(5.) This happens because Mean Absolute Error (MAE) applies a different penalty to deviations of expected from actual than does test set R^2. In particular, MAE is the 1-norm, and R^2 incorporates the 2-norm. For example, when $|y_i - \hat{y}_i| > 1$, the 2-norm in R^2 will penalize that error more heavily than will the 1-norm = MAE.

Below, I show a numerical example where (given actual values) predictions $\hat{y}_1$ from Model 1 have higher MAE but also higher test R^2 than predictions $\hat{y}_2$ from model 2. Thus, MAE would prefer model 2, but test set R^2 would prefer Model 1.

	Test Actual	$\hat{y}_1$	$\hat{y}_2$	$\lvert y-\hat{y}_1\rvert$	$\lvert y-\hat{y}_2\rvert$	$(y-\hat{y}_1)^2$	$(y-\hat{y}_2)^2$
	-2	-3	-3.5	1	1.5	1	2.25
	-1	-2	-1	1	0	1	0
	1	2	1	1	0	1	0
	2	3	3.5	1	1.5	1	2.25
Avg	0	0	0	$MAE_1 = \frac{4}{4} = 1.0$	$MAE_2 = 3/4$	$\Sigma = 4$	$\Sigma = 4.5$

$SST = 2^2 + 1^2 + 1^2 + 2^2 = 10$

$R_1^2 = \frac{10-4}{10} = 0.6$

$R_2^2 = \frac{10-4.5}{10} = 0.55$

Thus MAE would prefer Model 2, but test R^2 would prefer Model 1.