SIXTH-SENSE AGENCY HOUSE PRICE PREDICTION MODEL

BACKGROUND

- Home prices declining by approximately **10%** compared to the previous year due to interest rate increases and economic uncertainty.
- Despite the drop housing affordability remained a challenge for many potential buyers as well as Inventory shortages and a lack of new listings leading to increased competition among buyers.
- The number of available homes was considerably lower than the previous year, which impacted the overall sales activity in the market.

PROBLEM STATEMENT

Problem Statement:

- Which combinations of features provide the most accurate predictions for housing prices?
- By creating a multiple linear regression model, we can see how each coefficient works with pricing and better advise 6th sense agency on how to meet the changing needs of buyers and sellers while improving sales.

Objective: To develop a model that accurately predicts the optimal prices of houses based on a variety of price predictors.

DATA UNDERSTANDING

- Dataset: 21 variables, 21,597 records
- Key Predictors: Square foot of living, house grade, condition of the house, No. of floors, bedrooms and bathrooms, presence of a waterfront, year renovated, year the house was built etc.
- Features are a combination of numerical and categorical variables.
- Target Variable: Sales price

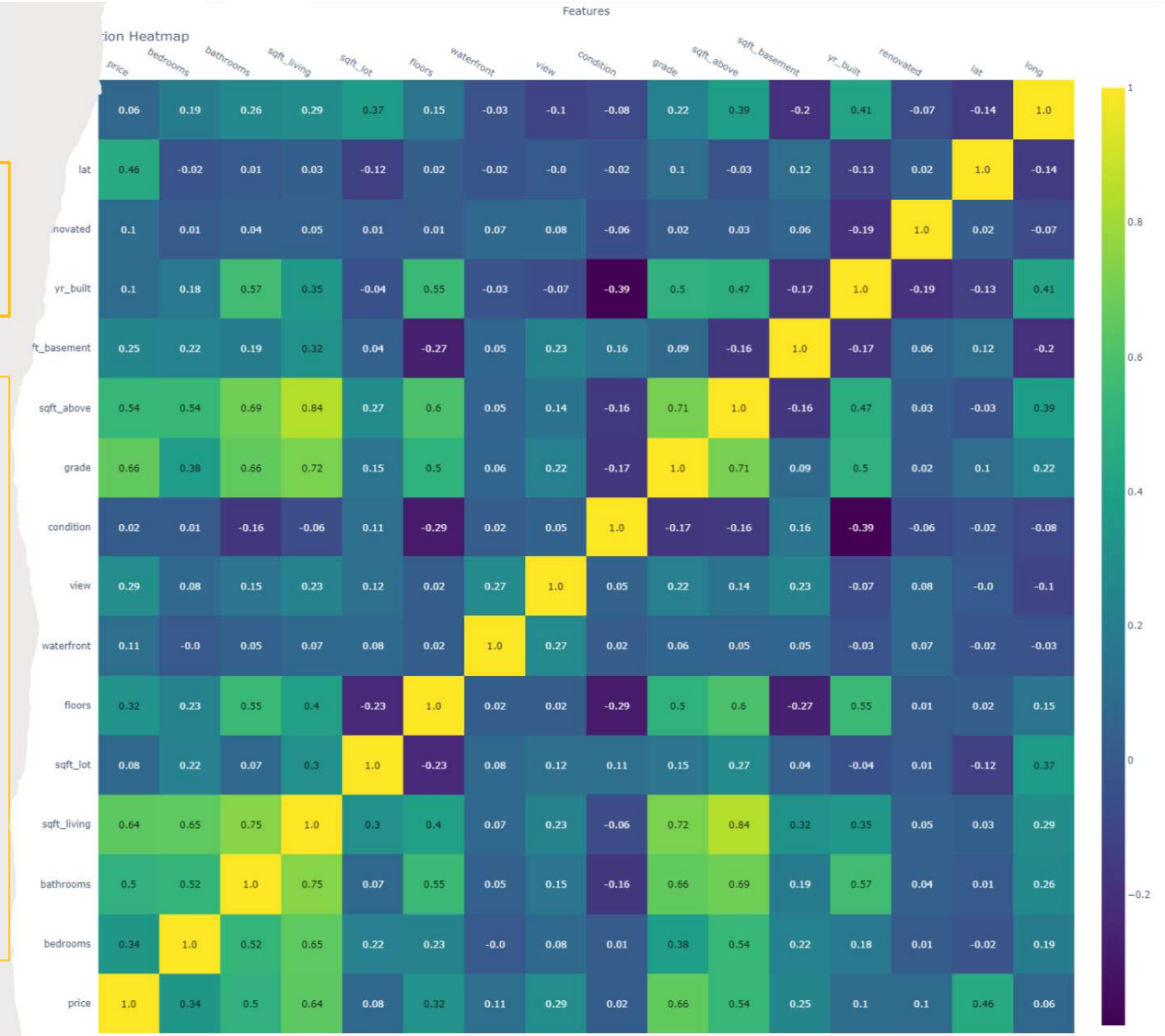
Methods

- Data Cleaning of null, outliers and placeholders
- Exploratory data Analysis – Univariate and Bivariate Analysis
- Feature Engineering: Introduce a new column ‘renovated’
- Perform a simple Linear Regression using high correlated predictor
- Perform a multiple Linear Regression using highly correlated predictors
- Standardization using logarithm transformation
- Perform the final Multiple linear regression using transformed values.
- Model Evaluation: Assess performance of our final multiple linear regression model

CORRELATION OF THE PREDICTOR VARIABLES

- sqft_living, bathrooms, grade and sqft_above have multicollinearity of 0.7 and above. This information is useful in deciding which variables to use in our model to avoid inaccuracies.

- Square fit leaving has the highest correlation with price



BASELINE MODEL

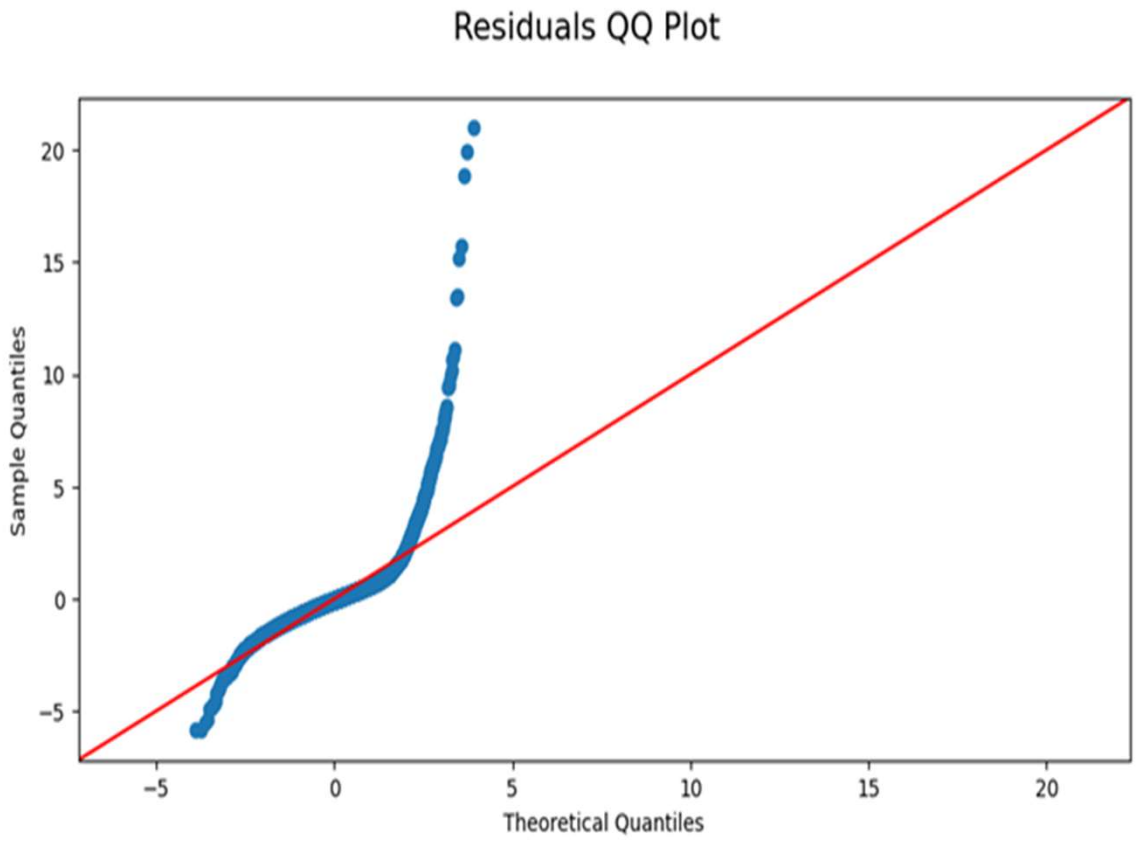
- Our baseline model is statistically significant shown by p-value less than alpha of 5%
- The model explains about 49% variance in price.
- For a unit increase in square foot of living, there is $280k increase in price.

MULTIPLE LINEAR REGRESSION MODEL

- The Model explains a variance of 67.7%
- The large condition indicates strong multicollinearity between the predictors.
- There is need to refine the model more by sorting out multicollinearity.

Baseline Model Residual Plot

- The Simple Linear Model captures only a small pattern in the data. It fails to capture most patterns and relationships in the data.
- Improvements needs to be done



IMPROVED MODEL

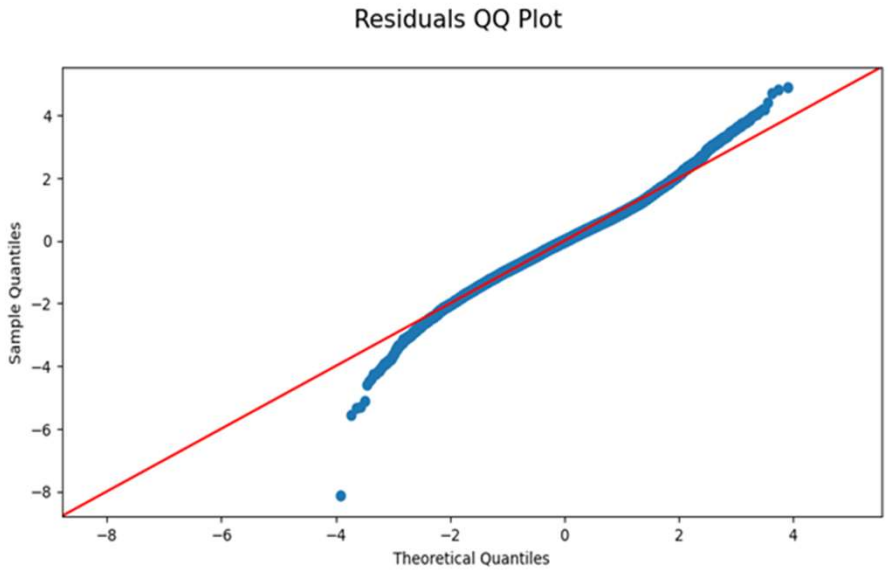
- Our multiple linear regression model was improved using log Transformation.
- From the results of the model, there is a significant change in our r-squared. The 76% explains the variance in price
- We can conclusively say that our model is statistically significant.

Dep. Variable:	price	R-squared:	0.759
Model:	OLS	Adj. R-squared:	0.759
Method:	Least Squares	F-statistic:	6168.
Date:	Fri, 02 Jun 2023	Prob (F-statistic):	0.00
Time:	13:12:45	Log-Likelihood:	-1442.1
No. Observations:	21597	AIC:	2908.
Df Residuals:	21585	BIC:	3004.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-47.1663	0.694	-67.940	0.000	-48.527	-45.806
bedrooms	-0.0141	0.003	-5.559	0.000	-0.019	-0.009
bathrooms	0.0672	0.004	16.270	0.000	0.059	0.075
sqft_living	0.0002	3.95e-06	45.498	0.000	0.000	0.000
sqft_lot	3.034e-07	4.37e-08	6.941	0.000	2.18e-07	3.89e-07
floors	0.0587	0.004	14.289	0.000	0.051	0.067
yr_built	-0.0032	8.62e-05	-37.546	0.000	-0.003	-0.003
renovated	0.0785	0.010	7.712	0.000	0.059	0.098
lat	1.3544	0.013	101.545	0.000	1.328	1.381
view	0.0816	0.002	33.027	0.000	0.077	0.086
condition	0.0660	0.003	22.128	0.000	0.060	0.072
grade	0.1795	0.003	69.033	0.000	0.174	0.185

Omnibus:	540.698	Durbin-Watson:	1.983
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1307.098
Skew:	0.023	Prob(JB):	1.47e-284
Kurtosis:	4.204	Cond. No.	1.74e+07

Residual plot for our Improved Model

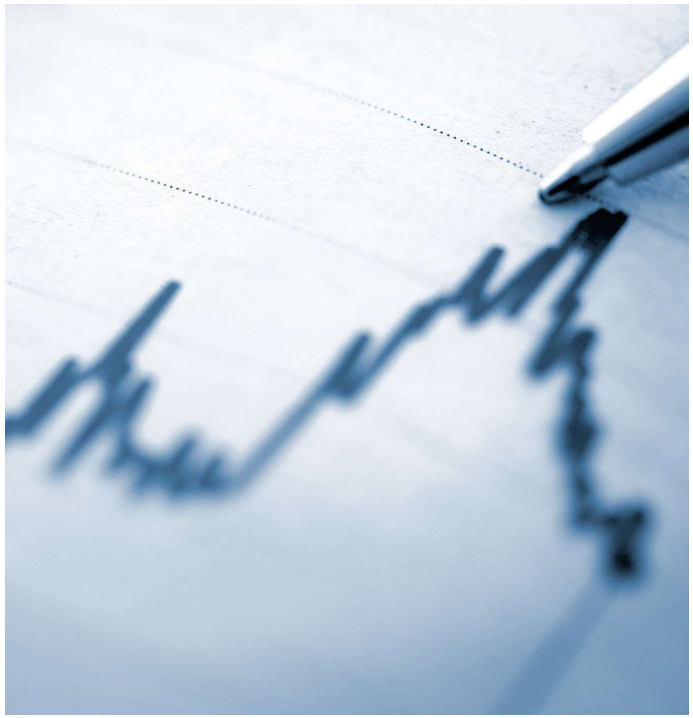


- From the plot we can now say that it follows the normality assumption.
- The spread/dispersion of residuals is more relatively consistent
- We have a close to perfect goodness of fit.

MODEL EVALUATION

- Our r-squared of about 76% explains the variance in price. This shows that a larger proportion of the variability in price is accounted for by the predictor variables.

- Our Mean Absolute Error, Mean squared Error and Root Mean Squared error of approximately 0.2, 0.1 and 0.3 is low meaning that our model performance is good.



Data Limitation

- Data is only from 2014 to 2015.
- Models to predict future sales price would need to be updated with newer data.
- Some data might be missing requiring us to make assumptions that might affect our model performance

CONCLUSIONS

- Square foot of living, grade, square foot above, bathrooms and view are the top 5 factors showing very high influence in the price of a house.
- The higher the house grade, the more price it fetches. Houses with average condition and above tend to fetch high prices. This could be because of several factors e.g. a house with average condition and above could have been renovated, recently build or have a higher square foot of living.
- From our analysis we can almost conclusively say that renovations have increased the quality of the house thereby increasing in price.

RECOMMENDATIONS

- Encourage renovations to improve the overall condition and raise the property's grade as this has a great impact on the value of a house.
- Highlight the significant impact of square footage of living space on house prices and use this information to justify higher listing prices for properties with more extensive square footage.
- The number of bathrooms and bedrooms also have a positive correlation with the value of a house. Therefore, during renovation adding a bedroom would increase the value of the house.

NEXT STEPS

- While the model provides insights into specific variables, it is important to consider broader market trends and factors influencing real estate prices. Keep track of market conditions, economic indicators, and buyer preferences to provide clients with up-to-date and accurate data

- Continuously Refine and Validate the Model: Understand the limitations and assumptions of the model and its applicability to specific markets, and update while refining the model based on new data and local market knowledge to improve its accuracy and relevance.

END

THANKYOU