My first real project employs numerical imputation, ordinal encoding, nominal dummy variables, and natural logarithmic transform to produce a clean, repressible dataset. After some EDA and DataViz, we address some potentially redundant features, and form null and alternative hypotheses about the potential multicollinearity about such features. Using both linear and ensemble models, we train on both datasets, then compare error metrics with root mean squared error and R2 score. Once we’ve detected a best-fit, we compare our predictions with the results, and check the distribution of our residuals.