
77	Sale Type	Type of sale

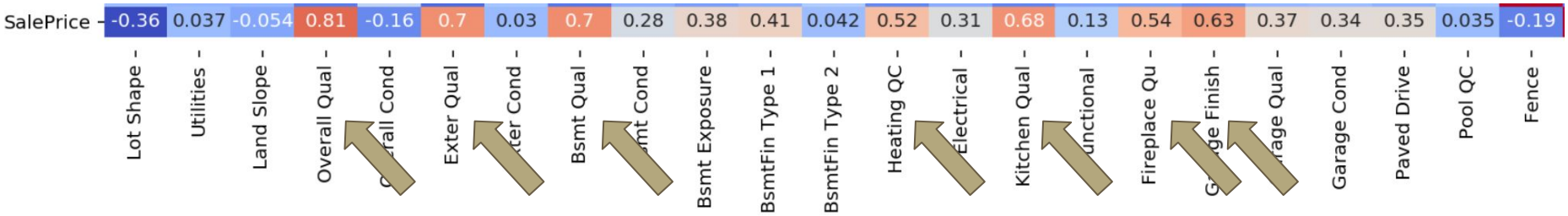
Correlation of Numerical Variables to Sale Price



Newness matters!

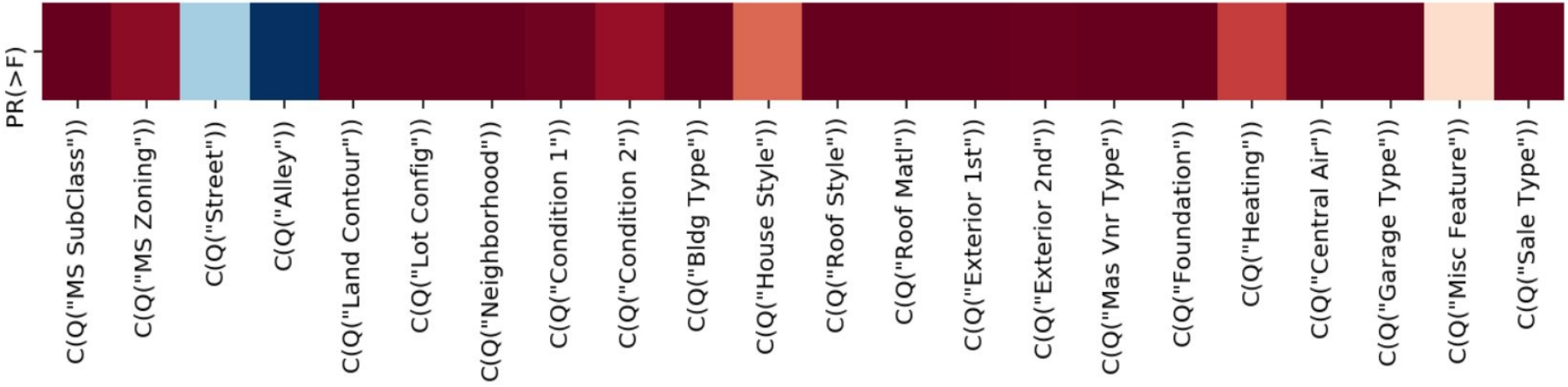
Size matters!

Correlation of Ordinal Variables to Sale Price

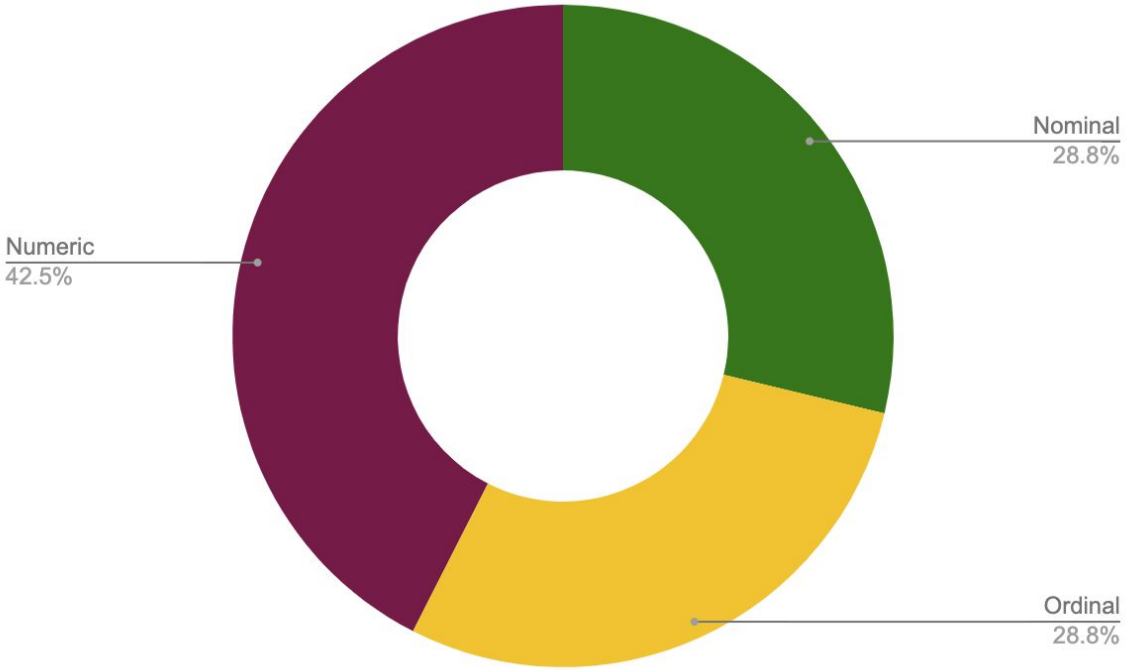


Quality matters!

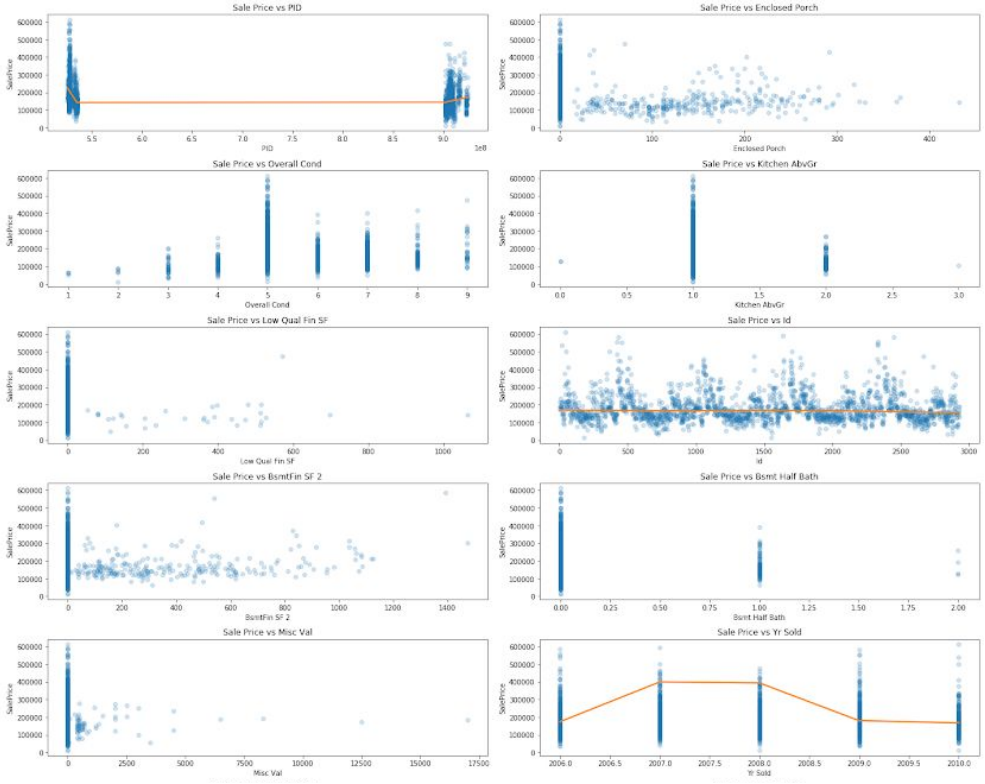
Nominal Variables Impacting Sale Price



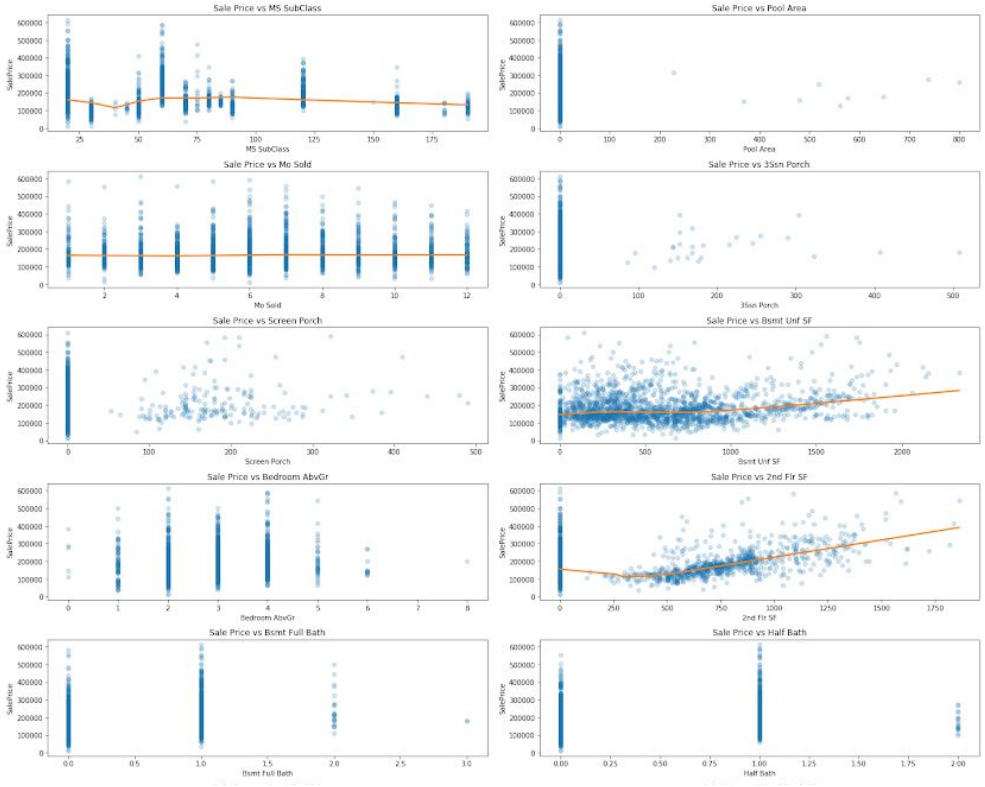
16 variables with difference in mean Sale Price


Nominal
28.8%
Numeric
42.5%
Ordinal
28.8%

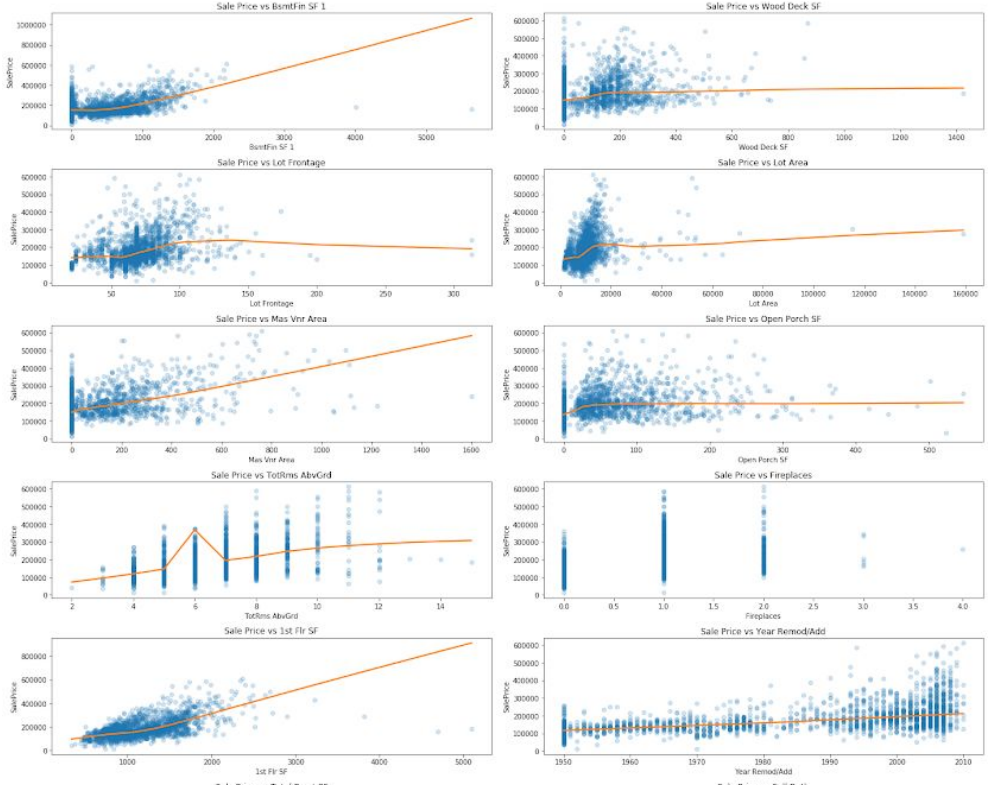
Numeric Variables



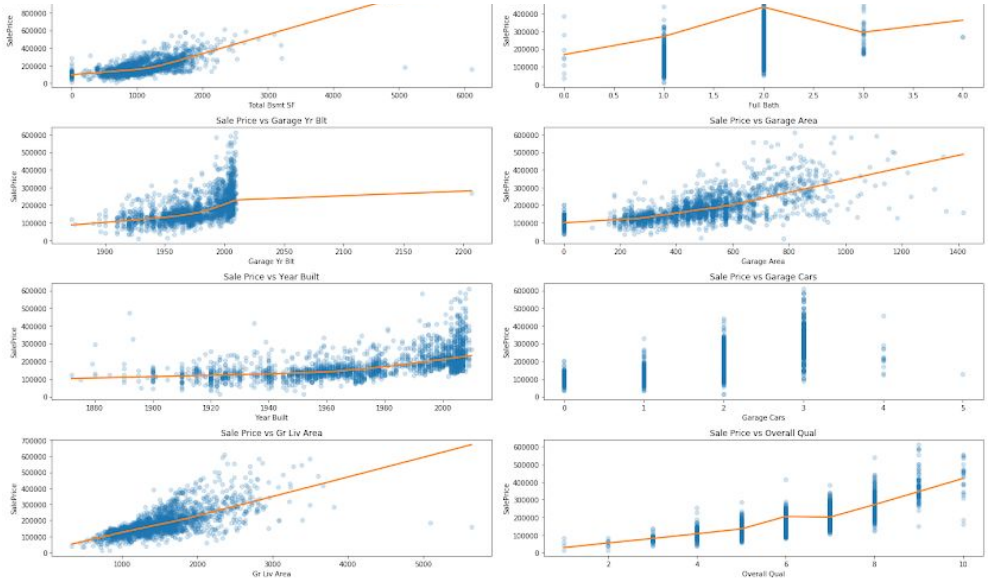
Numeric Variables



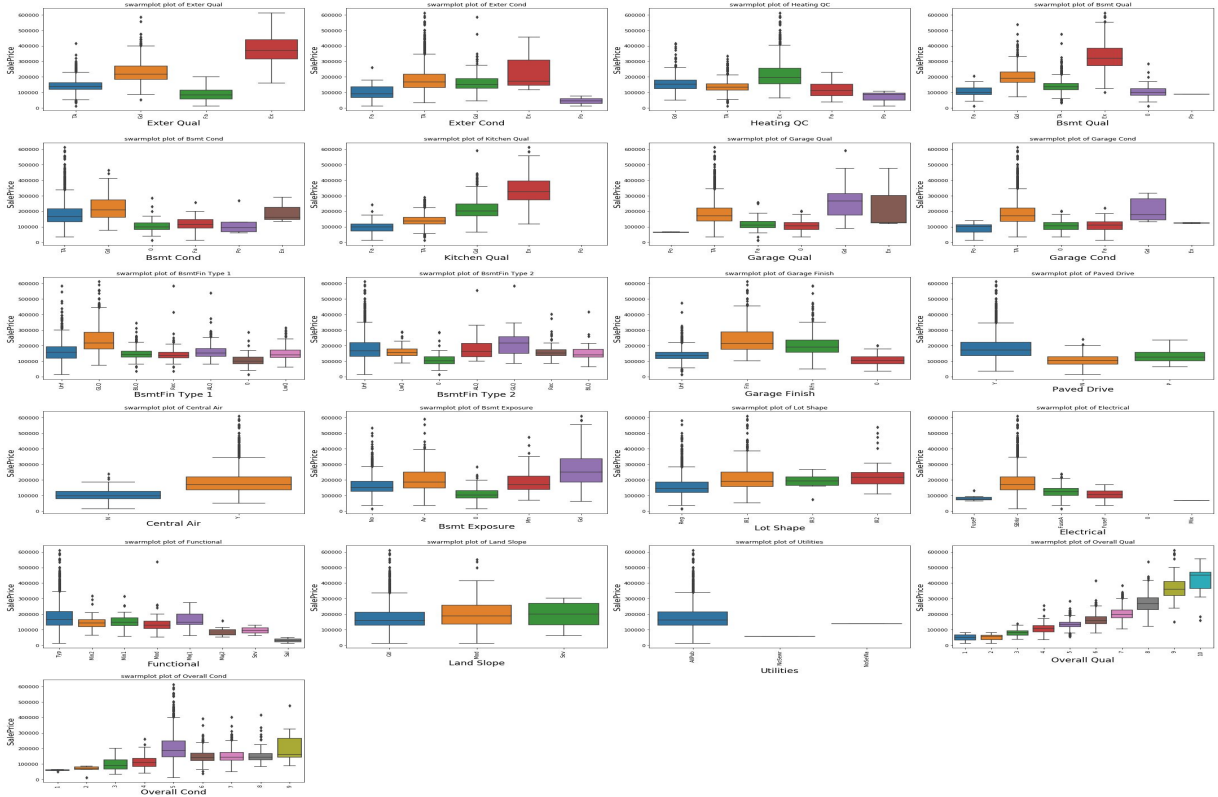
Numeric Variables



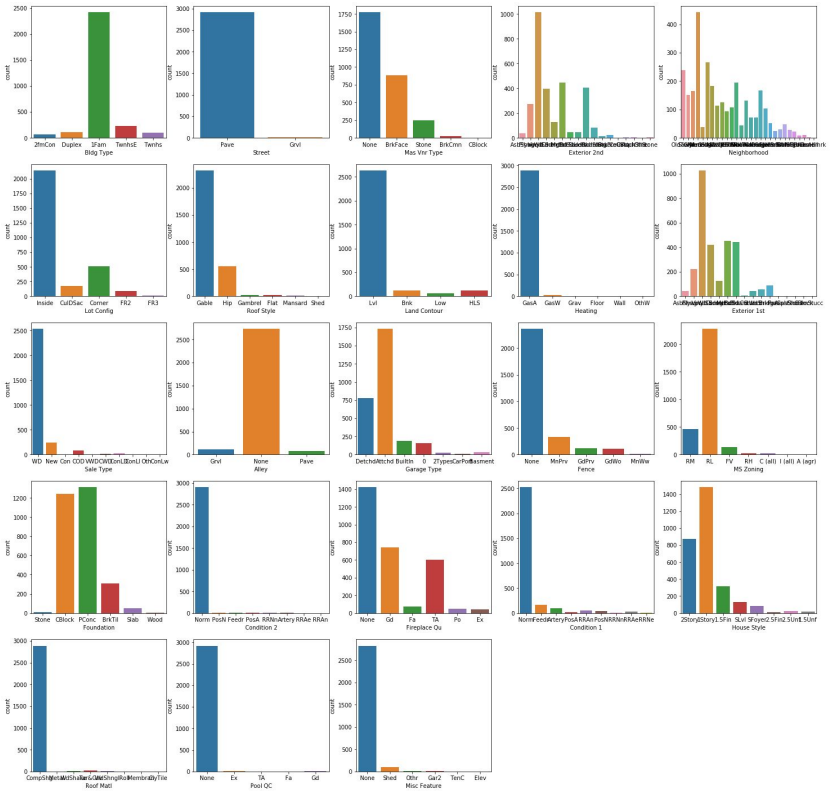
Numeric Variables



Ordinal Variables



Nominal Variables



Feature Engineering

Feature Engineering was implemented step by step and the results were tested at each stage to ascertain if it was making the model more accurate

Steps taken:

1. Certain related features were combined. E.g Additional columns were added for total Baths & total Square foot
2. Polynomial feature columns were created for the variables with highest correlation to output
3. Recursive feature elimination was implemented to reduce the number of variables used in the model

Other techniques explored:

1. Chi2 to extract categorical features that were most related to the target. Target (continuous) was split into ordinal bins (for chi2 to work)
2. VIF to eliminate highly correlated numerical features

Feature Engineering Results

All iterations of our feature engineering was run through a function that compared Ridge, Lasso & Normal Linear regression models before choosing the one which had the best score & Metrics. Hyper Parameters were tuned using GridSearch. The results for the various iterations are as follows:

Test Name	Model used	Alpha	R-squared	Mean Absolute Error	Public Score	Private Score
No Feature Engineering	ridge	250.4640	0.9025	16071.1293	29427	30164
1st Feature Engineering	ridge	289.4870	0.9025	15991.0874	29207	30160
2nd Feature Engineering	ridge	321.3835	0.9233	14407.5733	23529	32005
Recursive Feaeture Elimination - 200	ridge	9.1011	0.8805	16587.5144	28388	27568
Recursive Feaeture Elimination - 150	ridge	1.0797	0.8846	16292.8619	28395	27219
Recursive Feaeture Elimination - 250	ridge	3.5818	0.9132	15178.3269	25830	28446

Conclusion

Best 10 coefficients/features:

1	33821.9651	Neighborhood_NoRidge	6	19251.15009	Kitchen Qual_Ex
2	31521.58977	Roof Matl_WdShngl	7	17242.86124	Exter Qual_Ex
3	29624.73447	Neighborhood_StoneBr	8	16726.19396	Garage Type_BuiltIn
4	23916.54094	Garage Qual_Gd	9	16588.55204	Neighborhood_NridgHt
5	22269.32372	Total Bsmt SF	10	15368.51248	Bsmt Exposure_Gd

Conclusion

Worst 10 coefficients/features:

1	-14793.81473	Functional_Sal	6	-7958.539757	Bsmt Unf SF
2	-11655.56086	Heating QC_Po	7	-7812.518201	Bsmt Qual_Gd
3	-10951.82426	MS Zoning_A (agr)	8	-7775.999311	Kitchen Qual_TA
4	-9196.783825	Exter Qual_TA	9	-7549.908939	Neighborhood_Gilbert
5	-8928.719741	Heating_Grav	10	-7276.026174	Neighborhood_OldTown

Mean SalePrice of Best 10 vs Worst 10 features

Neighborhood_NoRidge	$	313,875
Roof Matl_WdShngl	$	354,250
Neighborhood_StoneBr	$	329,676
Garage Qual_Gd	$	262,989
Total Bsmt SF (Large)	$	335,802
Kitchen Qual_Ex	$	339,485
Exter Qual_Ex	$	379,588
Garage Type_BuiltIn	$	247,884
Neighborhood_NridgHt	$	322,831
Bsmt Exposure_Gd	$	272,549

Functional_Sal	$	31,550
Heating QC_Po	$	69,033
MS Zoning_A (agr)	$	47,300
Exter Qual_TA	$	142,840
Heating_Grav	$	65,180
Bsmt Unf SF (Small)	$	142,153
Bsmt Qual_Gd	$	201,858
Kitchen Qual_TA	$	139,502
Neighborhood_Gilbert	$	189,228
Neighborhood_OldTown	$	125,276

Recommendation

- To provide accurate house saleprices, focus on collecting data for these 14 features (consolidated from the 20 features)

1	Neighborhood
2	Roof Matl
3	Garage Qual
4	Total Bsmt SF
5	Kitchen Qual
6	Exter Qual
7	Garage Type

8	Bsmt Exposure
9	Functional
10	Heating QC
11	MS Zoning
12	Heating
13	Bsmt Unf SF
14	Bsmt Qual

Future work

Extra features we could include in future:

- Schools, amenities, expressways, factories/employment centres, major towns within a certain radius
- Vicinity to Iowa State University (largest employer)
- Air quality (eg. NOx levels) in the region