/\* import in the data set \*/

Proc print data = TRAIN;

Run;

/\*Look at only the neighborhoods of Edwards , North Ames, and BrookSide \*/

data train2;

set TRAIN;

where Neighborhood = “Edwards” | Neighborhood = “NAmes” | Neighborhood = “BrkSide;”

run;

**Table (1.1)**

Graphical user interface, application, table

Description automatically generated with medium confidence

/\*Look at influential points and outliers \*/

Proc reg data = train2 plots(only label) = (CooksD RStudentByPredicted RStudentByLeverage);

model SalePrice = GrLivArea;

run;

**Figure 2**

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

Table

Description automatically generatedTable

Description automatically generated

/\*remove outliers from dataset \*/

data train2;

modify train2;

if Id=1299 then remove;

if Id=524 then remove;

if Id=725 then remove;

if Id=643 then remove;

run;

/\* display of the four outliers that are removed to mitigate overfitting of high residuals \*/

data trainoutlier;

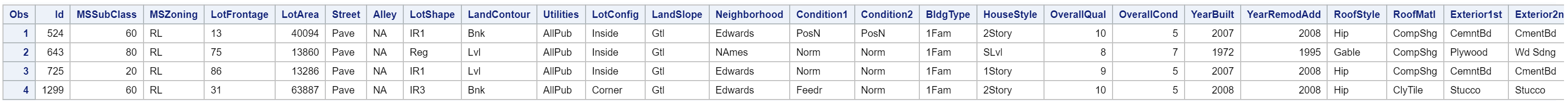
set train2;

where Id=643 or Id=725 or Id=1299 or Id=524;

proc print data = trainoutlier;

run;

**Table 1.1**



/\*Sort the data by Neighborhood \*/

Proc sort data = train2;

By Neighborhood;

Run;

/\* Get diagnostic plots and scatter plots for the neighborhoods \*/

Proc reg data = train2;

Model SalePrice = GrLivArea / vif cli clm clparm;

Run;

**Table 2**

Graphical user interface, text, application

Description automatically generatedTable

Description automatically generated

**Figure 3**

Chart, diagram, engineering drawing

Description automatically generatedChart, scatter chart

Description automatically generated

**Figure 3.1**

Chart, scatter chart

Description automatically generated

/\* Look at the scatterplots and residuals per neighborhood \*/

Proc reg data = train2;

By Neighborhood;

Model SalePrice = GrLivArea;

Run;

**Figure 4.1**

Diagram

Description automatically generatedChart, scatter chart

Description automatically generated

**Figure 4.2**

Graphical user interface, diagram, Word, engineering drawing

Description automatically generatedChart, scatter chart

Description automatically generated

**Figure 4.3**

Chart, diagram

Description automatically generatedChart, scatter chart

Description automatically generated

\*Based on the scatterplots, there is no evidence additional transformations needed.

/\* Look at the final model with respect to the neighborhood \*/

Proc glm data = train2;

Class Neighborhood;

Model SalePrice = GrLivArea | Neighborhood /solution clparm;

Run;

**Table 3**

Table

Description automatically generated

/\* Look at the final model with respect to the neighborhood with reference point BrkSide \*/

Proc glm data = train2;

Class Neighborhood(ref = “BrkSide”);

Model SalePrice = GrLivArea | Neighborhood /solution clparm;

Run;

**Table 3.1**

Table

Description automatically generated

**proc** **print** data = TRAIN;

**run**;

**data** example;

set TRAIN;

if Neighborhood = "Blmngtn" and LotFrontage = "NA" then LotFrontage = "47";

if Neighborhood = "Blueste" and LotFrontage = "NA" then LotFrontage = "24";

if Neighborhood = "BrDale" and LotFrontage = "NA" then LotFrontage = "22";

if Neighborhood = "BrkSide" and LotFrontage = "NA" then LotFrontage = "57";

if Neighborhood = "ClearCr" and LotFrontage = "NA" then LotFrontage = "83";

if Neighborhood = "CollgCr" and LotFrontage = "NA" then LotFrontage = "72";

if Neighborhood = "Crawfor" and LotFrontage = "NA" then LotFrontage = "72";

if Neighborhood = "Edwards" and LotFrontage = "NA" then LotFrontage = "68";

if Neighborhood = "Gilbert" and LotFrontage = "NA" then LotFrontage = "80";

if Neighborhood = "IDOTRR" and LotFrontage = "NA" then LotFrontage = "63";

if Neighborhood = "MeadowV" and LotFrontage = "NA" then LotFrontage = "28";

if Neighborhood = "Mitchel" and LotFrontage = "NA" then LotFrontage = "70";

if Neighborhood = "NAmes" and LotFrontage = "NA" then LotFrontage = "76";

if Neighborhood = "NoRidge" and LotFrontage = "NA" then LotFrontage = "92";

if Neighborhood = "NPkVill" and LotFrontage = "NA" then LotFrontage = "32";

if Neighborhood = "NridgHt" and LotFrontage = "NA" then LotFrontage = "82";

if Neighborhood = "NWAmes" and LotFrontage = "NA" then LotFrontage = "81";

if Neighborhood = "OldTown" and LotFrontage = "NA" then LotFrontage = "63";

if Neighborhood = "Sawyer" and LotFrontage = "NA" then LotFrontage = "74";

if Neighborhood = "SawyerW" and LotFrontage = "NA" then LotFrontage = "72";

if Neighborhood = "Somerst" and LotFrontage = "NA" then LotFrontage = "65";

if Neighborhood = "StoneBr" and LotFrontage = "NA" then LotFrontage = "63";

if Neighborhood = "SWISU" and LotFrontage = "NA" then LotFrontage = "59";

if Neighborhood = "Timber" and LotFrontage = "NA" then LotFrontage = "80";

if Neighborhood = "Veenker" and LotFrontage = "NA" then LotFrontage = "60";

if Alley = "NA" then Alley = "None";

if PoolQC = "NA" then PoolQC = "None";

if Fence = "NA" then Fence = "None";

if MiscFeature = "NA" then MiscFeature = "None";

if FireplaceQu = "NA" then FireplaceQu = "None";

**run**;

**data** example2;

set example;

LotFrontage2 = input(LotFrontage,**4.**);

**run**;

**proc** **print** data = example2;

**run**;

**data** example3;

set example2;

logprice = log(SalePrice);

**run**;

/\* forward selection \*/

ods graphics on;

**proc** **glmselect** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF \_1stFlrSF \_2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch \_3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold / selection = forward(choose = cv stop = cv) details=all stats=all;

**run**;

/\* looking pltos for assumptions \*/

**proc** **glm** data = example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd GarageCars WoodDeckSF ScreenPorch;

**run**;

/\* looking pltos for assumptions \*/

**proc** **reg** data = example2;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd GarageCars WoodDeckSF ScreenPorch / vif cli clm;

**run**;

/\* log trans \*/

**proc** **glm** data = example3 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model logprice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd GarageCars WoodDeckSF ScreenPorch;

**run**;

/\*backward selection \*/

ods graphics on;

**proc** **glmselect** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF \_1stFlrSF \_2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch \_3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold / selection = backward(choose = cv stop = cv) details=all stats=all;

**run**;

/\*check assumptions \*/

ods graphics on;

**proc** **glm** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars WoodDeckSF ScreenPorch;

**run**;

ods graphics on;

**proc** **reg** data=example2;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars WoodDeckSF ScreenPorch/vif cli clm;

**run**;

ods graphics on;

**proc** **glm** data=example3 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model logprice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars WoodDeckSF ScreenPorch;

**run**;

/\*stepwise selection \*/

ods graphics on;

**proc** **glmselect** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF \_1stFlrSF \_2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch \_3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold / selection = stepwise(choose = cv stop = cv) details=all stats=all;

**run**;

/\* check assumptions \*/

ods graphics on;

**proc** **glm** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BedroomAbvGr GarageCars;

**run**;

ods graphics on;

**proc** **reg** data=example2;

model SalePrice = MSSubClass LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BedroomAbvGr GarageCars / vif clm;

**run**;

ods graphics on;

**proc** **glm** data=example3 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model logprice = MSSubClass LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 GrLivArea BedroomAbvGr GarageCars;

**run**;

/\* Lasso model averaging) \*/

ods graphics on;

**proc** **glmselect** data=example2 plots=(EffectSelectPct ParmDistribution);

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF \_1stFlrSF \_2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF EnclosedPorch \_3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold / selection = lasso(adaptive stop=none choose=SBC);

modelAverage tables = (EffectSelectPct(all) ParmEst(all));

**run**;

/\* check assumptions from lasso \*/

ods graphics on;

**proc** **glm** data=example2 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars GarageArea WoodDeckSF ScreenPorch PoolArea GrLivArea;

**run**;

/\* transformation from lasso \*/

ods graphics on;

**proc** **glm** data=example3 plots=all;

class MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinType2 Heating HeatingQC CentralAir Electrical KitchenQual Functional FireplaceQu GarageType GarageFinish GarageQual GarageCond PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition;

model logprice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars GarageArea WoodDeckSF ScreenPorch PoolArea GrLivArea;

**run**;

ods graphics on;

**proc** **reg** data=example2;

model SalePrice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars GarageArea WoodDeckSF ScreenPorch PoolArea GrLivArea /vif clm;

**run**;

ods graphics on;

**proc** **reg** data=example3 plots(only label) =(CooksD RStudentbyLeverage);

model logprice = MSSubClass LotFrontage2 LotArea OverallQual OverallCond YearBuilt YearRemodAdd MasVnrArea BsmtFinSF1 \_1stFlrSF \_2ndFlrSF BsmtFullBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageCars GarageArea WoodDeckSF ScreenPorch PoolArea GrLivArea;

**run**;