```
*AMY AUMPANSUB's Final SAS Code;

*STEP1 import data from housedata_new;

Proc import datafile = 'S:\csc423\housedata_new.csv' out = housedata_amy
replace;
delimiter = ',';
datarow = 2;
getnames = yes;
run;

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


*STEP2 Explore y and x vars;
*Before Transformation Y-var (SalePrice)→ (skewed right);
proc univariate normal;
var SalePrice;
histogram/ normal (mu=est sigma=est);
run;

*STEP3 Transform y var and recode dummy variables;
*After Transformation Y-var (ln_saleprice)→ (normal distribution with few
outliers);

data house_log_amy;
set housedata_amy;
ln_saleprice = log(SalePrice);
run;

proc univariate normal data=house_log_amy;
var ln_saleprice;
histogram/ normal (mu=est sigma=est);
run;


* STEP4- Recoding text vars and create new dummy variables;
data house_log_amy;
set house_log_amy;

numMSSubClass1 = (MSSubClass="30");
numMSSubClass2 = (MSSubClass="40");
numMSSubClass3 = (MSSubClass="45");
numMSSubClass4 = (MSSubClass="50");
numMSSubClass5 = (MSSubClass="60");
numMSSubClass6 = (MSSubClass="70");
numMSSubClass7 = (MSSubClass="75");
numMSSubClass8 = (MSSubClass="80");
numMSSubClass9 = (MSSubClass="85");
numMSSubClass10 = (MSSubClass="90");
numMSSubClass11 = (MSSubClass="120");
numMSSubClass12 = (MSSubClass="160");
numMSSubClass13 = (MSSubClass="180");
numMSSubClass14 = (MSSubClass="190");
```

```
numMSZoning1 = (MSZoning="FV");
numMSZoning2 = (MSZoning="RH");
numMSZoning3 = (MSZoning="RL");
numMSZoning4 = (MSZoning="RM");

numStreet1 = (Street="Pave");

numAlley1 = (Alley="Pave");
numAlley2 = (Alley="NA");

numLotShape1 = (LotShape="IR2");
numLotShape2 = (LotShape="IR3");
numLotShape3 = (LotShape="Reg");

numLandContour1 = (LandContour="HLS");
numLandContour2 = (LandContour="Low");
numLandContour3 = (LandContour="Lvl");

numUtilities = (Utilities="NoSeWa");

numLotConfig1 = (LotConfig="CulDSac");
numLotConfig2 = (LotConfig="FR2");
numLotConfig3 = (LotConfig="FR3");
numLotConfig4 = (LotConfig="Inside");

numLandSlope1 = (LandSlope="Mod");
numLandSlope2 = (LandSlope="Sev");

numNeighborhood1 = (Neighborhood="Blueste");
numNeighborhood2 = (Neighborhood="BrDale");
numNeighborhood3 = (Neighborhood="BrkSide");
numNeighborhood4 = (Neighborhood="ClearCr");
numNeighborhood5 = (Neighborhood="CollgCr");
numNeighborhood6 = (Neighborhood="Crawfor");
numNeighborhood7 = (Neighborhood="Edwards");
numNeighborhood8 = (Neighborhood="Gilbert");
numNeighborhood9 = (Neighborhood="IDOTRR");
numNeighborhood10 = (Neighborhood="MeadowV");
numNeighborhood11 = (Neighborhood="Mitchel");
numNeighborhood12 = (Neighborhood="NAmes");
numNeighborhood13 = (Neighborhood="NoRidge");
numNeighborhood14 = (Neighborhood="NPkVill");
numNeighborhood15 = (Neighborhood="NridgHt");
numNeighborhood16 = (Neighborhood="NWAmes");
numNeighborhood17 = (Neighborhood="OldTown");
numNeighborhood18 = (Neighborhood="Sawyer");
numNeighborhood19 = (Neighborhood="SawyerW");
numNeighborhood20 = (Neighborhood="Somerst");
numNeighborhood21 = (Neighborhood="StoneBr");
numNeighborhood22 = (Neighborhood="SWISU");
numNeighborhood23 = (Neighborhood="Timber");
numNeighborhood24 = (Neighborhood="Veenker");

numCondition1_1 = (Condition1="Feedr");
numCondition1_2 = (Condition1="Norm");
numCondition1_3 = (Condition1="PosN");
numCondition1_4 = (Condition1="PosA");
```

```
numCondition1_5 = (Condition1="RRAe");
numCondition1_6 = (Condition1="RRAn");
numCondition1_7 = (Condition1="RRNe");
numCondition1_8 = (Condition1="RRNn");


numCondition2_1 = (Condition2="Feedr");
numCondition2_2 = (Condition2="Norm");
numCondition2_3 = (Condition2="PosN");
numCondition2_4 = (Condition2="PosA");
numCondition2_5 = (Condition2="RRAe");
numCondition2_6 = (Condition2="RRAn");
numCondition2_7 = (Condition2="RRNn");

numBldgType1 = (BldgType="2fmCon");
numBldgType2 = (BldgType="Duplex");
numBldgType3 = (BldgType="Twnhs");
numBldgType4 = (BldgType=" TwnhsE");

numHouseStyle1 = (HouseStyle="1.5Unf");
numHouseStyle2 = (HouseStyle="1Story");
numHouseStyle3 = (HouseStyle="2.5Fin");
numHouseStyle4 = (HouseStyle="2.5Unf");
numHouseStyle5 = (HouseStyle="2Story");
numHouseStyle6 = (HouseStyle="SFoyer");
numHouseStyle7 = (HouseStyle="SLvl");

numOverallQual1 = (OverallQual="medium");
numOverallQual2 = (OverallQual="high");

numOverallCond1 = (OverallCond ="medium");
numOverallCond2 = (OverallCond ="high");

numYearBuilt1 = (YearBuilt="1891-1910");
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");
numYearRemodAdd2 = (YearRemodAdd="1971-1980");
numYearRemodAdd3 = (YearRemodAdd="1981-1990");
numYearRemodAdd4 = (YearRemodAdd="1991-2000");
numYearRemodAdd5 = (YearRemodAdd="2001-2010");

numRoofStyle1 = (RoofStyle="Gable");
numRoofStyle2 = (RoofStyle="Gambrel");
numRoofStyle3 = (RoofStyle="Hip");
numRoofStyle4 = (RoofStyle="Mansard");
numRoofStyle5 = (RoofStyle="Shed");

numRoofMatl1 = (RoofMatl="CompShg");
numRoofMatl2 = (RoofMatl="Membran");
numRoofMatl3 = (RoofMatl="Metal");
numRoofMatl4 = (RoofMatl="Roll");
```

```
numRoofMatl5 = (RoofMatl="Tar");
numRoofMatl6 = (RoofMatl="WdShake");
numRoofMatl7 = (RoofMatl="WdShngl");
numExterior1st1 = (Exterior1st="AsphShn");
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="WdShing");


numExterior2nd1 = (Exterior2nd="AsphShn");
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");
numMasVnrType2 = (MasVnrType="None");
numMasVnrType3 = (MasVnrType="Stone");

numExterQual1 = (ExterQual="Fa");
numExterQual2 = (ExterQual="Gd");
numExterQual3 = (ExterQual="TA");

ExterCond1 = (ExterCond= 'Fa');
ExterCond2 = (ExterCond= 'Gd');
ExterCond3 = (ExterCond= 'Po');
ExterCond4 = (ExterCond= 'TA');

Foundation1 = (Foundation= 'CBlock');
Foundation2 = (Foundation= 'PConc');
Foundation3 = (Foundation= 'Slab');
Foundation4 = (Foundation= 'Stone');
Foundation5 = (Foundation= 'Wood');

BsmtQual1 = (BsmtQual= 'Fa');
BsmtQual2 = (BsmtQual= 'Gd');
BsmtQual3 = (BsmtQual= 'NA');
```

```
BsmtQual4 = (BsmtQual= 'TA');

BsmtCond1 = (BsmtCond= 'Gd');
BsmtCond2 = (BsmtCond= 'NA');
BsmtCond3 = (BsmtCond= 'Po');
BsmtCond4 = (BsmtCond= 'TA');

BsmtExposure1 = (BsmtExposure= 'Gd');
BsmtExposure2 = (BsmtExposure= 'Mn');
BsmtExposure3 = (BsmtExposure= 'NA');
BsmtExposure4 = (BsmtExposure= 'No');

BsmtFinType11 = (BsmtFinType1= 'BLQ');
BsmtFinType12 = (BsmtFinType1= 'GLQ');
BsmtFinType13 = (BsmtFinType1= 'LwQ');
BsmtFinType14 = (BsmtFinType1= 'NA');
BsmtFinType15 = (BsmtFinType1= 'Rec');
BsmtFinType16 = (BsmtFinType1= 'Unf');

BsmtFinType21 = (BsmtFinType2= 'BLQ');
BsmtFinType22 = (BsmtFinType2= 'GLQ');
BsmtFinType23 = (BsmtFinType2= 'LwQ');
BsmtFinType24 = (BsmtFinType2= 'NA');
BsmtFinType25 = (BsmtFinType2= 'Rec');
BsmtFinType26 = (BsmtFinType2= 'Unf');

Heating1 = (Heating= 'GasA');
Heating2 = (Heating= 'GasW');
Heating3 = (Heating= 'Grav');
Heating4 = (Heating= 'OthW');
Heating5 = (Heating= 'Wall');

HeatingQC1 = (HeatingQC= 'Fa');
HeatingQC2 = (HeatingQC= 'Gd');
HeatingQC3 = (HeatingQC= 'Po');
HeatingQC4 = (HeatingQC= 'TA');

NumCentAir = (CentralAir = 'Y');

Electrical1 = (Electrical= 'FuseF');
Electrical2 = (Electrical= 'FuseP');
Electrical3 = (Electrical= 'Mix');
Electrical4 = (Electrical= 'NA');
Electrical5 = (Electrical= 'SBrkr');

KitchenQual1 = (KitchenQual= 'Fa');
KitchenQual2 = (KitchenQual= 'Gd');
KitchenQual3 = (KitchenQual= 'TA');

numFunctional1 = (Functional='Min1');
numFunctional2 = (Functional='Min2');
numFunctional3 = (Functional='Mod');
numFunctional4 = (Functional='Maj1');
numFunctional5 = (Functional='Maj2');
numFunctional6 = (Functional='Sev');
numFunctional7 = (Functional='Sal');
```

```
numFireplaceQu1 = (FireplaceQu='Gd');
numFireplaceQu2 = (FireplaceQu='TA');
numFireplaceQu3 = (FireplaceQu='Fa');
numFireplaceQu4 = (FireplaceQu='Po');
numFireplaceQu5 = (FireplaceQu='NA');

numGarageType1 = (GarageType='Attchd');
numGarageType2 = (GarageType='Basment');
numGarageType3 = (GarageType='BuiltIn');
numGarageType4 = (GarageType='CarPort');
numGarageType5 = (GarageType='Detchd');
numGarageType6 = (GarageType='NA');

numGarageYrBlt1 = (GarageYrBlt='1921-1940');
numGarageYrBlt2 = (GarageYrBlt='1941-1960');
numGarageYrBlt3 = (GarageYrBlt='1961-1980');
numGarageYrBlt4 = (GarageYrBlt='1981-2000');
numGarageYrBlt5 = (GarageYrBlt='2000-2010');
numGarageYrBlt6 = (GarageYrBlt='NA');


numGarageFinish1 = (GarageFinish='RFn');
numGarageFinish2 = (GarageFinish='Unf');
numGarageFinish3 = (GarageFinish='NA');

numGarageQual1 = (GarageQual='Gd');
numGarageQual2 = (GarageQual='TA');
numGarageQual3 = (GarageQual='Fa');
numGarageQual4 = (GarageQual='Po');
numGarageQual5 = (GarageQual='NA');

numGarageCond1 = (GarageCond='Gd');
numGarageCond2 = (GarageCond='TA');
numGarageCond3 = (GarageCond='Fa');
numGarageCond4 = (GarageCond='Po');
numGarageCond5 = (GarageCond='NA');

numPavedDrive1 = (PavedDrive='P');
numPavedDrive2 = (PavedDrive='N');

numPoolQC1 = (PoolQC='Gd');
numPoolQC2 = (PoolQC='TA');
numPoolQC3 = (PoolQC='Fa');
numPoolQC4 = (PoolQC='NA');

numFence1 = (Fence='MnPrv');
numFence2 = (Fence='GdWo');
numFence3 = (Fence='MnWw');
numFence4 = (Fence='NA');

numMiscFeature1 = (MiscFeature='Gar2');
numMiscFeature2 = (MiscFeature='Othr');
numMiscFeature3 = (MiscFeature='Shed');
numMiscFeature4 = (MiscFeature='TenC');
numMiscFeature5 = (MiscFeature='NA');
```

```
numMoSold1 = (MoSold='spring');
numMoSold2 = (MoSold='summer');
numMoSold3 = (MoSold='winter');
numYrSold1 = (YrSold='2007');
numYrSold2 = (YrSold='2008');
numYrSold3 = (YrSold='2009');
numYrSold4 = (YrSold='2010');


numSaleType1 = (SaleType='CWD');
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');
numSaleCondition2 = (SaleCondition='AdjLand');
numSaleCondition3 = (SaleCondition='Alloca');
numSaleCondition4 = (SaleCondition='Family');
numSaleCondition5 = (SaleCondition='Partial');

run;

* STEP5- check corr among y and x vars (only numeric vars);
proc corr data = house_log_amy;
var ln_saleprice LotFrontage LotArea MasVnrArea BsmtFinSF1 BsmtFinSF2
BsmtUnfSF TotalBsmtSF _1stFlrSF _2ndFlrSF LowQualFinSF
GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr
KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars
GarageArea WoodDeckSF OpenPorchSF EnclosedPorch _3SsnPorch
ScreenPorch PoolArea MiscVal;
run;

* STEP6- fit the model to check vif, outliers and influential points;
Proc reg data = house_log_amy;
model ln_saleprice = numMSSubClass1 numMSSubClass2 numMSSubClass3
numMSSubClass4 numMSSubClass5 numMSSubClass6 numMSSubClass7 numMSSubClass8
numMSSubClass9 numMSSubClass10 numMSSubClass11 numMSSubClass12
numMSSubClass13 numMSSubClass14 numMSZoning1 numMSZoning2 numMSZoning3
numMSZoning4 LotFrontage LotArea numStreet1 numAlley2 numLotShape1
numLotShape2 numLotShape3 numLandContour1 numLandContour2 numLandContour3
numUtilities numLotConfig1 numLotConfig2 numLotConfig3 numLotConfig4
numLandSlope1 numLandSlope2 numNeighborhood1 numNeighborhood2
numNeighborhood3 numNeighborhood4 numNeighborhood5 numNeighborhood6
numNeighborhood7 numNeighborhood8 numNeighborhood9 numNeighborhood10
numNeighborhood11 numNeighborhood12 numNeighborhood13 numNeighborhood14
numNeighborhood15 numNeighborhood16 numNeighborhood17 numNeighborhood18
numNeighborhood19 numNeighborhood20 numNeighborhood21 numNeighborhood22
numNeighborhood23 numNeighborhood24 numCondition1_1 numCondition1_2
numCondition1_3 numCondition1_4 numCondition1_5 numCondition1_6
numCondition1_7 numCondition1_8 numCondition2_1 numCondition2_2
numCondition2_3 numCondition2_4 numCondition2_5 numCondition2_6
```

```
numCondition2_7 numBldgType1 numBldgType3 numHouseStyle1 numHouseStyle2
numHouseStyle3 numHouseStyle4 numHouseStyle5 numHouseStyle6 numHouseStyle7
numOverallQual1 numOverallQual2 numOverallCond1 numOverallCond2 numYearBuilt1
numYearBuilt2 numYearBuilt3 numYearBuilt4 numYearBuilt5 numYearBuilt6
numYearBuilt7 numYearRemodAdd1 numYearRemodAdd2 numYearRemodAdd3
numYearRemodAdd4 numYearRemodAdd5 numRoofStyle1 numRoofStyle3 numRoofStyle5
numRoofMatl1 numRoofMatl2 numRoofMatl3 numRoofMatl4 numRoofMatl5 numRoofMatl6
numRoofMatl7 numExterior1st1 numExterior1st2 numExterior1st3 numExterior1st4
numExterior1st5 numExterior1st6 numExterior1st7 numExterior1st8
numExterior1st9 numExterior1st10 numExterior1st11 numExterior1st12
numExterior1st13 numExterior1st14 numExterior2nd1 numExterior2nd2
numExterior2nd3 numExterior2nd6 numExterior2nd7 numExterior2nd8
numExterior2nd9 numExterior2nd10
numExterior2nd11 numExterior2nd12 numExterior2nd13 numExterior2nd14
numExterior2nd15 numMasVnrType1 numMasVnrType2 numMasVnrType3 MasVnrArea
numExterQual1 numExterQual2 numExterQual3 ExterCond1 ExterCond2 ExterCond3
ExterCond4 Foundation1 Foundation2 Foundation3 Foundation4 Foundation5
BsmtQual1 BsmtQual2 BsmtQual3 BsmtQual4 BsmtCond1 BsmtCond3
BsmtCond4 BsmtExposure1 BsmtExposure2 BsmtExposure3 BsmtExposure4
BsmtFinType11 BsmtFinType12 BsmtFinType13 BsmtFinType15
BsmtFinType16 BsmtFinType21 BsmtFinType22 BsmtFinType23 BsmtFinType24
BsmtFinType25 BsmtFinType26 TotalBsmtSF Heating1 Heating2 Heating3 Heating4
Heating5 HeatingQC1 HeatingQC2 HeatingQC3
HeatingQC4 NumCentAir Electrical1 Electrical2 Electrical3 Electrical4
Electrical5 GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 KitchenQual3 TotRmsAbvGrd
numFunctional1 numFunctional2
numFunctional3 numFunctional4 numFunctional5 numFunctional6
Fireplaces numFireplaceQu1 numFireplaceQu2 numFireplaceQu3 numFireplaceQu4
numFireplaceQu5 numGarageType1 numGarageType2 numGarageType3 numGarageType4
numGarageType5 numGarageType6 numGarageYrBlt1 numGarageYrBlt2 numGarageYrBlt3
numGarageYrBlt4 numGarageFinish1 numGarageFinish2 GarageCars GarageArea
numGarageQual1 numGarageQual2 numGarageQual3 numGarageQual4 numGarageCond1
numGarageCond2 numGarageCond3 numGarageCond4 numPavedDrive1 numPavedDrive2
WoodDeckSF OpenPorchSF  EnclosedPorch _3SsnPorch ScreenPorch PoolArea
numPoolQC1 numPoolQC3 numPoolQC4 numFence1 numFence2 numFence3 numFence4
numMiscFeature1 numMiscFeature2 numMiscFeature3 numMiscFeature4 MiscVal
numMoSold1 numMoSold2 numMoSold3 numYrSold1 numYrSold2 numYrSold3 numYrSold4
numSaleType1 numSaleType3 numSaleType4 numSaleType5 numSaleType9
numSaleCondition1 numSaleCondition2 numSaleCondition3 numSaleCondition4
numSaleCondition5/vif influence r;
plot npp.*student.;
plot student.*predicted.;
run;


* STEP7- Cleaning data, remove outliers and influential points with both red
and blue arrowheads;
data house_log_amy;
set house_log_amy;
if _n_=10 then delete;
if _n_=31 then delete;
if _n_=89 then delete;
if _n_=463 then delete;
if _n_=524 then delete;
if _n_=534 then delete;
if _n_=633 then delete;
```

```
if _n_=689 then delete;
if _n_=712 then delete;
if _n_=826 then delete;
if _n_=969 then delete;
if _n_=971 then delete;
if _n_=1187 then delete;
if _n_=1325 then delete;
if _n_=1433 then delete;
if _n_=1444 then delete;
if _n_=1454 then delete;
run;


* STEP8- Run model after Cleaning data to check the improvement;

Proc reg data = house_log_amy;
model ln_saleprice = numMSSubClass1 numMSSubClass2 numMSSubClass3
numMSSubClass4 numMSSubClass5 numMSSubClass6 numMSSubClass7 numMSSubClass8
numMSSubClass9 numMSSubClass10 numMSSubClass11 numMSSubClass12
numMSSubClass13 numMSSubClass14 numMSZoning1 numMSZoning2 numMSZoning3
numMSZoning4 LotFrontage LotArea numStreet1 numAlley2 numLotShape1
numLotShape2 numLotShape3 numLandContour1 numLandContour2 numLandContour3
numUtilities numLotConfig1 numLotConfig2 numLotConfig3 numLotConfig4
numLandSlope1 numLandSlope2 numNeighborhood1 numNeighborhood2
numNeighborhood3 numNeighborhood4 numNeighborhood5 numNeighborhood6
numNeighborhood7 numNeighborhood8 numNeighborhood9 numNeighborhood10
numNeighborhood11 numNeighborhood12 numNeighborhood13 numNeighborhood14
numNeighborhood15 numNeighborhood16 numNeighborhood17 numNeighborhood18
numNeighborhood19 numNeighborhood20 numNeighborhood21 numNeighborhood22
numNeighborhood23 numNeighborhood24 numCondition1_1 numCondition1_2
numCondition1_3 numCondition1_4 numCondition1_5 numCondition1_6
numCondition1_7 numCondition1_8 numCondition2_1 numCondition2_2
numCondition2_4 numCondition2_5 numCondition2_6
numBldgType1 numBldgType3 numHouseStyle2 numHouseStyle3 numHouseStyle4
numHouseStyle5 numHouseStyle6 numHouseStyle7 numOverallQual1 numOverallQual2
numOverallCond1 numOverallCond2 numYearBuilt1 numYearBuilt2 numYearBuilt3
numYearBuilt4 numYearBuilt5 numYearBuilt6 numYearBuilt7 numYearRemodAdd1
numYearRemodAdd2 numYearRemodAdd3 numYearRemodAdd4 numYearRemodAdd5
numRoofStyle1 numRoofStyle3 numRoofStyle5 numRoofMatl1 numRoofMatl2
numRoofMatl3 numRoofMatl4 numRoofMatl5 numRoofMatl6 numRoofMatl7
numExterior1st1 numExterior1st2 numExterior1st3 numExterior1st4
numExterior1st5 numExterior1st6 numExterior1st7 numExterior1st8
numExterior1st9 numExterior1st10 numExterior1st11 numExterior1st12
numExterior1st13 numExterior1st14 numExterior2nd1 numExterior2nd2
numExterior2nd3 numExterior2nd6 numExterior2nd7 numExterior2nd8
numExterior2nd9 numExterior2nd10
numExterior2nd11 numExterior2nd12 numExterior2nd13 numExterior2nd14
numExterior2nd15 numMasVnrType1 numMasVnrType2 numMasVnrType3 MasVnrArea
numExterQual1 numExterQual2 numExterQual3 ExterCond1 ExterCond2 ExterCond3
ExterCond4 Foundation1 Foundation2 Foundation3 Foundation4 Foundation5
BsmtQual1 BsmtQual2 BsmtQual3 BsmtQual4 BsmtCond1 BsmtCond3
BsmtCond4 BsmtExposure1 BsmtExposure2 BsmtExposure3 BsmtExposure4
BsmtFinType11 BsmtFinType12 BsmtFinType13 BsmtFinType15
BsmtFinType16 BsmtFinType21 BsmtFinType22 BsmtFinType23 BsmtFinType24
BsmtFinType25 BsmtFinType26 TotalBsmtSF Heating1 Heating2 Heating3 Heating4
Heating5 HeatingQC1 HeatingQC2 HeatingQC3
HeatingQC4 NumCentAir Electrical1 Electrical2 Electrical3 Electrical4
```

```
Electrical5 GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 KitchenQual3 TotRmsAbvGrd
numFunctional1 numFunctional2
numFunctional3 numFunctional4 numFunctional5 numFunctional6
Fireplaces numFireplaceQu1 numFireplaceQu2 numFireplaceQu3 numFireplaceQu4
numFireplaceQu5 numGarageType1 numGarageType2 numGarageType3 numGarageType4
numGarageType5 numGarageType6 numGarageYrBlt1 numGarageYrBlt2 numGarageYrBlt3
numGarageYrBlt4 numGarageFinish1 numGarageFinish2 GarageCars GarageArea
numGarageQual1 numGarageQual2 numGarageQual3 numGarageQual4 numGarageCond1
numGarageCond2 numGarageCond3 numGarageCond4 numPavedDrive1 numPavedDrive2
WoodDeckSF OpenPorchSF  EnclosedPorch _3SsnPorch ScreenPorch PoolArea
numPoolQC1 numPoolQC3 numPoolQC4 numFence1 numFence2 numFence3 numFence4
numMiscFeature1 numMiscFeature2 numMiscFeature3 numMiscFeature4 MiscVal
numMoSold1 numMoSold2 numMoSold3 numYrSold1 numYrSold2 numYrSold3 numYrSold4
numSaleType1 numSaleType3 numSaleType4 numSaleType5 numSaleType9
numSaleCondition1 numSaleCondition2 numSaleCondition3 numSaleCondition4
numSaleCondition5/vif influence r;
plot npp.*student.;
plot student.*predicted.;
run;

* STEP9- Split the cleaned data to train/test sets with seed#731425;
proc surveyselect data=house_log_amy out=train_amy seed=731425
samprate=0.80 outall;
run;

data train_amy;
set train_amy;
if selected then train_y=ln_saleprice;
run;

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

* STEP10- Fit the model with train set, using stepwise selection;
Proc reg data = train_amy;
model train_y = numMSSubClass1 numMSSubClass2 numMSSubClass3 numMSSubClass4
numMSSubClass5 numMSSubClass6 numMSSubClass7 numMSSubClass8 numMSSubClass9
numMSSubClass10 numMSSubClass11 numMSSubClass12 numMSSubClass13
numMSSubClass14 numMSZoning1 numMSZoning2 numMSZoning3 numMSZoning4
LotFrontage LotArea numStreet1 numAlley2 numLotShape1
numLotShape2 numLotShape3 numLandContour1 numLandContour2 numLandContour3
numUtilities numLotConfig1 numLotConfig2 numLotConfig3 numLotConfig4
numLandSlope1 numLandSlope2 numNeighborhood1 numNeighborhood2
numNeighborhood3 numNeighborhood4 numNeighborhood5 numNeighborhood6
numNeighborhood7 numNeighborhood8 numNeighborhood9 numNeighborhood10
numNeighborhood11 numNeighborhood12 numNeighborhood13 numNeighborhood14
numNeighborhood15 numNeighborhood16 numNeighborhood17 numNeighborhood18
numNeighborhood19 numNeighborhood20 numNeighborhood21 numNeighborhood22
numNeighborhood23 numNeighborhood24 numBldgType1 numBldgType3 numHouseStyle2
numHouseStyle3 numHouseStyle4 numHouseStyle5 numHouseStyle6 numHouseStyle7
numOverallQual1 numOverallQual2 numOverallCond1 numOverallCond2 numYearBuilt1
numYearBuilt2 numYearBuilt3 numYearBuilt4 numYearBuilt5 numYearBuilt6
numYearBuilt7 numYearRemodAdd1 numYearRemodAdd2 numYearRemodAdd3
numYearRemodAdd4 numYearRemodAdd5 numRoofStyle1 numRoofStyle3 numRoofStyle5
numRoofMatl1 numRoofMatl2 numRoofMatl3 numRoofMatl4 numRoofMatl5 numRoofMatl6
numRoofMatl7 numExterior1st1 numExterior1st2 numExterior1st3 numExterior1st4
```

```
numExterior1st5 numExterior1st6 numExterior1st7 numExterior1st8
numExterior1st9 numExterior1st10 numExterior1st11 numExterior1st12
numExterior1st13 numExterior1st14 numExterior2nd1 numExterior2nd2
numExterior2nd3 numExterior2nd6 numExterior2nd7 numExterior2nd8
numExterior2nd9 numExterior2nd10
numExterior2nd11 numExterior2nd12 numExterior2nd13 numExterior2nd14
numExterior2nd15 numMasVnrType1 numMasVnrType2 numMasVnrType3 MasVnrArea
numExterQual1 numExterQual2 numExterQual3 ExterCond1 ExterCond2 ExterCond3
ExterCond4 Foundation1 Foundation2 Foundation3 Foundation4 Foundation5
BsmtQual1 BsmtQual2 BsmtQual3 BsmtQual4 BsmtCond1 BsmtCond3
BsmtCond4 BsmtExposure1 BsmtExposure2 BsmtExposure3 BsmtExposure4
BsmtFinType11 BsmtFinType12 BsmtFinType13 BsmtFinType15
BsmtFinType16 BsmtFinType21 BsmtFinType22 BsmtFinType23 BsmtFinType24
BsmtFinType25 BsmtFinType26 TotalBsmtSF Heating1 Heating2 Heating3 Heating4
Heating5 HeatingQC1 HeatingQC2 HeatingQC3
HeatingQC4 NumCentAir Electrical1 Electrical2 Electrical3 Electrical4
Electrical5 GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 KitchenQual3 TotRmsAbvGrd
numFunctional1 numFunctional2
numFunctional3 numFunctional4 numFunctional5 numFunctional6
Fireplaces numFireplaceQu1 numFireplaceQu2 numFireplaceQu3 numFireplaceQu4
numFireplaceQu5 numGarageType1 numGarageType2 numGarageType3 numGarageType4
numGarageType5 numGarageType6 numGarageYrBlt1 numGarageYrBlt2 numGarageYrBlt3
numGarageYrBlt4 numGarageFinish1 numGarageFinish2 GarageCars GarageArea
numGarageQual1 numGarageQual2 numGarageQual3 numGarageQual4 numGarageCond1
numGarageCond2 numGarageCond3 numGarageCond4 numPavedDrive1 numPavedDrive2
WoodDeckSF OpenPorchSF  EnclosedPorch _3SsnPorch ScreenPorch PoolArea
numPoolQC1 numPoolQC3 numPoolQC4 numFence1 numFence2 numFence3 numFence4
numMiscFeature1 numMiscFeature2 numMiscFeature3 numMiscFeature4 MiscVal
numMoSold1 numMoSold2 numMoSold3 numYrSold1 numYrSold2 numYrSold3 numYrSold4
numSaleType1 numSaleType3 numSaleType4 numSaleType5 numSaleType9
numSaleCondition1 numSaleCondition2 numSaleCondition3 numSaleCondition4
numSaleCondition5/selection=stepwise stb;
run;

* STEP11- Remove some vars, which lower partial R2, then fit the FINAL model;
Proc reg data = train_amy;
model train_y = numMSSubClass1 numMSSubClass4 numMSSubClass7 numMSSubClass10
numMSSubClass11 numMSSubClass12 numMSZoning1 numMSZoning2 numMSZoning3
numMSZoning4 LotArea numLotShape1 numNeighborhood3 numNeighborhood4
numNeighborhood6 numNeighborhood9 numNeighborhood10
numNeighborhood13 numNeighborhood15 numNeighborhood20 numNeighborhood21
numOverallQual1 numOverallQual2 numOverallCond1 numOverallCond2 numYearBuilt4
numYearBuilt5 numYearBuilt6 numYearBuilt7 numYearRemodAdd2 numYearRemodAdd3
numYearRemodAdd4 numYearRemodAdd5 numExterior1st3 numExterior1st5
numExterior1st8 numExterior1st11 numExterior2nd13 MasVnrArea
BsmtQual2 BsmtQual4 BsmtExposure1 BsmtFinType16 BsmtFinType24 TotalBsmtSF
Heating1 Heating2 Heating5 NumCentAir GrLivArea BsmtFullBath FullBath
HalfBath BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 KitchenQual3
numFunctional1 numFunctional2 numFunctional3 numFunctional4
Fireplaces numGarageType1 numGarageYrBlt1 GarageCars GarageArea
numGarageQual1 numGarageQual3 numGarageCond1 WoodDeckSF EnclosedPorch
ScreenPorch PoolArea numPoolQC1 numPoolQC4 numSaleType3 numSaleCondition1
numSaleCondition5/stb;

output out=outamy(where=(train_y=.))  p=phat; *train_y=. save the test set;
run;
```

```
* STEP12 Compute performance stats (R2, RMSE, MAE);
data outamy_sum;
set outamy;
diff=ln_saleprice-phat; *diff is the difference b/t observed and predicted;
absd=abs(diff);
run;

proc summary data= outamy_sum;
var diff absd;
output out=outamy_stats std(diff)=rmse mean(absd)=mae;
run;

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

proc corr data=outamy;
var ln_saleprice phat;
run;

* STEP13 Create new datalines for 2 predictions;
data new_amy;
input
train_y numMSZoning1 numMSZoning2 numMSZoning3 GrLivArea KitchenQual3
PoolArea numPoolQC4 numSaleType3 numOverallQual2 numOverallCond2
numMSSubClass1 numMSSubClass4 numMSSubClass7 numMSSubClass10 numMSSubClass11
numMSSubClass12 numMSZoning4 LotArea numLotShape1 numNeighborhood3
numNeighborhood4 numNeighborhood6 numNeighborhood9 numNeighborhood10
numNeighborhood13 numNeighborhood15 numNeighborhood20 numNeighborhood21
numOverallQual1 numOverallCond1 numYearBuilt4 numYearBuilt5 numYearBuilt6
numYearRemodAdd5 numExterior1st3 numExterior1st5 numExterior1st8
numYearBuilt7 numYearRemodAdd2 numYearRemodAdd3 numYearRemodAdd4
numExterior1st11 numExterior2nd13 MasVnrArea
BsmtQual2 BsmtQual4 BsmtExposure1 BsmtFinType16 BsmtFinType24 TotalBsmtSF
Heating1 Heating2 Heating5 NumCentAir BsmtFullBath FullBath HalfBath
BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 numFunctional1
numFunctional2 numFunctional3 numFunctional4
Fireplaces numGarageType1 numGarageYrBlt1 GarageCars GarageArea
numGarageQual1 numGarageQual3 numGarageCond1 WoodDeckSF EnclosedPorch
ScreenPorch numPoolQC1 numSaleCondition1 numSaleCondition5;

datalines;

. 0 0 1 1000 1 300 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0

. 0 0 1 1200 1 200 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0

;
proc print data=new_amy (obs=2);
run;
```

```
* STEP14 Merge 2 datasets for new predictions;
data pred_amy;
set new_amy train_amy;
run;
proc print data=pred_amy (obs=5);
run;

* STEP15 Calculate predicted value, C.I., P.I. for new predictions;
Proc reg data = pred_amy;
model train_y = numMSSubClass1 numMSSubClass4 numMSSubClass7 numMSSubClass10
numMSSubClass11 numMSSubClass12 numMSZoning1 numMSZoning2 numMSZoning3
numMSZoning4 LotArea numLotShape1 numNeighborhood3 numNeighborhood4
numNeighborhood6 numNeighborhood9 numNeighborhood10
numNeighborhood13 numNeighborhood15 numNeighborhood20 numNeighborhood21
numOverallQual1 numOverallQual2 numOverallCond1 numOverallCond2 numYearBuilt4
numYearBuilt5 numYearBuilt6 numYearBuilt7 numYearRemodAdd2 numYearRemodAdd3
numYearRemodAdd4 numYearRemodAdd5 numExterior1st3 numExterior1st5
numExterior1st8 numExterior1st11 numExterior2nd13 MasVnrArea
BsmtQual2 BsmtQual4 BsmtExposure1 BsmtFinType16 BsmtFinType24 TotalBsmtSF
Heating1 Heating2 Heating5 NumCentAir GrLivArea BsmtFullBath FullBath
HalfBath BedroomAbvGr KitchenAbvGr KitchenQual1 KitchenQual2 KitchenQual3
numFunctional1 numFunctional2 numFunctional3 numFunctional4
Fireplaces numGarageType1 numGarageYrBlt1 GarageCars GarageArea
numGarageQual1 numGarageQual3 numGarageCond1 WoodDeckSF EnclosedPorch
ScreenPorch PoolArea numPoolQC1 numPoolQC4 numSaleType3 numSaleCondition1
numSaleCondition5/p clm cli;
run;
```