# Housing Prices

## Load required libraries for models that we are testing

#### Load the data sets

This loads up the data sets dfTraining and dfAnalysis so we can use them to create models

```
rm(list = ls())
df <- read.csv("dfTrain1.csv")</pre>
```

#### **Decision Tree**

Create a decision tree model

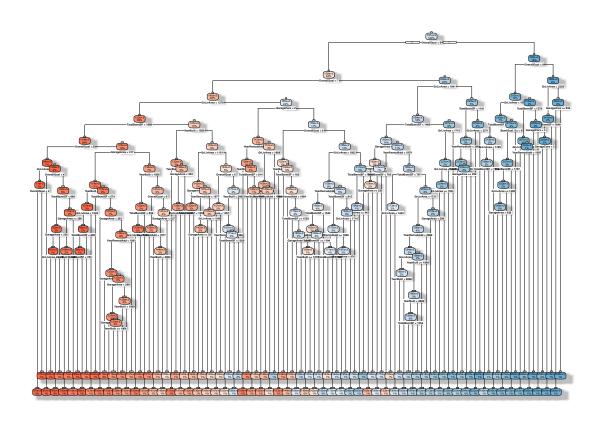
```
# Decision Tree

tree <- rpart(formula = SalePrice ~ OverallQual + YearBuilt + YearRemodAdd + TotalBsmtSF + GrLivArea +
```

Decision Tree plot

```
# Visualize the decision tree with rpart.plot
rpart.plot(tree, box.palette="RdBu", shadow.col="gray", nn=TRUE)
```

## Warning: labs do not fit even at cex 0.15, there may be some overplotting



### Linear Regression

Regression Model

```
# Linear Regression (using variables that are high correlation)
reg <- lm(formula = SalePrice ~ OverallQual + YearBuilt + YearRemodAdd + TotalBsmtSF + GrLivArea + Gara
# Model Performance
print(reg)
##
## lm(formula = SalePrice ~ OverallQual + YearBuilt + YearRemodAdd +
##
       TotalBsmtSF + GrLivArea + GarageCars + BsmtQual_Ex + KitchenQual_Ex,
##
       data = df)
##
## Coefficients:
                                        YearBuilt
                                                     YearRemodAdd
                                                                       TotalBsmtSF
##
      (Intercept)
                      OverallQual
##
       -1.004e+06
                        1.473e+04
                                        2.448e+02
                                                         2.442e+02
                                                                         2.231e+01
##
        GrLivArea
                       GarageCars
                                      BsmtQual_Ex KitchenQual_Ex
        4.999e+01
                        1.323e+04
                                        3.708e+04
                                                         3.593e+04
##
r2(reg)
```

```
## # R2 for Linear Regression
##
          R2: 0.802
##
     adj. R2: 0.801
model_performance(reg)
## # Indices of model performance
##
                                                    RMSE |
## AIC
                     BIC |
                              R2 | R2 (adj.) |
                                                                Sigma
## 34742.881 | 34795.743 | 0.802 |
                                       0.801 | 35328.983 | 35438.380
#Visualization of model checks
check_model(reg)
  Linearity
                                                Homogeneity of Variance
                                                Reference line should be flat and horizontal
  Reference line should be flat and horizontal
                                                residu
 (VIF)Residual
     -250000
     -500000
                                                Std.
             0e+00
                      2e+05
                               4e+05
                                        6e+05
                                                       0e+00
                                                                2e+05
                                                                         4e+05
                                                                                  6e+05
                       Fitted values
                                                                  Fitted values
  ©ollinearity
                                                Influential Observations
 'ariance Inflation
                                                Std.
          Bsmt Quad of 10 to 14 co 20 verted to Ballote SBR it mod
                                                       0.00
                                                                  0.05
                                                                            0.10
                                                                  Leverage (hii)
                      moderate (< 10)
           low (< 5)
                                         high (>
  Mormality of Residuals
                                                Normality of Residuals
 Distribution should be close to the normal curve
                                                Density
                                                               -250000
                             0
                                                     -500000
                                                                                   250000
           Standard Normal Distribution Quantile
                                                                   Residuals
```

#### **Neural Network**

Adding a neural network model

```
# Neural Network
nn <-nnet.formula(formula = SalePrice ~ OverallQual + YearBuilt + YearRemodAdd + TotalBsmtSF + GrLivAre</pre>
```

## # weights: 111

## initial value 56997365982750.078125 ## final value 9208736243659.042969

## converged