Home Buying

Abdelmalek Hajjam/ Monu Chacko

05/15/2020

Abstract

- In this study we will understand the factors affecting home buying
- How does location influence home buying
- What other conditions influence in the decision of home buying?

Introduction

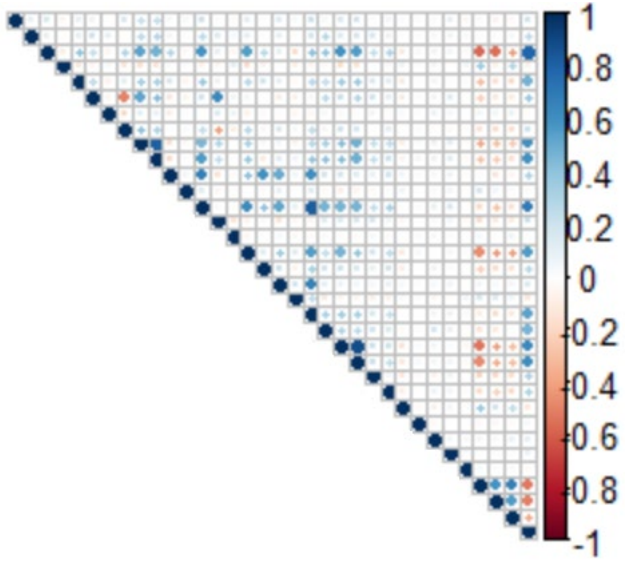
This data was published by Journal of Statistics Education.

There are 79 explanatory features describe attributes of the houses such as conditions, zoning, and neighborhood. They are both numeric and categorical

Methodology

- Data Imputation:

Missing data was a problem and efforts were made to smooth them out and fill with meaningful data.



- Data Split:

The data was split for training and testing.

- Categorical Variables:

Used dummy variables where necessary.

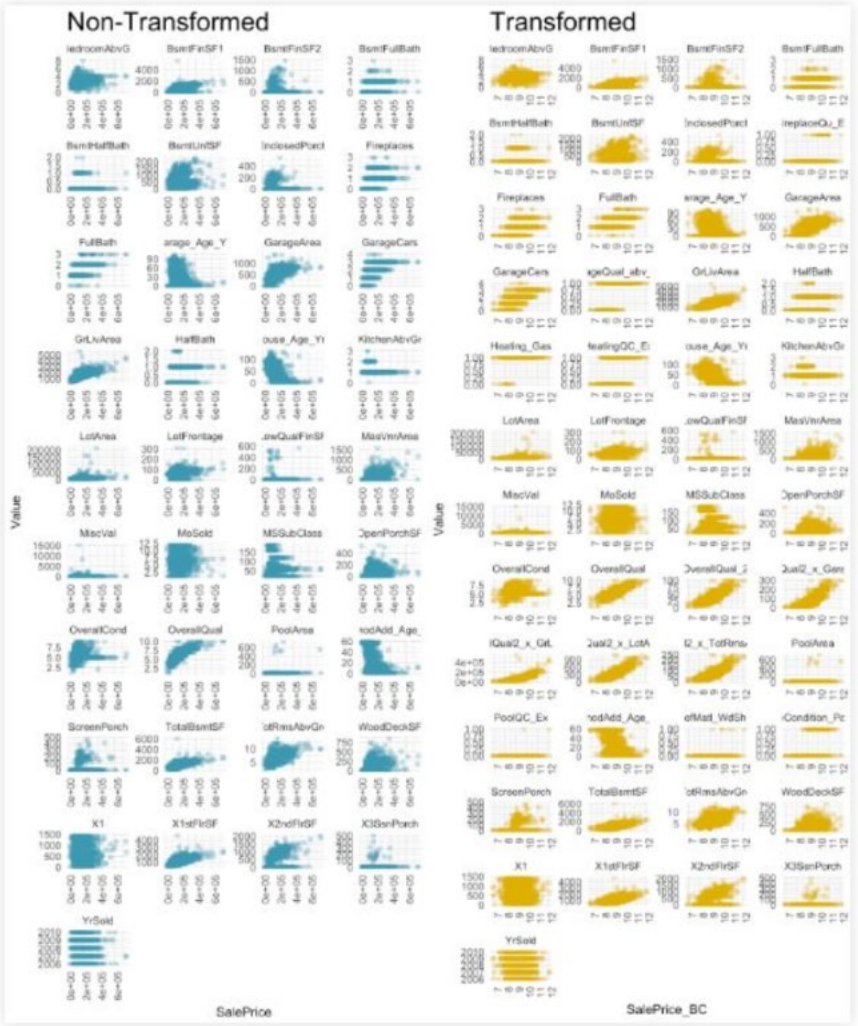
Transformation

- Various transformations were made during this project.

Box-Cox transformation was performed. The optimal λ was found to be 0.184. This means that the response variable SalePrice was raised to the 0.184 power.

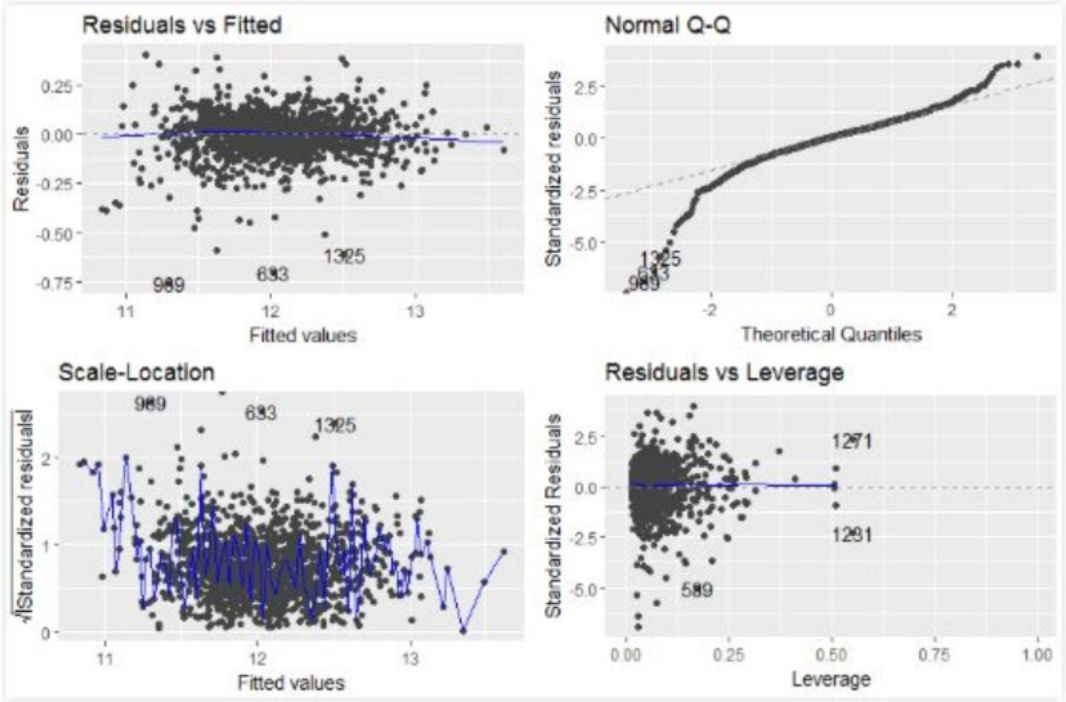
$$Y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ log(y), & \text{if } \lambda = 0 \end{cases}$$

Certain calculations had to be made like age etc. Negative values were set to 0.



Model Selection

Various models were evaluated. Model tuning was performed through K-fold validation. Four leverage points were removed, and categorical variables *Condition2* and *Utilities* were removed for poor sample distribution.

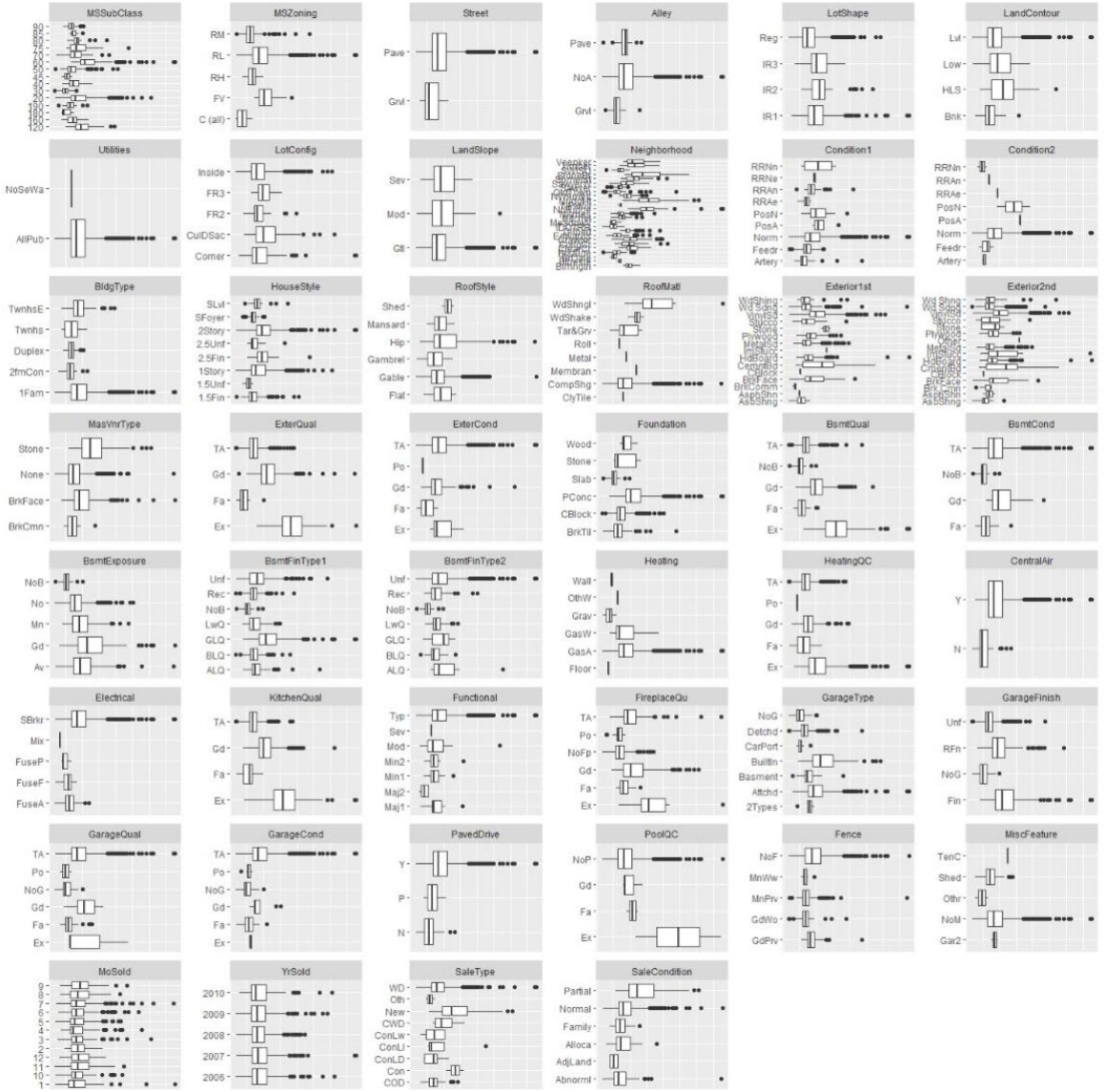


Selected model base on best performance on the training data. As such, we are not worried about over fitting.

Building model

- $model 1 :
  $All data from bcData
  $F-statistic: 79.22 on 227 and 1232 DF $p-value: < 2.2e-16
  $ Multiple R-squared: 0.9359
- $model 2 :
  $All data from bcData : Stepwise regression
  $F-statistic: 119.8 on 152 and 1307DF $p-value: < 2.2e-16
  $ Multiple R-squared: 0.933
- $model 3 :
  $Only 0 p-value significant data from bcData
  $F-statistic: 202.4 on 58 and 1401 DF $p-value: < 2.2e-16
  $Multiple R-squared: 0.8934

- $model 4 :
  $p-value <.01 significance data from bcData
  $F-statistic: 148.4 on 104 and 1355 DF $p-value: < 2.2e-16
  $Multiple R-squared: 0.9193



Building model

- $model 5 :
  $Same parameters from model 4 - interactions,
  log(SalePrice),log(LotArea)
  $F-statistic: 115.2 on 99 and 1360 DF $p-value:
  < 2.2e-16
  $Multiple R-squared: 0.8935

- $model 6 :
  $Same parameters from model 4 with
  log(SalePrice),log(LotArea)
  $F-statistic: 142.8 on 104 and 1355 DF $p-value:
  < 2.2e-16
  $Multiple R-squared: 0.9164

Literature Review

- Ecomena (ecomena.org) provides us a list of factors that incluence home buying. This includes Topography, Air and water quality, Climate, Sustainability etc.

- Daveramsey (daveramsey.com) in its article outlines some tips for buying first time home.

- Artificial intelligence (AI) is quietly infiltrating the real estate industry. Here is an article - https://www.bankrate.com/mortgages/how-ai-is-revolutionizing-the-real-estate-market/

Conclusion

- Location was the most important feature. Feature like garages, neighborhood, land and other amenities were important. Features like shapes, styles, and building materials were on the lower end.
- Other factors such as condition also played a role. Further, it is unlikely that this model will transfer to other geographic areas and should only be used to estimate houses in the mid west. In particular, university towns such as this one where off-campus housing may dictate a large portion of sale price.