ST301\_A1\_Report

S16\_806

10/12/2020

Introductin

A real estate agent wants to develop a model to predict the sale price of a house using data collected from 5575 sold houses. we assumes that the following variables which may useful in predicting the sale price of a given house. 1. soldPrice - sold price of house 2. sqftLiving - square footage of living area 3. sqftLand - square footage of land 4. sqftAbove - square footage of area above ground 5. sqftBasement - square footage of basement 6. numBedRooms - number of bed rooms 7. numBathRooms - number of bath rooms 8. numFloors - number of ﬂoors 9. builtYear - year of construction 10. grade - construction quality ranked from 1 to 4 where 1 is the lowest grade 11. waterFront- whether the house has a waterfront (1) or not (0) 12. condition - condition of the house (Excellent, Good, Average)

Exploratory Analysis

housePrice=read.csv("F:\\study\_note\\3rd year\\ST301\\A\\housePrice.csv")

variable.names(housePrice)

## [1] "soldPrice" "sqftLiving" "sqftLand" "sqftAbove" "sqftBasement"  
## [6] "numBedRooms" "numBathRooms" "numFloors" "builtYear" "grade"   
## [11] "waterFront" "condition"

Forward Selection Method Iteration 1

summary(lm(soldPrice ~ sqftLiving, data = housePrice))$adj.r.squared

## [1] 0.4940643

summary(lm(soldPrice ~ sqftLand, data = housePrice))$adj.r.squared

## [1] 0.005532735

summary(lm(soldPrice ~ sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.4104951

summary(lm(soldPrice ~ sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.1352688

summary(lm(soldPrice ~ numBedRooms, data = housePrice))$adj.r.squared

## [1] 0.1101489

summary(lm(soldPrice ~ numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.31588

summary(lm(soldPrice ~ numFloors, data = housePrice))$adj.r.squared

## [1] 0.1304164

summary(lm(soldPrice ~ builtYear, data = housePrice))$adj.r.squared

## [1] 0.01449347

summary(lm(soldPrice ~ grade, data = housePrice))$adj.r.squared

## [1] 0.2620642

summary(lm(soldPrice ~ waterFront, data = housePrice))$adj.r.squared

## [1] 0.03628621

summary(lm(soldPrice ~ condition, data = housePrice))$adj.r.squared

## [1] 0.002325881

we can add sqftLiving Iteration 2

summary(lm(soldPrice ~ sqftLiving+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.4950925

summary(lm(soldPrice ~ sqftLiving+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.4955874

summary(lm(soldPrice ~ sqftLiving+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.4955874

summary(lm(soldPrice ~ sqftLiving+numBedRooms, data = housePrice))$adj.r.squared

## [1] 0.5164011

summary(lm(soldPrice ~ sqftLiving+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.4939742

summary(lm(soldPrice ~ sqftLiving+numFloors, data = housePrice))$adj.r.squared

## [1] 0.4970073

summary(lm(soldPrice ~ sqftLiving+builtYear, data = housePrice))$adj.r.squared

## [1] 0.5205632

summary(lm(soldPrice ~ sqftLiving+grade, data = housePrice))$adj.r.squared

## [1] 0.5094433

summary(lm(soldPrice ~ sqftLiving+waterFront, data = housePrice))$adj.r.squared

## [1] 0.5151876

summary(lm(soldPrice ~ sqftLiving+condition, data = housePrice))$adj.r.squared

## [1] 0.495239

we can add builtYear Iteration 3

summary(lm(soldPrice ~ sqftLiving+builtYear+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5214174

summary(lm(soldPrice ~ sqftLiving+builtYear+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.5268141

summary(lm(soldPrice ~ sqftLiving+builtYear+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5268141

summary(lm(soldPrice ~ sqftLiving+builtYear+numBedRooms, data = housePrice))$adj.r.squared

## [1] 0.5377418

summary(lm(soldPrice ~ sqftLiving+builtYear+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.5241533

summary(lm(soldPrice ~ sqftLiving++builtYear+numFloors, data = housePrice))$adj.r.squared

## [1] 0.5299091

summary(lm(soldPrice ~ sqftLiving++builtYear+grade, data = housePrice))$adj.r.squared

## [1] 0.5485533

summary(lm(soldPrice ~ sqftLiving+builtYear+waterFront, data = housePrice))$adj.r.squared

## [1] 0.54023

summary(lm(soldPrice ~ sqftLiving+builtYear+condition, data = housePrice))$adj.r.squared

## [1] 0.5204939

we can add grade Iteration 4

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5488577

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.5536932

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5536932

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+numBedRooms, data = housePrice))$adj.r.squared

## [1] 0.5657548

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.5517589

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+numFloors, data = housePrice))$adj.r.squared

## [1] 0.5568888

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront, data = housePrice))$adj.r.squared

## [1] 0.5679138

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+condition, data = housePrice))$adj.r.squared

## [1] 0.5485626

we can add waterFront Iteration 5

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5682216

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.5728255

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5728255

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms, data = housePrice))$adj.r.squared

## [1] 0.5829823

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.5708217

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numFloors, data = housePrice))$adj.r.squared

## [1] 0.5753253

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+condition, data = housePrice))$adj.r.squared

## [1] 0.5679375

we can add numBedRooms Iteration 6

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5836906

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.58662

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.58662

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.5888093

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors, data = housePrice))$adj.r.squared

## [1] 0.5893767

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+condition, data = housePrice))$adj.r.squared

## [1] 0.5831776

we can add numFloors Iteration 7

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5899195

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.5907356

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5907356

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms, data = housePrice))$adj.r.squared

## [1] 0.5924343

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+condition, data = housePrice))$adj.r.squared

## [1] 0.5897794

we can add numBathRooms Iteration 8

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.592857

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove, data = housePrice))$adj.r.squared

## [1] 0.5942811

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5942811

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+condition, data = housePrice))$adj.r.squared

## [1] 0.5926938

we can add sqft Iteration 9

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+sqftLand, data = housePrice))$adj.r.squared

## [1] 0.5949049

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5942811

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+condition, data = housePrice))$adj.r.squared

## [1] 0.5946543

Iteration 10

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+sqftLand+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5949049

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+sqftLand+condition, data = housePrice))$adj.r.squared

## [1] 0.5952841

Iteration 11

summary(lm(soldPrice ~ sqftLiving+builtYear+grade+waterFront+numBedRooms+numFloors+numBathRooms+sqftAbove+sqftLand+condition+sqftBasement, data = housePrice))$adj.r.squared

## [1] 0.5952841

Model Fitting

We can take as a final fitted model from iteration 11.

Discussion and Conclusion

we can plot this model what we selected(below).summary 65 and summary 66 are got same value 0.5952841. if we go to predicting sales prices, All coefficient of the model are non negative then we can say that predicting values are always increasing from some unit.

plot(c(1,2,3,4,5,6,7,8,9,10,11),c(0.4940643,0.5205632,0.5485533,0.5679138,0.5829823,0.5893767,0.5924343,0.5942811,0.5949049,0.5952841,0.5952841))

