ST 516: Midterm Project

Due by 11:55pm, Thursday, October 7

Problem Statement:

Your team has been hired by the Ames, Iowa association of realtors to do an analysis of some historical data to understand if there are any differences the drivers of housing prices today and in the period before the housing price crash that occurred at the end of the 2000's. The first task they want you to perform is an analysis of housing data from the 2000'sunderstanding attributes of houses that will influence purchase price. You performance on this task will determine if you are hired to take on a follow on project to do a similar analysis for more current housing data.

Your team is tasked with creating a predictive model using this data. Your client has told you to prioritize accurate housing price predictions – this will be a trial to see how well you might be able to predict current house prices -- but also stressed it would be very valuable to identify a handful of attributes that have the most significant association with house prices, as well as have an estimate of the quantitative relationship between those attributes and price.

General Instructions:

- You will work in teams of four +/- 1 team members.
- Your team will submit one written report no more than 10 pages in length in the following format:
    - Executive summary of your findings
    - Introduction (explain what you are trying to learn and why) o Data (description of the data and any sources)
    - Methods (describe the modeling and error estimation methods you used and why you chose them; this is also the section to discuss results of diagnostics and steps needed to correct problems, if any)
    - Results (discuss how well your model model(s) fit the data and enables you to interpret the relationship between individual predictors and response; if you feel it is not possible to produce a reliably interpretable model that has accuracy that is competitive with all other models, be sure to state why this is the case).
- One team member should submit a final report and accompanying R code using the following format o Report: pdf named in the format 'lastname_firstname_Proj1.pdf' o Code: R script named 'lastname_firstname_Proj1.r'.

Housing Data Description

Number of instances (observations): 1437

Response: Sales Price

Number of Attributes (predictor variables): 25
Attribute breakdown: 15 quantitative predictor variables and 10 categorical predictor variables. Variable descriptions:

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

LotShape: General shape of property

Reg	Regular
IR1	Slightly irregular
IR2	Moderately Irregular
IR3	Irregular

LotConfig: Lot configuration

Inside	Inside lot
Corner	Corner lot
CulDSac	Cul-de-sac
FR2	Frontage on 2 sides of property
FR3	Frontage on 3 sides of property

Neighborhood: Physical locations within Ames city limits

Blmngtn	Bloomington Heights
Blueste	Bluestem
BrDale	Briardale
BrkSide	Brookside
ClearCr	Clear Creek
CollgCr	College Creek
Crawfor	Crawford
Edwards	Edwards

Gilbert	Gilbert
IDOTRR	Iowa DOT and Rail Road
MeadowV	Meadow Village
Mitchel	Mitchell
Names	North Ames
NoRidge	Northridge
NPkVill	Northpark Villa
NridgHt	Northridge Heights
NWAmes	Northwest Ames
OldTown	Old Town
SWISU	South & West of Iowa State University
Sawyer	Sawyer
SawyerW	Sawyer West
Somerst	Somerset
StoneBr	Stone Brook
Timber	Timberland
Veenker	Veenker

BldgType: Type of dwelling

1Fam	Single-family Detached
2FmCon	Two-family Conversion; originally built as one-family dwelling
Duplx	Duplex
TwnhsE	Townhouse End Unit
TwnhsI	Townhouse Inside Unit

HouseStyle: Style of dwelling

1Story	One story
1.5Fin	One and one-half story: 2nd level finished
1.5Unf	One and one-half story: 2nd level unfinished
2Story	Two story
2.5Fin	Two and one-half story: 2nd level finished
2.5Unf	Two and one-half story: 2nd level unfinished
SFoyer	Split Foyer
SLvl	Split Level

OverallQual: Rates the overall material and finish of the house

10	Very Excellent
9	Excellent
8	Very Good
7	Good

6 Above Average
5 Average
4 Below Average
3 Fair
2 Poor
1 Very Poor

OverallCond: Rates the overall condition of the house

10 Very Excellent
9 Excellent
8 Very Good
7 Good
6 Above Average
5 Average
4 Below Average
3 Fair
2 Poor
1 Very Poor

YearBuilt: Original construction date

YearRemodAdd: Remodel date (same as construction date if no remodeling or additions)

Exterior1st: Exterior covering on house

AsbShng	Asbestos Shingles
AsphShn	Asphalt Shingles
BrkComm	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block
CemntBd	Cement Board
HdBoard	Hard Board
ImStucc	Imitation Stucco
MetalSd	Metal Siding
Other	Other
Plywood	Plywood
PreCast	PreCast
Stone	Stone
Stucco	Stucco
VinylSd	Vinyl Siding
Wd Sdng	Wood Siding
WdShing	Wood Shingles

Foundation: Type of foundation

BrkTil	Brick & Tile
CBlock	Cinder Block
PConc	Poured Contrete
SlabSlab	
Stone	Stone
Wood	Wood

BsmtFinType1: Rating of basement finished area

GLQ	Good Living Quarters
ALQ	Average Living Quarters
BLQ	Below Average Living Quarters
Rec	Average Rec Room
LwQ	Low Quality
Unf	Unfinshed
NA	No Basement

TotalBsmtSF: Total square feet of basement area

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

BedroomAbvGr: Bedrooms above grade (does NOT include basement bedrooms)

KitchenQual: Kitchen quality

Ex	Excellent
Gd	Good
TA	Typical/Average
Fa	Fair
Po	Poor

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Fireplaces: Number of fireplaces

GarageType: Garage location

2Types	More than one type of garage

Attchd	Attached to home
Basment	Basement Garage
BuiltIn	Built-In (Garage part of house - typically has room above garage)
CarPort	Car Port
Detchd	Detached from home
NA	No Garage

GarageCars: Size of garage in car capacity

YrSold: Year Sold (YYYY)

SalePrice: Sales price in $