Mid-Term Project : PCA and Multiple Linear Regression

Smita Sukhadeve

# Appendix & Program Code

***A: Dataset summary***

ames0 = read.csv("http://people.usm.maine.edu/cpeng/datasets/amescsv.csv")  
# dimension of the dataset  
dim(ames0)

## [1] 2930 82

# Summary of the Dataset  
summary(ames0)

## Order PID MSSubCls MSZoning   
## Min. : 1.0 Min. :5.263e+08 Min. : 20.00 A (agr): 2   
## 1st Qu.: 733.2 1st Qu.:5.285e+08 1st Qu.: 20.00 C (all): 25   
## Median :1465.5 Median :5.355e+08 Median : 50.00 FV : 139   
## Mean :1465.5 Mean :7.145e+08 Mean : 57.39 I (all): 2   
## 3rd Qu.:2197.8 3rd Qu.:9.072e+08 3rd Qu.: 70.00 RH : 27   
## Max. :2930.0 Max. :1.007e+09 Max. :190.00 RL :2273   
## RM : 462   
## LotFrontage LotArea Street Alley LotShape   
## Min. : 21.00 Min. : 1300 Grvl: 12 Grvl: 120 IR1: 979   
## 1st Qu.: 58.00 1st Qu.: 7440 Pave:2918 Pave: 78 IR2: 76   
## Median : 68.00 Median : 9436 NA's:2732 IR3: 16   
## Mean : 69.22 Mean : 10148 Reg:1859   
## 3rd Qu.: 80.00 3rd Qu.: 11555   
## Max. :313.00 Max. :215245   
## NA's :490   
## LandContour Utilities LotConfig LandSlope Neighborhood   
## Bnk: 117 AllPub:2927 Corner : 511 Gtl:2789 NAmes : 443   
## HLS: 120 NoSeWa: 1 CulDSac: 180 Mod: 125 CollgCr: 267   
## Low: 60 NoSewr: 2 FR2 : 85 Sev: 16 OldTown: 239   
## Lvl:2633 FR3 : 14 Edwards: 194   
## Inside :2140 Somerst: 182   
## NridgHt: 166   
## (Other):1439   
## Condition1 Condition2 BldgType HouseStyle   
## Norm :2522 Norm :2900 1Fam :2425 1Story :1481   
## Feedr : 164 Feedr : 13 2fmCon: 62 2Story : 873   
## Artery : 92 Artery : 5 Duplex: 109 1.5Fin : 314   
## RRAn : 50 PosA : 4 Twnhs : 101 SLvl : 128   
## PosN : 39 PosN : 4 TwnhsE: 233 SFoyer : 83   
## RRAe : 28 RRNn : 2 2.5Unf : 24   
## (Other): 35 (Other): 2 (Other): 27   
## OverallQual OverallCond YearBuilt YearRemodAdd   
## Min. : 1.000 Min. :1.000 Min. :1872 Min. :1950   
## 1st Qu.: 5.000 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1965   
## Median : 6.000 Median :5.000 Median :1973 Median :1993   
## Mean : 6.095 Mean :5.563 Mean :1971 Mean :1984   
## 3rd Qu.: 7.000 3rd Qu.:6.000 3rd Qu.:2001 3rd Qu.:2004   
## Max. :10.000 Max. :9.000 Max. :2010 Max. :2010   
##   
## RoofStyle RoofMatl Exterior1st Exterior2nd   
## Flat : 20 CompShg:2887 VinylSd:1026 VinylSd:1015   
## Gable :2321 Tar&Grv: 23 MetalSd: 450 MetalSd: 447   
## Gambrel: 22 WdShake: 9 HdBoard: 442 HdBoard: 406   
## Hip : 551 WdShngl: 7 Wd Sdng: 420 Wd Sdng: 397   
## Mansard: 11 ClyTile: 1 Plywood: 221 Plywood: 274   
## Shed : 5 Membran: 1 CemntBd: 126 CmentBd: 126   
## (Other): 2 (Other): 245 (Other): 265   
## MasVnrType MasVnrArea ExterQual ExterCond Foundation   
## : 23 Min. : 0.0 Ex: 107 Ex: 12 BrkTil: 311   
## BrkCmn : 25 1st Qu.: 0.0 Fa: 35 Fa: 67 CBlock:1244   
## BrkFace: 880 Median : 0.0 Gd: 989 Gd: 299 PConc :1310   
## CBlock : 1 Mean : 101.9 TA:1799 Po: 3 Slab : 49   
## None :1752 3rd Qu.: 164.0 TA:2549 Stone : 11   
## Stone : 249 Max. :1600.0 Wood : 5   
## NA's :23   
## BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1   
## : 1 : 1 : 4 GLQ :859 Min. : 0.0   
## Ex : 258 Ex : 3 Av : 418 Unf :851 1st Qu.: 0.0   
## Fa : 88 Fa : 104 Gd : 284 ALQ :429 Median : 370.0   
## Gd :1219 Gd : 122 Mn : 239 Rec :288 Mean : 442.6   
## Po : 2 Po : 5 No :1906 BLQ :269 3rd Qu.: 734.0   
## TA :1283 TA :2616 NA's: 79 (Other):155 Max. :5644.0   
## NA's: 79 NA's: 79 NA's : 79 NA's :1   
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF   
## Unf :2499 Min. : 0.00 Min. : 0.0 Min. : 0   
## Rec : 106 1st Qu.: 0.00 1st Qu.: 219.0 1st Qu.: 793   
## LwQ : 89 Median : 0.00 Median : 466.0 Median : 990   
## BLQ : 68 Mean : 49.72 Mean : 559.3 Mean :1052   
## ALQ : 53 3rd Qu.: 0.00 3rd Qu.: 802.0 3rd Qu.:1302   
## (Other): 36 Max. :1526.00 Max. :2336.0 Max. :6110   
## NA's : 79 NA's :1 NA's :1 NA's :1   
## Heating HeatingQC CentralAir Electrical X1stFlrSF   
## Floor: 1 Ex:1495 N: 196 : 1 Min. : 334.0   
## GasA :2885 Fa: 92 Y:2734 FuseA: 188 1st Qu.: 876.2   
## GasW : 27 Gd: 476 FuseF: 50 Median :1084.0   
## Grav : 9 Po: 3 FuseP: 8 Mean :1159.6   
## OthW : 2 TA: 864 Mix : 1 3rd Qu.:1384.0   
## Wall : 6 SBrkr:2682 Max. :5095.0   
##   
## X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath   
## Min. : 0.0 Min. : 0.000 Min. : 334 Min. :0.0000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.:1126 1st Qu.:0.0000   
## Median : 0.0 Median : 0.000 Median :1442 Median :0.0000   
## Mean : 335.5 Mean : 4.677 Mean :1500 Mean :0.4314   
## 3rd Qu.: 703.8 3rd Qu.: 0.000 3rd Qu.:1743 3rd Qu.:1.0000   
## Max. :2065.0 Max. :1064.000 Max. :5642 Max. :3.0000   
## NA's :2   
## BsmtHalfBath FullBath HalfBath BedroomAbvGr   
## Min. :0.00000 Min. :0.000 Min. :0.0000 Min. :0.000   
## 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000   
## Median :0.00000 Median :2.000 Median :0.0000 Median :3.000   
## Mean :0.06113 Mean :1.567 Mean :0.3795 Mean :2.854   
## 3rd Qu.:0.00000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000   
## Max. :2.00000 Max. :4.000 Max. :2.0000 Max. :8.000   
## NA's :2   
## KitchenAbvGr KitchenQual TotRmsAbvGrd Functional   
## Min. :0.000 Ex: 205 Min. : 2.000 Typ :2728   
## 1st Qu.:1.000 Fa: 70 1st Qu.: 5.000 Min2 : 70   
## Median :1.000 Gd:1160 Median : 6.000 Min1 : 65   
## Mean :1.044 Po: 1 Mean : 6.443 Mod : 35   
## 3rd Qu.:1.000 TA:1494 3rd Qu.: 7.000 Maj1 : 19   
## Max. :3.000 Max. :15.000 Maj2 : 9   
## (Other): 4   
## Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish  
## Min. :0.0000 Ex : 43 2Types : 23 Min. :1895 : 2   
## 1st Qu.:0.0000 Fa : 75 Attchd :1731 1st Qu.:1960 Fin : 728   
## Median :1.0000 Gd : 744 Basment: 36 Median :1979 RFn : 812   
## Mean :0.5993 Po : 46 BuiltIn: 186 Mean :1978 Unf :1231   
## 3rd Qu.:1.0000 TA : 600 CarPort: 15 3rd Qu.:2002 NA's: 157   
## Max. :4.0000 NA's:1422 Detchd : 782 Max. :2207   
## NA's : 157 NA's :159   
## GarageCars GarageArea GarageQual GarageCond PavedDrive  
## Min. :0.000 Min. : 0.0 : 1 : 1 N: 216   
## 1st Qu.:1.000 1st Qu.: 320.0 Ex : 3 Ex : 3 P: 62   
## Median :2.000 Median : 480.0 Fa : 124 Fa : 74 Y:2652   
## Mean :1.767 Mean : 472.8 Gd : 24 Gd : 15   
## 3rd Qu.:2.000 3rd Qu.: 576.0 Po : 5 Po : 14   
## Max. :5.000 Max. :1488.0 TA :2615 TA :2665   
## NA's :1 NA's :1 NA's: 158 NA's: 158   
## WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch   
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.75 Mean : 47.53 Mean : 23.01 Mean : 2.592   
## 3rd Qu.: 168.00 3rd Qu.: 70.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :1012.00 Max. :508.000   
##   
## ScreenPorch PoolArea PoolQC Fence MiscFeature  
## Min. : 0 Min. : 0.000 Ex : 4 GdPrv: 118 Elev: 1   
## 1st Qu.: 0 1st Qu.: 0.000 Fa : 2 GdWo : 112 Gar2: 5   
## Median : 0 Median : 0.000 Gd : 4 MnPrv: 330 Othr: 4   
## Mean : 16 Mean : 2.243 TA : 3 MnWw : 12 Shed: 95   
## 3rd Qu.: 0 3rd Qu.: 0.000 NA's:2917 NA's :2358 TenC: 1   
## Max. :576 Max. :800.000 NA's:2824   
##   
## MiscVal MoSold YrSold SaleType   
## Min. : 0.00 Min. : 1.000 Min. :2006 WD :2536   
## 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007 New : 239   
## Median : 0.00 Median : 6.000 Median :2008 COD : 87   
## Mean : 50.63 Mean : 6.216 Mean :2008 ConLD : 26   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009 CWD : 12   
## Max. :17000.00 Max. :12.000 Max. :2010 ConLI : 9   
## (Other): 21   
## SaleCondition SalePrice   
## Abnorml: 190 Min. : 12789   
## AdjLand: 12 1st Qu.:129500   
## Alloca : 24 Median :160000   
## Family : 46 Mean :180796   
## Normal :2413 3rd Qu.:213500   
## Partial: 245 Max. :755000   
##

# *Appendix B: Section 3.1 : Full Model With all continuous predictors*

lm\_allCont = lm(SalePrice ~.-Order, data = contVar.dat)  
summary(lm\_allCont)

##   
## Call:  
## lm(formula = SalePrice ~ . - Order, data = contVar.dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -683359 -19899 282 18904 309782   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.565e+04 3.638e+03 -4.302 1.76e-05 \*\*\*  
## LotFrontage -9.085e+01 4.683e+01 -1.940 0.052487 .   
## LotArea 3.776e-01 1.624e-01 2.324 0.020184 \*   
## MasVnrArea 5.968e+01 5.797e+00 10.295 < 2e-16 \*\*\*  
## BsmtFinSF1 5.736e+01 3.679e+00 15.593 < 2e-16 \*\*\*  
## BsmtFinSF2 3.650e+01 6.331e+00 5.766 9.14e-09 \*\*\*  
## BsmtUnfSF 3.921e+01 3.593e+00 10.913 < 2e-16 \*\*\*  
## TotalBsmtSF NA NA NA NA   
## X1stFlrSF 6.417e+01 4.272e+00 15.022 < 2e-16 \*\*\*  
## X2ndFlrSF 6.518e+01 2.464e+00 26.452 < 2e-16 \*\*\*  
## LowQualFinSF -3.175e+00 1.822e+01 -0.174 0.861693   
## GrLivArea NA NA NA NA   
## GarageArea 8.932e+01 5.030e+00 17.756 < 2e-16 \*\*\*  
## WoodDeckSF 5.850e+01 7.867e+00 7.437 1.43e-13 \*\*\*  
## OpenPorchSF 4.471e+01 1.406e+01 3.180 0.001490 \*\*   
## EnclosedPorch -5.541e+01 1.412e+01 -3.924 8.97e-05 \*\*\*  
## X3SsnPorch 3.117e+01 3.562e+01 0.875 0.381622   
## ScreenPorch 5.321e+01 1.583e+01 3.362 0.000785 \*\*\*  
## PoolArea -8.633e+01 2.511e+01 -3.439 0.000594 \*\*\*  
## MiscVal -1.882e+01 1.776e+00 -10.600 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43090 on 2403 degrees of freedom  
## (509 observations deleted due to missingness)  
## Multiple R-squared: 0.7346, Adjusted R-squared: 0.7327   
## F-statistic: 391.2 on 17 and 2403 DF, p-value: < 2.2e-16

# Section 1.2 - Removed Outliers from the Working Dataset

# Full model result after removing Outliers  
workingSet2 = workingSet2[which(workingSet2$GrLivArea < 4000), ]  
contVar.dat <- subset(workingSet2, select = continuous.var)  
dim(contVar.dat)

## [1] 2925 21

lm\_noOutlier = lm(SalePrice~ .-Order, data = contVar.dat)  
summary(lm\_noOutlier)

##   
## Call:  
## lm(formula = SalePrice ~ . - Order, data = contVar.dat)  
## Residuals:  
## Min 1Q Median 3Q Max   
## -209825 -19547 1119 20867 212241   
##   
## Residual standard error: 37580 on 2398 degrees of freedom  
## (509 observations deleted due to missingness)  
## Multiple R-squared: 0.7904, Adjusted R-squared: 0.789   
## F-statistic: 532.1 on 17 and 2398 DF, p-value: < 2.2e-16

# *Appendix C : Section 3.2 - Full regression model using continuous predictors with NA's replaced with corresponding Mean values*

#Replaced NAs by corresponding mean values for variables in blue

## Order LotFrontage LotArea ***MasVnrArea*** ***BsmtFinSF1***   
## 0 490 0 0 0   
## ***BsmtFinSF2*** **BsmtUnfSF** ***TotalBsmtSF*** X1stFlrSF X2ndFlrSF   
## 0 0 0 0 0   
## LowQualFinSF GrLivArea  *GarageArea* WoodDeckSF OpenPorchSF   
## 0 0 0 0 0   
## EnclosedPorch X3SsnPorch ScreenPorch PoolArea MiscVal   
## 0 0 0 0 0   
## SalePrice   
## 0

#Fitting linear Model  
lm\_withNoNull = lm(contVar.dat$SalePrice ~.-Order, data = contVar.dat[,-2])  
summary(lm\_withNoNull )

##   
## Call:  
## lm(formula = contVar.dat$SalePrice ~ . - Order, data = contVar.dat[,   
## -2])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -213378 -18984 1202 19791 218122   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.738e+04 2.703e+03 -10.127 < 2e-16 \*\*\*  
## LotArea 2.205e-01 9.511e-02 2.318 0.0205 \*   
## MasVnrArea 5.344e+01 4.535e+00 11.782 < 2e-16 \*\*\*  
## BsmtFinSF1 7.216e+01 2.977e+00 24.240 < 2e-16 \*\*\*  
## BsmtFinSF2 4.374e+01 4.889e+00 8.946 < 2e-16 \*\*\*  
## BsmtUnfSF 4.780e+01 2.862e+00 16.701 < 2e-16 \*\*\*  
## TotalBsmtSF NA NA NA NA   
## X1stFlrSF 6.418e+01 3.258e+00 19.700 < 2e-16 \*\*\*  
## X2ndFlrSF 6.899e+01 1.937e+00 35.616 < 2e-16 \*\*\*  
## LowQualFinSF -4.375e+00 1.506e+01 -0.291 0.7714   
## GrLivArea NA NA NA NA   
## GarageArea 7.734e+01 4.049e+00 19.103 < 2e-16 \*\*\*  
## WoodDeckSF 4.175e+01 5.957e+00 7.009 2.96e-12 \*\*\*  
## OpenPorchSF 6.505e+01 1.113e+01 5.846 5.61e-09 \*\*\*  
## EnclosedPorch -5.721e+01 1.116e+01 -5.128 3.13e-07 \*\*\*  
## X3SsnPorch 1.285e+01 2.761e+01 0.466 0.6416   
## ScreenPorch 2.664e+01 1.259e+01 2.115 0.0345 \*   
## PoolArea -1.480e+01 2.143e+01 -0.690 0.4900   
## MiscVal -1.725e+00 1.470e+00 -1.174 0.2405   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 37420 on 2908 degrees of freedom  
## Multiple R-squared: 0.7743, Adjusted R-squared: 0.7731   
## F-statistic: 623.6 on 16 and 2908 DF, p-value: < 2.2e-16

# D: Result of Principal Component Analysis

# excluded SalePrice: Response variable and X1stFlrSF, X1stFlrSF  
ames.pc = princomp(contVar.dat[, c(-1, -2, -21, -9, -10)], cor = TRUE)  
summary(ames.pc, loadings = TRUE)

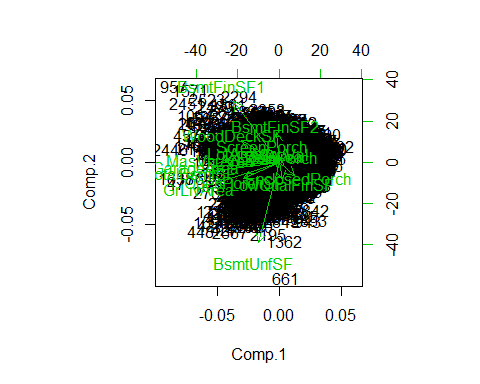
## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 1.7109750 1.2812616 1.1289493 1.03461167 1.01598803  
## Proportion of Variance 0.1829647 0.1026020 0.0796579 0.06690133 0.06451448  
## Cumulative Proportion 0.1829647 0.2855667 0.3652246 0.43212591 0.49664039  
## Comp.6 Comp.7 Comp.8 Comp.9  
## Standard deviation 1.00474754 0.98774499 0.97339698 0.96865569  
## Proportion of Variance 0.06309485 0.06097751 0.05921886 0.05864337  
## Cumulative Proportion 0.55973524 0.62071275 0.67993161 0.73857497  
**## *Rest of the result excluded***## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8  
## LotArea -0.231 -0.312 0.127 -0.235   
## MasVnrArea -0.358 0.172 0.110   
## BsmtFinSF1 -0.301 0.530 0.139 0.309 0.200   
## BsmtFinSF2 0.257 -0.371 -0.546 -0.447   
## BsmtUnfSF -0.134 -0.702 -0.132 -0.119   
## TotalBsmtSF -0.460   
## LowQualFinSF -0.152 -0.357 0.345 -0.334 0.683  
## GrLivArea -0.418 -0.189 -0.183   
## GarageArea -0.432   
## WoodDeckSF -0.245 0.190 -0.120 0.424 -0.305 0.125 0.189  
## OpenPorchSF -0.235 -0.145 -0.272 0.103 0.133 -0.267  
## EnclosedPorch -0.111 -0.483 0.248 -0.146 0.112 -0.191  
## X3SsnPorch 0.143 0.351 0.460 -0.128 -0.739 -0.145  
## ScreenPorch 0.114 -0.761 -0.247   
## PoolArea -0.500 0.149 -0.541  
## MiscVal -0.125 0.260 0.883 0.236  
## Comp.9   
## LotArea -0.291   
## MasVnrArea -0.109

## BsmtFinSF1 -0.133   
## BsmtFinSF2   
## BsmtUnfSF   
## TotalBsmtSF -0.199   
## LowQualFinSF 0.276   
## GrLivArea   
## GarageArea   
## WoodDeckSF 0.312

## OpenPorchSF 0.315   
## EnclosedPorch -0.608   
## X3SsnPorch 0.174   
## ScreenPorch   
## PoolArea 0.402   
## MiscVal

# Two dimensional view of the Data using first two principal components  
library(lattice)  
screeplot(ames.pc, npcs = 10, type = "lines", main = "Scree Plot of Principal Components" )

biplot (ames.pc , scale =1, col = c(1,3))

***Fig D : Biplot of First and Second principal components***

# Combining Saleprice and Perform Linear regression modelling

PCA\_SalePrice\_dat = cbind(amesTranformed.dat, SalePrice= contVar.dat[, 21])  
dim(PCA\_SalePrice\_dat)

## [1] 2925 10

# Handling Categorical Variables

categorical\_dat = workingSet2[ , sapply(workingSet2, is.factor)]  
names(categorical\_dat)

## [1] "MSZoning" "Street" "LotShape" "LandContour"   
## [5] "Utilities" "LotConfig" "LandSlope" "Neighborhood"   
## [9] "Condition1" "Condition2" "BldgType" "HouseStyle"   
## [13] "RoofStyle" "RoofMatl" "Exterior1st" "Exterior2nd"   
## [17] "MasVnrType" "ExterQual" "ExterCond" "Foundation"   
## [21] "Heating" "HeatingQC" "CentralAir" "Electrical"   
## [25] "KitchenQual" "Functional" "PavedDrive" "SaleType"   
## [29] "SaleCondition"

for (i in c(1:29)) {   
 categorical\_dat[,i] <- as.factor(categorical\_dat[,i])   
}

# Combine Continous and categorical variables

finalWorkingSet = cbind(PCA\_SalePrice\_dat, categorical\_dat)  
dim(finalWorkingSet)

## [1] 2925 39

# *Appendix E:*

# 3.4 Variable Selection

[1] "Comp.1" "Comp.2" "Comp.3" "Comp.4" "Comp.5" "Comp.6" "Comp.7"

[8] "Comp.8" "Comp.9" "SalePrice" "MSZoning" "Street" "LotShape" "LandContour"

[15] "Utilities" "LotConfig" "LandSlope" "Neighborhood" "Condition1" "Condition2" "BldgType"

[22] "HouseStyle" "RoofStyle" "RoofMatl" "Exterior1st" "Exterior2nd" "MasVnrType" "ExterQual"

[29] "ExterCond" "Foundation" "Heating" "HeatingQC" "CentralAir" "Electrical" "KitchenQual"

[36] "Functional" "PavedDrive" "SaleType" "SaleCondition"

Full model : lmFinal.full = lm(SalePrice ~ ., data = finalWorkingSet)  
Null Model : lmFinal.null = lm(SalePrice ~ 1, data = finalWorkingSet)

# 

Stepwise Forward Selection

step(lmFinal.null, scope = list(lower = lmFinal.null, upper = lmFinal.full ), direction = "forward")

**Best subset Model :**

forward = lm(formula = SalePrice ~ Comp.1 + Neighborhood + KitchenQual +   
 ExterQual + HouseStyle + BldgType + MasVnrType + SaleCondition +   
 Exterior1st + Functional + Foundation + Comp.8 + Comp.2 +   
 Condition1 + HeatingQC + LandSlope + RoofMatl + LandContour +   
 Street + Comp.5 + Comp.3 + LotConfig + ExterCond + Condition2 +   
 RoofStyle + Comp.7 + Utilities + Comp.4 + SaleType, data = finalWorkingSet)

summary(forward)

##   
## Call:  
## lm(formula = SalePrice ~ Comp.1 + Neighborhood + KitchenQual +   
## ExterQual + HouseStyle + BldgType + MasVnrType + SaleCondition +   
## Exterior1st + Functional + Foundation + Comp.8 + Comp.2 +   
## Condition1 + HeatingQC + LandSlope + RoofMatl + LandContour +   
## Street + Comp.5 + Comp.3 + LotConfig + ExterCond + Condition2 +   
## RoofStyle + Comp.7 + Utilities + Comp.4 + SaleType, data = finalWorkingSet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -157897 -13437 -364 12565 165097   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 202783.9 23454.1 8.646 < 2e-16 \*\*\*  
## Comp.1 -25578.5 494.1 -51.763 < 2e-16 \*\*\*  
## NeighborhoodBlueste -11107.4 9885.3 -1.124 0.261268   
## NeighborhoodBrDale -30134.5 7690.3 -3.919 9.12e-05 \*\*\*  
## NeighborhoodBrkSide -34626.7 6344.9 -5.457 5.26e-08 \*\*\*  
## NeighborhoodClearCr -28772.3 7083.6 -4.062 5.00e-05 \*\*\*  
## NeighborhoodCollgCr -31927.7 5472.1 -5.835 6.02e-09 \*\*\*  
## NeighborhoodCrawfor -13165.4 6182.4 -2.129 0.033301 \*   
## NeighborhoodEdwards -41721.1 5856.7 -7.124 1.33e-12 \*\*\*  
## NeighborhoodGilbert -26441.6 5778.5 -4.576 4.95e-06 \*\*\*  
## NeighborhoodGreens 11443.4 10743.7 1.065 0.286912   
## NeighborhoodGrnHill 107632.2 18634.8 5.776 8.50e-09 \*\*\*  
## NeighborhoodIDOTRR -45964.9 6388.8 -7.195 8.02e-13 \*\*\*  
## NeighborhoodLandmrk -21763.6 26022.3 -0.836 0.403033   
## NeighborhoodMeadowV -39899.5 7689.1 -5.189 2.27e-07 \*\*\*  
## NeighborhoodMitchel -38986.6 6018.2 -6.478 1.09e-10 \*\*\*  
## NeighborhoodNAmes -41174.3 5745.1 -7.167 9.79e-13 \*\*\*  
## NeighborhoodNoRidge 7603.4 6267.1 1.213 0.225151   
## NeighborhoodNPkVill -2955.1 8004.1 -0.369 0.712008   
## NeighborhoodNridgHt 1198.8 5640.5 0.213 0.831708   
## NeighborhoodNWAmes -35357.6 6038.9 -5.855 5.33e-09 \*\*\*  
## NeighborhoodOldTown -48870.5 5968.0 -8.189 3.97e-16 \*\*\*  
## NeighborhoodSawyer -36297.9 6025.0 -6.025 1.92e-09 \*\*\*  
## NeighborhoodSawyerW -32221.6 5811.2 -5.545 3.22e-08 \*\*\*  
## NeighborhoodSomerst -16883.3 5495.1 -3.072 0.002144 \*\*   
## NeighborhoodStoneBr 28762.7 6371.8 4.514 6.63e-06 \*\*\*  
## NeighborhoodSWISU -42938.5 6889.7 -6.232 5.29e-10 \*\*\*  
## NeighborhoodTimber -22292.7 6156.0 -3.621 0.000298 \*\*\*  
## NeighborhoodVeenker -22581.3 7680.9 -2.940 0.003310 \*\*   
## KitchenQualFa -40466.6 4351.9 -9.299 < 2e-16 \*\*\*  
## KitchenQualGd -32202.4 2564.8 -12.556 < 2e-16 \*\*\*  
## KitchenQualPo -21010.1 26555.6 -0.791 0.428909   
## KitchenQualTA -40336.8 2834.4 -14.231 < 2e-16 \*\*\*  
## ExterQualFa -37108.4 6589.0 -5.632 1.96e-08 \*\*\*  
## ExterQualGd -33058.3 3507.5 -9.425 < 2e-16 \*\*\*  
## ExterQualTA -42929.6 3902.9 -10.999 < 2e-16 \*\*\*  
## HouseStyle1.5Unf -2218.7 6193.1 -0.358 0.720183   
## HouseStyle1Story -9155.7 1923.8 -4.759 2.04e-06 \*\*\*  
## HouseStyle2.5Fin 4566.6 10168.8 0.449 0.653409   
## HouseStyle2.5Unf 12665.6 5605.4 2.260 0.023926 \*   
## HouseStyle2Story 7073.8 2026.6 3.490 0.000490 \*\*\*  
## HouseStyleSFoyer 639.9 3543.0 0.181 0.856685   
## HouseStyleSLvl 3064.0 2999.7 1.021 0.307140   
## BldgType2fmCon -12037.5 3531.1 -3.409 0.000661 \*\*\*  
## BldgTypeDuplex -10678.3 2920.1 -3.657 0.000260 \*\*\*  
## BldgTypeTwnhs -28838.3 3771.6 -7.646 2.83e-14 \*\*\*  
## BldgTypeTwnhsE -24371.2 2363.5 -10.312 < 2e-16 \*\*\*  
## MasVnrTypeBrkCmn -11089.9 7543.8 -1.470 0.141655   
## MasVnrTypeBrkFace -6486.6 5516.7 -1.176 0.239772   
## MasVnrTypeCBlock -152145.8 30971.7 -4.912 9.52e-07 \*\*\*  
## MasVnrTypeNone -417.4 5466.1 -0.076 0.939137   
## MasVnrTypeStone 3036.5 5628.8 0.539 0.589616   
## SaleConditionAdjLand 21882.7 8110.3 2.698 0.007015 \*\*   
## SaleConditionAlloca 16113.3 6068.8 2.655 0.007973 \*\*   
## SaleConditionFamily 3596.3 4283.6 0.840 0.401241   
## SaleConditionNormal 10998.9 2137.8 5.145 2.86e-07 \*\*\*  
## SaleConditionPartial 19036.4 10991.5 1.732 0.083400 .   
## Exterior1stAsphShn 22772.6 18664.4 1.220 0.222527   
## Exterior1stBrkComm 17048.1 11492.8 1.483 0.138088   
## Exterior1stBrkFace 30427.6 5079.2 5.991 2.36e-09 \*\*\*  
## Exterior1stCBlock 14107.6 18894.4 0.747 0.455335   
## Exterior1stCemntBd 15582.9 5199.7 2.997 0.002751 \*\*   
## Exterior1stHdBoard 4979.4 4439.5 1.122 0.262125   
## Exterior1stImStucc -2906.7 25880.7 -0.112 0.910583   
## Exterior1stMetalSd 9857.4 4297.0 2.294 0.021864 \*   
## Exterior1stPlywood 7493.3 4709.9 1.591 0.111730   
## Exterior1stPreCast 102340.3 26790.4 3.820 0.000136 \*\*\*  
## Exterior1stStone 44205.8 19267.9 2.294 0.021849 \*   
## Exterior1stStucco 14638.7 5683.1 2.576 0.010051 \*   
## Exterior1stVinylSd 10782.3 4367.1 2.469 0.013610 \*   
## Exterior1stWd Sdng 7690.9 4287.7 1.794 0.072968 .   
## Exterior1stWdShing 12314.5 5363.1 2.296 0.021741 \*   
## FunctionalMaj2 -359.5 10635.4 -0.034 0.973038   
## FunctionalMin1 10457.9 6917.7 1.512 0.130712   
## FunctionalMin2 5632.4 6891.9 0.817 0.413855   
## FunctionalMod -3251.3 7605.8 -0.427 0.669064   
## FunctionalSal -30633.2 22354.8 -1.370 0.170697   
## FunctionalSev -36597.9 19624.6 -1.865 0.062301 .   
## FunctionalTyp 18152.7 6148.4 2.952 0.003179 \*\*   
## FoundationCBlock 672.1 2097.9 0.320 0.748716   
## FoundationPConc 6589.4 2291.7 2.875 0.004068 \*\*   
## FoundationSlab 21847.6 4471.0 4.887 1.08e-06 \*\*\*  
## FoundationStone 9671.0 8003.1 1.208 0.226991   
## FoundationWood 685.9 11711.4 0.059 0.953304   
## Comp.8 859.4 534.2 1.609 0.107748   
## Comp.2 2666.8 433.3 6.155 8.58e-10 \*\*\*  
## Condition1Feedr 1501.4 3581.7 0.419 0.675120   
## Condition1Norm 10683.6 2946.4 3.626 0.000293 \*\*\*  
## Condition1PosA 22664.4 6870.4 3.299 0.000983 \*\*\*  
## Condition1PosN 13933.0 5303.8 2.627 0.008662 \*\*   
## Condition1RRAe -3634.1 5945.0 -0.611 0.541062   
## Condition1RRAn 5967.8 4932.7 1.210 0.226437   
## Condition1RRNe 7164.9 11083.0 0.646 0.518028   
## Condition1RRNn 451.4 9192.0 0.049 0.960833   
## HeatingQCFa -13365.6 3023.5 -4.421 1.02e-05 \*\*\*  
## HeatingQCGd -2912.0 1522.4 -1.913 0.055878 .   
## HeatingQCPo -9633.0 17046.2 -0.565 0.572045   
## HeatingQCTA -7149.3 1460.4 -4.896 1.04e-06 \*\*\*  
## LandSlopeMod 8365.3 2897.5 2.887 0.003919 \*\*   
## LandSlopeSev -35064.7 8015.9 -4.374 1.26e-05 \*\*\*  
## RoofMatlMembran 121691.1 28993.1 4.197 2.79e-05 \*\*\*  
## RoofMatlMetal 36206.1 28869.6 1.254 0.209901   
## RoofMatlRoll -18330.0 25967.0 -0.706 0.480313   
## RoofMatlTar&Grv 9021.0 9876.4 0.913 0.361118   
## RoofMatlWdShake 1300.4 9619.8 0.135 0.892476   
## RoofMatlWdShngl 48186.5 11095.7 4.343 1.46e-05 \*\*\*  
## LandContourHLS 8931.5 3663.7 2.438 0.014837 \*   
## LandContourLow -14095.1 4701.3 -2.998 0.002741 \*\*   
## LandContourLvl -780.4 2714.3 -0.288 0.773737   
## StreetPave 31309.7 8247.2 3.796 0.000150 \*\*\*  
## Comp.5 2090.0 492.0 4.248 2.23e-05 \*\*\*  
## Comp.3 -2330.1 487.3 -4.782 1.83e-06 \*\*\*  
## LotConfigCulDSac 10358.3 2357.8 4.393 1.16e-05 \*\*\*  
## LotConfigFR2 -1656.5 3119.3 -0.531 0.595420   
## LotConfigFR3 5998.8 6982.5 0.859 0.390354   
## LotConfigInside 1663.3 1303.2 1.276 0.201960   
## ExterCondFa -14205.4 8432.6 -1.685 0.092182 .   
## ExterCondGd 727.4 7711.9 0.094 0.924866   
## ExterCondPo -18783.3 18381.0 -1.022 0.306923   
## ExterCondTA -2274.3 7620.2 -0.298 0.765375   
## Condition2Feedr -10204.1 13758.2 -0.742 0.458347   
## Condition2Norm 131.2 11868.2 0.011 0.991182   
## Condition2PosA 55691.1 18368.1 3.032 0.002452 \*\*   
## Condition2PosN -8982.6 19491.3 -0.461 0.644943   
## Condition2RRAe -75093.9 33017.1 -2.274 0.023018 \*   
## Condition2RRAn 2973.3 28379.4 0.105 0.916567   
## Condition2RRNn 164.2 21726.8 0.008 0.993970   
## RoofStyleGable 4335.1 11331.7 0.383 0.702073   
## RoofStyleGambrel 324.4 12632.1 0.026 0.979513   
## RoofStyleHip 7980.3 11416.8 0.699 0.484611   
## RoofStyleMansard 269.7 14038.5 0.019 0.984672   
## RoofStyleShed 61562.4 19603.5 3.140 0.001705 \*\*   
## Comp.7 300.2 507.7 0.591 0.554333   
## UtilitiesNoSeWa -69687.6 25974.3 -2.683 0.007341 \*\*   
## UtilitiesNoSewr -20237.8 18872.6 -1.072 0.283663   
## Comp.4 -1331.0 470.3 -2.830 0.004685 \*\*   
## SaleTypeCon 41413.6 12018.9 3.446 0.000578 \*\*\*  
## SaleTypeConLD 4790.3 6021.1 0.796 0.426345   
## SaleTypeConLI -7332.5 9058.4 -0.809 0.418312   
## SaleTypeConLw 2787.5 9627.0 0.290 0.772184   
## SaleTypeCWD 21138.3 7964.1 2.654 0.007995 \*\*   
## SaleTypeNew 8211.0 11379.5 0.722 0.470622   
## SaleTypeOth 12202.0 10055.9 1.213 0.225074   
## SaleTypeVWD 13968.1 25768.2 0.542 0.587817   
## SaleTypeWD 3454.1 3069.7 1.125 0.260584   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 25230 on 2780 degrees of freedom  
## Multiple R-squared: 0.902, Adjusted R-squared: 0.8969   
## F-statistic: 177.6 on 144 and 2780 DF, p-value: < 2.2e-16

# *Appendix F*

# 3.6 Stepwise Backward Selection

step(lmFinal.full, data = finalWorkingSet, direction = "backward")

backward = lm(formula = SalePrice ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 +   
 Comp.5 + Comp.7 + Comp.8 + Street + LandContour + Utilities +   
 LotConfig + LandSlope + Neighborhood + Condition1 + Condition2 +   
 BldgType + HouseStyle + RoofStyle + RoofMatl + Exterior1st +   
 MasVnrType + ExterQual + ExterCond + Foundation + HeatingQC +   
 KitchenQual + Functional + SaleType + SaleCondition, data = finalWorkingSet)  
summary(backward)

##   
## Call:  
## lm(formula = SalePrice ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 +   
## Comp.5 + Comp.7 + Comp.8 + Street + LandContour + Utilities +   
## LotConfig + LandSlope + Neighborhood + Condition1 + Condition2 +   
## BldgType + HouseStyle + RoofStyle + RoofMatl + Exterior1st +   
## MasVnrType + ExterQual + ExterCond + Foundation + HeatingQC +   
## KitchenQual + Functional + SaleType + SaleCondition, data = finalWorkingSet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -157897 -13437 -364 12565 165097   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 202783.9 23454.1 8.646 < 2e-16 \*\*\*  
## Comp.1 -25578.5 494.1 -51.763 < 2e-16 \*\*\*  
## Comp.2 2666.8 433.3 6.155 8.58e-10 \*\*\*  
## Comp.3 -2330.1 487.3 -4.782 1.83e-06 \*\*\*  
## Comp.4 -1331.0 470.3 -2.830 0.004685 \*\*   
## Comp.5 2090.0 492.0 4.248 2.23e-05 \*\*\*  
## Comp.7 300.2 507.7 0.591 0.554333   
## Comp.8 859.4 534.2 1.609 0.107748   
## StreetPave 31309.7 8247.2 3.796 0.000150 \*\*\*  
## LandContourHLS 8931.5 3663.7 2.438 0.014837 \*   
## LandContourLow -14095.1 4701.3 -2.998 0.002741 \*\*   
## LandContourLvl -780.4 2714.3 -0.288 0.773737   
## UtilitiesNoSeWa -69687.6 25974.3 -2.683 0.007341 \*\*   
## UtilitiesNoSewr -20237.8 18872.6 -1.072 0.283663   
## LotConfigCulDSac 10358.3 2357.8 4.393 1.16e-05 \*\*\*  
## LotConfigFR2 -1656.5 3119.3 -0.531 0.595420   
## LotConfigFR3 5998.8 6982.5 0.859 0.390354   
## LotConfigInside 1663.3 1303.2 1.276 0.201960   
## LandSlopeMod 8365.3 2897.5 2.887 0.003919 \*\*   
## LandSlopeSev -35064.7 8015.9 -4.374 1.26e-05 \*\*\*  
## NeighborhoodBlueste -11107.4 9885.3 -1.124 0.261268   
## NeighborhoodBrDale -30134.5 7690.3 -3.919 9.12e-05 \*\*\*  
## NeighborhoodBrkSide -34626.7 6344.9 -5.457 5.26e-08 \*\*\*  
## NeighborhoodClearCr -28772.3 7083.6 -4.062 5.00e-05 \*\*\*  
## NeighborhoodCollgCr -31927.7 5472.1 -5.835 6.02e-09 \*\*\*  
## NeighborhoodCrawfor -13165.4 6182.4 -2.129 0.033301 \*   
## NeighborhoodEdwards -41721.1 5856.7 -7.124 1.33e-12 \*\*\*  
## NeighborhoodGilbert -26441.6 5778.5 -4.576 4.95e-06 \*\*\*  
## NeighborhoodGreens 11443.4 10743.7 1.065 0.286912   
## NeighborhoodGrnHill 107632.2 18634.8 5.776 8.50e-09 \*\*\*  
## NeighborhoodIDOTRR -45964.9 6388.8 -7.195 8.02e-13 \*\*\*  
## NeighborhoodLandmrk -21763.6 26022.3 -0.836 0.403033   
## NeighborhoodMeadowV -39899.5 7689.1 -5.189 2.27e-07 \*\*\*  
## NeighborhoodMitchel -38986.6 6018.2 -6.478 1.09e-10 \*\*\*  
## NeighborhoodNAmes -41174.3 5745.1 -7.167 9.79e-13 \*\*\*  
## NeighborhoodNoRidge 7603.4 6267.1 1.213 0.225151   
## NeighborhoodNPkVill -2955.1 8004.1 -0.369 0.712008   
## NeighborhoodNridgHt 1198.8 5640.5 0.213 0.831708   
## NeighborhoodNWAmes -35357.6 6038.9 -5.855 5.33e-09 \*\*\*  
## NeighborhoodOldTown -48870.5 5968.0 -8.189 3.97e-16 \*\*\*  
## NeighborhoodSawyer -36297.9 6025.0 -6.025 1.92e-09 \*\*\*  
## NeighborhoodSawyerW -32221.6 5811.2 -5.545 3.22e-08 \*\*\*  
## NeighborhoodSomerst -16883.3 5495.1 -3.072 0.002144 \*\*   
## NeighborhoodStoneBr 28762.7 6371.8 4.514 6.63e-06 \*\*\*  
## NeighborhoodSWISU -42938.5 6889.7 -6.232 5.29e-10 \*\*\*  
## NeighborhoodTimber -22292.7 6156.0 -3.621 0.000298 \*\*\*  
## NeighborhoodVeenker -22581.3 7680.9 -2.940 0.003310 \*\*   
## Condition1Feedr 1501.4 3581.7 0.419 0.675120   
## Condition1Norm 10683.6 2946.4 3.626 0.000293 \*\*\*  
## Condition1PosA 22664.4 6870.4 3.299 0.000983 \*\*\*  
## Condition1PosN 13933.0 5303.8 2.627 0.008662 \*\*   
## Condition1RRAe -3634.1 5945.0 -0.611 0.541062   
## Condition1RRAn 5967.8 4932.7 1.210 0.226437   
## Condition1RRNe 7164.9 11083.0 0.646 0.518028   
## Condition1RRNn 451.4 9192.0 0.049 0.960833   
## Condition2Feedr -10204.1 13758.2 -0.742 0.458347   
## Condition2Norm 131.2 11868.2 0.011 0.991182   
## Condition2PosA 55691.1 18368.1 3.032 0.002452 \*\*   
## Condition2PosN -8982.6 19491.3 -0.461 0.644943   
## Condition2RRAe -75093.9 33017.1 -2.274 0.023018 \*   
## Condition2RRAn 2973.3 28379.4 0.105 0.916567   
## Condition2RRNn 164.2 21726.8 0.008 0.993970   
## BldgType2fmCon -12037.5 3531.1 -3.409 0.000661 \*\*\*  
## BldgTypeDuplex -10678.3 2920.1 -3.657 0.000260 \*\*\*  
## BldgTypeTwnhs -28838.3 3771.6 -7.646 2.83e-14 \*\*\*  
## BldgTypeTwnhsE -24371.2 2363.5 -10.312 < 2e-16 \*\*\*  
## HouseStyle1.5Unf -2218.7 6193.1 -0.358 0.720183   
## HouseStyle1Story -9155.7 1923.8 -4.759 2.04e-06 \*\*\*  
## HouseStyle2.5Fin 4566.6 10168.8 0.449 0.653409   
## HouseStyle2.5Unf 12665.6 5605.4 2.260 0.023926 \*   
## HouseStyle2Story 7073.8 2026.6 3.490 0.000490 \*\*\*  
## HouseStyleSFoyer 639.9 3543.0 0.181 0.856685   
## HouseStyleSLvl 3064.0 2999.7 1.021 0.307140   
## RoofStyleGable 4335.1 11331.7 0.383 0.702073   
## RoofStyleGambrel 324.4 12632.1 0.026 0.979513   
## RoofStyleHip 7980.3 11416.8 0.699 0.484611   
## RoofStyleMansard 269.7 14038.5 0.019 0.984672   
## RoofStyleShed 61562.4 19603.5 3.140 0.001705 \*\*   
## RoofMatlMembran 121691.1 28993.1 4.197 2.79e-05 \*\*\*  
## RoofMatlMetal 36206.1 28869.6 1.254 0.209901   
## RoofMatlRoll -18330.0 25967.0 -0.706 0.480313   
## RoofMatlTar&Grv 9021.0 9876.4 0.913 0.361118   
## RoofMatlWdShake 1300.4 9619.8 0.135 0.892476   
## RoofMatlWdShngl 48186.5 11095.7 4.343 1.46e-05 \*\*\*  
## Exterior1stAsphShn 22772.6 18664.4 1.220 0.222527   
## Exterior1stBrkComm 17048.1 11492.8 1.483 0.138088   
## Exterior1stBrkFace 30427.6 5079.2 5.991 2.36e-09 \*\*\*  
## Exterior1stCBlock 14107.6 18894.4 0.747 0.455335   
## Exterior1stCemntBd 15582.9 5199.7 2.997 0.002751 \*\*   
## Exterior1stHdBoard 4979.4 4439.5 1.122 0.262125   
## Exterior1stImStucc -2906.7 25880.7 -0.112 0.910583   
## Exterior1stMetalSd 9857.4 4297.0 2.294 0.021864 \*   
## Exterior1stPlywood 7493.3 4709.9 1.591 0.111730   
## Exterior1stPreCast 102340.3 26790.4 3.820 0.000136 \*\*\*  
## Exterior1stStone 44205.8 19267.9 2.294 0.021849 \*   
## Exterior1stStucco 14638.7 5683.1 2.576 0.010051 \*   
## Exterior1stVinylSd 10782.3 4367.1 2.469 0.013610 \*   
## Exterior1stWd Sdng 7690.9 4287.7 1.794 0.072968 .   
## Exterior1stWdShing 12314.5 5363.1 2.296 0.021741 \*   
## MasVnrTypeBrkCmn -11089.9 7543.8 -1.470 0.141655   
## MasVnrTypeBrkFace -6486.6 5516.7 -1.176 0.239772   
## MasVnrTypeCBlock -152145.8 30971.7 -4.912 9.52e-07 \*\*\*  
## MasVnrTypeNone -417.4 5466.1 -0.076 0.939137   
## MasVnrTypeStone 3036.5 5628.8 0.539 0.589616   
## ExterQualFa -37108.4 6589.0 -5.632 1.96e-08 \*\*\*  
## ExterQualGd -33058.3 3507.5 -9.425 < 2e-16 \*\*\*  
## ExterQualTA -42929.6 3902.9 -10.999 < 2e-16 \*\*\*  
## ExterCondFa -14205.4 8432.6 -1.685 0.092182 .   
## ExterCondGd 727.4 7711.9 0.094 0.924866   
## ExterCondPo -18783.3 18381.0 -1.022 0.306923   
## ExterCondTA -2274.3 7620.2 -0.298 0.765375   
## FoundationCBlock 672.1 2097.9 0.320 0.748716   
## FoundationPConc 6589.4 2291.7 2.875 0.004068 \*\*   
## FoundationSlab 21847.6 4471.0 4.887 1.08e-06 \*\*\*  
## FoundationStone 9671.0 8003.1 1.208 0.226991   
## FoundationWood 685.9 11711.4 0.059 0.953304   
## HeatingQCFa -13365.6 3023.5 -4.421 1.02e-05 \*\*\*  
## HeatingQCGd -2912.0 1522.4 -1.913 0.055878 .   
## HeatingQCPo -9633.0 17046.2 -0.565 0.572045   
## HeatingQCTA -7149.3 1460.4 -4.896 1.04e-06 \*\*\*  
## KitchenQualFa -40466.6 4351.9 -9.299 < 2e-16 \*\*\*  
## KitchenQualGd -32202.4 2564.8 -12.556 < 2e-16 \*\*\*  
## KitchenQualPo -21010.1 26555.6 -0.791 0.428909   
## KitchenQualTA -40336.8 2834.4 -14.231 < 2e-16 \*\*\*  
## FunctionalMaj2 -359.5 10635.4 -0.034 0.973038   
## FunctionalMin1 10457.9 6917.7 1.512 0.130712   
## FunctionalMin2 5632.4 6891.9 0.817 0.413855   
## FunctionalMod -3251.3 7605.8 -0.427 0.669064   
## FunctionalSal -30633.2 22354.8 -1.370 0.170697   
## FunctionalSev -36597.9 19624.6 -1.865 0.062301 .   
## FunctionalTyp 18152.7 6148.4 2.952 0.003179 \*\*   
## SaleTypeCon 41413.6 12018.9 3.446 0.000578 \*\*\*  
## SaleTypeConLD 4790.3 6021.1 0.796 0.426345   
## SaleTypeConLI -7332.5 9058.4 -0.809 0.418312   
## SaleTypeConLw 2787.5 9627.0 0.290 0.772184   
## SaleTypeCWD 21138.3 7964.1 2.654 0.007995 \*\*   
## SaleTypeNew 8211.0 11379.5 0.722 0.470622   
## SaleTypeOth 12202.0 10055.9 1.213 0.225074   
## SaleTypeVWD 13968.1 25768.2 0.542 0.587817   
## SaleTypeWD 3454.1 3069.7 1.125 0.260584   
## SaleConditionAdjLand 21882.7 8110.3 2.698 0.007015 \*\*   
## SaleConditionAlloca 16113.3 6068.8 2.655 0.007973 \*\*   
## SaleConditionFamily 3596.3 4283.6 0.840 0.401241   
## SaleConditionNormal 10998.9 2137.8 5.145 2.86e-07 \*\*\*  
## SaleConditionPartial 19036.4 10991.5 1.732 0.083400 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 25230 on 2780 degrees of freedom  
## Multiple R-squared: 0.902, Adjusted R-squared: 0.8969   
## F-statistic: 177.6 on 144 and 2780 DF, p-value: < 2.2e-16

# *Appendix :G*

# 3.7 Model Improvement By removing insignificant variables

forward1 = update(forward, ~.-LotConfig)  
summary(forward1)

## Residual standard error: 25310 on 2784 degrees of freedom  
## Multiple R-squared: 0.9012, Adjusted R-squared: 0.8962   
## F-statistic: 181.3 on 140 and 2784 DF, p-value: < 2.2e-16

forward2= update(forward1, ~. -RoofStyle)  
summary(forward2)

##   
## Residual standard error: 25370 on 2789 degrees of freedom  
## Multiple R-squared: 0.9005, Adjusted R-squared: 0.8957   
## F-statistic: 186.9 on 135 and 2789 DF, p-value: < 2.2e-16

forward3= update(forward2, ~. -ExterCond)  
summary(forward3)

##   
## Residual standard error: 25430 on 2793 degrees of freedom  
## Multiple R-squared: 0.8999, Adjusted R-squared: 0.8952   
## F-statistic: 191.7 on 131 and 2793 DF, p-value: < 2.2e-16

forward4= update(forward3, ~. -Condition2)  
summary(forward4)

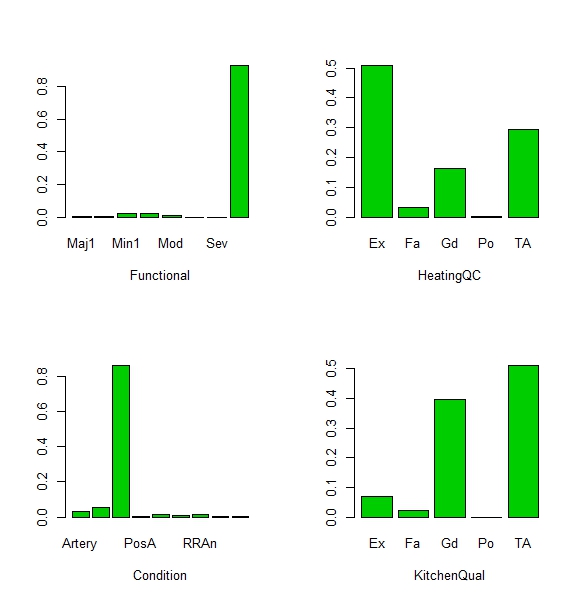
## Residual standard error: 25480 on 2800 degrees of freedom  
## Multiple R-squared: 0.8992, Adjusted R-squared: 0.8948   
## F-statistic: 201.5 on 124 and 2800 DF, p-value: < 2.2e-16

forward5= update(forward4, ~. -Exterior1st)  
summary(forward5)

## Residual standard error: 25830 on 2815 degrees of freedom  
## Multiple R-squared: 0.8959, Adjusted R-squared: 0.8919   
## F-statistic: 222.3 on 109 and 2815 DF, p-value: < 2.2e-16

# *Appendix H:*

Bar Plot of for selected set of categorical predictor variable as shown below:



# Combining the Categorical Variable levels

#Comining variable levels  
WS\_combLevel = finalWorkingSet  
levels(WS\_combLevel$Functional)<-c("NonTyp", "NonTyp", "NonTyp", "NonTyp", "NonTyp" , "NonTyp" , "NonTyp", "Typ")  
levels(WS\_combLevel$HeatingQC)<-c("Gd", "Po", "Gd", "Po", "Gd")  
levels(WS\_combLevel$Condition1)= c("AbNorm", "AbNorm" , "Norm", "AbNorm" , "AbNorm", "AbNorm", "AbNorm", "AbNorm", "AbNorm")  
levels(WS\_combLevel$KitchenQual) = c("Gd", "Po", "Gd", "Po", "Gd")  
par(mfrow = c(1,1))

***Appendix I:***

Final Model :

finalmodel = lm(formula = SalePrice ~ Comp.1 + Neighborhood + KitchenQual +   
 ExterQual + HouseStyle + BldgType + MasVnrType + SaleCondition +   
 Functional + Foundation + Comp.8 + Comp.2 + Condition1 +   
 HeatingQC + LandSlope + RoofMatl + LandContour + Street +   
 Comp.5 + Comp.3 + Comp.7 + Utilities + Comp.4 ,   
 data = WS\_combLevel)  
summary(finalmodel)

Appendix J:

#confint(finalmodel)

2.5 % 97.5 %

(Intercept) 182851.2857 225999.9542

Comp.1 -30346.6980 -28819.8951

NeighborhoodBlueste -31560.5476 2286.7231

NeighborhoodBrDale -41979.3515 -15928.9029

NeighborhoodBrkSide -39887.8499 -18591.0149

NeighborhoodClearCr -41168.4033 -17349.8359

NeighborhoodCollgCr -38231.8795 -19437.8399

NeighborhoodCrawfor -22020.0263 -1145.4013

NeighborhoodEdwards -43763.9221 -23841.9010

NeighborhoodGilbert -35023.6897 -15307.3974

NeighborhoodGreens -18786.9995 17554.0386

NeighborhoodGrnHill 61880.5227 125943.4282

NeighborhoodIDOTRR -50109.3737 -28481.3528

NeighborhoodLandmrk -68386.5131 21559.2437

NeighborhoodMeadowV -41779.4315 -17662.9635

NeighborhoodMitchel -48269.2556 -27818.0230

NeighborhoodNAmes -45361.8723 -25950.0874

NeighborhoodNoRidge -17888.8870 3494.3978

NeighborhoodNPkVill -27501.3974 -946.7607

NeighborhoodNridgHt -13019.3598 6266.8555

NeighborhoodNWAmes -44564.3181 -24250.9156

NeighborhoodOldTown -50913.3408 -30762.5693

NeighborhoodSawyer -43787.9970 -23415.9885

NeighborhoodSawyerW -41403.4533 -21667.9486

NeighborhoodSomerst -26941.0746 -8208.3137

NeighborhoodStoneBr 8425.5973 30129.0274

NeighborhoodSWISU -47985.3177 -24553.1885

NeighborhoodTimber -34497.9068 -13384.9365

NeighborhoodVeenker -34565.0164 -8899.9223

KitchenQualPo -5467.7554 5724.6881

ExterQualFa -63223.2194 -42736.5945

ExterQualGd -46245.0979 -35629.1277

ExterQualTA -59412.7966 -47486.4567

HouseStyle1.5Unf -10594.8021 10452.3233

HouseStyle1Story -9014.7803 -2582.6517

HouseStyle2.5Fin -8401.5527 27343.3718

HouseStyle2.5Unf -2104.4749 16949.7559

HouseStyle2Story 4984.4220 11789.5080

HouseStyleSFoyer -1692.2900 10337.7943

HouseStyleSLvl 1025.1453 11213.9796

BldgType2fmCon -22446.9040 -10630.5535

BldgTypeDuplex -20874.4069 -11187.3031

BldgTypeTwnhs -24667.5634 -11914.1169

BldgTypeTwnhsE -20016.0421 -12045.5596

MasVnrTypeBrkCmn -27869.8668 -2045.7592

MasVnrTypeBrkFace -19806.8468 -850.9340

MasVnrTypeCBlock -139866.0363 -49881.2420

MasVnrTypeNone -8571.4944 10245.0055

MasVnrTypeStone -11026.7073 8363.3518

SaleConditionAdjLand -421.8999 26427.7850

SaleConditionAlloca 1422.2729 21824.2167

SaleConditionFamily -4526.0586 9963.6633

SaleConditionNormal 6705.5807 13479.2520

SaleConditionPartial 17472.3803 26995.6299

FunctionalTyp 10103.3198 16926.7972

FoundationCBlock -3241.3625 3732.7590

FoundationPConc 3657.8328 11306.8546

FoundationSlab 20875.2857 35891.3065

FoundationStone -9765.0183 17549.8388

FoundationWood -21213.0575 19090.9813

Comp.8 2759.3551 4583.6490

Comp.2 1771.4262 3241.0990

Condition1Norm 2959.5543 7928.7409

HeatingQCPo -12841.0368 -3273.8355

LandSlopeMod -732.7108 9145.5899

LandSlopeSev -45787.7638 -19328.3697

RoofMatlMembran 51549.3196 142929.8103

RoofMatlMetal -9175.5236 82948.5255

RoofMatlRoll -70781.2701 17442.5519

RoofMatlTar&Grv -227.9124 19150.5599

RoofMatlWdShake -8662.8693 21011.2780

RoofMatlWdShngl 24330.9820 61653.0452

LandContourHLS 802.8874 13301.2070

LandContourLow -23207.5293 -7207.7699

LandContourLvl -6590.5451 2634.8431

StreetPave 13045.1092 40918.9112

Comp.5 -2756.6027 -1063.5845

Comp.3 -2413.7479 -762.5051

Comp.7 343.9032 2158.2173

UtilitiesNoSeWa -94543.3783 -5706.0471

UtilitiesNoSewr -44779.1323 19203.6890

Comp.4 173.0461 1783.5262

***Appendix K: Variation inflation factor***

> vif(finalmodel)

GVIF Df GVIF^(1/(2\*Df))

Comp.1 3.332651 1 1.825555

Neighborhood 140.980143 27 1.095971

KitchenQual 1.139315 1 1.067387

ExterQual 4.627359 3 1.290889

HouseStyle 3.908208 7 1.102260

BldgType 4.480834 4 1.206201

MasVnrType 2.307916 5 1.087231

SaleCondition 1.891901 5 1.065835

Functional 1.149456 1 1.072127

Foundation 5.460732 5 1.185018

Comp.8 1.199420 1 1.095180

Comp.2 1.360140 1 1.166251

Condition1 1.133461 1 1.064641

HeatingQC 1.104490 1 1.050947

LandSlope 2.158355 2 1.212078

RoofMatl 1.564670 6 1.038011

LandContour 2.642336 3 1.175794

Street 1.218978 1 1.104074

Comp.5 1.134756 1 1.065249

Comp.3 1.330244 1 1.153362

Comp.7 1.202917 1 1.096776

Utilities 1.112537 2 1.027019

Comp.4 1.066031 1 1.032488