EDA of the Ames Housing Report

Joel Riesen, SPS MSDS Northwestern University

04/11/2021

Section 1: Introduction

The Ames data set submitted by Dean De Cock of Truman State University. The original data contains individual residential properties sold in Ames, IA, from 2006 to 2010. The data includes 82 columns, 23 nominal, 23 ordinals, 14 discrete, and 20 continuous variables. It contains 2930 observations.

Table 1: Ames Housing descriptions from http://ise.amstat.org/v19n3/decock/DataDocumentation.txt

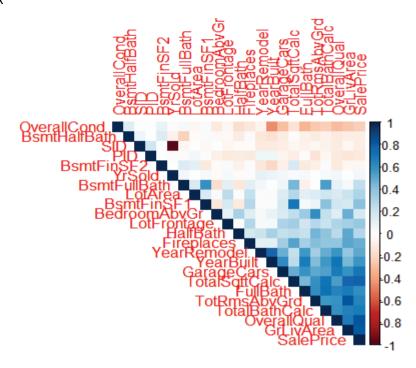
		using descriptions from http://jse.ams	_		
1	Order (Discrete)	Observation number	4 2	HeatingQC (Ordinal)	Heating quality and condition
2	PID (Nominal)	Parcel identification number - can be used	4	Central Air	Central air conditioning
	,	with city website for parcel review.	3	(Nominal)	5
3	MS SubClass	Identifies the type of dwelling involved in	4	Electrical	Electrical system
)		the sale.	4		Electrical system
1	(Nominal)			(Ordinal)	E: 4E1 C 4
4	MS Zoning	Identifies the general zoning classification	4	1st Flr SF	First Floor square feet
	(Nominal)	of the sale.	5	(Continuous)	
5	Lot Frontage (Continuous)	Linear feet of street-connected to property	4 6	2nd Flr SF (Continuous)	Second-floor square feet
6	Lot Area	Lot size in square feet	4	Low Qual Fin	Low quality finished square feet (all
	(Continuous)	-	7	SF (Continuous)	floors)
7	Street	Type of road access to property	4	Gr Liv Area	Above grade (ground) living area
′		Type of foad access to property			
0	(Nominal)	T C 11 4 4	8	(Continuous)	square feet
8	Alley	Type of alley access to property	4	Bsmt Full	Basement full bathrooms
	(Nominal)		9	Bath	
				(Discrete)	
9	Lot Shape	General shape of property	5	Bsmt Half	Basement half bathrooms
	(Ordinal)		0	Bath	
	*			(Discrete)	
1	Land Contour	Flatness of the property	5	Full Bath	Full bathrooms above grade
0	(Nominal)	Timmess of the property	1	(Discrete)	Tun cum comb ucc ve grade
1	Utilities	Type of utilities available	5	Half Bath	Half baths above grade
		Type of unifices available			Trair bails above grade
1	(Ordinal)	T	2	(Discrete)	D 1 1 1 /1 370m
1	Lot Config	Lot configuration	5	Bedroom	Bedrooms above grade (does NOT
2	(Nominal)		3	(Discrete)	include basement bedrooms)
1	Land Slope	Slope of property	5	Kitchen	Kitchens above grade
3	(Ordinal)		4	(Discrete)	
1	Neighborhood	Physical locations within Ames city limits	5	KitchenQual	Kitchen quality
4	(Nominal)	(map available)	5	(Ordinal)	<u> </u>
1	Condition 1	Proximity to various conditions	5	TotRmsAbvGr	Total rooms above grade (does not
5		110Minity to various conditions	6	d (Discrete)	include bathrooms)
	(Nominal)	Donation its 4s are since 112 CC			
1	Condition 2	Proximity to various conditions (if more	5	Functional	Home functionality (Assume typical
6	(Nominal)	than one is present)	7	(Ordinal)	unless deductions are warranted)
1	Bldg Type	Type of dwelling	5	Fireplaces	Number of fireplaces
7	(Nominal)		8	(Discrete)	
1	House Style	Style of dwelling	5	FireplaceQu	Fireplace quality
8	(Nominal)	-	9	(Ordinal)	
1	Overall Qual	Rates the overall material and finish of the	6	Garage Type	Garage location
9	(Ordinal)	house	0	(Nominal)	
2	Overall Cond	Rates the overall condition of the house	6	Garage Yr Blt	Year garage was built
		Rates the overall condition of the house	1		Toar garage was built
0	(Ordinal)		1	(Discrete)	T
2	Year Built	Original construction date	6	Garage Finish	Interior finish of the garage
1	(Discrete)		2	(Ordinal)	
2	Year	Remodel date (same as construction date if	6	Garage Cars	Size of garage in car capacity
2	Remod/Add	no remodeling or additions)	3	(Discrete)	
	(Discrete)				
2	Roof Style	Type of roof	6	Garage Area	Size of garage in square feet
3		1,700 01 1001	4	(Continuous)	Size of garage in square rect
	(Nominal)	D f 1			C1:6
2	Roof Matl	Roof material	6	Garage Qual	Garage quality
	(Nominal)	I	5	(Ordinal)	1

Society of the continuous Continuous	2	Exterior 1	Exterior covering on house	6	Garage Cond	Garage condition
Continuous Con	5	(Nominal)		6	(Ordinal)	
Mas Vnr Type (Nominal)				_		Paved driveway
Refers to walkout or garden level walls Semult Taylous Semult Taylou			Ź	<u> </u>		
Mas Vnr Area Masonry veneer area in square feet 6 Open Porch SF (Continuous)			Masonry veneer type	-		Wood deck area in square feet
Mas Vnr Area (Continuous)	7	(Nominal)		8		
Sexter Qual (Ordinal) Evaluates the quality of the material on the exterior Sexter Cond (Ordinal) Evaluates the present condition of the (Continuous) Type of foundation Type of foundatio						
Continuous Exert Qual Evaluates the quality of the material on the exterior Continuous			Masonry veneer area in square feet			Open porch area in square feet
Exter Qual (Ordinal) Evaluates the quality of the material on the exterior Continuous	8	(Continuous)		9		
Sexter Cond Evaluates the present condition of the Ordinal						
Continuous Sexter Cond Evaluates the present condition of the Continuous Type of foundation Type of fo				1 '		Enclosed porch area in square feet
Section Evaluates the present condition of the (Ordinal)	9	(Ordinal)	exterior	0		
0(Ordinal)material on the exterior1(Continuous)3Foundation (Nominal)Type of foundation7Screen Porch (Continuous)Screen porch area in square feet3Bsmt Qual (Ordinal)Evaluates the height of the basement7Pool Area (Continuous)Pool area in square feet4(Ordinal)Bsmt Cond (Ordinal)Evaluates the general condition of the basement7Pool QC (Ordinal)Pool quality3Bsmt Exposure (Ordinal)Refers to walkout or garden level walls7Fence (Ordinal)Fence quality4(Ordinal)5(Ordinal)Misc Feature (Nominal)Miscellaneous feature not covered in other categories3BsmtFin Type (Continuous)7Misc Val (Continuous)\$Value of miscellaneous feature3BsmtFin Type (Continuous)8(Discrete)3BsmtFin SF 2 (Ordinal)7Mo Sold (MM)4(Continuous)9(Discrete)3Bsmt Unf SF (Continuous)9(Discrete)4Total Bsmt SF (Continuous)8Sale Type (Nominal)4Total Bsmt SF (Continuous)0(Nominal)Condition of sale4Total Bsmt SF (Continuous)1(Nominal)Condition of sale						
Type of foundation Type of	3			7		Three season porch area in square feet
1 (Nominal) 2 (Continuous) 3 Bsmt Qual Evaluates the height of the basement 7 Pool Area (Continuous) 3 Bsmt Cond Evaluates the general condition of the 3 (Ordinal) basement 4 (Ordinal) Continuous 2 (Ordinal) Continuous 2 (Ordinal) Continuous Continuous Condition Condi	0			1	(Continuous)	
3 Bsmt Qual 2 (Ordinal) Evaluates the height of the basement 7 Pool Area Continuous	3	Foundation	Type of foundation	7		Screen porch area in square feet
2 (Ordinal) 3 Bsmt Cond 3 (Ordinal) 4 (Ordinal) 5 Bsmt Exposure 4 (Ordinal) 6 Refers to walkout or garden level walls 7 Fence 7 Fence 8 Fence quality 7 Fence 9 (Ordinal) 7 Fence 9 Fence quality 9 (Ordinal) 7 Fence 9 Fence quality 9 (Ordinal) 9 Rating of basement finished area 1 (Ordinal) 9 Rating of basement finished area 1 (Ordinal) 1 Type 1 finished square feet 1 (Continuous) 1 SwmtFin SF 1 1 (Continuous) 1 SwmtFin SF 2 1 (Ordinal) 2 (Ordinal) 3 BsmtFin SF 2 4 (Continuous) 1 Type 1 finished square feet 1 Misc Val 9 (Continuous) 1 SValue of miscellaneous feature 1 Month Sold (MM) 1 Month Sold (MM) 1 Year Sold (YYYY) 2 (Continuous) 3 Bsmt Unf SF 9 (Continuous) 4 Total Bsmt SF 1 Cotal square feet of basement area 1 Sale Condition 1 (Nominal) 2 Condition of sale	1	(Nominal)		2	(Continuous)	
Bsmt Cond Evaluates the general condition of the 3 (Ordinal) basement 4 (Ordinal)	3	Bsmt Qual	Evaluates the height of the basement	7	Pool Area	Pool area in square feet
3 (Ordinal) basement 4 (Ordinal)	2	(Ordinal)		3	(Continuous)	_
Refers to walkout or garden level walls Fence Fence quality	3	Bsmt Cond	Evaluates the general condition of the	7	Pool QC	Pool quality
4(Ordinal)5(Ordinal)3BsmtFin Type 5Rating of basement finished area7Misc Feature 6Miscellaneous feature not covered in 63BsmtFin SF 1 6Type 1 finished square feet7Misc Val 7\$Value of miscellaneous feature4(Continuous)Year Sold (MM)3BsmtFinType 7Rating of basement finished area (if multiple types)7Mo Sold 8Month Sold (MM)3BsmtFin SF 2 8Type 2 finished square feet7Yr Sold 9Year Sold (YYYY)4Total Bsmt SF 9Unfinished square feet of basement area8Sale Type 0Type of sale4Total Bsmt SF 9Total square feet of basement area8Sale Condition 1Condition of sale	3	(Ordinal)	basement	4	(Ordinal)	•
4 (Ordinal)5 (Ordinal)3 BsmtFin Type 5 1 (Ordinal)Rating of basement finished area7 Misc Feature 6 (Nominal)Miscellaneous feature not covered in 6 (Nominal)3 BsmtFin SF 1 6 (Continuous)Type 1 finished square feet7 Misc Val 7 (Continuous)\$Value of miscellaneous feature3 BsmtFinType 7 2 (Ordinal)Rating of basement finished area (if multiple types)7 Mo Sold 8 (Discrete)Month Sold (MM)3 BsmtFin SF 2 8 (Continuous)Type 2 finished square feet7 Yr Sold 9 (Discrete)Year Sold (YYYY)3 Bsmt Unf SF 9 (Continuous)Unfinished square feet of basement area 0 (Nominal)8 Sale Type 0 (Nominal)Type of sale4 Total Bsmt SF 0 (Continuous)Total square feet of basement area8 Sale Condition 1 (Nominal)Condition of sale	3	Bsmt Exposure	Refers to walkout or garden level walls	7	Fence	Fence quality
5 1 (Ordinal) 3 BsmtFin SF 1 (Continuous) 3 BsmtFin SF 1 (Continuous) 3 BsmtFinType Rating of basement finished area (if year Sold (MM)) 7 2 (Ordinal) multiple types) 3 BsmtFin SF 2 Type 2 finished square feet 7 Yr Sold Year Sold (YYYY) 8 (Continuous) 9 (Discrete) 3 Bsmt Unf SF (Continuous) 1 Bsmt Unf SF (Continuous) 1 Total Bsmt SF (Continuous) 4 Total Bsmt SF (Continuous) 1 Total square feet of basement area 0 (Continuous) 6 (Nominal) other categories 7 Misc Val (Continuous) 8 (Discrete) 7 Yr Sold Year Sold (YYYY) 9 (Discrete) 7 Type of sale 1 (Nominal) Condition of sale	4	(Ordinal)	_	5	(Ordinal)	
5 1 (Ordinal) 3 BsmtFin SF 1 (Continuous) 3 BsmtFin SF 1 (Continuous) 3 BsmtFinType Rating of basement finished area (if year Sold (MM)) 7 2 (Ordinal) multiple types) 3 BsmtFin SF 2 Type 2 finished square feet 7 Yr Sold Year Sold (YYYY) 8 (Continuous) 9 (Discrete) 3 Bsmt Unf SF (Continuous) 1 Bsmt Unf SF (Continuous) 1 Total Bsmt SF (Continuous) 4 Total Bsmt SF (Continuous) 1 Total square feet of basement area 0 (Continuous) 6 (Nominal) other categories 7 Misc Val (Continuous) 8 (Discrete) 7 Yr Sold Year Sold (YYYY) 9 (Discrete) 7 Type of sale 1 (Nominal) Condition of sale	3	BsmtFin Type	Rating of basement finished area	7	Misc Feature	Miscellaneous feature not covered in
3 BsmtFin SF 1 Type 1 finished square feet 7 Misc Val \$Value of miscellaneous feature 7 (Continuous) 3 BsmtFinType Rating of basement finished area (if 7 Mo Sold Month Sold (MM) Mo	5			6	(Nominal)	other categories
6 (Continuous) 7 (Continuous) 8 BsmtFinType Rating of basement finished area (if pulliple types) 8 (Discrete) 7 Wo Sold Month Sold (MM) 8 (Discrete) 9 (Continuous) 1 Type 2 finished square feet pulliple types			Type 1 finished square feet	7	Misc Val	
3 BsmtFinType Rating of basement finished area (if 7 Mo Sold Month Sold (MM) 7 2 (Ordinal) multiple types) 8 (Discrete) 8 (BsmtFin SF 2 Type 2 finished square feet 7 Yr Sold Year Sold (YYYY) 9 (Discrete) Type of sale 1 (Sold Year Sold (YYYY) 9 (Discrete) Type of sale 1 (Sold Year Sold (YYYY) 1 (Sold Year Sold (YYYY) 1 (Sold Year Sold (YYYY) 2 (Discrete) Type of sale 3 (Sold Year Sold (YYYY) 9 (Discrete) Type of sale 1 (Sold Year Sold (YYYY) 1 (Sold Year Sold (YYYY) 1 (Sold Year Sold (YYYY) 2 (Sold Year Sold (YYYY) 3 (Sold Year Sold (YYYY) 4 (Sold Year Sold (YYYY) 5 (Sold Year Sold (YYYY) 6 (Sold Year Sold (YYYY) 7 (Sold Year Sold (YYYY) 9 (Discrete) 1 (Sold Year Sold (YYYY) 1 (Sold Year Sold (YYYY)	6	(Continuous)		7	(Continuous)	
7 2 (Ordinal) multiple types) 8 (Discrete) 3 BsmtFin SF 2 (Continuous) Type 2 finished square feet 7 Yr Sold (YYYY) 8 (Continuous) 9 (Discrete) 3 Bsmt Unf SF (Continuous) Unfinished square feet of basement area 8 Sale Type (Nominal) 9 (Continuous) 0 (Nominal) 4 Total Bsmt SF (Continuous) Total square feet of basement area 8 Sale Condition (Nominal) 0 (Continuous) 1 (Nominal)	3		Rating of basement finished area (if	7		Month Sold (MM)
3 BsmtFin SF 2 Type 2 finished square feet 7 Yr Sold Year Sold (YYYY) 9 (Discrete) 3 Bsmt Unf SF Unfinished square feet of basement area 8 Sale Type Type of sale (Continuous) 0 (Nominal) 4 Total Bsmt SF Total square feet of basement area 8 Sale Condition Condition of sale (Nominal) 1 (Nominal) 1				8	(Discrete)	,
8 (Continuous) 9 (Discrete) 3 Bsmt Unf SF (Continuous) Unfinished square feet of basement area 8 Sale Type (Nominal) 9 (Continuous) 0 (Nominal) 4 Total Bsmt SF (Continuous) Total square feet of basement area 8 Sale Condition (Nominal) 0 (Continuous) 1 (Nominal)	3					Year Sold (YYYY)
3 Bsmt Unf SF Unfinished square feet of basement area 8 Sale Type Type of sale 0 (Nominal) 4 Total Bsmt SF Total square feet of basement area 8 Sale Condition Condition of sale 0 (Continuous) 5 Sale Type Type of sale 0 (Nominal) 6 Sale Type Type of sale 0 (Nominal) 7 Sale Type Type of sale 0 (Nominal) 8 Sale Type Type of sale 0 (Nominal) 9 Sale Type Type of sale 0 (Nominal) 1 Sale Type Type of sale 0 (Nominal) 2 Sale Type Type of sale 0 (Nominal) 3 Sale Type Type of sale 0 (Nominal) 4 Sale Type Type of sale 0 (Nominal) 5 Sale Type Type of sale 0 (Nominal) 5 Sale Type Type of sale 0 (Nominal) 6 Sale Type Type of sale 0 (Nominal) 7 Sale Type Type of sale 0 (Nominal) 8 Sale Type Type of sale 0 (Nominal) 8 Sale Type Type of sale 0 (Nominal) 8 Sale Type Type of sale 0 (Nominal) 9 Sale Type Type of sale	_		71			(= = = =)
9 (Continuous) 0 (Nominal) 4 Total Bsmt SF Total square feet of basement area 8 Sale Condition Condition of sale 1 (Nominal)			Unfinished square feet of basement area			Type of sale
4 Total Bsmt SF Total square feet of basement area 8 Sale Condition Condition of sale 1 (Nominal)	_			_		71
0 (Continuous) 1 (Nominal)			Total square feet of basement area			Condition of sale
	1 -			1		
	-		Type of heating	8		Sale price \$\$
1 (Nominal) 2 (Continuous)	1		1 JP2 01 nowing	_		2000 price 44

The reason for the analysis is to use the data to fit specific models. We will start with a Simple Linear Regression (SLR). SalesPrice will be our response variable. We will conduct an EDA to find two predictor variables to help us understand the data better. Next, we will complete a Multiple Linear Regression (MLR) using the two predictor variables from the SLR. We will add two more predictor variables which can be continuous or discrete.

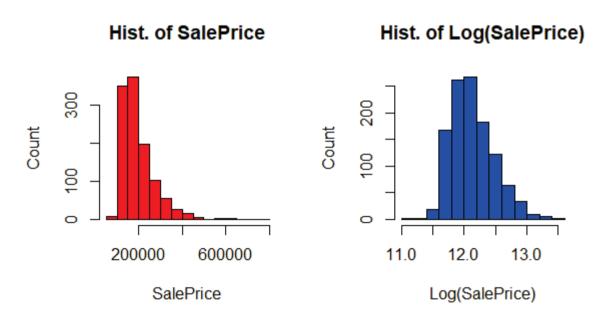
To find the most vital relationship with SalePrice, we will run a correlation matrix heat map to narrow down top relationships. We need to take out all the non-numeric columns first.

Fig1 Correlation Matrix



From this, we can see that GrLivArea, OverallQual, TotalBathCalc, and TotRmsAbvGrd. Looking at the data, we should see if Logging SalePrice will help to smooth out the data.

Fig 2: Histogram of logged SalesPrice v SalePrice

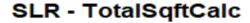


Looking at the graphs, the logged SalePrice looks more like a normal distribution than the non-logged SalePrice. So we will use the logged SalesPrice; the predictor variables will be GrLivArea, and TotalSqftCalc.

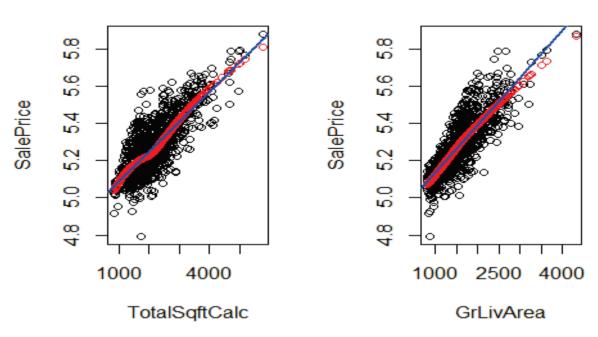
Section 2: Simple Linear Regression Models

The predictor variables are GrLivArea, and TotalSqftCalc. Both deal with the square feet of the house. The TotalSqftCalc also looks to include the basement.

Fig 3: SLR of the two predictor variables



SLR - GrLivArea



From the graphs, the blue line is the fitted linear model, and the red is the Loess local fit. TotalSqftCalc is a little more accurate in helping to predict the SalePrice.

Section 2.3 Model Comparison

Comparing the two predictors and how they relate to SalePrice

Table 2: Linear Regression Results

Fitted SLR

	Dependent variable:							
	SalePrice							
	TotalSqftCalc	GrLivArea						
	(1)	(2)						
TotalSqftCalc	0.0002*** (0.00000)						
GrLivArea	•	0.0003*** (0.00001)						
Constant	4.930*** (0.009)	4.879*** (0.008)						
Observations	1,135	1,135						
\mathbb{R}^2	0.589	0.694						
Adjusted R ²	0.588	0.694						
Residual Std. Error (df = 1133)	0.095	0.082						
F Statistic (df = 1; 1133)	1,620.627***	2,574.373***						
Note:	*p<0.1	; **p<0.05; ***p<0.01						

Both have an \mathbb{R}^2 above 50%; however, GrLivArea is almost 70%. So I would make sure to include it in the prediction.

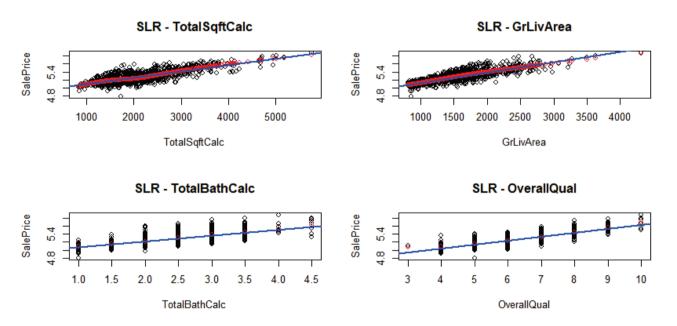
Putting both the TotalSqftCalc and GrLivArea for a multiple linear regression model. It does help a little bit pushing the R^2 up over 70%.

Table 3: The two predictors combined

Fitted MLR									
	Dependent variable:								
_	SalePrice								
	TotalSqftCalc	GrLivArea	Combined						
	(1)	(2)	(3)						
TotalSqftCalc	0.0002*** (0.00000)		0.0001*** (0.00001)						
GrLivArea		0.0003*** (0.00001)	0.0002*** (0.00001)						
Constant	4.930*** (0.009)	4.879*** (0.008)	4.854*** (0.008)						
Observations	1,135	1,135	1,135						
R ²	0.589	0.694	0.733						
Adjusted R ²	0.588	0.694	0.733						
Residual Std. Error	0.095 (df = 1133)	0.082 (df = 1133)	0.076 (df = 1132)						
F Statistic 1	,620.627*** (df = 1; 1133)	2,574.373*** (df = 1; 113	3) 1,554.249*** (df = 2; 1132)						
Note:			*p<0.1; **p<0.05; ***p<0.01						

To see if we can increase this by adding more predictor variables. Using the subsequent two strongest correlations, we will add OverallQual and TotalBathCalc.

Fig 4: The four predictors



The lines are in the above charts are very close together. Adding the extra two predictors for four predictors, we now have an \mathbb{R}^2 of almost 90%.

Table 4: Linear regression results of the four predictors

		Fit	ted MLR		
_			SalePrice		
	TotalSqftCalc	GrLivArea	TotalBathCalc	OverallQual	
	(1)	(2)	(3)	(4)	(5)
TotalSqftCalc	0.0002*** (0.00000)				0.00005*** (0.00000)
GrLivArea		0.0003*** (0.00001)			0.0001*** (0.00001)
TotalBathCalc			0.144*** (0.004)		0.021*** (0.003)
OverallQual				0.097*** (0.002)	0.057*** (0.002)
Constant	4.930*** (0.009)	4.879*** (0.008)	4.931*** (0.010)	4.655*** (0.011)	4.644*** (0.007)
Observations	1,135	1,135	1,135	1,135	1,135
\mathbb{R}^2	0.589	0.694	0.531	0.739	0.890
Adjusted R ²	0.588	0.694	0.530	0.739	0.890
Residual Std. Error	0.095 (df = 1133)	0.082 (df = 1133)	0.101 (df = 1133)	0.075 (df = 1133)	0.049 (df = 1130)
F Statistic 1	,620.627*** (df = 1; 1133)	2,574.373*** (df = 1; 1133)	1,280.411*** (df = 1; 1133)	3,215.188*** (df = 1; 1133) 2,283.792*** (df = 4; 1130
Note:					*p<0.1· **p<0.05· ***p<0.0

Section 5: Transformed MLR Model

So did we make the right call by transforming SalePrice and using it for this report?

Table 5: Comparison of logged v non logged.

SalesPrice not logged(transformed)

SID PI	D LotFrontag	ge LotArea	overallQual	OverallCond	YearBuilt	YearRemode	1 BsmtFinSF	1 BsmtFinSF	2 GrLivAre	a BsmtFullBat	h BsmtHalfBat	h FullBath	HalfBath	BedroomAbvC	r TotRmsAbvGrd
SalePrice -0.10 -0.	12 0.36	0.29	0.82	-0.24	0.58	0.50	0.40	-0.05	0.81	0.24	-0.10	0.61	0.27	0.18	0.65
SalesPrice logged															
SID PI	D LotFrontag	ge LotArea	o OverallQual	OverallCond	YearBuilt	YearRemode	l BsmtFinSF	1 BsmtFinSF	2 GrLivAre	a BsmtFullBa	th BsmtHalfBa	th FullBath	n HalfBath	BedroomAbv	Gr TotRmsAbvGrd
SalePrice -0.10 -0.	10 0.36	0.28	0.86	-0.26	0.65	0.58	0.36	-0.07	0.83	0.22	-0.11	0.69	0.31	0.21	0.68

The graphs confirm the log, so I think it was a good choice.

Fig 5: Log vs. non-log

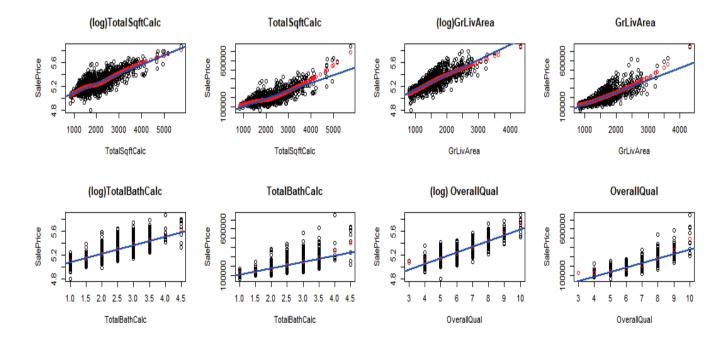


Table 6: The linear comparison

	Fitted MLR						
	Dependent variable:						
	SalePrice						
	log-SalePrice	SalePrice					
	(1)	(2)					
TotalSqftCalc	0.00005*** (0.00000)	41.437*** (2.152)					
GrLivArea	0.0001*** (0.00001)	30.377*** (3.328)					
TotalBathCalc	0.021*** (0.003)	-1,878.221 (1,807.801)					
OverallQual	0.057*** (0.002)	29,558.380*** (968.756)					
Constant	4.644*** (0.007)	-119,037.900*** (4,437.400)					
Observations	1,135	1,135					
\mathbb{R}^2	0.890	0.851					
Adjusted R ²	0.890	0.850					
Residual Std. Error (df = 1130)	0.049	29,913.560					
F Statistic (df = 4; 1130)	2,283.792***	1,611.802***					
Note:		*p<0.1; **p<0.05; ***p<0.01					

Conclusion

Running through the EDA process, I found a lot of data that I didn't lose. I also found a few outliers that I left in, and I am happy that I did. From the graph of the GrLivArea, vs. the log, you can see the outliers. I think adding a few more predictor variables would help a little bit more in explaining the SalePrice, but I think it would help with increasing its accuracy as well.