“Sum”: specific to the model that, which “contrast” is to use, to compare how the different levels of a categorical variable contributes to the overall model. For “sum”, meaning we are comparing each level of the categorical variable to the grand mean of the dependent variable.

Normally can be useful when answering the question “How does each level of the categorical variable differ from the overall average of the dependent variable?”

“no Sum”, is assumed to be using “treatment coding" contrast, for the categorical variable. Compares each level of the categorical variable to a reference level.

"sum\_sq": the sum of squares for each source of variation (or each term in the model).

The factors or variables that contribute to the variability observed in a response variable. In the context of ANOVA, the source of variation is typically broken down into two types:

1. "Within-group variation" or "error variation" is the variation that is due to random or uncontrolled factors that affect all the groups or levels of the independent variable in the same way.
2. "Between-group variation" is the variation that is due to the independent variable itself, or the factor being studied. This variation represents the differences in the mean values of the response variable across the levels or groups of the independent variable.

By separating the variation into these two sources, ANOVA helps to identify which factors or variables are contributing significantly to the variability in the response variable, and how much of the variability can be explained by each of these factors or variables

A high sum\_sq indicates that a large proportion of the variability in the dependent variable is accounted for by the independent variables, while a low sum\_sq indicates that the independent variables do not explain much of the variability in the dependent variable.

"df": the degrees of freedom associated with each source of variation. It represents the number of observations minus the number of parameters estimated in that term.

"F": the F-statistic for each source of variation. It is the ratio of the explained variation to the unexplained variation (residuals) and it tests the null hypothesis that the coefficients for that term are all zero.

A high F-value indicates that the variation between groups is large relative to the variation within groups, suggesting that the independent variables are highly significant in explaining the variation in the dependent variable. A low F-value suggests that the variation between groups is small relative to the variation within groups, indicating that the independent variables are not highly significant in explaining the variation in the dependent variable.

"PR(>F)": the p-value associated with each F-statistic. It represents the probability of obtaining an F-value as extreme as the observed one if the null hypothesis were true. A p-value less than the chosen significance level (e.g. 0.05) indicates that the null hypothesis can be rejected and the corresponding term is considered statistically significant.

“Residual row” is the overall significance of the model

P<0.05, meaning the term is statistically significant

F = SS between / SS within

“+” adds a new variable to the model. For example, y ~ x1 + x2 means that the response variable y is modeled as a linear combination of x1 and x2.

“\*” adds the interaction term between two variables to the model. For example, y ~ x1 + x2 + x1\*x2 means that the response variable y is modeled as a linear combination of x1, x2, and the interaction between x1 and x2.

Y = x1 \* x2 +x3, include all possible interactions between x1, x2 and x3.

Y = x1:x2 + x3, only include interactions between x1 and x2, and treat x3 as a separate variable