SMART GRID DATA ANALYSIS

A thesis

submitted by

Matam Srivatsav(EE11B026)

under the guidance of

Dr. K. Shanti Swarup

for the award of the degree

of

Bachelor of Technology

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

MAY 2015

THESIS CERTIFICATE

This is to certify that the thesis entitled **Smart Grid Data Analysis**, submitted by **Matam Srivatsav**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology**, is a bonafide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Place: Chennai
Date: 8th May 2015

Prof. K. S. Swarup
Research Guide
Professor
Dept. of Electrical Engineering
IIT-Madras, Chennai - 600 036

ACKNOWLEDGEMENTS

It gives me an immense pleasure to take the opportunity to acknowledge and express my gratitude to all those who helped me throughout my Project.

First and foremost, I express my sincere gratitude to my Research Supervisor, Prof.K. Shanti Swarup, for his consistent encouragement, guidance and support from the initial to the final level of my research work, which enabled me to carry out this work successfully. I also gratefully acknowledge the generosity and infinite patience shown by him. I wish to thank him for the personal care and affection showed to me, which made my research a lifetime memorable one.

Last but not the least, I thank all my friends, colleagues and all others who directly or indirectly helped me in carrying out this research work in a successful manner.

Above all, I thank the Almighty for His guidance, strength, protection and tremendous blessings bestowed throughout my life.

Place: Chennai
Date: 8th May 2015

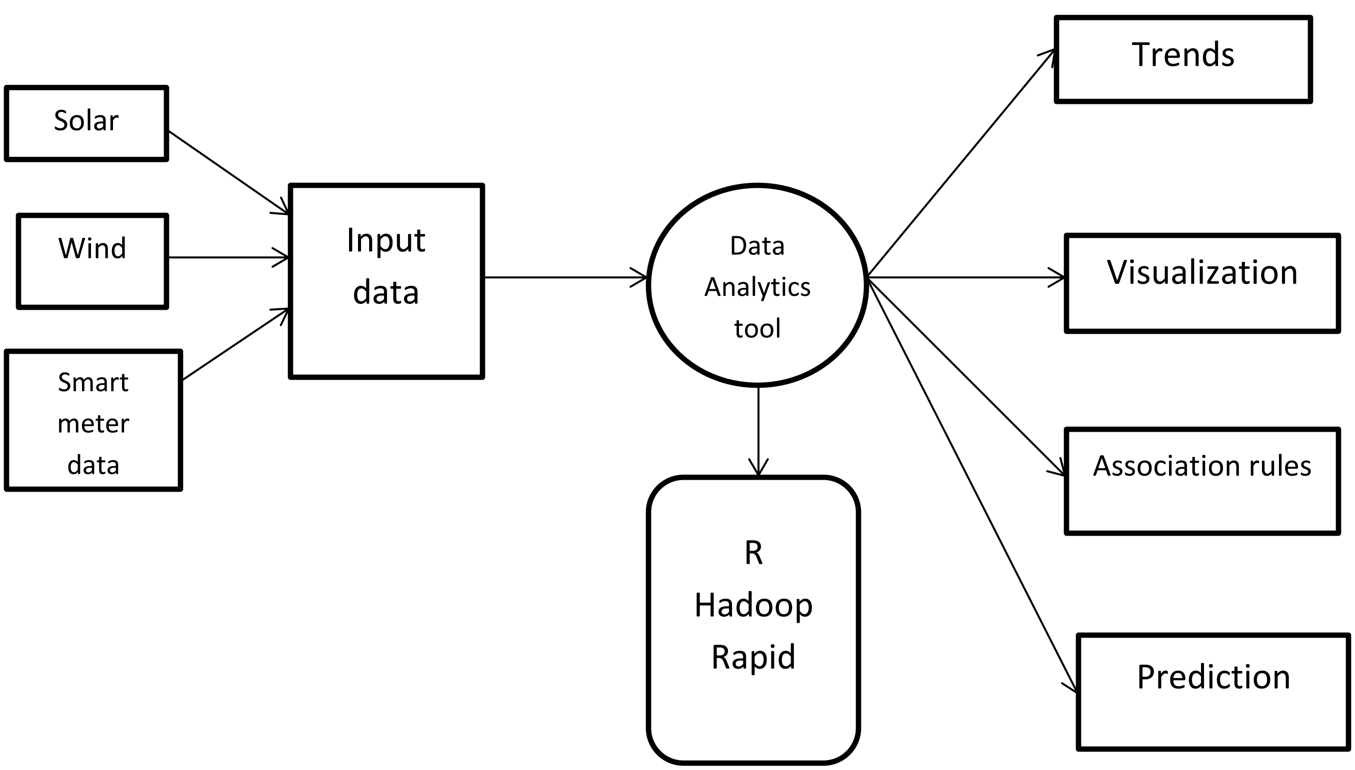
Contents

1. ABSTRACT

The smart grid we have chosen to analyze is an Australian smart grid from an area in New South Wales. This area has a tropical climate which is almost the same as Indian climate. We have the smart grid Load data, Price data and the Demand data for a year (2012).

The Data:

The smart grids read the load every half an hour in a day and we have this data. Each half an hour is called as a PERIOD. So in total we have 48 periods a day. And as we are analyzing for the year 2012, which is a leap year, we have 366 days. The demand and the price for the power utilized have also got in the same manner i.e. 48 samples per day and 366 days in total.



The analysis we have done is mostly of Trends, Association rules and Visualization. With this analysis, we can help the consumers with the analytics results to reduce their power usage or giving the results for their efficient usage of power. From there on we proceed for a pricing scheme which helps the company to increase their profit. The data analytics tool I'm using is **R.**

2. Analysis

For power producing companies:

We have analyzed the Price and demand trends for the year 2012. There were many instants where the demand for the power was very high but the price was in fact low. There were few instances where the Price was high but the demand for the power was very low compared to the high peaks. As a result, the company had less profit. We designed a better way of their pricing scheme which includes the demand and load that area had at that instant, and thereby increasing their profits.

For consumers:

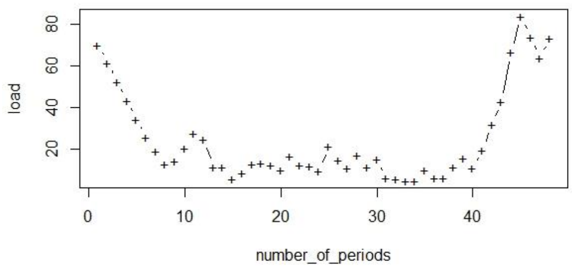
The Load and Price pattern of the power, with each period, with each day, with each week, with each month and each season is analyzed and given to the consumers. And also we give the approximate times of the peak loads and prices, so that they can reduce the usage of their high power appliances at those times to save more money.

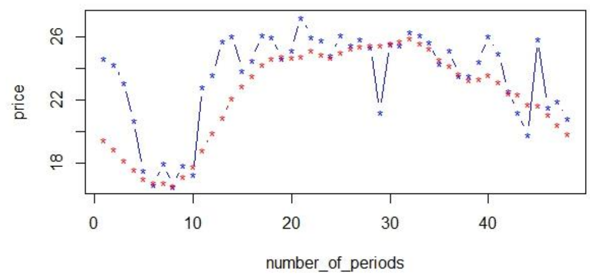
2.1 Analysis for the Consumer

2.1.1. Load, Price and Demand profile for a day:

We have studied the load, price and demand profiles for many days in random. They all appeared to be similar. Given below are some of the graphs with the load, price and demand profiles.

Case 1:





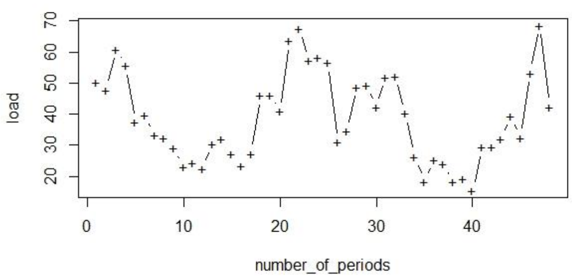
The above graph is the graph for January 10th 2012 (Tuesday). The first graph shows the load profile for the day and the second one show the price (blue) and demand (red) profiles. The demand magnitude being very high is reduced by a factor of 380.

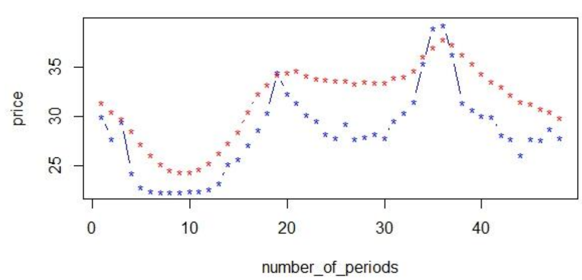
Inferences :

- Being a weekday, the load values are low during the day times and the load is high at the nights.
- We can see small peaks (of about 15-20 Mwh) at 10th period i.e. 5a.m, when most people start their day. Usage of exercise machines, washing machines and geysers is the cause for such a peak at that time.
- There is a small peak again (10-15 Mwh) at 25th period i.e. at 12-1p.m - lunch time. Usage of ovens, grills and toasters is the reason.
- The load is almost uniform (at around 5-10 Mwh) throughout till 40th period i.e. 8p.m. Then we can see the load value suddenly starting to blow up. Usage of televisions for entertainment, geysers, ovens and toasters for dinner are the reasons.
- The highest peak (80-85 Mwh) of the day is observed at 46nd period i.e. at 11p.m. Usage of air conditioners for sleeping is the reason.

- The load is almost the same throughout the daytime. People working in offices and staying away from their houses reduces the load at their house. So we can see no peaks but a constant little load during this time.
- The demand and the price are going up and coming down simultaneously. When the demand is high, the price is high and when there is low demand, the price is low.
- The highest demand is 9800 units and the highest price is 26 $ /unit .

Case 2 :





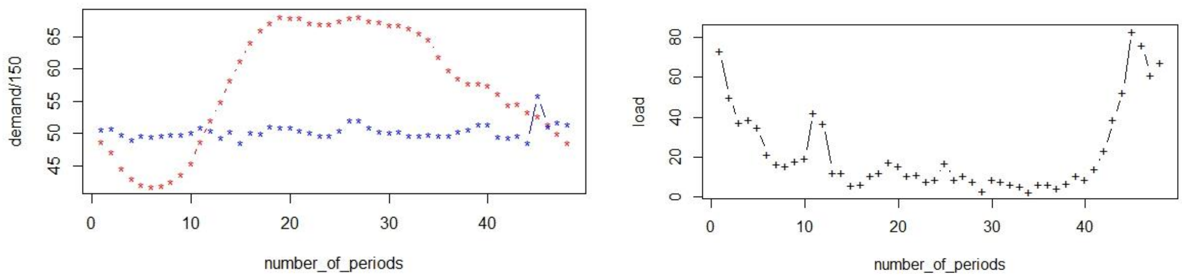
The above graph is the graph for 3rd of June 2012 (Saturday). The first graph shows the load profile for the day and the second one show the price (blue) and demand (red) profiles. The demand magnitude being very high is reduced by a factor of 270.

Inferences :

- Being a weekend, the load values are low during the evening times and the load is high during the mid-day and the night.
- We can see small peaks (25-30 Mwh) at 14th period i.e. 7a.m, when most people start their day (a little late compared to a weekday). Usage of exercise machines, washing machines and geysers is the cause for such a peak at that time.
- The load is reaches a peak (65-70 Mwh) by 24th period i.e. at 12 p.m. People staying at home and watching televisions, gardening, use of washing machines for laundry are the reasons. Cooking lunch also adds up for the peak.
- There is a small peak (45-50 Mwh) again till 32th period i.e. at 4p.m - lunch time. Relaxing and power naps with the usage of fans and televisions/computers is the reason.
- The load is almost uniform (20-30 Mwh) throughout till 40th period i.e. 8p.m. Then we can see the load value suddenly starting to blow up. Usage of televisions for entertainment, geysers, ovens and toasters for dinner are the reasons.

- The highest peak (65-70 Mwh) of the day is observed at 46nd period i.e. at 11p.m. Usage of air conditioners for sleeping is the reason.
- The demand and the price are going up and coming down simultaneously. When the demand is high, the price is high and when there is low demand, the price is low.
- We can observe two peaks in the price and demand also. One at 20th period at price of 35 $ / 9450 units and i.e. at 10a.m and the other at 32nd period i.e. at 4p.m at 40 $ and 10800 units.

Case 3:



The above graph is the graph for December 20th 2012 (Thursday). The first graph shows the price (blue) and demand (red) profiles for the day and the second one show the load profile. The demand magnitude being very high is reduced by a factor of 150.

Inferences :

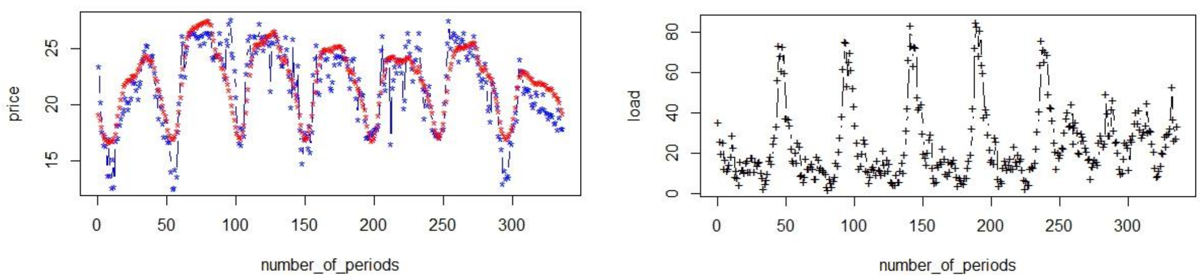
- Being a weekday, the load values are low during the day times and the load is high at the nights.
- We can see small peaks (40 Mwh) at 10th period i.e. 5a.m, when most people start their day. Usage of exercise machines, washing machines and geysers is the cause for such a peak at that time.
- There is a small peak (15-20 Mwh) again at 25th period i.e. at 12-1p.m - lunch time. Usage of ovens, grills and toasters is the reason.
- The load is almost uniform (5-10 Mwh) throughout till 40th period i.e. 8p.m. Then we can see the load value suddenly starting to blow up. Usage of televisions for entertainment, geysers, ovens and toasters for dinner are the reasons.
- The highest peak (85-90 Mwh) of the day is observed at 45th period i.e. at 10:30p.m. Usage of air conditioners for sleeping is the reason.

- The load is almost the same throughout the daytime. People working in offices and staying away from their houses reduces the load at their house. So we can see no peaks but a constant little load during this time.
- The demand and the price are following different trend here. Irrespective of the demand, the price is almost a constant at 50-55 $ throughout the day.

2.1.2. Load, Price and Demand profile for a week :

The Load, Demand and price profiles for a day are observed. Now we would like to see how these profiles change with the week.

<u>Case 1:</u>



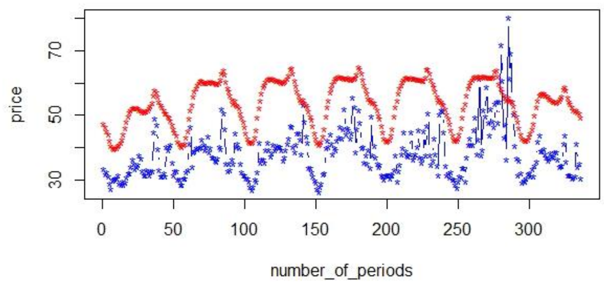
The above graph is the graph for a week (January 8th to 14th -Sunday to Saturday). The first graph shows the price (blue) and demand (red) profiles for the week and the second one show the load profile. The demand magnitude being very high is reduced by a factor of 370.

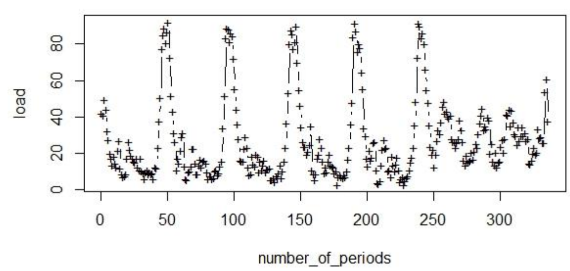
<u>Inferences:</u>

- The week starts with a Sunday. The load values are low during the day times and the loads are high at the nights.
- Weekday nights have higher loads than weekend nights. Parties and weekend late night activities are the reasons.
- The highest load of 80-85 Mwh is observed on Tuesday and Wednesday nights.

- The least load of 0-5 Mwh is observed at the dinner times of 8-10 p.m everyday.
- Weekend daytimes have higher loads than weekday daytimes. Staying in the houses in the hot afternoons is the reason. Usage of fans, televisions etc. cause the higher loads.
- Weekdays are similar with low and almost equal loads in the afternoons and higher loads during the nights.
- Loads are the least in the evenings throughout the week.
- The demand and the price are raise and fall almost simultaneously most of the times.
- The demand is less in the mornings and it reaches a peak by the afternoon. And again reaches the minimum the next morning.
- Similarly lower price is observed in the mornings.

<u>Case 2:</u>





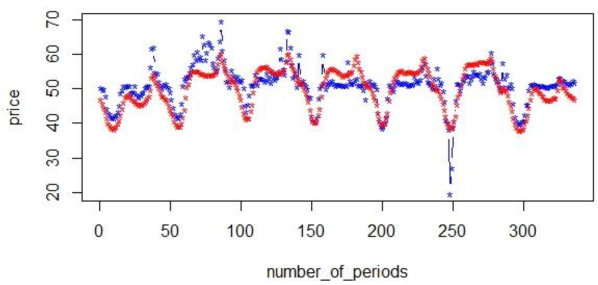
The above graph is the graph for a week (April 15th to 21st -Sunday to Saturday). The first graph shows the price (blue) and demand (red) profiles for the week and the second one show the load profile. The demand magnitude being very high is reduced by a factor of 150.

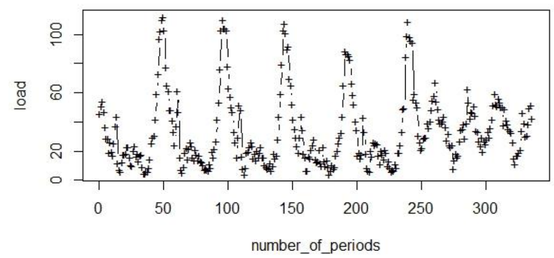
<u>Inferences:</u>

- The week starts with a Sunday. The load values are low during the day times and the loads are high at the nights.
- Weekday nights have higher loads than weekend nights. Parties and weekend late night activities are the reasons.
- The highest load of 90-95 Mwh is observed on Monday, Thursday and Friday nights.
- The least load of 0-5 Mwh is observed at the dinner times of 8-10 p.m everyday.

- Weekend daytimes have higher loads (20-50 Mwh) than weekday daytimes. Staying in the houses in the hot afternoons is the reason. Usage of fans, televisions etc. cause the higher loads.
- Weekdays are similar with low and almost equal loads in the afternoons and higher loads during the nights.
- Loads are the least in the evenings throughout the week.
- The demand and the price are raise and fall almost simultaneously most of the times.
- The demand is less in the mornings and it reaches a flat peak by the afternoon. And again reaches the highest peak by 3-4 p.m everyday.
- Lower price is observed in the mornings and throughout the daytime the price values are with full variation but they do not reach any peak anywhere.

Case 3:





The above graph is the graph for a week (September 23rd to 29th -Sunday to Saturday). The first graph shows the price (blue) and demand (red) profiles for the week and the second one show the load profile. The demand magnitude being very high is reduced by a factor of 150.

Inferences:

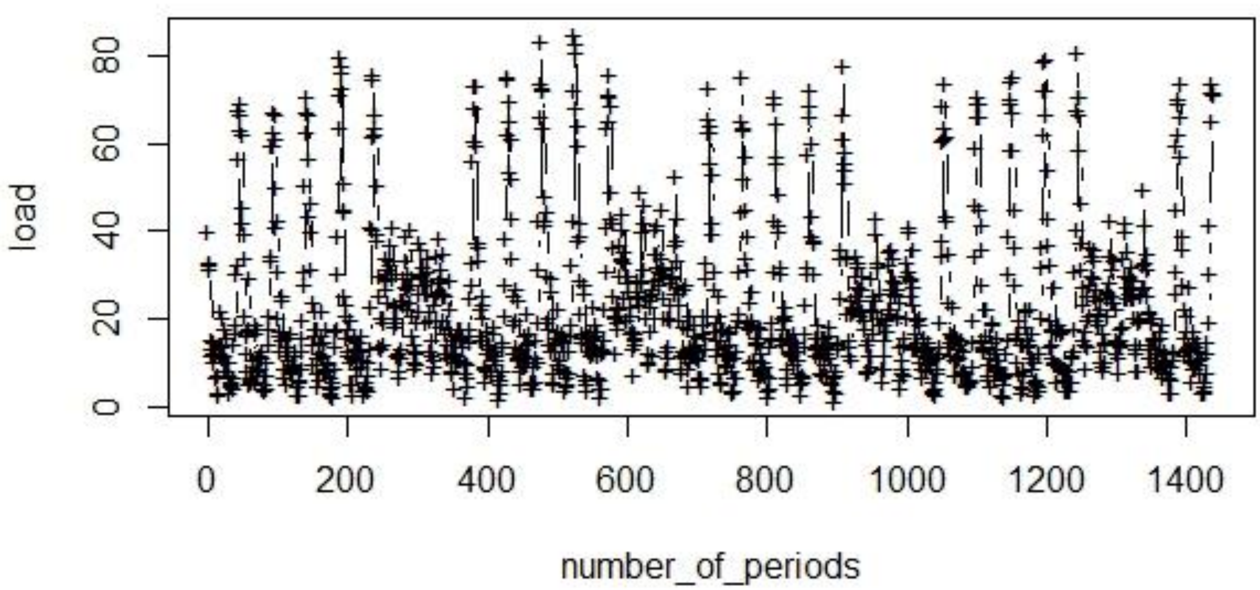
- The week starts with a Sunday. The load values are low during the day times and the loads are high at the nights.
- Weekday nights have higher loads than weekend nights. Parties and weekend late night activities are the reasons.
- The highest load of 115-120 Mwh is observed on Monday and Tuesday nights.
- The least load of 0-5 Mwh is observed at the dinner times of 8-10 p.m everyday.

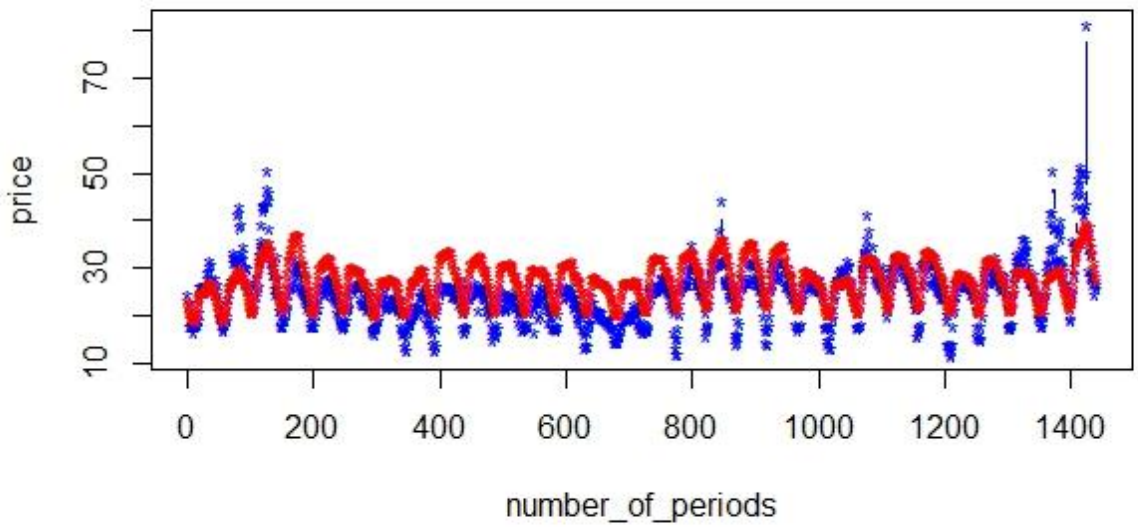
- Weekend daytimes have higher loads (30-70 Mwh) than weekday daytimes. Staying in the houses in the hot afternoons is the reason. Usage of fans, televisions etc. cause the higher loads.
- Weekdays are similar with low and almost equal loads in the afternoons and higher loads during the nights.
- Loads are the least in the evenings throughout the week.
- The demand and the price raise and fall almost simultaneously most of the times.
- The demand is less in the mornings and it reaches a flat peak by 10 a.m and then again decreases slightly. And again reaches the highest peak by 5-7 p.m everyday.
- Lower price is observed in the early mornings and throughout the daytime the price is almost a constant, reaches higher peaks at around 5-7 p.m everyday.

2.1.3. Load, Price and Demand profile for a month:

The observation of the load, price and demand profiles for a day and for a week is done. We would like to see how these vary within a month.

<u>January:</u>



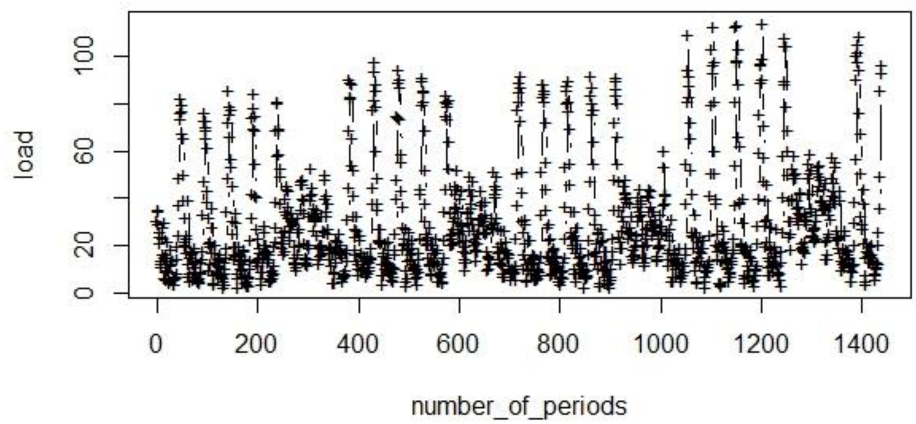


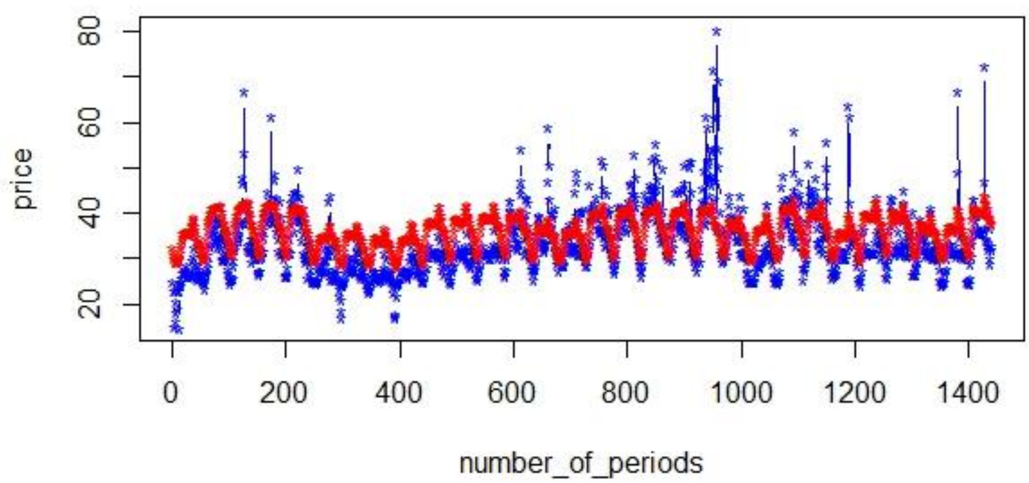
January starts with a Sunday. The first graph is the load vs the period of the day graph. The second one is the demand (red) and price (blue) vs the period of the day graph. The magnitude of demand is very high. So a reduction factor of 300 is used to reduce the demand and fit into the graph.

Inferences:

- Loads are higher on a weekend afternoon than on a weekday afternoon.
- Loads are higher on a weekday nights than on a weekend night.
- The highest load is around 75-80 Mwh observed on some weekday night and the least load is 0-5 Mwh observed on some weekday daytime.
- Price and demand almost follow the same trend. Price increses when demand is increased and decreases with the decrease of demand.
- Highest prices are observed in the afternoons and the least prices are observed in the early mornings.
- Price on some of the afternoons blew up, where the demand is not the highest.
- The highest demand is 10500 units and the least is 5100 units.
- The highest price is 80$ and the least is 12 $.

April:



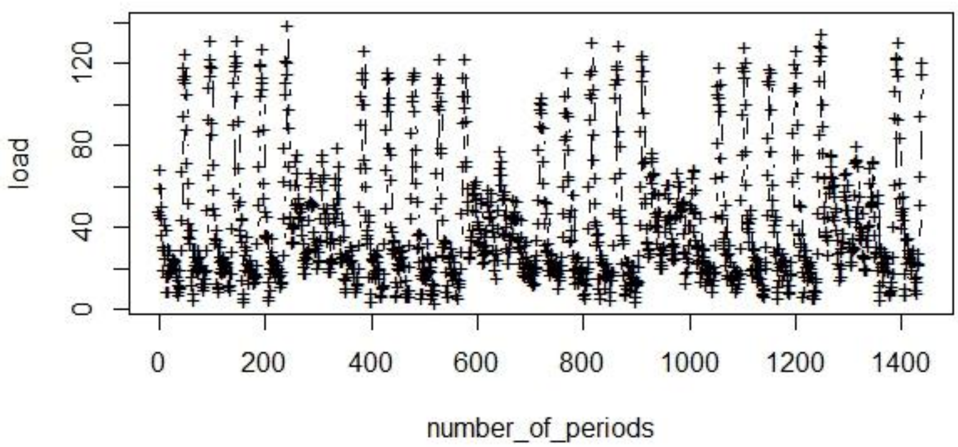


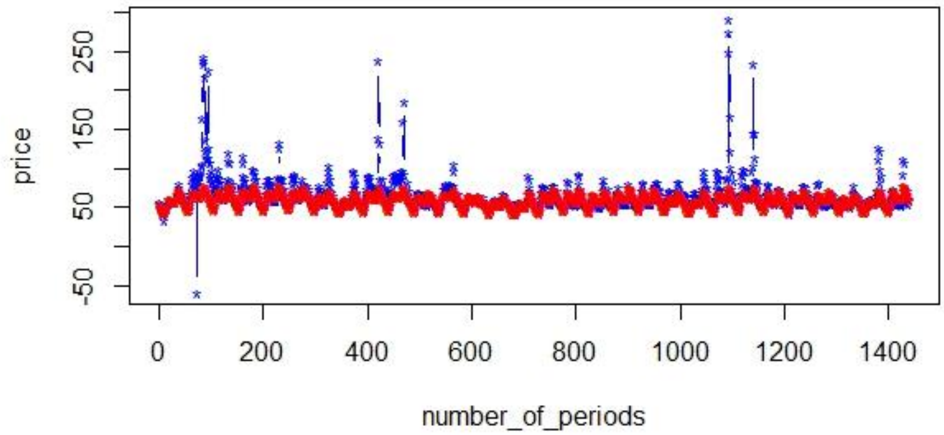
April starts with a Sunday. The first graph is the load vs the period of the day graph. The second one is the demand (red) and price (blue) vs the period of the day graph. The magnitude of demand is very high. So a reduction factor of 300 and an offset of 10 is used to reduce the demand and fit into the graph.

Inferences:

- Loads are higher on a weekend afternoon than on a weekday afternoon.
- Loads are higher on a weekday nights than on a weekend night.
- The highest load is around 115-120 Mwh observed on some of the weekday nights and the least load is 0-5 Mwh observed on some of the weekday's daytime.
- Price and demand almost follow the same trend. Price increases when demand is increased and decreases with the decrease of demand.
- Highest prices are observed in the afternoons and the least prices are observed in the early mornings.
- Price on some of the afternoons blew up, where the demand is not the highest.
- The highest demand is 9600 units and the least is 5100 units.
- The highest price is 80$ and the least is 12 $.

July:



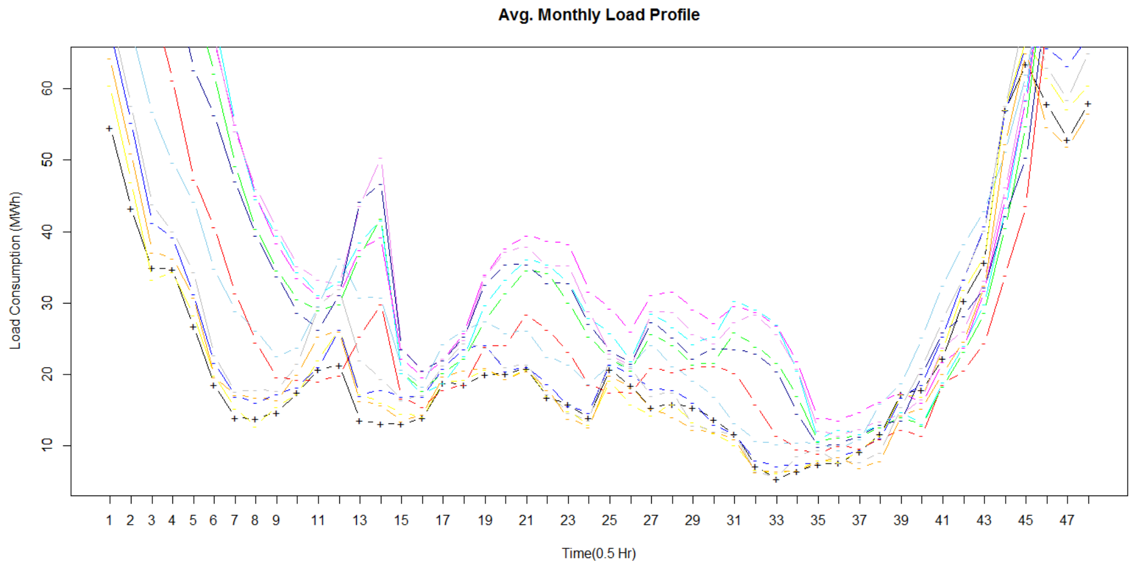


July starts with a Sunday. The first graph is the load vs the period of the day graph. The second one is the demand (red) and price (blue) vs the period of the day graph. The magnitude of demand is very high. So a reduction factor of 150 is used to reduce the demand and fit into the graph.

Inferences:

- Loads are higher on a weekend afternoon than on a weekday afternoon.
- Loads are higher on a weekday nights than on a weekend night.
- Being summer, the highest load is moved up to 135-140 Mwh observed on some of the weekday nights and the least load is 0-5 Mwh observed on some of the weekday's daytime.
- Price and demand almost follow the same trend. Price increases when demand is increased and decreases with the decrease of demand.
- Highest prices are observed in the afternoons and the least prices are observed in the early mornings.
- Price on some of the afternoons blew up, where the demand is not the highest.
- The highest demand is 11800 units and the least is 6100 units.
- The highest price is 291$ and the least is -59 $.
- Here we observe that the least price is a negative value, which implies that the company is paying the consumers for the usage of electricity.
- The demand is higher when low price is observed. This results in losses to the company. For correcting this and bringing out the profits for the company, a new price schema is designed, which will be explained later.

2.1.4. Load, Price and Demand profile for the year 2012:



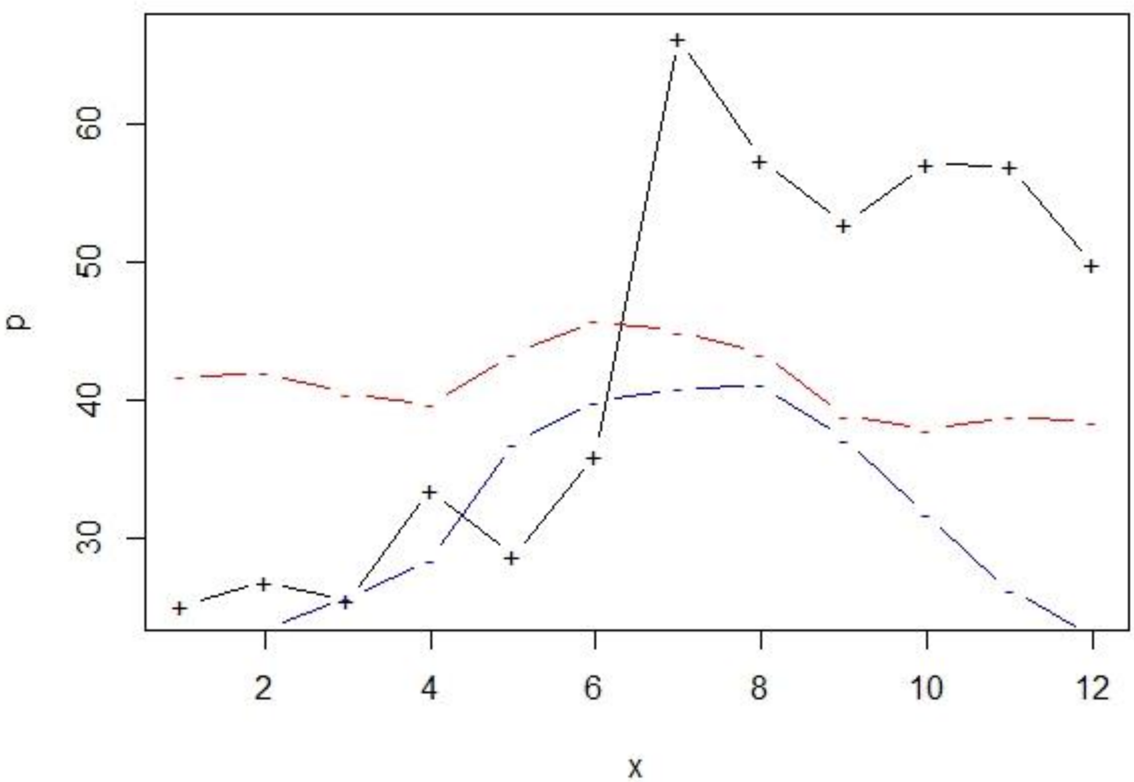
Each line is the average load profile for a month in 2012.

Black - january
Yellow – february
Blue – march
Red – april
Green - may
Cyan - June
Magenta - July
Violet - August
Dark blue - September
Skyblue - October
Grey -November
Orange – December

Inferences:

1. Cyan and magenta dominate at the noon periods. Summer noon have the highest loads.
2. All the plots look similar but with different offset.
3. Dark blue ,skyblue dominate in the morning times. Morning time consumption of power is higher during september and october. Usage of Geysers might be the reason for this.
4. There is a sudden raise in load at 11-16 periods i.e. 5:30 am to 8:00 am. This might be due to the cooking and other routine works at the start of the day.
5. Periods 33-41 i.e. 4:30 pm to 8:00 pm has always recorded the least loads. This might be due to the evening time jogging, playing and other works which do not require electricity.

2.1.5. Average load, Price and demand for each month :



The above graph is the average price load and demand, averaged on each month.

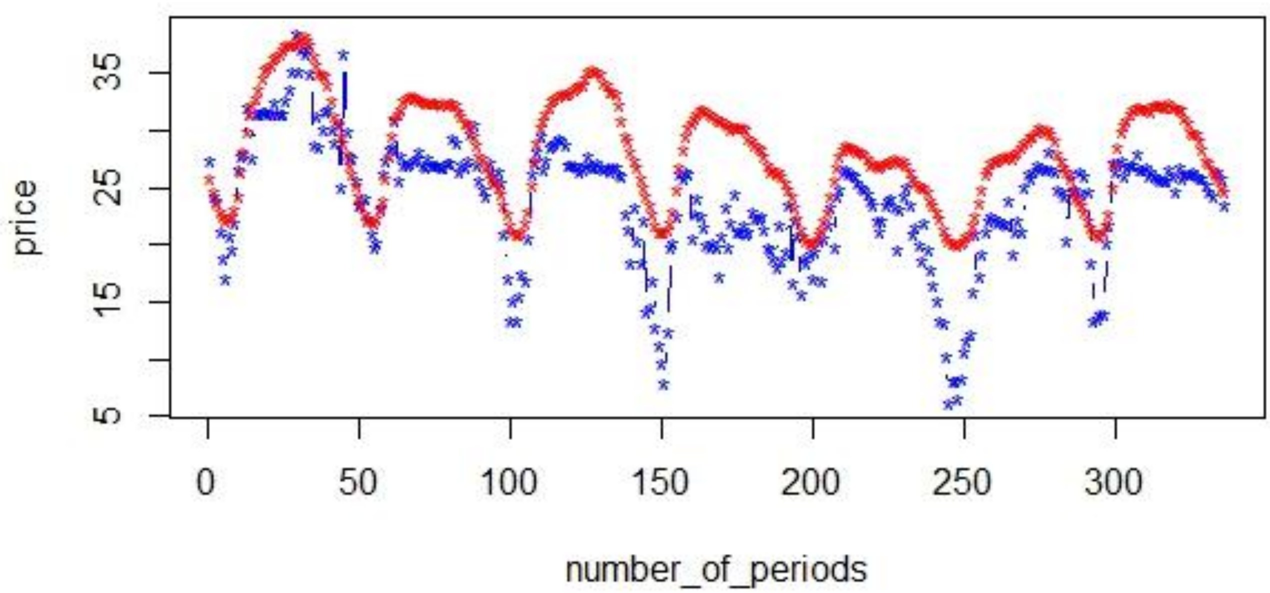
1. The blue graph is the average load for each month.
2. The red one is the average demand reduced by a factor of 200 for each month.
3. The black one is the average price for each month.
4. The average load is a bell shaped curve with the highest 40 MWh during the summer i.e. May - September and the least 20 MWh during winter i.e. November - March.
5. The demand is also a bell shaped curve with the highest during the summer i.e. May – September.
6. The demand increases and decreases with the load.
7. The price also increases with demand, but sometimes enormously.
8. The price is also the highest during the summer (or more precisely July and august) . The price of the power at the start and end of the year are not same which means hike in fuel prices may also be the reason for the increase in the price of the power utilized.

2.2. Analysis for the Power Producing company

Power companies are more about their profits. They cannot handle how the demand and loads are. But what they can handle is the price per unit of the power. If the price is higher than the normal when there is more demand, the company can cash out of it. And when there is less demand, the price can be reduced. Thus the company can gain something out of nothing. Here are some examples where the demand and price are not raising and falling simultaneously.

In the graphs to follow, the blue colored *'s are the price values and the red *'s are the demand values reduced by 300 times.

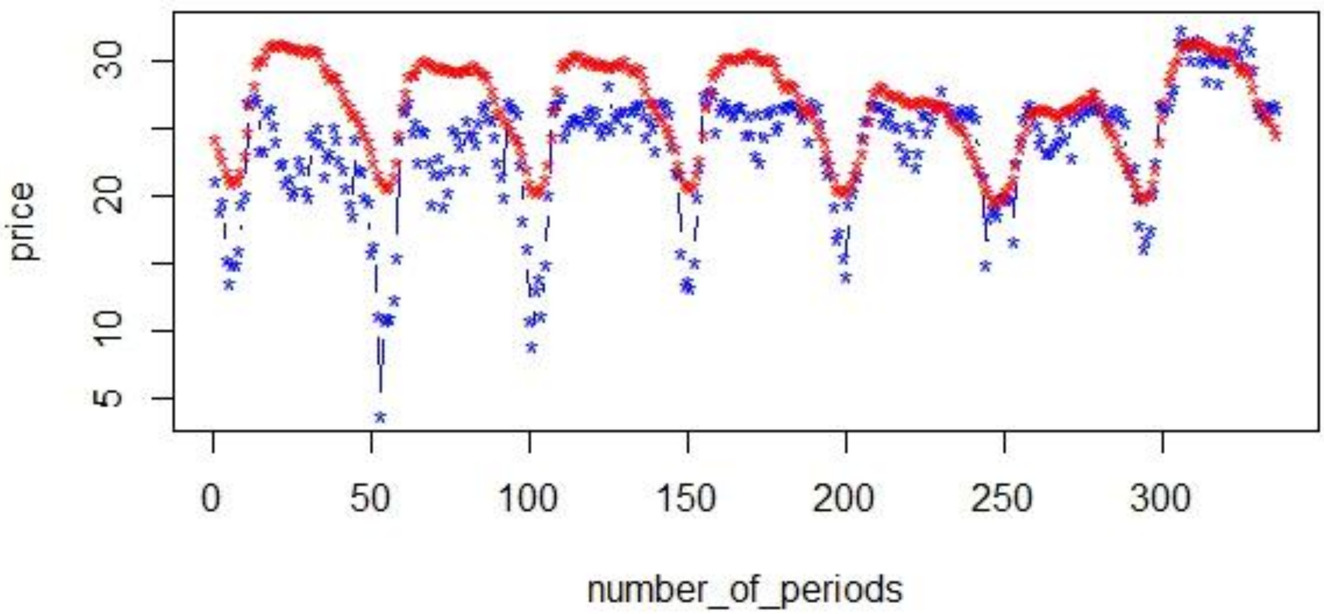
Case 1:



Faults:

- In between 150 and 200th periods and 200 and 250th periods the demand is high, but the price is going down at that point. This decreases the profit the company gets.

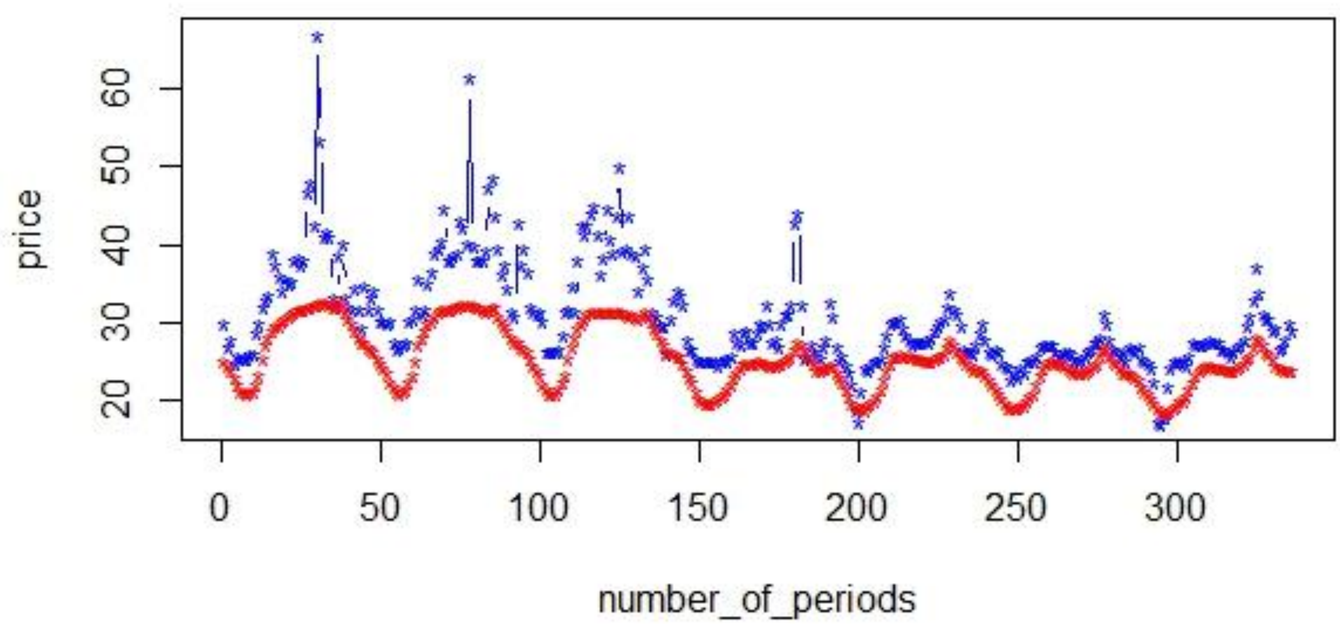
Case 2:



Faults:

- In this graph, at every point where the demand is high, the price is going down. Price should be high whenever demand is high. But here that is not happening. This decreases the profit the company gets.

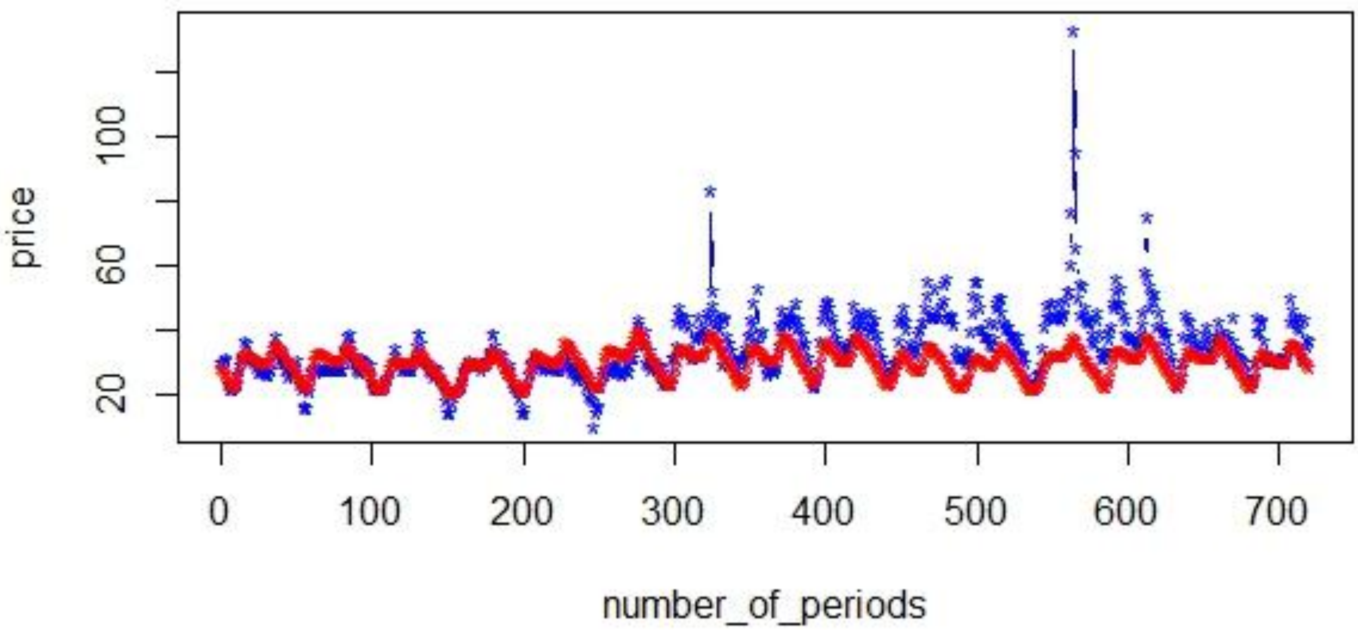
Case 3:



Faults:

- This might look as a perfect graph to maximize profit. But with a very blew up price, at the peak between 0-50th periods and also 50-100th periods, this might look bit awkward for the customers. Sometimes it so happens that they choose other companies for the power than this company. So this blowing up should also not happen.

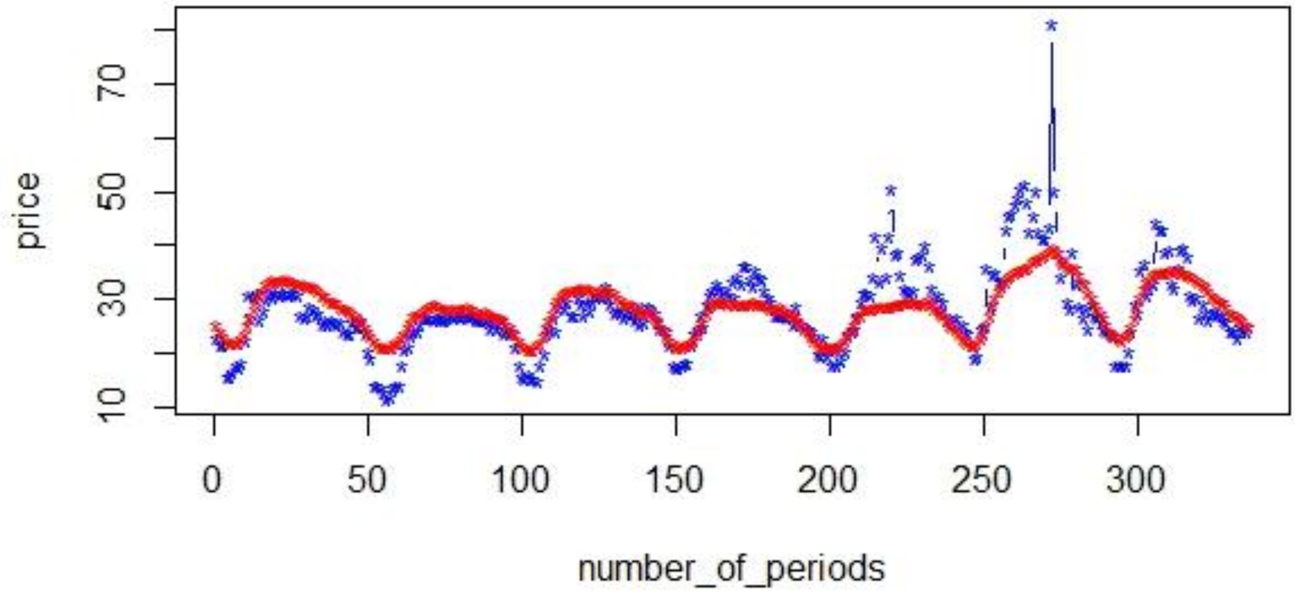
Case 4:



Faults:

- This might look as a perfect graph to maximize profit. But with a very blew up price, starting at the peak between 300-400th periods and also 500-600th periods, this might look bit awkward for the customers. Sometimes it so happens that they choose other companies for the power than this company. So this blowing up should also not happen.

Case 5:

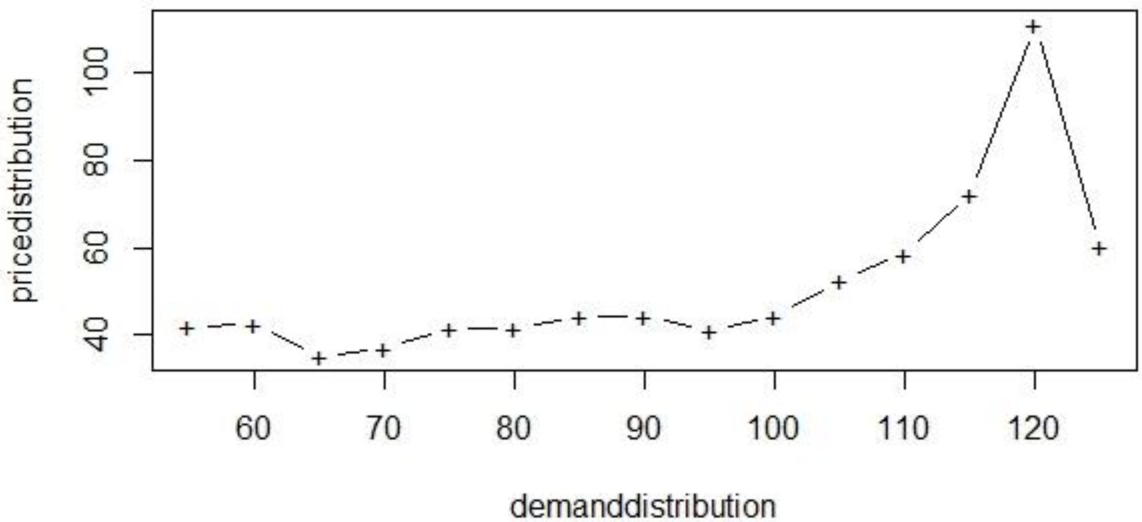


Faults:

- This graph looks PERFECT till 200th period. Then the blowing up of price started. We should try to create a graph which looks like the above graph till 200th period.

2.2.1. Price vs Demand

So, price vs demand curve should be linear or almost linear for the best pricing scheme. But the graph which shows the price vs demand is.



The demand distribution on the x-axis is demand reduced by 100. And the y-axis is unchaged value of price in $/ unit.

Observations:

- The demand vs price graph is neither linear nor almost linear. It is some kind of curvy graph which resembles a parabola.
- At the highest demand the price just got fallen down. This is the first thing that shouldn't happen.
- The price is almost a constant from demand values of 55-100. There should be an increase in price when demand increases.
- Price value is given as per load value i.e. price is $ / unit of load, which implies that price increases with load. Price can be a linear function on load.
- So the price we got should be a linear function on load and demand.

2.2.2. PRICING SCHEME

The company should change the way they price the power. The best pricing schema would be the one in which the price is comparably higher when demand is higher and lower when demand is lower. As seen before, price of the power should be linearly proportional to both demand and load.

P_new = New price

P_old = Old price

D = demand value at the same period as of price.

L = Load value at the same period as of price.

a = proportionality constant for demand

b = proportionality constant for load

So a new pricing schema should look like

P_new = P_old + (a*D) + (b*L)

Our aim is to increase the average price, in a way that the consumers cannot observe. So, if the price is higher than a certain value, we would like to increase the price with a "D" which is the difference of the **<u>D = (original demand) – (a reference value)</u>**. So, to choose this reference value, we have the statistics here.

```
> max(demandmat)
[1] 12152.87
> min(demandmat)
[1] 5074.63
> mean(demandmat)
[1] 8253.31
> quantile(demandmat)
      0%       25%       50%       75%      100% 
 5074.630  7307.505  8318.120  9127.810 12152.870 
```

The maximum of all the demands occurred on 366 days and 48 periods is 12152.87 units. The minimum of all the demands occurred on 366 days and 48 periods is 5074.63 units. The mean of all the demands occurred on 366 days and 48 periods is 8253.31 units. The 1st quartile value is 7307.505 units, median is 8318.120 units , 3rd quartile is 9127.810 units.

In the same way, we would like to increase the price with a “L” above the reference value.

L = (original load) – (a reference value).

Here are the statistics for load.

```
> max(loadmat)
[1] 147.629
> min(loadmat)
[1] 0.801
> mean(loadmat)
[1] 31.44064
> quantile(loadmat)
       0%       25%       50%       75%      100%
  0.80100  13.64300  23.12000  41.38725 147.62900
```

The maximum of all the loads occurred on 366 days and 48 periods is 147.629 Mwh. The minimum of all the loads occurred on 366 days and 48 periods is 0.801 Mwh. The mean of all the loads occurred on 366 days and 48 periods is 31.44064 Mwh. The 1st quartile value is 13.643 Mwh, median is 23.12 Mwh , 3rd quartile is 147.629 Mwh.

The median is always the center of any distribution. There would be equal number of values which are above median and the same number of values less than the median. So I would like to choose median for both demand and load reference values.

So, the equation thus transforms to

P_new = P_old + a*[(demand)-(median_demand)] +b*[(load)-(median_load)]

This result of this would be, if demand is higher than median, then with a positive “a” the price increases, and with the load lesser the price would again be decreased. So when there is higher demand and higher load, the price would be higher, and with higher/lower demand and lower/higher load, the price would be almost near the old price. And with lower demand and lower load, the price would decrease.

But we need to choose a value for checking if the demand and load are higher/lower. For checking if the load is higher, we would like to see load compared with the third quartile value of it. For checking if the load is lower, we compare it with the first quartile. Same thing applies for the demand. Third quartile is used for checking the higher values and first quartile is used for checking the lower values. And the other values in between first and third quartiles are taken the same way but with a factor of **(0.5)** i.e. “a” would be “a/2” for this set of values and “b” would be “b/2” for this set of values.

And amidst all these increasing prices and decreasing prices, we would like to see if we have not blown up the already blown up price. So, we keep a Maximum price beyond which the price shouldn’t go. This tells us that the price cannot go higher than some value but it can be even negative i.e. we can pay the customers for buying the power. Range of “a” and “b” values are chosen, and a loop is ran over all the values of demand, load and price. We finally get the “a” and “b” values.

The price statistics before were

```
> max(pricemat)
[1] 317.97
> min(pricemat)
[1] -59.28
> mean(pricemat)
[1] 43.05135
> quantile(pricemat)
    0%     25%     50%     75%    100%
-59.28   28.11   46.11   53.66  317.97
```

The new_price statistics are

```
> max(new_price)
[1] 252.422
> min(new_price)
[1] -167.2448
> mean(new_price)
[1] 46.8441
> quantile(new_price)
         0%         25%         50%         75%        100%
-167.244833    5.490771   40.604792   88.080250  252.422000
```

3. Conclusions

The maximum price in the previous case was 317.97 $ and in new scheme maximum price is 252.422 $, which is something exciting for the consumers.

The minimum price in the previous case was -59.28 $ and in the new scheme, minimum price is -167.244 $, which means the company pays you money for your usage of power, which excites the customer a little more.

The median of the prices in previous case was 46.11 $ and now it is 40.6 $. This says that median is also decreased, and so our company is preferred among all other companies.

But the mean in previous case was 43.05 $ and in the new scheme it is 46.84 $, which means the average price is increased by ~ 4$ on each period, which means the price is increased by 192 $ per day, and 70080 $ per year, and all this happens without the knowledge of the customers, who still think that the company has decreased its price.

References:

- X. Yu et al., “The New Frontier of Smart Grids,” IEEE Industrial Electronics Magazine, vol. 5, no. 3, 2011, pp. 49–63.
- S.D. Ramchurn et al., “Putting the ‘Smarts’ into the Smart Grid: A Grand Challenge for Artificial Intelligence,” Comm. ACM, vol. 55, no. 4, Apr. 2012, pp. 86–97.
- Dejan Ilic, Stamatis Karnouskos, and Martin Wilhelm et al., “A Comparative Analysis of Smart Metering Data Aggregation Performance,” IEEE conference , SAP Research, Karlsruhe, Germany.
- Stanley E. McHann, Jr et al., “Grid Analytics: How Much Data Do You Really Need? ,” IEEE conference , 2013.
- Damminda Alahakoon and Xinghuo Yu et al., “Advanced Analytics for Harnessing the Power of Smart Meter Big Data.,” IEEE conference, 2013
- R. Gerwen, S. Jaarsma and R. Wilhite, “Smart Metering,” Distributed Generation., pp 1-9, June 2006.