

KINGS_CONTY HOUSE PRICE PREDICTION

BY: JASWINDER_SINGH

DATA

The data for this project belongs to KAGGLE.

PROBLEM STATEMENT:

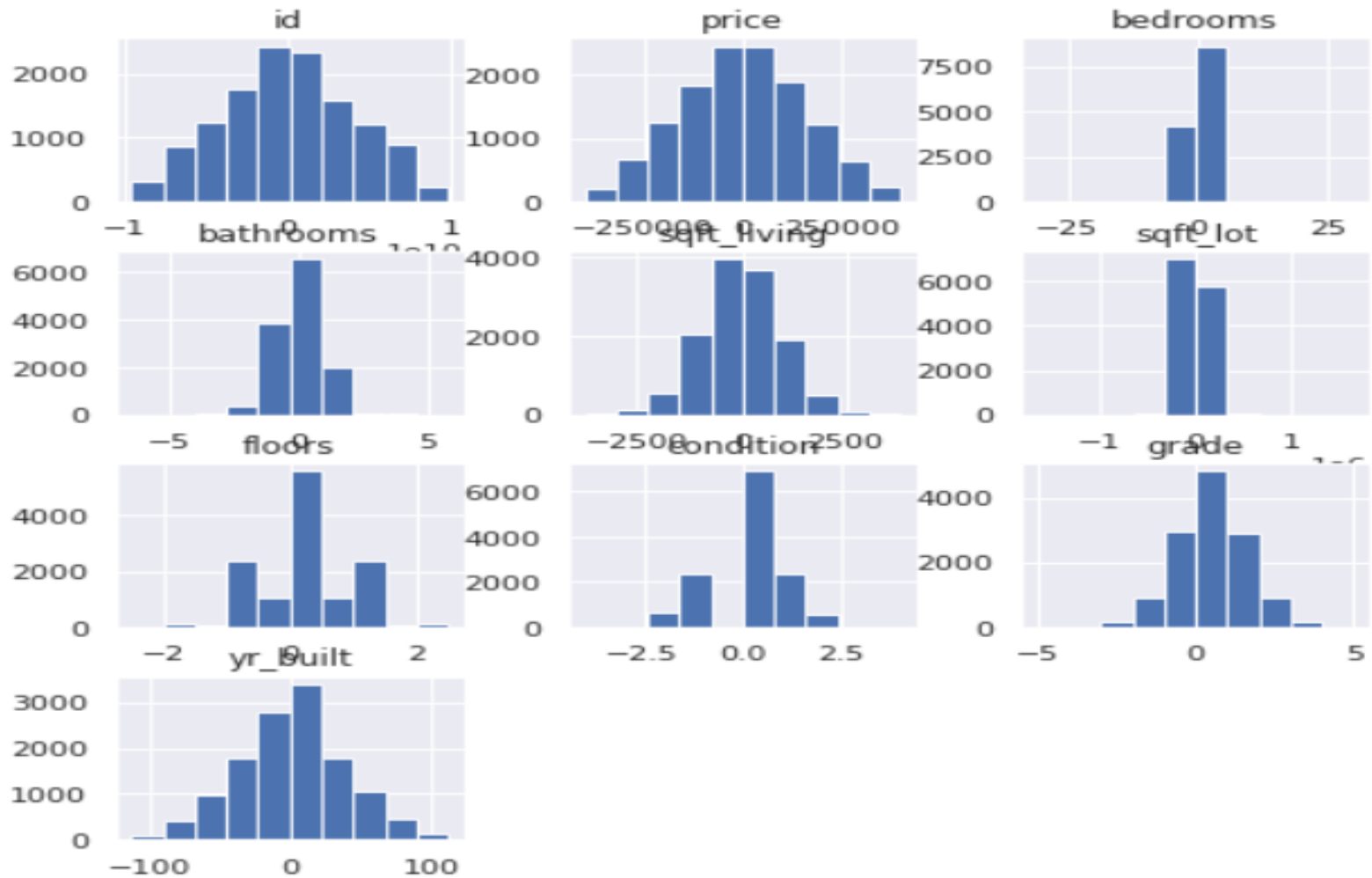
The stakeholder is expecting an exceptionally high increment in the prices of houses, When the attributes of the houses are subjected to change by single units.

```
data.corr()
```

	id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	yr_built
id	1.000000	0.041012	0.001205	0.038139	0.025295	-0.145108	0.036880	-0.046310	0.050729	0.025829
price	0.041012	1.000000	0.133458	0.210216	0.330074	0.060274	0.098756	0.059382	0.361308	-0.044638
bedrooms	0.001205	0.133458	1.000000	0.446571	0.572897	0.019763	0.073894	0.024514	0.205709	0.136699
bathrooms	0.038139	0.210216	0.446571	1.000000	0.644385	0.025689	0.464352	-0.159436	0.517827	0.582732
sqft_living	0.025295	0.330074	0.572897	0.644385	1.000000	0.132691	0.226658	-0.070331	0.552788	0.347224
sqft_lot	-0.145108	0.060274	0.019763	0.025689	0.132691	1.000000	-0.051319	0.027309	0.038840	0.022745
floors	0.036880	0.098756	0.073894	0.464352	0.226658	-0.051319	1.000000	-0.310970	0.419620	0.551684
condition	-0.046310	0.059382	0.024514	-0.159436	-0.070331	0.027309	-0.310970	1.000000	-0.219287	-0.382615
grade	0.050729	0.361308	0.205709	0.517827	0.552788	0.038840	0.419620	-0.219287	1.000000	0.537574
yr_built	0.025829	-0.044638	0.136699	0.582732	0.347224	0.022745	0.551684	-0.382615	0.537574	1.000000

- Certain columns are dropped before any use
- Outliers were present, removed before modeling

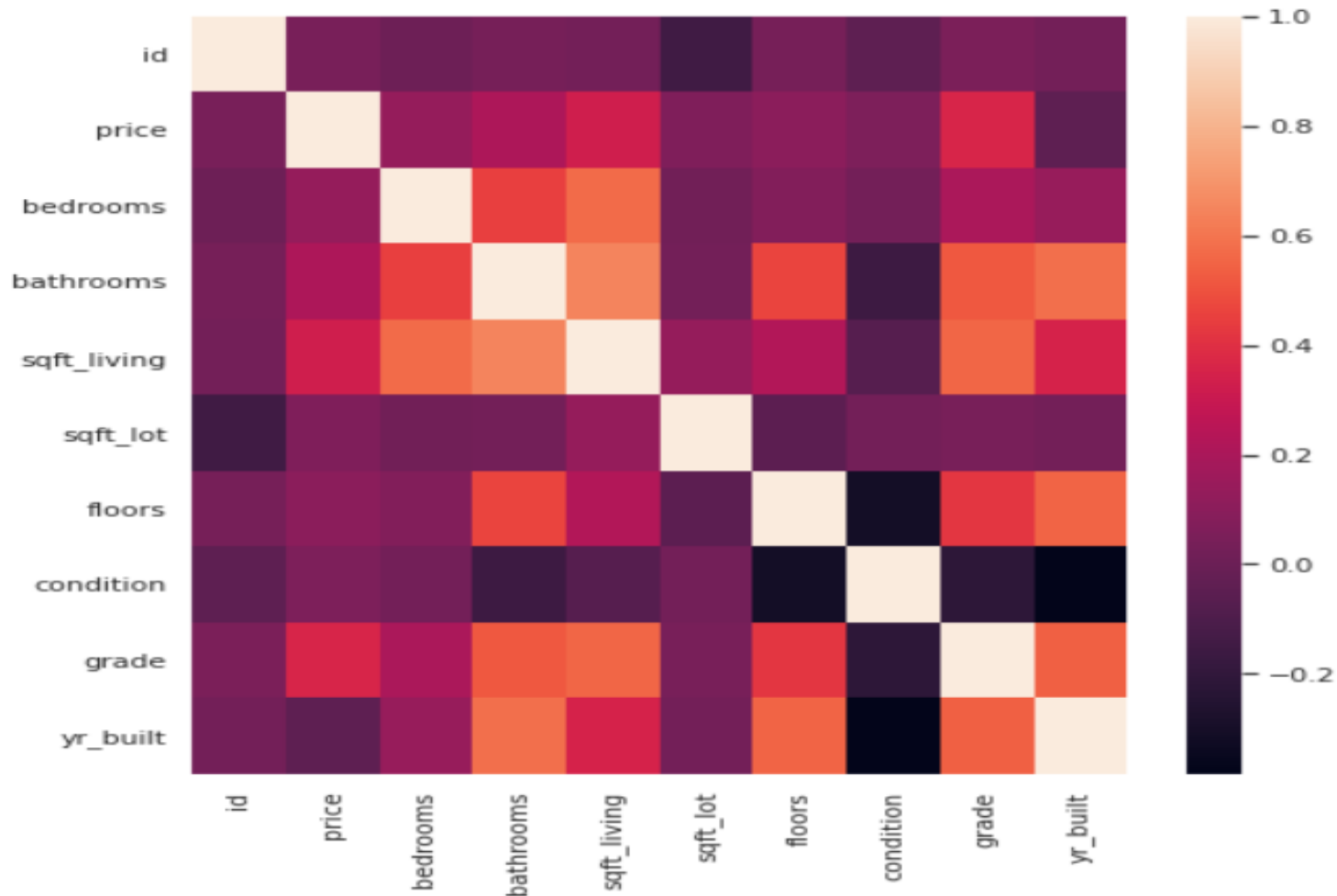
DISTRIBUTION OF THE DATA



HEATMAP

SHOWING THE CORRELATION VISUALLY

```
# Use the .heatmap method to depict the relationships visually!  
sns.heatmap(data.corr());
```



FIRST MODEL SUMMARY

OLS Regression Results

Dep. Variable:	price	R-squared (uncentered):	0.844
Model:	OLS	Adj. R-squared (uncentered):	0.844
Method:	Least Squares	F-statistic:	1.952e+04
Date:	Mon, 23 May 2022	Prob (F-statistic):	0.00
Time:	12:38:38	Log-Likelihood:	-2.9974e+05
No. Observations:	21597	AIC:	5.995e+05
Df Residuals:	21591	BIC:	5.995e+05
Df Model:	6		
Covariance Type:	nonrobust		

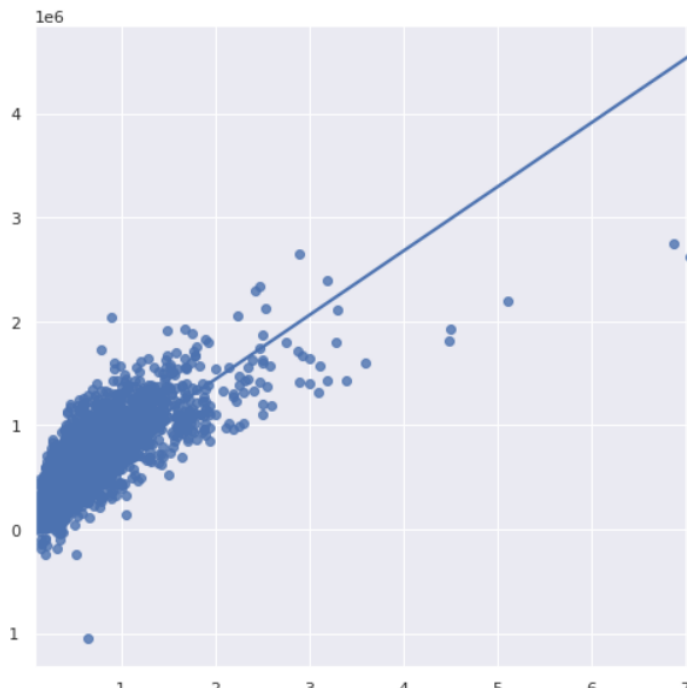
	coef	std err	t	P> t	[0.025	0.975]
bedrooms	-5.63e+04	2357.316	-23.882	0.000	-6.09e+04	-5.17e+04
bathrooms	6685.3449	3837.541	1.742	0.082	-836.518	1.42e+04
sqft_living	314.0643	3.152	99.628	0.000	307.885	320.243
sqft_lot	-0.3728	0.043	-8.606	0.000	-0.458	-0.288
floors	2234.4329	3817.106	0.585	0.558	-5247.376	9716.242
yr_built	32.4664	4.014	8.088	0.000	24.599	40.334

Omnibus:	14201.215	Durbin-Watson:	1.984
Prob(Omnibus):	0.000	Jarque-Bera (JB):	468535.936
Skew:	2.683	Prob(JB):	0.00
Kurtosis:	25.178	Cond. No.	1.13e+05

- DIFFERENCE IN PREDICTED AND ACTUAL VALUE IS 50K
- R² VALUE IS GOOD BUT NOT SIGNIFICANT
- COEFFICIENTS ARE NOT IN DESIRED CONDITON

SECOND MODEL(r2_value:0.60)

TRAIN/TEST SPLIT BY RATIO 3:7



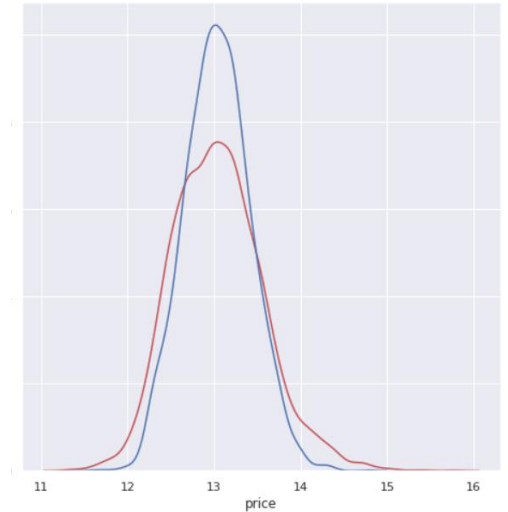
- THE REGRESSION IS BETTER AND THE POINTS ARE DISTRIBUTED EVENLY
- COEFFICIENTS ARE BETTER THAN THE PREVIOUS MODELS
- POSITIVE RELATION AMONG VARIABLES
- R2 VALUE IS GREATER
- MODEL SCORE IS DESIRABLE
- DIFFERENCE BETWEEN PREDICTED AND ACTUAL VALUE IS AROUND 30K

```
model1.coef_
```

```
array([-2.29025909e-06, -5.12971820e+04,  5.87523832e+04,  1.837  
       -3.27830585e-01,  2.06060871e+04,  2.16935350e+04,  1.339  
       -4.09190776e+03])
```

Strong predictors

- Bedrooms
 - Bathrooms
 - Floors
 - Area(dimensions)
 - Condition
 - Year of formation
- The negative coefficients suggesting:
bathrooms and yr_built are inversely proportional
- Sq_ft area accross all dimension is positively related to the price



Column	Change/unit	Change in price
Living area	1.83	1000
Floors	2.06	1000
Condition	By 1	1000

RECOMENDATIONS

•Through the visual of correlations, we can see that price is highly impacted by the area of house. Across all the dimensions.

Thus, the new houses should be bigger in area to attain maximum value.

- second strong predictor is grade of the house. The houses should undergo complete cleaning. And all the things within the house should be maintained properly.

- The condition of the house can not be compromised as it has a great impact on price. Thus, The old houses must be checked for repairs for any kind of damage.

- If the house is under construction, the architect should check the number of utility rooms as more number of bathrooms are not helping with the value of price.

- On the contrary, the bedrooms added more value to the house.

- Last, but not least, the number of floors should be checked for maximum availability.

THANK YOU