**PCA**

To perform PCA we used the label encoder to convert non-numeric features to numeric values. We then applied PCA without limiting the number of components to get a sense of the data. The quick tailing off of variance in the dataset explained by each subsequent component suggests that not many features in the housing training dataset are strongly correlated. We then limited the number of components for the analysis. It is worth noting that a reduction in the dimensionality of the data by half -- from 80 columns in the dataset to 40 components -- still encompasses 80% of the total variation in the dataset. If one wishes to drastically reduce dimensionality we would suggest using 15 components, which is sufficient to incorporate ~50% of the total variance.