Kaggle: House Prices with Advanced Regression Techniques

Trevor Isaacson, Hawas Alsadeed, Evan Kessler
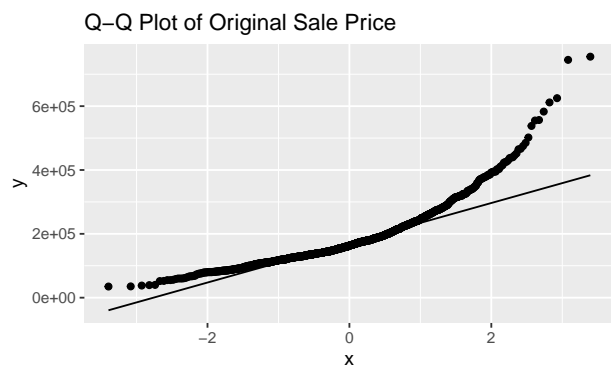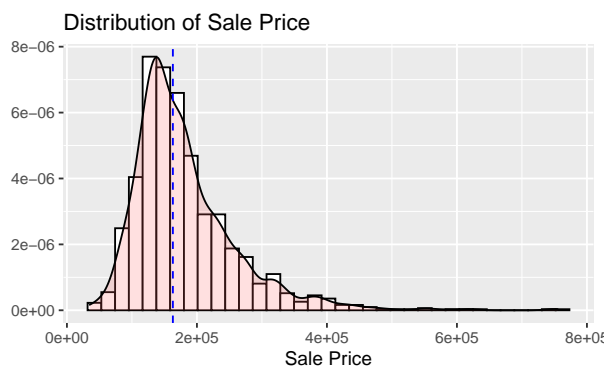
2/28/2022

Introduction

Imagine you're looking to purchase a home in the near future. How do you know the price of that home is accurate and fair. What factors or variables are you looking at to gauge the final price? For many, home prices are a number associated with a few key variables. For example, many buyers are looking at the number of bedrooms, bathrooms and the color of the fence. What other things affect the price of a home? Location, building materials, condition and quality are just a few that could potentially impact the price of a house. In this Kaggle competition, we look at home prices in Ames, Iowa and over 80 variables associated with those homes.
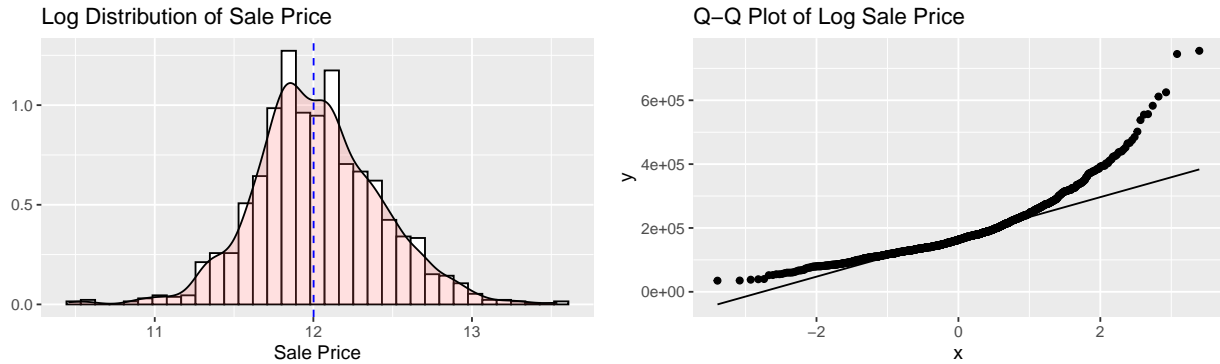
Problem Discription

Can we accurately predict the price of a home?

Data Cleaning

Because the original sales price data is very heavily skewed, we needed to log transform the prices. As shown in the histogram, we have a very heavy right tail because there are a some listings with very high prices compared to the median price of $180,921 (blue line). This non-normal shape and distribution is clearly evident in the Q-Q plot. The observations clearly curve away from the line depicting a large amount of skew.

Log Distribution of Sale Price — **Q–Q Plot of Log Sale Price**

Multiple Linear Regression

```
##
## Call:
## lm(formula = SalePrice ~ ., data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01183 -0.05998  0.00040  0.05940  0.43111
##
## Coefficients: (10 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.853e+00  7.052e+00   1.114 0.265755
## MSSubClass          -8.125e-05  6.619e-04  -0.123 0.902334
## MSZoningFV           4.374e-01  7.472e-02   5.854 6.89e-09 ***
## MSZoningRH           4.659e-01  7.056e-02   6.603 7.18e-11 ***
## MSZoningRL           4.465e-01  6.094e-02   7.327 5.57e-13 ***
## MSZoningRM           3.744e-01  6.121e-02   6.117 1.46e-09 ***
## LotFrontage         -3.353e-04  1.474e-04  -2.274 0.023204 *
## LotArea              1.517e-06  8.388e-07   1.808 0.070937 .
## StreetPave           1.821e-01  9.531e-02   1.911 0.056374 .
## LotShapeIR2          3.557e-02  2.724e-02   1.305 0.192091
## LotShapeIR3         -1.318e-01  5.725e-02  -2.303 0.021531 *
## LotShapeReg          1.230e-02  1.043e-02   1.180 0.238472
## LandContourHLS       1.285e-01  3.231e-02   3.976 7.63e-05 ***
## LandContourLow       5.984e-02  4.240e-02   1.412 0.158470
## LandContourLvl       6.877e-02  2.263e-02   3.039 0.002451 **
## LotConfigCulDSac     3.455e-02  2.106e-02   1.641 0.101251
## LotConfigFR2        -6.027e-02  2.637e-02  -2.285 0.022536 *
## LotConfigFR3        -1.568e-01  1.303e-01  -1.204 0.228933
## LotConfigInside     -1.431e-02  1.151e-02  -1.242 0.214433
## LandSlopeMod         3.979e-02  2.609e-02   1.525 0.127604
## LandSlopeSev        -6.532e-02  7.007e-02  -0.932 0.351508
## NeighborhoodEdwards -9.082e-02  2.496e-02  -3.639 0.000291 ***
## NeighborhoodNAmes   -3.826e-05  2.184e-02  -0.002 0.998602
## NeighborhoodOldTown -4.658e-02  2.928e-02  -1.591 0.111999
## NeighborhoodSomerst  8.064e-02  4.127e-02   1.954 0.051030 .
## NeighborhoodOther    1.536e-02  1.666e-02   0.922 0.356619
## Condition1Feedr      2.899e-02  3.348e-02   0.866 0.386809
## Condition1Norm       1.028e-01  2.800e-02   3.672 0.000256 ***
```

```
## Condition1PosN        5.869e-04  4.652e-02   0.013 0.989937
## Condition1RRAn        4.860e-02  4.215e-02   1.153 0.249199
## Condition1Other       3.969e-02  4.373e-02   0.908 0.364364
## BldgType2fmCon        1.084e-01  1.022e-01   1.061 0.289184
## BldgTypeDuplex        4.085e-02  5.978e-02   0.683 0.494613
## BldgTypeTwnhs        -9.642e-02  7.304e-02  -1.320 0.187179
## BldgTypeTwnhsE       -6.350e-02  6.865e-02  -0.925 0.355226
## HouseStyle1.5Unf     -1.514e-02  4.964e-02  -0.305 0.760375
## HouseStyle1Story      1.084e-02  2.952e-02   0.367 0.713422
## HouseStyle2.5Fin     -1.286e-01  8.254e-02  -1.558 0.119703
## HouseStyle2.5Unf      4.389e-02  5.599e-02   0.784 0.433302
## HouseStyle2Story     -4.758e-02  2.312e-02  -2.058 0.039873 *
## HouseStyleSFoyer     -7.391e-02  4.729e-02  -1.563 0.118471
## HouseStyleSLvl       -3.590e-02  3.758e-02  -0.955 0.339692
## OverallQual           5.575e-02  6.695e-03   8.327 3.39e-16 ***
## OverallCond           3.876e-02  5.812e-03   6.668 4.72e-11 ***
## YearBuilt             6.627e-04  4.766e-04   1.391 0.164743
## YearRemodAdd          8.254e-04  3.714e-04   2.222 0.026538 *
## RoofStyleGable       -3.053e-02  5.168e-02  -0.591 0.554885
## RoofStyleGambrel     -3.022e-02  8.263e-02  -0.366 0.714698
## RoofStyleHip         -5.885e-03  5.268e-02  -0.112 0.911081
## RoofStyleMansard     -1.818e-02  8.143e-02  -0.223 0.823352
## RoofStyleShed         1.801e-03  1.260e-01   0.014 0.988601
## Exterior1stMetalSd    2.828e-02  5.828e-02   0.485 0.627590
## Exterior1stPlywood    6.845e-02  3.082e-02   2.221 0.026615 *
## Exterior1stVinylSd    9.095e-02  5.494e-02   1.656 0.098186 .
## Exterior1stWd Sdng   -2.045e-02  3.680e-02  -0.556 0.578689
## Exterior1stOther      6.890e-02  3.317e-02   2.077 0.038075 *
## Exterior2ndMetalSd    1.352e-02  5.845e-02   0.231 0.817156
## Exterior2ndPlywood   -2.686e-02  2.765e-02  -0.972 0.331487
## Exterior2ndVinylSd   -5.499e-02  5.519e-02  -0.996 0.319344
## Exterior2ndWd Sdng    2.664e-02  3.703e-02   0.719 0.472189
## Exterior2ndOther     -5.710e-02  3.342e-02  -1.709 0.087890 .
## MasVnrTypeBrkFace     5.518e-02  4.292e-02   1.286 0.198910
## MasVnrTypeNone        5.922e-02  4.349e-02   1.362 0.173600
## MasVnrTypeStone       7.202e-02  4.557e-02   1.580 0.114374
## MasVnrArea            2.446e-05  3.513e-05   0.696 0.486420
## ExterQualFa           5.683e-02  8.042e-02   0.707 0.480009
## ExterQualGd           8.137e-02  3.156e-02   2.579 0.010093 *
## ExterQualTA           5.535e-02  3.509e-02   1.578 0.115057
## ExterCondFa          -1.013e-01  1.391e-01  -0.728 0.466527
## ExterCondGd          -7.163e-02  1.332e-01  -0.538 0.590924
## ExterCondPo          -8.326e-02  1.994e-01  -0.417 0.676424
## ExterCondTA          -5.119e-02  1.329e-01  -0.385 0.700148
## FoundationCBlock      3.482e-02  2.128e-02   1.636 0.102229
## FoundationPConc       5.413e-02  2.247e-02   2.409 0.016217 *
## FoundationSlab       -8.560e-02  7.727e-02  -1.108 0.268254
## FoundationStone       2.125e-01  8.687e-02   2.446 0.014632 *
## FoundationWood       -1.025e-01  8.313e-02  -1.234 0.217720
## BsmtQualFa           -9.829e-02  4.235e-02  -2.321 0.020549 *
## BsmtQualGd           -6.420e-02  2.118e-02  -3.031 0.002513 **
## BsmtQualNone         -3.964e-02  1.612e-01  -0.246 0.805764
## BsmtQualTA           -7.611e-02  2.640e-02  -2.883 0.004046 **
## BsmtCondGd            3.485e-02  3.626e-02   0.961 0.336800
```

```
## BsmtCondNone                 NA         NA       NA        NA
## BsmtCondPo          -2.888e-01  2.592e-01   -1.114 0.265511
## BsmtCondTA           3.165e-02  2.904e-02    1.090 0.276049
## BsmtExposureGd       3.250e-02  1.957e-02    1.661 0.097127 .
## BsmtExposureMn      -2.245e-02  2.024e-02   -1.109 0.267621
## BsmtExposureNo      -3.212e-02  1.445e-02   -2.222 0.026534 *
## BsmtExposureNone    -4.410e-02  1.281e-01   -0.344 0.730761
## BsmtFinType1BLQ     -8.384e-03  1.854e-02   -0.452 0.651207
## BsmtFinType1GLQ      1.338e-02  1.642e-02    0.815 0.415344
## BsmtFinType1LwQ     -1.038e-02  2.463e-02   -0.421 0.673570
## BsmtFinType1None                NA         NA       NA        NA
## BsmtFinType1Rec     -1.827e-02  1.951e-02   -0.936 0.349319
## BsmtFinType1Unf     -4.848e-02  1.852e-02   -2.617 0.009023 **
## BsmtFinSF1          -3.047e-06  3.168e-05   -0.096 0.923409
## BsmtFinType2BLQ     -2.783e-02  4.988e-02   -0.558 0.577015
## BsmtFinType2GLQ      2.451e-02  5.544e-02    0.442 0.658547
## BsmtFinType2LwQ      8.146e-03  4.641e-02    0.176 0.860703
## BsmtFinType2None                NA         NA       NA        NA
## BsmtFinType2Rec     -3.689e-03  4.428e-02   -0.083 0.933629
## BsmtFinType2Unf      4.660e-02  4.719e-02    0.988 0.323637
## BsmtFinSF2           6.135e-05  5.873e-05    1.045 0.296493
## BsmtUnfSF            1.207e-05  2.961e-05    0.408 0.683707
## TotalBsmtSF                     NA         NA       NA        NA
## HeatingGasA          1.341e-01  1.520e-01    0.882 0.377908
## HeatingGasW          1.784e-01  1.583e-01    1.126 0.260282
## HeatingGrav         -1.630e-01  1.801e-01   -0.905 0.365800
## HeatingOthW          8.607e-02  1.837e-01    0.469 0.639533
## HeatingWall          2.206e-01  1.738e-01    1.269 0.204781
## HeatingQCFa         -3.725e-02  2.988e-02   -1.246 0.212957
## HeatingQCGd         -2.205e-02  1.388e-02   -1.589 0.112545
## HeatingQCPo         -2.997e-01  1.482e-01   -2.022 0.043546 *
## HeatingQCTA         -3.811e-02  1.390e-02   -2.742 0.006232 **
## CentralAirY          6.552e-02  2.493e-02    2.628 0.008751 **
## ElectricalFuseF      2.693e-02  4.153e-02    0.648 0.516859
## ElectricalFuseP      4.089e-02  1.216e-01    0.336 0.736691
## ElectricalMix                   NA         NA       NA        NA
## ElectricalNone       8.114e-02  1.329e-01    0.610 0.541736
## ElectricalSBrkr     -3.056e-02  1.988e-02   -1.537 0.124569
## X1stFlrSF            2.065e-04  3.449e-05    5.986 3.20e-09 ***
## X2ndFlrSF            1.918e-04  3.603e-05    5.324 1.30e-07 ***
## LowQualFinSF         2.328e-04  1.171e-04    1.988 0.047128 *
## GrLivArea                       NA         NA       NA        NA
## BsmtFullBath         4.675e-02  1.343e-02    3.480 0.000527 ***
## BsmtHalfBath1        1.195e-02  2.098e-02    0.570 0.568891
## BsmtHalfBath2       -1.505e-01  1.294e-01   -1.163 0.245201
## FullBath1           -1.187e-01  9.119e-02   -1.302 0.193386
## FullBath2           -7.919e-02  9.200e-02   -0.861 0.389616
## FullBath3           -7.892e-02  9.954e-02   -0.793 0.428114
## HalfBath1            3.550e-02  1.451e-02    2.447 0.014598 *
## HalfBath2           -2.887e-02  6.172e-02   -0.468 0.640066
## BedroomAbvGr         3.869e-03  8.952e-03    0.432 0.665697
## KitchenAbvGr        -1.253e-01  4.297e-02   -2.917 0.003633 **
## KitchenQualFa       -5.481e-02  4.032e-02   -1.360 0.174349
## KitchenQualGd       -8.286e-02  2.270e-02   -3.650 0.000278 ***
```

```
## KitchenQualTA        -7.476e-02  2.535e-02  -2.949 0.003274 **
## TotRmsAbvGrd          1.600e-02  6.035e-03   2.652 0.008160 **
## FunctionalMaj2       -1.569e-01  1.488e-01  -1.054 0.292155
## FunctionalMin1        4.774e-02  6.008e-02   0.794 0.427134
## FunctionalMin2        2.922e-02  6.108e-02   0.478 0.632495
## FunctionalMod        -5.277e-02  7.332e-02  -0.720 0.471919
## FunctionalSev        -2.841e-01  1.755e-01  -1.619 0.105754
## FunctionalTyp         7.443e-02  5.384e-02   1.382 0.167227
## Fireplaces1           1.147e-01  3.796e-02   3.022 0.002585 **
## Fireplaces2           1.498e-01  4.089e-02   3.664 0.000264 ***
## Fireplaces3          -2.036e-01  8.598e-02  -2.368 0.018132 *
## FireplaceQuFa        -6.466e-02  4.697e-02  -1.377 0.169025
## FireplaceQuGd        -7.196e-02  3.636e-02  -1.979 0.048135 *
## FireplaceQuNone             NA         NA      NA       NA
## FireplaceQuPo        -8.207e-02  5.181e-02  -1.584 0.113591
## FireplaceQuTA        -7.754e-02  3.760e-02  -2.062 0.039519 *
## GarageTypeAttchd      5.027e-02  6.943e-02   0.724 0.469245
## GarageTypeBasment     3.125e-02  7.914e-02   0.395 0.693063
## GarageTypeBuiltIn     3.425e-02  7.264e-02   0.471 0.637410
## GarageTypeCarPort     1.184e-01  9.087e-02   1.303 0.192978
## GarageTypeDetchd      4.465e-02  6.896e-02   0.647 0.517500
## GarageTypeNone       -1.644e+00  8.107e-01  -2.027 0.042944 *
## GarageYrBlt          -8.261e-04  4.072e-04  -2.028 0.042830 *
## GarageFinishNone            NA         NA      NA       NA
## GarageFinishRFn      -2.230e-03  1.264e-02  -0.176 0.860017
## GarageFinishUnf      -1.153e-02  1.608e-02  -0.717 0.473386
## GarageCars            6.771e-02  1.442e-02   4.695 3.12e-06 ***
## GarageArea            4.545e-05  4.942e-05   0.920 0.358017
## GarageQualFa         -2.916e-01  1.638e-01  -1.780 0.075450 .
## GarageQualGd         -2.124e-01  1.684e-01  -1.261 0.207600
## GarageQualNone              NA         NA      NA       NA
## GarageQualPo         -7.845e-02  2.343e-01  -0.335 0.737815
## GarageQualTA         -2.605e-01  1.617e-01  -1.611 0.107557
## GarageCondFa          1.537e-01  1.902e-01   0.808 0.419228
## GarageCondGd          2.654e-01  2.002e-01   1.325 0.185457
## GarageCondNone              NA         NA      NA       NA
## GarageCondPo          3.406e-01  2.255e-01   1.510 0.131370
## GarageCondTA          1.954e-01  1.878e-01   1.041 0.298365
## PavedDriveP           2.589e-02  3.769e-02   0.687 0.492284
## PavedDriveY           3.341e-02  2.318e-02   1.442 0.149805
## WoodDeckSF            1.455e-04  3.779e-05   3.850 0.000127 ***
## OpenPorchSF          -7.410e-05  7.863e-05  -0.942 0.346247
## EnclosedPorch         1.364e-04  8.368e-05   1.630 0.103515
## X3SsnPorch            2.422e-04  1.391e-04   1.741 0.082083 .
## ScreenPorch           3.075e-04  9.046e-05   3.400 0.000707 ***
## PoolArea             -3.067e-04  1.364e-04  -2.248 0.024828 *
## MiscVal              -2.842e-06  7.813e-06  -0.364 0.716141
## MoSold                1.611e-04  1.653e-03   0.097 0.922367
## YrSold                4.151e-04  3.457e-03   0.120 0.904473
## SaleTypeCon           1.377e-01  9.681e-02   1.423 0.155219
## SaleTypeConLD         2.662e-01  6.916e-02   3.849 0.000128 ***
## SaleTypeConLI         2.675e-02  6.530e-02   0.410 0.682231
## SaleTypeConLw         3.983e-02  7.451e-02   0.535 0.593085
## SaleTypeCWD           8.887e-02  7.386e-02   1.203 0.229261
```

```
## SaleTypeNew           1.585e-01  8.919e-02   1.777 0.076001 .
## SaleTypeOth           7.252e-03  9.559e-02   0.076 0.939542
## SaleTypeWD           -6.655e-03  2.683e-02  -0.248 0.804183
## SaleConditionAdjLand  8.473e-02  1.104e-01   0.767 0.443091
## SaleConditionAlloca   1.366e-01  6.090e-02   2.243 0.025146 *
## SaleConditionFamily   7.333e-02  3.935e-02   1.863 0.062766 .
## SaleConditionNormal   7.092e-02  1.844e-02   3.846 0.000129 ***
## SaleConditionPartial -4.561e-02  8.492e-02  -0.537 0.591387
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1253 on 833 degrees of freedom
## Multiple R-squared:  0.9207, Adjusted R-squared:  0.9029
## F-statistic: 51.72 on 187 and 833 DF,  p-value: < 2.2e-16


## [1] "MSE of Testing Set:  0.02"
```

Splines and GAMs

Sources