From the elbow curve, it could be noted that the best K-mean value is 3. Plots were done although with the K-mean values of 3 and 4 to compare the output.

The variance ratio of the first principal component is approximately 99.76%. This shows that the dataset is highly skewed towards the first principal component indicating a strong underlying patterns or structure of the dataset. Therefore, a PCA dimensionality reduction on the dataset can be done with just the first principal component.

An additional analysis to gain a deeper understanding of the model’s performance was done by visualizing the predicted versus the actual values. The residuals were examined and other evaluation metrics such as R-squared, intercept and the coefficient.

A mean squared error value 0.0864958441960215 which lower and closer to zero for this analysis predicting the target variable, which is the age bracket, indicates the model is very accurate.

The intercept value of -2.010 might not have a direct interpretation in terms of age since the age bracket is categorical. Since the value of the coefficient is negative an increase in predictor leads to a decrease in age bracket therefore it has less influence in predicting the age bracket.