Data Issues

Brandon Upper

The first major issue when forecasting sales is that there is a huge proportion of missing points. In the MarkDown1 col alone there is almost half of the data are just NA, which means there missing and we'll need to solve that issue somehow.

The second issue when forecasting sales is the fact that all of the holidays have been lumped into one separate column. This would cause issues as not all holidays are being considered according to that data set and the majors ones are being considered the same which is also an incorrect assumption as sales should massively differ across the major holidays.

The last issue I want to discuss here is high collinearity between features. Markdown4 and Markdown1 have a value of 0.819041866 for collinearty which on the natural scale of 0 - 1 is extremely high an indicates the variables are highly related and would pose a problem being separate features in forcasting sales