Table of Contents

Appendix – SAS and R Code for Project	
Analysis 1	
Analysis 2	
Appendix 1: Vanilla Regression – Raw Data Model Original	
Appendix 2: Vanilla Regression Model: Log-Linear	8
Appendix 3: Vanilla Regression Model: Linear - Log	9
Appendix 4: Vanilla Regression Model: Log-Log	10
Appendix 5: Vanilla Regression Model: Outliers Addressed*	11
Appendix 6: Vanilla Regression Model: Log-Linear (Outliers Addressed)	12
Appendix 7: Vanilla Regression Model: Linear-Log (Outliers Addressed)	13
Appendix 8: Vanilla Regression Model: Log-Log (Outliers Addressed)	14
Appendix 9: Analysis 2 Checking Assumptions	15
Appendix 10 : Forward Selection SAS Results	16
	16
Appendix 11 : Analysis 2 Backwards Elimination	17
	17
	17
	17
Appendix 12: Analysis 2 Stepwise Selection	
Appendix 13 : Analysis 2 Custom Selection	
Appendix 13 : Analysis 2 Custom Selection	
APPENDIX 14 · Kaggle Suhmission Score	20

Appendix – SAS and R Code for Project

Analysis 1

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/test.csv';
PROC IMPORT DATAFILE=REFFILE
        DBMS=CSV
        OUT=test;
        GETNAMES=YES;
RUN;
FILENAME REFFILE '/folders/myfolders/sasuser.v94/train.csv';
PROC IMPORT DATAFILE=REFFILE
        DBMS=CSV
        OUT=train;
        GETNAMES=YES;
RUN;
/*Filter only relevant Neighborhoods for Q1. NAmes, Edwards, BrkSide*/
data neighborhoodSS;
        set train;
        where Neighborhood = 'NAmes' or Neighborhood = 'Edwards' or Neighborhood='BrkSide';
run;
/* Create GrLivAreaper100 variable, log Transform the SalesPrice and GrLivAreaper100 to re-run model */
data logneighborhoodSS;
        set neighborhoodSS;
        GrLivAreaper100 = GrLivArea/100;
        logSalePrice = log(SalePrice);
        logGrLivAreaper100 = log(GrLivArea/100);
run;
/*proc print data = logneighborhoodSS;
run;*/
/*Scatterplot the Saleprice to GrLivArea for the three requested neighborhoods*/
proc sgplot data=logneighborhoodSS;
scatter x=GrLivAreaper100 y=SalePrice / group=Neighborhood;
run;
/* Model for Salesprice to GrLivArea*/
proc glm data = logneighborhoodSS plots=all;
class Neighborhood;
model SalePrice = GrLivAreaper100 | Neighborhood / cli clm clparm;
run;
/*Filter out all homes greater than 40 100 sq.ft.*/
data logneighborhoodSSsmall;
        set logneighborhoodSS;
        where GrLivAreaper100 <= 40;
run;
```

```
/* Model for Salesprice to GrLivArea*/
proc glm data = logneighborhoodSSsmall plots=all;
class Neighborhood;
model SalePrice = GrLivAreaper100 | Neighborhood / cli clm clparm;
/*Filter out all homes that costs more than $300,000*/
data logneighborhoodSScheap;
        set logneighborhoodSS;
        where SalePrice <= 300000;
run;
/* Model for Salesprice to GrLivArea*/
proc glm data = logneighborhoodSScheap plots=all;
class Neighborhood;
model SalePrice = GrLivAreaper100 | Neighborhood / cli clm clparm;
run;
/*Filter out all homes that costs more than $300,000 and smaller than 40 100 sq. ft.*/
data logneighborhoodSScheapsmall;
        set logneighborhoodSS;
        where SalePrice <= 300000 and GrLivAreaper100 <= 40;
run;
/* Model for Salesprice to GrLivArea*/
proc glm data = logneighborhoodSScheapsmall plots=all;
class Neighborhood;
model SalePrice = GrLivAreaper100 | Neighborhood / cli clm clparm solution;
/* Model with log transformed SalePrice*/
proc glm data = logneighborhoodSS plots=all;
class Neighborhood;
model logSalePrice = GrLivAreaper100 | Neighborhood / cli clm clparm;
run;
/* Model with log transformed GrLivAreaper100*/
proc glm data = logneighborhoodSS plots=all;
class Neighborhood;
model SalePrice = logGrLivAreaper100 | Neighborhood / cli clm clparm;
/* Model with log-log transformed SalePrice-GrLivAreaper100*/
proc glm data = logneighborhoodSS plots=all;
class Neighborhood;
model logSalePrice = logGrLivAreaper100 | Neighborhood / cli clm clparm;
run;
/* Model with log-log transformed SalePrice-GrLivAreaper100 on homes that are less than $300,000 and 40 100 sq. ft.*/
proc glm data = logneighborhoodSScheapsmall plots=all;
class Neighborhood;
model logSalePrice = logGrLivAreaper100 | Neighborhood / cli clm clparm; run;
```

Analysis 2

/* Export Kaggle Submission csv file*/ proc export data=forward_results2

```
#masterAN.csv is combination of train.csv and test.csv
         trainDF <- read.csv("masterAN.csv", header = T, stringsAsFactors = F)
         trainDF <- subset(trainDF[1:1470,])
         trainDF$SalePrice <- as.integer(trainDF$SalePrice)
         titles <- names(trainDF)
         for (i in 1:length(trainDF)){
          if (is.character(trainDF[[i]])){
           boxplot(trainDF$SalePrice ~ trainDF[[i]], main = paste("Sale price vs ", titles[i], sep=""), ylab = titles[i], xlab = "Sale price", horizontal = T,
           else if (is.integer(trainDF[[i]])){
           plot(x = trainDF[[i]], y = trainDF$SalePrice, main = paste("Sale price vs ", titles[i], sep=""), xlab = titles[i], ylab = "Sale price")
         }
/*SAS code for model selection (Forward, Backward, Stepwise, and custom user defined);*/
         data train;
         infile '/folders/myfolders/sasuser.v94/masterAN.csv' dlm=',' firstobs=2;
         input Id MSSubClass MSZoning $ LotFrontage LotArea Street $ Alley $ LotShape $ LandContour $
                   Utilities $ LotConfig $ LandSlope $ Neighborhood $ Condition1 $ Condition2 $ BldgType $
                   HouseStyle $ OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle $ RoofMatl $
                   Exterior1st $ Exterior2nd $ MasVnrType $ MasVnrArea ExterQual $ ExterCond $ Foundation $
                   BsmtQual $ BsmtCond $ BsmtExposure $ BsmtFinType1 $ BsmtFinSF1 BsmtFinType2 $ BsmtFinSF2 BsmtUnfSF TotalBsmtSF
                   Heating $ HeatingQC $ CentralAir $ Electrical $ _1stFlrSF _2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath
                   FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual $ TotRmsAbvGrd Functional $ Fireplaces FireplaceQu $
                   GarageType $ GarageYrBlt GarageFinish $ GarageCars GarageArea GarageQual $ GarageCond $ PavedDrive $ WoodDeckSF
                   OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch PoolArea PoolQC $ Fence $ MiscFeature $ MiscVal MoSold YrSold
                   SaleType $ SaleCondition $ SalePrice;
         /*Forward selection*/
         proc glmselect data=train seed=1;
                   class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope
                            Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl
                            Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual
                            BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir
                            Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish
                            GarageQual GarageCond PavedDrive Fence MiscFeature SaleType SaleCondition PoolQC;
                   model SalePrice=MSSubClass--SaleCondition / selection=forward (choose=CV
                             stop=CV) cvmethod=split(10) CVdetails;
                   output out=forward results p=Predict;
         run;
         /* Subset only Test data with salesprice prediction*/
         data forward_results2;
                   set forward results;
                   if Predict > 0 then
                            SalePrice=Predict:
                   if Predict < 0 then
                            SalePrice=10000;
                   keep id SalePrice;
                   where id > 1460;
         run;
```

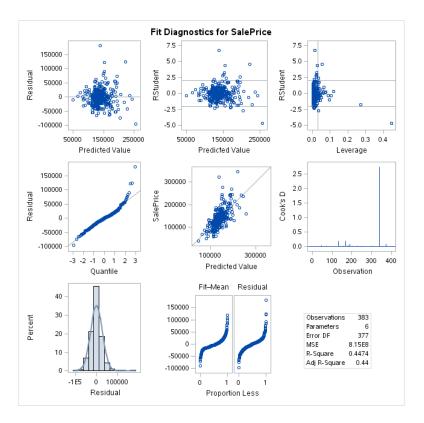
```
outfile='/folders/myfolders/sasuser.v94/submit_forward2.csv' dbms=csv replace;
run;
proc print data=forward_results2;
/*Backward elimination*/
proc glmselect data=train seed=1;
         class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope
                   Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl
                   Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual
                   BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir
                   Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish
                   GarageQual GarageCond PavedDrive Fence MiscFeature SaleType SaleCondition PoolQC;
         model SalePrice=MSSubClass--SaleCondition / selection=backward(choose=CV
                   stop=CV) cvmethod=split(10) CVdetails;
         output out=backward_results p=Predict;
run;
/* Subset only Test data with salesprice prediction*/
data backward_results2;
         set backward_results;
         if Predict > 0 then
                   SalePrice=Predict;
         if Predict < 0 then
                   SalePrice=10000;
         keep id SalePrice;
         where id > 1460;
run;
/* Export Kaggle Submission csv file*/
proc export data=backward_results2
                  outfile='/folders/myfolders/sasuser.v94/submit_backwards.csv' dbms=csv replace;
run;
proc print data=backward_results2;
run:
/*Stepwise selection*/
proc glmselect data=train seed=1;
         class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope
                   Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl
                   Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual
                   BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir
                   Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish
                   GarageQual GarageCond PavedDrive Fence MiscFeature SaleType SaleCondition PoolQC;
         model SalePrice=MSSubClass--SaleCondition / selection=stepwise(choose=CV
                   stop=CV) cvmethod=split(10) CVdetails;
         output out=stepwise_results p=Predict;
run;
/* Subset only Test data with salesprice prediction*/
data stepwise results2;
         set stepwise_results;
         if Predict > 0 then
                  SalePrice=Predict;
```

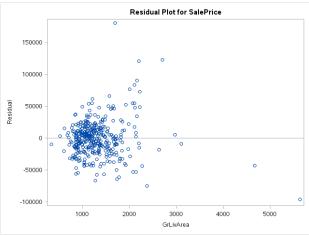
```
if Predict < 0 then
                   SalePrice=10000;
         keep id SalePrice;
         where id > 1460;
run;
/* Export Kaggle Submission csv file*/
proc export data=stepwise_results2
                   outfile='/folders/myfolders/sasuser.v94/submit stepwise.csv' dbms=csv replace;
run;
proc print data=stepwise_results2;
/*Custom model*/
proc glmselect data=train2 plots=all;
         class KitchenQual GarageFinish BsmtQual ExterQual MasVnrType Neighborhood;
         model SalePrice=LotFrontage LotArea OverallQual OverallCond YearBuilt
                   BsmtFinSF1 TotalBsmtSF _1stFlrSF _2ndFlrSF GrLivArea FullBath TotRmsAbvGrd
                   GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF KitchenQual
                   GarageFinish BsmtQual ExterQual MasVnrType Neighborhood / selection=none;
         output out=results p=Predict; run;
/* Subset only Test data with salesprice prediction*/
data results3;
         set results;
         if Predict > 0 then
                   SalePrice=Predict;
         if Predict < 0 then
                   SalePrice=10000;
         keep id SalePrice;
         where id > 1460;
run;
/* Export Kaggle Submission csv file*/
proc export data=results3 outfile='/folders/myfolders/sasuser.v94/submit custom2.csv'
                   dbms=csv replace;run;
/*Custom model, use for obtaining Cook's D, Leverate, and Studentized residuals;*/
proc glm data=train plots=all;
         class KitchenQual GarageFinish BsmtQual ExterQual MasVnrType Neighborhood;
         model SalePrice=LotFrontage LotArea OverallQual OverallCond YearBuilt
                   BsmtFinSF1\ TotalBsmtSF\ \_1stFlrSF\ \_2ndFlrSF\ GrLivArea\ FullBath\ TotRmsAbvGrd
                   GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF KitchenQual
                   GarageFinish BsmtQual ExterQual MasVnrType Neighborhood / solution;
         output out=results student=res cookd = cookd h = lev;run;
/*Display individuals with high Cook's D, large absolute studentized residuals or high leverage;*/
data results2;
         where id \leq 1460 and (cookd > 1 or res > 10 or res < -10 or lev > 1);
run;
proc print data = results2;run;
```

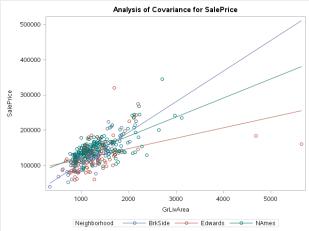
Appendix 1: Vanilla Regression – Raw Data | Model Original

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

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.447376	20.68070	28552.30	138062.5



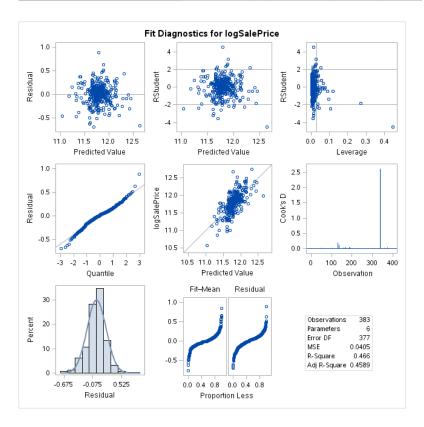


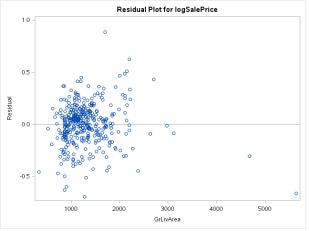


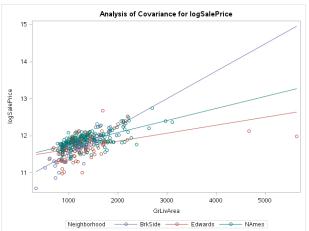
Appendix 2: Vanilla Regression | Model: Log-Linear

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	11.44334070	В	0.04465161	256.28	<.0001
GrLivArea	0.00032412	В	0.00003250	9.97	<.0001
Neighborhood BrkSide	-0.65174673	В	0.09780355	-6.66	<.0001
Neighborhood Edwards	-0.02139976	В	0.06409406	-0.33	0.7387
Neighborhood NAmes	0.00000000	В		_	
GrLivArea*Neighborho BrkSide	0.00041410	В	0.00007620	5.43	<.0001
GrLivArea*Neighborho Edwards	-0.00010744	В	0.00004482	-2.40	0.0170
GrLivArea*Neighborho NAmes	0.00000000	В		_	

R-Square	Coeff Var	Root MSE	logSalePrice Mean
0.465985	1.704877	0.201156	11.79887



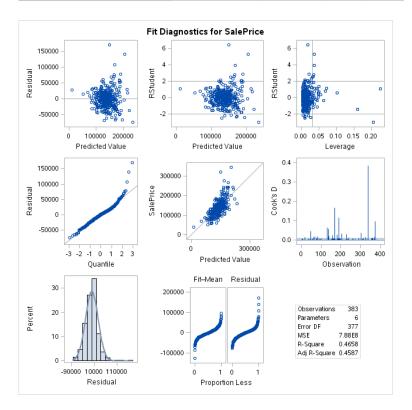


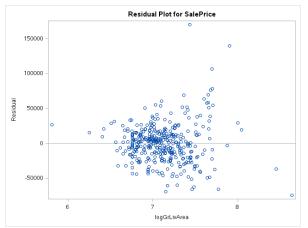


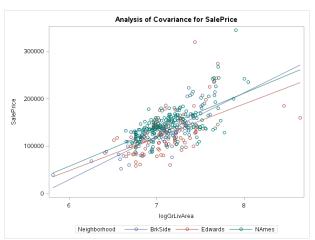
Appendix 3: Vanilla Regression | Model: Linear - Log

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-405474.3769	В	47364.43769	-8.56	<.0001
logGrLivArea	77263.2789	В	6632.56245	11.65	<.0001
Neighborhood BrkSide	-115363.9795	В	87581.82712	-1.32	0.1886
Neighborhood Edwards	29405.8761	В	75555.48597	0.39	0.6974
Neighborhood NAmes	0.0000	В			
logGrLivA*Neighborho BrkSide	14507.4723	В	12383.61185	1.17	0.2421
logGrLivA*Neighborho Edwards	-6546.6413	В	10581.99330	-0.62	0.5365
logGrLivA*Neighborho NAmes	0.0000	В			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.465809	20.33287	28072.06	138062.5



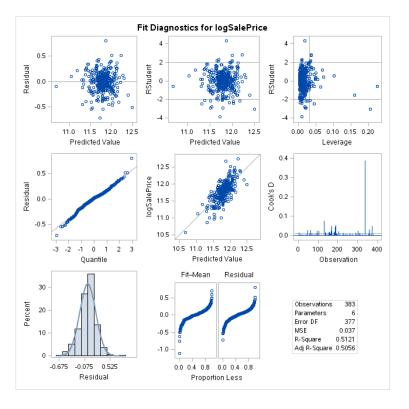


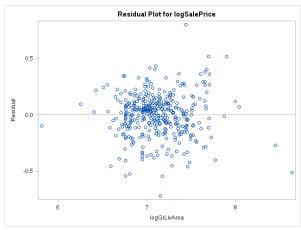


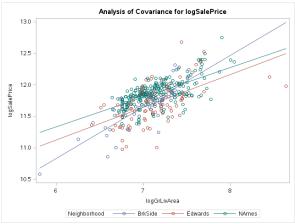
Appendix 4: Vanilla Regression | Model: Log-Log

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

R-Square	Coeff Var	Root MSE	logSalePrice Mean
0.512092	1.629617	0.192276	11.79887



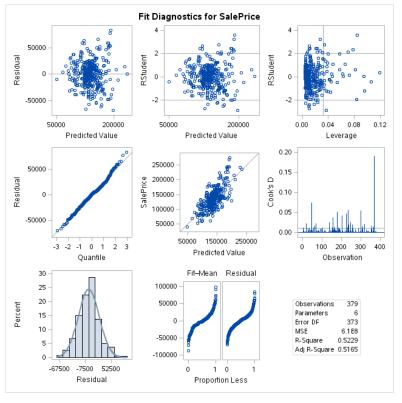


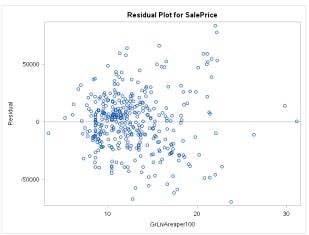


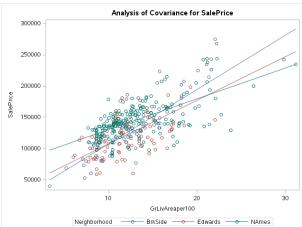
Appendix 5: Vanilla Regression | Model: Outliers Addressed*

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	80325.71230	В	5592.03832	14.36	<.0001
GrLivAreaper100	4956.12477	В	409.70671	12.10	<.0001
Neighborhood BrkSide	-60354.19850	В	12060.03479	-5.00	<.0001
Neighborhood Edwards	-43225.29073	В	10837.81644	-3.99	<.0001
Neighborhood NAmes	0.00000	В			
GrLivArea*Neighborho BrkSide	3760.12849	В	940.21789	4.00	<.0001
GrLivArea*Neighborho Edwards	2059.71212	В	820.38610	2.51	0.0125
GrLivArea*Neighborho NAmes	0.00000	В			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.522897	18.04909	24701.16	136855.4



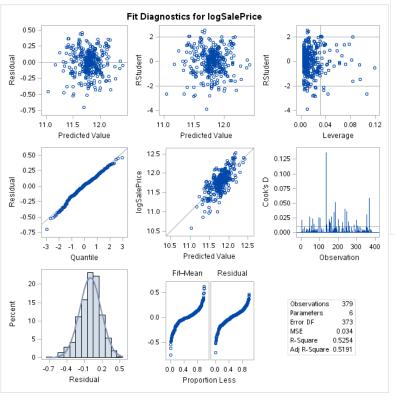


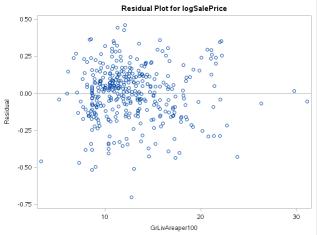


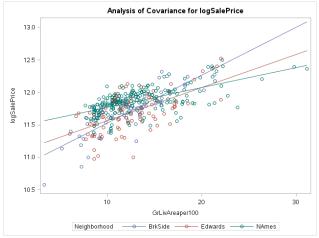
Appendix 6: Vanilla Regression | Model: Log-Linear (Outliers Addressed)

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	11.46308762	В	0.04175292	274.55	<.0001
GrLivAreaper100	0.03075050	В	0.00305907	10.05	<.0001
Neighborhood BrkSide	-0.67149365	В	0.09004618	-7.46	<.0001
Neighborhood Edwards	-0.41144266	В	0.08092049	-5.08	<.0001
Neighborhood NAmes	0.00000000	В		_	
GrLivArea*Neighborho BrkSide	0.04307178	В	0.00702013	6.14	<.0001
GrLivArea*Neighborho Edwards	0.02043149	В	0.00612541	3.34	0.0009
GrLivArea*Neighborho NAmes	0.00000000	В			

R-Square	Coeff Var	Root MSE	logSalePrice Mean
0.525413	1.563944	0.184431	11.79269



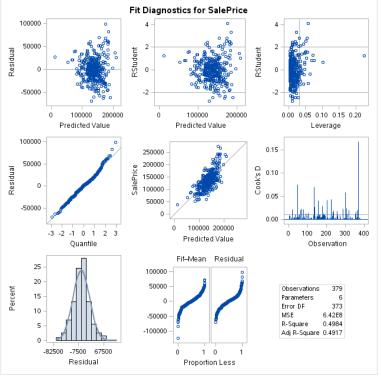


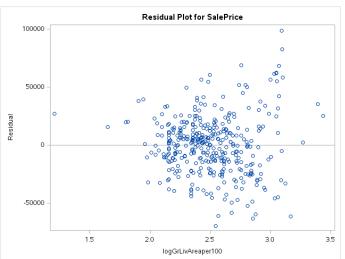


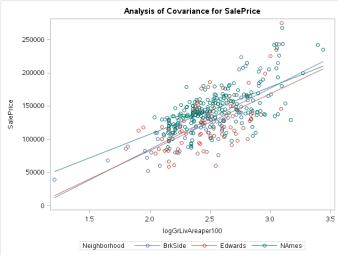
Appendix 7: Vanilla Regression | Model: Linear-Log (Outliers Addressed)

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-34567.92221	В	15470.76545	-2.23	0.0260
logGrLivAreaper100	71042.34430	В	6085.38391	11.67	<.0001
Neighborhood BrkSide	-63650.50684	В	27862.76048	-2.28	0.0229
Neighborhood Edwards	-52636.52333	В	28037.21033	-1.88	0.0612
Neighborhood NAmes	0.00000	В			
logGrLivA*Neighborho BrkSide	20728.40689	В	11227.58949	1.85	0.0657
logGrLivA*Neighborho Edwards	14188.73230	В	11130.71521	1.27	0.2032
logGrLivA*Neighborho NAmes	0.00000	В			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.498378	18.50708	25327.93	136855.4



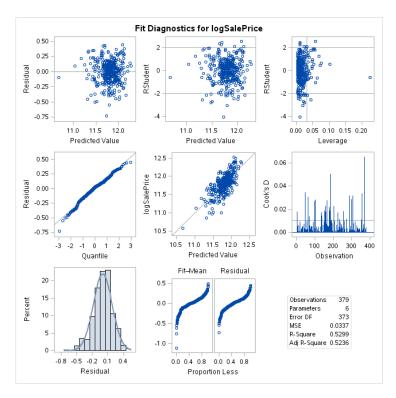




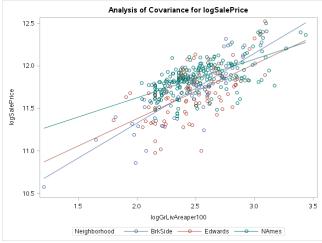
Appendix 8: Vanilla Regression | Model: Log-Log (Outliers Addressed)

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	10.72724642	В	0.11211841	95.68	<.0001
logGrLivAreaper100	0.44987850	В	0.04410147	10.20	<.0001
Neighborhood BrkSide	-1.03970690	В	0.20192461	-5.15	<.0001
Neighborhood Edwards	-0.62585626	В	0.20318887	-3.08	0.0022
Neighborhood NAmes	0.00000000	В			
logGrLivA*Neighborho BrkSide	0.36976955	В	0.08136763	4.54	<.0001
logGrLivA*Neighborho Edwards	0.18931410	В	0.08066557	2.35	0.0195
logGrLivA*Neighborho NAmes	0.00000000	В			

R-Square	Coeff Var	Root MSE	logSalePrice Mean
0.529914	1.556511	0.183554	11.79269



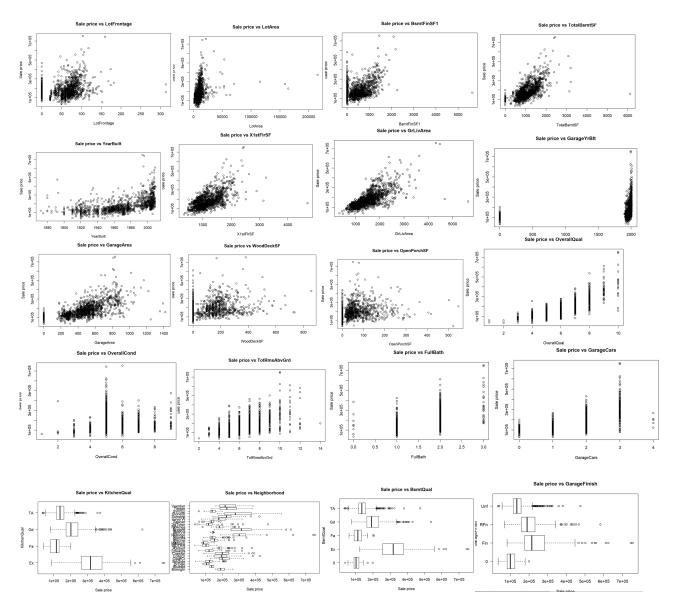




Appendix 9: Analysis 2 | Checking Assumptions

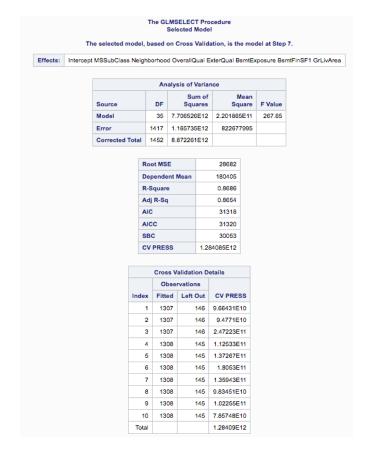
(Pairwise Variables with Linearity)

We disregarded the values of '0' as it was originally 'NA'. For a few of the plots, if '0' was not included in judging the plots, there would be sufficient evidence of correlation between the pairwise variables.



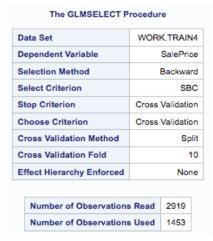
Appendix 10: Forward Selection SAS Results

Selection Method Forward Select Criterion SBC Stop Criterion Cross Validation Choose Criterion Cross Validation Cross Validation Method Split Cross Validation Fold 10	Data Set	WOR	K.TRAIN4		
Select Criterion SBC Stop Criterion Cross Validation Choose Criterion Cross Validation Cross Validation Method Split Cross Validation Fold 10	Dependent Variable		SalePrice		
Stop Criterion Cross Validation Choose Criterion Cross Validation Cross Validation Method Split Cross Validation Fold 10	Selection Method		Forward		
Choose Criterion Cross Validation Cross Validation Method Split Cross Validation Fold 10	Select Criterion		SBC		
Cross Validation Method Split Cross Validation Fold 10	Stop Criterion	Cross Validation			
Cross Validation Fold 10	Choose Criterion	Cross	Validation		
10	Cross Validation Method	Split			
Effect Hierarchy Enforced None	Cross Validation Fold	10			
	Effect Hierarchy Enforced		None		
	Number of Observations	Read	2919		
Number of Observations Read 2919	Number of Observations	Used	1453		



F	aram	eter Estimates		
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	23083	10518	2.19
MSSubClass	1	-275.366217	21.280080	-12.94
Neighborhood Blmngtn	1	-7602.992318	11296	-0.67
Neighborhood Blueste	1	-13329	22296	-0.60
Neighborhood BrDale	1	-28498	11717	-2.43
Neighborhood BrkSide	1	-34635	9732.085281	-3.56
Neighborhood ClearCr	1	-22912	10333	-2.22
Neighborhood CollgCr	1	-20846	9085.523663	-2.29
Neighborhood Crawfor	1	-11832	9735.585870	-1.22
Neighborhood Edwards	1	-44564	9384.758640	-4.78
Neighborhood Gilbert	1	-20714	9389.515096	-2.21
Neighborhood IDOTRR	1	-49267	10218	-4.82
Neighborhood MeadowV	1	-21426	11605	-1.88
Neighborhood Mitchel	1	-31967	9776.667107	-3.27
Neighborhood NAmes	1	-37252	9109.777197	-4.09
Neighborhood NPkVill	1	-10921	13243	-0.82
Neighborhood NWAmes	1	-30911	9542.657568	-3.24
Neighborhood NoRidge	1	18607	9985.134713	1.86
Neighborhood NridgHt	1	24169	9538.973472	2.53
Neighborhood OldTown	1	-47004	9380.734580	-5.01
Neighborhood SWISU	1	-52040	10704	-4.86
Neighborhood Sawyer	1	-33888	9558.483583	-3.5
Neighborhood SawyerW	1	-27646	9529.227186	-2.90
Neighborhood Somerst	1	1083.589725	9375.122303	0.12
Neighborhood StoneBr	1	33689	10549	3.19
Neighborhood Timber	1	-17183	9873.938952	-1.74
Neighborhood Veenker	0	0		
OverallQual	1	13747	1010.396593	13.61
ExterQual Ex	1	61646	5435.303177	11.34
ExterQual Fa	1	-4045.246585	8079.321547	-0.50
ExterQual Gd	1	9922.462958	2536.197499	3.91
ExterQual TA	0	0		
BsmtExposure 0	1	-789.054205	4985.392953	-0.16
BsmtExposure Av	1	11100	2363.318582	4.70
BsmtExposure Gd	1	33250	3062.316378	10.86
BsmtExposure Mn	1	8970.798814	2889.601355	3.10
BsmtExposure No	0	0		
BsmtFinSF1	1	24.467273	2.048623	11.94
GrLivArea	1	60.931671	2.057459	29.62

Appendix 11 : Analysis 2 | Backwards Elimination



	Ar	alysis of V	ariar	108			
Source	DF	Sun Squa			Mean quare	F Value	
Model	231	8.2735938	12	358164	21120	73.05	
Error	1221	5.9866788	E11	4903	09403		
Corrected Total	1452	8.8722618	12				
	R-Squa Adj R-S			180405 0.9325 0.9198 30742			
	AICC			30831			
	SBC			30512			
	CV PRI	nee	1.2082E12				

	Cross V	alidation D	etails
	Obse	rvations	
Index	Fitted	Left Out	CV PRESS
1	1307	146	6.29604E10
2	1307	146	7.0841E10
3	1307	146	2.43882E11
4	1308	145	8.98807E10
5	1308	145	1.03041E11
6	1308	145	1.38758E11
7	1308	145	9.06665E10
8	1308	145	7.13204E10
9	1308	145	8.98909E10
10	1308	145	2.46959E11
Total			1.2082E12

	Param	eter Estimates			Neighborhood Sawyer	- 1	-10370	8254.090541	-1.26	RoofMatl WdShngl	0	0			BsmtFinType1 ALQ	1	-2737.478979	2832.743013	-0.9
			Standard		Neighborhood SawyerW	- 1	-3338.198855	8263.887725	-0.40	Exterior1st AsbShng	- 1	6100.049267	7851.526579	0.78	BsmtFinType1 BLQ	1	-227.379663	3035.373339	-0.0
Parameter	DF	Estimate	Error	t Value	Neighborhood Somerst	- 1	-2959.142333	9572.621799	-0.31	Exterior1st AsphShn	1	-4987.949758	24740	-0.20	BsmtFinType1 GLQ	1	3275.584175	2646.285029	1.2
Intercept	1	455801	1019977	0.45	Neighborhood StoneBr	1	36704	9190.754496	3.99	Exterior1st BrkComm	- 1	3312.177857	18942	0.17	BsmtFinType1 LwQ	- 1	-5793.857420	3644.072557	-4.5
MSSubClass	1	-47.661185	80.414938	-0.59	Neighborhood Timber	- 1	-9994.277509	8830,439128	-1.13	Exterior1st BrkFace	1	17966	6066.708599	2.96	BsmtFinType1 Rec	- 1	-2745.080595	3085.563373	-0.1
MSZoning C (all)	1	-21515	9341.518072	-2.30	Neighborhood Veenker	0	0			Exterior1st CBlock	1	-12389	26309	-0.47	BsmtFinType1 Unf	0	0		
MSZoning FV	1	11452	7418.640168	1.54	Condition1 Arter	1	-11959	12318	.0.07	Exterior1st CemntBd	- 1	5013.425301	6347.241189	0.79	BsmtFinSF1	- 1	36.415168	5.173747	7.0
MSZoning RH	1	565.002446	7373.814431	0.08	Condition1 Feedr	÷	-7370.708354	11920		Exterior1st HdBoard	1	-1041 228705	5279.512800	-0.20	BsmtFinType2 0	- 1	-18004	24089	-0.7
MSZoning RL	1	3593.746453	3661.969781	0.98		1	1734 118922		0.62	Exterior1st ImStuce	1	-3536.489817	24201	-0.15	BsmtFinType2 ALQ	1	8291.756428	7386.126840	1.1
MSZoning RM	0	0			Condition1 Norm	1	-4834.036921	11631		Exterior1st MetalSd	-	3260.308700	5124,593809	0.64	BsmtFinType2 BLQ	- 1	-4236.939712	4970.933202	-0.8
LotFrontage	1	11.804637	22.370992	0.53	Condition1 PosA	-		14577		Exterior1st Plywood	1	-3040.923820	5631.892835	-0.54	BsmtFinType2 GLQ	1	5966,414679	8670.493999	0.6
Li trea	- 1	0.730481	0.104492	6.99	Condition1 PosN	- 1	-1093.918382	13026	-0.08	Exterior1st Stone	1	-3040.923820	19658	-0.54	BsmtFinType2 LwQ	1	-2462.341131	4500.949358	-0.5
St et GrvI	1	-32557	11902	-2.74	Condition1 RRAe	- 1	-29795	14101	-2.11		- 1				BsmtFinType2 Rec	1	-1234.668209	4676.981038	-0.2
St et Pave	0	0			Condition1 RRAn	1	-1370.463981	12236	-0.11	Exterior1st Stucco	1	6761.844773	7118.470764	0.95	BsmtFinType2 Unf	0	0		
Alley 0	1	529.020919	4612.754680	0.11	Condition1 RRNe	- 1	-17656	20135	-0.88	Exterior1st VinylSd	1	2524.268367	5170.980167	0.49	BsmtFinSF2	1	28.707654	8.815251	3.2
Alley Gryl	1	832.339977	5868.060953	0.14	Condition1 RRNn	0	0			Exterior1st Wd Sdng	1	53.264056	5096.300373	0.01	BsmtUnfSF	1	19 544998	4 734109	4.1
Alley Pave	0	0			Condition2 Arter	- 1	1427.552024	26445	0.05	Exterior1st WdShing	0	0	-		Heating Floo	-	-9167.346822	28759	-0.3
LotShape IR1	1	-2005.359904	1588 572325	-1.26	Condition2 Feedr	- 1	-4885.224225	20600	-0.24	MasVnrType 0	1	-8764.851594	8547.222067	-1.03	Heating GasA	-	-5921.954456	13793	-0.4
LotShape IR2	1	1984.673674	4221.582585	0.47	Condition2 Norm	- 1	-6278.701688	17299	-0.36	MasVnrType BrkCm	1	-10394	7055.660922	-1.47	Heating GasW	-	-9798 974734	14863	-0.6
LotShape IR3	1	3717.705375	8637.540302	0.43	Condition2 PosA	1	40450	36178	1.12	MasVnrType BrkFa	- 1	-5028.349186	2808.745293	-1.79	Heating Gasw	1	-9/98.9/4/34	16901	-0.6
LotShape Reg	0	0			Condition2 PosN	- 1	-233580	25246	-9.25	MasVnrType None	1	-2198.935374	3058.571806	-0.72	Heating OthW	-	-11122	22723	-1.2
LandContour Bok	1	-5012.242698	3589.675608	-1.40	Condition2 RRAe	1	-117890	63175	-1.87	MasVnrType Stone	0	0			Heating Wall	0	-20793	22123	-1.2
LandContour HLS	1	2831.838534	3972.579777	0.71	Condition2 RRAn	1	-17351	28866	-0.60	MasVnrArea	1	22.479543	5.658221	3.97			3029 300818	2013 989125	
LandContour Low		-15611	5518 448937	-2.83	Condition2 RRNn		0	20000	-0.00	ExterQual Ex	- 1	23014	5206.536432	4.42	HeatingQC Ex	- 1	3029.300818 4364.228857	2013.989125 4392.626582	1.5
LandContour Lvl	0	-15011	0016.440937	-2.03	BldgType 1Fam	1	15581	8742.768486	1.78	ExterQual Fa	1	11948	9435.239124	1.27	HeatingQC Fa	- 1			
	_	-	05700			-	15501		1.74	ExterQual Gd	- 1	535.792299	2397.200372	0.22	HeatingQC Gd	1	-636.612877	2102.823900	-0.3
Utilities AllPub	1	39760	25792	1.54	BldgType 2fmCon	-1	12821	7383.062078		ExterQual TA	0	0			HeatingQC Po	1	6061.782721	25988	0.2
Utilities NoSeWa	0	0			BldgType Duplex	1	8418.536751	8491.364783	0.99	ExterCond Ex	1	2832.038076	16804	0.17	HeatingQC TA	0	0		
LotConfig Corner	1	1463.381584	1711.234849	0.86	BldgType Twnhs	-1	-4763.643432	4803.273042	-0.99	ExterCond Fa	1	-264.852697	5594.939143	-0.05	CentralAir N	1	-560.747923	3753.312070	-0.1
LotConfig CuIDSa	1	10029	2936.803256	3.42	BldgType TwnhsE	0	0			ExterCond Gd	1	-1907.991800	2328.520292	-0.82	CentralAir Y	0	0		
LotConfig FR2	1	-5660.332208	3672.918171	-1.54	HouseStyle 105Fin	1	-3958.447514	5434.998208	-0.73	ExterCond Po	-	5626.576929	25364	0.22	Electrical 0	1	13970	23362	0.6
LotConfig FR3	1	-14168	12199	-1.16	HouseStyle 105Unf	1	8677.255532	8635.834236	1.00	ExterCond TA	0	0	25554	0.22	Electrical FuseA	1	2696.431433	2861.122451	0.9
LotConfig Inside	0	0			HouseStyle 1Story	1	2147.503108	5883.017492	0.37	Foundation BrkTil	1	23051	14432	1.60	Electrical FuseF	1	1188.287942	5343.537515	0.2
LandSlope Gtl	1	43224	11116	3.89	HouseStyle 205Fin	- 1	-29241	13303	-2.20	Foundation CBlock	1	26363	14211	1.86	Electrical FuseP	1	-2847.900137	17734	-0.1
LandSlope Mod	1	50495	11048	4.57	HouseStyle 205Unf	1	-13427	9736.291789	-1.38	Foundation CBlock	1	28236	14113	2.00	Electrical Mix	1	-43972	41637	-1.0
LandSlope Sev	0	0			HouseStyle 2Story	- 1	-8753.513870	5181,853335	-1.69	Foundation PConc	1	18817			Electrical SBrkr	0	0		
Neighborhood Blmngtr	1	578.779303	10214	0.06	HouseStyle SFover	1	-2380.377156	5384.269778	-0.44				16826	1.12	A1stFirSF	1	44.321615	5.490132	8.0
Neighborhood Blueste	1	7496.630490	19000	0.39	HouseStyle SLvI	0	0			Foundation Stone	- 1	30801	17874	1.72	A2ndFlrSF	1	60.971619	5.550750	10.9
Neighborhood BrDale	1	-1303.307897	11245	-0.12	OverallQual	1	6804.476213	985.285364	6.91	Foundation Wood	0	0			LowQualFinSF	- 1	25.015558	18.958772	1.3
Neighborhood BrkSide	- 1	-5699.351008	9189.583066	-0.62	OverallCond	1	5486.284727	849.804231	6.46	BsmtQual 0	1	40926	35066	1.17	BsmtFullBath	- 1	1720.024480	1926.620555	0.8
Neighborhood ClearCr	1	-14707	9246.047859	-1.59	YearBuilt	+	299.197264	74.323153		BsmtQual Ex	- 1	14189	4024.210428	3.53	BsmtHalfBath	1	801.885105	2923.785455	0.2
Neighborhood CollgCr	1	-9906.376941	8037.091543	-1.23		-			4.03	BsmtQual Fa	1	2138.427037	4748.397502	0.45	FullBath	- 1	3697.729967	2138.813106	1.7
Neighborhood Crawfor	- 1	9583.684201	8662.312413	1.11	YearRemodAdd	1	118.998282	54.367681		BsmtQual Gd	1	-2862.885884	2477.067201	-1.16	HalfBath	- 1	1733.927607	2029.680834	0.8
Neighborhood Edward	1	-20526	8318.071740	-2.47	RoofStyle Flat	1	-93641	33577	-2.79	BsmtQual TA	0	0			BedroomAbvGr	1	-3387.471228	1319.941030	-2.5
Neighborhood Gilbert	1	-10245	8402.251023	-1.22	RoofStyle Gable	- 1	-89142	29129	-3.06	BsmtCond 0	0	0			KitchenAbvGr	- 1	-14485	5508.261594	-2.6
Neighborhood IDOTRR	1	-13648	10378	-1.32	RoofStyle Gambrel	- 1	-85564	30081		BsmtCond Fa	1	-3231.985232	4132.307362	-0.78	KitchenQual Ex	- 1	20740	3831.324413	5.4
Neighborhood Meadow	V 1	-6840.749053	11375	-0.60	RoofStyle Hip	- 1	-89462	29157	-3.07	BsmtCond Gd	1	-3177.292960	3171.340280	-1.00	KitchenQual Fa	1	3286.230145	4720.097108	0.7
Neighborhood Mitchel	1	-20826	8409 582820	-2.45	RoofStyle Mansard	- 1	-79003	28867	-2.74	BsmtCond Po	1	72796	29028	2.51	KitchenQual Gd	1	-725.042008	2082.906848	-0.3
Neighborhood NAmes	1	-16562	7950.338677	-2.08	RoofStyle Shed	0	0			BsmtCond TA	0	0			KitchenQual TA	0	0		-
Neighborhood NAMes Neighborhood NPkVill	1	13214	11373	1.16	RoofMatl CompShg	- 1	-54207	11057	-4.90	BsmtExposure 0	- 1	-6243.459554	22480	-0.28	TotRmsAbvGrd	1	1729.918503	929.125504	1.8
-		13214	8182.603346	-2.08	RoofMatl Membran	1	35702	33963		BsmtExposure Av	- 1	4900.025199	2129.387256	2.30	Functional Mai	1	.20989	6262.465087	-3.1
Neighborhood NWAme	_				RoofMati Metal	1	6662.328806	33003		BsmtExposure Gd	1	19382	2897.942328	6.69	Functional Min	1	20000	3320.987389	-3.
Neighborhood NoRidg		22769	8825.154154	2.58	RoofMati Roll	1	-68297	27218		BsmtExposure Mn	-	1418 849766	2456.641154	0.58		1			-
Neighborhood NridgHt	1	17620	8572.122664	2.06	RoofMatl Tar&Gry	-	-54162	20798		BsmtExposure No	0	1418.849700	2400.041104	0.56	Functional Mod	1	-21589	7383.777413	-2.9
Neighborhood OldTow	. 1	-15622	9343.086167	-1.67	Rooman laragev	1	+54162	20798	-2.60	BSmtExposure No	0	0			Functional Sev	1	-60914	28127	-2.1

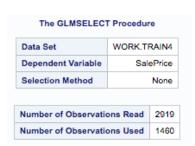
Appendix 12: Analysis 2 | Stepwise Selection

Data Set	WORK.TRAIN4			
Dependent Variable	SalePrice			
Selection Method	Stepwise			
Select Criterion	SBC			
Stop Criterion	Cross Validation			
Choose Criterion	Cross Validation			
Cross Validation Method	Spli			
Cross Validation Fold	10			
Effect Hierarchy Enforced	None			

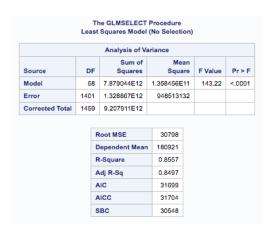
		An	alysis o	f Varia	nce		
Source		DF		Sum of quares		lean uare	F Value
Model		35	7.7065	26E12	2.201869	E11	267.65
Error	1	1417	1.1657	35E12	82267	7995	
Corrected Tot	al '	1452	8.8722	61E12			
	Ro	ot MSI	E		28682		
	De	pende	nt Mea	1	180405		
	R-9	Square	•		0.8686		
	Ad	R-Sq			0.8654		
	AIC	;			31318		
	AIC	cc			31320		
	SB	С			30053		
	cv	PRES	S	1.28	34085E12		
		Cros	s Valida	tion D	etails		
		Ob	Observation				
In	dex	Fitte	d Le	t Out	CV PRES	SS	
	1	130	7	146	9.66431E	10	
	2	130	7	146	9.4771E	10	
	3	130	7	146	2.47223E11		
					1.12533E11		
	4	130	8	145	1.12533E	11	
	5	130		145 145	1.12533E 1.37267E		
)8			11	
	5	130	8 8	145	1.37267E	11	
	5	130)8)8)8	145 145	1.37267E 1.8053E	11 11 11	
	5 6 7	130 130)8)8)8)8	145 145 145	1.37267E 1.8053E 1.35943E	11 11 11 10	
	5 6 7 8	130 130 130)8)8)8)8	145 145 145 145	1.37267E 1.8053E 1.35943E 9.83451E	11 11 11 10	

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	
Intercept	1	23083	10518	2.19	
MSSubClass	1	-275.366217	21.280080	-12.94	
Neighborhood Blmngtn	1	-7602.992318	11296	-0.67	
Neighborhood Blueste	1	-13329	22296	-0.60	
Neighborhood BrDale	1	-28498	11717	-2.43	
Neighborhood BrkSide	1	-34635	9732.085281	-3.56	
Neighborhood ClearCr	1	-22912	10333	-2.22	
Neighborhood CollgCr	1	-20846	9085.523663	-2.29	
Neighborhood Crawfor	1	-11832	9735.585870	-1.22	
Neighborhood Edwards	1	-44564	9384.758640	-4.75	
Neighborhood Gilbert	1	-20714	9389.515096	-2.2	
Neighborhood IDOTRR	1	-49267	10218	-4.83	
Neighborhood MeadowV	1	-21426	11605	-1.8	
Neighborhood Mitchel	1	-31967	9776.667107	-3.2	
Neighborhood NAmes	1	-37252	9109.777197	-4.09	
Neighborhood NPkVill	1	-10921	13243	-0.8	
Neighborhood NWAmes	1	-30911	9542.657568	-3.2	
Neighborhood NoRidge	1	18607	9985.134713	1.8	
Neighborhood NridgHt	1	24169	9538.973472	2.5	
Neighborhood OldTown	1	-47004	9380.734580	-5.0	
Neighborhood SWISU	1	-52040	10704	-4.8	
Neighborhood Sawyer	1	-33888	9558.483583	-3.5	
Neighborhood SawyerW	1	-27646	9529.227186	-2.9	
Neighborhood Somerst	1	1083.589725	9375.122303	0.1	
Neighborhood StoneBr	1	33689	10549	3.1	
Neighborhood Timber	1	-17183	9873.938952	-1.7	
Neighborhood Veenker	0	0			
OverallQual	1	13747	1010.396593	13.6	
ExterQual Ex	1	61646	5435.303177	11.3	
ExterQual Fa	1	-4045.246585	8079.321547	-0.5	
ExterQual Gd	1	9922.462958	2536.197499	3.9	
ExterQual TA	0	0			
BsmtExposure 0	1	-789.054205	4985.392953	-0.1	
BsmtExposure Av	1	11100	2363.318582	4.7	
BsmtExposure Gd	1	33250	3062.316378	10.8	
BsmtExposure Mn	1	8970.798814	2889.601355	3.1	
BsmtExposure No	0	0			
BsmtFinSF1	1	24.467273	2.048623	11.9	
GrLivArea	1	60.931671	2.057459	29.6	

Appendix 13: Analysis 2 | Custom Selection



	Pa	arameter Estima	tes		
Parameter	DF	Estimate	Standard Error	t Value	Pr > [t]
Intercept	1	-427134	165878	-2.57	0.0101
LotFrontage	1	-20.289697	25.933358	-0.78	0.4341
LotArea	1	0.545618	0.096061	5.68	<.0001
OverallQual	1	10486	1163.841790	9.01	<.0001
OverallCond	1	6775.936796	881.410022	7.69	<.0001
YearBuilt	1	298.450330	76.443384	3.90	<.0001
BsmtFinSF1	1	11.857029	2.310627	5.13	<.0001
TotalBsmtSF	1	3.199482	4.882422	0.66	0.5124
A1stFirSF	1	27.624703	18.947882	1.46	0.1451
A2ndFirSF	1	18.648615	18.375416	1.01	0.3103
GrLivArea	1	19.129055	18.322085	1.04	0.2966
FullBath	1	2430.130029	2385.894354	1.02	0.3086
TotRmsAbvGrd	1	1163.576980	973.104965	1.20	0.2320
GarageYrBlt	1	-95.689282	68.537013	-1.40	0.1629
GarageCars	1	14015	2772.172788	5.06	<.0001
GarageArea	1	1.331760	9.463098	0.14	0.8881
WoodDeckSF	1	23.713846	7.144184	3.32	0.0009
OpenPorchSF	1	5.858968	13.737009	0.43	0.6698
KitchenQual Ex	1	30562	4820.201639	6.34	<.0001
KitchenQual Fa	1	673.600459	5631.775551	0.12	0.9048
KitchenQual Gd	1	4211.597007	2560.005327	1.65	0.1002
KitchenQual TA	0	0			
GarageFinish 0	1	-172383	133386	-1.29	0.1964
GarageFinish Fin	1	6447.686766	2862.799050	2.25	0.0245
GarageFinish RFn	1	2305.719977	2470.049925	0.93	0.3507
GarageFinish Unf	0	0			
BsmtQual 0	1	-9626.708062	7309.116629	-1.32	0.1880
BsmtQual Ex	1	30387	5025.225166	6.05	<.0001
BsmtQual Fa	1	-1825.588841	5688.341768	-0.32	0.7483
BsmtQual Gd	1	-2523.238208	2957.712595	-0.85	0.3937
BsmtQual TA	0	0			
ExterQual Ex	1	19183	6503.615057	2.95	0.0032
ExterQual Fa	1	2890.634534	9229.274923	0.31	0.7542
ExterQual Gd	1	1655.229412	3047.200080	0.54	0.5871



BsmtQual TA	0	0			
ExterQual Ex	1	19183	6503.615057	2.95	0.0032
ExterQual Fa	1	2890.634534	9229.274923	0.31	0.7542
ExterQual Gd	1	1655.229412	3047.200080	0.54	0.5871
ExterQual TA	0	0			
MasVnrType 0	1	-727.980970	11377	-0.06	0.9490
MasVnrType BrkCm	1	-13212	8855.522649	-1.49	0.1359
MasVnrType BrkFa	1	-1912.709008	3508.983949	-0.55	0.5858
MasVnrType None	1	470.556562	3628.403230	0.13	0.8968
MasVnrType Stone	0	0			
Neighborhood Blmngtn	1	-30558	12415	-2.46	0.0140
Neighborhood Blueste	1	-50121	24004	-2.09	0.0370
Neighborhood BrDale	-1	-45077	12846	-3.51	0.0005
Neighborhood BrkSide	1	-24276	10848	-2.24	0.0254
Neighborhood ClearCr	1	-13600	11278	-1.21	0.2281
Neighborhood CollgCr	- 1	-15807	9902.689966	-1.60	0.1107
Neighborhood Crawfor	1	-1818.648751	10774	-0.17	0.8660
Neighborhood Edwards	1	-39691	10274	-3.86	0.0001
Neighborhood Gilbert	1	-22840	10355	-2.21	0.0276
Neighborhood IDOTRR	1	-35346	11470	-3.08	0.0021
Neighborhood MeadowV	1	-44209	12376	-3.57	0.0004
Neighborhood Mitchel	1	-36508	10566	-3.46	0.0006
Neighborhood NAmes	1	-28200	9988.832172	-2.82	0.0048
Neighborhood NPkVill	1	-38212	14239	-2.68	0.0074
Neighborhood NWAmes	1	-31493	10326	-3.05	0.0023
Neighborhood NoRidge	1	38068	10852	3.51	0.0005
Neighborhood NridgHt	1	8794.347363	10577	0.83	0.4059
Neighborhood OldTown	1	-40255	10637	-3.78	0.0002
Neighborhood SWISU	1	-33778	12118	-2.79	0.0054
Neighborhood Sawyer	1	-29777	10361	-2.87	0.0041
Neighborhood SawyerW	- 1	-22278	10356	-2.15	0.0316
Neighborhood Somerst	- 1	-8774.462064	10342	-0.85	0.3963
Neighborhood StoneBr	1	25030	11373	2.20	0.0279
Neighborhood Timber	1	-18263	10808	-1.69	0.0913
Neighborhood Veenker	0	0			

APPENDIX 14 : Kaggle Submission Score

Submission and Description	Public Score	Use for Final Score
submit_stepwise3.csv 11 minutes ago by Andy H add submission details	0.15937	♂
submit_forward2.csv a day ago by Andy H add submission details	0.15937	♂
submit_custom2.csv a day ago by Andy H add submission details	0.18171	♂
submit_backwards.csv a day ago by Andy H add submission details	0.17309	♂