

Machine Learning Foundations for Product Managers

Course Project

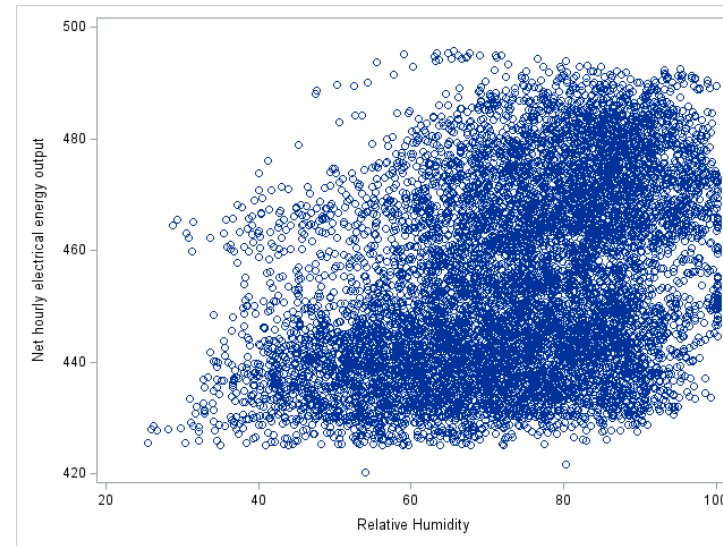
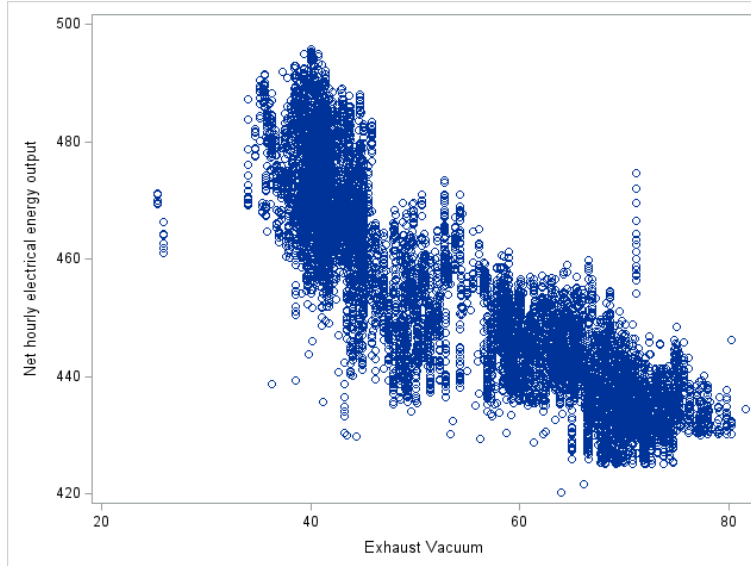
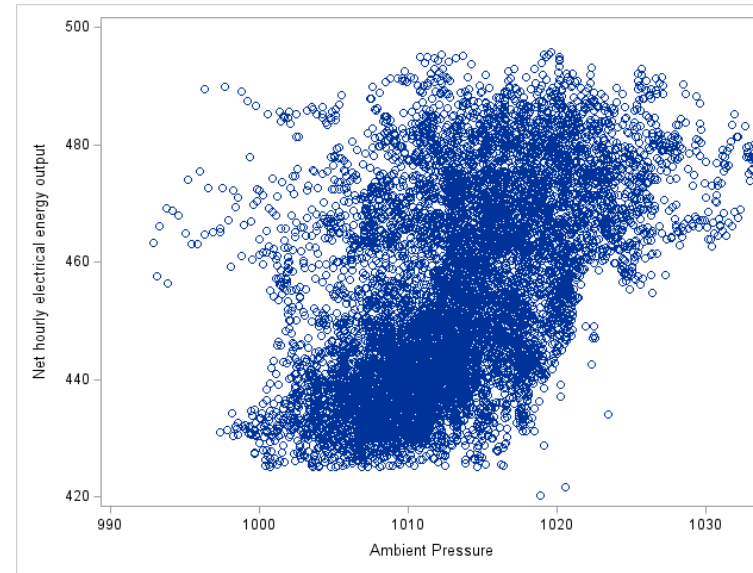
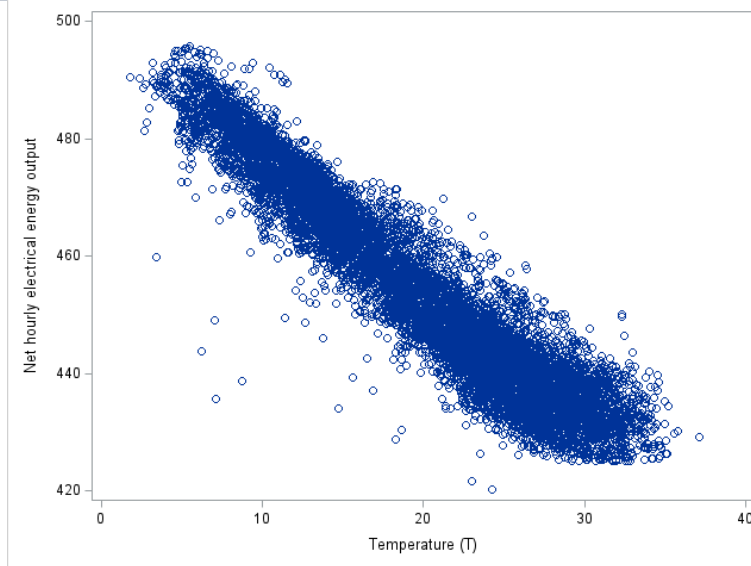
PROJECT TOPIC

- Build a model to predict the electrical energy output of a [Combined Cycle Power Plant](#).
- We have a set of 9568 hourly average ambient environmental readings from sensors at the power plant which we will use in our model.
- The columns in the data consist of hourly average ambient variables:
 - Temperature (T) in the range 1.81°C to 37.11°C
 - Ambient Pressure (AP) in the range 992.89-1033.30 millibar
 - Relative Humidity (RH) in the range 25.56% to 100.16%
 - Exhaust Vacuum (V) in the range 25.36-81.56 cm Hg
 - Net hourly electrical energy output (PE) 420.26-495.76 MW (Target to predict)

DATASET (shape)

	Ambient_Temperature	Exhaust_Vacuum	Ambient_Pressure	Relative_Humidity	electrical_energy
1	14.96	41.76	1024.07	73.17	463.26
2	25.18	62.96	1020.04	59.08	444.37
3	5.11	39.4	1012.16	92.14	488.56
4	20.86	57.32	1010.24	76.64	446.48
5	10.82	37.5	1009.23	96.62	473.9
6	26.27	59.44	1012.23	58.77	443.67
7	15.89	43.96	1014.02	75.24	467.35
8	9.48	44.71	1019.12	66.43	478.42
9	14.64	45	1021.78	41.25	475.98
10	11.74	43.56	1015.14	70.72	477.5
11	17.99	43.72	1008.64	75.04	453.02
12	20.14	46.93	1014.66	64.22	453.99
13	24.34	73.5	1011.31	84.15	440.29
14	25.71	58.59	1012.77	61.83	451.28
15	26.19	69.34	1009.48	87.59	433.99
16	21.42	43.79	1015.76	43.08	462.19
17	18.21	45	1022.86	48.84	467.54
18	11.04	41.74	1022.6	77.51	477.2
19	14.45	52.75	1023.97	63.59	459.85
20	13.97	38.47	1015.15	55.28	464.3

Visualizing Relationships

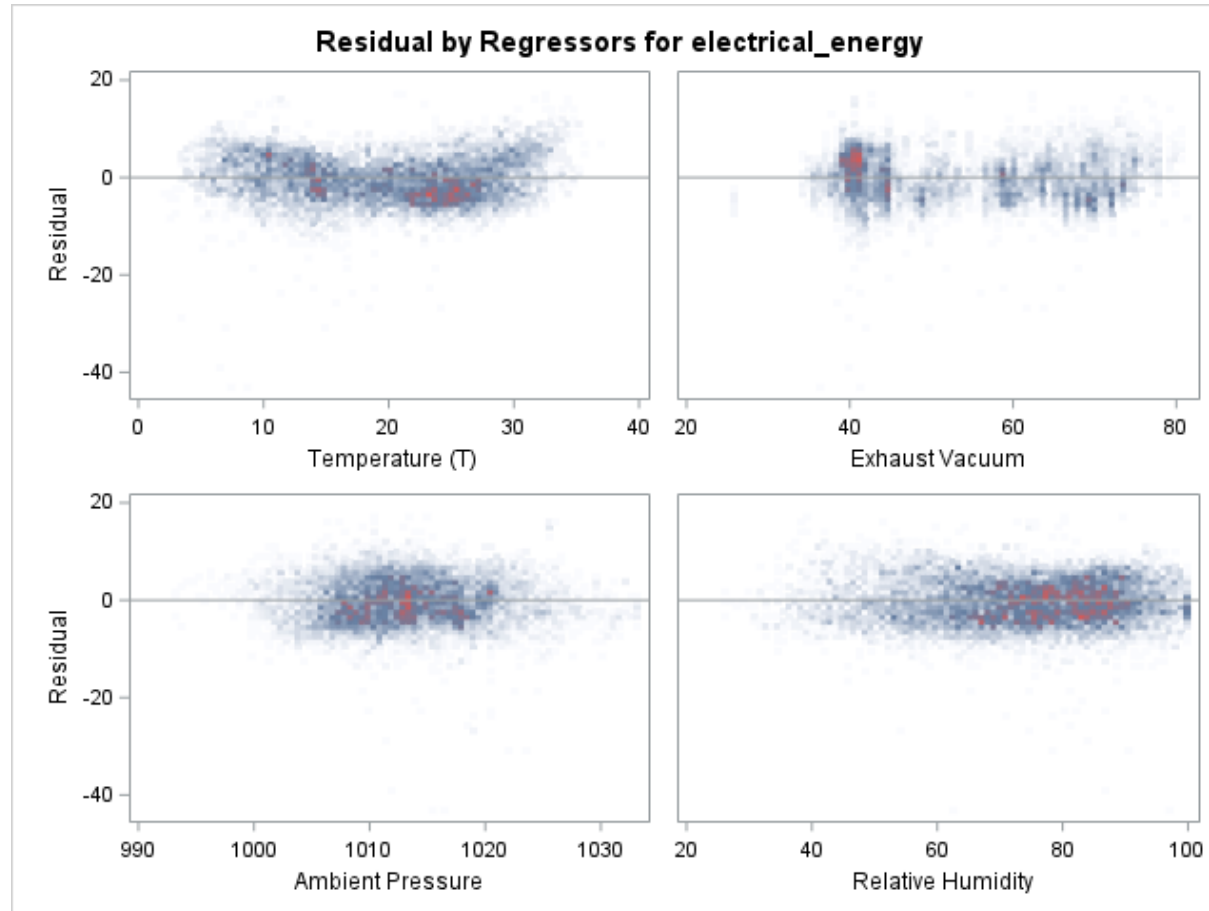


SELECTION OF VARIABLES

- **STEPWISE SELECTION**
 - Step by step selection to add/remove variables at each step
 - Based on statistics value: significance level for entry=0.05 / to keep in the model : stay=0.05
- **Results :**
 - All variables left in the model are significant at the 0.0500 level.
 - All variables have been entered into the model

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Ambient_Temperature		Temperature (T)	1	0.8989	0.8989	3988.75	85097.8	<.0001
2	Relative_Humidity		Relative Humidity	2	0.0220	0.9209	1040.13	2661.98	<.0001
3	Exhaust_Vacuum		Exhaust Vacuum	3	0.0074	0.9284	46.0872	991.68	<.0001
4	Ambient_Pressure		Ambient Pressure	4	0.0003	0.9287	5.0000	43.09	<.0001

LINEAR REGRESSION MODEL



MODEL ESTIMATES

- overall model fit statistics such as R^2 , adjusted R^2 , and Akaike Information Criterion (AIC), which can be interpreted to assess the quality of the model as a whole.
- A high adjusted R^2 indicates a good fit

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2587995	646999	31138.3	<.0001
Error	9563	198702	20.77826		
Corrected Total	9567	2786698			

Root MSE	4.55832	R-Square	0.9287
Dependent Mean	454.36501	Adj R-Sq	0.9287
Coeff Var	1.00323		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	454.60927	9.74851	46.63	<.0001	0
Ambient_Temperature	Temperature (T)	1	-1.97751	0.01529	-129.34	<.0001	5.97760
Exhaust_Vacuum	Exhaust Vacuum	1	-0.23392	0.00728	-32.12	<.0001	3.94300
Ambient_Pressure	Ambient Pressure	1	0.06208	0.00946	6.56	<.0001	1.45264
Relative_Humidity	Relative Humidity	1	-0.15805	0.00417	-37.92	<.0001	1.70529