MVP

Data has been successfully scraped from both Steam and SteamSpy to create a baseline model. Initial modeling showed a heavily uneven distribution of the target variable, however, causing heteroskedasticity in the residuals. To fix this, I have begun trying different transformations on the target variable (owners) to re-interpret the linear correlations with the model's features. Included in my Steam_Analysis repository is a notebook labeled "target_transformations" that show a more detailed look into how each of these transformations are plotting and scoring. After selecting which transformation produces the lowest error cost, categorical features and interactions will be added, as well as other feature transformations to find the right level of complexity. Below are the residual plots for a simple regression, log-transformed regression, and root-transformed regression, respectively.