/\*\* Import an XLSX file. \*\*/

PROC IMPORT DATAFILE="/folders/myfolders/sasuser.v94/house\_price.xlsx"

OUT=WORK.HOUSE\_PRICE

DBMS=XLSX

REPLACE;

RUN;

/\* Run Means Procedure to check for Missing Values \*/

proc means data=House\_price NMISS N;

RUN;

/\* identify data sets \*/

data House\_price\_train;

set House\_price;

RUN;

/\*log transformation due to outliers \*/

data log\_House\_price\_train;

set House\_price\_train;

logGrLivArea = log(GrLivArea);

logSalePrice = log(SalePrice);

RUN;

/\*exploratory data analysis\*/

proc univariate data=log\_House\_price\_train;

var SalePrice logSalePrice;

histogram SalePrice logSalePrice;

RUN;

proc univariate data=log\_House\_price\_train;

var GrLivArea logGrLivArea;

histogram GrLivArea logGrLivArea;

RUN;

/\* frequency table to show size of neighborhoods in dataset \*/

proc freq data= House\_price\_train;

tables Neighborhood

/plots=freqplot;

RUN;

/\*boxplots of saleprice by neighborhood\*/

proc sgplot data=House\_price\_train;

hbox SalePrice / category=Neighborhood;

RUN;

/\*boxplots of GrLivArea by neighborhood \*/

proc sgplot data= House\_price\_train;

hbox GrLivArea / category= Neighborhood;

RUN;

ods graphics on / attrpriority=none;

Title 'SalePrice by Neighborhood';

proc sgplot data=House\_price\_train;

scatter y=SalePrice x=GrLivArea /group= Neighborhood markerattrs=(size=10px);

styleattrs datasymbols=(circle square star);

RUN;

proc glm data=House\_price\_train plots =all;

class Neighborhood(ref = 'BrkSide');

model SalePrice = GrLivArea | Neighborhood /solution;

RUN;

/\* lm for log transformed data \*/

proc glm data=log\_House\_price\_train plots =all;

class Neighborhood(ref = 'BrkSide');

model logSalePrice = logGrLivArea | Neighborhood /solution;

RUN;

**The MEANS Procedure**

| **Variable** | **Label** | **N Miss** | **N** |
| --- | --- | --- | --- |
| Id  MSSubClass  LotFrontage  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  SalePrice | Id  MSSubClass  LotFrontage  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  SalePrice | 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 | 1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460  1460 |