Title: Panel Data Analysis of Airbnb Listings

Author: Rohan chawla

Identifying ID and Time Variables: In the provided Airbnb dataset, the variable propertyid uniquely identifies each listing and serves as the panel ID. The variable time denotes the monthly period (unbalanced panel, T = 2 to 49), making the data suitable for panel data techniques. The dependent variable of interest is avg_price, with rating as the primary independent variable.

Influence of Rating Score on Average Price. To explore how a listing's rating affects its average price, we estimated three models:

1. **Pooled OLS (Ordinary Least Squares)**
2. **Fixed Effects (FE)**
3. **Random Effects (RE)**

Pooled OLS: This model treats all listings as identical, ignoring property-specific traits. According to the OLS results, a one-point increase in a property's rating leads to an average price increase of $38.68. Importantly, this effect is **statistically significant** (p = 0.0299), suggesting that if we trusted the OLS model's assumptions, we might believe that better-rated listings command higher prices. However, this model oversimplifies reality, ignoring that a 4.8 rating might mean very different things for a downtown studio versus a beach villa. Moreover, the R^2 is extremely low (less than 1%), indicating that ratings explain almost none of the price variation. So, while the coefficient is statistically significant, the model lacks reliability due to poor overall fit.

Fixed Effects: This model focuses on changes within each listing over time, controlling for their unique, constant traits. It estimates that a one-point increase in rating correlates with a $13.22 *decrease* in average price. However, this estimate is **not statistically significant** (p = 0.1548), meaning that this negative relationship could be due to random variation in the data. The R^2 is near zero and even negative in the adjusted version, confirming that the model struggles to explain price changes.

Random Effects: This model blends listing-specific effects with general patterns. It estimates a $13.17 decrease in price per one-point rating increase—very similar to the FE result. Again, the estimate is **not statistically significant**(p = 0.1574). Although this model has the highest R^2 among the three (~1%), it still suggests that ratings have minimal influence on prices when considering both individual listing traits and market-level factors.

In conclusion, while the OLS model shows a statistically significant relationship, it does so at the cost of realism, ignoring listing-specific differences. The FE and RE models, which more appropriately handle panel structure, suggest that rating has no meaningful impact on price. Among them, RE is slightly better in model fit, though still weak.

Model Comparison Using Fitted Plots and Individual Heterogeneity:

Model comparison involves understanding how each method handles differences between listings and fits the data.

1. **Individual Heterogeneity**: *Plot means (Figure A1)* display the average price for each listing. The significant variation confirms that each Airbnb property is different. OLS assumes no such variation, which is unrealistic. In contrast, FE and RE models account for these differences, which is critical when analyzing a marketplace like Airbnb.
2. **Fixed Effects Intercepts**: *Intercept Coefficient Plot (Figure A2)* visualizes the baseline pricing level of each property. This further emphasizes that each listing starts from a different price point. FE captures this heterogeneity, making it more appropriate than OLS.
3. **Fitted Estimates**: *Fitted Lines Comparison (Figure A3)* shows how well each model predicts prices across ratings. The RE line appears more stable and balanced. OLS provides a flat, over-generalized prediction, while FE is more erratic due to its strict within-property focus.

In conclusion, the individual heterogeneity and how well predictions match the data, both RE and FE outperform OLS. RE's smoother, more stable predictions give it an edge in this context.

Choosing the Most Appropriate Model:

From a modeling perspective, the **Random Effects** model stands out as the most practical and statistically reliable option for this dataset.

- It accounts for individual differences between listings (unlike OLS).
- It provides more stable, interpretable predictions than Fixed Effects.
- It balances between capturing general patterns and respecting listing-level uniqueness.

In platforms like Airbnb, where each listing has its own story, amenities, and pricing logic, a model that combines both specificity and generalization is ideal. Random Effects fulfills this role effectively.

Model Selection Justified Through Intuition, Visualizations, and Hausman Test:

We assess model selection through three lenses: visual interpretation, statistical testing, and intuitive business reasoning.

1. The *Plotmeans (Figure A1)* shows considerable price variation across listings, confirming the need to model individual property differences. OLS does not accommodate this.
    - The *Intercept Plot (Figure A2)* displays diverse fixed effects, again highlighting variation that OLS ignores.
    - *Fitted Value Plot (Figure A3)* indicates RE's smoother and more stable prediction line compared to OLS and FE.
2. **Hausman Test:** The Hausman test helps determine whether the FE or RE model provides more consistent estimates. With a **p-value of 0.9445**, the test suggests no meaningful difference in consistency between FE and RE. This supports the use of the more efficient RE model.
3. **Intuition:** Airbnb properties vary widely in location, amenities, and seasonal pricing. It is logical to believe that both listing-specific traits and market-wide factors

influence price. RE is the only model that considers both these dimensions simultaneously.

References

- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica.*
- Torres-Reyna, O. (2007). Panel Data 101: https://www.princeton.edu/~otorres/Panel101.pdf

Appendices

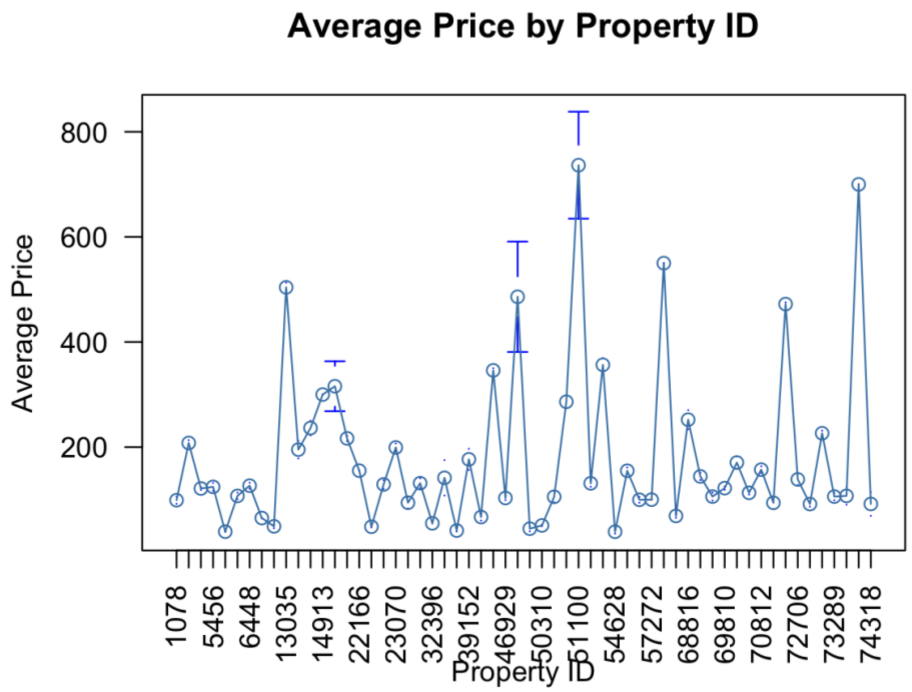

Figure A1 Plot means of Property ID and Average Price

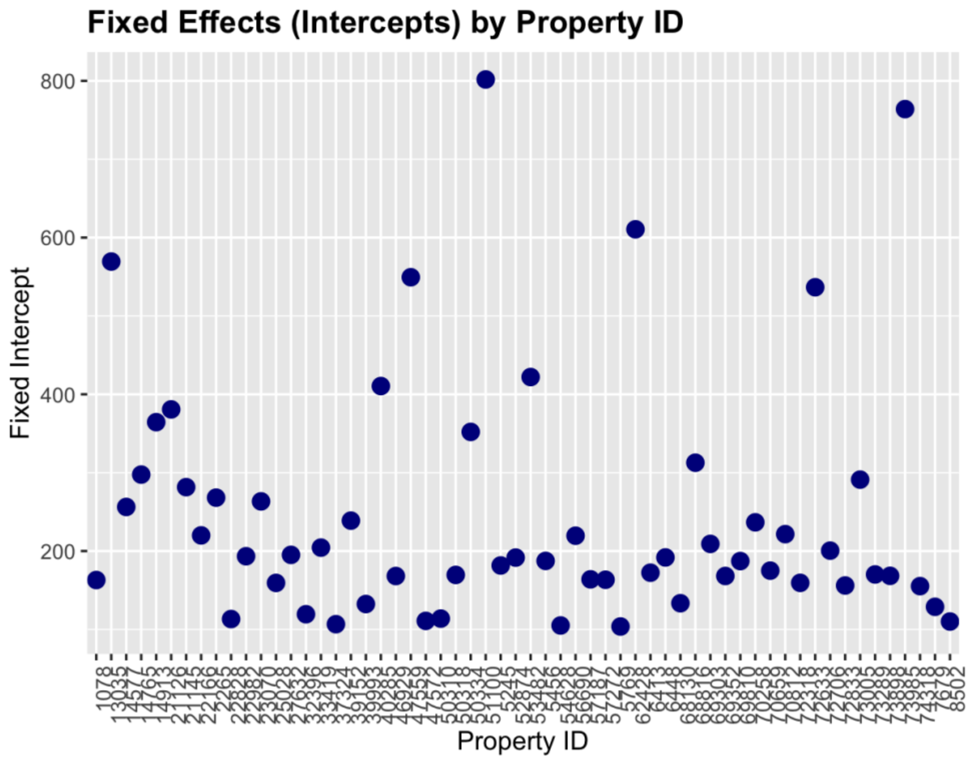

Figure A2: *Intercept Coefficient Plot for Fixed effect model*

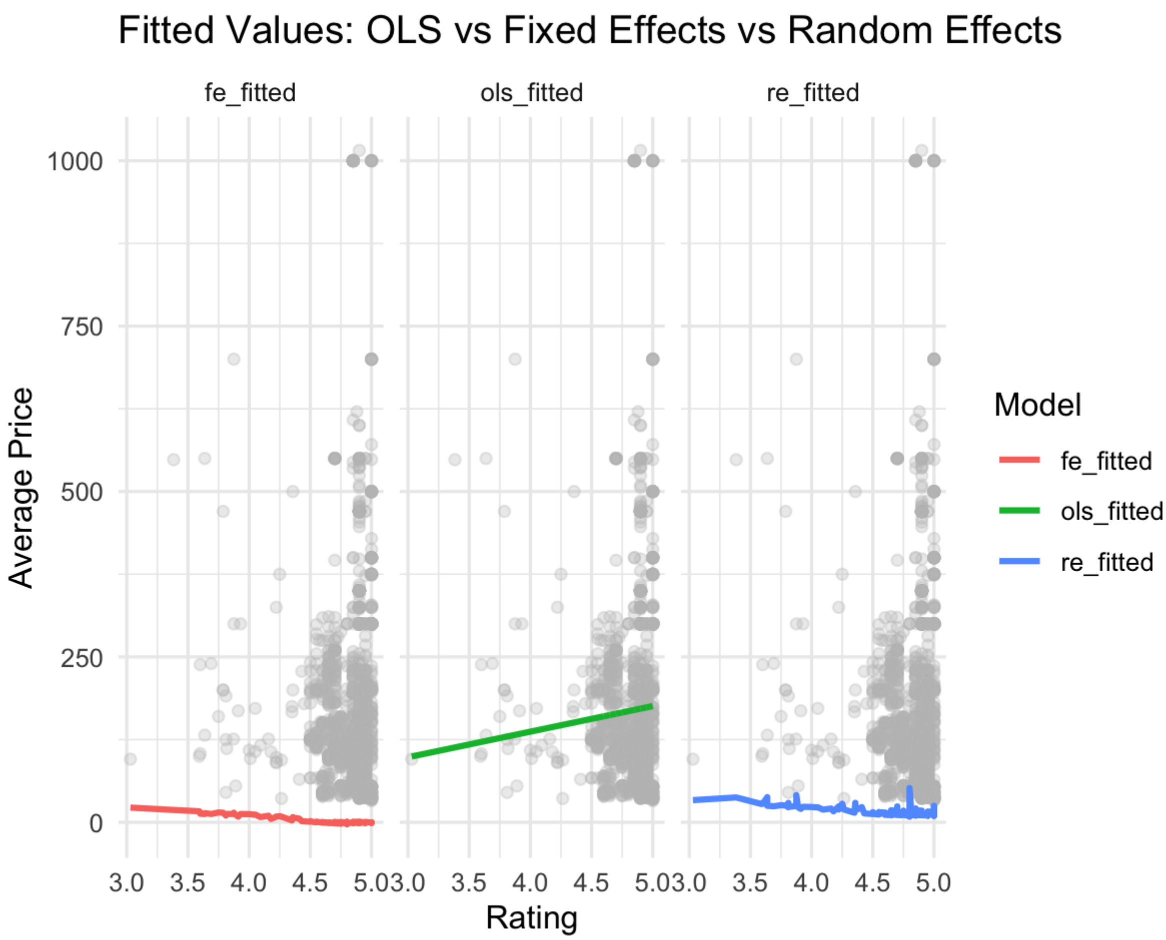

Figure A3: Fitted values plot for OLD, Random effect and Fixed effect