

Linear Model Summaries

Model 1

Predictors	Estimate	Std. Error	t value	p-value
Intercept	144.96537	17.53432	8.268	2.65e-15 ***
GHI	6.13850	0.06355	96.597	< 2e-16 ***
Temp	-11.02146	0.57909	-19.032	< 2e-16 ***
WS	-1.97711	4.11843	-0.480	0.631

Note. GHI = Global Horizontal Irradiance, Temp = Ambient Temperature (C), WS = Wind Speed (mph). *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 86.48 on 361 degrees of freedom

Multiple R-squared: 0.9725, Adjusted R-squared: 0.9723

F-statistic: 4259 on 3 and 361 DF, p-value: < 2.2e-16

Model 2

Predictors	Estimate	Std. Error	t value	p-value
Intercept	25.429126	9.038885	2.813	0.00517 **
GHI	7.038609	0.071641	98.248	< 2e-16 ***
GHI:Temp	-0.065587	0.002401	-27.319	< 2e-16 ***

Note. GHI = Global Horizontal Irradiance, Temp = Ambient Temperature (C), : denotes interaction (product of two variables). *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 69.96 on 362 degrees of freedom

Multiple R-squared: 0.982, Adjusted R-squared: 0.9819

F-statistic: 9855 on 2 and 362 DF, p-value: < 2.2e-16

Model 3

Predictors	Estimate	Std. Error	t value	p-value
Intercept	141.14925	13.95269	10.12	<2e-16 ***
GHI	5.40070	0.06853	78.81	< 2e-16 ***

Note. GHI = Global Horizontal Irradiance. *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 122.2 on 363 degrees of freedom

Multiple R-squared: 0.9448, Adjusted R-squared: 0.9446

F-statistic: 6211 on 1 and 363 DF, p-value: < 2.2e-16

Takeaways: Basing our decision off of Multiple R-squared (correlation coefficient) and lowest p-values, we can deduce that Model 2 with the highest R² of 0.982 is the best model to use going forward. The coefficients for this model are 25.429126 for intercept, 7.038609 for GHI and -0.065587 for the interaction effect of GHI and Temp. This can be written into a linear equation of $y = 25.429126 + 7.038609(\text{GHI}) - 0.065587(\text{GHI:Temp})$. Where $y = \text{kWh}$, and GHI:Temp is the product of the two.

Summary Stats on Predicted vs Observed Values (Using Model 2)

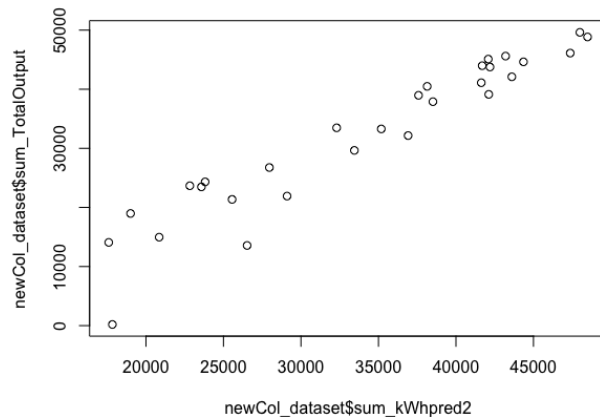
Five number summary stats:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Energy Ratio (monthly)	0.01096	0.88654	0.98721	0.91662	1.03623	1.07161
kWh Predicted (monthly)	17597	25556	36913	34189	42114	48489
kWh Observed (monthly)	195.5	23478.3	33472.3	32385.3	43759.5	49621.3

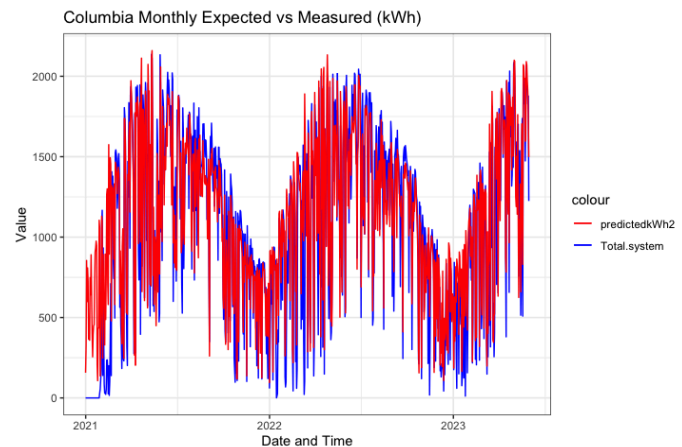
Model 2 predicted values summary stats:

	RMSE	R
Model 2	4863.94	0.9503407

Expected vs Observed (kWh)



Monthly Expected vs Measured (kWh)



Note. newCol_dataset\$sum_TotalOutput = Observed/Actual Monthly sum (kWh), newCol_dataset2\$sum_kWhpred2 = Monthly sum predicted kWh. Total.system = TotalOutput observed kWh.

Takeaways: As we can see from the above plots and five number summary statistics, the predicted/expected kWh monthly values fit exceptionally well to the observed ones. The high R value indicates a strong linear relationship between predicted values and observed ones. The minimum ER of 0.01096 can be disregarded as it comes from an incomplete month, this also explains the minimum Total Output sum value of 195.5.