King County Housing Price Analysis

Lucas Wilkerson

Summary

Data from King County housing sales was analyzed to determine what housing characteristics correlate to higher housing prices.

Regression Modeling was utilized to come up with a predictive model to determine housing prices.

Insight from the analysis will be used to generate actionable recommendations for the stakeholder.

Outline

- Business Problem
- Data/Methods
- Results
- Conclusions

Business Problem

A King County real estate company wants to increase client acquisition and retention through:

- Identifying key housing price characteristics
- Giving sound recommendations
- Improving client's home sale price

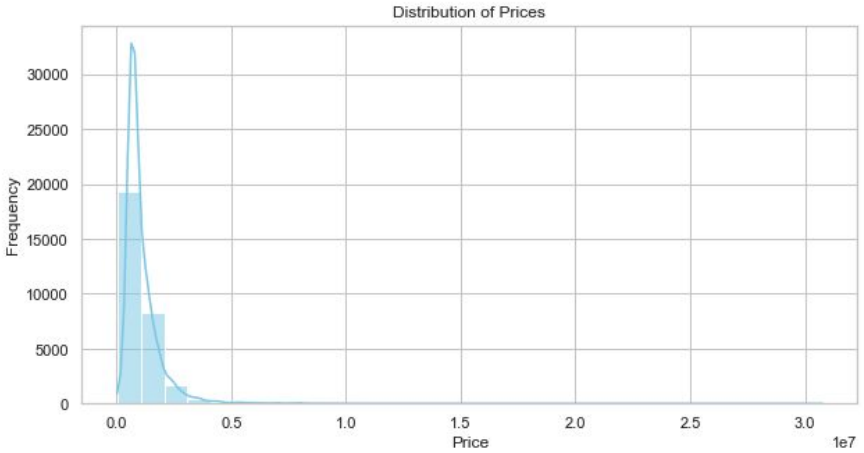
Data/Methods

EDA:

- King County House Sales dataset
- Final dataset:
  - 30,062 entries
  - 21 columns (features)
  - Sqft_living: **correlation = 0.61**

Regression Modeling:

- 6 model iterations were ran
- Target Variable: Price



Results: Baseline Model

R-Squared:

- **0.375** : model explains 37.5% variance in price

Sqft_living:

- 1 sqft increase = **~ 562.53 increase in price (USD)**

High Error Metrics**:**

- MAE of $ 395,915.33
- MSE of $ 706,874.49

OLS Regression Results

Dep. Variable:	price	R-squared:	0.375
Model:	OLS	Adj. R-squared:	0.375
Method:	Least Squares	F-statistic:	1.800e+04
Date:	Thu, 10 Aug 2023	Prob (F-statistic):	0.00
Time:	20:46:46	Log-Likelihood:	-4.4755e+05
No. Observations:	30062	AIC:	8.951e+05
Df Residuals:	30060	BIC:	8.951e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-8.076e+04	9758.256	-8.276	0.000	-9.99e+04	-6.16e+04
sqft_living	562.5261	4.193	134.171	0.000	554.308	570.744

Omnibus:	43093.441	Durbin-Watson:	1.860
Prob(Omnibus):	0.000	Jarque-Bera (JB):	47238386.360
Skew:	8.103	Prob(JB):	0.00
Kurtosis:	196.520	Cond. No.	5.57e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.57e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Results

Final Model:

- R-Squared: **0.503**
- **Grade and Waterfront** = highest in price increase per unit
- **Square foot of living space**:
  - Each 1 square foot increase, there is a **0.02% increase** in average price
- **Condition of home**:
  - Improve by 1 rating = **about 4.92% increase** in average price

OLS Regression Results

Dep. Variable:	price	R-squared:	0.503
Model:	OLS	Adj. R-squared:	0.503
Method:	Least Squares	F-statistic:	1791.
Date:	Thu, 10 Aug 2023	Prob (F-statistic):	0.00
Time:	20:54:40	Log-Likelihood:	-15850.
No. Observations:	30062	AIC:	3.174e+04
Df Residuals:	30044	BIC:	3.189e+04
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	11.5676	0.030	379.636	0.000	11.508	11.627
bedrooms	-0.0113	0.003	-3.453	0.001	-0.018	-0.005
sqft_living	0.0002	4.52e-06	45.393	0.000	0.000	0.000
sqft_lot	3.442e-07	4.31e-08	7.981	0.000	2.6e-07	4.29e-07
floors	0.0369	0.006	6.659	0.000	0.026	0.048
grade	0.2164	0.003	62.540	0.000	0.210	0.223
Basement	0.0463	0.005	8.942	0.000	0.036	0.056
Garage	-0.0167	0.006	-2.755	0.006	-0.029	-0.005
Patio	0.0196	0.006	3.254	0.001	0.008	0.031
Waterfront	0.2819	0.021	13.733	0.000	0.242	0.322
Nuisance	0.0211	0.006	3.311	0.001	0.009	0.034
view_encoded	0.0365	0.003	11.158	0.000	0.030	0.043
condition_encoded	0.0492	0.004	13.342	0.000	0.042	0.056
heat_source_encoded	-0.0104	0.003	-3.451	0.001	-0.016	-0.004
sewer_system_encoded	-0.1144	0.007	-15.263	0.000	-0.129	-0.100
Month	-0.0149	0.001	-19.526	0.000	-0.016	-0.013
Age	0.0029	0.000	25.512	0.000	0.003	0.003
renovated	0.0564	0.012	4.664	0.000	0.033	0.080

Omnibus:	8730.615	Durbin-Watson:	1.963
Prob(Omnibus):	0.000	Jarque-Bera (JB):	110100.416
Skew:	-1.039	Prob(JB):	0.00
Kurtosis:	12.142	Cond. No.	8.04e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.04e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Conclusions/Recommendations

Overall Condition: Optimize the condition of their home.

Square Feet of Living Space: Increase the square footage of living space.

- **1 sqft = ~0.02 % increase in price**
- Scaled out: **1000 sqft = ~ 20% increase in price**

Grade: Hire a high quality contractor and invest in high quality materials when building on to the home or making structure improvements/repairs.

- 1 increase in grade level = **~ 21.6% increase in price**

Limitations/ Future Considerations

- The final R-Squared value is 0.503 which suggests that approximately only 50.3% of the variance. Ideally for confidence in the model we want this higher.

- There were columns eliminated from the dataset which could have impact.

- There are other factors of influence that could be explored in further detail such as location and time of year sold.

Thank You!

Email: ldwilker10@gmail.com
GitHub: @ldwilker10
LinkedIn: https://www.linkedin.com/in/lucasdukewilkerson/