SDM A2 Hunters Green Home Sale

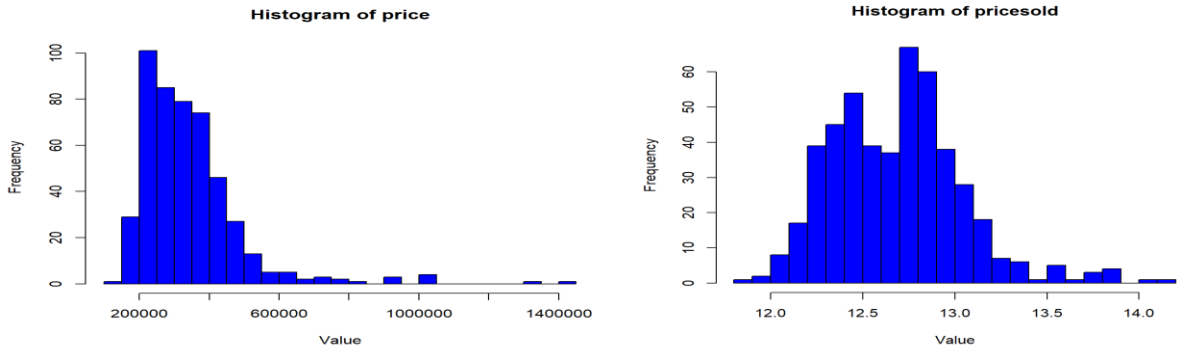
A. PREDICTOR TABLE: Y= PriceSold

Predictor	Effect	Rationale
Beds	+	A property with more bedrooms or beds will usually have a higher price than a property with fewer bedrooms or beds as it can accommodate more people.
Bath Total	+	A house with more bathrooms generally has a higher price than a similar house with fewer bathrooms as having more bathrooms provides added convenience and comfort for the homeowners and their guests, especially in larger households or when entertaining.
sqft	+	In general, larger houses with more square footage tend to be more expensive than smaller houses with less square footage as larger house offers more living space, which is generally more desirable for homeowners.
garages	+	A house with more garages, especially if they are attached and provide additional storage space, generally has a higher price as they provide valuable space to protect vehicles and can be used for storage or as a workshop
roof	+/-	The type of roofing material used can have an impact on the overall value of a property due to the aesthetic appeal, durability, and energy efficiency of the roof. Tile, slate, and other high-end roofing materials are generally more expensive than traditional asphalt shingle roofs
Lotsqft	+	larger lots tend to be more desirable and can command a higher price as then can offer more privacy, more outdoor space, and more potential for future development
Age of house "yearsold-Yrblt"	-	newer homes may command a higher price than older homes, all other things being equal. This is because newer homes often have more modern features and amenities, are built to more stringent building codes and energy standards
pool	+/-	The type of pool, whether it is a community or private pool, can also affect the price of a house. A private pool, which is only accessible to the residents of the house, can often command a higher price than a community pool, which is shared by multiple households
lppersqft	+	As the list price per square ft increase, the price increases
Cdom_cumulative daysonmarket	-	A longer time on the market can sometimes indicate that the property is overpriced or has other issues that are making it less attractive to potential buyers. As a result, sellers may have to lower their asking price in order to attract more interest and secure a sale. This also includes the days in which the property is off the list
List price	+	List price increases, the price of the house sold increases as the selling price may be close to listing price
Spa	+	Having spa could increase the price of the house
splsale	+/-	Having a special sale could increase the price of the house if it is a special one or could decrease if any discounts were given
Excluded Variables		
Slnoskm	0	NA
Status	0	This attribute has the same value "Sold"

Address	0	Address does not affect the price of the property unless the geographical coordinates are given
Subdivn	0	This category value does not affect price
bathfull	0	This can be included with bath total and having it as an exclusive variable does not make any significant changes
bathhalf	0	This can be included with bath total and having it as an exclusive variable does not make any significant changes
sppersqft	0	Sold price per sq ft directly dependent on the sold price

Here the target variable is (y = pricesold). Using the literature search, we consider number of beds, number of baths, area of house, extra area (diff between lotsqft and sqft), number of garages, type of roof, type of pool, age of house, list price per sqft, cumulative days on market, list price, presence of extra features like spa, and type of the sale. We have calculated the age of house instead of including the pending date and year build as different factors as they may not individually affect the price of house. Also we created a new filed named “extra sqft” which is difference between the sqft and lotsqft as this area can help in doing further constructions and increase the price of the house.

Let us see how the histogram is behaving (pricesold) both normal and logarithmic



The normal histogram shows us there are some outliners in the graph

There are 4 missing values in number of garages. I have replaced these null values with mode i.e, 2. Also there are multiple missing values in spa category. So I have not used that as a factor in price sold as it may affect the predictions due to huge null values.

Using all the assumptions the models are

Model 1

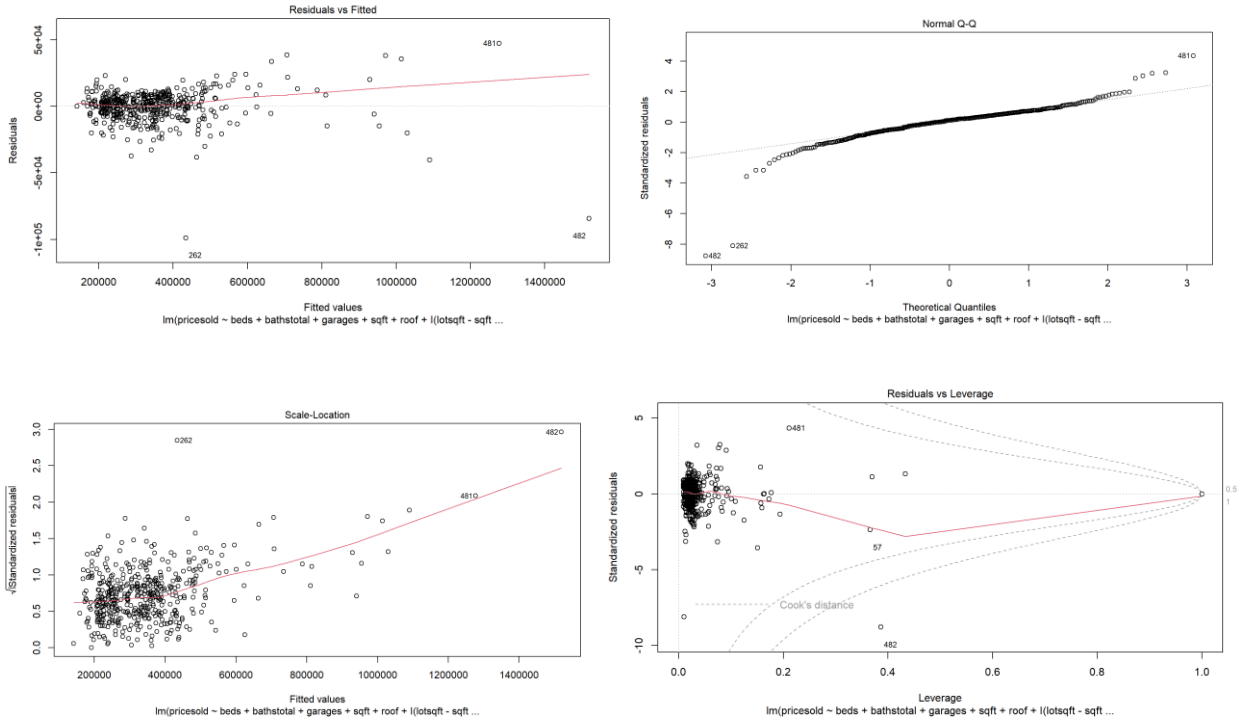
```
psmod1=lm(pricesold~beds+bathstotal+sqft+roof+garages+lotsqft+yrblt+pool+lppersqft+listprice+spa+splsale, data = house_data)
```

Model 2

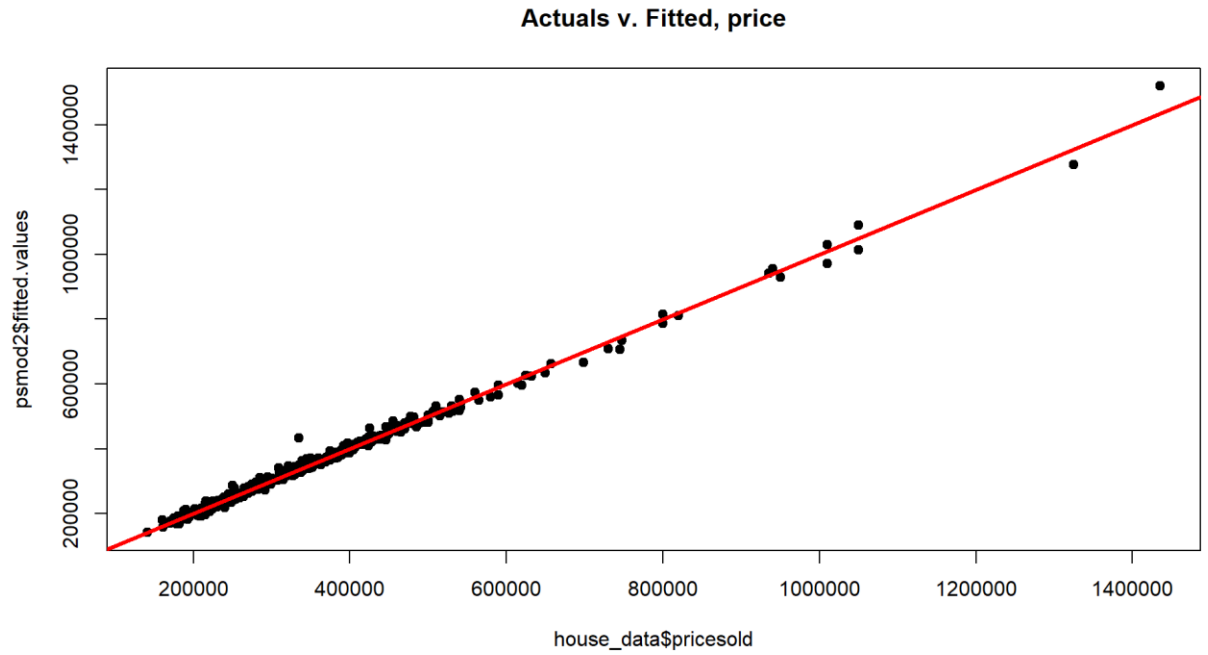
```
psmod2=lm(pricesold~beds+bathstotal+garages+roof+I(lotsqft - sqft)+houseage
+pool+lppersqft+listprice+splsale, data = house_data)
```

model 2 is better fit when compared to model 1 as it have broader attributes and transformed attributes.

Model 2.



Lets see the linearity of model.



	pricesold			pricesold		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	544745.65	-435044.85 – 1524536.15	0.275	-35679.30	-65530.29 – -5828.31	**0.019**
beds	1126.50	-1164.28 – 3417.29	0.334	2868.77	612.50 – 5125.04	**0.013**
bathstotal	4614.30	2067.65 – 7160.95	**<0.001**	5915.95	3323.32 – 8508.58	**<0.001**
sqft	25.47	16.97 – 33.97	**<0.001**			
roof [Concrete, Tile]	-280.61	-34884.71 – 34323.48	0.987	5601.09	-30059.09 – 41261.26	0.758
roof [Other]	980.48	-33425.79 – 35386.75	0.955	5868.81	-29647.76 – 41385.38	0.746
roof [Shake, Shingle]	6056.06	-28695.48 – 40807.59	0.732	5536.06	-29793.42 – 40865.53	0.758
roof [Shingle]	-5700.05	-30083.59 – 18683.49	0.646	-2839.23	-28006.17 – 22327.70	0.825
roof [Shingle, Tile]	-3089.42	-37452.94 – 31274.11	0.860	-105.38	-35612.24 – 35401.48	0.995
roof [Slate]	-1494.98	-35986.82 – 32996.86	0.932	132.97	-35458.55 – 35724.50	0.994
roof [Slate, Tile]	-24706.88	-60243.04 – 10829.29	0.173	-26580.16	-62149.76 – 8989.44	0.143
roof [Tile]	-6146.14	-30683.71 – 18391.42	0.623	-1371.64	-26656.76 – 23913.48	0.915
garages	57.49	-2757.57 – 2872.56	0.968	1694.65	-1152.17 – 4541.46	0.243
lotsqft	0.55	0.20 – 0.90	**0.002**			
yrblt	-318.50	-808.05 – 171.06	0.202			
pool [None]	859.49	-3298.94 – 5017.91	0.685	1574.98	-2702.69 – 5852.65	0.470
pool [Private]	1889.20	-2221.69 – 6000.09	0.367	5121.32	1099.10 – 9143.53	**0.013**
pool [Private, Community]	2441.26	-2183.62 – 7066.15	0.300	6206.70	1631.82 – 10781.59	**0.008**
lppersqft	657.80	482.79 – 832.81	**<0.001**	181.68	99.32 – 264.04	**<0.001**
listprice	0.71	0.66 – 0.76	**<0.001**	0.86	0.84 – 0.87	**<0.001**
splsale [Bank Owned/REO]	9953.18	-5696.87 – 25603.24	0.212			
splsale [None]	9368.85	-5323.71 – 24061.41	0.211			
splsale [Short Sale]	11283.56	-5870.48 – 28437.60	0.197			
lotsqft - sqft				0.46	0.12 – 0.81	**0.009**
houseage				619.37	214.99 – 1023.76	**0.003**
Observations	482			482		
R^2 / R^2 adjusted	0.994 / 0.993			0.993 / 0.993		

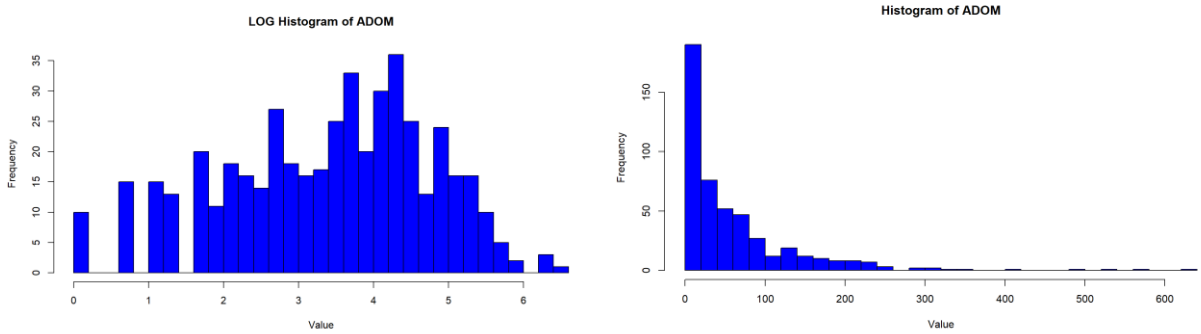
Interpretations:

1. Increase in count of one bed, increase the price of house by 1035$
2. Increase in 1 bath increase the house price by 4629 $
3. Increase in one garage increase the price by 40$
4. For increase in 100 sqft, the price increase by 2560$
5. From the plot between types of roof and price, rooftype “tile” has more price than others
6. Every 1000 increase in extra lotsqft, the price increased by 571$

B. PREDICTOR TABLE: Y= Agent Days on Market

Predictor	Effect	Rationale
Beds	+/-	Larger homes with more bedrooms may have a longer ADOM than smaller homes with fewer bedrooms as larger homes tend to have higher listing prices, which can make them less accessible to some buyers. They may be sold early if the price is listed less
Bath Total	+/-	larger homes with more bathrooms may have a longer ADOM than smaller homes with fewer bath. This is because larger homes tend to have higher listing prices, which can make them less accessible to some buyers
sqft	+/-	In general, larger houses with more square footage tend to be more expensive than smaller houses with less square footage as larger house offers more living space, larger homes may have a smaller pool of potential buyers, as they may be too big for some buyers or out of their price range
garages	+/-	Homes with garages are in more demand as most of the people have their own vehicle and prefer having garage. Some times they tend to be more expensive and increase ADOM
roof	+/-	Roof types with more reliable types are in more demand and they tend to be sold earlier than the other types
Lotsqft	-	Most families with children prefer having play area and houses providing this may be sold earlier
Age of house "yearsold-Yrblt"	+	New houses are sold earlier than the old ones
pool	+/-	Type of pool and the listing price including may affect the ADOM
lppersqft	+	As the list price per square ft increase, the price increases and ADOM increases
List price	+	List price increases, the number of people tends to buy houses with low listing prices and may not prefer houses with more listing price
splsale	-	Houses in special sale are advertised more than in normal sale and they may be sold first
Excluded Variables		
Slnoskm	0	NA
Status	0	This attribute has the same value "Sold"
Address	0	Address does not affect the price of the property and therefore the ADOM
Subdivn	0	This category value does not affect ADOM
bathfull	0	This can be included with bath total and having it as an exclusive variable does not make any significant changes
bathhalf	0	This can be included with bath total and having it as an exclusive variable does not make any significant changes
sppersqft	0	Sold price per sq ft is after ADOM and it cannot be a factor predicting ADOM

The same data which is already preprocessed is being used. We can see that number of beds, baths, sqft, list price per square ft are directly in relation with the list price which is the major factor affecting ADOM. We know that higher beds, higher baths and higher sqft with less list price are tend to have more demand and hence less ADOM. Let us see how the histogram of ADOM is.



We can see that normal histogram is highly “positive Skewed” but log curve is having bell shaped curve.

To work on logarithmic data, we need to transform values that are not finite.

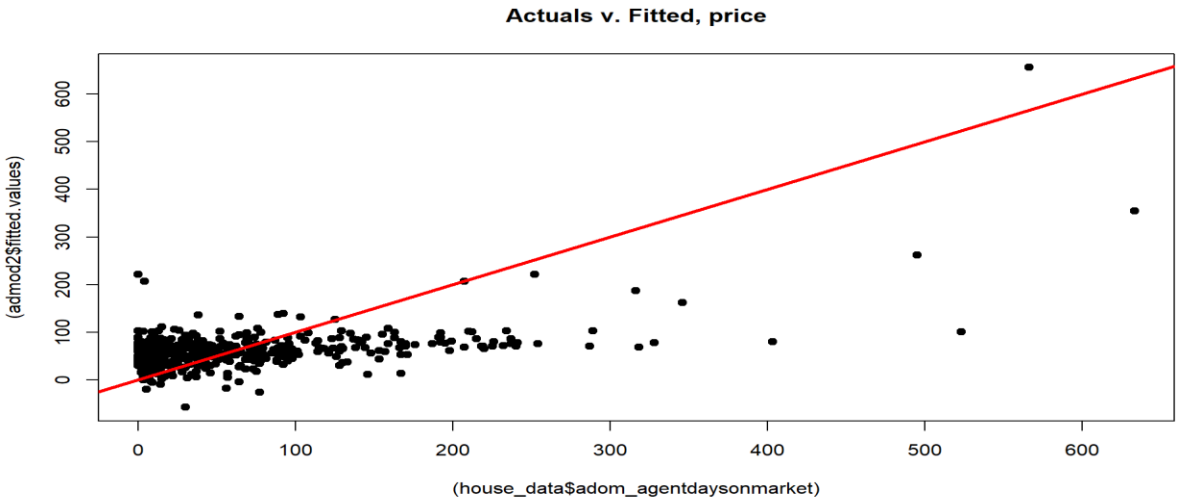
Model 1.

admod1=lm(adom_agentdaysonmarket~beds+bathstotal+sqft+roof+garages+lotsqft+yrblt +pool+lppersqft+listprice+splsale, data = house_data)

model 2.

admod2=lm(adom_agentdaysonmarket~I(((beds+bathstotal))/lppersqft)+I(listprice*listprice)+I(lppersqft *lppersqft)+pool+garages+houseage+roof+I(lotsqft-sqft)+pool+splsale, data = house_data)

here in this model, we created two new attributes . ((Bed+bath)/listprice per sqft) and square of list price as these influence the ADOM directly where other factors influence it indirectly. The fit of linear module is as follows.



The summary of the models is as follows.

	adom_agentdaysonmarket			adom_agentdaysonmarket		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-4085.62	-9705.47 – 1534.23	0.154	132.79	-40.86 – 306.44	0.134
beds	-5.76	-18.90 – 7.37	0.389			
bathstotal	-13.21	-27.81 – 1.40	0.076			
sqft	-0.07	-0.12 – -0.02	**0.004**			
roof [Concrete, Tile]	74.61	-123.88 – 273.09	0.460	93.66	-99.60 – 286.91	0.341
roof [Other]	14.00	-183.34 – 211.35	0.889	18.63	-173.54 – 210.81	0.849
roof [Shake, Shingle]	-48.73	-248.05 – 150.60	0.631	-41.38	-235.79 – 153.04	0.676
roof [Shingle]	12.08	-127.78 – 151.94	0.865	12.15	-123.93 – 148.24	0.861
roof [Shingle, Tile]	-29.70	-226.80 – 167.40	0.767	-30.81	-222.99 – 161.38	0.753
roof [Slate]	-12.39	-210.23 – 185.44	0.902	-15.43	-208.10 – 177.24	0.875
roof [Slate, Tile]	119.93	-83.89 – 323.76	0.248	124.69	-74.38 – 323.75	0.219
roof [Tile]	24.66	-116.08 – 165.40	0.731	27.59	-109.13 – 164.31	0.692
garages	-7.31	-23.46 – 8.83	0.374	-8.47	-23.60 – 6.67	0.272
lotsqft	0.00	-0.00 – 0.00	0.868			
yrblt	2.23	-0.58 – 5.04	0.119			
pool [None]	12.91	-10.94 – 36.77	0.288	11.45	-11.77 – 34.67	0.333
pool [Private]	9.33	-14.25 – 32.91	0.437	11.59	-10.35 – 33.53	0.300
pool [Private, Community]	9.42	-17.10 – 35.95	0.485	11.67	-13.32 – 36.66	0.359
lppersqft	-3.31	-4.32 – -2.31	**<0.001**			
listprice	0.00	0.00 – 0.00	**<0.001**			
splsale [Bank Owned/REO]	53.36	-36.40 – 143.13	0.243	65.67	-22.28 – 153.61	0.143
splsale [None]	84.80	0.52 – 169.07	**0.049**	89.89	7.99 – 171.79	**0.032**
splsale [Short Sale]	74.07	-24.32 – 172.47	0.140	93.95	-1.97 – 189.87	0.055
((beds + bathstotal))/lppersqft				-663.53	-1454.21 – 127.15	0.100
listprice * listprice				0.00	0.00 – 0.00	**<0.001**
lppersqft * lppersqft				-0.01	-0.01 – -0.00	**<0.001**
houseage				-2.25	-4.47 – -0.03	**0.047**
lotsqft - sqft				0.00	-0.00 – 0.00	0.696
Observations	482			482		
R^2 / R^2 adjusted	0.269 / 0.234			0.299 / 0.269		

Interpretation :

1. With increase in the number of beds and decrease in list price simultaneously, reduces the ADOM and helps to sell the properity sooner.
2. As the age of house increases, the ADOM decreases from the model
3. Having a private pool or community pool helps in decreasing ADOM
4. Roof type tile or slate have more impact on ADOM