Housing price guidance for real estate market in Ames Iowa

Justin Cleveland
Sanjay Pillay
Yang Zhang
Malcolm Carlson
12/06/2019

Introduction 2
Data Description 2
Analysis Question 1: 2
Assumptions: 2
Build and Fit the Model 2
Interpretation: 4
Conclusion: 4
Analysis Question 2 5
Assumptions: Use similar EDA assumptions from analysis 1 by employing a reduced range on the above grade (ground) living area square feet, and limited our analysis to homes with less than 4500sqft. 5
Variable Selection 5
Forward Selection 5
Backwards Selection 6
Stepwise Selection 6
Custom Model 7
Model Statistics 7
Conclusion 8
APPENDIX 9
SAS Code for Analysis 1 9
Matrix Plot Investating Variable Relationship 12
Variable Selection 15
Variable Imputation Code 16
Forward Regression VIF 20
Forward Regression Parameters 21
Forward Regression Residual Plots 23
Forward Regression Prediction Code 26
Backwards Regression Parameter Estimates 27
Backward Regression Plots 28
Backward Regression Code Chunks 29
Stepwise Regression Parameter Estimates 30
Stepwise Regression Plots 31
Stepwise Regression Code Chunks 31
Custom Model Parameter Estimates 33
Custom Model Regression Plots 33

Introduction

The purpose of this study is to analyze a real estate data set to explore relationships between key factors that can influence the final sales price for a house. These data describe house sales information in Ames Iowa from 2006 to 2010. Yang, Justin, Sanjay and I will use various statistical techniques to answer questions of interest such as what would an estimate of the sale price be for a house considering square footage of living space and does that relationship to square footage depend on neighborhood location? We will also build four different models, Forward Pass, Backwards Pass, Stepwise and a custom model to identify the most predictive model for home pricing in all neighborhoods of Ames Iowa.

Data Description

The data set used for this study was compiled by Dean De Cock and is hosted by kaggle.com. There are 2930 observations (23 nominal, 23 ordinal, 14 discreet and 20 continuous). De Cock,Dean “Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project” Journal of Statistics Education Volume 19, Number 3(2011), www.amstat.org/publications/jse/v19n3/decock.pdf

Analysis Question 1:

Our first question of interest is what effect does square footage have on the sales price and is there a relationship between the sales price of homes between the Brookside, Edwards and North Ames neighborhoods in Ames Iowa?

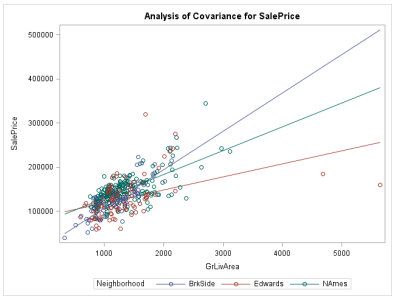
Assumptions:

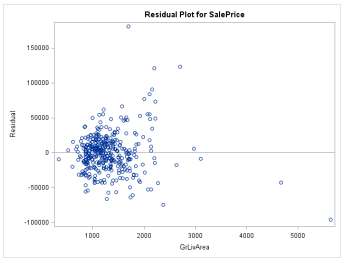
Our initial look at the residual and fit plots point to not meeting all of the assumptions for regression of linearity, normality of residuals, constant variance and independence. There appears to be a lack of constant variance as seen in the plots below. There also appears to be some outliers that may have influence on the data. We will assume the data is independent and that we will be able to employ adjustments methods to address the constant variance and outliers.

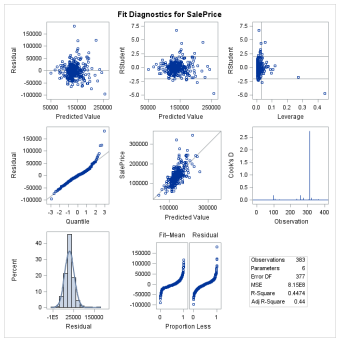
Build and Fit the Model

The first model used in the investigation is μ{(SalePrice|GrLivArea,BrkSide|Edwards+NAmes+GrLivArea*BrkSide+GrLivArea*Edwards+GrLiveArea*NAmes)}=8.50+.47GrLivArea-2.6BrkSide-1.6Edwards+NAmes+.35GrLivArea*BrkSide+2.0GrLivArea*Edwards. Code can be found here.

Also the residual plot does not show good variance and also indicates outliers that could affect sale price estimates. The residuals look normally distributed and the variables are assumed to be independent of each other.





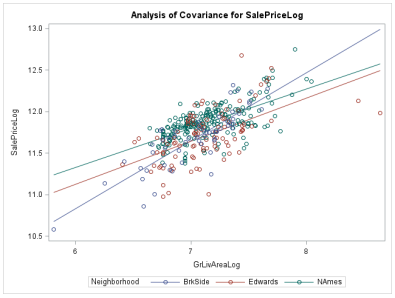


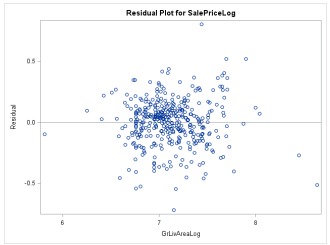
Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		74676.40154	B	6337.89399	11.78	<.0001
GrLivArea		54.31586	B	4.61364	11.77	<.0001
Neighborhood	BrkSide	-54704.88774	B	13882.33364	-3.94	<.0001
Neighborhood	Edwards	13676.70324	B	9097.57465	1.50	0.1336
Neighborhood	NAmes	0.00000	B	.	.	.
GrLivArea*Neighborho	BrkSide	32.84667	B	10.81538	3.04	0.0026
GrLivArea*Neighborho	Edwards	-24.56556	B	6.36139	-3.86	0.0001
GrLivArea*Neighborho	NAmes	0.00000	B	.	.	.

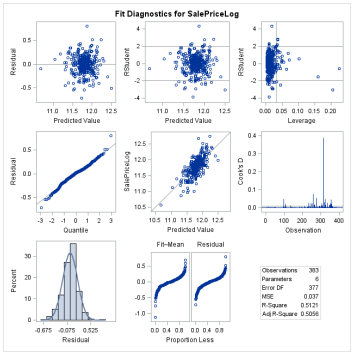
Sum of Residuals	1.8771971E-9
Sum of Squared Residuals	307343083951
Sum of Squared Residuals - Error SS	-0.000183105
PRESS Statistic	340967460744
First Order Autocorrelation	-0.033584549
Durbin-Watson D	2.0654468952

Observations	383
Parameters	6
Error DF	377
MSE	8.15E8
R-Square	0.4474
Adj R-Square	0.44

Because of the lack of constant variance and non linearity to this model. A log of both the SalePrice and GrLivArea parameters was used to meet the assumptions. μ{log(SalePrice|GrLivAreaLog,BrkSide|Edwards+NAmes+GrLivArea*BrkSide+GrLivArea*Edwards+GrLiveArea*NAmes)}=8.50+.47GrLivAreaLog-2.6BrkSide-1.6Edwards+NAmes+.35GrLivAreaLog*BrkSide+2.0GrLivAreaLog*Edwards.





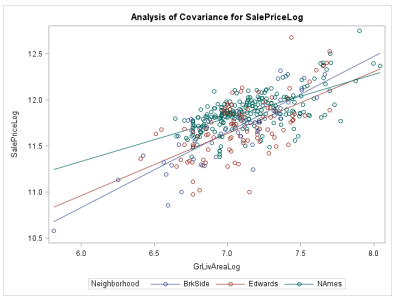


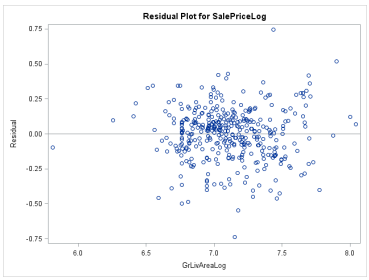
Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		8.492727641	B	0.32441709	26.18	<.0001
GrLivAreaLog		0.473023602	B	0.04542895	10.41	<.0001
Neighborhood	BrkSide	-2.579806905	B	0.59988132	-4.30	<.0001
Neighborhood	Edwards	-0.486220461	B	0.51750833	-0.94	0.3481
Neighborhood	NAmes	0.000000000	B	.	.	.
GrLivArea*Neighborho	BrkSide	0.346624454	B	0.08482008	4.09	<.0001
GrLivArea*Neighborho	Edwards	0.046643642	B	0.07248011	0.64	0.5203
GrLivArea*Neighborho	NAmes	0.000000000	B	.	.	.

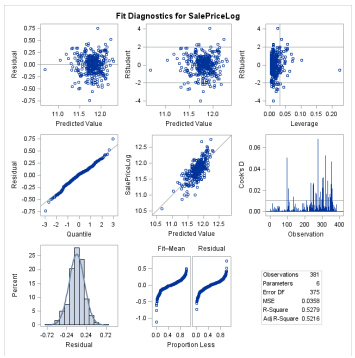
Sum of Residuals	-0.0
Sum of Squared Residuals	13.9
Sum of Squared Residuals - Error SS	-0.0
PRESS Statistic	14.6
First Order Autocorrelation	-0.0
Durbin-Watson D	2.0

Observations	383
Parameters	6
Error DF	377
MSE	0.037
R-Square	0.5121
Adj R-Square	0.5056

Due to the large Cook's D indicated an influential outlier that needed to be addressed in order to ensure the fit of the model. We address the outlier by reducing the range of our analysis to exclude GrLivArea values larger than 4500 square feet.







Parameter		Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept		8.492727641 B	0.31916233	26.61	<.0001	7.865155507	9.120299776
GrLivAreaLog		0.473023602 B	0.04469311	10.58	<.0001	0.385143083	0.560904121
Neighborhood	BrkSide	-2.579806905 B	0.59016472	-4.37	<.0001	-3.740253790	-1.419360020
Neighborhood	Edwards	-1.569634752 B	0.58560055	-2.68	0.0077	-2.721107064	-0.418162440
Neighborhood	NAmes	0.000000000 B	.	.	.	.	.
GrLivArea*Neighborho BrkSide		0.346624454 B	0.08344620	4.15	<.0001	0.182543340	0.510705568
GrLivArea*Neighborho Edwards		0.200314013 B	0.08227532	2.43	0.0154	0.038535213	0.362092813
GrLivArea*Neighborho NAmes		0.000000000 B	.	.	.	.	.

Sum of Residuals	-0.0000000
Sum of Squared Residuals	13.4183281
Sum of Squared Residuals - Error SS	-0.0000000
PRESS Statistic	13.9480700
First Order Autocorrelation	-0.0433249
Durbin-Watson D	2.0838189

Observations	381
Parameters	6
Error DF	375
MSE	0.0358
R-Square	0.5279
Adj R-Square	0.5216

Interpretation:

To achieve the best fit, we log transformed the SalePrice and GrvLivArea variables. We also needed to make adjustments to our range in order to account for an influential outlier that caused our Cook's D value to be above 2.5. The combination of reducing our square footage range for NAmes, Edwards and Brookside neighborhoods along with log transforming our response and explanatory variable allowed us to meet assumptions. Our final equation:

μ{log(SalePrice|GrLivAreaLog,BrkSide|Edwards+NAmes+GrLivArea*BrkSide+GrLivArea*Edwards+GrLiveArea*NAmes)}=8.50+.47GrLivAreaLog-2.6BrkSide-1.6Edwards+.35GrLivAreaLog*BrkSide+2.0GrLivAreaLog*Edwards.

Conclusion:

The square footage and neighborhood explains ~52% of the sale price of a house in North Ames, Edwards and Brookside neighborhoods in Iowa. A doubling in square feet is associated with a $2^{.473}$ or 1.39 multiplicative change in the median sales price (95% confidence interval of 1.47 and 1.75). Houses in BrkSide are 2.6 times and houses in Edwards are 1.6 times less than the ones in NAmes for the same Sq footage.

Analysis Question 2

Our Second question of interest is to build the most predictive model for sales prices of homes in all of Ames Iowa. We will build four models based on

1. Forward Selection
2. Backward Selection
3. Stepwise Selection
4. Custom Model

We will finally compare the Adjusted R^2 , CV PRESS and the Kaggle Score

Assumptions: Use similar EDA assumptions from analysis 1 by employing a reduced range on the above grade (ground) living area square feet, and limited our analysis to homes with less than 4500sqft.

Variable Selection

In our EDA of the full dataset to identify a list of explanatory variables, we initially found 27 numerical variables for our test model and 54 character variables which were excluded. However, upon further investigation of the data descriptions we identified 18 ordinal variables out of the 54 character variables which showed a ranking structure of numerical importance. We imputed these 18 ordinal variables based on the ranking structure in the data_description.txt file, and included these into our test model. The list of variables is filed in the Appendix under Variable Selection.

In addition, we manually imputed the variable “neighborhood“ from the dataset. In this process we converted neighborhood from a character variable to an ordinal variable based on the average sale price of the 25 neighborhoods. A chart of this grouping can be found in the Variable Selection appendix page.

Forward Selection

After investigating the relationship between the variables a subset of variables having relation to the house sales prices were analysed using Forward Selection process, and re analysed to make sure the ones used satisfied with Variation Inflation Factor below 10.

The graphical analysis of the forward model after accounting for the outliers from Analysis 1, we see see normality being satisfied with a normal residual plot and a fairly straight QQ plot. Distribution is accounted for looking at the random cloud residual plots for all identified variables, also we do not see any out liners with significant leverage in the cock’s D plot.

Regression Equation using selected parameters

μ{log(SalePrice|Hood ExteriorQual MSSubClass GarageC HmFunctional HeatingQ BsmtExposur Fireplaces WoodDeckSF ScreenPorch BsmtFullBath KitchenAbvGr GarageCars GarageArea BsmtFinSF1 OverallQual OverallCond TotalBsmtSF GrLivArea YearBuilt YearRemodAdd LotArea GrLivAreaLog) = 2.96 + 0.065*Hood + 0.024*ExteriorQual + -0.0004*MSSubClass + 0.012*GarageC + 0.03*HmFunctional + 0.011HeatingQ + 0.016*BsmtExposur + 0.03*Fireplaces + 0.000053*WoodDeckSF + 0.00024*ScreenPorch + 0.025*BsmtFullBath - 0.052*KitchenAbvGr +

0.022*GarageCars + 0.000089* GarageArea + 0.00007*BsmtFinSF1 + 0.053*OverallQual + 0.044*OverallCond + 0.00009*TotalBsmtSF + 0.0001*GrLivArea +0.002*YearBuilt + 0.0008*YearRemodAdd + 0.000001*LotArea

Code for predictions

Backwards Selection

In the backwards selection model, we utilized the earlier investigation used for the Forward selection. After running the Backwards selection modeling, we checked the Variation Inflation Factors again to verify no values exceeded 10.

The graphical analysis of the backward model after accounting for the outliers from Analysis 1 normality is being satisfied with a normal residual plot and QQ plot. Random cloud residual plots look good and we do not see any out outliers with significant leverage in the cock's D plot.

Regression Equation using Backwards selected parameters

μ{log(SalePrice|Intercept,Hood,ExteriorQual,KitchQual,MSSubClass,BsmtFinSF2, LowQualFinSF,FenceQ,PoolQ,GarageC,GarageQ,FireplaceQ,HmFunctional,HeatingQ, BsmtFinType2q,BsmtExposur,ExterCondQ,Fireplaces,WoodDeckSF,EnclosedPorch, ScreenPorch,BsmtFullBath,HalfBath,KitchenAbvGr,GarageCars,BedroomAbvGr TotRmsAbvGrd,OpenPorchSF,BsmtUnfSF,GarageArea,BsmtFinSF1,OverallQual,OverallCond GrLivArea,YearBuilt,YearRemodAdd,YrSold LotArea _3SsnPorch _2ndFlrSF, GrLivAreaLog)}=11.728889+0.064455*Hood+0.018219*ExteriorQual+0.008328*KitchQual-0.000423*MSSubClass+0.000144*BsmtFinSF2-0.000080303*LowQualFinSF-0.003412*FenceQ+0.022557*PoolQ+0.009194*GarageC+0.005593*GarageQ+0.004821*FireplaceQ+0.031659*HmFunctional+0.011941*HeatingQ-0.006065*BsmtFinType2q+0.015673*BsmtExposur-0.005471*ExterCondQ+0.018064*Fireplaces+0.000059990*WoodDeckSF+0.000072726*EnclosedPorch+0.000246*ScreenPorch+0.023737*BsmtFullBath+0.013956*HalfBath-0.055591*KitchenAbvGr+0.020627*GarageCars-0.007991*BedroomAbvGr+0.008234*TotRmsAbvGrd+0.000074386*OpenPorchSF+0.000088185*BsmtUnfSF+0.000080001*GarageArea+0.000163*BsmtFinSF1+0.051710*OverallQual+0.045942*OverallCond+0.000143*GrLivArea+0.002066*YearBuilt+0.000748*YearRemodAdd-0.004241*YrSold+0.000001905*LotArea+0.000125*_3SsnPorch-0.000018858*_2ndFlrSF+0.222701*GrLivAreaLog

Code for Backwards Model Predictions

Stepwise Selection

We continue to use the initial investigation before the Forward selection model for our Stepwise selection model. Our Variation Inflation Factors stay within acceptable limits under 10.

The graphical analysis of the stepwise model after accounting for the outliers from Analysis 1 normality is being satisfied with a normal residual plot and QQ plot. Random cloud residual plots look good and we do not see any out outliers with significant leverage in the cock's D plot.

Stepwise Regression equation selected parameters

μ{log(SalePrice| Hood,ExteriorQual,MSSubClass,GarageC,HmFunctional,HeatingQ BsmtExposur,Fireplaces,ScreenPorch,BsmtFullBath,KitchenAbvGr,GarageArea,BsmtFinSF1 OverallQual,OverallCond,TotalBsmtSF,GrLivArea,YearBuilt,YearRemodAd, LotArea GrLivAreaLog)}=2.720683+0.066412*Hood+0.024241*ExteriorQual-0.000433*MSSubClass+0.013708*GarageC+0.031567*HmFunctional+0.011427*HeatingQ+0.017292*BsmtExposur+0.032778*Fireplaces+0.000232*ScreenPorch+0.025858*BsmtFullBath-0.052170*KitchenAbvGr+0.000148*GarageArea+0.000070102*BsmtFinSF1+0.053768*OverallQual+0.044227*OverallCond+0.000093987*TotalBsmtSF+0.000143*GrLivArea+0.002143*YearBuilt+0.000867*YearRemodAd+0.000001946*LotArea+0.240951*GrLivAreaLog

Custom Model

Again, with the custom model we continued to use the initial investigation as in the Forward selection model. Our primary factor in variable selection was to keep coefficients with a p-values less than .0099 in the model and remove all others. We first started with all numerical variables and then trimmed the model variables based on non-significant parameter estimates. Our final model resulted in 21 variables which can be reviewed in the equation below or in the appendix Custom Model Parameters Estimates.

The graphical analysis of the custom model after accounting for the outliers from Analysis 1, the outliers being homes with square feet less than 4500, we found normality being satisfied with the residual plot and QQ plot. Random cloud residual plots look good and we do not see any out outliers with significant leverage in the cock's D plot.

Custom Regression equation:

μ{log(SalePrice|Hood,ExteriorQual,GarageC,HmFunctional,HeatingQ,BsmtExposur,Fireplaces,ScreenPorch,TotRmsAbvGrd,GarageArea,OverallQual,OverallCond,YearBuilt,YearRemodAdd,LotArea,MSSubClass,BsmtUnfSF,BsmtFinSF2,BsmtFinSF1,KitchenAbvGr,GrLivAreaLog)}

=1.84843 + 0.06954(Hood) + 0.02476(ExteriorQual) + 0.01104(GarageC) + 0.03151(HmFunctional) + 0.01214(HeatingQ) + 0.01838(BsmtExposur) + 0.03337(Fireplaces) + 0.00022354(ScreenPorch) + 0.00959(TotRmsAbvGrd) + 0.00016074(GarageArea) + 0.05425(OverallQual) + 0.0436(OverallCond) + 0.002(YearBuilt) + 0.00091241(YearRemodAdd) + 0.00000197(LotArea) + -0.00037265(MSSubClass) + 0.00009064(BsmtUnfSF) + 0.0001392(BsmtFinSF2) + 0.00018859(BsmtFinSF1) + -0.06477(KitchenAbvGr) + 0.40728(GrLivAreaLog)

Model Statistics

Model	RMSE	Adj R-Sq	CVPRESS	Kaggle
Forward	0.11772	0.9146	20.7	0.12885
Backward	0.11733	0.9138	20.67	0.12905
Stepwise	0.11802	0.9128	20.78	0.12918
Custom	0.11904	0.9113	21.13	0.12990

Conclusion

From the four models, the lowest Kaggle score was generated from the Forward Pass model although there was not much separation in scores between the models. We found the RMSE to be the lowest for the Forward pass model as well but the best Adj R-sq score was the Custom model. The lowest CVPress came from the Backwards model. From the results, it appears all of the models perform similarly. Other models may prove to be more effective such as Logistic regression and Random Forest.

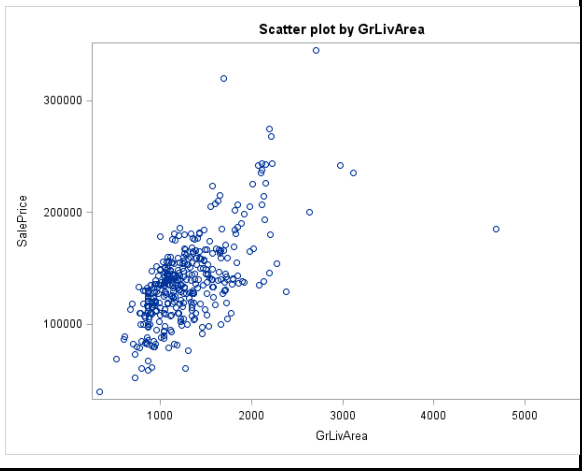
APPENDIX

SAS Code for Analysis 1

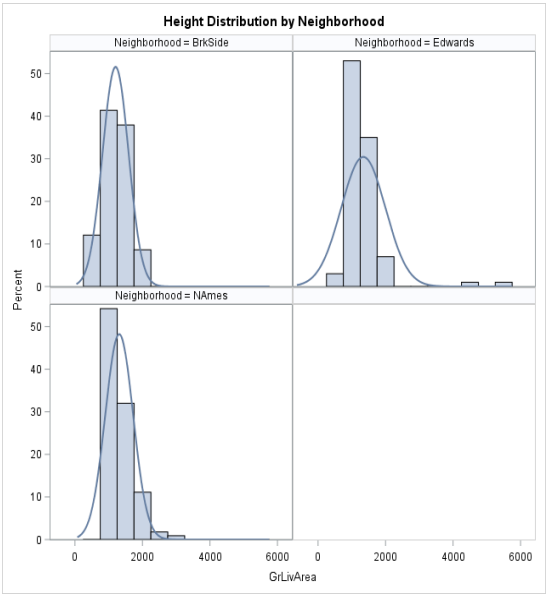
```
/* Sorted to enable removing all but the three neigborhoods for Q1 */
proc sort data = train out=trainsorted;
      by descending Neighborhood;
run;
proc print data = trainsorted;run;

/* Train data set with only NAmes, Edwards and BrkSide neighborhoods */
data trainredux;
set trainsorted;
if (Neighborhood = "NAmes") or (Neighborhood = "Edwards") or (Neighborhood = "BrkSide") then output;
run;
proc print data = trainredux;run;

/* Outliers study from the three neighborhoods */
data outliers;
set trainredux;
if (Neighborhood = "Edwards") AND (GrLivArea > 5000) then output;
if (Neighborhood = "Edwards") AND (SalePrice > 300000) then output;
run;
```

/* Code to identify the outliers by ID */
data outliers;
set trainredux;
if (Neighborhood = "Edwards") AND (GrLivArea > 5000) then output;
run;
proc print data = outliers;	

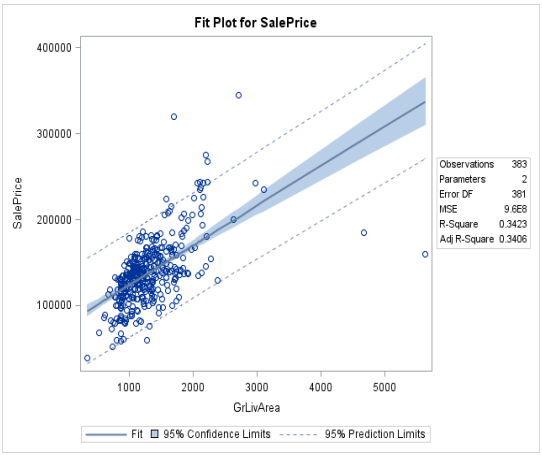
/** Basic Statistical Information **/
proc means data = trainredux nway;
class Neighborhood;
var GrLivArea;
output out = SalePrice median = median1 mean = mean1;
run;	

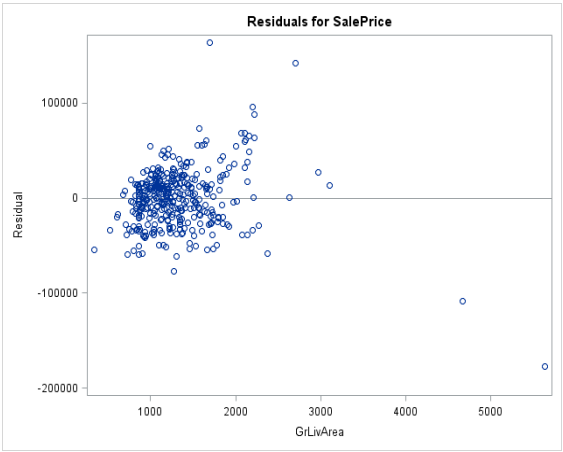
proc sgscatter data=trainredux; plot SalePrice*GrLivArea; run;	

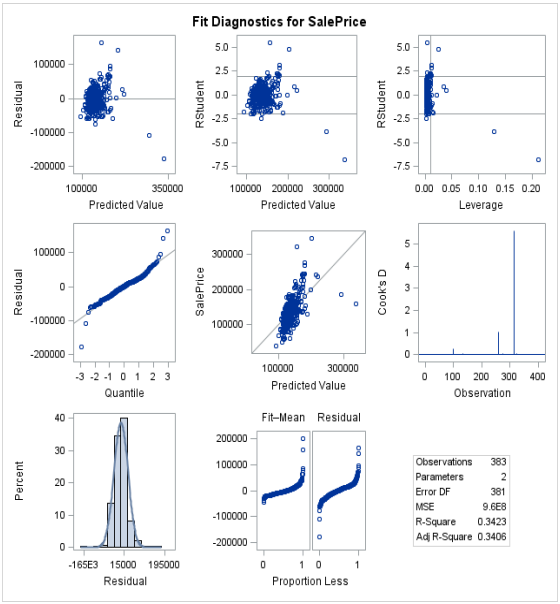
```
proc sgpanel data=trainredux
noautolegend;
title "Height Distribution by
Neighborhood";
panelby Neighborhood;
histogram GrLivArea;
Density GrLivArea;
run;
end;
/** Model Fitting**/
proc reg data= trainredux;
```




```
/** Model Fitting**/
proc reg data= trainredux;
model SalePrice = GrLivArea /
clm cli;
run;
```



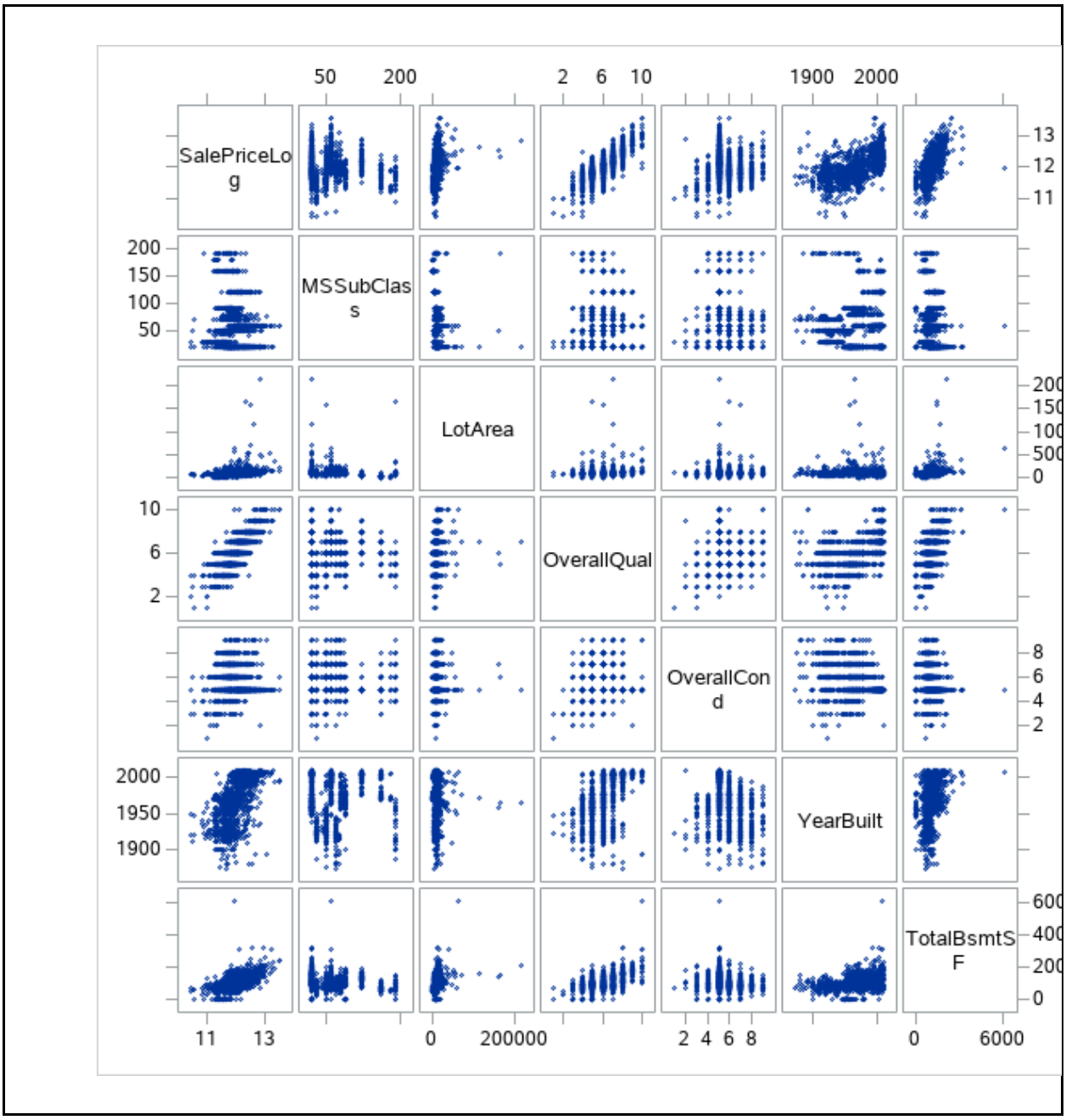


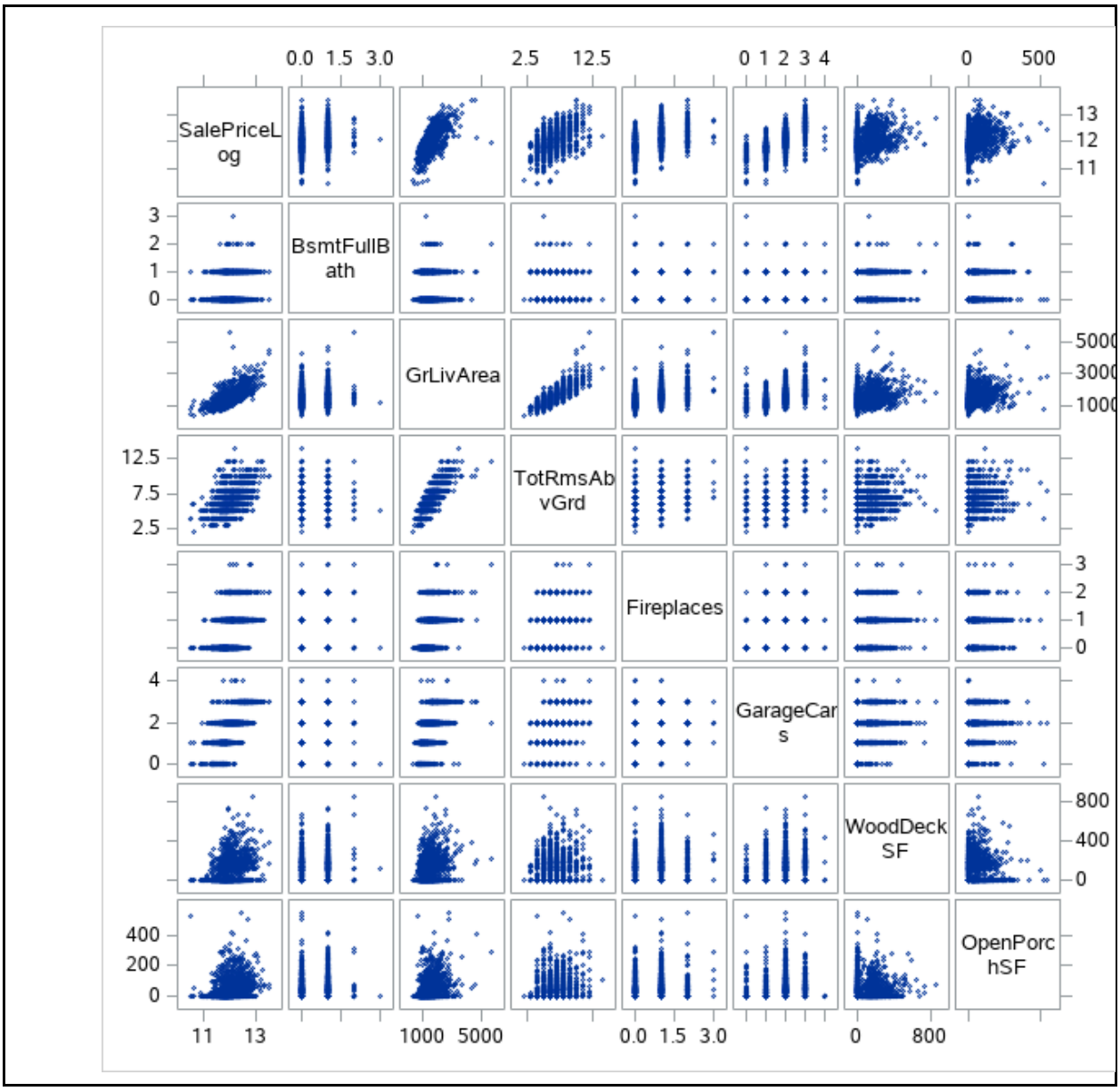
	
Question1	

data trainredux_xylog;
set trainredux;
GrLivAreaLog = log(GrLivArea);
SalePriceLog = log(SalePrice);
run;
 	
proc glm data = trainredux_xylog plots = all;
class Neighborhood;
model SalePriceLog = Neighborhood | GrLivAreaLog / CLM solution;
run;	

Matrix Plot Investating Variable Relationship

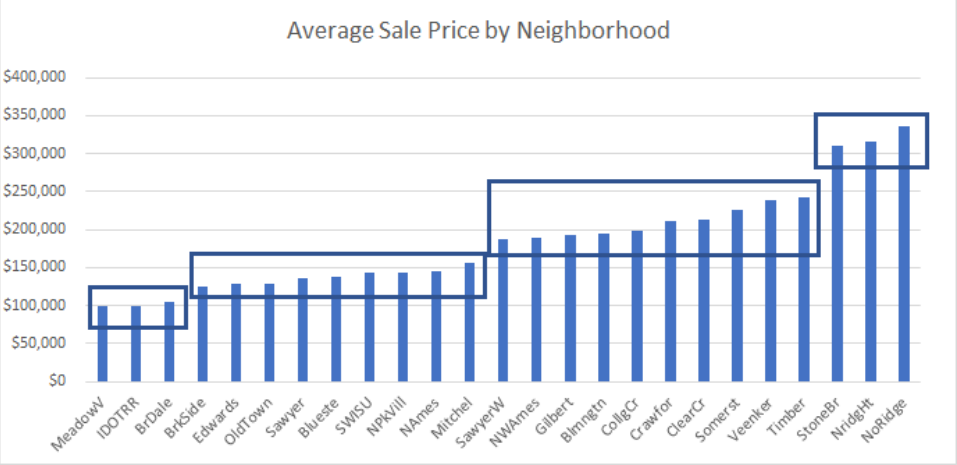
OverallCond


SalePriceLog
MSSubClass
LotArea
OverallQual
OverallCond
YearBuilt
TotalBsmtSF
50
200
2
6
10
1900
2000
13
12
11
200
150
100
50
10
6
2
8
6
4
2
2000
1950
1900
600
400
200
0
11
13
0
200000
2 4 6 8
0
6000


SalePriceLog
BsmtFullBath
GrLivArea
TotRmsAbvGrd
Fireplaces
GarageCars
WoodDeckSF
OpenPorchSF

Variable Selection

Variable	Variable Type	Desc
BedroomAbvGr	Num	Bedroom: Bedrooms above grade (does NOT include basement bedrooms)
EnclosedPorch	Num	Enclosed porch area in square feet
Fireplaces	Num	Number of fireplaces
FullBath	Num	Full bathrooms above grade
GrLivArea	Num	Above grade (ground) living area square feet
HalfBath	Num	Half baths above grade
KitchenAbvGr	Num	Kitchen: Kitchens above grade
LotArea	Num	Lot size in square feet
LowQualFinSF	Num	Low quality finished square feet (all floors)
MiscVal	Num	$Value of miscellaneous feature
MoSold	Num	Month Sold (MM)
MSSubClass	Num	Identifies the type of dwelling involved in the sale.
OpenPorchSF	Num	Open porch area in square feet
OverallCond	Num	Rates the overall condition of the house
OverallQual	Num	Rates the overall material and finish of the house
PoolArea	Num	Pool area in square feet
ScreenPorch	Num	Screen porch area in square feet
TotRmsAbvGrd	Num	Total rooms above grade (does not include bathrooms)
WoodDeckSF	Num	Wood deck area in square feet
YearBuilt	Num	Original construction date
YearRemodAdd	Num	Remodel date (same as construction date if no remodeling or additions)
YrSold	Num	Year Sold (YYYY)
1stFlrSF	Num	First Floor square feet
2ndFlrSF	Num	Second floor square feet
3SsnPorch	Num	Three season porch area in square feet
BasemtQual	Imputed	Evaluates the general condition of the basement
BsmtCondQ	Imputed	Evaluates the height of the basement
BsmtExposur	Imputed	Refers to walkout or garden level walls
BsmtFinType1q	Imputed	Rating of basement finished area
BsmtFinType2q	Imputed	Rating of basement finished area (if multiple types)
ExterCondQ	Imputed	Evaluates the present condition of the material on the exterior
ExteriorQual	Imputed	Evaluates the quality of the material on the exterior
FenceQ	Imputed	Fence quality
FireplaceQ	Imputed	Fireplace quality
GarageC	Imputed	Garage condition
GarageFin	Imputed	Interior finish of the garage
GarageQ	Imputed	Garage quality
HeatingQ	Imputed	Heating quality and condition
HmFunctional	Imputed	Home functionality (Assume typical unless deductions are warranted)
Hood	Imputed	Physical locations within Ames city limits
KitchQual	Imputed	Kitchen quality
PoolQ	Imputed	Pool quality



Variable Imputation Code

```
/*Impute Variables NA variables to 0 and LOG of SQFT and SaleProce*/
data DataTable;
   set DataTable;
   if MasVnrArea="NA" then MasVnrArea=0;
   if GarageYrBlt='NA' then GarageYrBlt=YearBuilt;
   if GarageCars='NA' then GarageCars=0;
   if GarageArea='NA' then GarageArea=0;
                                                                    if BsmtFinSF1='NA' then BsmtFinSF1=0;
                                                                    if BsmtFinSF2='NA' then BsmtFinSF2=0;
                                                                    if BsmtFullBath='NA' then BsmtFullBath=0;
                                                                    if BsmtHalfBath='NA' then BsmtHalfBath=0;
                                                                    if BsmtUnfSF='NA' then BsmtUnfSF=0;
                                                                    if TotalBsmtSF='NA' then TotalBsmtSF=0;
                                                                    if LotFrontage='NA' then LotFrontage=sqrt(LotArea);
   SalePriceLog = Log(SalePrice);
   GrLivAreaLog = Log(GrLivArea);
run;

/*Impute Neighborhood Variables*/
data DataTable;
   set DataTable;
                                                                                                Hood = 0;
                                                                    if Neighborhood in('MeadowV','IDOTRR','BrDale') then Hood=1;
      if Neighborhood in('BrkSide','Edwards','OldTown','Sawyer','Blueste','SWISU','NPkVill','NAmes','Mitchel') then Hood=2;
  if Neighborhood in('SawyerW','NWAmes','Gilbert','Blmngtn','CollgCr','Crawfor','ClearCr','Somerst','Veenker','Timber') then
Hood=3;
                                                                    if Neighborhood in('StoneBr','NridgHt','NoRidge') then Hood=4;
Run;

/*Impute ExterQual Variables*/
data DataTable;
   set DataTable;
                                                                                                ExteriorQual = 0;
                                                                    if ExterQual = 'Ex' then ExteriorQual =5;
                                                                    Else if ExterQual = 'Gd' then ExteriorQual =4;
                                                                    Else if ExterQual = 'TA' then ExteriorQual =3;
                                                                    Else if ExterQual = 'FA' then ExteriorQual =2;
                                                                    Else if ExterQual = 'Po' then ExteriorQual =1;
                                                                                                Else ExteriorQual =0;
Run;

/*Impute KitchenQual Variables*/
data DataTable;
   set DataTable;
                                                                                                KitchQual = 0;
                                                                    if KitchenQual = 'Ex' then KitchQual =5;
                                                                    Else if KitchenQual = 'Gd' then KitchQual =4;
                                                                    Else if KitchenQual = 'TA' then KitchQual =3;
                                                                    Else if KitchenQual = 'FA' then KitchQual =2;
                                                                    Else if KitchenQual = 'Po' then KitchQual =1;
                                                                                                Else KitchQual =0;
Run;

/*Impute BsmtQual Variables*/
data DataTable;
   set DataTable;
                                                                                                BasemtQual = 0;
                                                                    if BsmtQual = 'Ex' then BasemtQual =6;
                                                                    Else if BsmtQual = 'Gd' then BasemtQual =5;
                                                                    Else if BsmtQual = 'TA' then BasemtQual =4;
                                                                    Else if BsmtQual = 'FA' then BasemtQual =3;
                                                                    Else if BsmtQual = 'Po' then BasemtQual =2;
                                                                    Else if BsmtQual = 'NA' then BasemtQual =1;
                                                                                                Else BasemtQual =0;
Run;

/*Impute GarageFinish Variables*/
```

```
data DataTable;
   set DataTable;
                                                                GarageFin = 0;
                                                if GarageFinish = 'Fin' then GarageFin =4;
                                            Else if GarageFinish = 'RFn' then GarageFin =3;
                                             Else if GarageFinish = 'Unf' then GarageFin =2;
                                              Else if GarageFinish = 'NA' then GarageFin =1;
                                                                Else GarageFin =0;
Run;

/*Impute Fence Variables*/
data DataTable;
   set DataTable;
                                                                FenceQ = 0;
                                                if Fence = 'GdPrv' then FenceQ=5;
                                              Else if Fence= 'MnPrv' then FenceQ=4;
                                              Else if Fence= 'GdWo' then FenceQ=3;
                                              Else if Fence= 'MnWw' then FenceQ=2;
                                                Else if Fence= 'NA' then FenceQ=1;
                                                                Else FenceQ=0;
Run;

/*Impute PoolQC Variables*/
data DataTable;
   set DataTable;
                                                                PoolQ = 0;
                                                if PoolQC= 'Ex' then PoolQ=5;
                                              Else if PoolQC= 'Gd' then PoolQ=4;
                                              Else if PoolQC= 'TA' then PoolQ=3;
                                              Else if PoolQC= 'Fa' then PoolQ=2;
                                              Else if PoolQC= 'NA' then PoolQ=1;
                                                                Else PoolQ=0;
Run;

/*Impute GarageCond Variables*/
data DataTable;
   set DataTable;
                                                                GarageC = 0;
                                                if GarageCond= 'Ex' then GarageC=6;
                                              Else if GarageCond= 'Gd' then GarageC=5;
                                              Else if GarageCond= 'TA' then GarageC=4;
                                              Else if GarageCond= 'FA' then GarageC=3;
                                              Else if GarageCond= 'Po' then GarageC=2;
                                              Else if GarageCond= 'NA' then GarageC=1;
                                                                Else GarageC=0;
Run;

/*Impute GarageQual Variables*/
data DataTable;
   set DataTable;
                                                                GarageQ = 0;
                                                if GarageQual= 'Ex' then GarageQ=6;
                                              Else if GarageQual= 'Gd' then GarageQ=5;
                                              Else if GarageQual= 'TA' then GarageQ=4;
                                              Else if GarageQual= 'FA' then GarageQ=3;
                                              Else if GarageQual= 'Po' then GarageQ=2;
                                              Else if GarageQual= 'NA' then GarageQ=1;
                                                                Else GarageQ=0;
Run;

/*Impute FireplaceQu Variables*/
data DataTable;
   set DataTable;
                                                                FireplaceQ = 0;
                                                if FireplaceQu= 'Ex' then FireplaceQ=6;
                                              Else if FireplaceQu= 'Gd' then FireplaceQ=5;
                                              Else if FireplaceQu= 'TA' then FireplaceQ=4;
                                              Else if FireplaceQu= 'FA' then FireplaceQ=3;
                                              Else if FireplaceQu= 'Po' then FireplaceQ=2;
                                              Else if FireplaceQu= 'NA' then FireplaceQ=1;
```

```
                                                
```

```
/*Impute BsmtCond Variables*/
data DataTable;
   set DataTable;
                                                                                BsmtCondQ = 0;
                                                                    if BsmtCond= 'Ex' then BsmtCondQ=6;
                                                                  Else if BsmtCond= 'Gd' then BsmtCondQ=5;
                                                                  Else if BsmtCond= 'TA' then BsmtCondQ=4;
                                                                  Else if BsmtCond= 'FA' then BsmtCondQ=3;
                                                                  Else if BsmtCond= 'Po' then BsmtCondQ=2;
                                                                  Else if BsmtCond= 'NA' then BsmtCondQ=1;
                                                                                Else BsmtCond=0;
Run;

/*Impute ExterCond Variables*/
data DataTable;
   set DataTable;
                                                                                ExterCondQ = 0;
                                                                    if ExterCond= 'Ex' then ExterCondQ=6;
                                                                  Else if ExterCond= 'Gd' then ExterCondQ=5;
                                                                  Else if ExterCond= 'TA' then ExterCondQ=4;
                                                                  Else if ExterCond= 'FA' then ExterCondQ=3;
                                                                  Else if ExterCond= 'Po' then ExterCondQ=2;
                                                                  Else if ExterCond= 'NA' then ExterCondQ=1;
                                                                                Else ExterCond=0;
Run;


/*Change Character Columns to Numeric for fitting the model*/
data DataTable;
set DataTable;
BsmtFinSF1a = Input(BsmtFinSF1,BEST12.);
BsmtFinSF2a = Input(BsmtFinSF2,BEST12.);
BsmtFullBatha = Input(BsmtFullBath,BEST12.);
BsmtHalfBatha = Input(BsmtHalfBath,BEST12.);
BsmtUnfSFa = Input(BsmtUnfSF,BEST12.);
GarageAreaa = Input(GarageArea,BEST12.);
GarageCarsa = Input(GarageCars,BEST12.);
GarageYrBlta = Input(GarageYrBlt,BEST12.);
MasVnrAreaa = Input(MasVnrArea,BEST12.);
TotalBsmtSFa = Input(TotalBsmtSF,BEST12.);
LotFrontageA = Input(LotFrontage,BEST12.);
  drop LotFrontage BsmtFinSF1                                                   BsmtFinSF2
                                                                                BsmtFullBath
                                                                                BsmtHalfBath
                                                                                 BsmtUnfSF
                                                                                 GarageArea
                                                                                 GarageCars
                                                                                 GarageYrBlt
                                                                                 MasVnrArea
                                                                                TotalBsmtSF;
  rename LotFrontageA=LotFrontage BsmtFinSF1a=BsmtFinSF1                  BsmtFinSF2a=BsmtFinSF2
                                                                        BsmtFullBatha=BsmtFullBath
                                                                       BsmtHalfBatha=BsmtHalfBath
                                                                          BsmtUnfSFa=BsmtUnfSF
                                                                         GarageAreaa=GarageArea
                                                                         GarageCarsa=GarageCars
                                                                         GarageYrBlta=GarageYrBlt
                                                                         MasVnrAreaa=MasVnrArea
                                                                        TotalBsmtSFa=TotalBsmtSF;
run;
```

Forward Regression VIF

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	2.51997	0.50158	5.02	<.0001	0
Hood	1	0.06439	0.00674	9.56	<.0001	2.52861
ExteriorQual	1	0.02383	0.00727	3.28	0.0011	2.28419
MSSubClass	1	-0.00041460	0.00008621	-4.81	<.0001	1.37508
HmFunctional	1	0.03132	0.00494	6.35	<.0001	1.12298
HeatingQ	1	0.01219	0.00329	3.71	0.0002	1.49291
BsmtExposur	1	0.01753	0.00345	5.09	<.0001	1.39006
Fireplaces	1	0.03484	0.00592	5.89	<.0001	1.49133
ScreenPorch	1	0.00024030	0.00005737	4.19	<.0001	1.05782
KitchenAbvGr	1	-0.05608	0.01622	-3.46	0.0006	1.32030
GarageArea	1	0.00017708	0.00001948	9.09	<.0001	1.76417
BsmtFinSF1	1	0.00009056	0.00000862	10.51	<.0001	1.43751
OverallQual	1	0.05345	0.00432	12.38	<.0001	3.64651
OverallCond	1	0.04577	0.00348	13.15	<.0001	1.55078
TotalBsmtSF	1	0.00009240	0.00001121	8.24	<.0001	2.23648
GrLivArea	1	0.00013021	0.00002690	4.84	<.0001	19.27951
YearBuilt	1	0.00228	0.00018446	12.34	<.0001	3.20316
YearRemodAdd	1	0.00081469	0.00022898	3.56	0.0004	2.30671
LotArea	1	0.00000203	3.475473E-7	5.85	<.0001	1.21236
GrLivAreaLog	1	0.25504	0.04241	6.01	<.0001	20.27011

Forward Regression Parameters

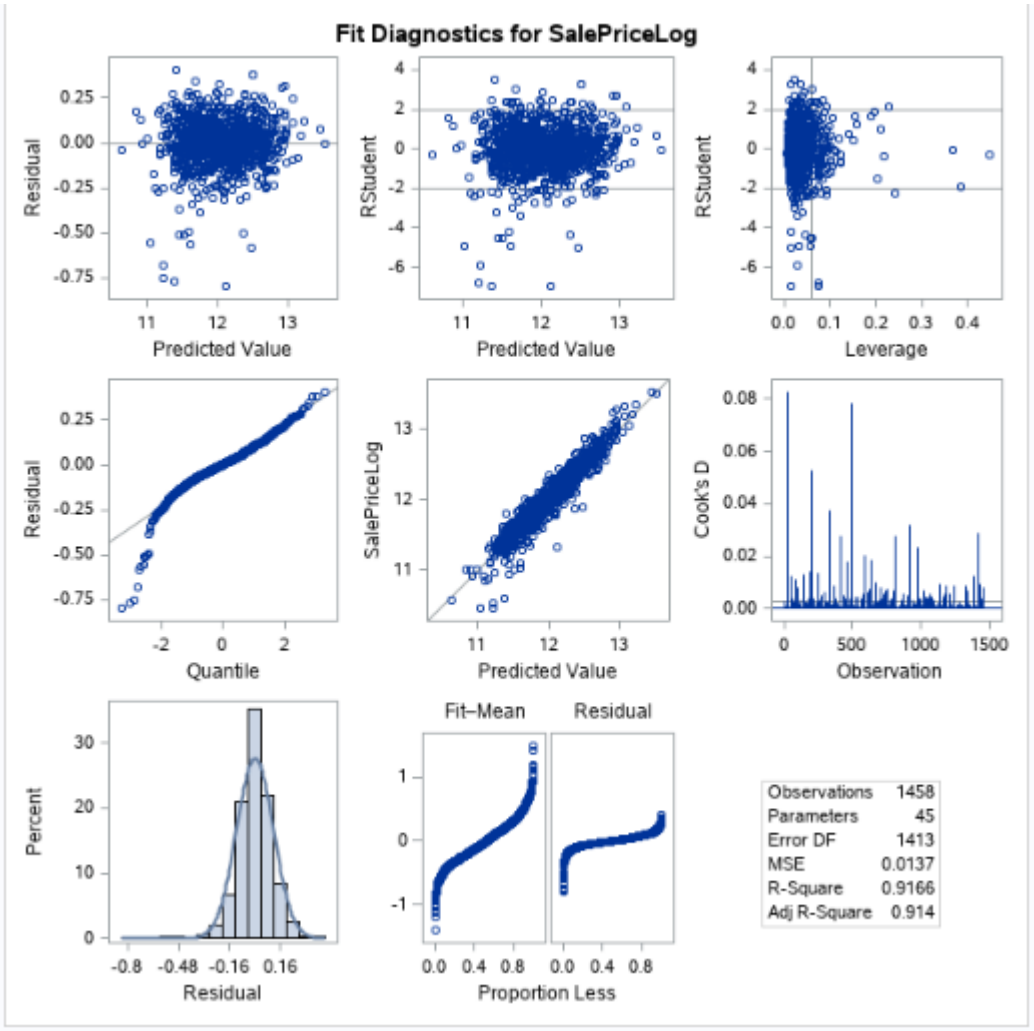
Root MSE	0.11772
Dependent Mean	12.02401
R-Square	0.9146
Adj R-Sq	0.9133
AIC	-4754.77558

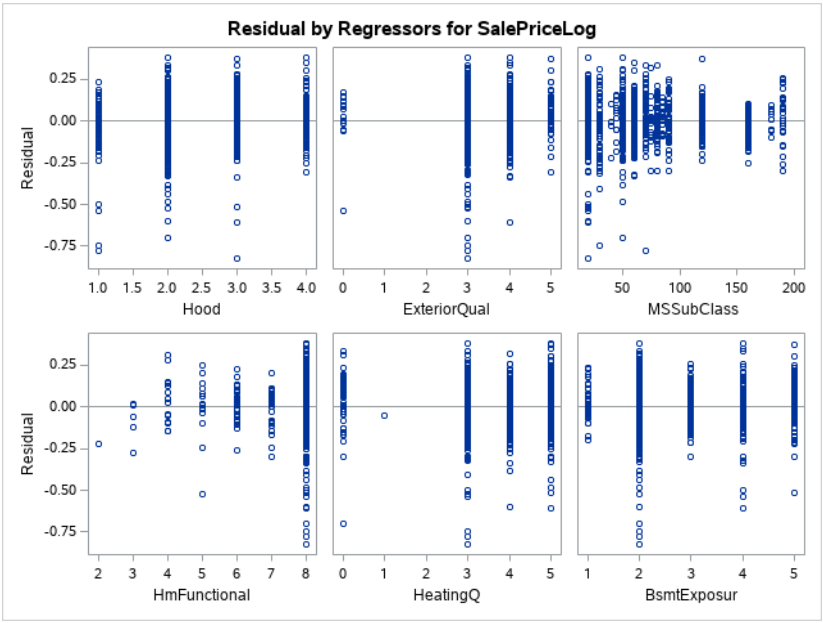
AICC	-4753.86776
SBC	-6087.93988
CV PRESS	20.70752

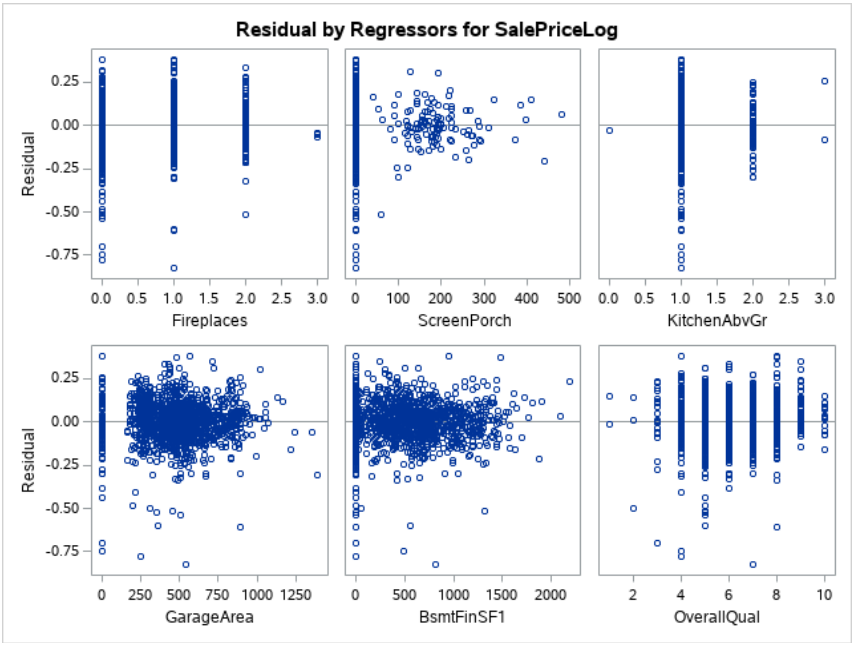
Summary of Forward Selection

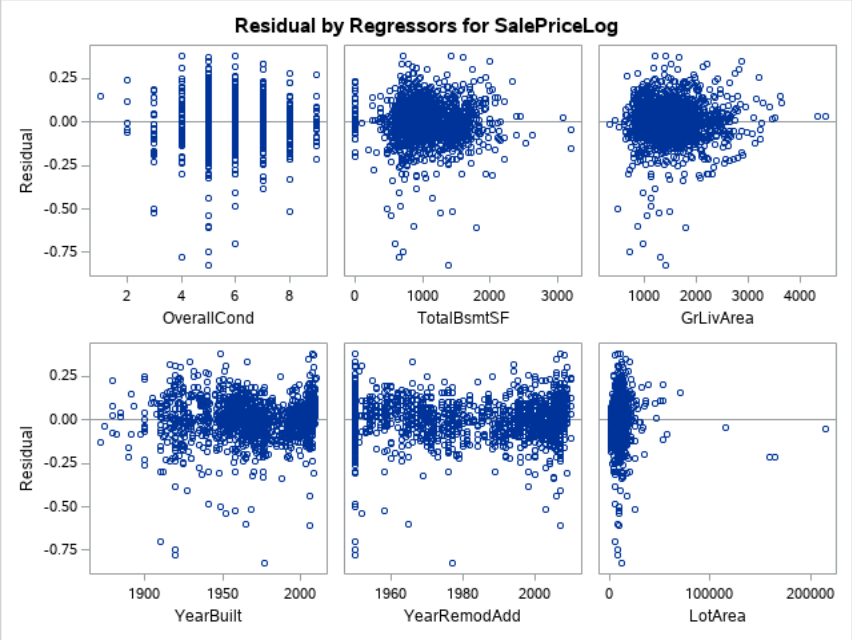
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	OverallQual	1	0.6747	0.6747	4036.08	3019.94	<.0001
2	GrLivAreaLog	2	0.0895	0.7642	2527.88	552.11	<.0001
3	BsmtFinSF1	3	0.0471	0.8113	1734.81	362.98	<.0001
4	Hood	4	0.0319	0.8432	1198.18	295.76	<.0001
5	GarageArea	5	0.0133	0.8565	975.609	134.65	<.0001
6	OverallCond	6	0.0107	0.8672	796.573	117.24	<.0001
7	YearBuilt	7	0.0163	0.8835	523.861	202.63	<.0001
8	TotalBsmtSF	8	0.0105	0.8940	348.710	143.51	<.0001
9	LotArea	9	0.0041	0.8981	281.279	58.48	<.0001
10	Fireplaces	10	0.0027	0.9008	237.448	39.63	<.0001
11	KitchenAbvGr	11	0.0019	0.9028	206.801	28.77	<.0001
12	HmFunctional	12	0.0019	0.9047	176.467	29.05	<.0001
13	HeatingQ	13	0.0016	0.9063	150.804	25.27	<.0001
14	BsmtExposur	14	0.0013	0.9077	130.161	20.97	<.0001
15	MSSubClass	15	0.0012	0.9089	111.883	19.01	<.0001
16	ScreenPorch	16	0.0010	0.9098	97.7675	15.26	<.0001
17	YearRemodAdd	17	0.0010	0.9108	83.7263	15.34	<.0001
18	ExteriorQual	18	0.0006	0.9114	75.0483	10.28	0.0014
19	BsmtFullBath	19	0.0005	0.9119	68.1631	8.60	0.0034
20	GarageC	20	0.0004	0.9124	62.5982	7.35	0.0068
21	TotRmsAbvGrd	21	0.0004	0.9128	57.9498	6.49	0.0110
22	WoodDeckSF	22	0.0003	0.9131	54.6960	5.14	0.0235
23	PoolQ	23	0.0003	0.9134	52.1099	4.50	0.0341
24	BedroomAbvGr	24	0.0003	0.9136	49.5379	4.50	0.0342
25	GarageCars	25	0.0002	0.9139	47.3203	4.16	0.0417
26	YrSold	26	0.0002	0.9141	45.5829	3.69	0.0550
27	HalfBath	27	0.0002	0.9143	45.0010	2.55	0.1104
28	_1stFlrSF	28	0.0002	0.9144	44.1368	2.83	0.0925
29	_2ndFlrSF	29	0.0010	0.9154	29.3789	16.77	<.0001
30	KitchQual	30	0.0001	0.9156	28.8641	2.52	0.1127
31	FireplaceQ	31	0.0001	0.9157	28.6422	2.23	0.1358
32	BsmtFinSF2	32	0.0001	0.9158	28.6083	2.04	0.1534
33	OpenPorchSF	33	0.0001	0.9159	28.8895	1.72	0.1893
34	FenceQ	34	0.0001	0.9160	29.3545	1.54	0.2147
35	GarageQ	35	0.0001	0.9161	29.8400	1.52	0.2177
36	_3SsnPorch	36	0.0001	0.9162	30.6390	1.21	0.2722
37	EnclosedPorch	37	0.0001	0.9163	31.3989	1.25	0.2645
38	ExterCondQ	38	0.0001	0.9163	32.2266	1.18	0.2780
39	FullBath	39	0.0001	0.9164	33.1525	1.08	0.2990
40	BsmtFinType2q	40	0.0001	0.9164	34.1697	0.99	0.3205
41	BsmtCondQ	41	0.0001	0.9165	35.0495	1.13	0.2889
42	LowQualFinSF	42	0.0001	0.9166	36.1686	0.89	0.3469
43	LotFrontage	43	0.0000	0.9166	37.4979	0.67	0.4119
44	GarageFin	44	0.0000	0.9166	39.0316	0.47	0.4939

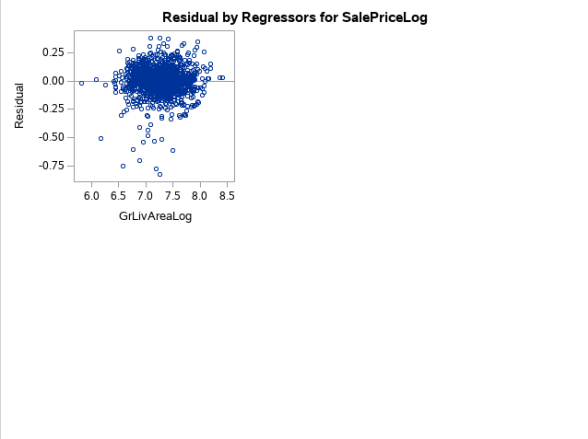
Forward Regression Residual Plots




Residual by Regressors for SalePriceLog
Residual
Hood
ExteriorQual
MSSubClass
HmFunctional
HeatingQ
BsmtExposur


Residual by Regressors for SalePriceLog
Residual
Fireplaces
ScreenPorch
KitchenAbvGr
GarageArea
BsmtFinSF1
OverallQual





Forward Regression Code Chunks

```
*/////////////////////////////////////////////Forward Selection Model  *//////////////////////////////////////////////;
proc GLMSELECT  data=train plots=all;
model SalePriceLog = Hood ExteriorQual KitchQual BasemtQual GarageFin LotFrontage
```

```
MSSubClass BsmtFinSF2 LowQualFinSF
FenceQ PoolQ GarageC GarageQ FireplaceQ HmFunctional HeatingQ BsmtFinType1q
BsmtFinType2q BsmtExposur BsmtCondQ ExterCondQ
BsmtHalfBath Fireplaces WoodDeckSF EnclosedPorch ScreenPorch PoolArea BsmtFullBath
HalfBath KitchenAbvGr FullBath GarageCars MoSold
BedroomAbvGr  TotRmsAbvGrd OpenPorchSF BsmtUnfSF GarageArea BsmtFinSF1
OverallQual OverallCond
TotalBsmtSF GrLivArea YearBuilt YearRemodAdd YrSold LotArea _3SsnPorch _2ndFlrSF
_1stFlrSF GarageYrBlt MasVnrArea GrLivAreaLog
/selection=forward(stop=cv) CVMETHOD=RANDOM(5) stats=ADJRSQ;
store out=TrainModelFW;
run;
/*Get VIF*/
proc reg data=train;
                model SalePriceLog =  Hood ExteriorQual MSSubClass HmFunctional HeatingQ
                  BsmtExposur Fireplaces ScreenPorch KitchenAbvGr GarageArea BsmtFinSF1
                                                                    OverallQual OverallCond
                   TotalBsmtSF GrLivArea YearBuilt YearRemodAdd LotArea GrLivAreaLog/ vif;
run;
```

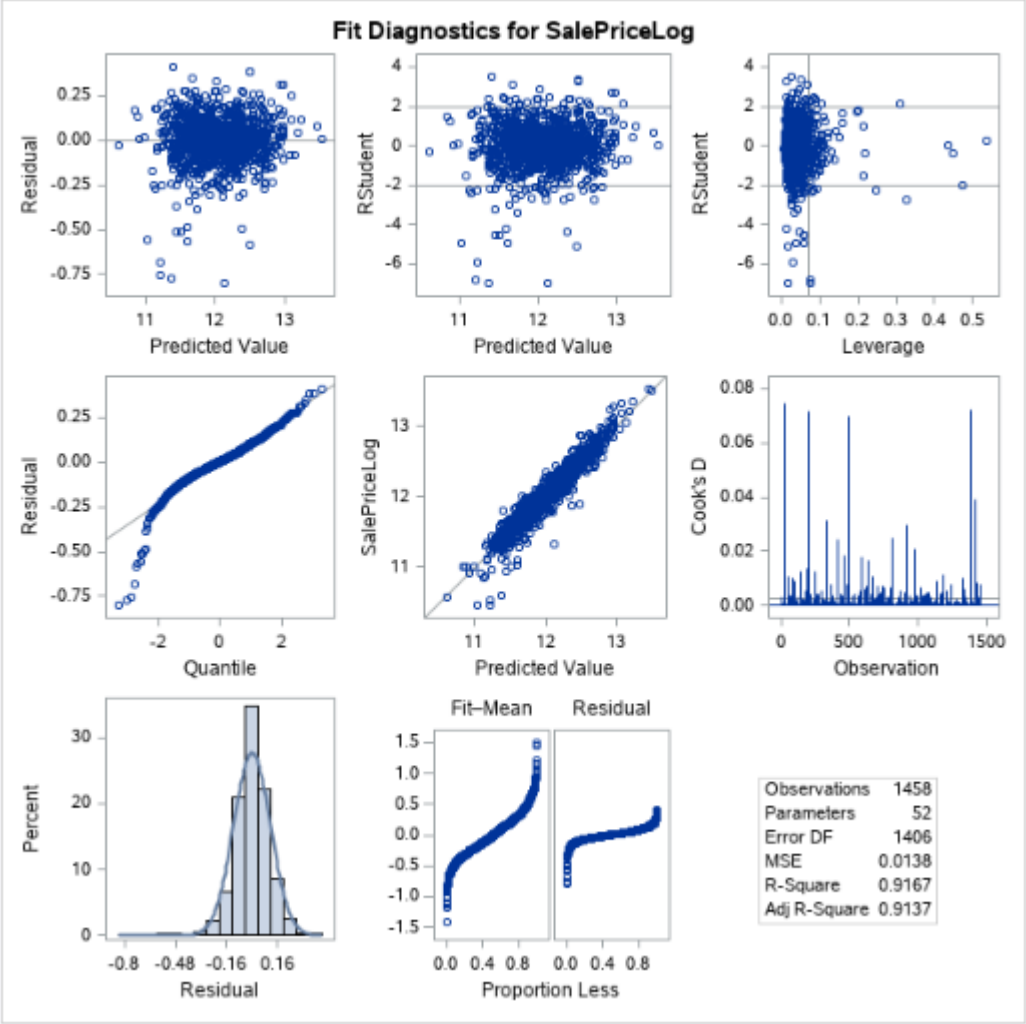
Forward Regression Prediction Code

```
/*Score Data, create predicted values for Forward Model*/
proc plm restore=TrainModelFW;
score data=Test out=TestModelFW
 pred=Predicted lcl=Lower ucl=Upper;
 run;

data ExportKaggleFW;
                                                    set TestModelFW (keep=ID Predicted);
                                                          psale = exp(Predicted);
                                                               SalePrice = psale;
                                                            drop psale Predicted;
run;
```

Backwards Regression Parameter Estimates

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	11.971679	4.742682	2.52
Hood	1	0.064281	0.006848	9.39
ExteriorQual	1	0.017588	0.007567	2.32
KitchQual	1	0.008368	0.005346	1.57
MSSubClass	1	-0.000430	0.000091627	-4.69
BsmtFinSF2	1	0.000141	0.000036493	3.87
FenceQ	1	-0.003396	0.002773	-1.22
PoolQ	1	0.025206	0.017147	1.47
GarageC	1	0.008451	0.005371	1.57
GarageQ	1	0.006212	0.005055	1.23
FireplaceQ	1	0.004918	0.003101	1.59
HmFunctional	1	0.031455	0.005022	6.26
HeatingQ	1	0.011939	0.003304	3.61
BsmtFinType2q	1	-0.006511	0.005898	-1.10
BsmtExposur	1	0.015693	0.003512	4.47
BsmtCondQ	1	0.004497	0.004238	1.06
ExterCondQ	1	-0.006423	0.005309	-1.21
Fireplaces	1	0.017610	0.008905	1.98
WoodDeckSF	1	0.000059933	0.000027269	2.20
EnclosedPorch	1	0.000068709	0.000057476	1.20
ScreenPorch	1	0.000248	0.000058196	4.26
BsmtFullBath	1	0.024618	0.008578	2.87
HalfBath	1	0.016532	0.009228	1.79
KitchenAbvGr	1	-0.058415	0.017737	-3.29
FullBath	1	0.009702	0.009670	1.00
GarageCars	1	0.019493	0.009869	1.98
BedroomAbvGr	1	-0.008910	0.005869	-1.52
TotRmsAbvGrd	1	0.008697	0.004179	2.08
OpenPorchSF	1	0.000072385	0.000051495	1.41
BsmtUnfSF	1	0.000083957	0.000015568	5.39
GarageArea	1	0.000083469	0.000033530	2.49
BsmtFinSF1	1	0.000159	0.000017189	9.23
OverallQual	1	0.051579	0.004387	11.76
OverallCond	1	0.045264	0.003817	11.86
YearBuilt	1	0.001943	0.000223	8.71
YearRemodAdd	1	0.000750	0.000239	3.14
YrSold	1	-0.004269	0.002355	-1.81
LotArea	1	0.000001889	0.000000347	5.44
_3SsnPorch	1	0.000122	0.000106	1.15
_2ndFlrSF	1	0.000113	0.000027819	4.07
_1stFlrSF	1	0.000137	0.000031700	4.33
GrLivAreaLog	1	0.229285	0.042690	5.37

Backward Regression Plots



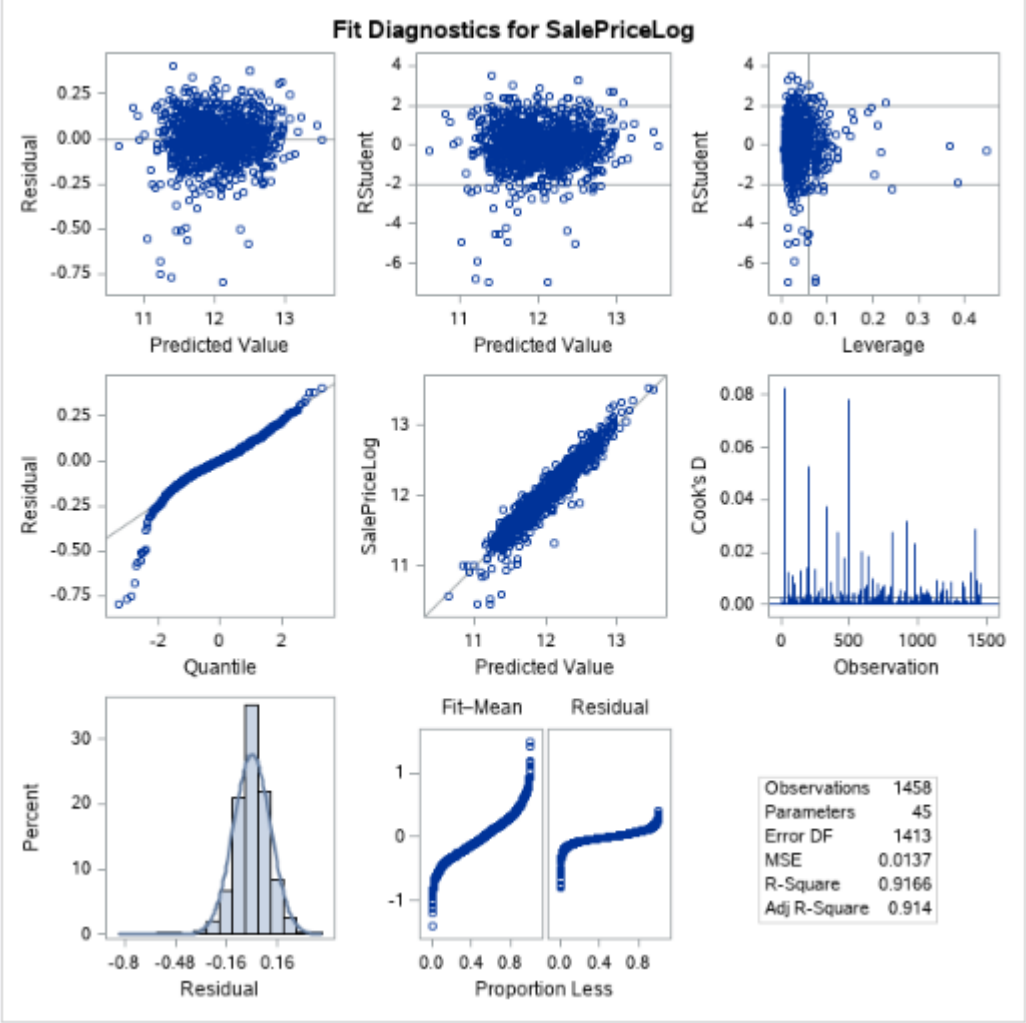
Backward Regression Code Chunks

```
proc GLMSELECT  data=train plots=all;
model SalePriceLog = Hood ExteriorQual KitchQual BasemtQual GarageFin LotFrontage
MSSubClass BsmtFinSF2 LowQualFinSF FenceQ PoolQ GarageC GarageQ FireplaceQ
HmFunctional HeatingQ BsmtFinType1q BsmtFinType2q BsmtExposur BsmtCondQ
ExterCondQ BsmtHalfBath Fireplaces WoodDeckSF EnclosedPorch ScreenPorch PoolArea
BsmtFullBath HalfBath KitchenAbvGr FullBath GarageCars MoSold BedroomAbvGr
TotRmsAbvGrd OpenPorchSF BsmtUnfSF GarageArea BsmtFinSF1 OverallQual
OverallCond TotalBsmtSF GrLivArea YearBuilt YearRemodAdd YrSold LotArea _3SsnPorch
_2ndFlrSF _1stFlrSF GarageYrBlt MasVnrArea GrLivAreaLog
/selection=backward(stop=cv) CVMETHOD=RANDOM(5) stats=ADJRSQ;
store out=TrainModelBW;
run;

/*Score Data, create predicted values for Backward Model*/
proc plm restore=TrainModelBW;
score data=Test out=TestModelBW
 pred=Predicted lcl=Lower ucl=Upper;
 run;

data ExportKaggleBW;
                                                    set TestModelBW (keep=ID Predicted);
                                                          psale = exp(Predicted);
                                                               SalePrice = psale;
                                                          drop psale Predicted;
run;
```

Stepwise Regression Parameter Estimates

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	1.593377	0.467198	3.41
Hood	1	0.067317	0.006762	9.96
ExteriorQual	1	0.023492	0.007328	3.21
MSSubClass	1	-0.000408	0.000086868	-4.69
HmFunctional	1	0.030884	0.004973	6.21
HeatingQ	1	0.012050	0.003311	3.64
BsmtExposur	1	0.018480	0.003468	5.33
Fireplaces	1	0.034557	0.005965	5.79
ScreenPorch	1	0.000236	0.000057807	4.09
KitchenAbvGr	1	-0.058542	0.016336	-3.58
GarageArea	1	0.000181	0.000019609	9.23
BsmtFinSF1	1	0.000093245	0.000008667	10.76
OverallQual	1	0.053883	0.004350	12.39
OverallCond	1	0.046172	0.003507	13.17
TotalBsmtSF	1	0.000091735	0.000011301	8.12
YearBuilt	1	0.002143	0.000184	11.66
YearRemodAdd	1	0.000802	0.000231	3.48
LotArea	1	0.000002044	0.000000350	5.84
GrLivAreaLog	1	0.447745	0.014705	30.45

Stepwise Regression Plots



Stepwise Regression Code Chunks

```
/*Stepwise model */
proc GLMSELECT  data=train plots=all;
model SalePriceLog = Hood ExteriorQual KitchQual BasemtQual GarageFin
LotFrontage MSSubClass BsmtFinSF2 LowQualFinSF FenceQ PoolQ GarageC
GarageQ FireplaceQ HmFunctional HeatingQ BsmtFinType1q BsmtFinType2q
BsmtExposur BsmtCondQ ExterCondQ BsmtHalfBath Fireplaces WoodDeckSF
EnclosedPorch ScreenPorch PoolArea BsmtFullBath HalfBath KitchenAbvGr FullBath
```

```
GarageCars MoSold BedroomAbvGr  TotRmsAbvGrd OpenPorchSF BsmtUnfSF
GarageArea BsmtFinSF1 OverallQual OverallCond TotalBsmtSF GrLivArea YearBuilt
YearRemodAdd YrSold LotArea _3SsnPorch _2ndFlrSF _1stFlrSF GarageYrBlt
MasVnrArea GrLivAreaLog
/selection=stepwise(stop=cv) CVMETHOD=RANDOM(5) stats=ADJRSQ;
store out=TrainModelSW;
run;

/*Score Data, create predicted values for Stepwise Model*/
proc plm restore=TrainModelSW;
score data=Test out=TestModelSW
 pred=Predicted lcl=Lower ucl=Upper;
 run;

data ExportKaggleSW;
                                                    set TestModelSW (keep=ID Predicted);
                                                          psale = exp(Predicted);
                                                               SalePrice = psale;
                                                          drop psale Predicted;
run;
```

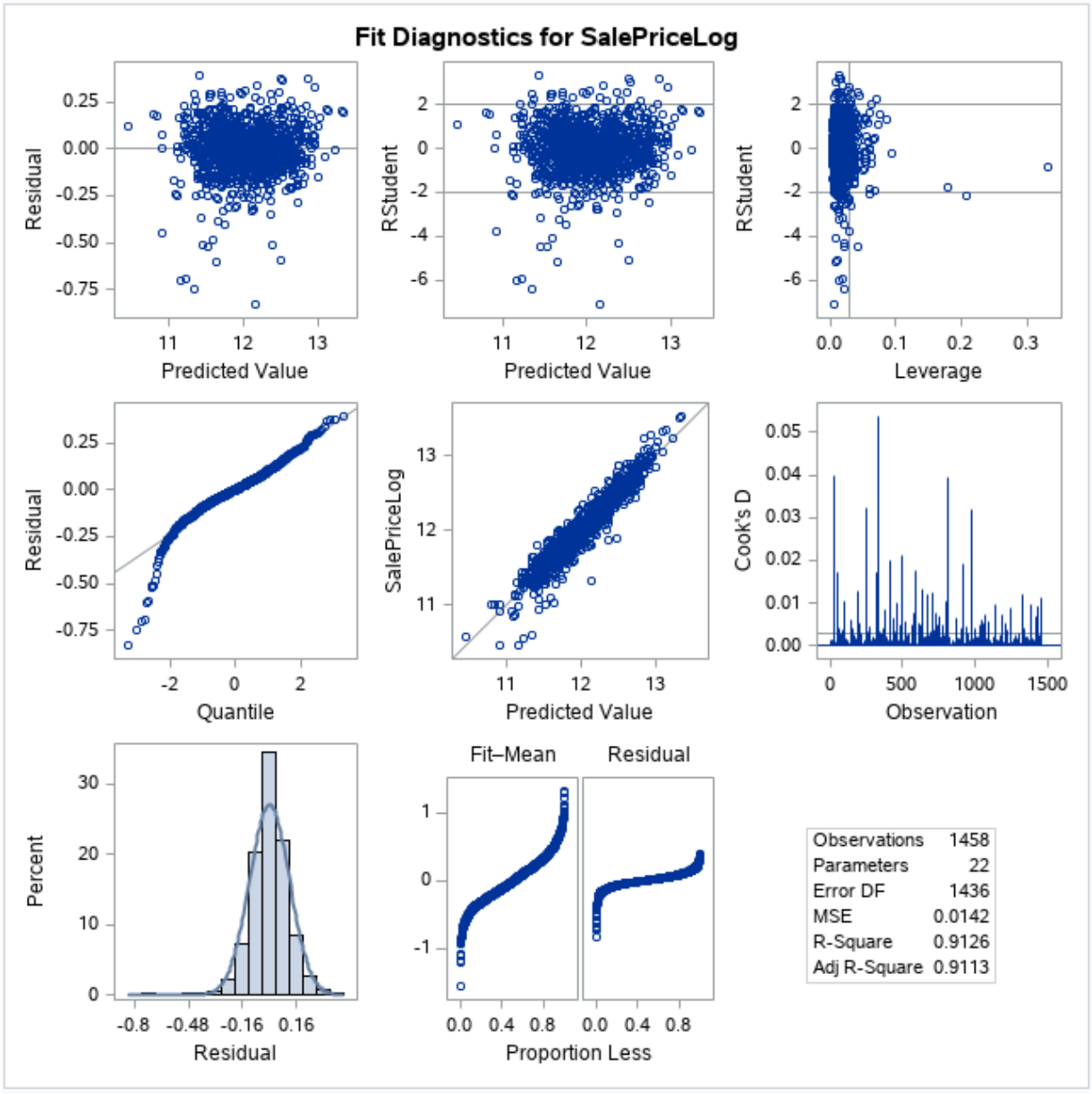
Custom Model Parameter Estimates

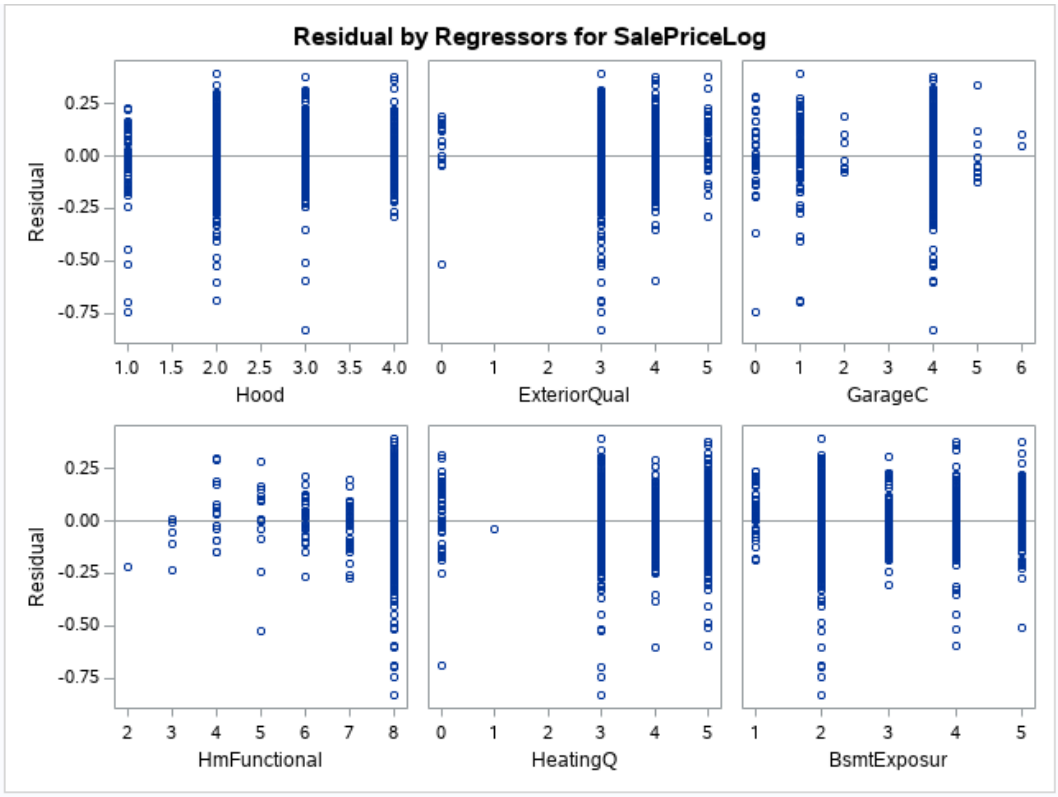
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.84843	0.47297	3.91	<.0001
Hood	1	0.06954	0.00675	10.3	<.0001
ExteriorQual	1	0.02476	0.0073	3.39	0.0007
GarageC	1	0.01104	0.00412	2.68	0.0074
HmFunctional	1	0.03151	0.00498	6.32	<.0001
HeatingQ	1	0.01214	0.0033	3.68	0.0002
BsmtExposur	1	0.01838	0.00348	5.28	<.0001
Fireplaces	1	0.03337	0.00596	5.6	<.0001
ScreenPorch	1	0.000224	5.77E-05	3.87	0.0001
TotRmsAbvGrd	1	0.00959	0.00369	2.6	0.0094
GarageArea	1	0.000161	2.11E-05	7.61	<.0001
OverallQual	1	0.05425	0.00434	12.51	<.0001
OverallCond	1	0.0436	0.00355	12.27	<.0001
YearBuilt	1	0.002	0.000189	10.61	<.0001
YearRemodAdd	1	0.000912	0.000232	3.94	<.0001
LotArea	1	1.97E-06	3.49E-07	5.65	<.0001
MSSubClass	1	-0.00037	8.76E-05	-4.25	<.0001
BsmtUnfSF	1	9.06E-05	1.15E-05	7.9	<.0001
BsmtFinSF2	1	0.000139	2.15E-05	6.47	<.0001
BsmtFinSF1	1	0.000189	1.18E-05	16	<.0001
KitchenAbvGr	1	-0.06477	0.01694	-3.82	0.0001
GrLivAreaLog	1	0.40728	0.0216	18.86	<.0001

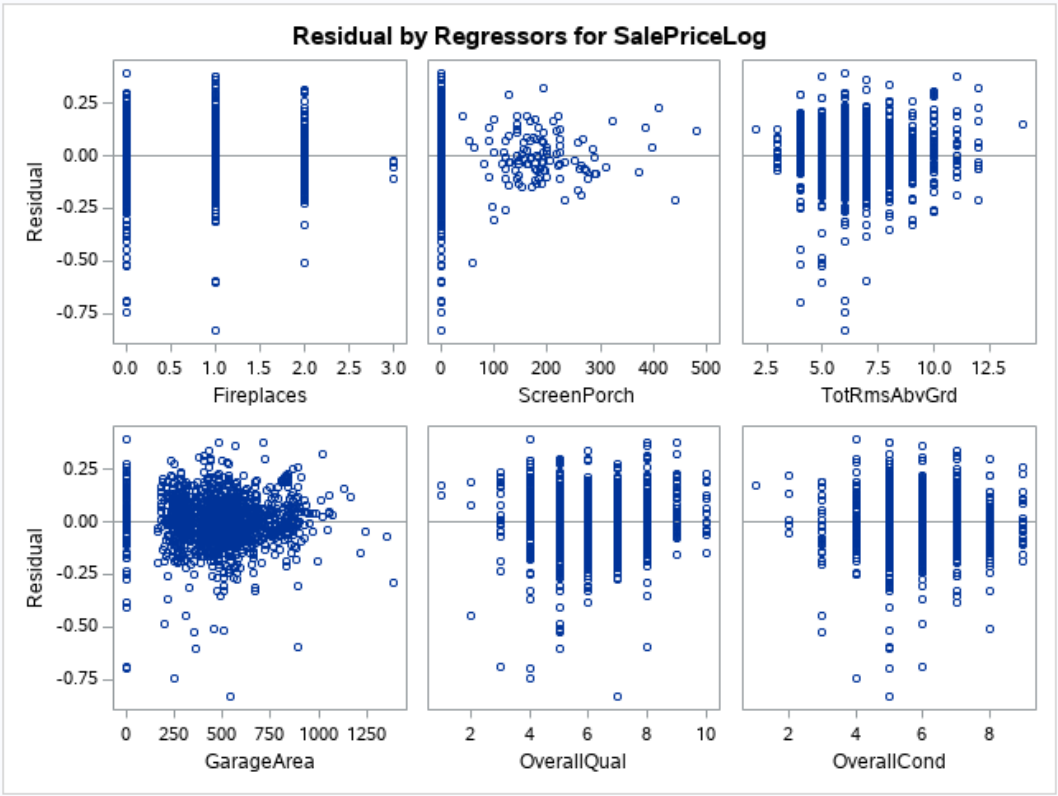
Custom Model Regression Plots

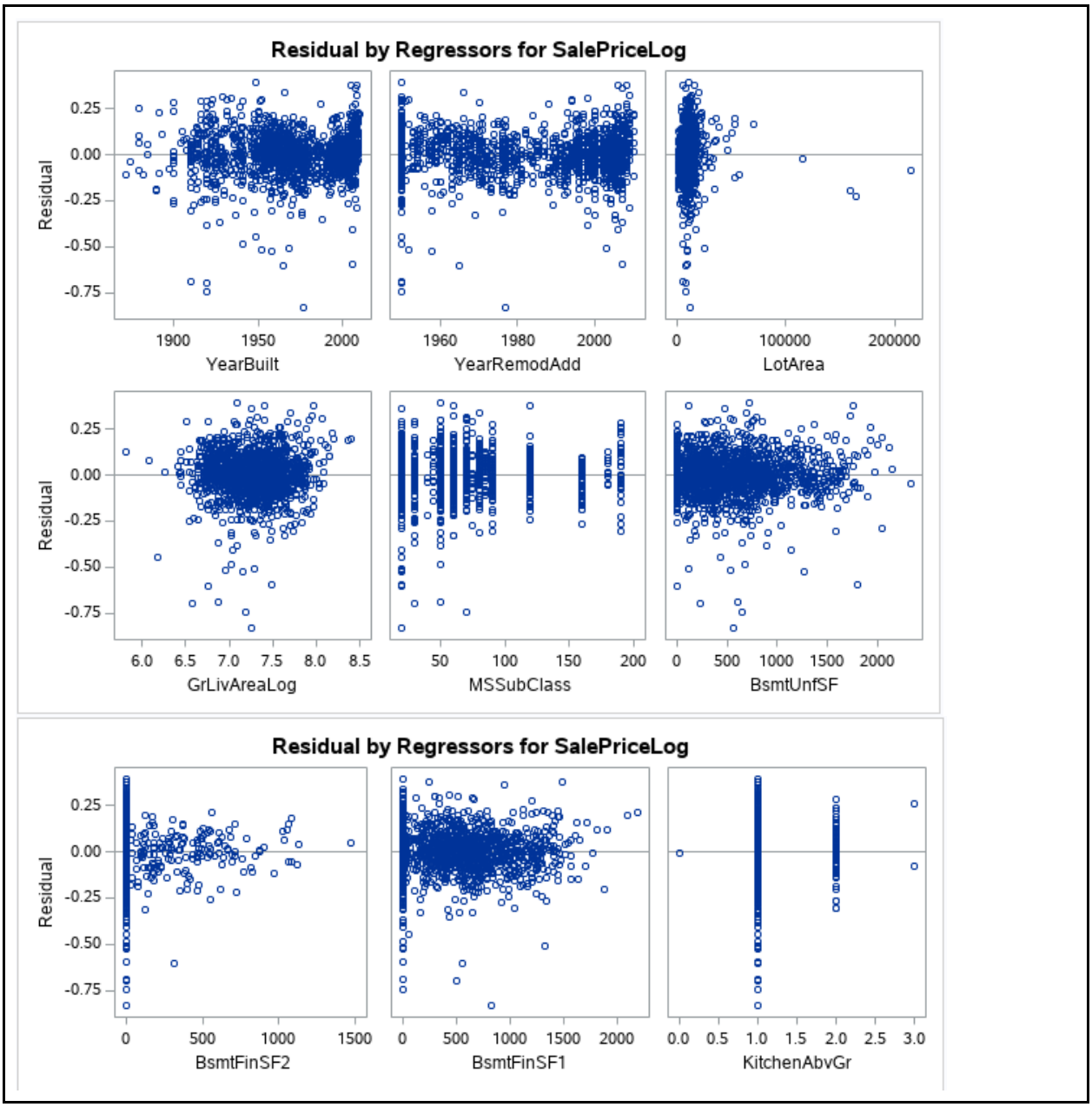
```
*////////////////////////////////////////////////////Custom Model//////////////////////////////////////////////////////;
proc reg data=train;
                                                    model SalePriceLog = Hood ExteriorQual GarageC
HmFunctional   HeatingQ BsmtExposur Fireplaces ScreenPorch TotRmsAbvGrd  GarageArea
OverallQual OverallCond YearBuilt YearRemodAdd LotArea rLivAreaLog MSSubClass
BsmtUnfSF BsmtFinSF2 BsmtFinSF1 KitchenAbvGr / clm cli press cp;
                                                output out=preddata predicted=pred lcl=lower ucl=upper;
                                                                                    store ExportKaggleCustom;
run;

/*Score Data, create predicted values for Custom Model*/
proc plm restore=ExportKaggleCustom;
score data=Test out=TestModelCustom
 pred=Predicted lcl=Lower ucl=Upper;
 run;
```

```
data ExportKaggleCustom;
                                        set TestModelCustom (keep=ID Predicted);
                                            psale = exp(Predicted);
                                            SalePrice = psale;
                                            drop psale Predicted;
run;
```




Residual by Regressors for SalePriceLog
Residual
Hood
ExteriorQual
GarageC
HmFunctional
HeatingQ
BsmtExposur


Residual by Regressors for SalePriceLog
Residual
Fireplaces
ScreenPorch
TotRmsAbvGrd
GarageArea
OverallQual
OverallCond


Residual by Regressors for SalePriceLog
Residual
0.25
0.00
-0.25
-0.50
-0.75
1900
1950
2000
YearBuilt
1960
1980
2000
YearRemodAdd
0
100000
200000
LotArea
6.0
6.5
7.0
7.5
8.0
8.5
GrLivAreaLog
50
100
150
200
MSSubClass
0
500
1000
1500
2000
BsmtUnfSF
Residual by Regressors for SalePriceLog
BsmtFinSF2
BsmtFinSF1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
KitchenAbvGr