Assignment	**2020-21**
Title	**Ames Iowa Housing dataset**
Training data file	**ames_iowa_housing_##. csv**
Test data file	**ames_iowa_housing_test.csv**

The data of this assignment refer to the database of the Ames City Assessor's Office. It includes a large number of variables and observations within the data set and they refer to 2930 property sales that had occurred in Ames, Iowa between 2006 and 2010. A mix of 82 nominal, ordinal, continuous, and discrete variables were used in the calculation of the assessed values. The dataset included physical property measurements in addition to computation variables used in the city's assessment process. All variables focus on the quality and quantity of many physical attributes of the property. Most of the variables are exactly the type of information that a typical home buyer would want to know about a potential property (e.g. When was it built? How big is the lot? How many feet of living space is in the dwelling? Is the basement finished? How many bathrooms are there?). For more details see at De Cock (2011).

Each student will receive a random sub-sample of 1500 observations to use it for training their model and for inference. All students will use a common evaluation/test dataset of 500 observations.

1. You should first do some exploratory data analysis. Visualizing the data should give you some insight into certain particularities of this dataset. Pairwise comparisons will help you also learn about the association implied by the data.
2. The main aim is to identify the best model for predicting the prices of the properties. Select the appropriate features to predict your model. Be careful, your model should not be over-parameterized.
3. Check the assumptions of the model and revise your procedure
4. Use leave-one-out and 10-fold cross-validation to select your model and assess the out-of-sample predictive ability of the model.
5. Use the test dataset to select your model and assess the out-of-sample predictive ability of the model.
6. Compare results obtained by different methods under 2, 3 and 4.
7. Select your final model and features and justify your choice.
8. Interpret the parameters and the predicting performance of the final model.

9. Describe the typical profile of a property sale.
10. Write a report summarizing your results (see attached directions for this)

A detailed list of the variables follows:

NAME: AmesHousing.txt
SIZE: 2930 observations, 82 variables
ARTICLE TITLE: Ames Iowa: Alternative to the Boston Housing Data Set

DESCRIPTIVE ABSTRACT: Data set contains information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

SOURCES:
Ames, Iowa Assessor's Office

VARIABLE DESCRIPTIONS:
Tab characters are used to separate variables in the data file. The data has 82 columns which include 23 nominal, 23 ordinal, 14 discrete, and 20 continuous variables (and 2 additional observation identifiers).

1) Order (Discrete): Observation number
2) PID (Nominal): Parcel identification number - can be used with city web site for parcel review.
3) MS SubClass (Nominal): Identifies the type of dwelling involved in the sale.
    a) 020 1-STORY 1946 & NEWER ALL STYLES
    b) 030 1-STORY 1945 & OLDER
    c) 040 1-STORY W/FINISHED ATTIC ALL AGES
    d) 045 1-1/2 STORY - UNFINISHED ALL AGES
    e) 050 1-1/2 STORY FINISHED ALL AGES
    f) 060 2-STORY 1946 & NEWER
    g) 070 2-STORY 1945 & OLDER
    h) 075 2-1/2 STORY ALL AGES
    i) 080 SPLIT OR MULTI-LEVEL
    j) 085 SPLIT FOYER
    k) 090 DUPLEX - ALL STYLES AND AGES
    l) 120 1-STORY PUD (Planned Unit Development) - 1946 & NEWER
    m) 150 1-1/2 STORY PUD - ALL AGES
    n) 160 2-STORY PUD - 1946 & NEWER
    o) 180 PUD - MULTILEVEL - INCL SPLIT LEV/FOYER
    p) 190 2 FAMILY CONVERSION - ALL STYLES AND AGES
4) MS Zoning (Nominal): Identifies the general zoning classification of the sale.
    a) A Agriculture
    b) C Commercial
    c) FV Floating Village Residential
    d) I Industrial
    e) RH Residential High Density
    f) RL Residential Low Density
    g) RP Residential Low Density Park
    h) RM Residential Medium Density
5) Lot Frontage (Continuous): Linear feet of street connected to property

6) Lot Area (Continuous): Lot size in square feet
7) Street (Nominal): Type of road access to property
    a) Grvl Gravel
    b) Pave Paved
8) Alley (Nominal): Type of alley access to property
    a) Grvl Gravel
    b) Pave Paved
    c) NA No alley access
9) Lot Shape (Ordinal): General shape of property
    a) Reg Regular
    b) IR1 Slightly irregular
    c) IR2 Moderately Irregular
    d) IR3 Irregular
10) Land Contour (Nominal): Flatness of the property
    a) Lvl Near Flat/Level
    b) Bnk Banked - Quick and significant rise from street grade to building
    c) HLS Hillside - Significant slope from side to side
    d) Low Depression
11) Utilities (Ordinal): Type of utilities available
    a) AllPub All public Utilities (E,G,W,& S)
    b) NoSewr Electricity, Gas, and Water (Septic Tank)
    c) NoSeWa Electricity and Gas Only
    d) ELO Electricity only
12) Lot Config (Nominal): Lot configuration
    a) Inside Inside lot
    b) Corner Corner lot
    c) CulDSac Cul-de-sac
    d) FR2 Frontage on 2 sides of property
    e) FR3 Frontage on 3 sides of property
13) Land Slope (Ordinal): Slope of property
    a) Gtl Gentle slope
    b) Mod Moderate Slope
    c) Sev Severe Slope
14) Neighborhood (Nominal): Physical locations within Ames city limits (map available)
    a) Blmngtn Bloomington Heights
    b) Blueste Bluestem
    c) BrDale Briardale
    d) BrkSide Brookside
    e) ClearCr Clear Creek
    f) CollgCr College Creek
    g) Crawfor Crawford
    h) Edwards Edwards
    i) Gilbert Gilbert
    j) Greens Greens
    k) GrnHill Green Hills
    l) IDOTRR Iowa DOT and Rail Road
    m) Landmrk Landmark
    n) MeadowV Meadow Village
    o) Mitchel Mitchell
    p) Names North Ames
    q) NoRidge Northridge
    r) NPkVill Northpark Villa
    s) NridgHt Northridge Heights
    t) NWAmes Northwest Ames
    u) OldTown Old Town
    v) SWISU South & West of Iowa State University
    w) Sawyer Sawyer
    x) SawyerW Sawyer West
    y) Somerst Somerset
    z) StoneBr Stone Brook
    aa) Timber Timberland

bb) Veenker Veenker

15) Condition 1 (Nominal): Proximity to various conditions
    a) Artery Adjacent to arterial street
    b) Feedr Adjacent to feeder street
    c) Norm Normal
    d) RRNn Within 200' of North-South Railroad
    e) RRAn Adjacent to North-South Railroad
    f) PosN Near positive off-site feature--park, greenbelt, etc.
    g) PosA Adjacent to postive off-site feature
    h) RRNe Within 200' of East-West Railroad
    i) RRAe Adjacent to East-West Railroad
16) Condition 2 (Nominal): Proximity to various conditions (if more than one is present)
    a) Artery Adjacent to arterial street
    b) Feedr Adjacent to feeder street
    c) Norm Normal
    d) RRNn Within 200' of North-South Railroad
    e) RRAn Adjacent to North-South Railroad
    f) PosN Near positive off-site feature--park, greenbelt, etc.
    g) PosA Adjacent to postive off-site feature
    h) RRNe Within 200' of East-West Railroad
    i) RRAe Adjacent to East-West Railroad
17) Bldg Type (Nominal): Type of dwelling
    a) 1Fam Single-family Detached
    b) 2FmCon Two-family Conversion; originally built as one-family dwelling
    c) Duplx Duplex
    d) TwnhsE Townhouse End Unit
    e) TwnhsI Townhouse Inside Unit
18) House Style (Nominal): Style of dwelling
    a) 1Story One story
    b) 1.5Fin One and one-half story: 2nd level finished
    c) 1.5Unf One and one-half story: 2nd level unfinished
    d) 2Story Two story
    e) 2.5Fin Two and one-half story: 2nd level finished
    f) 2.5Unf Two and one-half story: 2nd level unfinished
    g) SFoyer Split Foyer
    h) SLvl Split Level
19) Overall Qual (Ordinal): Rates the overall material and finish of the house
    a) 10 Very Excellent
    b) 9 Excellent
    c) 8 Very Good
    d) 7 Good
    e) 6 Above Average
    f) 5 Average
    g) 4 Below Average
    h) 3 Fair
    i) 2 Poor
20) Very Poor
21) Overall Cond (Ordinal): Rates the overall condition of the house
    a) 10 Very Excellent
    b) 9 Excellent
    c) 8 Very Good
    d) 7 Good
    e) 6 Above Average
    f) 5 Average
    g) 4 Below Average
22) Fair
23) Poor
24) Very Poor
25) Year Built (Discrete): Original construction date
26) Year Remod/Add (Discrete): Remodel date (same as construction date if no remodeling or additions)
27) Roof Style (Nominal): Type of roof

a) Flat Flat
b) Gable Gable
c) Gambrel Gabrel (Barn)
d) Hip Hip
e) Mansard Mansard
f) Shed Shed

28) Roof Matl (Nominal): Roof material
a) ClyTile Clay or Tile
b) CompShg Standard (Composite) Shingle
c) Membran Membrane
d) Metal Metal
e) Roll Roll
f) Tar&Grv Gravel & Tar
g) WdShake Wood Shakes
h) WdShngl Wood Shingles

29) Exterior 1 (Nominal): Exterior covering on house
a) AsbShng Asbestos Shingles
b) AsphShn Asphalt Shingles
c) BrkComm Brick Common
d) BrkFace Brick Face
e) CBlock Cinder Block
f) CemntBd Cement Board
g) HdBoard Hard Board
h) ImStucc Imitation Stucco
i) MetalSd Metal Siding
j) Other Other
k) Plywood Plywood
l) PreCast PreCast
m) Stone Stone
n) Stucco Stucco
o) VinylSd Vinyl Siding
p) Wd Sdng Wood Siding
q) WdShing Wood Shingles

30) Exterior 2 (Nominal): Exterior covering on house (if more than one material)
a) AsbShng Asbestos Shingles
b) AsphShn Asphalt Shingles
c) BrkComm Brick Common
d) BrkFace Brick Face
e) CBlock Cinder Block
f) CemntBd Cement Board
g) HdBoard Hard Board
h) ImStucc Imitation Stucco
i) MetalSd Metal Siding
j) Other Other
k) Plywood Plywood
l) PreCast PreCast
m) Stone Stone
n) Stucco Stucco
o) VinylSd Vinyl Siding
p) Wd Sdng Wood Siding
q) WdShing Wood Shingles

31) Mas Vnr Type (Nominal): Masonry veneer type
a) BrkCmn Brick Common
b) BrkFace Brick Face
c) CBlock Cinder Block
d) None None
e) Stone Stone

32) Mas Vnr Area (Continuous): Masonry veneer area in square feet

33) Exter Qual (Ordinal): Evaluates the quality of the material on the exterior
a) Ex Excellent
b) Gd Good

c) TA Average/Typical
d) Fa Fair
e) Po Poor

34) Exter Cond (Ordinal): Evaluates the present condition of the material on the exterior
a) Ex Excellent
b) Gd Good
c) TA Average/Typical
d) Fa Fair
e) Po Poor

35) Foundation (Nominal): Type of foundation
a) BrkTil Brick & Tile
b) CBlock Cinder Block
c) PConc Poured Contrete
d) Slab Slab
e) Stone Stone
f) Wood Wood

36) Bsmt Qual (Ordinal): Evaluates the height of the basement
a) Ex Excellent (100+ inches)
b) Gd Good (90-99 inches)
c) TA Typical (80-89 inches)
d) Fa Fair (70-79 inches)
e) Po Poor (<70 inches
f) NA No Basement

37) Bsmt Cond (Ordinal): Evaluates the general condition of the basement
a) Ex Excellent
b) Gd Good
c) TA Typical - slight dampness allowed
d) Fa Fair - dampness or some cracking or settling
e) Po Poor - Severe cracking, settling, or wetness
f) NA No Basement

38) Bsmt Exposure (Ordinal): Refers to walkout or garden level walls
a) Gd Good Exposure
b) Av Average Exposure (split levels or foyers typically score average or above)
c) Mn Mimimum Exposure
d) No No Exposure
e) NA No Basement

39) BsmtFin Type 1 (Ordinal): Rating of basement finished area
a) GLQ Good Living Quarters
b) ALQ Average Living Quarters
c) BLQ Below Average Living Quarters
d) Rec Average Rec Room
e) LwQ Low Quality
f) Unf Unfinshed
g) NA No Basement

40) BsmtFin SF 1 (Continuous): Type 1 finished square feet

41) BsmtFinType 2 (Ordinal): Rating of basement finished area (if multiple types)
a) GLQ Good Living Quarters
b) ALQ Average Living Quarters
c) BLQ Below Average Living Quarters
d) Rec Average Rec Room
e) LwQ Low Quality
f) Unf Unfinshed
g) NA No Basement

42) BsmtFin SF 2 (Continuous): Type 2 finished square feet

43) Bsmt Unf SF (Continuous): Unfinished square feet of basement area

44) Total Bsmt SF (Continuous): Total square feet of basement area

45) Heating (Nominal): Type of heating
a) Floor Floor Furnace
b) GasA Gas forced warm air furnace
c) GasW Gas hot water or steam heat
d) Grav Gravity furnace

e) OthW Hot water or steam heat other than gas
f) Wall Wall furnace

46) HeatingQC (Ordinal): Heating quality and condition
a) Ex Excellent
b) Gd Good
c) TA Average/Typical
d) Fa Fair
e) Po Poor

47) Central Air (Nominal): Central air conditioning
a) N No
b) Y Yes

48) Electrical (Ordinal): Electrical system
a) SBrkr Standard Circuit Breakers & Romex
b) FuseA Fuse Box over 60 AMP and all Romex wiring (Average)
c) FuseF 60 AMP Fuse Box and mostly Romex wiring (Fair)
d) FuseP 60 AMP Fuse Box and mostly knob & tube wiring (poor)
e) Mix Mixed

49) 1st Flr SF (Continuous): First Floor square feet
50) 2nd Flr SF (Continuous) : Second floor square feet
51) Low Qual Fin SF (Continuous): Low quality finished square feet (all floors)
52) Gr Liv Area (Continuous): Above grade (ground) living area square feet
53) Bsmt Full Bath (Discrete): Basement full bathrooms
54) Bsmt Half Bath (Discrete): Basement half bathrooms
55) Full Bath (Discrete): Full bathrooms above grade
56) Half Bath (Discrete): Half baths above grade
57) Bedroom (Discrete): Bedrooms above grade (does NOT include basement bedrooms)
58) Kitchen (Discrete): Kitchens above grade
59) KitchenQual (Ordinal): Kitchen quality
a) Ex Excellent
b) Gd Good
c) TA Typical/Average
d) Fa Fair
e) Po Poor

60) TotRmsAbvGrd (Discrete): Total rooms above grade (does not include bathrooms)
61) Functional (Ordinal): Home functionality (Assume typical unless deductions are warranted)
a) Typ Typical Functionality
b) Min1 Minor Deductions 1
c) Min2 Minor Deductions 2
d) Mod Moderate Deductions
e) Maj1 Major Deductions 1
f) Maj2 Major Deductions 2
g) Sev Severely Damaged
h) Sal Salvage only

62) Fireplaces (Discrete): Number of fireplaces
63) FireplaceQu (Ordinal): Fireplace quality
a) Ex Excellent - Exceptional Masonry Fireplace
b) Gd Good - Masonry Fireplace in main level
c) TA Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement
d) Fa Fair - Prefabricated Fireplace in basement
e) Po Poor - Ben Franklin Stove
f) NA No Fireplace

64) Garage Type (Nominal): Garage location
a) 2Types More than one type of garage
b) Attchd Attached to home
c) Basment Basement Garage
d) BuiltIn Built-In (Garage part of house - typically has room above garage)
e) CarPort Car Port
f) Detchd Detached from home
g) NA No Garage

65) Garage Yr Blt (Discrete): Year garage was built

66) Garage Finish (Ordinal) : Interior finish of the garage
    a) Fin Finished
    b) RFn Rough Finished
    c) Unf Unfinished
    d) NA No Garage
67) Garage Cars (Discrete): Size of garage in car capacity
68) Garage Area (Continuous): Size of garage in square feet
69) Garage Qual (Ordinal): Garage quality
    a) Ex Excellent
    b) Gd Good
    c) TA Typical/Average
    d) Fa Fair
    e) Po Poor
    f) NA No Garage
70) Garage Cond (Ordinal): Garage condition
    a) Ex Excellent
    b) Gd Good
    c) TA Typical/Average
    d) Fa Fair
    e) Po Poor
    f) NA No Garage
71) Paved Drive (Ordinal): Paved driveway
    a) Y Paved
    b) P Partial Pavement
    c) N Dirt/Gravel
72) Wood Deck SF (Continuous): Wood deck area in square feet
73) Open Porch SF (Continuous): Open porch area in square feet
74) Enclosed Porch (Continuous): Enclosed porch area in square feet
75) 3-Ssn Porch (Continuous): Three season porch area in square feet
76) Screen Porch (Continuous): Screen porch area in square feet
77) Pool Area (Continuous): Pool area in square feet
78) Pool QC (Ordinal): Pool quality
    a) Ex Excellent
    b) Gd Good
    c) TA Average/Typical
    d) Fa Fair
    e) NA No Pool
79) Fence (Ordinal): Fence quality
    a) GdPrv Good Privacy
    b) MnPrv Minimum Privacy
    c) GdWo Good Wood
    d) MnWw Minimum Wood/Wire
    e) NA No Fence
80) Misc Feature (Nominal): Miscellaneous feature not covered in other categories
    a) Elev Elevator
    b) Gar2 2nd Garage (if not described in garage section)
    c) Othr Other
    d) Shed Shed (over 100 SF)
    e) TenC Tennis Court
    f) NA None
81) Misc Val (Continuous): $Value of miscellaneous feature
82) Mo Sold (Discrete): Month Sold (MM)
83) Yr Sold (Discrete): Year Sold (YYYY)
84) Sale Type (Nominal): Type of sale
    a) WD Warranty Deed - Conventional
    b) CWD Warranty Deed - Cash
    c) VWD Warranty Deed - VA Loan
    d) New Home just constructed and sold
    e) COD Court Officer Deed/Estate
    f) Con Contract 15% Down payment regular terms
    g) ConLw Contract Low Down payment and low interest

h) ConLI Contract Low Interest
i) ConLD Contract Low Down
j) Oth Other

85) Sale Condition (Nominal): Condition of sale
a) Normal Normal Sale
b) Abnorml Abnormal Sale - trade, foreclosure, short sale
c) AdjLand Adjoining Land Purchase
d) Alloca Allocation - two linked properties with separate deeds, typically condo with a garage unit
e) Family Sale between family members
f) Partial Home was not completed when last assessed (associated with New Homes)

86) SalePrice (Continuous): Sale price $$

STORY BEHIND THE DATA:

This data set was constructed for the purpose of an end of semester project for an undergraduate regression course. The original data (obtained directly from the Ames Assessor's Office) is used for tax assessment purposes but lends itself directly to the prediction of home selling prices. The type of information contained in the data is similar to what a typical home buyer would want to know before making a purchase and students should find most variables straightforward and understandable.

PEDAGOGICAL NOTES:

Outside of the general issues associated with multiple regression discussed in this article, this particular data set offers several opportunities to discuss how the purpose of a model might affect the type of modeling done. User of this data may also want to review another JSE article related directly to real estate pricing:

> *Pardoe , I. (2008), "Modeling home prices using realtor data", Journal of Statistics Education Volume 16, Number 2 (2008).*