bnittalee / KC-Housing-Price-Project Public

0 stars 0 forks

Star Unwatch

Code Issues Pull requests Actions Projects Wiki Security Insights Se

main

bnittalee Merge branch 'main' of https://github.com/bnittalee/KC-Housing-Price-... ... now 40

View code

README.md

King County Housing Prices: A Multiple Regression Analysis

Author: Brittney Nitta-Lee

Overview

Compass Real Estate wants to host a class for homeowners who are interested in selling their home. This analysis will be focusing on the King County House Sales dataset to find information on ways to increase the value of their home.

Questions

1. Which features are most highly correlated with price?
2. Which feature has the strongest correlation with the value of a home?
3. Does grade or condition of a house contribute to the value of a home??

Data

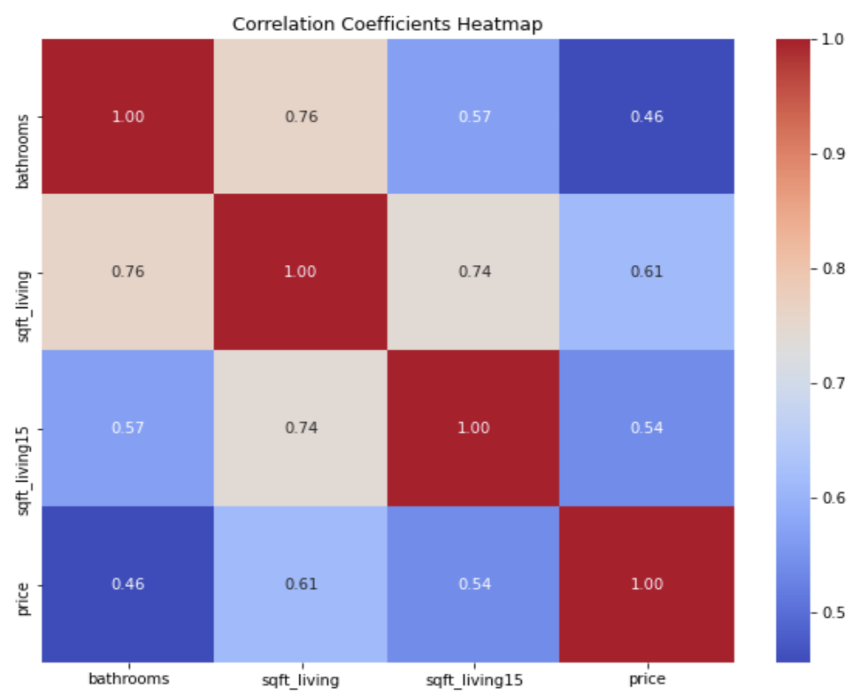
The King County Housing Data Set includes information about the size, location, condition, and other features of houses in Washington's King County. The dataset and variable descriptions can be found on Kaggle.

Methods

Methods included preprocessing the data, performing a train-test split to create a linear reagression models with price as the target variable.

Results

The heatmap and the scatterplot matrix of the correlation coefficients between the the value of a home and features (bathrooms, sqft_living and sqft_living15, helps visualize the correlation with price. According to the correlation coefficients, sqft_living has the strongest linear relationship with price while the number of bathrooms is the lowest.



The OLS Regression Result shows a R-squared value of 0.507. This was before the log transformation of my independent variables. Condition_3 and Condition_4 shows a negative correlation to price while bathrooms, sqft_living and sqft_living 15 have positive correlations. The RMSE is 0.38742367880763834 which means that, on average, the predicted values of your multiple linear regression model are off by approximately 0.387 units from the actual values. That's not bad at all.

Out[30]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.507
Model:	OLS	**Adj. R-squared:**	0.506
Method:	Least Squares	**F-statistic:**	4434.
Date:	Mon, 06 Mar 2023	**Prob (F-statistic):**	0.00
Time:	07:46:19	**Log-Likelihood:**	-2.9976e+05
No. Observations:	21597	**AIC:**	5.995e+05
Df Residuals:	21591	**BIC:**	5.996e+05
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.692e+04	8019.487	-3.357	0.001	-4.26e+04	-1.12e+04
condition_3	-1.043e+05	6357.139	-16.401	0.000	-1.17e+05	-9.18e+04
condition_4	-6.604e+04	6851.158	-9.640	0.000	-7.95e+04	-5.26e+04
bathrooms	3209.5992	3543.538	0.906	0.365	-3735.998	1.02e+04
sqft_living15	73.1732	3.929	18.622	0.000	65.471	80.875
sqft_living	240.3983	3.684	65.259	0.000	233.178	247.619

Omnibus:	15670.814	**Durbin-Watson:**	1.984
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	708449.398
Skew:	3.003	**Prob(JB):**	0.00
Kurtosis:	30.408	**Cond. No.**	1.91e+04

The OLS Regression Result for building grade and home condition has somewhat similar results. The R-squared shows th 50% of the variability in the price of the house can be explained by sqft_living, condition_5 and building grade 10 through 13. The RMSE is 0.371 which is similar to our previous model. The RMSE is also small which tells me the predicted home price is closer to the actual price of a home.

Out[49]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.503
Model:	OLS	**Adj. R-squared:**	0.503
Method:	Least Squares	**F-statistic:**	3647.
Date:	Mon, 06 Mar 2023	**Prob (F-statistic):**	0.00
Time:	07:46:26	**Log-Likelihood:**	-9234.0
No. Observations:	21597	**AIC:**	1.848e+04
Df Residuals:	21590	**BIC:**	1.854e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.6796	0.050	152.139	0.000	7.581	7.779
sqft_living	0.7049	0.007	104.820	0.000	0.692	0.718
grade_10	0.3758	0.012	30.948	0.000	0.352	0.400
grade_11	0.5516	0.020	28.056	0.000	0.513	0.590
grade_12	0.7839	0.040	19.551	0.000	0.705	0.862
grade_13	1.0857	0.103	10.501	0.000	0.883	1.288
condition_5	0.1556	0.009	16.580	0.000	0.137	0.174

Omnibus:	100.507	**Durbin-Watson:**	1.983
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	84.363
Skew:	0.092	**Prob(JB):**	4.79e-19
Kurtosis:	2.755	**Cond. No.**	313.

Recommendations

Here are a few recommendations for homeowners who want to sell their home.

1. Increase the square footage of your home
2. Build additional bathrooms
3. The building grade should be at least grade 10. Which means, homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage.
4. The condition of the home must be a 5, according to King County standards. All items well maintained, many having been overhauled and repaired as they have shown signs of wear, increasing the life expectancy and lowering the effective age with little deterioration or obsolescence evident with a high degree of utility.

Limitations and Next Steps

There was a lot of preprocessing and variables we had to perform log transformations on variables to satisfy regression assumptions. Therefore, the model may not accurately predict a home's value. A future analysis could include looking at data in other counties and using an updated dataset. King County is also a huge county and it would be better to narrodown differen't areas within King County to get better results.

For more information

For additional information, contact Brittney Nitta-Lee at bnittalee@gmail.com

Repository Structure

```
├── .ipynb_checkpoints/
├── Images
├── PDFS
├── .DS_Store
├── .gitattributes
├── Project_notebook.ipynb
├── column_names.md
└── kc_house_data.csv
```

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%