**Analysis 1 - Price and Square Foot Relationship Report**

**Purpose**

Century 21 Ames asked for clear evidence of how a home’s living area translates into sale price in the three neighborhoods where the company operates. Using the Ames Housing training data, all records outside NAmes, Edwards, and BrkSide were removed. The living area measurement was rescaled to represent hundreds of square feet so that an increase of one unit corresponds to 100 sq ft (GrLivArea100).

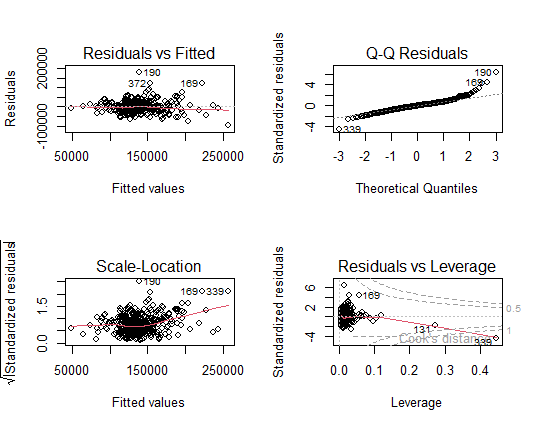
**Model**

BrkSide is the reference neighborhood.

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Estimate | 95% CI | p-value |
| Intercept () | $19,971 | [-4,314:44,257] | 0.107 |
| GrLivArea100 () | $8,716 | [6,793:10,640] | <0.001 |
| Edwards () | $68,382 | [40,914:95,850] | <0.001 |
| Names () | $54,705 | [27,408:82,001] | <0.001 |
| GrLivArea100 x Edwards () | -$5,741 | [-7,849:-3,634] | <0.001 |
| GrLivArea100 x Names () | -$3,285 | [-5,411:-1,158] | 0.0026 |

The baseline price for a home when size is 0 (not practical, simply needed for calculations), is estimated to be $19,971. The estimated added value of each 100 sqft in BrkSide is $8,716. Independent of size, the premium to live in Edwards opposed to BrkSide is estimated to be $68,382. Independent of size, the premium to live in NAmes opposed to BrkSide is estimated to be $54,705. In Edwards, every additional 100 sq ft raises the sale price by about $5,741 less than the same space would add to a BrkSide home. In NAmes, each extra 100 sq ft increases the price by roughly $3,285 less than it does in BrkSide.

**Diagnostics**

****

Residuals versus fitted values show a cloud of points centered around zero with roughly constant spread, suggesting reasonable linearity and homoscedasticity.

The Q–Q plot follows the line closely in the middle and only departs slightly in the tails, indicating acceptable normality of residuals. Scale-location results support the finding of stable variance across predicted prices. Taken together, the assumptions appear satisfied for practical decision-making.

**Conclusion**

* The relationship between finished living area and sale price is strongly positive in all three neighborhoods but the magnitude of that relationship varies.
* BrkSide shows the steepest pay-off for adding square footage, Edwards the flattest, NAmes moderate.
* Location premiums are sizeable: about $68,382 for Edwards and $54,705 for NAmes relative to BrkSide before accounting for size.
* Model diagnostics confirm that linear regression provides an adequate summary for this data.

**Analysis 2 – Prediction Model**

**Objective and Data Preparation**

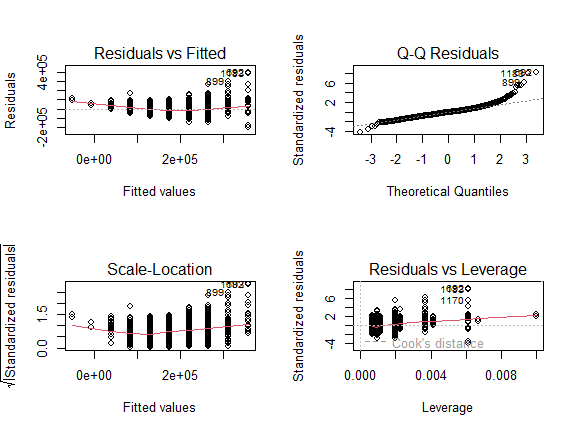
The goal is to build the most accurate linear-regression model for predicting the sale price of every home in the Ames Housing data set. The full Ames Housing training set contains 1,460 sales. All character columns were converted to factors. Columns that begin with a number, were renamed so that every variable can be processed by olsrr.

|  |  |
| --- | --- |
| **Model Name** |  |
| Simple | SalePrice = -96,206 + 45,435.8 \* OverallQual |
| Fixed | SalePrice = 3,162.993 + 89.091 \* GrLivArea + 27,311.09 \* FullBath |
| Stepwise (MLR Model) | 49 Predictors selected with “ols\_step\_both\_p” at α = 0.05 |

**Model diagnostics and interpretation**

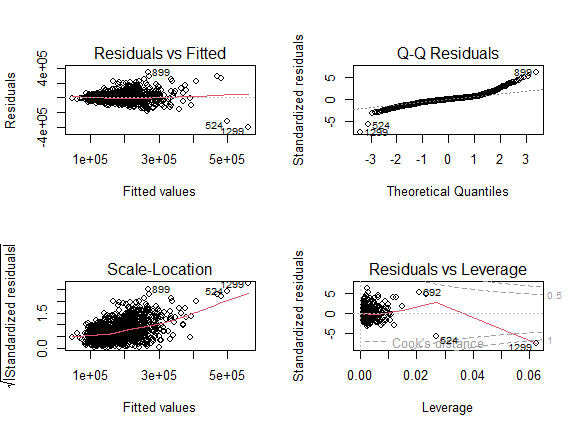
*Simple model*

Adjusted R² equals 0.626. Each one-point rise in OverallQual adds roughly $45,400 to the sale price. Residuals widen slightly for prices above $300,000, but the overall pattern is a random scatter.



*Fixed two-term model*

Adjusted R² equals 0.524, the weakest of the three. Finished living area contributes about $89 per square foot. Each additional full bath adds about $27,300. Residual standard error is $54,900.

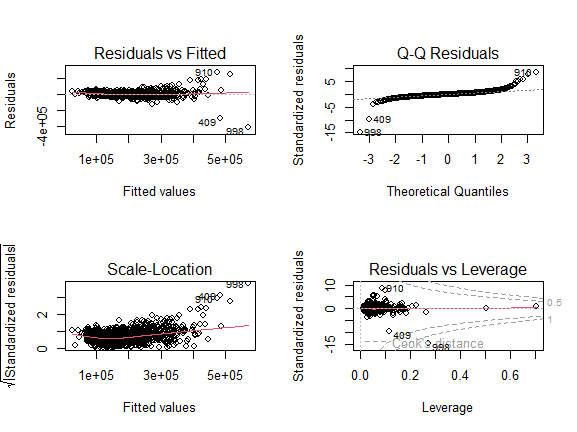


*Stepwise-selected model*

Adjusted R² equals 0.845 and the residual standard error is $32,700. Key numeric drivers are

* OverallQual: +$11,700 per point
* GrLivArea: +$45 per finished square foot
* GarageCars: +$14,100 per indoor parking bay
* YearBuilt: +$270 per calendar year

Exterior quality has a large penalty when lower than Excellent. A good rating reduces price by about $42,100, Typical by $43,100, Fair by $48,100. Five neighborhoods show significant positive premiums once physical features are included, led by Stone Brook and North Ridge.



The PRESS was obtained with ols\_press(). For the stepwise fit the two perfect-leverage rows had to be omitted before calculation to avoid an “inf” result.

|  |  |  |  |
| --- | --- | --- | --- |
| Predictive Models | Adjusted | CV PRESS | Kaggle Score |
| Simple Model | 0.6254 | 3.464089e+12 | 1.00432 |
| Fixed Model | 0.5231 | 4.441743e+12 | 0.28586 |
| MLR Model | 0.8451 | 1.436939e+12 | 0.18099 |

**Recommendation**

The stepwise model predicts sale price with the lowest error and the highest adjusted R². It improves RMSE by more than nine per cent compared with the simple specification and by nearly twenty-three per cent compared with the fixed two-term equation. Although the equation contains forty-nine slopes, every term has a clear real-estate meaning.

Therefore, the stepwise model is recommended for future prediction of Ames home prices. The simple model may be preferred when a single-variable explanation is needed, but it carries an expected prediction error about $5,000 larger than the stepwise selected model. The fixed model is not advised because it is both less accurate and less interpretable than the other two options.

**Key takeaways**

* Workmanship quality remains the strongest single predictor, yet floor area, garage capacity, basement finish and exterior grade add independent value.
* Location effects mostly disappear once house characteristics are controlled, except for a small set of premium neighborhoods.
* The recommended model delivers an expected absolute error near $30,000.

**Appendix**

[**RShiny App**](https://devin476.shinyapps.io/HousingExplorer/)

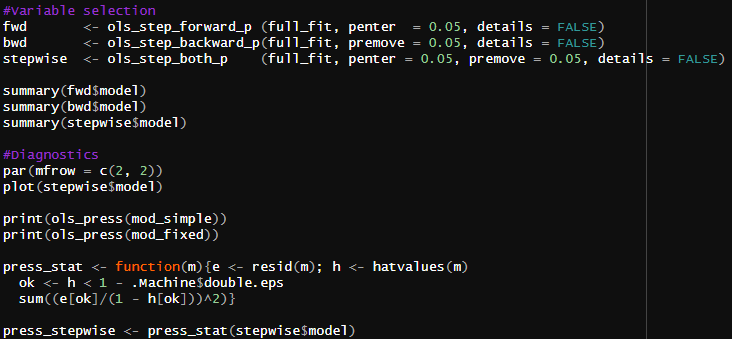
[**GitHub Repo**](https://github.com/devin476/Ames-Housing-Model)

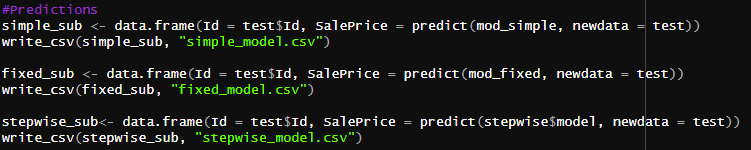
**A screen shot of a computer

AI-generated content may be incorrect.**

**A computer screen with white and green text

AI-generated content may be incorrect.**

****

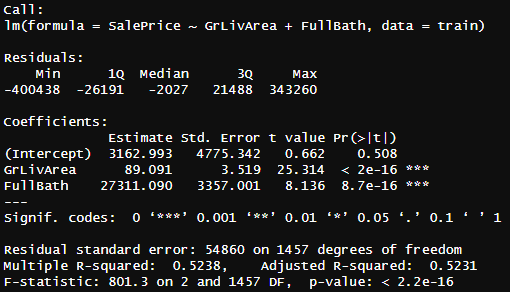
****

Simple-Model Summary

**A screenshot of a computer

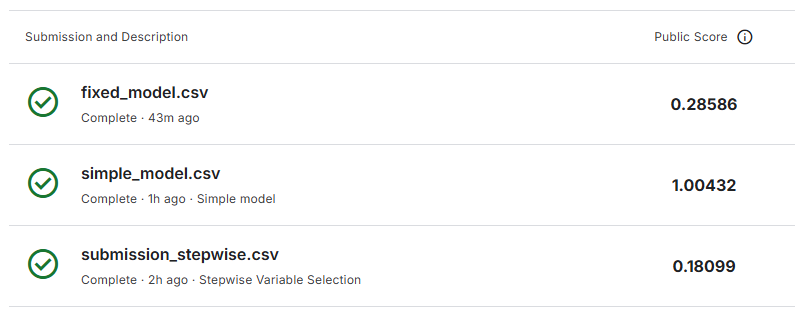
AI-generated content may be incorrect.**

Fixed-MLR Model Summary

****

**A screen shot of a computer

AI-generated content may be incorrect.**Stepwise Variable-Selected Model Sum

****