**Aaron King** -- .code file that corresponds to this: **austin\_linear\_regression.ipynb**

**SCENARIO ONE: PREDICTORS OF PRICE**  
**1. What analytical technique(s) did you choose to address this scenario?**

Linear regression using a logistically transformed y-variable (price). I used a stepwise selection method but modified the confidence interval

**2. Why did you choose this technique? In other words, why was this technique a good  
match for this scenario?**

Our y-variable, price, was numerical and skewed in a way that made sense for a log transformation to take place. I used stepwise selection on the normal y form and logged form, and the R-squared jumped from ~0.45 to ~0.60 without changing any other characteristics.  
**3. If you are using a supervised learning technique to answer this question, copy and paste  
the output from your final model below.**

<https://pastebin.com/rgNvvNHg>

The regression analysis explains the factors influencing the **price** of properties with an R-squared value of **0.601**, indicating that about 60.1% of the variability in price is explained by the model.  
**4. If you are using a supervised learning technique to answer this question, please explain  
how the model that you chose satisfies all the requirements of a “good model.”**

* All predictors of the y-variable are significant.
* Handles outliers well. See residual plot:A graph with blue dots

  Description automatically generated
* Adjusted r-squared of 0.597, a high enough predictor to be considered relevant. (The purely linear model had an adjusted r-squared of 0.450.)
* Strong Q-Q plot.

A line graph with a red line

Description automatically generated

**Please list at least two key business takeaways from your analysis that the real estate  
investor can use to help him address the chosen business scenario**

* What increases prices:
  + Larger properties: More bedrooms and the ability to accommodate more guests boost prices.
  + Prime neighborhoods: Properties in Rainey Street, Downtown, Tarrytown, and Zilker command higher prices.
  + Unique property types: Houses tend to perform better compared to less traditional options.
* What decreases prices:
  + Room types: Private and shared rooms have lower prices compared to entire homes.
  + Cancellation policies: Stricter policies and those labeled as "moderate" tend to reduce demand and pricing.
  + Superhost status: Interestingly, being a superhost slightly lowers the price, perhaps reflecting more affordable pricing strategies to maintain occupancy.

**SCENARIO TWO: PREDICTORS OF BOOKING AVAILABILITY**

for output see austin\_booking.ipynb  
**1. What analytical technique(s) did you choose to address this scenario?**

A classification tree with a minimum node size of 75, no depth limit, and an alpha level of 0.

**2. Why did you choose this technique? In other words, why was this technique a good  
match for this scenario?**

Our y-variable, whether the Airbnb listing was booked (represented by variable *booked*) was a binomial variable, so I chose to use a decision tree to model it, since we wanted to find types of listings that would be more available than others.

**3. If you are using a supervised learning technique to answer this question, copy and paste  
the output from your final model below.**

See decision\_tree\_final.pdf in folder

-- train set --

Accuracy : 0.7056

Precision: 0.5925

Recall : 0.3208

ROC: 0.7147

-- test set --

Accuracy : 0.6985

Precision: 0.5404

Recall : 0.2829

ROC: 0.6849

**4. If you are using a supervised learning technique to answer this question, please explain  
how the model that you chose satisfies all the requirements of a “good model.”**

It had the best overall testing averages of all the trees I tried. The tuned model with the same tuning constraints as the code examples provided for class produced a tuned test set (min node size 100, no depth limit, and ccp alpha = 0) with lower overall scores for all four evaluation statistics.

Although recall is low for both train and test, the other statistics are strong enough to consider the model to be at least a bit informative.

**Please list at least two key business takeaways from your analysis that the real estate  
investor can use to help him address the chosen business scenario**

* The **number of reviews** a listing has had is the most influential factor for predicting bookings. Listings with more reviews are more likely to be booked consistently.
  + Prioritize properties with a strong history of guest reviews or implement strategies to encourage guests to leave reviews, such as providing exceptional experiences and follow-up reminders.
* **Price** is the second most significant factor influencing bookings. Competitive pricing correlates with higher booking rates.
  + Use dynamic pricing tools to adjust prices based on seasonality, demand, and competition, ensuring the property remains attractive to prospective guests.