Dummy Variable

Dataset Background:

This dataset contains information on the sales price of houses based on the specifications of the house. The goal is to predict the sale price of the house based on the lot size, number of bedrooms, number of baths, number of stories, and the availability of a basement, driveway and recreation room.

Dataset Glimpse:

sale price	lot size	#bedroom	#bath	#stories	driveway	rec room	basement
42000	5850	3	1	2	1	0	1
38500	4000	2	1	1	1	0	0
49500	3060	3	1	1	1	0	0
60500	6650	3	1	2	1	1	0
61000	6360	2	1	1	1	0	0
66000	4160	3	1	1	1	1	1
66000	3880	3	2	2	1	0	1
69000	4160	3	1	3	1	0	0
83800	4800	3	1	1	1	1	1
88500	5500	3	2	4	1	1	0

Total Number of Rows: 546

Total Number of Columns: 8

Column Details:

- sale_price: the price of the house in dollars.
- lot_size: the size of the lot in acres.
- #bedroom: the number of bedrooms in the house.
- #bath: the number of baths in the house.
- #stories: the number of stories in the house.
- driveway: the availability of a driveway in the house, 1-yes, 0-no.
- rec_room: the availability of a recreation room in the house, 1-yes, 0-no.
- basement: the availability of a basement in the house, 1-yes, 0-no.

Main Dependent Variable: sale_price.

Using SPSS Software, we have analysed the data:

Descriptive Statistics:

	SALE PRICE	LOT SIZE	BEDROOM	BATH	STORIES	DRIVEWAY	REC ROOM	BASEMENT
Mean	68121.60	5150.266	2.965201	1.285714	1.807692	0.858974	0.177656	0.349817
Median	62000.00	4600.000	3.000000	1.000000	2.000000	1.000000	0.000000	0.000000
Maximum	190000.0	16200.00	6.000000	4.000000	4.000000	1.000000	1.000000	1.000000
Minimum	25000.00	1650.000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000
Std. Dev.	26702.67	2168.159	0.737388	0.502158	0.868203	0.348367	0.382573	0.477349
Skewness	1.206503	1.319121	0.494509	1.587718	1.071702	-2.062787	1.686684	0.629815
Kurtosis	4.930871	5.726370	3.717276	5.144182	3.639259	5.255088	3.844902	1.396667
Jarque-Bera	217.2820	327.4503	33.95759	333.9908	113.8144	502.9064	275.1263	94.57959
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	37194392	2812045.	1619.000	702.0000	987.0000	469.0000	97.00000	191.0000
Sum Sq. Dev.	3.89E+11	2.56E+09	296.3388	137.4286	410.8077	66.14103	79.76740	124.1850
Observations	546	546	546	546	546	546	546	546

Inferences:

- The variable sale_price is right skewed, ranging between 25000 to 190000 dollars.
- The variable lot_size is right skewed, ranging between 1650 to 16200 acres of land.
- The variable bedroom is slightly right skewed, ranging between 1 to 6 bedrooms.
- The variable bath is right skewed, ranging between 1 to 4 baths.
- The variable stories is right skewed, ranging between 1 to 4 stories.
- The variable driveway is left skewed, with more houses without a driveway.
- The variable rec_room is right skewed, with more houses with a recreation room.
- The variable basement is slightly right skewed, with more houses with a basement.
- There is no missing data.

Correlation Analysis:

	SALE PRICE	LOT SIZE	BEDROOM	BATH	STORIES	DRIVEWAY	REC ROOM	BASEMENT
SALE...	1.000000	0.535796	0.366447	0.516719	0.421190	0.297167	0.254960	0.186218
LOT S...	0.535796	1.000000	0.151851	0.193833	0.083675	0.288778	0.140327	0.047487
_BED...	0.366447	0.151851	1.000000	0.373769	0.407974	-0.011996	0.080492	0.097201
BATH	0.516719	0.193833	0.373769	1.000000	0.324066	0.041955	0.126892	0.102791
_STO...	0.421190	0.083675	0.407974	0.324066	1.000000	0.122499	0.042281	-0.173860
DRIV...	0.297167	0.288778	-0.011996	0.041955	0.122499	1.000000	0.091959	0.043428
REC_...	0.254960	0.140327	0.080492	0.126892	0.042281	0.091959	1.000000	0.372434
BASE...	0.186218	0.047487	0.097201	0.102791	-0.173860	0.043428	0.372434	1.000000

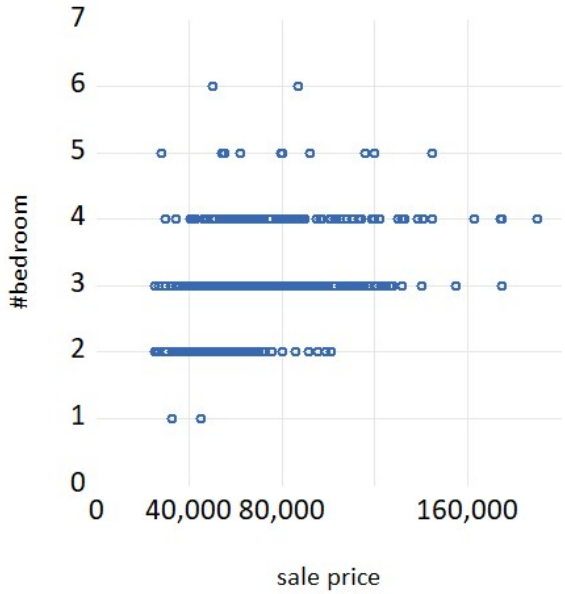
Inferences:

- The variables sale_price and lot_size have only the possibility of positive linear correlation, having correlation coefficient 0.53.
- The variables sale_price and bedroom have only the possibility of positive linear correlation, having correlation coefficient 0.36.
- The variables sale_price and bath have only the possibility of positive linear correlation, having correlation coefficient 0.52.
- The variables sale_price and stories have only the possibility of positive linear correlation, having correlation coefficient 0.42.
- The variables sale_price and driveway have no possible linear correlation, having correlation coefficient 0.30.
- The variables sale_price and rec_room have no possible linear correlation, having correlation coefficient 0.25.
- The variables sale_price and basement have no possible linear correlation, having correlation coefficient 0.19.
- The variables lot_size and bedroom have no possible linear correlation, having correlation coefficient 0.15.
- The variables lot_size and bath have no possible linear correlation, having correlation coefficient 0.19.
- The variables lot_size and stories have no possible correlation, having correlation coefficient 0.08.
- The variables lot_size and driveway have only the possibility of positive linear correlation, having correlation coefficient 0.29.
- The variables lot_size and rec_room have no possible linear correlation, having correlation coefficient 0.14.
- The variables lot_size and basement have no possible linear correlation, having correlation coefficient 0.05.
- The variables bedroom and bath have only the possibility of positive linear correlation, having correlation coefficient 0.38.

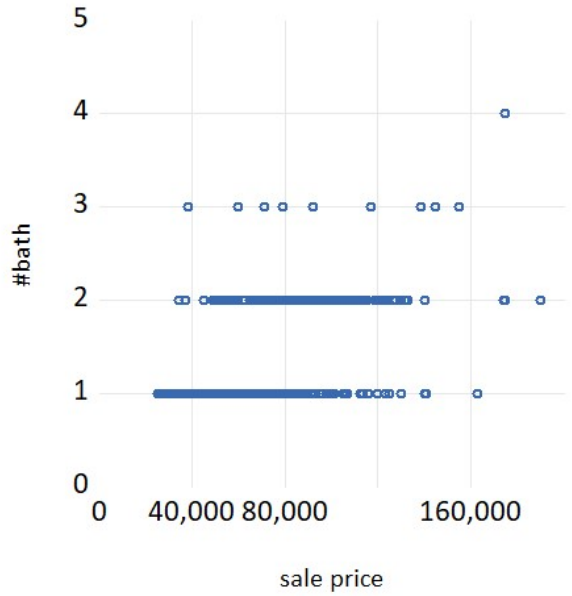
- The variables bedroom and stories have only the possibility of positive linear correlation, having correlation coefficient 0.41.
- The variables bedroom and driveway have no possible linear correlation, having correlation coefficient -0.01.
- The variables bedroom and rec_room have no possible linear correlation, having correlation coefficient 0.08.
- The variables bedroom and basement have no possible linear correlation, having correlation coefficient 0.1.
- The variables bath and stories have only the possibility of positive linear correlation, having correlation coefficient 0.32.
- The variables bath and driveway have no possible linear correlation, having correlation coefficient 0.04.
- The variables bath and rec_room have no possible linear correlation, having correlation coefficient 0.13.
- The variables bath and basement have no possible linear correlation, having correlation coefficient 0.10.
- The variables stories and driveway have no possible linear correlation, having correlation coefficient 0.12.
- The variables stories and rec_room have no possible linear correlation, having correlation coefficient 0.04.
- The variables stories and basement have no possible linear correlation, having correlation coefficient -0.17.
- The variables driveway and rec_room have no possible linear correlation, having correlation coefficient 0.09.
- The variables driveway and basement have no possible linear correlation, having correlation coefficient 0.04.
- The variables rec_room and basement have only the possibility of positive linear correlation, having correlation coefficient 0.37.

Scatter Plots:

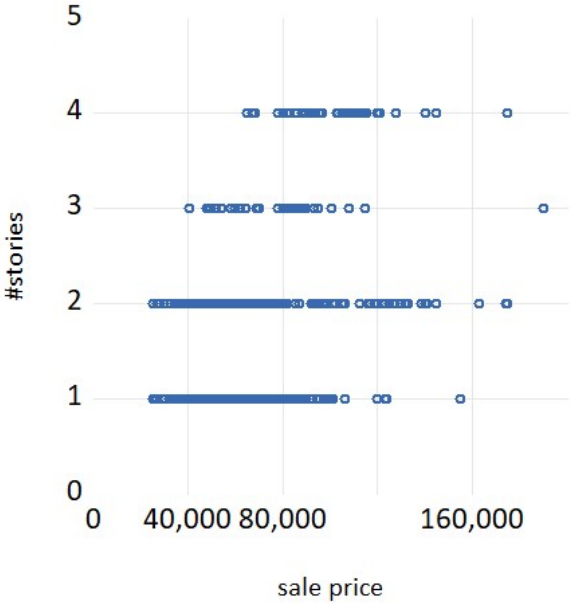
Inference: the variables sale_price and lot_size have only the possibility of positive linear correlation.



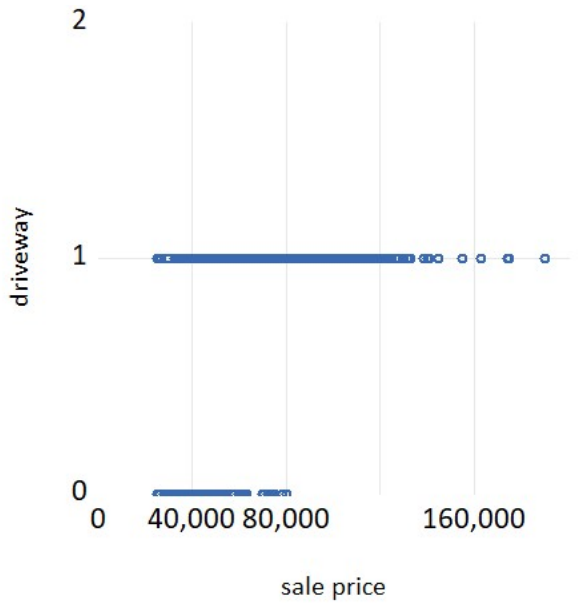
Inference: most of the house have 2, 3 or 4 rooms and houses with 4 rooms have the highest prices.



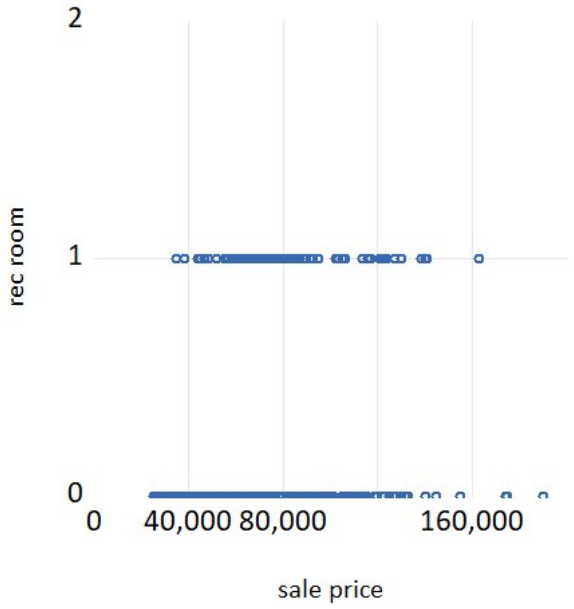
Inference: most of the houses have 1 or 2 baths, and houses with 2 baths have the highest prices.



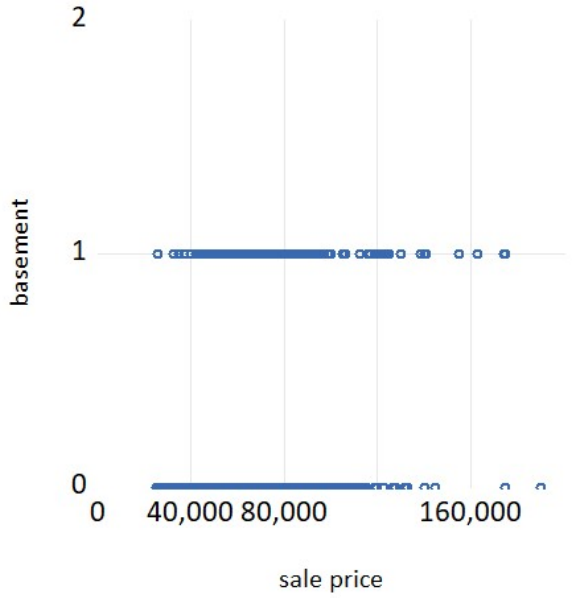
Inference: most of the houses have 1 or 2 stories and houses with 3 stories have the highest prices.



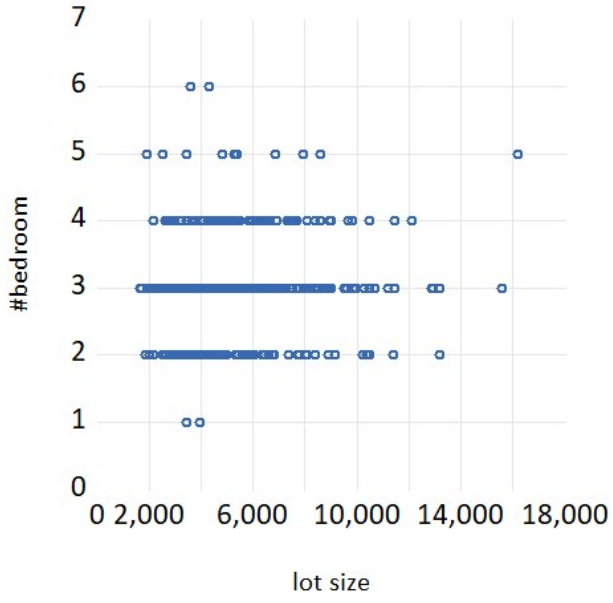
Inference: most of the houses have a driveway and the houses having a driveway have higher prices.



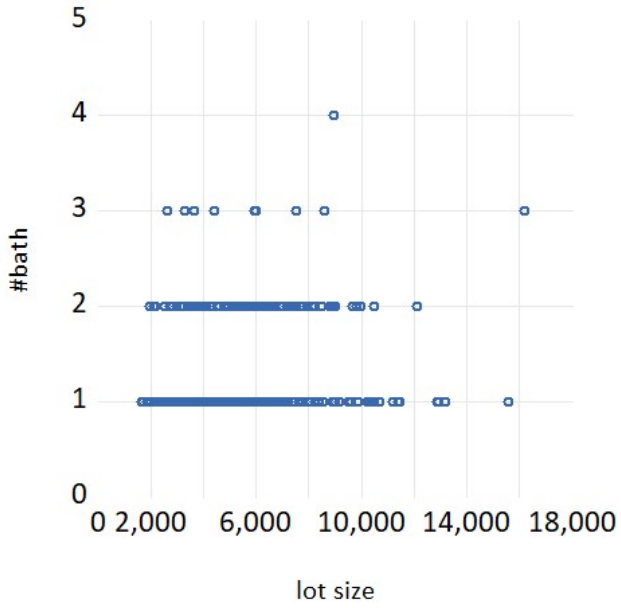
Inference: there are equal number of houses with and without recreation rooms, and houses without recreation rooms have higher prices.



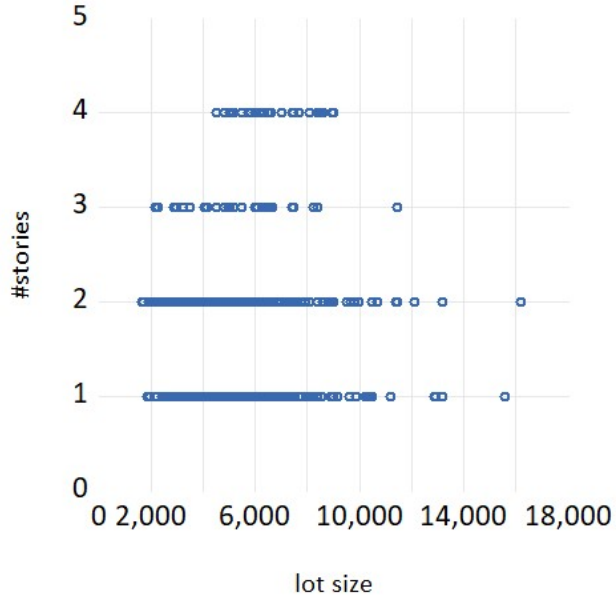
Inference: there is an equal distribution of houses with and without a basement, and houses without a basement have higher prices.



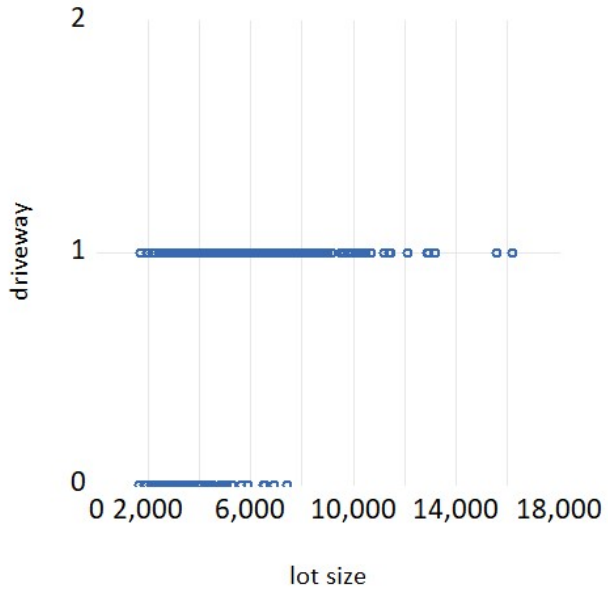
Inference: most of the houses have 2, 3 or 4 rooms and houses with 3 and 5 bedrooms have bigger lot sizes.



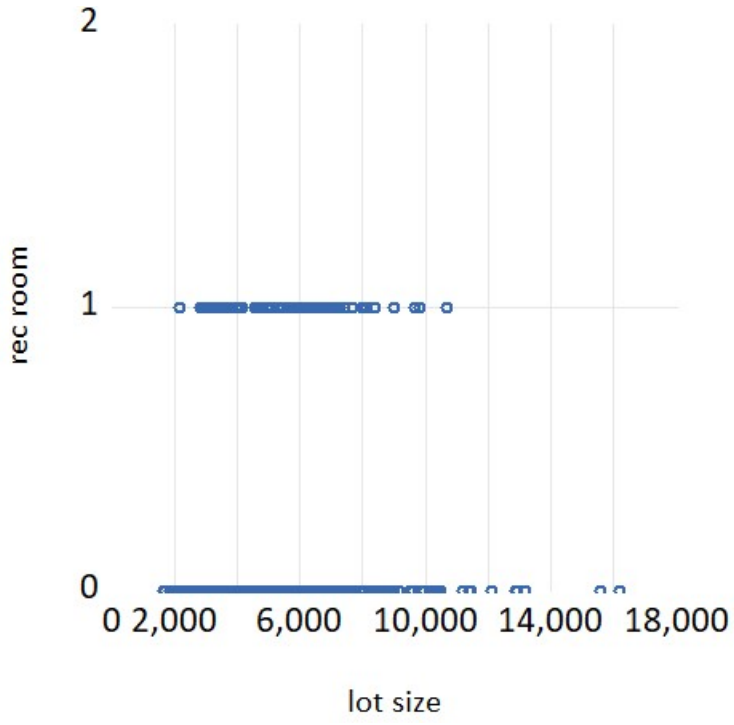
Inference: most of the houses have 1 or2 baths and houses with 1 or 3 baths have bigger lot sizes.



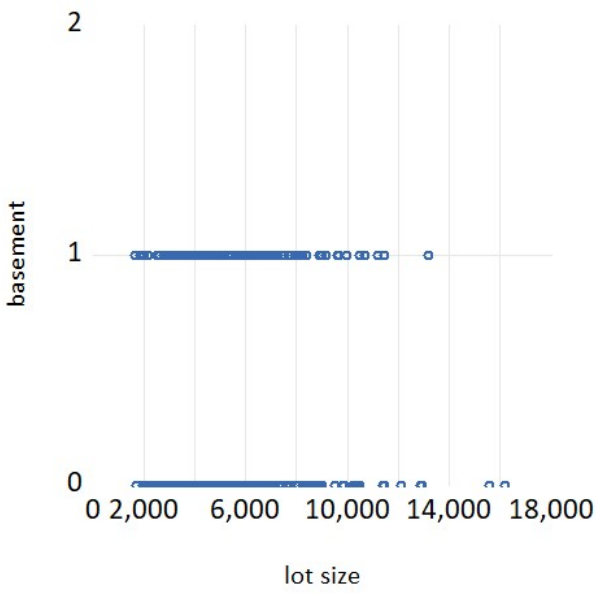
Inference: most of the houses have 1 or 2 stories and houses with 1 or 2 stories have bigger lot sizes.



Inference: most of the houses have a driveway and houses with a driveway have bigger lot sizes.



Inference: most of the houses donot have recreation room and houses without a recreation room have bigger lot sizes.



Inference: there is an equal distribution of houses with and without a basement and houses without a basement have bigger lot sizes.

Regression Model:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11303.18	3746.753	-3.016793	0.0027
LOT_SIZE	4.812987	0.367472	13.09757	0.0000
_BEDROOM	2405.010	1169.481	2.056475	0.0402
_BATH	15863.18	1656.209	9.578010	0.0000
_STORIES	8321.279	1002.326	8.301972	0.0000
DRIVEWAY	9645.331	2253.924	4.279350	0.0000
REC_ROOM	5657.131	2119.685	2.668855	0.0078
BASEMENT	7939.389	1744.925	4.549989	0.0000

R-squared	0.586313	Mean dependent var	68121.60
Adjusted R-squared	0.580931	S.D. dependent var	26702.67
S.E. of regression	17286.13	Akaike info criterion	22.36774
Sum squared resid	1.61E+11	Schwarz criterion	22.43078
Log likelihood	-6098.393	Hannan-Quinn criter.	22.39238
F-statistic	108.9286	Durbin-Watson stat	1.485707
Prob(F-statistic)	0.000000		

Estimate Equation:

Sale Price = -11303.18 + (4.81)(Lot Size) + (2405.01)(Bedroom) + (15863.18)(Bath) + (8321.28)(Stories) + (9645.33)(Driveway) + (5657.13)(Rec Room) + (7939.39)(Basement)

Inferences:

- Since the model has an R^2 value of 0.58, it means that the model has moderate explanatory power.
- All the variables are statistically significant, having p-values under 0.05.

Equation when all the dummy variables have value = 1:

⇨ Sale Price = -11303.18 + (4.81)(Lot Size) + (2405.01)(Bedroom) + (15863.18)(Bath) + (8321.28)(Stories) + (9645.33)(1) + (5657.13)(1) + (7939.39)(1)

⇨ Sale Price = -11303.18 + (4.81)(Lot Size) + (2405.01)(Bedroom) + (15863.18)(Bath) + (8321.28)(Stories) + (9645.33)(1) + (5657.13)(1) + (7939.39)(1)

⇨ Sale Price = 11938.67 + (4.81)(Lot Size) + (2405.01)(Bedroom) + (15863.18)(Bath) + (8321.28)(Stories)