

```
*AKBAR AIDAROV - FINAL REGRESSION MODEL
```

```
*For clarity and comprehensiveness, the following steps include the initial  
data import, transformation/recoding, and cleaning that may differ from my  
teammates' work.;
```

```
*STEP 1 - Import data;
```

```
proc import datafile = 'housedata_new.csv' out = housedata replace;  
delimiter = ',';  
datarow = 2;  
getnames = yes;  
run;
```

```
*STEP 2 - Transform the dependent variable, and create dummy variables and  
new variables with grouped initial values;
```

```
data house_log;  
set housedata;  
ln_saleprice = log(SalePrice);
```

```
numMSSubClass1 = (MSSubClass="30"); *Base "020" = 1-story 1946 and newer all  
styles;
```

```
numMSSubClass1 = (MSSubClass="40");  
numMSSubClass1 = (MSSubClass="45");  
numMSSubClass1 = (MSSubClass="50");  
numMSSubClass2 = (MSSubClass="60");  
numMSSubClass3 = (MSSubClass="70");  
numMSSubClass3 = (MSSubClass="75");  
numMSSubClass3 = (MSSubClass="80");  
numMSSubClass3 = (MSSubClass="85");  
numMSSubClass3 = (MSSubClass="90");  
numMSSubClass4 = (MSSubClass="120");  
numMSSubClass4 = (MSSubClass="160");  
numMSSubClass4 = (MSSubClass="180");  
numMSSubClass4 = (MSSubClass="190");
```

```
numMSZoning1 = (MSZoning="FV"); *Base "C" = Commercial;  
numMSZoning2 = (MSZoning="RH");  
numMSZoning3 = (MSZoning="RL");  
numMSZoning4 = (MSZoning="RM");
```

```
numStreet1 = (Street="Pave"); *Base "Grvl" = Gravel;
```

```
numAlley1 = (Alley="Pave"); *Base "Grvl" = Gravel;  
numAlley2 = (Alley="NA");
```

```
numLotShapel = (LotShape="Reg"); *Base "Irr" = Irregular (IR1, IR2, and IR3  
combined);
```

```
numLandContour1 = (LandContour="HLS"); *Base "Bnk" = Banked;  
numLandContour2 = (LandContour="Low");  
numLandContour3 = (LandContour="Lvl");
```

```
numUtilities1 = (Utilities="NoSeWa"); *Base "AllPub" = All public Utilities  
(E, G, W, & S);
```

```

numLotConfig1 = (LotConfig="Inside"); *Base "MFr" = Frontage on multiple
sides;

numLandSlope1 = (LandSlope="Mod"); *Base "Gtl" = Gentle slope;
numLandSlope2 = (LandSlope="Sev");

numNeighborhood1 = (Neighborhood="CollgCr"); *Base "Crawford" = Crawford,
Bluestem, SWISU, and Timberland grouped by physical proximity, as are all
other groups;
numNeighborhood1 = (Neighborhood="Edwards"); *numNeighborhood1 consists of
College Creek and Edwards neighborhoods;
numNeighborhood2 = (Neighborhood="Blmngtn"); *numNeighborhood2 consists of
Bloomington Heights, Clear Creek, Gilbert, and Stone Brooke neighborhoods;
numNeighborhood2 = (Neighborhood="ClearCr");
numNeighborhood2 = (Neighborhood="Gilbert");
numNeighborhood2 = (Neighborhood="StoneBr");
numNeighborhood3 = (Neighborhood="MeadowV"); *numNeighborhood3 consists of
Meadow Village and Mitchel;
numNeighborhood3 = (Neighborhood="Mitchel");
numNeighborhood4 = (Neighborhood="BrDale"); *numNeighborhood4 consists of
Briardale, North Ames, North Park Village, and North West Ames;
numNeighborhood4 = (Neighborhood="NAMES");
numNeighborhood4 = (Neighborhood="NPkVill");
numNeighborhood4 = (Neighborhood="NWAmes");
numNeighborhood5 = (Neighborhood="NoRidge"); *numNeighborhood5 consists of
Northridge, North Ridge Heights, and Somerset;
numNeighborhood5 = (Neighborhood="NridgHt");
numNeighborhood5 = (Neighborhood="Somerst");
numNeighborhood6 = (Neighborhood="BrkSide"); *numNeighborhood6 consists of
Brookside, Iowa Department of Transportation and Railroad (IDOTRR), and Old
Town;
numNeighborhood6 = (Neighborhood="IDOTRR");
numNeighborhood6 = (Neighborhood="OldTown");
numNeighborhood7 = (Neighborhood="Sawyer");
numNeighborhood7 = (Neighborhood="SawyerW");
numNeighborhood7 = (Neighborhood="Veenker");

numCondition1_1 = (Condition1="Artery"); *Base "Norm" = normal proximity to
streets, parks, and railroads.;
numCondition1_1 = (Condition1="Feedr"); *If Condition1 is not "Norm", all
other labels =1, which means the property is nearby main streets, railroads,
and off-site features like park and greenery;
numCondition1_1 = (Condition1="PosN");
numCondition1_1 = (Condition1="PosA");
numCondition1_1 = (Condition1="RRAe");
numCondition1_1 = (Condition1="RRAn");
numCondition1_1 = (Condition1="RRNe");
numCondition1_1 = (Condition1="RRNn");

numCondition2_1 = (Condition2="Artery"); *Base "Norm" = normal proximity to
more than one positive feature.;
numCondition2_1 = (Condition2="Feedr"); *If Condition2 is not "Norm", all
other labels = 1, which means there are two or more positive features
nearby.;
numCondition2_1 = (Condition2="PosN");
numCondition2_1 = (Condition2="PosA");

```

```

numCondition2_1 = (Condition2="RRAe");
numCondition2_1 = (Condition2="RRAn");
numCondition2_1 = (Condition2="RRNn");

numBldgType1 = (BldgType="2fmCon"); *Base "1Fam" = Single-Family detached
type of dwelling;
numBldgType1 = (BldgType="Duplex"); *All other labels are grouped into one
"Multi-Family" label ("Two-family conversion, originally built as one-family
dwelling", "Duplex", "Townhouse End Unit", "Townhouse Inside Unit").;
numBldgType1 = (BldgType="Twnhs");
numBldgType1 = (BldgType=" TwnhsE");

numHouseStyle1 = (HouseStyle="1Story"); *Base "2Story" = Two story style of
dwelling;
numHouseStyle2 = (HouseStyle="1.5Fin"); *numHouseStyle1 is 1-story dwelling
style;
numHouseStyle2 = (HouseStyle="1.5Unf"); *numHouseStyle2 is all other labels
(1.5- and 2.5-story, finished and unfinished, split foyer and split level);
numHouseStyle2 = (HouseStyle="2.5Fin");
numHouseStyle2 = (HouseStyle="2.5Unf");
numHouseStyle2 = (HouseStyle="SFoyer");
numHouseStyle2 = (HouseStyle="SLvl");

numOverallQual1 = (OverallQual="medium"); *Base "low" = Low overall quality
of material and finish of the house;
numOverallQual2 = (OverallQual="high");

numOverallCond1 = (OverallCond ="medium"); *Base "low" = Low overall
condition of the house;
numOverallCond2 = (OverallCond ="high");

numYearBuilt1 = (YearBuilt="1891-1910"); *Base "1870-1890" = a period in
which the building was built;
numYearBuilt2 = (YearBuilt="1911-1930");
numYearBuilt3 = (YearBuilt="1931-1950");
numYearBuilt4 = (YearBuilt="1951-1970");
numYearBuilt5 = (YearBuilt="1971-1990");
numYearBuilt6 = (YearBuilt="1991-2000");
numYearBuilt7 = (YearBuilt="2001-2010");

numYearRemodAdd1 = (YearRemodAdd="1961-1970"); *Base "1950-1960" = decade in
which the building was last remodeled (same as construction year, if no
remodeling or additions);
numYearRemodAdd2 = (YearRemodAdd="1971-1980");
numYearRemodAdd3 = (YearRemodAdd="1981-1990");
numYearRemodAdd4 = (YearRemodAdd="1991-2000");
numYearRemodAdd5 = (YearRemodAdd="2001-2010");

numRoofStyle1 = (RoofStyle="Gable"); *Base "Other" = All other roof styles in
one label, if not RoofStyle is not Gable or Hip;
numRoofStyle2 = (RoofStyle="Hip");

numRoofMatl1 = (RoofMatl="CompShg"); *Base "ClyTile" = Clay or Tile;
numRoofMatl2 = (RoofMatl="Membran");
numRoofMatl3 = (RoofMatl="Metal");
numRoofMatl4 = (RoofMatl="Roll");
numRoofMatl5 = (RoofMatl="Tar");

```

```

numRoofMatl6 = (RoofMatl="WdShake");
numRoofMatl7 = (RoofMatl="WdShngl");

numExterior1st1 = (Exterior1st="AsphShn"); *Base "AsbShng" = Asbestos
Shingles exterior covering on house;
numExterior1st2 = (Exterior1st="BrkComm");
numExterior1st3 = (Exterior1st="BrkFace");
numExterior1st4 = (Exterior1st="CBlock");
numExterior1st5 = (Exterior1st="CemntBd");
numExterior1st6 = (Exterior1st="HdBoard");
numExterior1st7 = (Exterior1st="ImStucc");
numExterior1st8 = (Exterior1st="MetalSd");
numExterior1st9 = (Exterior1st="Plywood");
numExterior1st10 = (Exterior1st="Stone");
numExterior1st11 = (Exterior1st="Stucco");
numExterior1st12 = (Exterior1st="VinylSd");
numExterior1st13 = (Exterior1st="Wd Sdng");
numExterior1st14 = (Exterior1st="WdShng");

numExterior2nd1 = (Exterior2nd="AsphShn"); *Base "AsbShng" = Asbestos
Shingles exterior covering on house (if more than one material);
numExterior2nd2 = (Exterior2nd="Brk Cmn");
numExterior2nd3 = (Exterior2nd="BrkFace");
numExterior2nd4 = (Exterior2nd="CBlock");
numExterior2nd5 = (Exterior2nd="CemntBd");
numExterior2nd6 = (Exterior2nd="HdBoard");
numExterior2nd7 = (Exterior2nd="ImStucc");
numExterior2nd8 = (Exterior2nd="MetalSd");
numExterior2nd9 = (Exterior2nd="Plywood");
numExterior2nd10 = (Exterior2nd="Stone");
numExterior2nd11 = (Exterior2nd="Other");
numExterior2nd12 = (Exterior2nd="Stucco");
numExterior2nd13 = (Exterior2nd="VinylSd");
numExterior2nd14 = (Exterior2nd="Wd Sdng");
numExterior2nd15 = (Exterior2nd="Wd Shng");

numMasVnrType1 = (MasVnrType="BrkFace"); *Base "BrkCmn" = Brick Common;
numMasVnrType2 = (MasVnrType="None");
numMasVnrType3 = (MasVnrType="Stone");

numExterQual1 = (ExterQual="Fa"); *Base "Ex" = Excellent;
numExterQual2 = (ExterQual="Gd");
numExterQual3 = (ExterQual="TA");

numExterCond1 = (ExterCond= 'Fa'); *Base "Ex" = Excellent;
numExterCond2 = (ExterCond= 'Gd');
numExterCond3 = (ExterCond= 'Po');
numExterCond4 = (ExterCond= 'TA');

numFoundation1 = (Foundation= 'CBlock'); *Base "BrkTil" = Brick & Tile;
numFoundation2 = (Foundation= 'PConc');
numFoundation3 = (Foundation= 'Slab');
numFoundation4 = (Foundation= 'Stone');
numFoundation5 = (Foundation= 'Wood');

```

```

numBsmtQual1 = (BsmtQual= 'Fa'); *Base "Ex" = Excellent (100+ inches) the
height of the basement;
numBsmtQual2 = (BsmtQual= 'Gd');
numBsmtQual3 = (BsmtQual= 'NA');
numBsmtQual4 = (BsmtQual= 'TA');

numBsmtCond1 = (BsmtCond= 'Gd'); *Base "Ex" = Excellent general condition of
the basement;
numBsmtCond2 = (BsmtCond= 'NA');
numBsmtCond3 = (BsmtCond= 'Po');
numBsmtCond4 = (BsmtCond= 'TA');

numBsmtExposure1 = (BsmtExposure= 'Gd'); *Base "Av" = Average Exposure to
walkout or garden level walls (split levels or foyers typically score average
or above);
numBsmtExposure2 = (BsmtExposure= 'Mn');
numBsmtExposure3 = (BsmtExposure= 'NA');
numBsmtExposure4 = (BsmtExposure= 'No');

numBsmtFinType11 = (BsmtFinType1= 'BLQ'); *Base "ALQ" = Average Living
Quarters (rating of basement finished area);
numBsmtFinType12 = (BsmtFinType1= 'GLQ');
numBsmtFinType13 = (BsmtFinType1= 'LwQ');
numBsmtFinType14 = (BsmtFinType1= 'NA');
numBsmtFinType15 = (BsmtFinType1= 'Rec');
numBsmtFinType16 = (BsmtFinType1= 'Unf');

numBsmtFinType21 = (BsmtFinType2= 'BLQ'); *Base "ALQ" = Average Living
Quarters (rating of basement finished area, if multiple types);
numBsmtFinType22 = (BsmtFinType2= 'GLQ');
numBsmtFinType23 = (BsmtFinType2= 'LwQ');
numBsmtFinType24 = (BsmtFinType2= 'NA');
numBsmtFinType25 = (BsmtFinType2= 'Rec');
numBsmtFinType26 = (BsmtFinType2= 'Unf');

TotalHouseSF = TotalBsmtSF + GrLivArea; *Total House area including basement
and bathrooms, and excluding low quality areas;

numHeating1 = (Heating= 'GasA'); *Base "Floor" = Floor Furnace;
numHeating2 = (Heating= 'GasW');
numHeating3 = (Heating= 'Grav');
numHeating4 = (Heating= 'OthW');
numHeating5 = (Heating= 'Wall');

numHeatingQC1 = (HeatingQC= 'Fa'); *Base "Ex" = Excellent quality of heating;
numHeatingQC2 = (HeatingQC= 'Po');
numHeatingQC3 = (HeatingQC= 'Gd');
numHeatingQC4 = (HeatingQC= 'TA');

numCentAir1 = (CentralAir = 'Y'); *Base "No" = No central air conditioning;

numElectrical1 = (Electrical= 'FuseF'); *Base "FuseA" = Fuse Box over 60 AMP
and all Romex wiring (Average);
numElectrical2 = (Electrical= 'FuseP');
numElectrical3 = (Electrical= 'Mix');
numElectrical4 = (Electrical= 'NA');
numElectrical5 = (Electrical= 'SBrkr');

```

```

TotalBaths = BsmtFullBath + BsmtHalfBath + FullBath + HalfBath; *Total baths
in the house;

numKitchenQual1 = (KitchenQual= 'Fa'); *Base 'Ex' = Excellent Kitchen
quality;
numKitchenQual2 = (KitchenQual= 'Gd');
numKitchenQual3 = (KitchenQual= 'TA');

numFunctional1 = (Functional='Min1'); *Base 'Typ' = Typical Functionality;
numFunctional1 = (Functional='Min2');
numFunctional2 = (Functional='Mod');
numFunctional2 = (Functional='Maj1');
numFunctional2 = (Functional='Maj2');
numFunctional2 = (Functional='Sev');

numFireplaceQu1 = (FireplaceQu='Gd'); *Base 'Ex' = Excellent - Exceptional
Masonry Fireplace;
numFireplaceQu2 = (FireplaceQu='TA');
numFireplaceQu3 = (FireplaceQu='Fa');
numFireplaceQu4 = (FireplaceQu='Po');
numFireplaceQu5 = (FireplaceQu='NA');

numGarageType1 = (GarageType='Detchd'); *Base 'Attchd' = Attached type of
garage;
numGarageType2 = (GarageType='Basment'); *numGarageType1 is for detached
garages;
numGarageType2 = (GarageType='BuiltIn'); *numGarageType2 is a dummy variable
for all other types of garages (More than one type of garage, Basement
garage, Car port);
numGarageType2 = (GarageType='CarPort');
numGarageType2 = (GarageType='2Types');
numGarageType3 = (GarageType='NA'); *numGarageType3 = no garage;

numGarageYrBlt1 = (GarageYrBlt='1921-1940'); *Base "1910-1920" = a decade, in
which the garage was built;
numGarageYrBlt2 = (GarageYrBlt='1941-1960');
numGarageYrBlt3 = (GarageYrBlt='1961-1980');
numGarageYrBlt4 = (GarageYrBlt='1981-2000');
numGarageYrBlt5 = (GarageYrBlt='2000-2010');
numGarageYrBlt6 = (GarageYrBlt='NA');

numGarageFinish1 = (GarageFinish='RFn'); *Base 'Fin' = Finished;
numGarageFinish2 = (GarageFinish='Unf');
numGarageFinish3 = (GarageFinish='NA');

numGarageQual1 = (GarageQual='Gd'); *Base 'Ex' = Excellent garage quality;
numGarageQual2 = (GarageQual='TA');
numGarageQual3 = (GarageQual='Fa');
numGarageQual4 = (GarageQual='Po');
numGarageQual5 = (GarageQual='NA');

numGarageCond1 = (GarageCond='Gd'); *Base 'Ex' = Excellent garage condition;
numGarageCond2 = (GarageCond='TA');
numGarageCond3 = (GarageCond='Fa');
numGarageCond4 = (GarageCond='Po');

```

```

numGarageCond5 = (GarageCond='NA');

numPavedDrive1 = (PavedDrive='P'); *Base 'Y' = Paved driveway;
numPavedDrive2 = (PavedDrive='N'); *numPavedDrive1 is for partial pavement,
and numPavedDrive2 is for Dirt/Gravel;

numPoolQC1 = (PoolQC='Gd'); *Base 'Ex' = Excellent;
numPoolQC2 = (PoolQC='Fa');
numPoolQC3 = (PoolQC='NA');

numFence1 = (Fence='MnPrv'); *Base 'NA' = No Fence.;
numFence1 = (Fence='GdWo'); *The other values mean there is fence.;
numFence1 = (Fence='MnWw');
numFence1 = (Fence='GdPrv');

numMiscFeature1 = (MiscFeature='Gar2'); *Base 'Elev' = Elevator
(Miscellaneous features);
numMiscFeature2 = (MiscFeature='Othr');
numMiscFeature3 = (MiscFeature='Shed');
numMiscFeature4 = (MiscFeature='TenC');
numMiscFeature5 = (MiscFeature='NA');

numMoSold1 = (MoSold='spring'); *Base 'autumn' = Sep-Nov;
numMoSold2 = (MoSold='summer');
numMoSold3 = (MoSold='winter');

numYrSold1 = (YrSold='2007'); *Base '2006' = Year Sold;
numYrSold2 = (YrSold='2008');
numYrSold3 = (YrSold='2009');
numYrSold4 = (YrSold='2010');

numSaleType1 = (SaleType='CWD'); *Base 'WD' = Warranty Deed - Conventional;
numSaleType2 = (SaleType='VWD');
numSaleType3 = (SaleType='New');
numSaleType4 = (SaleType='COD');
numSaleType5 = (SaleType='Con');
numSaleType6 = (SaleType='ConLw');
numSaleType7 = (SaleType='ConLI');
numSaleType8 = (SaleType='ConLD');
numSaleType9 = (SaleType='Oth');

numSaleCondition1 = (SaleCondition='Abnorml'); *Base 'Normal' = Normal Sale;
numSaleCondition2 = (SaleCondition='AdjLand');
numSaleCondition3 = (SaleCondition='Alloca');
numSaleCondition4 = (SaleCondition='Family');
numSaleCondition5 = (SaleCondition='Partial');

run;

proc print data=house_log (obs=10);
run;

*STEP 3 - Fit initial regression model including all variables;
proc reg data=house_log;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numAlley1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numBldgType1

```

```

numHouseStyle1 numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1
numRoofStyle1 numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1
MasVnrArea numExterQual1 numExterCond1 numFoundation1 numBsmtQual1
numBsmtCond1 numBsmtExposure1 numBsmtFinType11 BsmtFinSF1 numBsmtFinType21
BsmtFinSF2 BsmtUnfSF TotalBsmtSF TotalHouseSF numHeating1 numHeatingQC1
numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF LowQualFinSF GrLivArea
BsmtFullBath BsmtHalfBath FullBath HalfBath TotalBaths BedroomAbvGr
KitchenAbvGr numKitchenQual1 TotRmsAbvGrd numFunctional1 Fireplaces
numFireplaceQu1 numGarageType1 numGarageYrBlt1 numGarageFinish1 GarageCars
GarageArea numGarageQual1 numGarageCond1 numPavedDrive1 WoodDeckSF
OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch PoolArea numPoolQC1
numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1 numSaleType1
numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2 numLandContour2
numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2 numOverallCond2
numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2 numExterior1st2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleType2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtQual3 numBsmtCond3 numBsmtExposure3
numBsmtFinType13 numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3
numKitchenQual3 numFireplaceQu3 numGarageType3 numGarageYrBlt3
numGarageFinish3 numGarageQual3 numGarageCond3 numPoolQC3 numMiscFeature3
numMoSold3 numYrSold3 numSaleType3 numSaleCondition3 numMSSubClass4
numMSZoning4 numNeighborhood4 numYearBuilt4 numYearRemodAdd4 numRoofMatl4
numExterior1st4 numExterior2nd4 numExterCond4 numFoundation4 numBsmtQual4
numBsmtCond4 numBsmtExposure4 numBsmtFinType14 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numExterior2nd5 numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5
numElectrical5 numFireplaceQu5 numGarageYrBlt5 numGarageQual5 numGarageCond5
numMiscFeature5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numGarageYrBlt6 numSaleType6 numNeighborhood7 numYearBuilt7
numRoofMatl7 numExterior1st7 numExterior2nd7 numSaleType7 numExterior1st8
numExterior2nd8 numSaleType8 numExterior1st9 numExterior2nd9 numSaleType9
numExterior1st10 numExterior2nd10 numExterior1st11 numExterior2nd11
numExterior1st12 numExterior2nd12 numExterior1st13 numExterior2nd13
numExterior1st14 numExterior2nd14 numExterior2nd15 / stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

```

*STEP 4 - Delete parameters that have been set to 0, since the variables are a linear combination of other variables and re-run the updated model;

*FIRST RE-RUN;

```

proc reg data=house_log;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea

```



```

numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalBsmtSF TotalHouseSF
numHeating1 numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF
LowQualFinSF GrLivArea TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1
TotRmsAbvGrd numFunctionall1 Fireplaces numFireplaceQu1 numGarageType1
numGarageYrBlt1 numGarageFinish1 GarageCars GarageArea numGarageQual1
numGarageCond1 numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch
ScreenPorch PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1
numYrSold1 numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2
numAlley2 numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2
numOverallQual2 numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2
numRoofMatl2 numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2
numExterCond2 numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2
numBsmtFinType12 numBsmtFinType22 numHeating2 numHeatingQC2 numElectrical2
numKitchenQual2 numFunctionall2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numSaleType6 numNeighborhood7 numYearBuilt7 numRoofMatl7
numExterior1st7 numExterior2nd7 numSaleType7 numExterior1st8 numExterior2nd8
numSaleType8 numExterior1st9 numExterior2nd9 numSaleType9 numExterior1st10
numExterior2nd10 numExterior1st11 numExterior2nd11 numExterior1st12
numExterior2nd12 numExterior1st13 numExterior2nd13 numExterior1st14
numExterior2nd14 numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

```

*Deleting 5 more parameters that have been set to 0, since the variables are linear combination of other variables.;

*SECOND RE-RUN;

```

proc reg data=house_log;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel1 numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF LowQualFinSF
GrLivArea TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctionall1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1

```

```

numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2
numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12
numBsmtFinType22 numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2
numFunctional2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

```

```

*Deleting 1 parameter LowQualFinSF;
*THIRD RE-RUN;

```

```

proc reg data=house_log;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAirl numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2
numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12

```

```

numBsmFinType22 numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2
numFunctional2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmFinType3 numBsmFinType3 numBsmFinType3
numBsmFinType3 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmQual4 numBsmQual4 numBsmQual4 numBsmQual4 numBsmQual4 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmFinType5 numBsmFinType5 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmFinType6
numBsmFinType6 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;

```

```

plot student.*predicted.;

```

```

plot npp.*student.;

```

```

run;

```

```

*Deleting 8 outliers with both red and blue arrowheads;

```

```

data house_log_clean;

```

```

set house_log;

```

```

if _n_=1433 then delete;

```

```

if _n_=1325 then delete;

```

```

if _n_=711 then delete;

```

```

if _n_=633 then delete;

```

```

if _n_=524 then delete;

```

```

if _n_=463 then delete;

```

```

if _n_=89 then delete;

```

```

if _n_=31 then delete;

```

```

run;

```

```

*Re-running the model with the updated dataset house_log_clean;

```

```

*Fourth RE-RUN;

```

```

proc reg data=house_log_clean;

```

```

model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea

```

```

numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1

```

```

numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1

```

```

numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1

```

```

numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea

```

```

numExterQual1 numExterCond1 numFoundation1 numBsmQual1 numBsmQual1

```

```

numBsmQual1 numBsmQual1 numBsmQual1 numBsmQual1 numBsmQual1 numHeating1

```

```

numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea

```

```

TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd

```

```

numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1

```

```

numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1

```

```

numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch

```

```

PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2
numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12
numBsmtFinType22 numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2
numFunctional2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

*Deleting 10 outliers with both red and blue arrowheads;
data house_log_clean;
set house_log_clean;
if _n_=1436 then delete;
if _n_=1265 then delete;
if _n_=995 then delete;
if _n_=965 then delete;
if _n_=684 then delete;
if _n_=632 then delete;
if _n_=493 then delete;
if _n_=477 then delete;
if _n_=409 then delete;
if _n_=180 then delete;
run;

*Fitting yet another regression model to the updated dataset;
*Fifth RE-RUN;
proc reg data=house_log_clean;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1

```

```

numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2
numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType2
numBsmtFinType22 numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2
numFunctional2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType3
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

*Deleting 2 outliers;
data house_log_clean;
set house_log_clean;
if _n_=1436 then delete;
if _n_=1168 then delete;
run;

*Fitting another regression model to the updated dataset;
*Sixth RE-RUN;
proc reg data=house_log_clean;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1

```

```

numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior1st2 numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2
numFoundation2 numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12
numBsmtFinType22 numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2
numFunctional2 numFireplaceQu2 numGarageType2 numGarageYrBlt2
numGarageFinish2 numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2
numMiscFeature2 numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3
numMSZoning3 numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3
numRoofMatl3 numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3
numExterCond3 numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;

```

```

plot student.*predicted.;

```

```

plot npp.*student.;

```

```

run;

```

```

*Deleting 2 more outliers and eliminating numExterior1st2 from the model;

```

```

data house_log_clean;

```

```

set house_log_clean;

```

```

if _n_=924 then delete;

```

```

if _n_=703 then delete;

```

```

run;

```

```

*Fitting a new regression model to the updated dataset;

```

```

*Seventh RE-RUN;

```

```

proc reg data=house_log_clean;

```

```

model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea

```

```

numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1

```

```

numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1

```

```

numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1

```

```

numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3 numMSZoning3
numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3 numRoofMatl3
numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3 numExterCond3
numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;

```

```
plot student.*predicted.;
```

```
plot npp.*student.;
```

```
run;
```

```

*Due to abnormally big Cook's D values, we delete observations #1405 and
#1155;

```

```

data house_log_clean;
set house_log_clean;
if _n_=1405 then delete;
if _n_=1155 then delete;
run;

```

```
*Fitting a new regression model to the updated dataset;
```

```
*Eighth RE-RUN;
```

```

proc reg data=house_log_clean;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1

```

```

numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctionall1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3 numMSZoning3
numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3 numRoofMatl3
numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3 numExterCond3
numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numRoofMatl7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

```

*A parameter with a very low variability is detected - numRoofMatl7 - and should be eliminated from the regression model;

*Ninth RE-RUN;

```

proc reg data=house_log_clean;
model ln_saleprice = numMSSubClass1 numMSZoning1 LotFrontage LotArea
numStreet1 numLotShapel numLandContour1 numUtilities1 numLotConfig1
numLandSlope1 numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd

```



```

numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3 numMSZoning3
numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3 numRoofMatl3
numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3 numExterCond3
numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/ stb r VIF;
plot student.*predicted.;
plot npp.*student.;
run;

```

```

*STEP 5 - Explore linear associations and correlations between the x-vars and
ln_saleprice;
*No multicollinearity detected;

```

```

proc corr;
var ln_saleprice LotFrontage LotArea MasVnrArea TotalHouseSF _1stFlrSF
_2ndFlrSF GrLivArea TotalBaths BedroomAbvGr KitchenAbvGr TotRmsAbvGrd
Fireplaces GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch
_3SsnPorch ScreenPorch PoolArea MiscVal;
run;

```

```

*Explore 10 highest linear associations and correlations between the x-vars
and ln_saleprice;
proc corr best=10;
var ln_saleprice LotFrontage LotArea MasVnrArea TotalHouseSF _1stFlrSF
_2ndFlrSF GrLivArea TotalBaths BedroomAbvGr KitchenAbvGr TotRmsAbvGrd
Fireplaces GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch
_3SsnPorch ScreenPorch PoolArea MiscVal;
run;

```

```

*Use the best 7 out of the ten to create scatter plot matrix;
*The scatterplots look linear;
proc sgscatter;
matrix ln_saleprice TotalHouseSF GrLivArea GarageCars GarageArea TotalBaths
_1stFlrSF TotRmsAbvGrd;
run;

*STEP 6 - Split the data;
*80/20 split (train/test);
proc surveyselect data=house_log_clean out=train_akbar seed=132408
samprate=0.80 outall;
run;

*STEP 6 - Create a new variable tr_y = ln_saleprice for the training set, and
tr_y = NA for the testing set;
data train_akbar;
set train_akbar;
if selected then tr_y=ln_saleprice;
run;

*STEP 7 - Run STEPWISE model selection on the training data set using tr_y as
a dependent variable;
proc reg data=train_akbar;
model tr_y = numMSSubClass1 numMSZoning1 LotFrontage LotArea numStreet1
numLotShapel numLandContour1 numUtilities1 numLotConfig1 numLandSlope1
numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1 _1stFlrSF _2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3 numMSZoning3
numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3 numRoofMatl3
numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3 numExterCond3
numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5

```

```

numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/selection=stepwise vif stb;
run;

```

**Try another STEPWISE with SLE and SLS = 0.05;*

```

proc reg data=train_akbar;
model tr_y = numMSSubClass1 numMSZoning1 LotFrontage LotArea numStreet1
numLotShape1 numLandContour1 numUtilities1 numLotConfig1 numLandSlope1
numNeighborhood1 numCondition1_1 numCondition2_1 numHouseStyle1
numOverallQual1 numOverallCond1 numYearBuilt1 numYearRemodAdd1 numRoofStyle1
numRoofMatl1 numExterior1st1 numExterior2nd1 numMasVnrType1 MasVnrArea
numExterQual1 numExterCond1 numFoundation1 numBsmtQual1 numBsmtCond1
numBsmtExposure1 numBsmtFinType11 numBsmtFinType21 TotalHouseSF numHeating1
numHeatingQC1 numCentAir1 numElectrical1_1stFlrSF_2ndFlrSF GrLivArea
TotalBaths BedroomAbvGr KitchenAbvGr numKitchenQual1 TotRmsAbvGrd
numFunctional1 Fireplaces numFireplaceQu1 numGarageType1 numGarageYrBlt1
numGarageFinish1 GarageCars GarageArea numGarageQual1 numGarageCond1
numPavedDrive1 WoodDeckSF OpenPorchSF EnclosedPorch_3SsnPorch ScreenPorch
PoolArea numPoolQC1 numFence1 numMiscFeature1 MiscVal1 numMoSold1 numYrSold1
numSaleType1 numSaleCondition1 numMSSubClass2 numMSZoning2 numAlley2
numLandContour2 numLandSlope2 numNeighborhood2 numHouseStyle2 numOverallQual2
numOverallCond2 numYearBuilt2 numYearRemodAdd2 numRoofStyle2 numRoofMatl2
numExterior2nd2 numMasVnrType2 numExterQual2 numExterCond2 numFoundation2
numBsmtQual2 numBsmtCond2 numBsmtExposure2 numBsmtFinType12 numBsmtFinType22
numHeating2 numHeatingQC2 numElectrical2 numKitchenQual2 numFunctional2
numFireplaceQu2 numGarageType2 numGarageYrBlt2 numGarageFinish2
numGarageQual2 numGarageCond2 numPavedDrive2 numPoolQC2 numMiscFeature2
numMoSold2 numYrSold2 numSaleCondition2 numMSSubClass3 numMSZoning3
numLandContour3 numNeighborhood3 numYearBuilt3 numYearRemodAdd3 numRoofMatl3
numExterior1st3 numExterior2nd3 numMasVnrType3 numExterQual3 numExterCond3
numFoundation3 numBsmtCond3 numBsmtExposure3 numBsmtFinType13
numBsmtFinType23 numHeating3 numHeatingQC3 numElectrical3 numKitchenQual3
numFireplaceQu3 numGarageType3 numGarageYrBlt3 numGarageQual3 numGarageCond3
numPoolQC3 numMiscFeature3 numMoSold3 numYrSold3 numSaleType3
numSaleCondition3 numMSSubClass4 numMSZoning4 numNeighborhood4 numYearBuilt4
numYearRemodAdd4 numRoofMatl4 numExterior1st4 numExterCond4 numFoundation4
numBsmtQual4 numBsmtCond4 numBsmtExposure4 numBsmtFinType24 numHeating4
numHeatingQC4 numElectrical4 numFireplaceQu4 numGarageYrBlt4 numGarageQual4
numGarageCond4 numMiscFeature4 numYrSold4 numSaleType4 numSaleCondition4
numNeighborhood5 numYearBuilt5 numYearRemodAdd5 numRoofMatl5 numExterior1st5
numFoundation5 numBsmtFinType15 numBsmtFinType25 numHeating5 numElectrical5
numFireplaceQu5 numSaleType5 numSaleCondition5 numNeighborhood6 numYearBuilt6
numRoofMatl6 numExterior1st6 numExterior2nd6 numBsmtFinType16
numBsmtFinType26 numNeighborhood7 numYearBuilt7 numExterior1st7
numExterior2nd7 numExterior1st8 numExterior2nd8 numExterior1st9
numExterior2nd9 numSaleType9 numExterior1st10 numExterior2nd10
numExterior1st11 numExterior2nd11 numExterior1st12 numExterior2nd12
numExterior1st13 numExterior2nd13 numExterior1st14 numExterior2nd14
numExterior2nd15/selection=stepwise sle=0.05 sls=0.05 vif stb;
run;

```

```

*STEP 8 - Fit the model from the last step of STEPWISE with SLE and SLS =
0.05 and run in on the test set;
proc reg data=train_akbar;
model tr_y = TotalHouseSF GarageCars numPoolQC1 numFoundation2 numCentAir1
TotalBaths numOverallQual2 KitchenAbvGr numYearRemodAdd5 Fireplaces
numMSZoning4 numOverallQual1 numYearRemodAdd4 numSaleCondition1
numNeighborhood1 numSaleType3 numBsmFinType16 GrLivArea numBsmFinType1
numExterQual1 numExterQual3 numExterQual2 LotArea ScreenPorch numHeating3
numOverallCond2 numOverallCond1 numRoofMat15 numStreet1 numGarageFinish2
numGarageYrBlt1 GarageArea numYearBuilt5 numHeatingQC4 numGarageQual3
numGarageYrBlt3 numExterior1st3 numYearRemodAdd1 numMasVnrType3
numGarageQual4 numExterior1st11 WoodDeckSF OpenPorchSF MasVnrArea
numNeighborhood2 numYearBuilt2 numBsmFinType21 numFunctional1 numFunctional2
numExterior1st10 numBsmFinType12 numFoundation5 numYearBuilt1 LotFrontage
numLotShapel numKitchenQual1;
output out=outakbar(where=(tr_y=.) ) p=phat;
run;

proc print data=outakbar;
run;

*STEP 9 - Summarize the results of the cross-validations for model-1;
title "Difference between Observed and Predicted in Test Set";
data outakbar_sum;
set outakbar;
diff=ln_saleprice-phat; *diff is the difference between observed and
predicted values in test set;
absd=abs(diff);
run;

*STEP 10 - Compute predictive statistics: root mean square error (RMSE) and
mean absolute error (MAE);
proc summary data= outakbar_sum;
var diff absd;
output out=outakbar_stats std(diff)=RMSE mean(absd)=MAE;
run;

proc print data=outakbar_stats;
title 'Validation statistics for Model';
run;

*STEP 11 - Compute correlation of observed and predicted values in test set;
proc corr data=outakbar;
var ln_saleprice phat;
run;

*STEP 12 - Create new datalines for 2 predictions;
data new_akbar;
input
tr_y Fireplaces GarageArea GarageCars GrLivArea KitchenAbvGr LotArea
LotFrontage MasVnrArea numBsmFinType1 numBsmFinType12 numBsmFinType16
numBsmFinType21 numCentAir1 numExterior1st10 numExterior1st11
numExterior1st3 numExterQual1 numExterQual2 numExterQual3 numFoundation2
numFoundation5 numFunctional1 numFunctional2 numGarageFinish2 numGarageQual3
numGarageQual4 numGarageYrBlt1 numGarageYrBlt3 numHeating3 numHeatingQC4
numKitchenQual1 numLotShapel numMasVnrType3 numMSZoning4 numNeighborhood1

```

```

numNeighborhood2 numOverallCond1 numOverallCond2 numOverallQual1
numOverallQual2 numPoolQC1 numRoofMatl5 numSaleCondition1 numSaleType3
numStreet1 numYearBuilt1 numYearBuilt2 numYearBuilt5 numYearRemodAdd1
numYearRemodAdd4 numYearRemodAdd5 OpenPorchSF ScreenPorch TotalBaths
TotalHouseSF WoodDeckSF;
datalines;
. 1 1000 2 2800 1 8000 140 800 1 1 0 0 1 1 0 0 1 0 0 1 0 0 1 1 1 0 1 0 1 1 1
1 1 1 1 0 1 0 1 0 0 1 1 0 1 1 0 0 1 0 0 300 0 2 3500 200
. 0 700 1 3000 1 9800 200 1000 1 0 0 1 1 0 1 0 0 1 0 0 1 1 0 1 1 0 0 1 1 0 0
0 1 0 0 1 0 1 0 1 0 0 0 1 1 0 1 0 0 1 0 450 300 3 3900 0
;
proc print data=new_akbar;
run;

*STEP 13 - Merge the two datasets for predictions;
data pred_akbar;
set new_akbar train_akbar;
run;

proc print data=pred_akbar (obs=5);
run;

*STEP 14 - Calculate predicted values, Confidence Interval and Prediction
Intervals;
proc reg data=pred_akbar;
model tr_y = Fireplaces GarageArea GarageCars GrLivArea KitchenAbvGr LotArea
LotFrontage MasVnrArea numBsmtExposure1 numBsmtFinType12 numBsmtFinType16
numBsmtFinType21 numCentAirl1 numExterior1st10 numExterior1st11
numExterior1st3 numExterQual1 numExterQual2 numExterQual3 numFoundation2
numFoundation5 numFunctional1 numFunctional2 numGarageFinish2 numGarageQual3
numGarageQual4 numGarageYrBlt1 numGarageYrBlt3 numHeating3 numHeatingQC4
numKitchenQual1 numLotShapel1 numMasVnrType3 numMSZoning4 numNeighborhood1
numNeighborhood2 numOverallCond1 numOverallCond2 numOverallQual1
numOverallQual2 numPoolQC1 numRoofMatl5 numSaleCondition1 numSaleType3
numStreet1 numYearBuilt1 numYearBuilt2 numYearBuilt5 numYearRemodAdd1
numYearRemodAdd4 numYearRemodAdd5 OpenPorchSF ScreenPorch TotalBaths
TotalHouseSF WoodDeckSF / p clm cli;
run;
quit;

```