Airbnb Price Predictions

Sungwan (PK) Kim

Data

- **Observations:** 3,000 postings

- **Independent Variables (57):** Location, Capacity, Amenities, Ratings, Communication, Cleanliness, Host Info, and so on

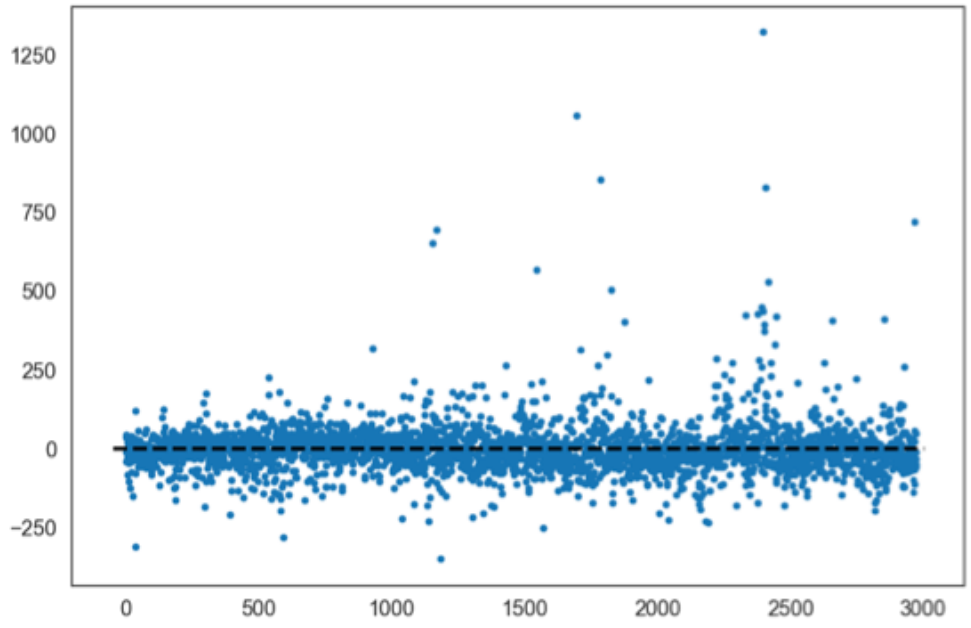
- **Dependent Variable:** Price

Linear Regression Results

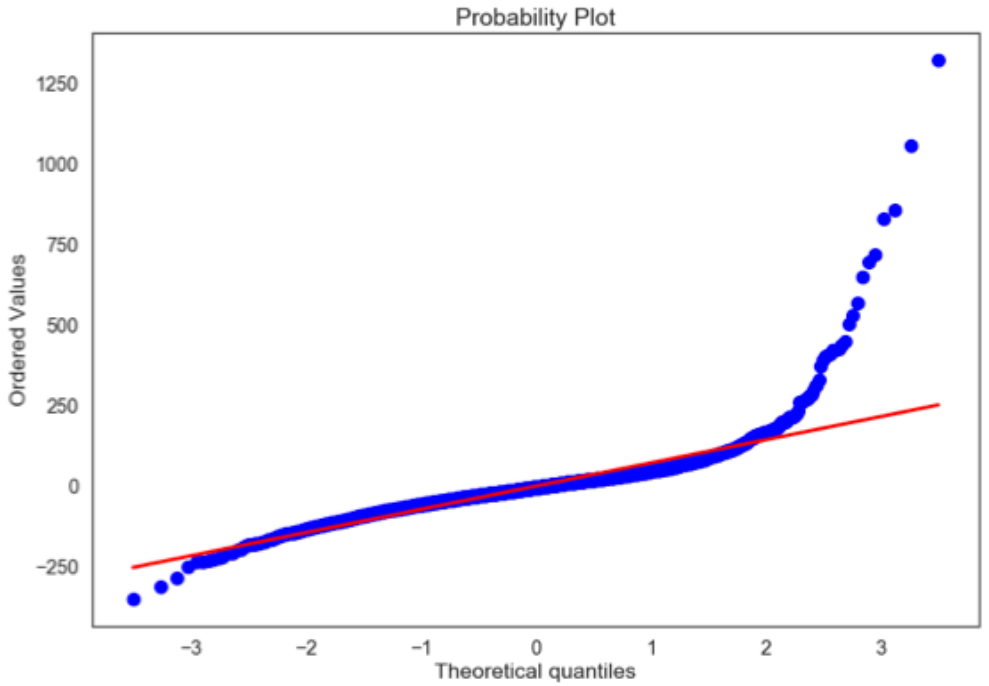
- R^2 = 0.523
- Adjusted R^2 = 0.514

Normalization Test - P(Omnibus): 0.000 | P(JB): 0.00

Residual Plot

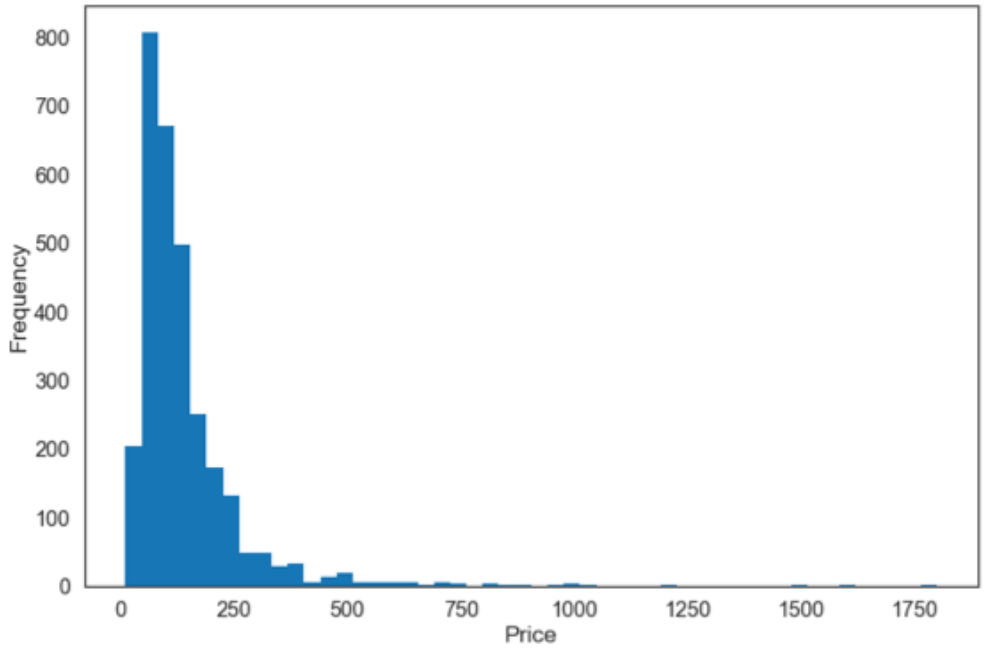


Normal Q-Q plot

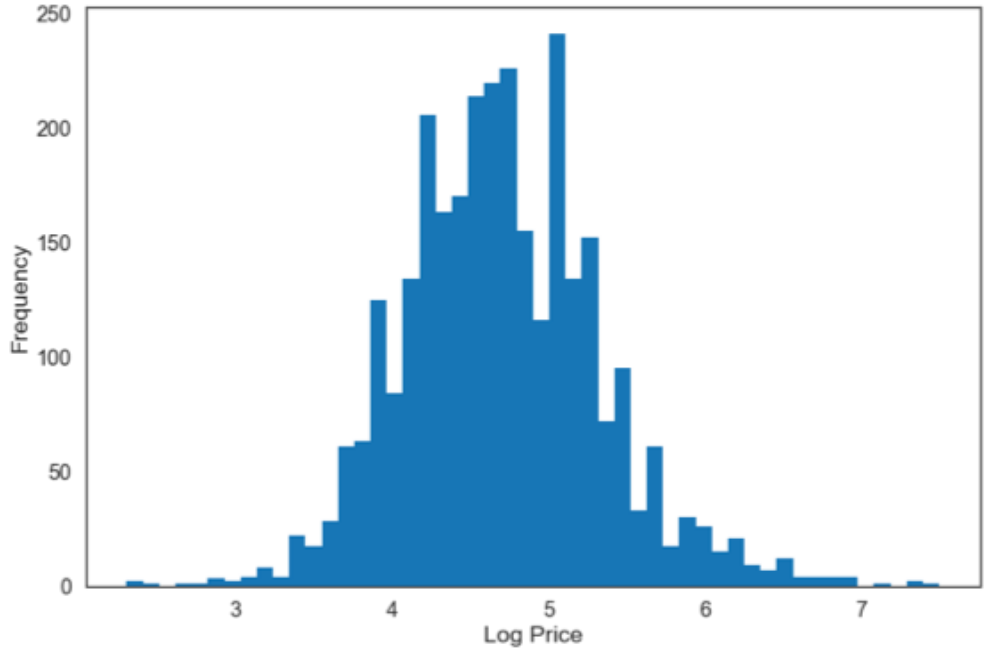


Price Distribution

Before

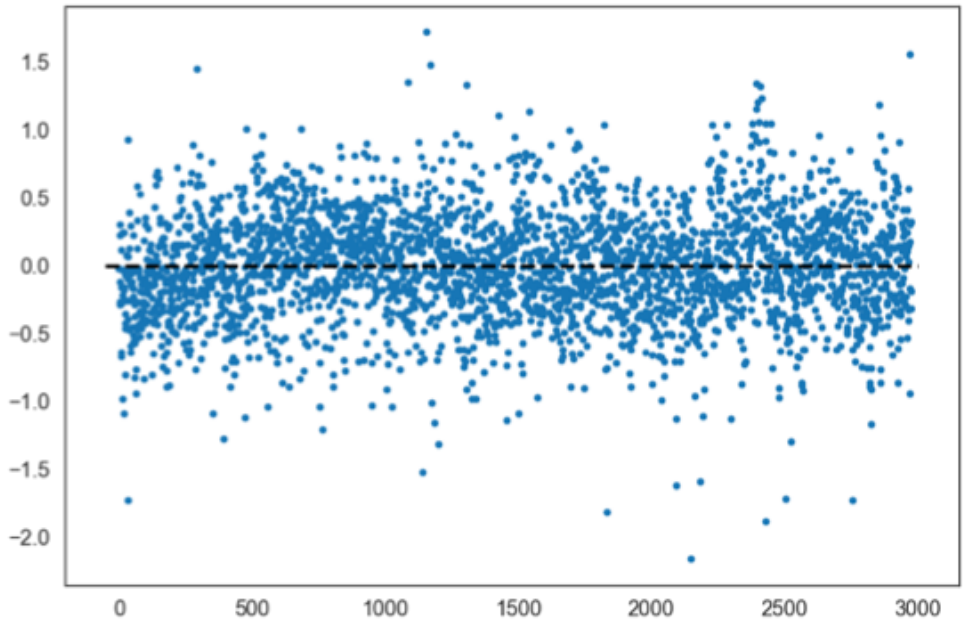


After Log Transformation

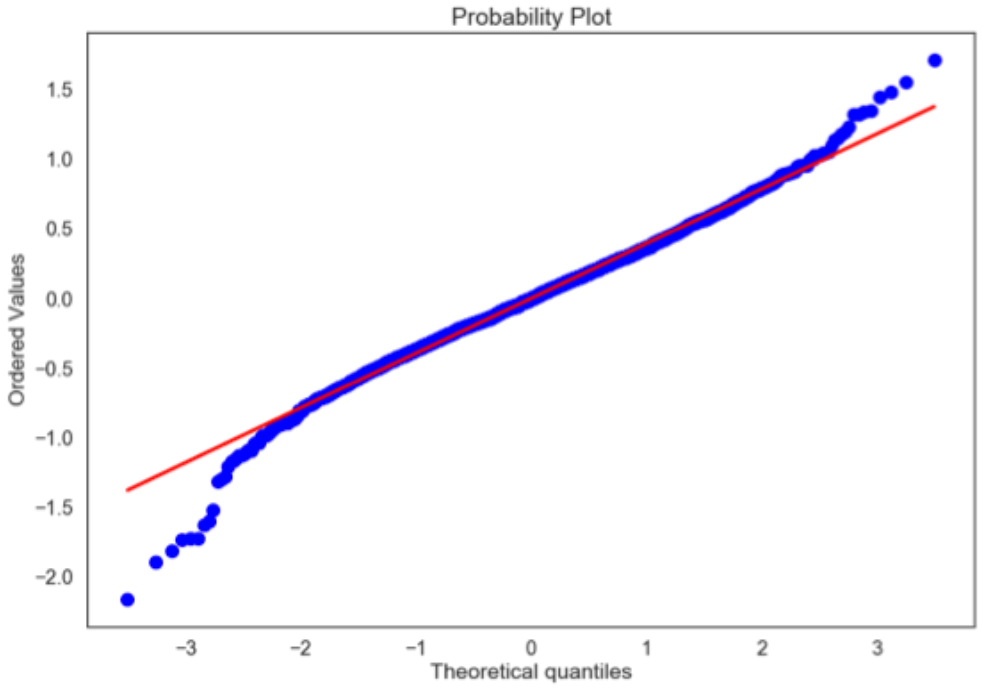


Residuals Normalized

Residual Plot



Normal Q-Q plot



Cross-Validation

Before

- Test / Train Split : $R^2 = 0.419$
- 10 - Fold CV : $R^2 = 0.364$

After Log-Transformation

- Test / Train Split : $R^2 = 0.515$
- 10 - Fold CV : $R^2 = 0.507$

Regularization - (10 - Fold CV)

Before

- Ridge : $R^2 = 0.365$
- Lasso : $R^2 = 0.364$
- ElasticNet : $R^2 = 0.364$

After Log-Transformation

- Ridge : $R^2 = 0.499$
- Lasso : $R^2 = 0.498$
- ElasticNet : $R^2 = 0.498$

Takeaways

- Location | Capacity | Pool
- Entire Place | Communication

Questions?