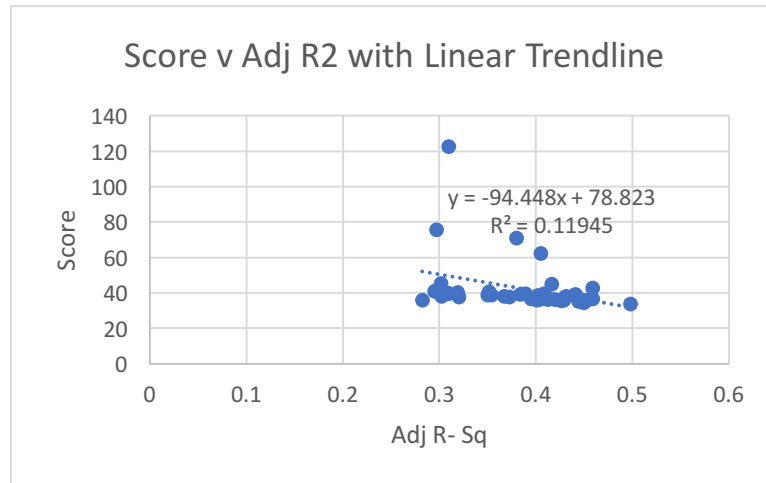


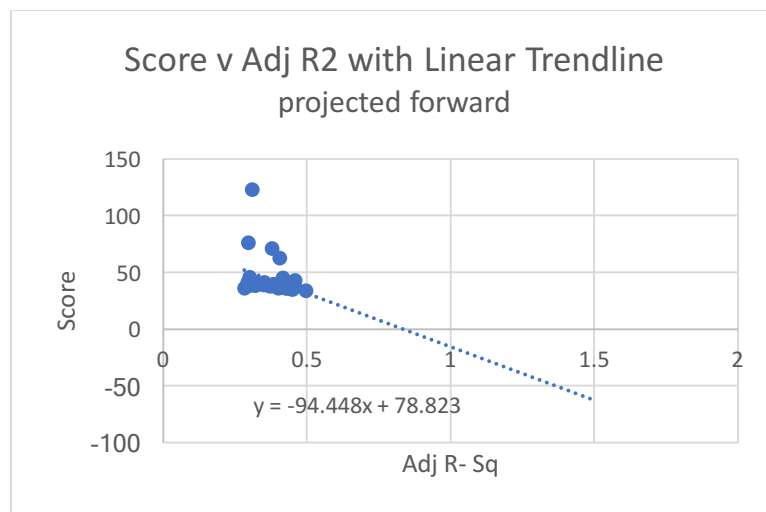
I started by plotting the data (Figure 1) we were given, and adding a trend line, which is the Excel way to make a linear model.

Figure 1



That result wasn't terribly informative, so I projected the line forward and got Figure 2

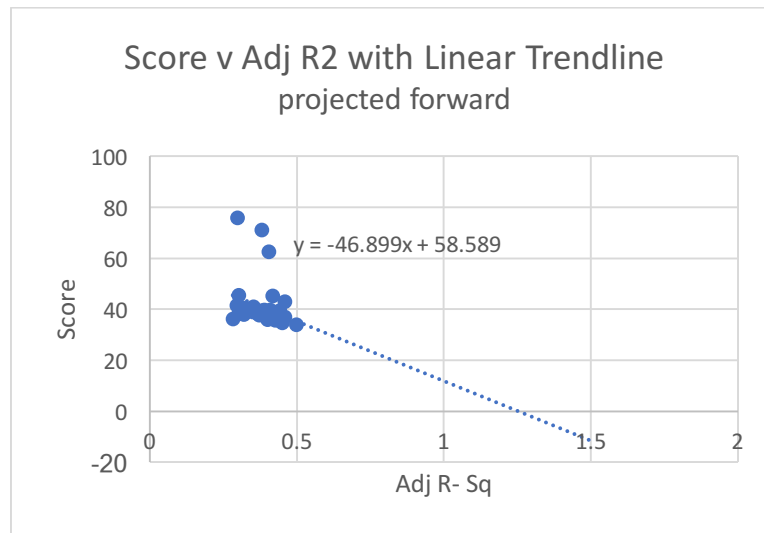
Figure 2



We know an Adjusted- $R^2$  value of 1.0 is a "perfect fit", we are told in the problem statement that an error of 0.0 is likewise a perfect fit. The line looks like there should be 0 error for Adjusted- $R^2$  values of  $\sim .84$ . However, there is an interesting question about the effect of those seeming outliers. Removing the error score of 122.656 changed the results significantly. Now, the line for the linear relationship between Adjusted- $R^2$  and error crosses the y-axis around

$x=1.25$  which is beyond the values Adjusted- $R^2$  can have. Removing additional seeming outliers pushes the x-intercept even higher. As seen in Figure 3. I conclude from this that no, on its own, Adjusted- $R^2$  is not a great predictor of performance on new data.

Figure 3



I think there is value in looking at Adjusted- $R^2$  when comparing 2 models to each other, but it is not a robust predictor of overall model goodness. To decide if a model is “ready for prime time” and deployment, I think there needs to be information on the performance the model showed with test data. Perhaps the area under an ROC curve for the test set.