

```
/* Generated Code (IMPORT) */  
/* Source File: House Sales (Without Outliers).csv */  
/* Source Path: /folders/myfolders/sasuser.v94 */  
/* Code generated on: 4/7/20, 2:07 PM */
```

```
%web_drop_table(WORK.house_sales);
```

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/House Sales (Without Outliers).csv';
```

```
PROC IMPORT DATAFILE=REFFILE
```

```
    DBMS=CSV
```

```
    OUT=WORK.house_sales;
```

```
    GETNAMES=YES;
```

```
RUN;
```

```
PROC CONTENTS DATA=WORK.house_sales; RUN;
```

```
%web_open_table(WORK.house_sales);
```

```
ods noproctitle;
```

```
ods graphics / imagemap=on;
```

```
proc glmselect data=house_sales outdesign(addinputvars)=Work.reg_design
```

```
    plots=(criterionpanel);
```

```
model SalePrice=LotArea ShapeIR1 ShapeIR2 ShapeIR3 AllUtilities AgeOfHouse
```

```
    TotalBsmtSF GrLivArea FullBath KitchenAbvGr TotRmsAbvGrd GarageArea
```

```
    PoolArea / showpvalues
```

```
    selection=stepwise
```

```
    (slentry=0.05 slstay=0.05 select=sbc stop=s1) details=steps(anova  
    fitstats parmest);
```

```
run;
```

```
proc reg data=Work.reg_design alpha=0.05 plots(only)=(diagnostics residuals  
    observedbypredicted);
```

```
ods select DiagnosticsPanel ResidualPlot ObservedByPredicted;
```

```
    model SalePrice=&_GLSMOD /;  
run;  
quit;
```

```
proc delete data=Work.reg_design;  
run;
```

```
*Histogram of Target Variable: SalePrice;
```

```
proc sgplot data=House_sales;  
title "SalePrice";  
histogram SalePrice ;  
density SalePrice;  
density SalePrice / type=kernel;  
run;
```

```
*Univariate procedure on GarageArea variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on GarageCars";  
var GarageArea ;  
run;
```

```
*Univariate procedure on Age of house variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on Age of House";  
var AgeOfHouse ;  
run;
```

```
*Univariate procedure on GrLivArea variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on Ground Living Area";  
var GrLivArea ;  
run;
```

```
*QQ plot of residuals;
```

```
PROC REG DATA=house_sales plots(only)=qq;  
MODEL SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
RUN;
```

*Distribution of residues;

```
PROC REG data = house_sales;  
model SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
output out=diag p=pred r=resid;  
RUN;
```

*Check normality assumption;

```
PROC UNIVARIATE data=diag normal;  
var resid;  
histogram resid/normal(mu=est sigma=est);  
qqplot resid/normal;  
title 'Check for normality'  
RUN;
```

*Residual plot vs predictors;

```
PROC REG DATA=house_sales plots(only)=ResidualPlot;  
MODEL SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
RUN;
```

```
proc glmselect data=house_sales outdesign(addinputvars)=Work.reg_design  
    plots=(criterionpanel);  
    model SalePrice=LotArea ShapeIR1 ShapeIR2 ShapeIR3 AllUtilities AgeOfHouse  
        TotalBsmtSF GrLivArea FullBath KitchenAbvGr TotRmsAbvGrd GarageArea PoolArea / showpvalues  
    selection=stepwise
```

```
(slentry=0.05 slstay=0.05 select=sbc stop=sl) details=steps(anova  
    fitstats parmest);  
run;
```

```
proc reg data=house_sales;
```

```
model SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea / vif tol collin;  
title 'Multicollinearity Investigation of VIF and Tol';  
run;  
quit;
```

```
proc corr data=house_sales plots=matrix(histogram) PLOTS(MAXPOINTS=NONE);  
var SalePrice LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
run;
```

```
/* Generated Code (IMPORT) */  
/* Source File: House Sales (Without Outliers).csv */  
/* Source Path: /folders/myfolders/sasuser.v94 */  
/* Code generated on: 4/7/20, 2:07 PM */
```

```
%web_drop_table(WORK.house_sales);
```

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/House Sales (Without Outliers).csv';
```

```
PROC IMPORT DATAFILE=REFFILE
```

```
  DBMS=CSV
```

```
  OUT=WORK.house_sales;
```

```
  GETNAMES=YES;
```

```
RUN;
```

```
PROC CONTENTS DATA=WORK.house_sales; RUN;
```

```
%web_open_table(WORK.house_sales);
```

```
ods noproctitle;
```

```
ods graphics / imagemap=on;
```

```
proc glmselect data=house_sales outdesign(addinputvars)=Work.reg_design
```

```
  plots=(criterionpanel);
```

```
  model SalePrice=LotArea ShapeIR1 ShapeIR2 ShapeIR3 AllUtilities AgeOfHouse
```

```
    TotalBsmtSF GrLivArea FullBath KitchenAbvGr TotRmsAbvGrd GarageArea
```

```
    PoolArea / showpvalues
```

```
  selection=stepwise
```

```
  (slentry=0.05 slstay=0.05 select=sbc stop=s1) details=steps(anova  
    fitstats parmest);
```

```
run;
```

```
proc reg data=Work.reg_design alpha=0.05 plots(only)=(diagnostics residuals  
  observedbypredicted);
```

```
ods select DiagnosticsPanel ResidualPlot ObservedByPredicted;
```

```
    model SalePrice=&_GLSMOD /;  
    run;  
quit;
```

```
proc delete data=Work.reg_design;  
run;
```

```
*Histogram of Target Variable: SalePrice;
```

```
proc sgplot data=House_sales;  
title "SalePrice";  
histogram SalePrice ;  
density SalePrice;  
density SalePrice / type=kernel;  
run;
```

```
*Univariate procedure on GarageArea variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on GarageCars";  
var GarageArea ;  
run;
```

```
*Univariate procedure on Age of house variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on Age of House";  
var AgeOfHouse ;  
run;
```

```
*Univariate procedure on GrLivArea variable;
```

```
proc univariate data=House_sales modes;  
title "Univariate Analysis on Ground Living Area";  
var GrLivArea ;  
run;
```

```
*QQ plot of residuals;
```

```
PROC REG DATA=house_sales plots(only)=qq;  
MODEL SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
RUN;
```

*Distribution of residues;

```
PROC REG data = house_sales;  
model SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
output out=diag p=pred r=resid;  
RUN;
```

*Check normality assumption;

```
PROC UNIVARIATE data=diag normal;  
var resid;  
histogram resid/normal(mu=est sigma=est);  
qqplot resid/normal;  
title 'Check for normality'  
RUN;
```

*Residual plot vs predictors;

```
PROC REG DATA=house_sales plots(only)=ResidualPlot;  
MODEL SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;  
RUN;
```

*creating an interaction term

```
data house_sales;  
    set house_sales;  
    interactionterm= GrLivArea*GarageArea;  
run;
```

*regression with interaction term

```
proc glmselect data=house_sales plot=CriterionPanel;  
model SalePrice = LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea interactionterm  
/selection =forward(select=SL) stats = all;
```

```
/* Generated Code (IMPORT) */  
/* Source File: House Sales (Without Outliers).csv */  
/* Source Path: /folders/myfolders/sasuser.v94 */  
/* Code generated on: 4/7/20, 2:07 PM */
```

```
%web_drop_table(WORK.house_sales);
```

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/House Sales (Without Outliers).csv';
```

```
PROC IMPORT DATAFILE=REFFILE
```

```
    DBMS=CSV
```

```
    OUT=WORK.house_sales;
```

```
    GETNAMES=YES;
```

```
RUN;
```

```
PROC CONTENTS DATA=WORK.house_sales; RUN;
```

```
%web_open_table(WORK.house_sales);
```

```
ods noproctitle;
```

```
ods graphics / imagemap=on;
```

```
*CODE FOR CORRELATION MATRIX
```

```
PROC CORR DATA=house_sales PLOTS=MATRIX;
```

```
    VAR LotArea AgeOfHouse TotalBsmtSF GrLivArea GarageArea;
```

```
RUN;
```

```
QUIT;
```


Stata codes

```
regress saleprice lotarea ageofhouse totalbsmtsf grlivarea garagearea
```

```
quietly reg saleprice lotarea ageofhouse totalbsmtsf grlivarea garagearea
```

```
estat hettest
```

```
estat imtest, white
```

```
reg saleprice lotarea ageofhouse totalbsmtsf grlivarea garagearea, vce(robust)
```