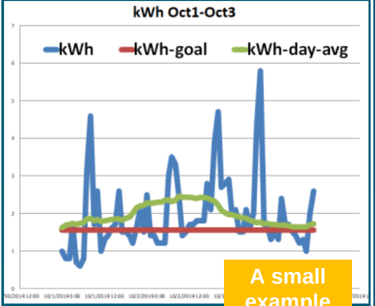
IT’S ELECTRIC!

[Reducing an Above-Average Electric Bill] – By Leland Ball

Dates:	Project Launch	Define	Measure	Analyze	Improve	Control
	Oct 3, 2019	Oct 5, 2019	Oct 12, 2019	Nov 16, 2019	Nov 20, 2019	Ongoing

Define

PROBLEM:
Expensive electric bill: **24% above average** for this zip code and is costing tenants money! A year of billing data shows every month’s/day’s electric usage is significantly **above average**!



A small example

Frequent thermostat adjustments indicate broken automated process with many **defects**

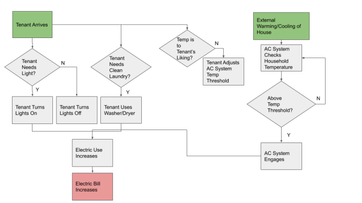
SQL: 2.1

Need to identify possible causes of most electric usage!

IMPACT:
Achieving average electric usage **could save over $720/year**

Measure

Process map shows both **manual** and **automated** steps that need measuring



Tenant presence data can be used to determine strength of human effect upon processes

Data Stratification Tree shows which data can be found, and where

Data Stratification Tree

Questions about process		Stratification Factors	Measurements	Samples
Does sunlight affect house electric usage independent of temperature?		External Temperature	Temperature in Fahrenheit by hour as measured by weather services (continuous)	For one month: 24*30=720 (per month)
Who is the biggest user of electricity, when home?		Thermostat Temperature	Fahrenheit values from thermostat's schedule when changed, by hour (continuous)	7 sets. Interpolated per hour for 720
What is the ideal thermostat setting (temperature @ hour)?	kWh	Who is Home	Presence of any of 4 tenants by hour, measured by presence-detecting tool (phone network monitoring) (discrete)	~40k samples, down-sampled to 720
Is energy use predominantly the fault of the AC system (internal/external temp mismatches)?	output	External Cloud/Sunlight Conditions	1-20 rating of sunlight intensity from cloudy/raining (1) to sunny (20) by hour from weather services (discrete)	720

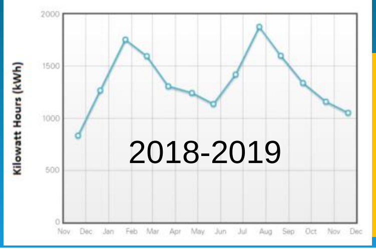
Temperature data, and other weather data obtained from weather services to see magnitude of effect upon processes

Analyze

WHERE TO LOOK FIRST?
Multiple Linear Regression was used to determine the biggest factors in electricity usage. A **hypothesis test** was carried out with 95% confidence (a=0.05).

	Coefficients	*Std. Err.*	*P-value*
Intercept	2.1605	0.1971	0.00000
tenant 1	0.1296	0.0766	0.09099
tenant 2	-0.1684	0.0674	**0.01262**
tenant 3	0.0712	0.0786	0.36518
tenant 4	0.0683	0.0851	0.42223
temp	-0.0107	0.0026	**0.00004**
cloudiness	-0.0099	0.0063	0.11251

Tenant 2’s presence and **outside temperature** have a statistically significant effect at this confidence level. With a p-value of < a=0.05, **H0 is rejected** for tenant 2, and outside temperatures. **This implies the presence of an AC unit in Tenant 2’s 3’rd story room, which is indeed the case.**
Descriptive statistics of each data-set shows large standard deviations for kWh measurements. This was the beginning of the **extreme seasonal variability** discovered in the data. This is expected even if process improvements are successfully made!



kWh usage varies significantly from month-to-month

Improve

Address thermostat settings. **Fewer manual adjustments show a process improvement**.

SQL: 3.2

Do the process adjustments (thermostat changes) also improve energy usage (and therefore electric bill improvements?)

pre nov-20 thermostat adjustments	
kWh/day	
Mean	34.34038462
Standard Error	1.374529289
Median	33.45
Mode	29.1
Standard Deviation	9.911871662
Sample Variance	98.24519985
Kurtosis	-0.408454332
Skewness	0.634962308
Range	36.9
Minimum	20
Maximum	56.9
Sum	1785.7
Count	52
Confidence Level(95.0%)	2.759482692

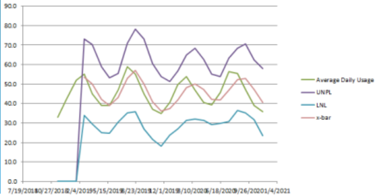
post nov-20 thermostat adjustments	
kWh/day	
Mean	41.4
Standard Error	1.364194136
Median	41.1
Mode	#N/A
Standard Deviation	4.918671907
Sample Variance	24.19333333
Kurtosis	0.161446019
Skewness	0.745350751
Range	16.6
Minimum	35.4
Maximum	52
Sum	538.2
Count	13
Confidence Level(95.0%)	2.972323686

Hypothesis test results: H0 is not rejected. Process changes did NOT result in energy savings!

Control

Modified control chart accounting for seasonality of data can now be used to track any future improvements that are made to the process!

Modified Exponential Smoothing used to predict seasonality



Define

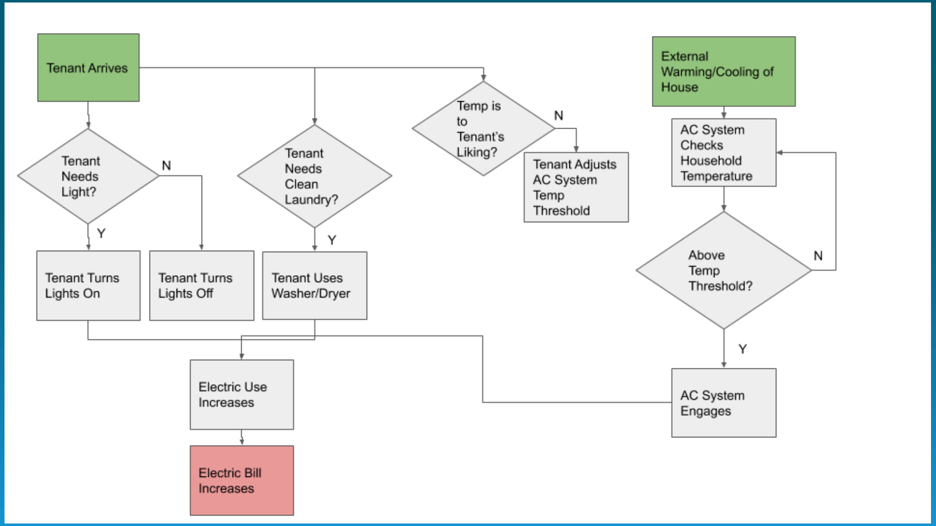
- Goal: Reduce electricity bill (cost is directly related to kWh)
  - 0% reduction (current state) - 1,268 kWh/mo
  - **12% reduction (goal) - 1,116 kWh/mo**
  - Identify and control sources of electrical use
  - **Maintain livable standards (simply turning off the electricity is not an option)**
- Operational Definitions:
  - For determining Sigma Quality Level (SQL) scores related to thermostat adjustments, the following definitions will be needed:
    - Unit: A single day (24 hours)
    - Defect: A manual adjustment of the thermostat, whether up or down, that does not match the scheduled setting
    - Opportunity: A given hour of a day. 24 opportunities exist in a given day
  - For measurements related to collecting data
    - Power data will be measured in terms of kWh, and total kWh per day
    - Presence data will consist of the yes/no presence of each of 4 tenants per hour
    - Temperature data will be measured in Fahrenheit/hr
    - Sun/light levels will be measured using standard weather terms, translated to a scale from 1 to 20, with 20 being the brightest and 1 being the most overcast and rainy
    - Thermostat changes will be recorded the hour they are made, with the difference in temperature (F) from the automated setting, to the manual value it was set to

Define (cont.)

- **<u>Process and process steps</u>**
- We are trying to improve the process where electricity use increases, **<u>while maintaining livability.</u>** In the process chart below, we have a variety of automated and manual processes that use electricity, started in a manual or automatic fashion

An improvement in this process is a decrease in electricity use, while maintaining a livable state inside house

Livability is reflected in the SQL score, which tracks thermostat adjustments as "failures"



Measure

Data Stratification Tree

Questions about process
Does sunlight affect house electric usage?
Who is the biggest user of electricity, when home?
What is the ideal thermostat setting (temperature @ hour)?
Is energy use predominantly the fault of the AC system (internal/external temp mismatches)?
What kWh values should we expect to know when we have made a significant change?

kWh

output

Stratification Factors	**Measurements**	**Samples**
External Temperature	Temperature in Fahrenheit by hour as measured by weather services **(continuous)**	For one month: 24*30=720 (per month)
Thermostat Temperature	Fahrenheit values from thermostat's schedule when changed, by hour **(continuous)**	7 sets. Interpolated per hour for 720
Who is Home (tenant presence data)	Presence of any of 4 tenants by hour, measured by presence-detecting tool (phone network monitoring) **(discrete)**	~40k samples, down-sampled to 720
External Cloud/Sunlight Conditions	1-20 rating of sunlight intensity from cloudy/raining (1) to sunny (20) by hour from weather services **(discrete)**	720
Historic energy measurements	kWh/day averages for each month **(continuous)**	12

Samples were collected every hour (or summed/interpolated/down-sampled to fit into an every-hour sampling) to facilitate easy comparison across time. Due to the cyclical nature of what is suspected to be the main contributor to energy use (the AC system) happening in 24 hour cycles, a large enough number of samples across each part of the day was required for regression analysis using the tenant presence data

Measure - Sample Sizes and Errors)

n = 577 samples required for us to be 95% confident about our estimates +/- 1 degree

n = ((z* s) / E)^2   z*=1.960, s=12.25, E=1

Samples collected for these operations was about 720

Fewer samples means less certainty about our results, should we choose to use the descriptive statistics to make other statements about our data, or use this data in a hypothesis test!

Data was collected from:

- Temperature and Cloud/Sunshine data – Manually scraped from WeatherUnderground.com
- Tenant Presence Data – Pulled from the "who-is-home" device that monitors cellphone presence on the house network
- Power usage was pulled from Dominion Power's website. A request was submitted prior to this assignment for them to log power data every 30 minutes
- Additional historic kWh/day averages for each month are available from Dominion Power. These were scraped by hand
- Thermostat information was recorded manually every time an adjustment was made by a tenant

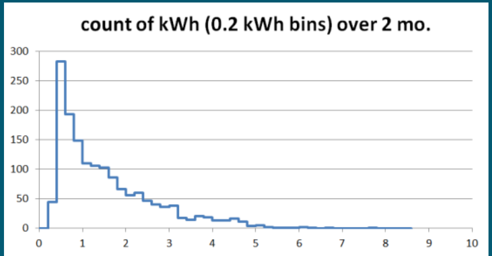
Measurement error can be seen in the tenant presence data, which depends on every tenant leaving their phone's WiFi connection enabled. This is not always the case. Likewise thermostat adjustments rely on manual record keeping and is prone to record-keeping errors. Temperature data is pulled from a 3rd party service and only represents status of the general area, not necessarily the weather at the house.

Additionally, the **"Hawthorne effect"** may be present in the thermostat temperature data, as this is adjusted by hand, the tenant who was recording change information knows he is being observed, and that may affect the ranges and frequency that they adjust the thermostat to!

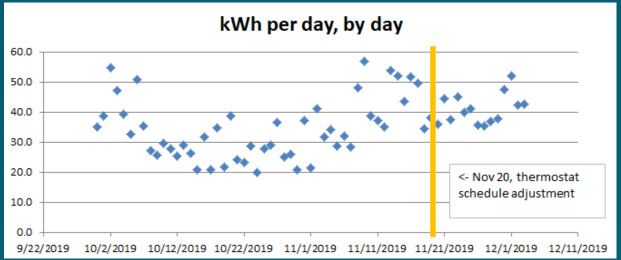
In general, these measurement errors can be addressed by increasing sample sizes!

Measure - Data Distribution

kWh values seem to show a heavily skewed distribution. This histogram shows that the data is perhaps in a Poisson distribution



When scatter-plotted by date, there is evidence of a non-linear relationship, which could be an effect of the seasons

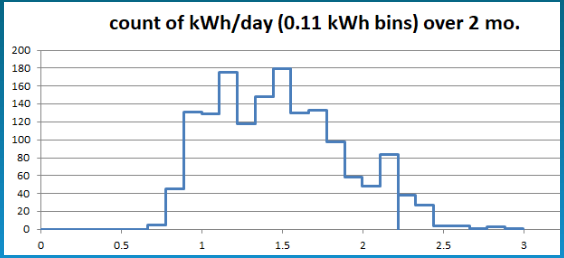


Data for kWh/day is approximately normally distributed. This histogram is the sum of all kWh measurements for a 24h period

Descriptive statistics for kWh, temperatures, sunlight amounts, and hourly themostat settings

pre-nov-20-stats	kWh/h	Temp (F)	sun-amount	Thermostat (F)
Mean	1.431	56.89	14.67	65.82
Median	1.100	58.00	13	68
Mode	0.400	62.00	11	68
Standard Deviation	1.114	12.45	4.64	3.58
Range	8.400	97.00	20	10
n=	1248.000			

post-nov-20-stats	kWh/h	Temp (F)	sun-amount	Thermostat (F)
Mean	1.725	45.63	13.62	66.08
Median	1.500	45	13	68
Mode	0.400	41	11	68
Standard Deviation	1.220	5.38	4.35	3.87
Range	7.500	26	20	10
n=	312.000			



Analyze – Biggest Contributors

A **multiple linear regression analysis** was performed upon the kWh data, tenant presence data, and two types of weather data (sunlight/cloudiness amounts, and outside temperature)

With a p-value of <0.05, **we can see that tenant 2 and the outside temperature have an effect on energy use.**

	Coefficients	*Std. Err.*	*P-value*
Intercept	2.1605	0.1971	0.00000
tenant 1	0.1296	0.0766	0.09099
tenant 2	-0.1684	0.0674	**0.01262**
tenant 3	0.0712	0.0786	0.36518
tenant 4	0.0683	0.0851	0.42223
temp	-0.0107	0.0026	**0.00004**
cloudiness	-0.0099	0.0063	0.11251

More investigation was warranted. As it turns out, Tenant 2 is the person that lives in the highest room on the third floor. This room has its own AC unit, which is turned on manually. **This separate A/C unit can be seen reflected in the data!**

Analyze - Livability

Energy use in households is predominantly used to heat and cool the house.* With this in mind, shutting off the HVAC would undoubtedly save energy (and therefore, money) but would make things unlivable. **How do we balance budget and livability?**

- **Manual thermostat adjustments** were recorded and used as evidence of an **unfavorable** temperature configuration. This was used to determine SQL values.
- On Nov 20th, updates were made to the thermostat schedule, which resulted in a process that had fewer manual adjustments

Are these thermostat changes saving energy, even though energy use has appeared to increase?

A single-upper/right-tailed Hypothesis test for continuous, two-sample data was carried out to answer this question:
H0: energy use is the same or less (=<34.34 kW/day)
Ha: energy use has increased (>34.34 kW/day)
s and n are taken from the tables to the right
a = 0.05

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Upper/right-tail

p = area right of Z or t

z = -3.95
p = 0.9999
Since p > a, we cannot reject H0

Pre-Schedule-Adjustment	Post-Schedule-Adjustment
SQL Value: 2.1	**SQL Value: 3.2**

pre nov-20 thermostat adjustments	
kWh/day	
Mean	34.34038462
Standard Error	1.374529289
Median	33.45
Mode	29.1
Standard Deviation	9.911871662
Sample Variance	98.24519985
Kurtosis	-0.408454332
Skewness	0.634962308
Range	36.9
Minimum	20
Maximum	56.9
Sum	1785.7
Count	52
Confidence Level(95.0%)	2.759482692

post nov-20 thermostat adjustments	
kWh/day	
Mean	41.4
Standard Error	1.364194136
Median	41.1
Mode	#N/A
Standard Deviation	4.918671907
Sample Variance	24.19333333
Kurtosis	0.161446019
Skewness	0.745350751
Range	16.6
Minimum	35.4
Maximum	52
Sum	538.2
Count	13
Confidence Level(95.0%)	2.972323686

SQL 2.1 is less than 3.2, so: **PROCESS QUALITY IMPROVED**
But… H0 is not rejected, so: **ENERGY USE HAS NOT DECREASED**

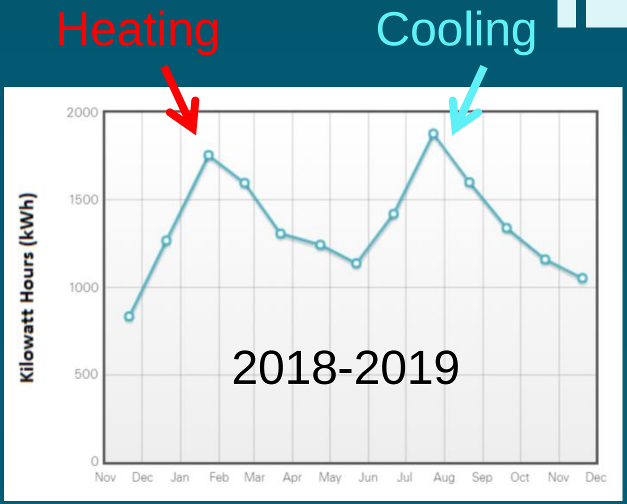
* Direct Energy - What Uses the Most Electricity in My Home?
https://www.directenergy.com/learning-center/energy-efficiency/what-uses-most-electricity-in-my-home

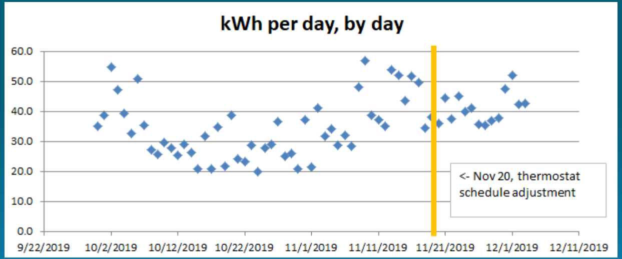
Analyze – Seasonality

There appears to be heavy variation by time of year

Insufficient data exists at the granularity measured at for the year prior. Data from the year prior would be useful, so as to rule out seasonal effects more easily. Perhaps the data currently available can be normalized using the month-by-month data available from the power company?

The effects of the seasons can even be seen in a plotting of ~two months of data, as seen on the right. This is the trend going from October into December. Note the trend upward.





Seasonal switch from air conditioning to heating, causing a trough to be seen in kWh electric usage

Improve

Livability improvements SQL proved to be a successful measuring device to maintain a quality of living while making adjustments to other parts of the system.

The data did highlight the non-negligible effect that the secondary AC system has. This AC unit will need to be included in the considerations for later. It is a manual unit. Perhaps getting a thermostat or timed model would prevent overuse?

To improve this mostly-automated process, thermostat schedule adjustments were made. As the hypothesis test in the Analyze portion of the DMAIC process has shown, <u>**this was not sufficient to lower my energy expenditures**</u>.

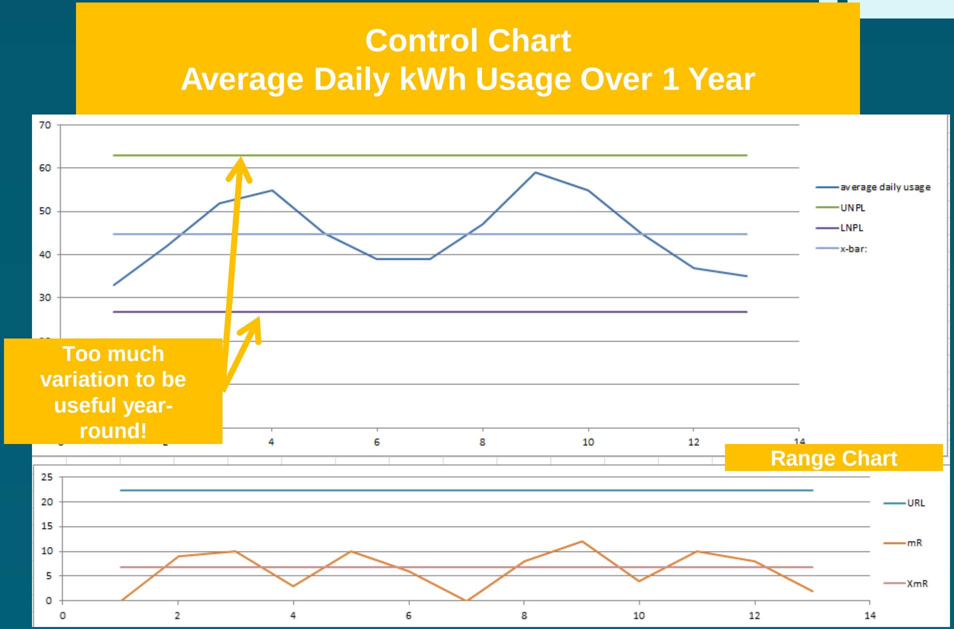
However, due to the seasonal effects on this system, <u>**I do not have the necessary past-year data to verify that my changes did in fact improve things**</u>. It could be that my changes improved efficiencies, but not enough to show through the seasonal shift that always happens this time of year, which drives energy use up.

Control

An additional data-set was available to use for **historical** daily averages, courtesy of the power company. An attempt was made to use this data for the basis of a control chart for the average daily kWh in the future. However, a naïve linear control chart has bounds that are too permissive to be of much use year-round

Because of the large seasonal variations, this chart is insufficient

A control chart that accounts for seasonal variation may be able to alert us to information beyond the noise. Such events could signal process failures or process successes



Control – Potential Seasonal Tool?

This version of a control chart takes the **seasonal variations** into account. The boundaries (UNPL, LNL) for the chart can be tightened by averaging fewer months of data.

This is accomplished using a **modified exponential smoothing** formula, which is offset by 12 months

Now that the process has been adjusted and a new baseline of data is present. **We can use control charts as a guide to see if our process is changing**

Meter Read Date	Days	Usage	Average Daily Usage	UNPL	LNL	URL	mR	x-bar	XmR
11/20/2018	25	834	33.0	0.0	0	0			
12/20/2018	30	1268	42.0	0.0	0	0	9.0		
1/23/2019	34	1757	52.0	0.0	0	0	10.0		
2/21/2019	29	1598	55.0	73.0	33.99333	23.98	3.0	53.5	[illegible]
3/22/2019	29	1308	45.0	70.4	29.60667	25.07	10.0	[illegible]	7.7
4/23/2019	32	1244	39.0	58.8	25.15333	20.71	6.0	[illegible]	6.3
5/22/2019	29	1138	39.0	53.2	24.81333	17.44	[illegible]	39.0	5.3
6/21/2019	30	1421	47.0	55.4	30.58667	15.26	8.0	43.0	4.7
7/23/2019	32	1880	59.0	70.7	35.26667	[illegible]	12.0	53.0	6.7
8/21/2019	29	1602	55.0	78.3	35.72	26.16	4.0	57.0	8.0
9/20/2019	30	1340	45.0	73.1	[illegible]	28.34	10.0	50.0	8.7
10/21/2019	31	1160	37.0	60.5	[illegible]	23.98	8.0	41.0	7.3
11/20/2019	30	1053	35.0	[illegible]	18.26667	21.8	2.0	36.0	=AVERAGE
12/20/2019			40.6	[illegible]	23.70133	17.004	5.6	[illegible]	5.2
1/19/2020			49.7	56.6	26.94827	18.2248	9.1	[illegible]	5.6
2/18/2020			53.9	64.9	31.29099	20.64896	4.2	48.1	[illegible]
3/19/2020			46.8	68.3	31.97498	22.34413	7.2	50.2	[illegible]
4/18/2020			40.6	62.7	31.48244	19.19516	6.2	47.1	5.9
5/18/2020			39.3	55.2	29.24527	15.94937	1.2	42.2	4.9
6/17/2020			45.5	53.9	29.69371	14.85453	6.2	41.8	4.5
7/17/2020			56.3	63.2	30.86073	19.86756	10.8	47.0	6.1
8/16/2020			55.3	68.3	36.36456	19.63622	1.0	52.3	6.0
9/15/2020			47.1	70.7	35.07134	21.8773	8.2	52.9	6.7
10/15/2020			39.0	62.4	31.78336	18.83751	8.0	47.1	5.8
11/14/2020			35.8	57.9	23.37001	21.2075	3.2	40.6	6.5

Exponential Smoothing with 12-month offset for past data. Predicts next year with seasonality considered. Yellow values are predicted

XmR (and a few other) columns only average values from past 3 months, instead of all past values

Seasonal Control Chart
Average Daily kWh Usage Over 1 Year