\section{Association Between Variables}

\subsubsection{ANOVA Test}

The relationship between sentiment and both match result and PPG Category were analyzed using Analysis of Variance (ANOVA). ANOVA is a widely employed statistical method designed to test for differences in means across multiple groups. This is achieved by comparing the variability within each group to the variability that exists among different groups. This comparison helps to determine if there exists a statistically significant difference in the examined variables.

\subsubsection{Eta-Squared ($\eta^2$)}

In order to quantify the strength of the relationships, Eta-squared $\eta^2$ was calculated. Eta-squared quantifies the proportion of the total variance in sentiment that can be explained by both match results and PPG category. This measure offers valuable insight into the practical significance of the relationship, enabling the assessment of the magnitude of the effect in addition to the statistical significance provided by the ANOVA p-value.

The formula for calculating Eta-squared is:

\[

\eta^2 = \frac{{\text{{SSB}}}}{{\text{{SST}}}}

\]

\textit{where SSB is the sum of squares between groups and SST is the total sum of squares.}

In order to accommodate the disparate ranges of tweets, the sentiment data was transformed into proportions, expressed as decimals. This normalization procedure facilitated more meaningful comparisons across teams characterized by varying tweet volumes. The calculations for Eta-squared and the ANOVA were executed utilizing the Python programming language, in conjunction with the pingouin statistical package. This method enabled a streamlined and efficient examination of the relationship while concurrently providing the requisite flexibility and computation power necessary for managing decimal data.

By incorporating Eta-squared alongside ANOVA, the significant conclusions were successfully extracted pertaining to the relationship between sentiment and outcome, while guaranteeing the analytical robustness and appropriate adjustment for the distinct features inherent in our dataset.