Linear Model Summaries

Model 1

Predictors	Estimate	Std. Error	t value	p-value
Intercept	92.26927	9.88023	9.339	<2e-16 ***
GHI	2.78757	0.03492	79.819	< 2e-16 ***
Temp	-5.36250	0.31936	-16.791	< 2e-16 ***
WS	-2.99576	2.42173	-1.237	0.217

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: 48.31 on 361 degrees of freedom
Multiple R-squared: 0.9595, Adjusted R-squared: 0.9592
F-statistic: 2853 on 3 and 361 DF, p-value: < 2.2e-16

Model 2

Predictors	Estimate	Std. Error	t value	p-value
Intercept	28.546200	5.403783	5.283	2.2e-07 ***
GHI	3.209817	0.042739	75.103	< 2e-16 ***
GHI:Temp	-0.030875	0.001439	-21.453	< 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: 42.72 on 362 degrees of freedom
Multiple R-squared: 0.9683, Adjusted R-squared: 0.9681
F-statistic: 5521 on 2 and 362 DF, p-value: < 2.2e-16

Model 3

Predictors	Estimate	Std. Error	t value	p-value
Intercept	80.52250	7.26954	11.08	<2e-16 ***
GHI	2.44827	0.03582	68.35	< 2e-16 ***

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

Residual standard error: 64.3 on 363 degrees of freedom
Multiple R-squared: 0.9279, Adjusted R-squared: 0.9277
F-statistic: 4672 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 R^2 of 0.9683 is the best model to use going forward. The coefficients for this model are 28.5462 for intercept, 3.209817 for GHI and -0.030875 for the interaction effect of GHI and Temp. This can be written into a linear equation of y = 28.5462 + 3.209817(GHI) - 0.030875(GHI:Temp). Where y=kWh and GHI:Temp is the product of the two.

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

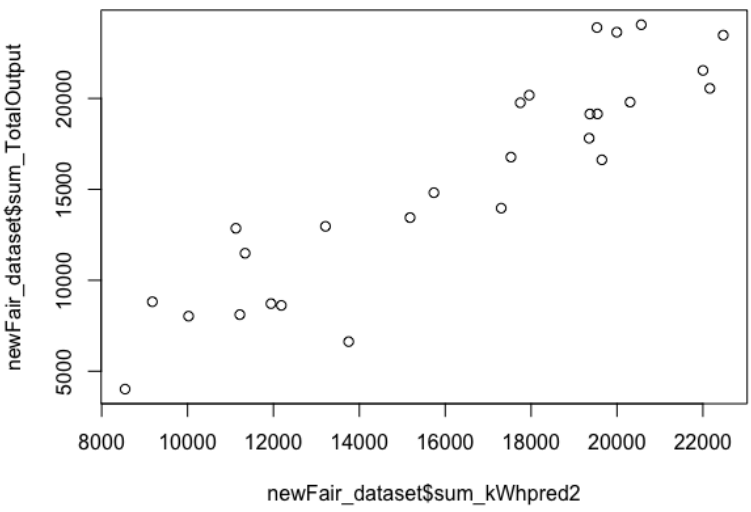
Five number summary stats:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Energy Ratio (monthly)	0.4699	0.8266	0.9608	0.9292	1.0288	1.2235
kWh Predicted (monthly)	8545	12063	17530	16256	19596	22473
kWh Observed (monthly)	4016	10153	16624	15514	19988	24053

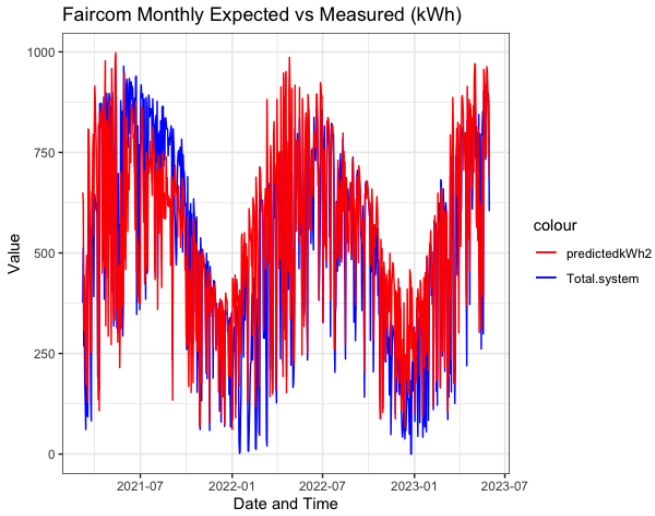
Model 2 predicted values summary stats:

	RMSE	R
Model 2	2695.129	0.9157884

Expected vs Observed (kWh)



Monthly Expected vs Measured (kWh)



Note. newFair_dataset$sum_TotalOutput = Observed/Actual Monthly sum (kWh), newFair_dataset2$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 are higher on average than the observed ones. However, the high R value indicates a strong linear relationship between predicted values and observed ones so we know the model fits fairly well. This could be due to a discrepancy in the Oiko weather data used to predict kWh values or numerous other factors. We also see a minimum ER of 0.4699 which is well below where we'd like to be, we can check the excel output chart to see which month this occurred in and investigate further.