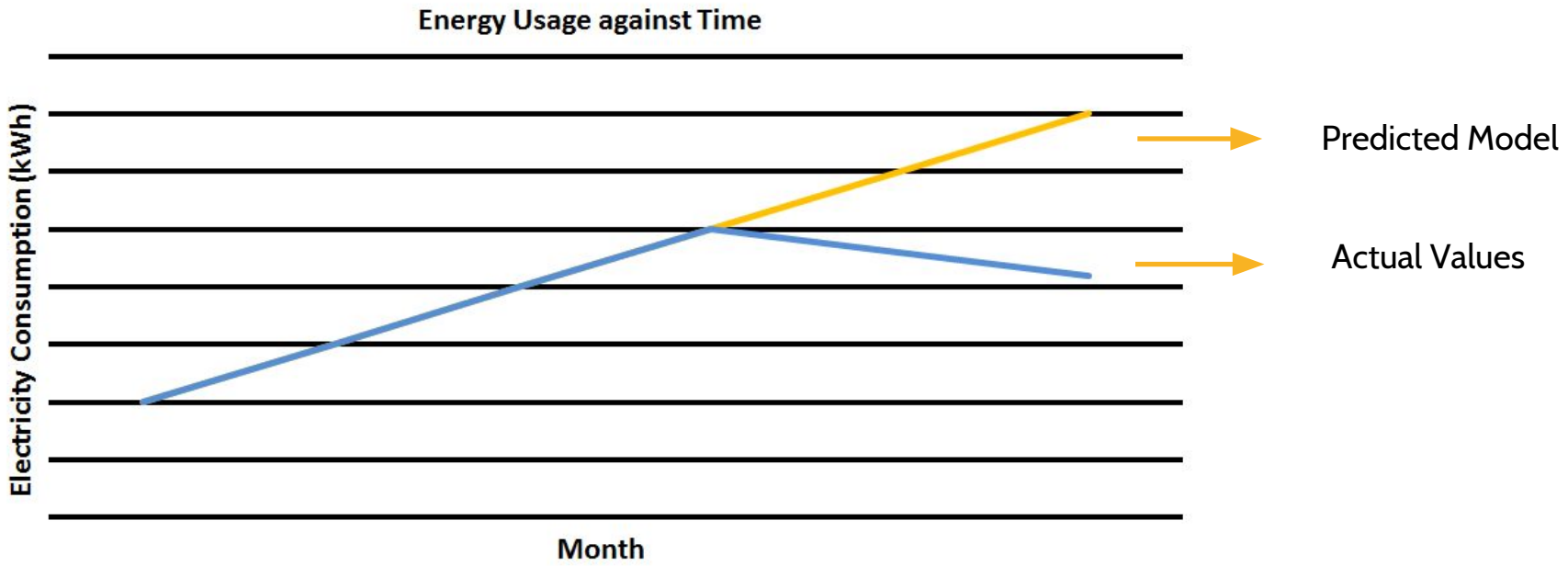
FORECASTING ELECTRICAL CONSUMPTION OF COMMERCIAL BUILDINGS

JISU LEE, ALEXIUS KOK, LIU SIDIAN, DANIEL TONG, WANG YIRAN

OUTLINE

- Project Description
- Project Deliverables
- Methodology
- Data Sets
- Linear Regression Models & Analysis
- Limitations
- Recommendations

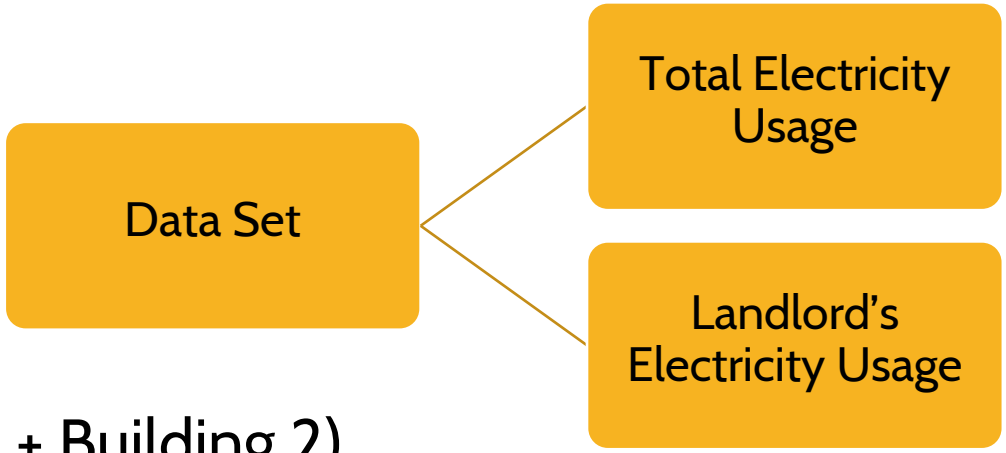
PROJECT DESCRIPTION



- Analyze the properties of different commercial buildings
- Model the monthly energy consumption
- Calculate the energy savings

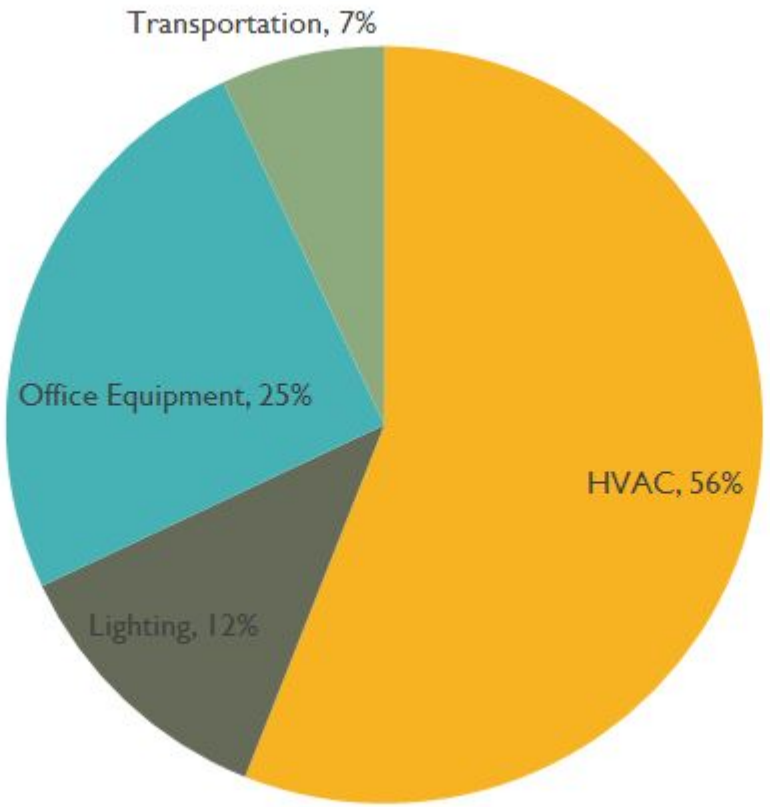
PROJECT DELIVERABLES

- Six Regression Models
  - Building 1
  - Building 2
  - Combined Data (Building 1 + Building 2)



- Requirements
  - Logical Linear Regression Equations
  - Meet the Significance Test Thresholds

METHODOLOGY



FACTORS OF ENERGY CONSUMPTION (COMMERCIAL BUILDINGS)

INDEPENDENT VARIABLES

Variables	Justification
Rainfall	↑ Humidity ↓ Temperature ↑ Energy Usage
Humidity	↑ Humidity ↑ Dehumidification ↑ Energy Usage
Temperature	↑ Temperature ↑ Cooling Energy
Cooling Degree Days	↑ CDD ↑ Cooling Energy
Time Series	Capture Trend
Time Series Squared	Compensate for large contribution as time ↑
Monthly Variables	Capture Seasonality

INDEPENDENT VARIABLES

Variables	Justification
Occupancy	↑ More Area to Cool ↑ Energy Usage
Human Traffic	↑ Human Traffic ↑ Heat ↑ Cooling Energy
Opening Hours	↑ Opening Hours ↑ Energy Usage
Building Type	↑ Electricity Needs ↑ Energy Usage
Gross Floor Area	↑ Gross Floor Area ↑ Energy Usage

Linear Regression Modelling

Public Source Data
Average Monthly Temperature
Average Monthly Humidity
Average Monthly Cooling Degree Days
Average Monthly Rainfall

Provided Data
Gross Floor Area
Net Lettable Area
Human Traffic
Occupancy Rate
Opening Hours

Correlation Analysis
Pearson Product Moment Correlation Analysis (Linear Assumption)

Assumptions
Choosing Independent Variables
Linearity of Variables
Independence of Data
Incomplete Data Set

Trend & Seasonality
Time Series Variable
11 Binary Variables for Months

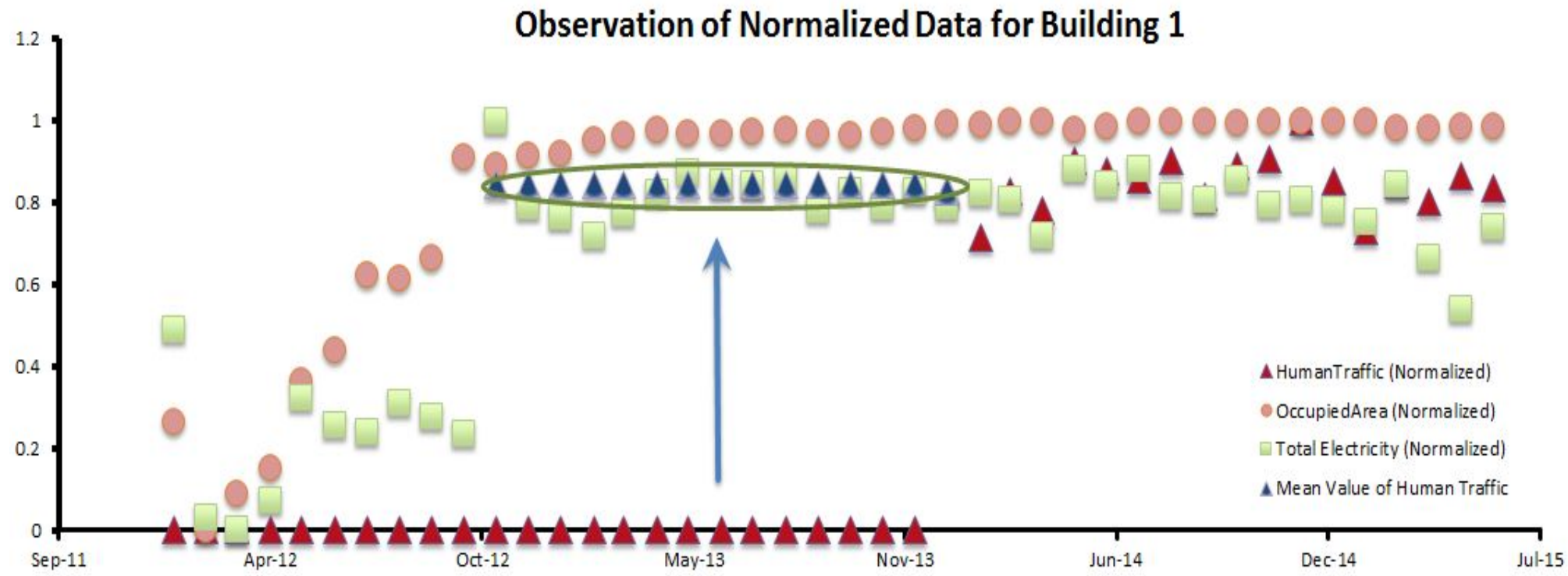
Artificial Variables
Testing Variables
Binary Variable for Buildings

Improve the Model
Transformation of Variables
Substitution of Variables
Addition of Variables

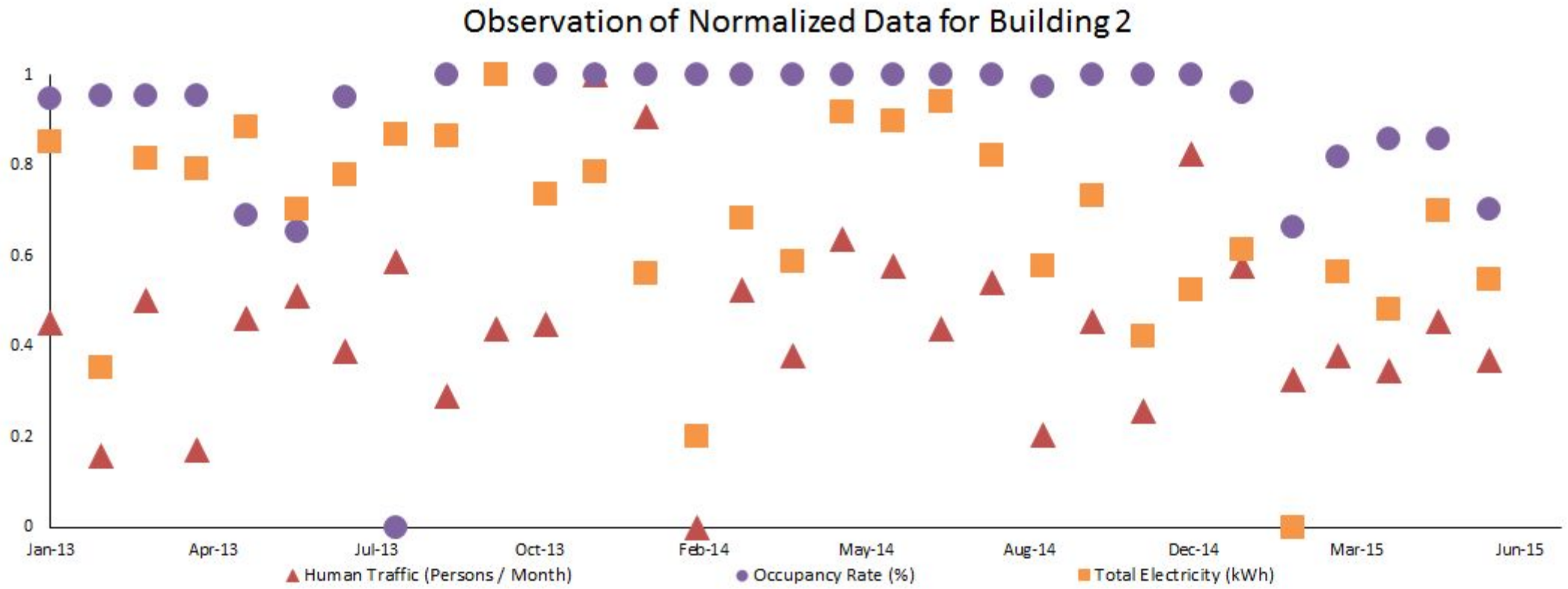
Final Regression Model

R Square Value greater than 0.75
Pass Significance Tests (P Value, T Stat, Significance F)
Coefficients of variables match hypothesis

DATA SET: BUILDING ONE



DATA SET: BUILDING TWO



CORRELATION BETWEEN VARIABLES

Correlation Coefficient	Occupied Area	Monthly Temperature	Rainfall	Average Humidity	Cooling Degree Days	Occupancy Rate	Human Traffic
Occupied Area	1.000	0.108	-0.015	-0.232	0.107	0.070	0.297
Monthly Temperature		1.000	-0.490	-0.530	0.865	0.038	0.287
Rainfall			1.000	0.669	-0.434	0.027	-0.225
Average Humidity				1.000	-0.509	-0.284	-0.473
Cooling Degree Days					1.000	-0.005	0.189
Occupancy Rate						1.000	0.598
Human Traffic							1.000

JUSTIFICATION OF FINAL REGRESSION MODEL

- Select one of two correlated variables
  - Higher R Square
  - More significant variables

- Variables: Monthly Binary, Time Series, Time Series Squared
  - Always kept to capture the seasonality and trend

- Insignificant Variables
  - Justify based on 3 conditions

LINEAR REGRESSION RESULTS

Coefficients of Variables	Building 1 + Building 2		Building 1		Building 2	
	Total	Landlord	Total	Landlord	Total	Landlord
Intercept	419,000	88,740	-57,040	-163,800	1,069,000	682,600
Occupied Area	10.65*	17.17**	6.96	14.33*	8.37	-
Human Traffic	0.22**	0.10**	0.18**	0.07*	-	0.06
Rainfall	7,275	4,678	8,827	-	-	-
Humidity	-	-	-	-	267900*	222,100
Time Series	20600**	7,468	37250**	19320*	2,197	-5,351
Time Series^2	-430.40**	-184.60	-725.90**	-407.50*	-151.70	42.64
Testing	408400**	420500**	468700**	461400**	-	-
Binary Building	-446900**	-268900**	-	-	-	-
January	27,270	27,620	65,910	31,160	10,020	8,741
February	-14,010	-17,570	64,020	28,390	-123,600	-95700**
March	35,620	76770*	64,930	55,430	23,230	94620**
May	48,840	70390*	57,920	68,980	60,470	76740*
June	66,330	56250*	102,100	59,370	63,080	56,560
July	54,440	73550*	68,250	71,770	57,190	85850*
August	37,070	62,170	50,650	69,520	62,190	58740*
September	15,470	25,840	11,660	27,940	25,120	32,390
October	24,110	34,690	11,440	28,830	70,900	50,510
November	-52,640	-12,290	-52,830	-8,950	-37,850	2,983
December	-39,040	7,569	11,890	64,450	-15,850	-25,270
R Square	0.95	0.82	0.93	0.87	0.81	0.75

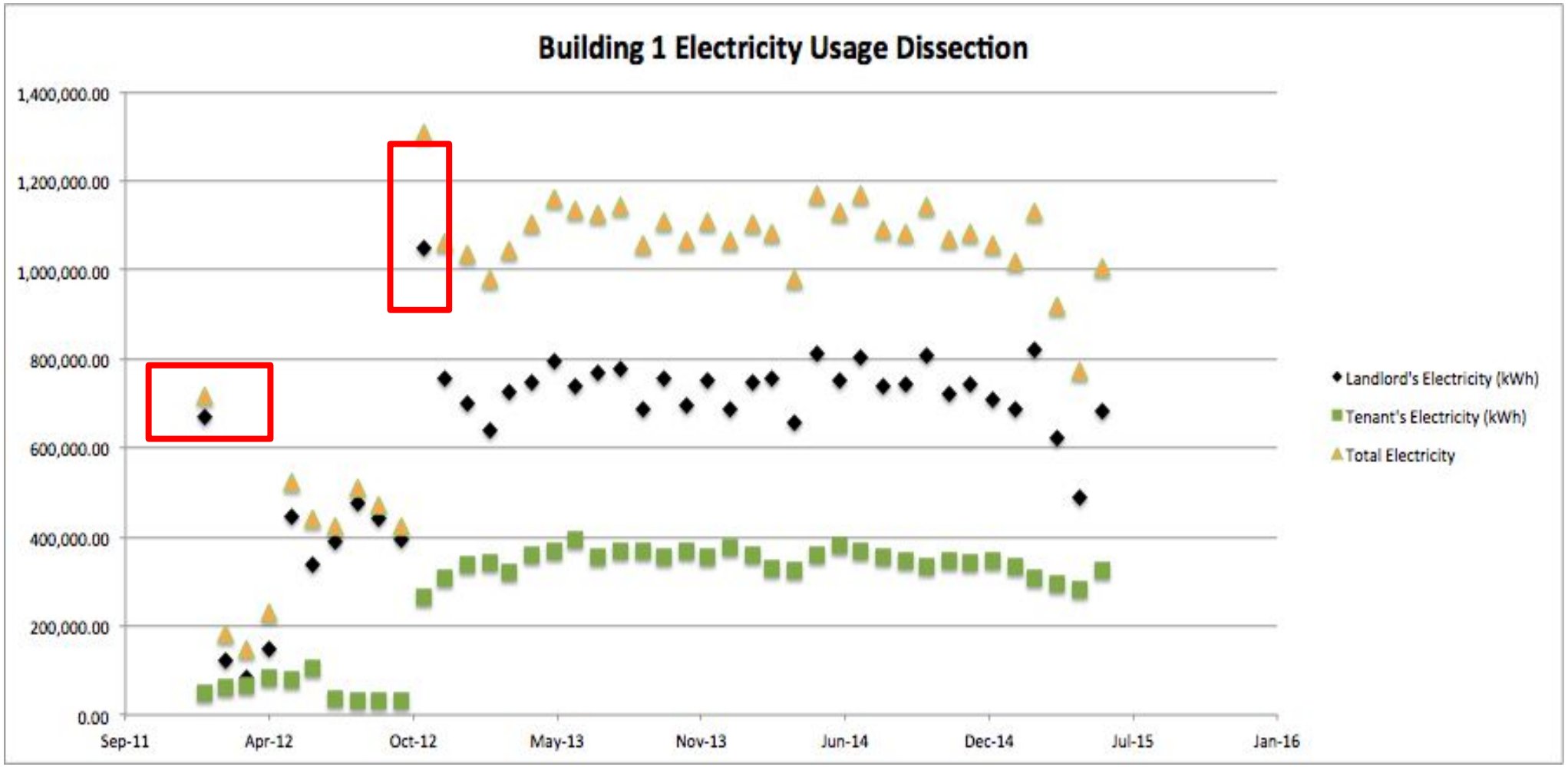
** Strongly Significant

* Significant

ANALYSIS

Coefficients of Variables	Building 1 + Building 2		Building 1		Building 2	
	Total	Landlord	Total	Landlord	Total	Landlord
Intercept	419,000	88,740	-57,040	-163,800	1,069,000	682,600
Occupied Area	10.65*	17.17**	6.96	14.33*	8.37	-
Human Traffic	0.22**	0.10**	0.18**	0.07*	-	0.06
Rainfall	7,275	4,678	8,827	-	-	-
Humidity	-	-	-	-	267900*	222,100
Time Series	20600**	7,468	37250**	19320*	2,197	-5,351
Time Series^2	-430.40**	-184.60	-725.90**	-407.50*	-151.70	42.64
Testing	408400**	420500**	468700**	461400**	-	-
Binary Building	-446900**	-268900**	-	-	-	-
January	27,270	27,620	65,910	31,160	10,020	8,741
February	-14,010	-17,570	64,020	28,390	-123,600	-95700**
March	35,620	76770*	64,930	55,430	23,230	94620**
May	48,840	70390*	57,920	68,980	60,470	76740*
June	66,330	56250*	102,100	59,370	63,080	56,560
July	54,440	73550*	68,250	71,770	57,190	85850*
August	37,070	62,170	50,650	69,520	62,190	58740*
September	15,470	25,840	11,660	27,940	25,120	32,390
October	24,110	34,690	11,440	28,830	70,900	50,510
November	-52,640	-12,290	-52,830	-8,950	-37,850	2,983
December	-39,040	7,569	11,890	64,450	-15,850	-25,270
R Square	0.95	0.82	0.93	0.87	0.81	0.75

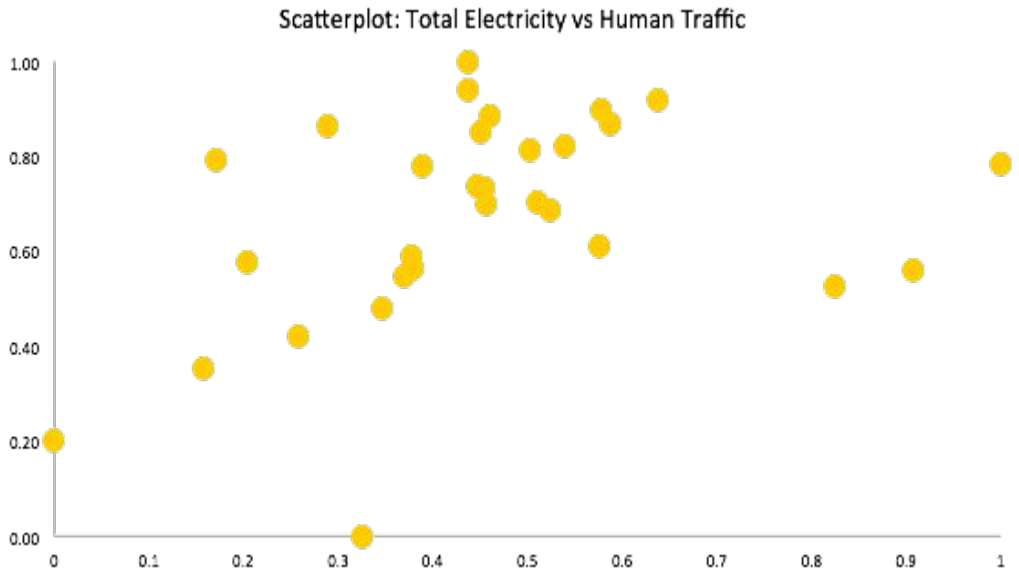
ANALYSIS

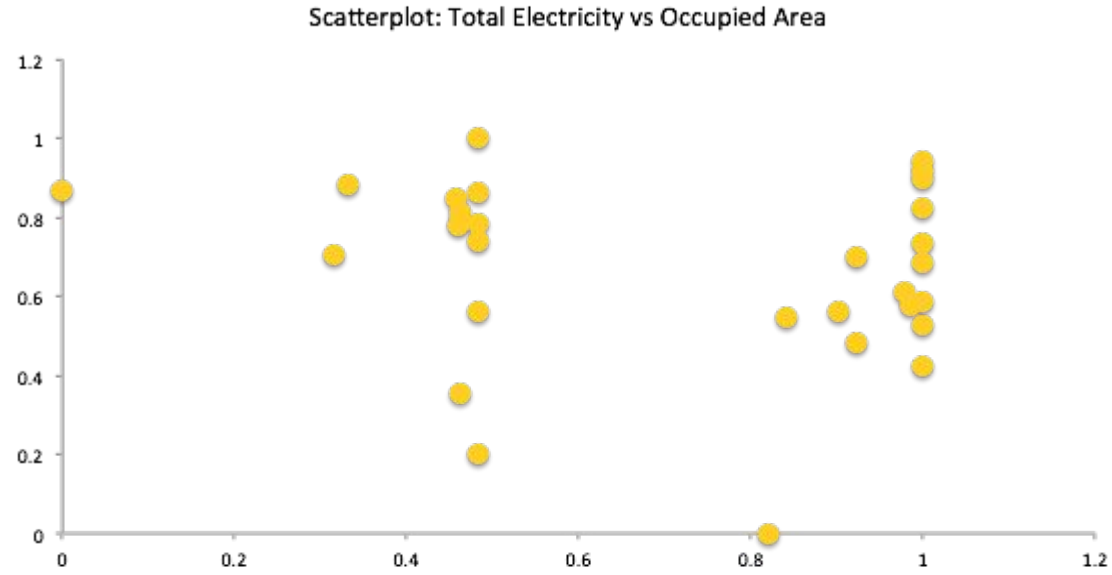

Building 1 Electricity Usage Dissection
1,400,000.00
1,200,000.00
1,000,000.00
800,000.00
600,000.00
400,000.00
200,000.00
0.00
Sep-11
Apr-12
Oct-12
May-13
Nov-13
Jun-14
Dec-14
Jul-15
Jan-16
Landlord's Electricity (kWh)
Tenant's Electricity (kWh)
Total Electricity

Analysis

Coefficients of Variables	Building 1 + Building 2		Building 1		Building 2	
	Total	Landlord	Total	Landlord	Total	Landlord
Intercept	419,000	88,740	-57,040	-163,800	1,069,000	682,600
Occupied Area	10.65*	17.17**	6.96	14.33*	8.37	-
Human Traffic	0.22**	0.10**	0.18**	0.07*	-	0.06
Rainfall	7,275	4,678	8,827	-	-	-
Humidity	-	-	-	-	267900*	222,100
Time Series	20600**	7,468	37250**	19320*	2,197	-5,351
Time Series^2	-430.40**	-184.60	-725.90**	-407.50*	-151.70	42.64
Testing	408400**	420500**	468700**	461400**	-	-
Binary Building	-446900**	-268900**	-	-	-	-
January	27,270	27,620	65,910	31,160	10,020	8,741
February	-14,010	-17,570	64,020	28,390	-123,600	-95700**
March	35,620	76770*	64,930	55,430	23,230	94620**
May	48,840	70390*	57,920	68,980	60,470	76740*
June	66,330	56250*	102,100	59,370	63,080	56,560
July	54,440	73550*	68,250	71,770	57,190	85850*
August	37,070	62,170	50,650	69,520	62,190	58740*
September	15,470	25,840	11,660	27,940	25,120	32,390
October	24,110	34,690	11,440	28,830	70,900	50,510
November	-52,640	-12,290	-52,830	-8,950	-37,850	2,983
December	-39,040	7,569	11,890	64,450	-15,850	-25,270
R Square	0.95	0.82	0.93	0.87	0.81	0.75

Analysis


Scatterplot: Total Electricity vs Human Traffic
1.00
0.80
0.60
0.40
0.20
0.00
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Scatterplot: Total Electricity vs Occupied Area
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2

Limitations

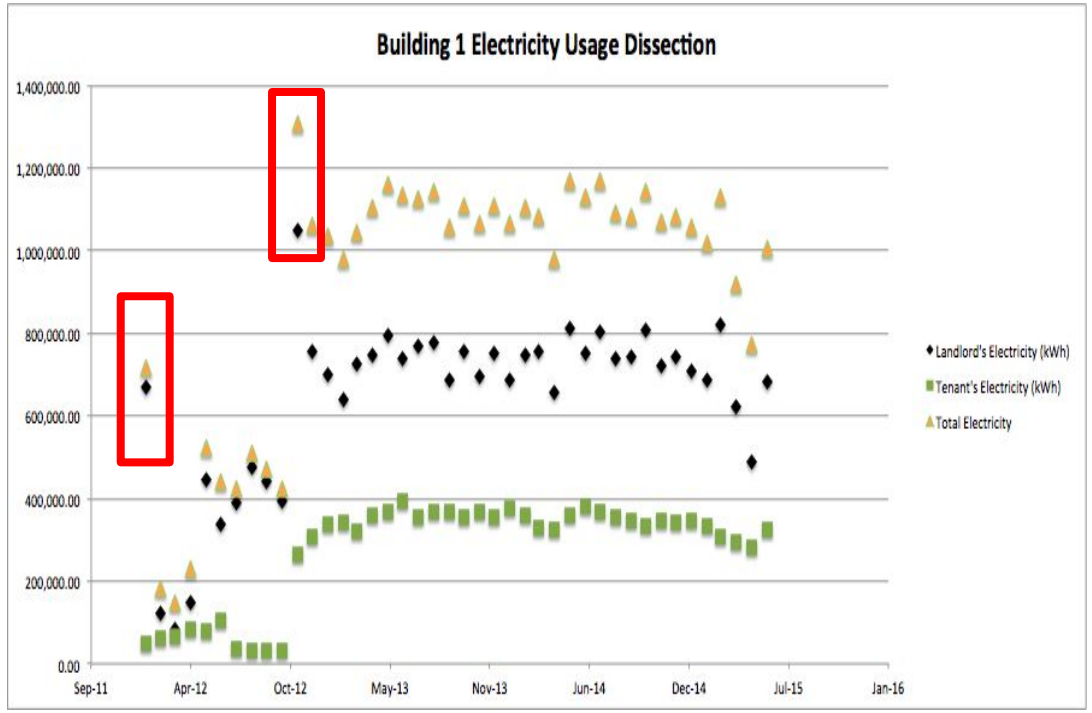
Assumptions for any Multiple Regression

Cases : Predictors Ratio (i.e. Sample Size)	10 to 20 cases for every independent variable
Normality (i.e. Approx. normally distributed)	Stem-and-leaf plots, Q-Q Plot, Shapiro-Wilk test
Outliers	Graphical Analysis
Multicollinearity	Correlation Analysis
Linearity of Variables with respect to Dependent Variable	Scatter Plot
Normality, Linearity and Homoscedasticity of Residuals	P-P Plot, Scatter Plot

Limitations

Assumptions that are not violated

Outliers



Multicollinearity

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TotalElectricity	.239	72	.000	.776	72	.000
Occupied_Area	.257	72	.000	.789	72	.000
AverageHumidity	.122	72	.010	.958	72	.017
HumanTraffic	.302	72	.000	.743	72	.000
Rainfall	.099	72	.077	.959	72	.018

a. Lilliefors Significance Correction

Shapiro-Wilk Test; Sig. < 0.05

Therefore, all Variables are not Normally Distributed

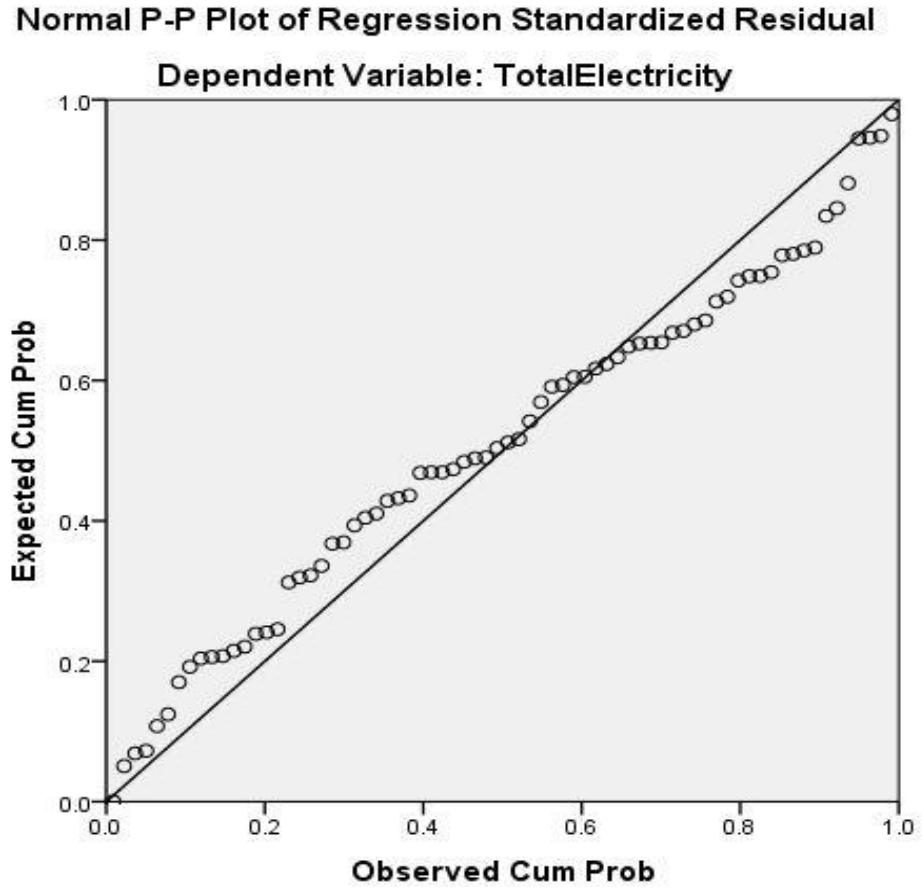
Spearman's Rank-Order Correlation Used

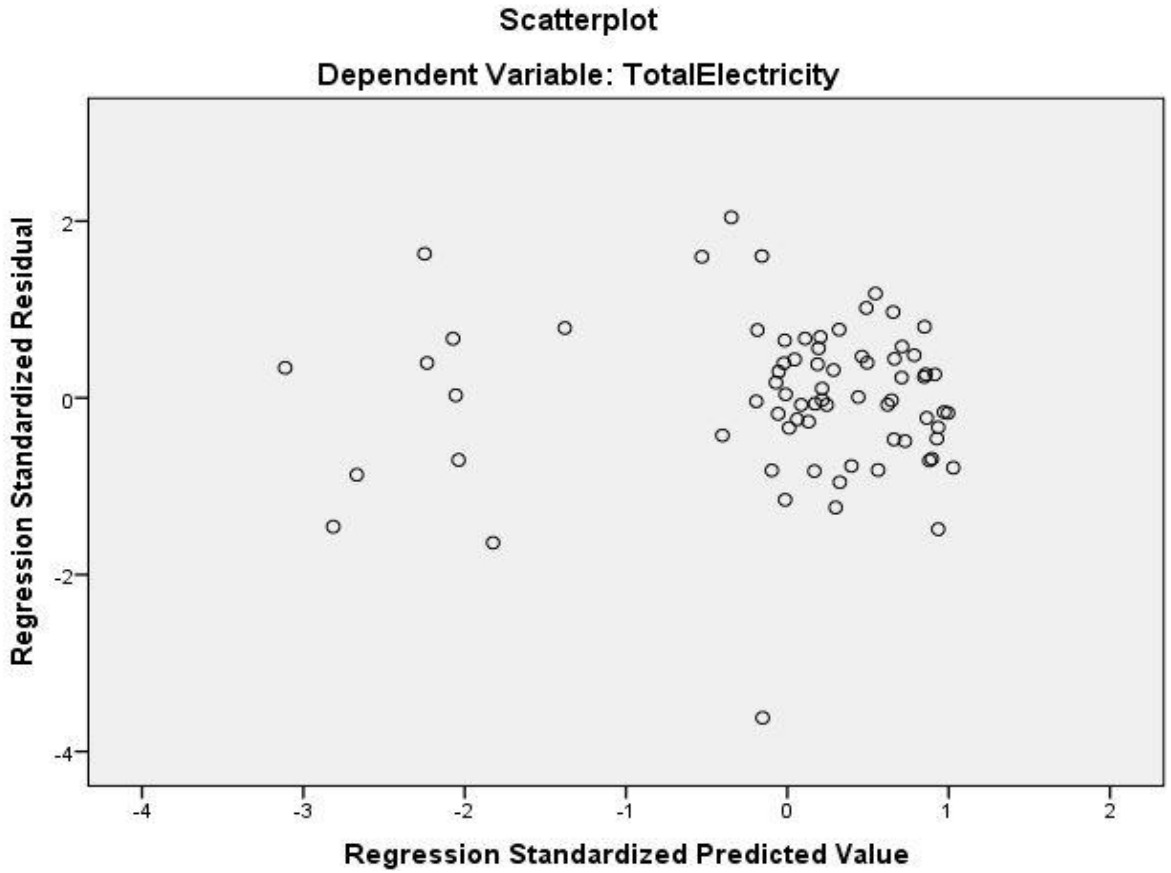
No Significant difference with Pearson Product-Moment Correlation

Limitations

Assumptions that are not violated

Normality, Linearity and homoscedasticity of Residual





Limitations

Assumptions that are violated

Sample Size

10~20 Cases for every Independent Variable

17 Independent Variable : 170 - 340 Cases

Given: 72 Cases Combined

Normality

```
Occupied_Area Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

    33.00        1 .  566777777777777789999999999999999
     3.00        2 .  023
     3.00        2 .  778
    22.00        3 .  2233344444444444444444
    11.00        3 .  55556666666

 Stem width:  10000.00
 Each leaf:       1 case(s)
```

Stem-and-leaf plot shows that non-normality is an issue, hence the coefficient of the linear equation could be inaccurate

Recommendations

- Removal of the two violated assumptions
  - Collect more data to meet the minimum required sample size
  - Perform Variable Transformation to create Normality of Variables with respect to Dependent Variable
- Calculate the subsequent month's electrical consumption
  - Compare the calculated value with the real value
- Reliability Analysis
  - To measure how long the calculated values will hold over time

CONCLUSION

- Six scenarios, Six assumptions, Six equations
- Both combined data set and Building 1 data set shows promising results
  - Example: Jun 15 data point was removed and regression was performed
    - Calculated Value : Total Electricity Usage (kWh) = 966,XXX
    - Observed Value : Total Electricity Usage (kWh) = 1,005,XXX
  - Observed value is within the prediction interval (±2 x 68,XXX)
- Given the questionable coefficients of Building 2
  - Linear Equation might be erratic
  - More work to be done in accordance with the recommendations as stated

Q&A