HOME RENOVATIONS AND THEIR BENEFICIAL EFFECT ON HOME ESTIMATED VALUE

OVERVIEW

This project is focused on home renovations and how it increases the estimated value of the homes.

The first step getting the business understanding, data understanding and the objective of the project.

We begin by importing the data from the provided databases(csv file and .md file), opening and reading data.We then prepare the data by cleaning it. This is by identifying any missing value, dropping the missing values and then checking for duplicates. Moreover, we formulate the hypothesis, null and alternative hypothesis.

We build models, get the results from the models, represent them using different visualizations such as scatterplots with fitted regression lines.

Finally, we deduced conclusions from the model findings.

BUSINESS UNDERSTANDING

Many homeowners face obstacles when it comes to home renovations resulting in outcomes such as delays, cost overrun among others.

They may need a slight push for them to believe that home renovations can help them get the maximum estimated value of their homes.

This is where real estate agency that help homeowners sell or/and buy houses comes in. The agency provides guidance on how renovating their homes can help them reach their objective of optimum home estimated value.

DATA UNDERSTANDING

We need to explore how increasing measures/sizes of various house features such as floors will increase the overall price of the house/s.

This is by analysing the dataset provided by formulating the hypothesis of the project and building various regression models to aid in analysing the relationship between the dependable variable and independent variables.

MAIN OBJECTIVE

The main objective of the project is to determine how renovations of various features such as square footing of living area leads to an increase in home estimated values.

MODELING

Dep. Variable:	price	**R-squared:**	0.499
Model:	OLS	**Adj. R-squared:**	0.499
Method:	Least Squares	**F-statistic:**	1.568e+04
Date:	Sun, 09 Jul 2023	**Prob (F-statistic):**	0.00
Time:	13:01:44	**Log-Likelihood:**	-2.1911e+05
No. Observations:	15762	**AIC:**	4.382e+05
Df Residuals:	15760	**BIC:**	4.382e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-5.516e+04	5205.744	-10.597	0.000	-6.54e+04	-4.5e+04
sqft_living	286.1494	2.285	125.214	0.000	281.670	290.629

Omnibus:	11136.965	**Durbin-Watson:**	1.970
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	457340.013
Skew:	2.910	**Prob(JB):**	0.00
Kurtosis:	28.739	**Cond. No.**	5.65e+03

Baseline modeling with target variable, Price, and feature variable, Sqft_livings, with the highest correlation.Together with visualizations.

By getting to know the percentage of variance of the model being 50% and that a unit increase in square foot of the living space leads to a corresponding increase of price by 286.15 as shown by the regression results.

.

Dep. Variable:	price	R-squared:	0.499
Model:	OLS	Adj. R-squared:	0.499
Method:	Least Squares	F-statistic:	7845.
Date:	Sun, 09 Jul 2023	Prob (F-statistic):	0.00
Time:	13:27:41	Log-Likelihood:	-2.1910e+05
No. Observations:	15762	AIC:	4.382e+05
Df Residuals:	15759	BIC:	4.382e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-5.327e+04	5260.979	-10.125	0.000	-6.36e+04	-4.3e+04
sqft_living	296.4372	4.741	62.532	0.000	287.145	305.729
sqft_above	-13.0205	5.257	-2.477	0.013	-23.324	-2.717

Omnibus:	11124.255	Durbin-Watson:	1.970
Prob(Omnibus):	0.000	Jarque-Bera (JB):	456859.299
Skew:	2.905	Prob(JB):	0.00
Kurtosis:	28.727	Cond. No.	7.52e+03

Second modeling with target variable, Price, with features Sqft_living and Sqft_above, that had the second highest correlation. Together with visualizations.

By getting the same percentage of variance as the baseline model of 50% and that a unit increase in square foot of the living space leads to a corresponding increase of price by 296.44 while an increase of one square foot of house apart from basement corresponds to a price decrease of 13.02 as shown by the regression results.

Dep. Variable:	price	R-squared:	0.523
Model:	OLS	Adj. R-squared:	0.523
Method:	Least Squares	F-statistic:	2467.
Date:	Sun, 09 Jul 2023	Prob (F-statistic):	0.00
Time:	13:52:58	Log-Likelihood:	-2.1871e+05
No. Observations:	15762	AIC:	4.374e+05
Df Residuals:	15754	BIC:	4.375e+05
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.731e+04	9055.678	7.433	0.000	4.96e+04	8.51e+04
bedrooms	-6.1e+04	2728.056	-22.359	0.000	-6.63e+04	-5.56e+04
bathrooms	2525.9420	4500.192	0.561	0.575	-6294.951	1.13e+04
sqft_living	334.9997	5.802	57.735	0.000	323.626	346.373
sqft_lot	-0.3679	0.050	-7.317	0.000	-0.467	-0.269
floors	6672.9219	5070.678	1.316	0.188	-3266.188	1.66e+04
yr_renovated	79.2145	5.172	15.316	0.000	69.077	89.352
sqft_above	-19.2082	5.924	-3.242	0.001	-30.820	-7.596

Omnibus:	10689.805	Durbin-Watson:	1.975
Prob(Omnibus):	0.000	Jarque-Bera (JB):	392884.313
Skew:	2.770	Prob(JB):	0.00
Kurtosis:	26.822	Cond. No.	2.08e+05

Third modeling with target variable, Price and all the feature variables with numeric datatype.

The percentage of variance increased to 52.3% and that a unit increase in square foot of the living space leads to a corresponding increase of price by 335,an increase in number of bathrooms and floors leads to price increase of 2525.94 and 6672.92 respectively, as shown by the regression results.

REGRESSION RESULTS

After analysing using both the coefficients and graphs of the models, the more befitting results were from the coefficients due to better understanding and easier interpretation. The following are the results:

From baseline model:

1. The model explains about 50% variance in price.

2. The model is statistically significant overall with an F-statistic p-value below 0.05

2. The model's coefficients(const and sqft_living) are both statistically significant with a t-statistic p-values below 0.05

3. For an increase in sqft_living there is an associate increase in price by 286.15.

4. When square foot of living space sqft_living = 0, the price would be be -55,160.00.

From second model:

1. The model is statistically significant overall, with an F-statistic p-value well below 0.05

1. The model explains about 50% variance in price.

2. The model coefficients(const, sqft_living and sqt_above) are statistically significant with a t-statistic p-values below 0.05

3. For an increase in sqft_living there is an associate increase in price by 296.44.\

We see that there is an increase from the previous model therefore showing that sqft_above has a meaningful confounding in the relationship between both sqft_living and price.

4. For an increase of 1 sqft_above, there is an associate decrease in price by -13.02.

From third model:

1. The model is statistically significant with a F-statistic p-value below 0.05

2. The model explains a 52.3% variance in price

   The R-squared increased by 2.3%

3. Most of the model coefficients are statistically significant:

   const, bedrooms, sqft_living, sqft_lot, yr_renovated and sqft_above have p-values below 0.05.

   bathrooms and floors have a coefficient greater than 0.05 meaning they have no effect on price thus statistically insignificant.

4. For an increase in sqft_living there is an associate increase in price by 335.\

   We see that there is an increase from the previous model therefore showing that other features have a meaningful confounding in the relationship between both sqft_living,sqft_above and price.

RECOMMENDATIONS

From the results of the three models, using their coefficients rather than the graphs, there is an increase in the variation(R-Squared) from 50% to 52.3%.Though it is a small difference we can see that features such as bathrooms and floors, though with higher p-value of more than 0.05 have contributed to the increase in variance in price. We can conclude that upgrade/increase of the features, such as , increasing number of bedrooms, sqft_lot, sqft_living and bathrooms increases the predicted target of price resulting in enhancement of livability, convenience and becomes more appealing to buyers.

Moreover,, we reject our null hypothesis that home renovations does not increase the estimated value of houses and accept the alternative that it actually does increase the estimated value of houses.

From the above, the stakeholders can then inform the homeowners that renovating their homes does increase the price of their homes.

NEXT STEP

The stakeholder should advice the homeowners on increasing the square footing of living area, renovating the bathrooms and the bedrooms as it leads to an increase in price.

THANKYOU

CONTACT INFO

NAME: LYNN WANJIKU NDERO

EMAIL: LNNWANJIKU@GMAIL.COM

LINKEDIN: Lynn Ndero