/\* after you've imported the training and test sets \*/

/\*log transformation due to outliers \*/

data train5;

set work.log\_house\_price\_filter\_out\_train;

RUN;

data test5;

set work.log\_house\_price\_filter\_out\_train;

SalePrice = .;

run;

data projection5;

set train5 test5;

run;

proc glm data = projection5 plots = all;

class RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType;

model SalePrice = RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType LotArea BedroomAbvGr YearBuilt / cli;

output out = results p = Predict;

run;

/\* Can't have negative predictions because of RMSLE \*/

/\* Also must have only two columns with appropriate labels. \*/

data projection5\_results;

set results;

if Predict > 0 then Salerice = Predict;

if Predict < 0 then SalePrice = 100000;

keep id SalePrice;

where id> 1460;

;

proc means data = projection5\_results;

var SalePrice;

run;









