Richard Swann | 823902440
Jason Cao | 200107583
Matt Fung | 301329703

CMPT318 Assignment 1

Q1.

Feature	Arithmetic Mean	Geometric Mean	Median	Mode	Standard Deviation
Feature A: Global Active Power	0.771729	0.5530869	0.489075	0.122	0.7072015
Feature B: Global Reactive Power	0.1344998	0	0.112	0	0.1172652
Feature C: Voltage	239.6841	239.6771	239.64	239.5	1.828296

Table 1. Calculated Values for Features A - C For Week 29

Time Window	Hours	Minimum	Maximum
Weekday Day Hours	07:30 - 17:00	0.114	3.976
Weekday Night Hours	17:01 - 07:29	0.112	5.686
Weekend Day Hours	07:30 - 17:00	0.12	6.596
Weekend Night Hours	17:01 - 07:29	0.11	5.75

Table 2. Min and Max of Feature A: Global Active Power For Week 29

Time Window	Hours	Minimum	Maximum
Weekday Day Hours	07:30 - 17:00	0	0.81
Weekday Night Hours	17:01 - 07:29	0	0.712
Weekend Day Hours	07:30 - 17:00	0	0.566
Weekend Night Hours	17:01 - 07:29	0	0.642

Table 3. Min and Max of Feature B: Global Reactive Power For Week 29

Comments:

We separated the 29th week of the year, segregated the data into weekdays and weekends, then finally differentiated into day (07:30 AM - 17:00 PM) and night (17:01 PM - 07:29 AM) inclusive.

Q2.

	Global active power	Global reactive power	Voltage	Global intensity	Sub metering 1	Sub metering 2	Sub metering 3
Global active power	1.000	-0.008	-0.095	0.034	-0.041	-0.005	0.066
Global reactive power	-0.008	1.000	-0.060	0.373	0.157	0.188	0.067
Voltage	-0.095	-0.060	1.000	-0.470	-0.279	-0.181	-0.379
Global intensity	0.034	0.373	-0.470	1.000	0.555	0.474	0.616
Sub metering 1	-0.041	0.157	-0.279	0.555	1.000	0.113	0.057
Sub metering 2	-0.005	0.188	-0.181	0.474	0.113	1.000	0.052
Sub metering 3	0.066	0.067	-0.379	0.616	0.057	0.052	1.000

Table 4. Correlation Matrix of Variables For Week 29

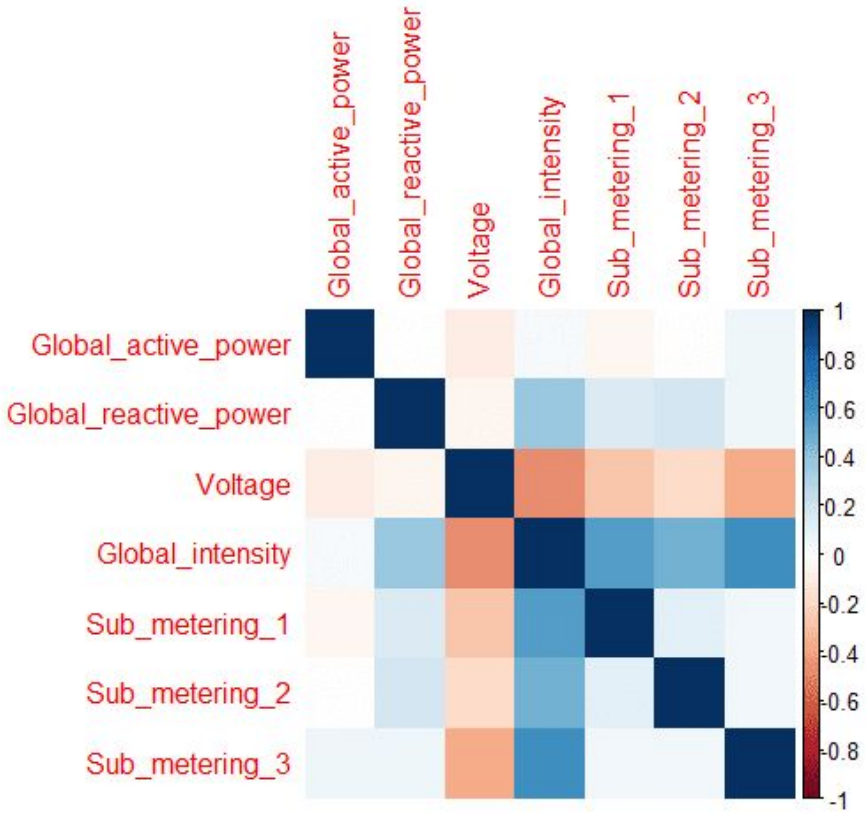

Figure 1. Correlation Matrix[2] Depicting the Correlation For Each Disjoint Pair of Features A - G For Week 29

Q3.

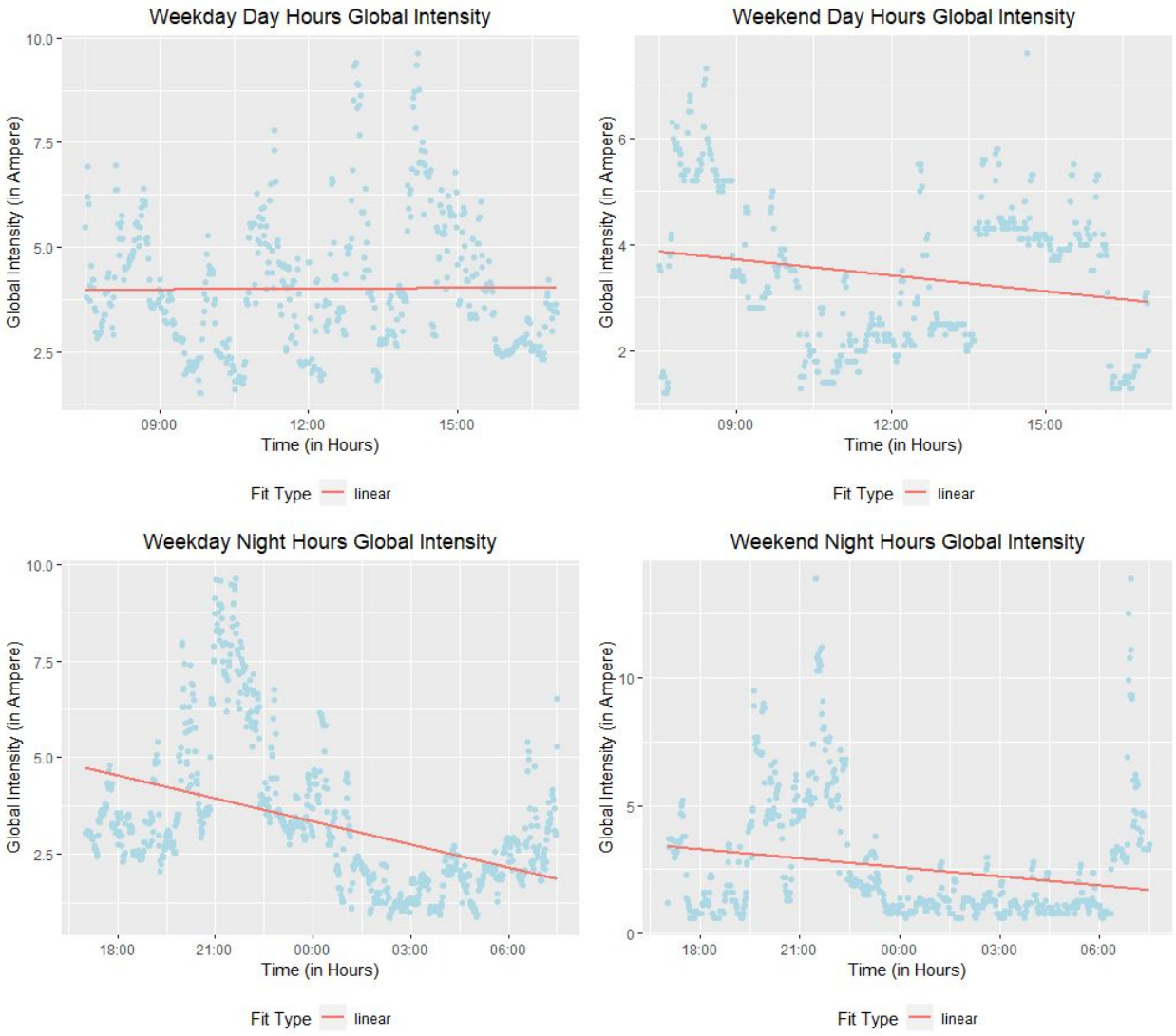

Figure 2. Diagram Depicting the Linear Regression Lines for Global Intensity of Four Time Representative Windows For Week 29

Comments:

We felt that the linear regression lines were not helpful for undermining our data.

Figure 3. Diagram Depicting the Polynomial Regression Lines for Global Intensity of Four Time Representative Windows For Week 29

Comments:

We chose two different degrees of freedom when we plotted the polynomial regression lines for each time representative window. We started with 3 degrees of freedom to depict the general trend of the data in each time representative plot. We noticed that this was not sufficient enough to adequately model the trend for each plot. As a result, we continued to increase the degrees of freedom until we discovered that the plots could no longer be optimized. Hence, we ended up using 10 degrees of freedom.

Summary

From Figure 1, there are several disjoint pairs that show evidence of correlation. There is a moderate negative correlation between global intensity and voltage. There is also a moderate negative correlation between sub metering 3 and voltage, but this correlation is weaker than the correlation between global intensity and voltage. (Former has a strength of ~0.3 and latter has a strength of ~0.4 - 0.5) Similarly, there is a moderate positive correlation between sub metering 3 and global intensity. Another moderate positive correlation exists between sub metering 1 and global intensity. However, this correlation is weaker than the correlation between sub metering 3 and global intensity. (Former has a strength of ~0.5 and latter has a strength of ~0.6) All the other correlations between the variables included in the correlation matrix show weak correlations. (strength < 0.3)

For week 29, there are some interesting observations worth noting for global intensity. From Figure 2, we can observe that there is a trend where global intensity decreases throughout the nighttime hours for both weekdays and weekends. We can also observe that this trend also holds true for daytime hours for weekends. This trend, however, is not shared with the daytime hours during weekdays. For daytime hours for weekdays, the trend for global intensity slightly increases. From Figure 3, we observed that there are more fluctuations in the daytime hours on weekdays than on weekends. On the contrary, nighttime hours during weekdays and weekends share a similar trend. For daytime hours, there is a surge in intensity at 15:00 for both weekdays and weekends. There is a large peak prior to 9:00 for the weekend. In fact, the peak prior to 9:00 for weekends is higher in magnitude than the peak that occurs later at 15:00. As for the nighttime hours, the biggest peak for both weekdays and weekends occurs around 21:00. A smaller peak then occurs around 7:00 for both weekdays and weekends as well. Overall, global intensity tends to surge sometime during the morning (Between 7:00 and 9:00), afternoon (Prior to 15:00) and at night (Shortly before 21:00).