Housing Price Prediction Model using modeldata package - a subset of ames housing data

Saurav Mukherjee

2023-02-16

Introduction

I am building home price prediction model. I am using Ames Housing dataset to explore the attributes which have been identified somehow influencing the housing cost.

Initially I wanted to use the 'Ames Housing Data" - a data set describing the sale of individual residential property in Ames, Iowa from 2006 to 2010. The data set contains 2930 observations and a large number of explanatory variables (23 nominal, 23 ordinal, 14 discrete, and 20 continuous) involved in assessing home values. Howvever, I looked at a dataset which is a subset of this dataset and which is available within modeldata package created by https://modeldata.tidymodels.org/. I did some research and looked at the model - Hedonic Pricing Method to predict the house price. The Hedonic Pricing Method talks about internal characteristics as well as the external factors affecting the price of a good. Based on the idea of hedonic price modeling I am looking the is that neighborhood-specific and unit-specific characteristics help determine house prices.

Data - Ames Housing Data

A data set from De Cock (2011) has 82 fields were recorded for 2,930 properties in Ames IA. I used a version from the package modeldata dataset name as ames which is copies from the original AmesHousing package but does not include a few quality columns that appear to be outcomes rather than predictors.

Load required Libraries

Load ames dataset

Exploratory Data Analysis

Table 1: Ames Housing Dataset dimension

x
2930
74

```
## tibble [2,930 x 74] (S3: tbl_df/tbl/data.frame)
##  $ MS_SubClass       : Factor w/ 16 levels "One_Story_1946_and_Newer_All_Styles",..: 1 1 1 1 6 6 12
##  $ MS_Zoning         : Factor w/ 7 levels "Floating_Village_Residential",..: 3 2 3 3 3 3 3 3 3 3 ...
##  $ Lot_Frontage      : num [1:2930] 141 80 81 93 74 78 41 43 39 60 ...
```

```
## $ Lot_Area           : int [1:2930] 31770 11622 14267 11160 13830 9978 4920 5005 5389 7500 ...
## $ Street             : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...
## $ Alley              : Factor w/ 3 levels "Gravel","No_Alley_Access",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ Lot_Shape          : Factor w/ 4 levels "Regular","Slightly_Irregular",..: 2 1 2 1 2 2 1 2 2 1 ...
## $ Land_Contour       : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 4 4 4 2 4 4 ...
## $ Utilities          : Factor w/ 3 levels "AllPub","NoSeWa",..: 1 1 1 1 1 1 1 1 1 1 ...
## $ Lot_Config         : Factor w/ 5 levels "Corner","CulDSac",..: 1 5 1 1 5 5 5 5 5 5 ...
## $ Land_Slope         : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...
## $ Neighborhood       : Factor w/ 29 levels "North_Ames","College_Creek",..: 1 1 1 1 7 7 17 17 17 7 .
## $ Condition_1        : Factor w/ 9 levels "Artery","Feedr",..: 3 2 3 3 3 3 3 3 3 3 ...
## $ Condition_2        : Factor w/ 8 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 3 ...
## $ Bldg_Type          : Factor w/ 5 levels "OneFam","TwoFmCon",..: 1 1 1 1 1 1 5 5 5 1 ...
## $ House_Style        : Factor w/ 8 levels "One_and_Half_Fin",..: 3 3 3 3 8 8 3 3 3 8 ...
## $ Overall_Cond       : Factor w/ 10 levels "Very_Poor","Poor",..: 5 6 6 5 5 6 5 5 5 5 ...
## $ Year_Built         : int [1:2930] 1960 1961 1958 1968 1997 1998 2001 1992 1995 1999 ...
## $ Year_Remod_Add     : int [1:2930] 1960 1961 1958 1968 1998 1998 2001 1992 1996 1999 ...
## $ Roof_Style         : Factor w/ 6 levels "Flat","Gable",..: 4 2 4 4 2 2 2 2 2 2 ...
## $ Roof_Matl          : Factor w/ 8 levels "ClyTile","CompShg",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ Exterior_1st       : Factor w/ 16 levels "AsbShng","AsphShn",..: 4 14 15 4 14 14 6 7 6 14 ...
## $ Exterior_2nd       : Factor w/ 17 levels "AsbShng","AsphShn",..: 11 15 16 4 15 15 6 7 6 15 ...
## $ Mas_Vnr_Type       : Factor w/ 5 levels "BrkCmn","BrkFace",..: 5 4 2 4 4 2 4 4 4 4 ...
## $ Mas_Vnr_Area       : num [1:2930] 112 0 108 0 0 20 0 0 0 0 ...
## $ Exter_Cond         : Factor w/ 5 levels "Excellent","Fair",..: 5 5 5 5 5 5 5 5 5 5 ...
## $ Foundation         : Factor w/ 6 levels "BrkTil","CBlock",..: 2 2 2 2 3 3 3 3 3 3 ...
## $ Bsmt_Cond          : Factor w/ 6 levels "Excellent","Fair",..: 3 6 6 6 6 6 6 6 6 6 ...
## $ Bsmt_Exposure      : Factor w/ 5 levels "Av","Gd","Mn",..: 2 4 4 4 4 4 3 4 4 4 ...
## $ BsmtFin_Type_1     : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 2 6 1 1 3 3 3 1 3 7 ...
## $ BsmtFin_SF_1       : num [1:2930] 2 6 1 1 3 3 3 1 3 7 ...
## $ BsmtFin_Type_2     : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 7 4 7 7 7 7 7 7 7 7 ...
## $ BsmtFin_SF_2       : num [1:2930] 0 144 0 0 0 0 0 0 0 0 ...
## $ Bsmt_Unf_SF        : num [1:2930] 441 270 406 1045 137 ...
## $ Total_Bsmt_SF      : num [1:2930] 1080 882 1329 2110 928 ...
## $ Heating            : Factor w/ 6 levels "Floor","GasA",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ Heating_QC         : Factor w/ 5 levels "Excellent","Fair",..: 2 5 5 1 3 1 1 1 1 3 ...
## $ Central_Air        : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...
## $ Electrical         : Factor w/ 6 levels "FuseA","FuseF",..: 5 5 5 5 5 5 5 5 5 5 ...
## $ First_Flr_SF       : int [1:2930] 1656 896 1329 2110 928 926 1338 1280 1616 1028 ...
## $ Second_Flr_SF      : int [1:2930] 0 0 0 0 701 678 0 0 0 776 ...
## $ Gr_Liv_Area        : int [1:2930] 1656 896 1329 2110 1629 1604 1338 1280 1616 1804 ...
## $ Bsmt_Full_Bath     : num [1:2930] 1 0 0 1 0 0 1 0 1 0 ...
## $ Bsmt_Half_Bath     : num [1:2930] 0 0 0 0 0 0 0 0 0 0 ...
## $ Full_Bath          : int [1:2930] 1 1 1 2 2 2 2 2 2 2 ...
## $ Half_Bath          : int [1:2930] 0 0 1 1 1 1 0 0 0 1 ...
## $ Bedroom_AbvGr      : int [1:2930] 3 2 3 3 3 3 2 2 2 3 ...
## $ Kitchen_AbvGr      : int [1:2930] 1 1 1 1 1 1 1 1 1 1 ...
## $ TotRms_AbvGrd      : int [1:2930] 7 5 6 8 6 7 6 5 5 7 ...
## $ Functional         : Factor w/ 8 levels "Maj1","Maj2",..: 8 8 8 8 8 8 8 8 8 8 ...
## $ Fireplaces         : int [1:2930] 2 0 0 2 1 1 0 0 1 1 ...
## $ Garage_Type        : Factor w/ 7 levels "Attchd","Basment",..: 1 1 1 1 1 1 1 1 1 1 ...
## $ Garage_Finish      : Factor w/ 4 levels "Fin","No_Garage",..: 1 4 4 1 1 1 1 3 3 1 ...
## $ Garage_Cars        : num [1:2930] 2 1 1 2 2 2 2 2 2 2 ...
## $ Garage_Area        : num [1:2930] 528 730 312 522 482 470 582 506 608 442 ...
## $ Garage_Cond        : Factor w/ 6 levels "Excellent","Fair",..: 6 6 6 6 6 6 6 6 6 6 ...
## $ Paved_Drive        : Factor w/ 3 levels "Dirt_Gravel",..: 2 3 3 3 3 3 3 3 3 3 ...
```

```
##  $ Wood_Deck_SF      : int [1:2930] 210 140 393 0 212 360 0 0 237 140 ...
##  $ Open_Porch_SF     : int [1:2930] 62 0 36 0 34 36 0 82 152 60 ...
##  $ Enclosed_Porch    : int [1:2930] 0 0 0 0 0 0 170 0 0 0 ...
##  $ Three_season_porch: int [1:2930] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Screen_Porch      : int [1:2930] 0 120 0 0 0 0 0 144 0 0 ...
##  $ Pool_Area         : int [1:2930] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Pool_QC           : Factor w/ 5 levels "Excellent","Fair",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Fence             : Factor w/ 5 levels "Good_Privacy",..: 5 3 5 5 3 5 5 5 5 5 ...
##  $ Misc_Feature      : Factor w/ 6 levels "Elev","Gar2",..: 3 3 2 3 3 3 3 3 3 3 ...
##  $ Misc_Val          : int [1:2930] 0 0 12500 0 0 0 0 0 0 0 ...
##  $ Mo_Sold           : int [1:2930] 5 6 6 4 3 6 4 1 3 6 ...
##  $ Year_Sold         : int [1:2930] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ Sale_Type         : Factor w/ 10 levels "COD","Con","ConLD",..: 10 10 10 10 10 10 10 10 10 10 ...
##  $ Sale_Condition    : Factor w/ 6 levels "Abnorml","AdjLand",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ Sale_Price        : int [1:2930] 215000 105000 172000 244000 189900 195500 213500 191500 236500 18
##  $ Longitude         : num [1:2930] -93.6 -93.6 -93.6 -93.6 -93.6 ...
##  $ Latitude          : num [1:2930] 42.1 42.1 42.1 42.1 42.1 ...
```

Table: Ames Housing Dataset

|| || || ||

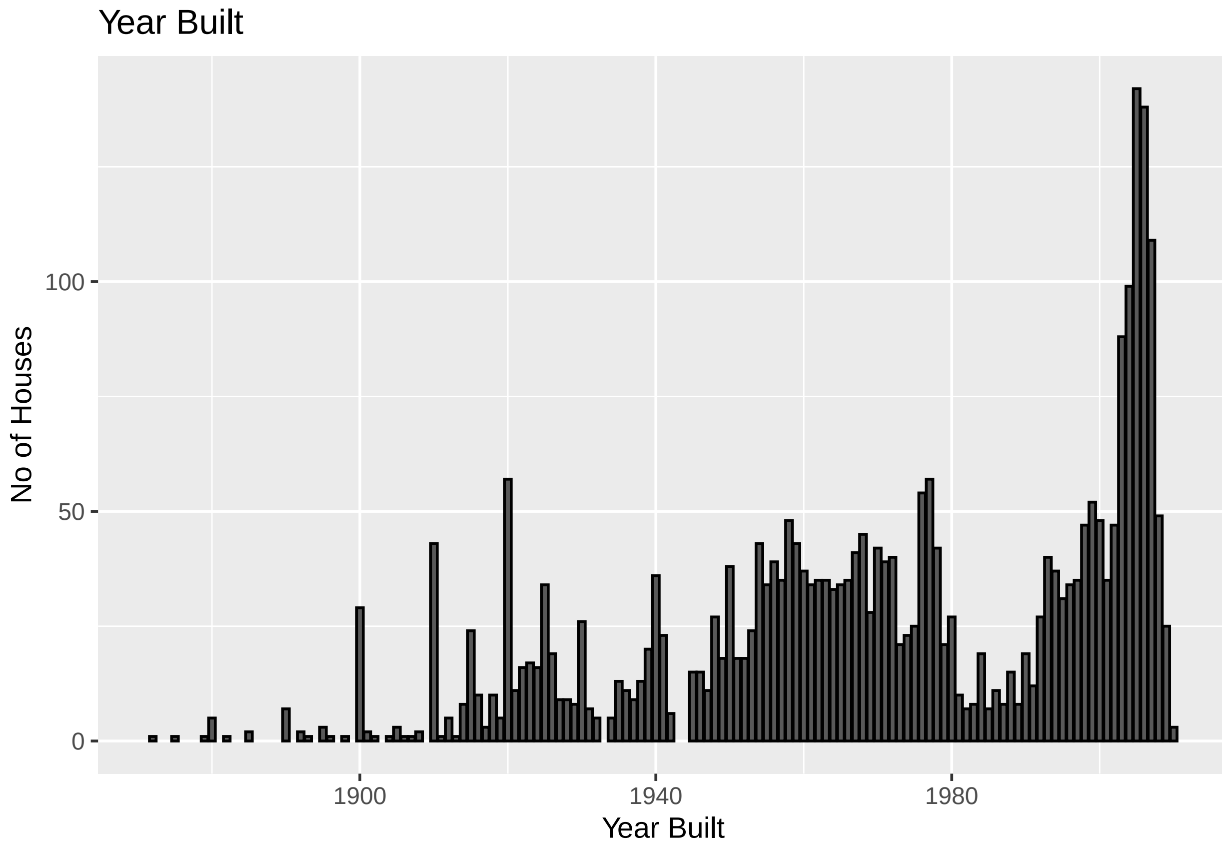

```
##
## Sale Price skewness : 1.742607
```

```
##
## Sale Price kurtosis : 8.108122
```

Sale Price Observation

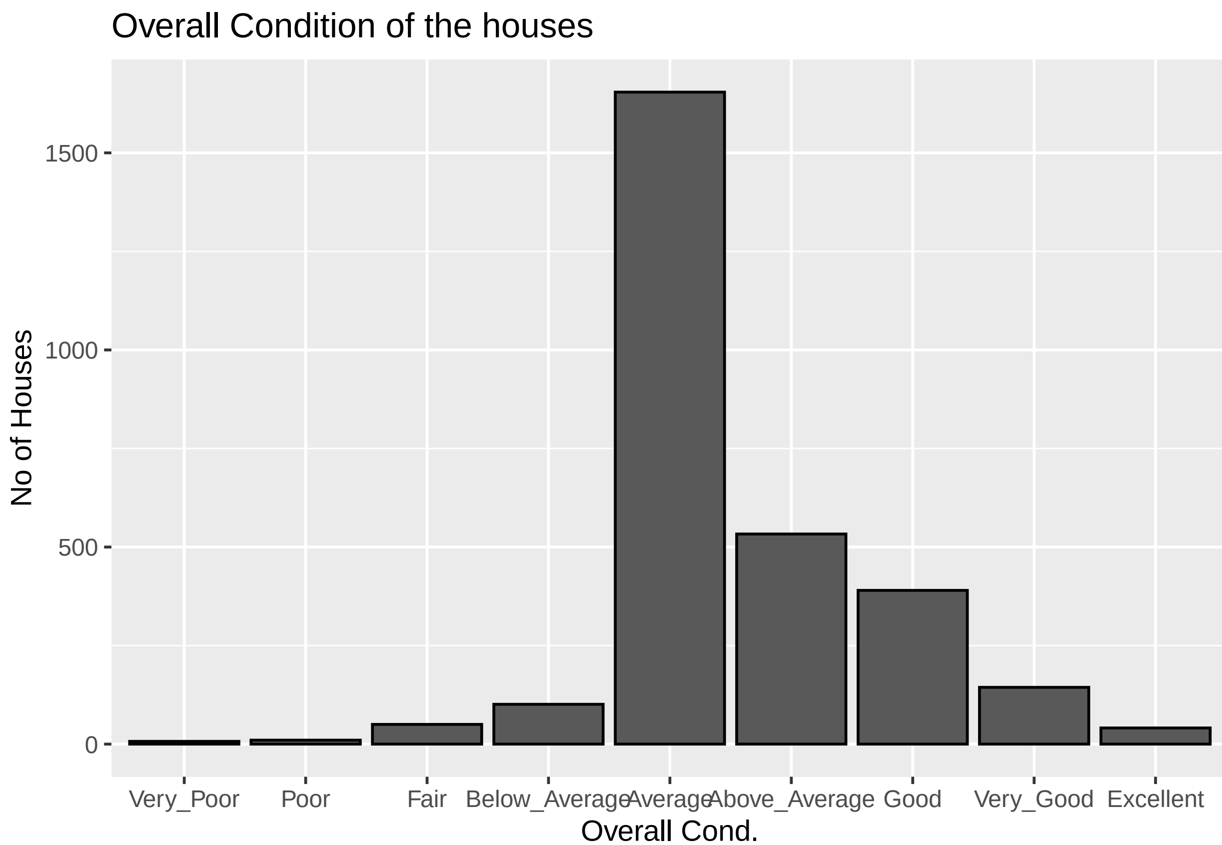
The Sale Price is right-skewed

Age of the Building



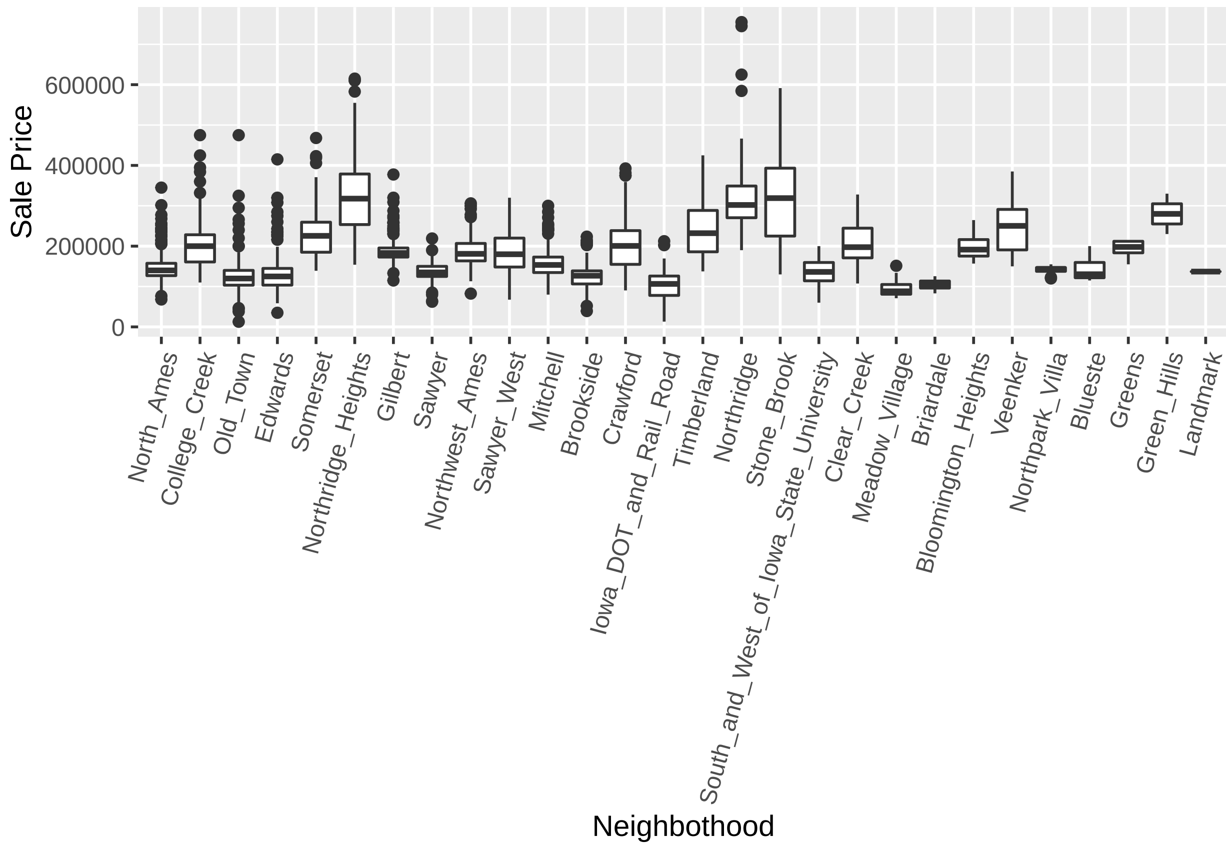
It looks that we have more houses were built at hte begining of 2000

Condition of the houses



House condition - most of the houses are of average condition

Neighborhood and House Price



House Price varies with the neighborhood with few outliers by neighborhood. Also, the median house price by neighborhood is roughly between 200,000 and 400,000. It seems Neighborhood would have some impact on housing price.

Correlation between Sale Price and other variables

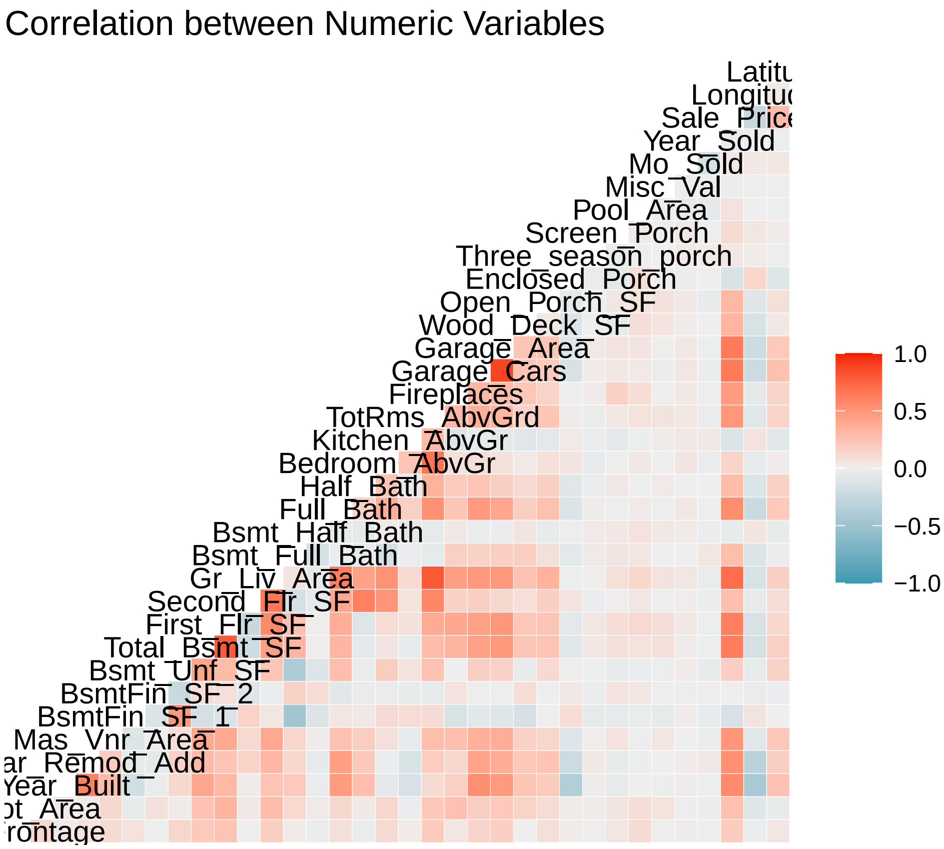


Table 2: Ames Housing Dataset - correlated numeric variables with the Sale Price

	x
Lot_Frontage	0.2018745
Lot_Area	0.2665492
Year_Built	0.5584261
Year_Remod_Add	0.5329738
Mas_Vnr_Area	0.5021960
BsmtFin_SF_1	-0.1349055
BsmtFin_SF_2	0.0060176
Bsmt_Unf_SF	0.1833076
Total_Bsmt_SF	0.6325288
First_Flr_SF	0.6216761
Second_Flr_SF	0.2693734
Gr_Liv_Area	0.7067799
Bsmt_Full_Bath	0.2758227
Bsmt_Half_Bath	-0.0358166
Full_Bath	0.5456039
Half_Bath	0.2850560
Bedroom_AbvGr	0.1439134
Kitchen_AbvGr	-0.1198137
TotRms_AbvGrd	0.4954744

	x
Fireplaces	0.4745581
Garage_Cars	0.6475616
Garage_Area	0.6401383
Wood_Deck_SF	0.3271432
Open_Porch_SF	0.3129505
Enclosed_Porch	-0.1287874
Three_season_porch	0.0322246
Screen_Porch	0.1121512
Pool_Area	0.0684032
Misc_Val	-0.0156915
Mo_Sold	0.0352588
Year_Sold	-0.0305691
Sale_Price	1.0000000
Longitude	-0.2513973
Latitude	0.2908914

There are some high correlations between variables mostly positive but with some negative. I did further analysis and added pairwise correlation between other numeric valiables and sales price. Thus, I dentified variables which has higher correlations (correlation > 0.5 and < - 0.2)

I also looked at some non-numeric variables and their relatins with the Sale Price

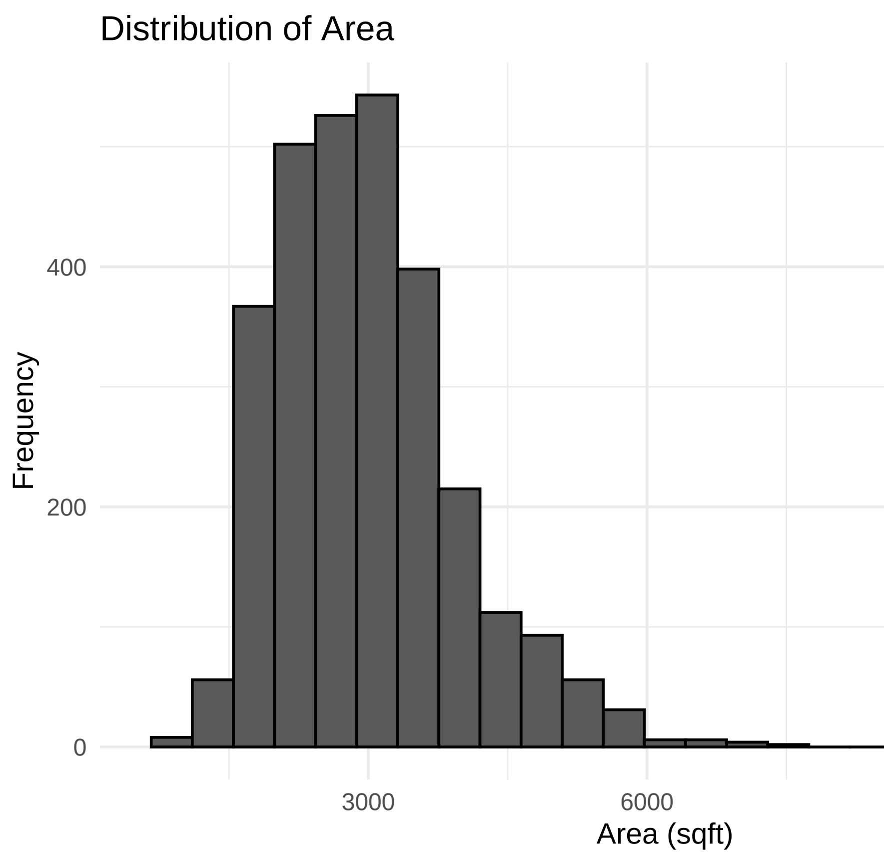
Table 3: Ames Housing Dataset - correlated non-numeric variables with the Sale Price

	x
MS_SubClass	-0.0347748
MS_Zoning	-0.3064225
Street	0.0595193
Alley	0.1088436
Lot_Shape	0.3026647
Land_Contour	-0.0693388
Utilities	-0.0310365
Lot_Config	-0.0587875
Land_Slope	0.0685534
Neighborhood	0.1575002
Condition_1	0.1590773
Condition_2	0.1048063
Bldg_Type	-0.0952280
House_Style	0.2310546
Overall_Cond	-0.1635790
Roof_Style	0.2546450
Roof_Matl	0.0720760
Exterior_1st	0.0550217
Exterior_2nd	0.0535448
Mas_Vnr_Type	-0.0763142
Exter_Cond	0.1206939
Foundation	0.4579558
Bsmt_Cond	0.1095363

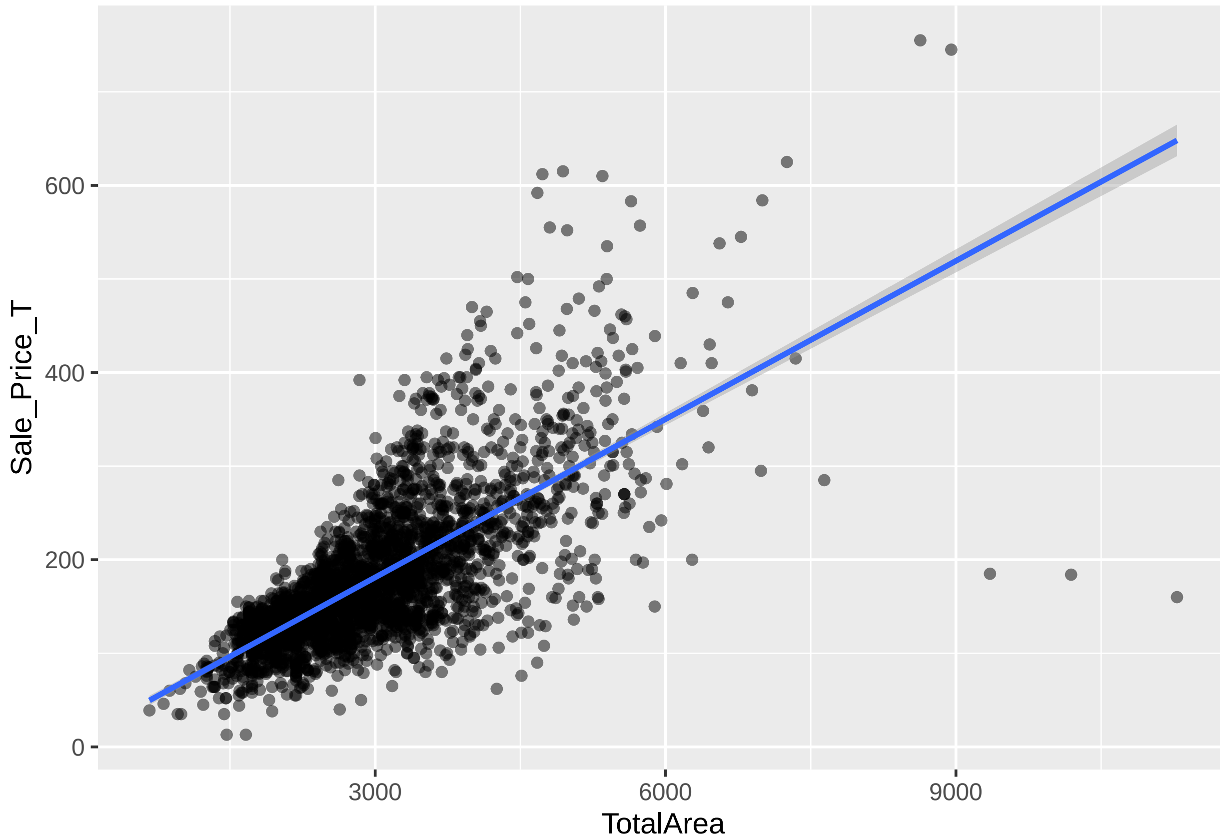
	x
Bsmt_Exposure	-0.3519094
BsmtFin_Type_1	-0.0975925
BsmtFin_Type_2	0.1074020
Heating	-0.0728977
Heating_QC	-0.4426972
Central_Air	0.2645064
Electrical	0.2378218
Functional	0.1192451
Garage_Type	-0.4061833
Garage_Finish	-0.4494826
Garage_Cond	0.2750657
Paved_Drive	0.2749134
Pool_QC	-0.0919699
Fence	0.1745827
Misc_Feature	-0.0574683
Sale_Type	-0.1845079
Sale_Condition	0.3330831

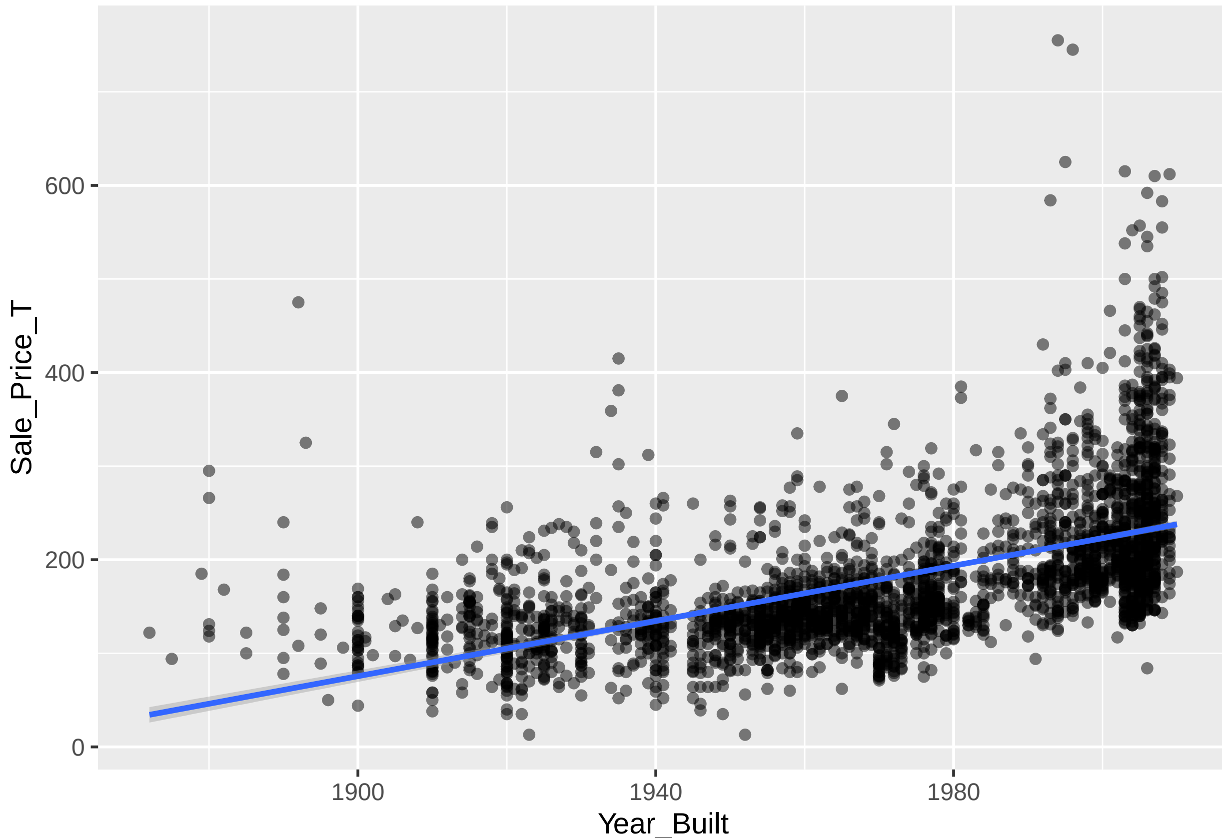
Looking at the non-numeric variable, I identified few variables which are highly correlated -

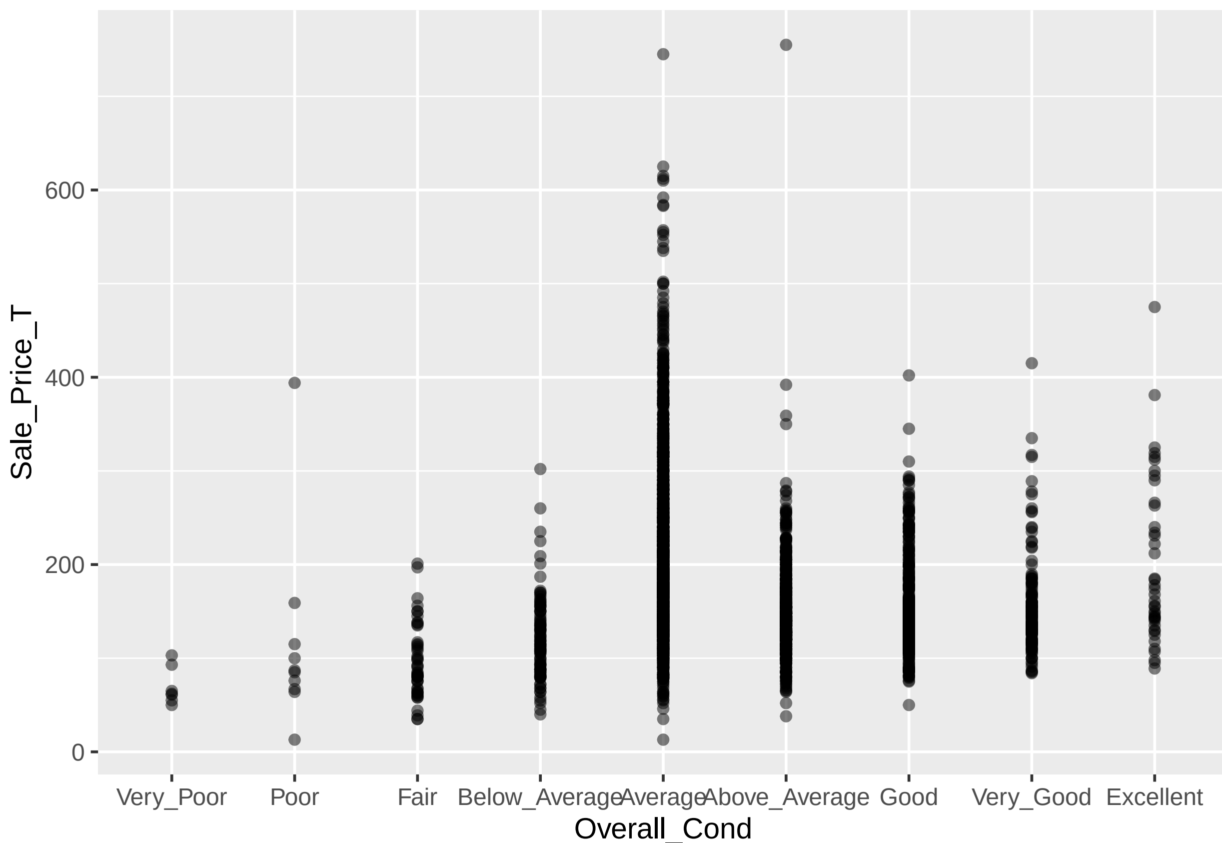
MS_Zoning, Lot_Shape, Foundation, Sale_Condition , Garage_Finish, House_Style, Heating_QC,



#Data Wrangling and Some more visualizations


Sale_Price_T
600
400
200
0
3000
6000
9000
TotalArea


Sale_Price_T
600
400
200
0
1900
1940
1980
Year_Built



Create Test Set and Training set for building Linear Models # test set will be 20% of housing_data data

Table 4: Ames Housing Dataset dimension

x
2930
16

Table 5: Ames Housing Dataset

Sale_Price	Total_Area	Year_Built	Overall_Cond	Garage_Cars	Garage_Area	Total_Bsmt_SF	Year_Remod_Add	Mas_Vnr_Area	MS_Zoning	Lot_Shape	Foundation	Sale_Condition	Garage_Finish	House_Style	Heating_QC
215	3312	1960	Average	2	528	1080	1960	112	Residential_Low_Density	Slightly_Irregular	CBlock	Normal	Fin	One_Story	Fair
105	1792	1961	Above_Average	1	730	882	1961	0	Residential_High_Density	Regular	CBlock	Normal	Unf	One_Story	Typical
172	2658	1958	Above_Average	1	312	1329	1958	108	Residential_Low_Density	Slightly_Irregular	CBlock	Normal	Unf	One_Story	Typical
244	4220	1968	Average	2	522	2110	1968	0	Residential_Low_Density	Regular	CBlock	Normal	Fin	One_Story	Excellent
190	3258	1997	Average	2	482	928	1998	0	Residential_Low_Density	Slightly_Irregular	PConc	Normal	Fin	Two_Story	Good
196	3208	1998	Above_Average	2	470	926	1998	20	Residential_Low_Density	Slightly_Irregular	PConc	Normal	Fin	Two_Story	Excellent

Table 6: Ames Housing Dataset Summary

Sale_Price_T	TotalArea	Year_Built	Overall_Cond	Garage_Cars	Garage_Area	Total_Bsmt_SF	Year_Remod_Add	Mas_Vnr_Area	MS_Zoning	Lot_Shape	Foundation	Sale_Condition	Garage_Finish	House_Style	Heating_QC
Min. : 13.0	Min. : 668	Min. :1872	Average :1654	Min. :0.000	Min. : 0.0	Min. : 0	Min. :1950	Min. : 0.0	Floating_Village_Residential: 139	Regular :1859	BrkTil: 311	Abnorml: 190	Fin : 728	One_Story :1481	Excellent:1495
1st Qu.:130.0	1st Qu.: 2248	1st Qu.:1954	Above_Average: 533	1st Qu.:1.000	1st Qu.: 320.0	1st Qu.: 793	1st Qu.:1965	1st Qu.: 0.0	Residential_High_Density : 27	Slightly_Irregular : 979	CBlock :1244	AdjLand: 12	No_Garage: 159	Two_Story : 873	Fair : 92
Median :160.0	Median : 2880	Median :1973	Good : 390	Median :2.000	Median : 480.0	Median : 990	Median :1993	Median : 0.0	Residential_Low_Density :2273	Moderately_Irregular: 76	PConc :1310	Alloca: 24	RFn : 812	One_and_Half_Fin: 314	Good : 476
Mean :180.8	Mean : 2995	Mean :1971	Very_Good : 144	Mean :1.766	Mean : 472.7	Mean :1051	Mean :1984	Mean : 101.1	Residential_Medium_Density : 462	Irregular : 16	Slab : 49	Family : 46	Unf :1231	SLvl : 128	Poor : 3
3rd Qu.:214.0	3rd Qu.: 3483	3rd Qu.:2001	Below_Average: 101	3rd Qu.:2.000	3rd Qu.: 576.0	3rd Qu.:1302	3rd Qu.:2004	3rd Qu.: 162.8	A_agr : 2	NA	Stone : 11	Normal :2413	NA	SFoyer : 83	Typical : 864
Max. :755.0	Max. :11284	Max. :2010	Fair : 50	Max. :5.000	Max. :1488.0	Max. :6110	Max. :2010	Max. :1600.0	C_all : 25	NA	Wood : 5	Partial: 245	NA	Two_and_Half_Unf: 24	NA
NA	NA	NA	(Other) : 58	NA	NA	NA	NA	NA	I_all : 2	NA	NA	NA	NA	(Other) : 27	NA

Recommendation System Model - develop, train and test

Build Linear Models

```
##
## Naive RMSE in ,000 : 75.25
```

```
## # A tibble: 6 x 16
##   Sale_Price_T TotalArea Year_~1 Overa~2 Garag~3 Garag~4 Total~5 Year_~6 Mas_V~7
##          <dbl>     <int>   <int> <fct>     <dbl>   <dbl>   <dbl>   <int>   <dbl>
## 1          105      1792    1961 Above_~       1     730     882    1961       0
## 2          172      2658    1958 Above_~       1     312    1329    1958     108
## 3          244      4220    1968 Average       2     522    2110    1968       0
## 4          190      3258    1997 Average       2     482     928    1998       0
## 5          196      3208    1998 Above_~       2     470     926    1998      20
## 6          236      3232    1995 Average       2     608    1595    1996       0
## # ... with 7 more variables: MS_Zoning <fct>, Lot_Shape <fct>,
## #   Foundation <fct>, Sale_Condition <fct>, Garage_Finish <fct>,
## #   House_Style <fct>, Heating_QC <fct>, and abbreviated variable names
## #   1: Year_Built, 2: Overall_Cond, 3: Garage_Cars, 4: Garage_Area,
## #   5: Total_Bsmt_SF, 6: Year_Remod_Add, 7: Mas_Vnr_Area
```

```
## [1] 54.84304
```

method	RMSE
Just the average in ,000	75.25000

method	RMSE
Area Effect Model in in ,000	54.84304

```
## [1] 37.32972
```

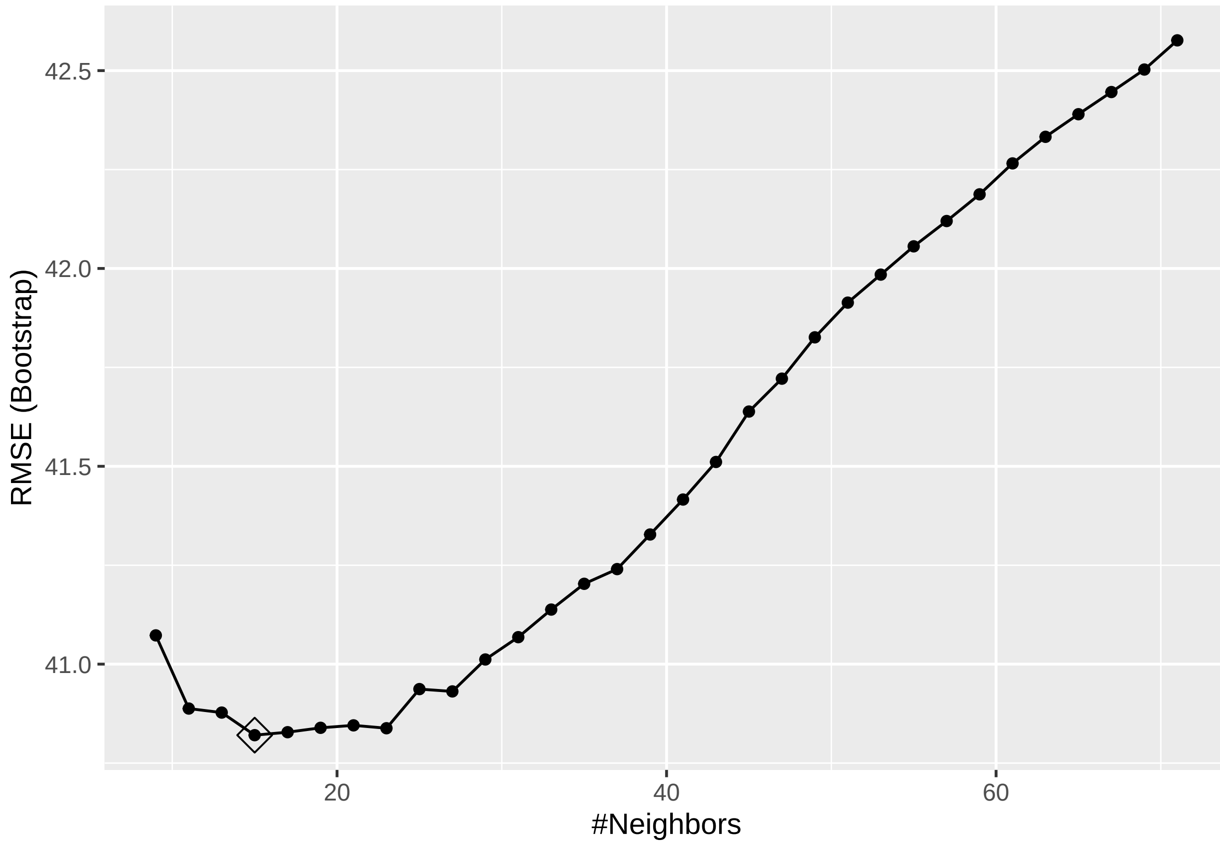
```
## # A tibble: 16 x 7
##    term                     estimate std.e~1 stati~2   p.value conf.low conf.h~3
##    <chr>                       <dbl>   <dbl>   <dbl>     <dbl>    <dbl>    <dbl>
##  1 (Intercept)              -1.87e+3 8.51e+1 -22.0   5.13e- 99 -2.04e+3 -1.70e+3
##  2 TotalArea                 3.16e-2 9.07e-4  34.9   5.40e-223  2.98e-2  3.34e-2
##  3 Year_Built                5.32e-1 4.18e-2  12.7   3.50e- 36  4.50e-1  6.14e-1
##  4 Overall_CondPoor          8.51e+0 1.91e+1   0.445 6.56e-  1 -2.90e+1  4.60e+1
##  5 Overall_CondFair          1.17e+1 1.57e+1   0.744 4.57e-  1 -1.91e+1  4.24e+1
##  6 Overall_CondBelow_Avera~  1.60e+1 1.52e+1   1.05  2.93e-  1 -1.38e+1  4.58e+1
##  7 Overall_CondAverage       3.18e+1 1.48e+1   2.14  3.22e-  2  2.71e+0  6.09e+1
##  8 Overall_CondAbove_Avera~  3.43e+1 1.48e+1   2.31  2.09e-  2  5.20e+0  6.33e+1
##  9 Overall_CondGood          4.44e+1 1.48e+1   2.99  2.81e-  3  1.53e+1  7.35e+1
## 10 Overall_CondVery_Good     4.93e+1 1.51e+1   3.27  1.11e-  3  1.97e+1  7.89e+1
## 11 Overall_CondExcellent     6.72e+1 1.60e+1   4.21  2.69e-  5  3.59e+1  9.86e+1
## 12 Garage_Cars               1.06e+1 2.19e+0   4.82  1.48e-  6  6.27e+0  1.49e+1
## 13 Garage_Area               2.79e-2 7.56e-3   3.70  2.23e-  4  1.31e-2  4.28e-2
## 14 Total_Bsmt_SF             4.10e-2 2.03e-3  20.2   3.49e- 85  3.70e-2  4.50e-2
## 15 Year_Remod_Add            3.99e-1 5.06e-2   7.87  4.85e- 15  2.99e-1  4.98e-1
## 16 Mas_Vnr_Area              5.13e-2 4.63e-3  11.1   6.09e- 28  4.22e-2  6.04e-2
## # ... with abbreviated variable names 1: std.error, 2: statistic, 3: conf.high
```

method	RMSE
Just the average in ,000	75.25000
Area Effect Model in in ,000	54.84304
Area + Year Built Effects Model in ,000	37.32972

Non-linear Models

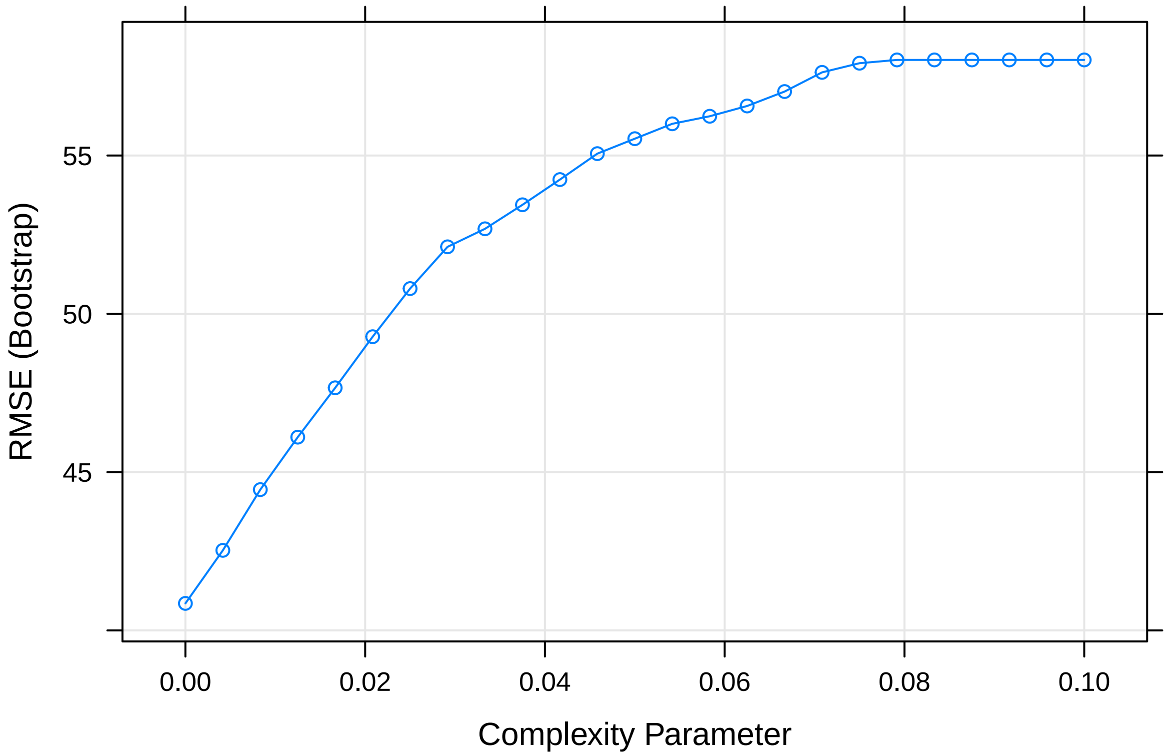
I took the optimum lamda for which the RMSE was the lowest. I built the model and ran the model against the final holdout set to validate the model performance

Train the final model



method	RMSE
Just the average in ,000	75.25000
Area Effect Model in in ,000	54.84304
Area + Year Built Effects Model in ,000	37.32972
Knn Model in ,000	37.09808

```
##  Accuracy
## 0.0239726
```




```
##         1         2         3         4         5         6         7         8
## 171.61538 195.41176 181.00000 205.71429 353.35714 217.57143 128.57895 444.55556
##         9        10        11        12        13        14        15        16
## 241.52941 227.50000 298.80000 173.00000 200.85714 274.00000 195.41176 119.21429
##        17        18        19        20        21        22        23        24
## 212.83333 278.46667 153.05556 144.71429 195.00000 150.35714 160.20000 165.63636
##        25        26        27        28        29        30        31        32
## 121.55556 102.15789 132.68750 160.20000 124.57895 179.45455 213.81818  73.93333
##        33        34        35        36        37        38        39        40
## 137.18182 132.66667 132.66667  73.93333 117.33333 141.12500 134.44444 274.71429
##        41        42        43        44        45        46        47        48
## 235.16667 179.47059 179.47059 114.20000 135.86667 194.20000 149.25000 320.07143
##        49        50        51        52        53        54        55        56
## 123.05556 259.11111  79.07143 126.76923 124.70000  90.00000 144.76471 239.26667
##        57        58        59        60        61        62        63        64
## 105.58333 286.12500 195.00000 259.11111 239.26667 179.93333 320.07143 201.43750
##        65        66        67        68        69        70        71        72
## 362.72727 194.20000  94.16667  73.93333 141.41667 291.92308 327.31579 259.11111
##        73        74        75        76        77        78        79        80
## 253.44444 149.57143 173.00000 235.16667 134.72727 152.33333 229.44444 224.00000
##        81        82        83        84        85        86        87        88
## 208.22222 134.71429 117.33333 126.00000 105.46667 117.20000 151.00000 146.00000
##        89        90        91        92        93        94        95        96
## 154.00000 134.44444 241.52941 320.07143 298.80000 310.11111 555.38462 225.62500
##        97        98        99       100       101       102       103       104
## 322.83333 168.66667 179.71429 205.36842 205.36842 204.76923 211.45455 193.14286
```

```
##       105       106       107       108       109       110       111       112
## 225.62500 264.00000 211.45455 224.00000 167.14286 353.35714 273.25000 320.07143
##       113       114       115       116       117       118       119       120
## 298.80000 253.00000 239.28571 255.77778 222.85714 229.44444 241.33333 173.00000
##       121       122       123       124       125       126       127       128
## 104.00000 151.00000 116.75000 353.11111 161.30000 200.85714 253.44444 168.62500
##       129       130       131       132       133       134       135       136
## 171.61538 171.61538 151.00000 236.90000 144.76471 126.76923 136.81250 160.20000
##       137       138       139       140       141       142       143       144
## 153.78571 208.22222 135.27273 156.75000 132.68750 113.50000 179.45455 146.00000
##       145       146       147       148       149       150       151       152
## 141.41667 135.27273 104.00000 109.60000 153.55556  90.90000 141.41667 175.00000
##       153       154       155       156       157       158       159       160
## 129.05556 116.86667 121.57895 116.00000 117.20000 141.12500 129.05556 101.61538
##       161       162       163       164       165       166       167       168
## 102.00000 132.66667 132.66667 179.47059 144.71429  56.81250 170.52941  92.83333
##       169       170       171       172       173       174       175       176
## 112.20000 155.10000 134.78947 168.62500 353.11111 310.11111 172.00000 149.25000
##       177       178       179       180       181       182       183       184
## 212.83333 181.00000 130.35294 239.76923 278.46667 229.44444 227.50000 182.75000
##       185       186       187       188       189       190       191       192
## 235.16667  90.90000  74.81250 112.20000  56.81250 121.57895 116.86667 156.75000
##       193       194       195       196       197       198       199       200
## 132.66667 143.90909 200.00000 105.58333 105.46667 229.44444 164.15789 200.00000
##       201       202       203       204       205       206       207       208
## 362.72727 320.07143 264.00000 327.31579 134.71429 169.71429  94.16667 135.23077
##       209       210       211       212       213       214       215       216
## 164.15789 181.00000 195.00000 212.83333 154.00000 105.46667 117.20000 154.00000
##       217       218       219       220       221       222       223       224
## 119.21429 444.55556 179.93333 177.87500 274.71429 273.25000 320.07143 310.11111
##       225       226       227       228       229       230       231       232
## 206.44444 200.00000 181.00000 180.69231 212.00000 110.30000 116.75000 229.44444
##       233       234       235       236       237       238       239       240
## 195.00000 253.44444 195.00000 153.78571 146.00000 139.00000 236.90000 174.73333
##       241       242       243       244       245       246       247       248
## 174.73333 116.75000 171.61538 160.88889 143.84211  56.81250 109.60000 181.08333
##       249       250       251       252       253       254       255       256
## 141.41667 132.68750 171.61538 145.00000 135.86667 124.57895 119.21429 128.57895
##       257       258       259       260       261       262       263       264
## 291.92308 117.33333 156.75000 129.05556 137.23529 136.81250 110.30000 134.78947
##       265       266       267       268       269       270       271       272
## 136.81250 136.81250 132.66667 116.00000  81.87500 170.52941  74.81250 129.05556
##       273       274       275       276       277       278       279       280
## 137.23529 146.00000 152.77778 152.77778 145.55556 109.60000 126.00000  56.81250
##       281       282       283       284       285       286       287       288
## 142.30769 135.86667  90.90000 126.76923 144.76471 212.00000 241.52941 327.31579
##       289       290       291       292       293       294       295       296
## 205.36842 200.00000 200.00000 164.15789 270.64706 224.00000 204.76923 241.52941
##       297       298       299       300       301       302       303       304
## 135.23077 135.23077 135.23077  97.00000 151.00000 138.00000 116.00000 176.00000
##       305       306       307       308       309       310       311       312
## 194.92308 155.10000 109.60000 175.00000 168.66667  89.30000 291.53846 152.33333
##       313       314       315       316       317       318       319       320
## 273.25000 212.00000 194.20000 278.46667 173.00000 153.00000  94.16667 123.05556
```

```
##       321       322       323       324       325       326       327       328 
## 123.05556 176.00000 144.71429 239.26667 239.26667 177.87500 180.69231 185.50000 
##       329       330       331       332       333       334       335       336 
## 299.88889 416.57143 205.36842 154.00000 150.35714 153.78571 146.00000 135.27273 
##       337       338       339       340       341       342       343       344 
## 189.25000 327.31579 362.72727 362.72727 278.46667 327.31579 222.85714 286.12500 
##       345       346       347       348       349       350       351       352 
## 225.62500 227.50000 194.00000 172.00000 182.15385 205.36842 153.05556 201.43750 
##       353       354       355       356       357       358       359       360 
## 201.43750 193.14286 161.30000 193.14286 458.00000 353.35714 299.88889 323.62500 
##       361       362       363       364       365       366       367       368 
## 255.77778 270.64706 239.28571 286.12500 139.37500 149.25000 213.44444 135.27273 
##       369       370       371       372       373       374       375       376 
## 274.00000 274.00000 177.87500 175.18750 175.18750 211.11765 274.71429 179.47059 
##       377       378       379       380       381       382       383       384 
## 153.00000 134.71429 136.81250 121.55556 102.00000 121.12500 150.35714 236.90000 
##       385       386       387       388       389       390       391       392 
## 160.88889 235.16667 134.44444 170.52941 112.20000 160.88889 110.30000 168.62500 
##       393       394       395       396       397       398       399       400 
## 128.57895 145.00000 116.75000 181.08333 353.11111 109.60000 128.00000 134.71429 
##       401       402       403       404       405       406       407       408 
## 170.52941  92.83333 110.30000 134.44444 102.00000 101.61538 132.66667 194.92308 
##       409       410       411       412       413       414       415       416 
##  94.16667 102.00000 121.57895 135.80000 124.87500 105.58333 117.20000  74.81250 
##       417       418       419       420       421       422       423       424 
## 291.92308 138.00000 124.57895 153.00000 143.90909 194.92308 146.00000 143.84211 
##       425       426       427       428       429       430       431       432 
## 116.86667 104.00000 145.00000 278.46667 252.00000 444.55556 205.36842 241.52941 
##       433       434       435       436       437       438       439       440 
## 149.25000 212.83333 224.00000 142.30769 142.30769 123.05556 126.00000 212.00000 
##       441       442       443       444       445       446       447       448 
## 212.83333 182.75000 140.10000 208.22222 555.38462 144.71429 135.27273 116.75000 
##       449       450       451       452       453       454       455       456 
## 117.33333 124.57895 194.92308 155.10000 181.08333 135.23077 137.23529 137.23529 
##       457       458       459       460       461       462       463       464 
## 310.11111 310.11111 179.93333 139.37500 153.00000 136.81250 130.35294 105.46667 
##       465       466       467       468       469       470       471       472 
##  89.30000  94.16667 130.35294 179.47059 171.61538 135.23077 123.05556 136.81250 
##       473       474       475       476       477       478       479       480 
## 134.72727 182.15385 264.00000 175.18750 182.15385 298.80000 164.15789 211.00000 
##       481       482       483       484       485       486       487       488 
## 299.88889 458.00000 255.07692 154.00000 264.00000 153.00000 153.00000 278.46667 
##       489       490       491       492       493       494       495       496 
## 278.46667 444.55556 194.00000 189.25000 189.25000 204.76923 182.15385 278.46667 
##       497       498       499       500       501       502       503       504 
## 211.45455 253.00000 362.72727 310.11111 206.44444 195.41176 182.15385 121.12500 
##       505       506       507       508       509       510       511       512 
## 175.18750 110.30000 139.00000 104.00000 217.57143  90.00000 121.55556 112.20000 
##       513       514       515       516       517       518       519       520 
## 239.26667 134.71429 135.86667 145.00000 160.20000 141.41667 142.30769 121.55556 
##       521       522       523       524       525       526       527       528 
##  90.90000  56.81250 101.61538 136.81250 134.71429 179.47059 179.47059  89.30000 
##       529       530       531       532       533       534       535       536 
## 126.76923 121.57895 137.23529 145.55556 154.00000 141.12500 109.60000 105.58333
```

```
##       537       538       539       540       541       542       543       544
## 153.78571 112.20000 123.50000 116.00000 181.08333 122.85714 137.18182 129.05556
##       545       546       547       548       549       550       551       552
##  90.00000  74.81250 101.61538 134.71429 153.78571 134.78947 134.44444 179.37500
##       553       554       555       556       557       558       559       560
## 231.14286 143.84211 144.76471  74.81250 270.64706 239.28571 149.25000 227.50000
##       561       562       563       564       565       566       567       568
## 200.85714 118.25000 164.15789 274.71429 322.83333 135.23077 253.00000 241.52941
##       569       570       571       572       573       574       575       576
## 154.00000 255.07692 105.46667 155.10000 156.75000 208.22222 181.08333 217.57143
##       577       578       579       580       581       582       583       584
##  56.81250  89.30000 235.16667 175.00000 396.29412 195.00000 118.25000 105.46667
##       585       586       587
##  73.93333 144.71429 241.71429
```

method	RMSE
Just the average in ,000	75.25000
Area Effect Model in in ,000	54.84304
Area + Year Built Effects Model in ,000	37.32972
Knn Model in ,000	37.09808
Knn Model in ,000	37.66704

```
##   Accuracy
## 0.01718213
```

Final Result and improvements over time

RMSEs over Model Accuracies of the Models

Conclusion

I have used linear model with regularization to build this recommendation system. I came to a reasonable level of accuracy. Linear model is relatively simple to start with but not the best and we realized that during our study. We need more sophisticated models to enhance the accuracy - may be the random forest would be better suited for this prediction.

Reference -

Introduction to Data Science https://modeldata.tidymodels.org/reference/ames.html - Ames Housing Data https://www.investopedia.com