Name: Muhammad Jawad

ERP# 18087

Dataset Chosen: Smart Home Dataset with Weather Information (Kaggle)

Business Intelligence by Dr. Tariq Mahmood - Final Course Project

---

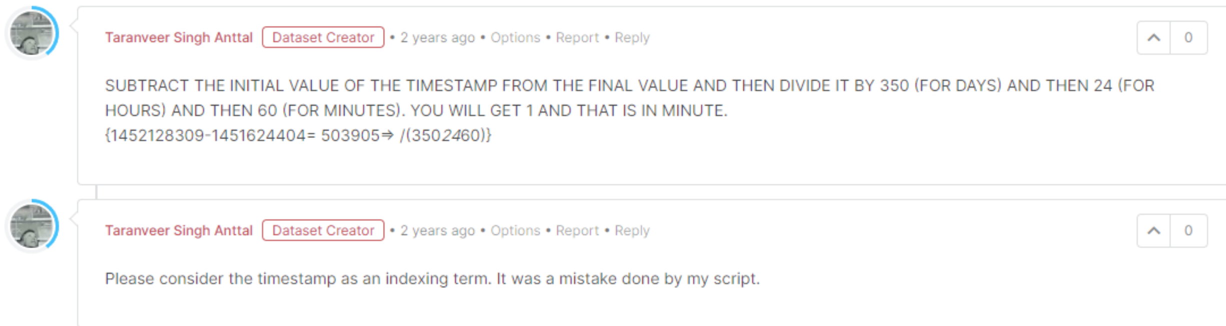
Tableau Packaged File (containing stories, dashboards, and all the charts) size is ~27 MB, exceeding IBA LMS limit. So, the file has been uploaded to Google Drive with complete view access. The file can be accessed from here.

Tool Chosen: Tableau (for the following reasons)

- More experience in Tableau
- Clean and easy-to-use interface
- Dataset is time series heavy, and I personally think that time series data analysis is more easily performed in Tableau than any other tool.
- Dataset's size is ~130 MB, so there's a clear signal to avoid at least cloud-based BI tools.

Data Cleaning:

- The missing value and outlier handling descriptions are briefly provided in the excel sheet. The remaining important steps are mentioned in this document.
- The column 'cloudCover' has strings and floats. There is one alphabetic string value 'cloudCover' occurring 58 times, since 58 << the dataset size, we replace all of them with np.nan (missing value). Now we delete all such rows, since they are very less compared to the dataset size. Then, we convert numbers in string type to float type to get the final column having all float values.
- The column time looks like a column of timestamp values. When we convert these values, we get dates from 1st January 2016 to 7th January 2016. But, the following comments are taken from the discussion section of kaggle's dataset. The dataset creator says to treat this column as indices rather than timestamp because of a mistake in script. So, we can't make sure the date, month, and year, since we can't treat this column as a timestamp. But what we know is that every row represents the data for a single minute. The data is consecutive. And it is for 350 days as per the dataset creator.



- The other continuous columns seem like giving better insights with continuous data, hence no need for discretization. Only cloudCover, visibility, and windBearings were discretized to be able to know easily the direction of wind and the level of cloud covers and visibility.

Fact checks:

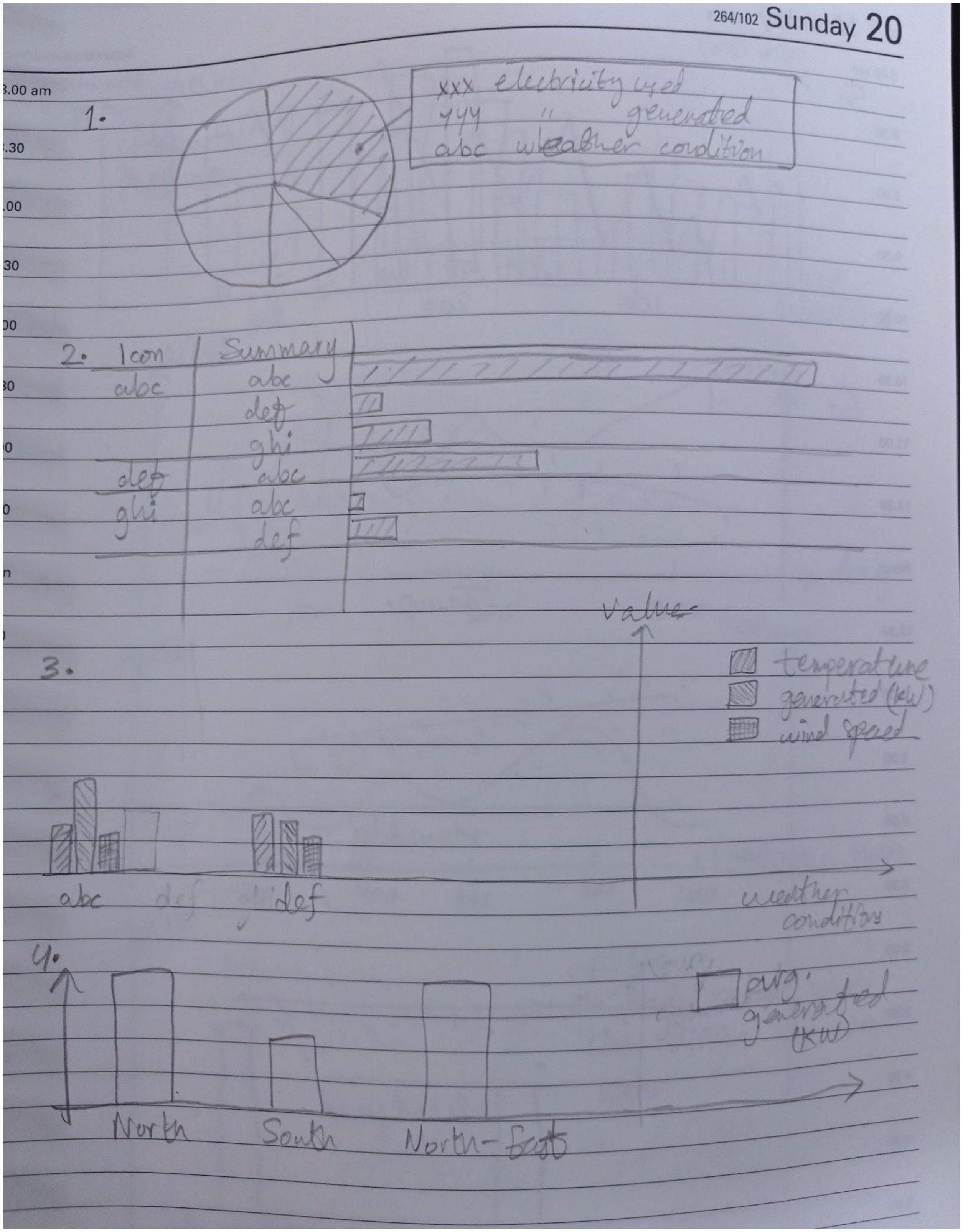
- Observation shows that at low relative humidity, output current increases as well as power efficiency of solar modules hence solar electricity generation is high and vice versa. [source]
- A southerly wind can increase the output of solar panels by up to 43% [source]
- With a moderate rise in dew point temperature, electricity demand could increase up to 20 percent. [source]
- Around 30% to 40% of electricity consumption in DCs is used for space cooling, thus leading to very inefficient DC operation. [source]

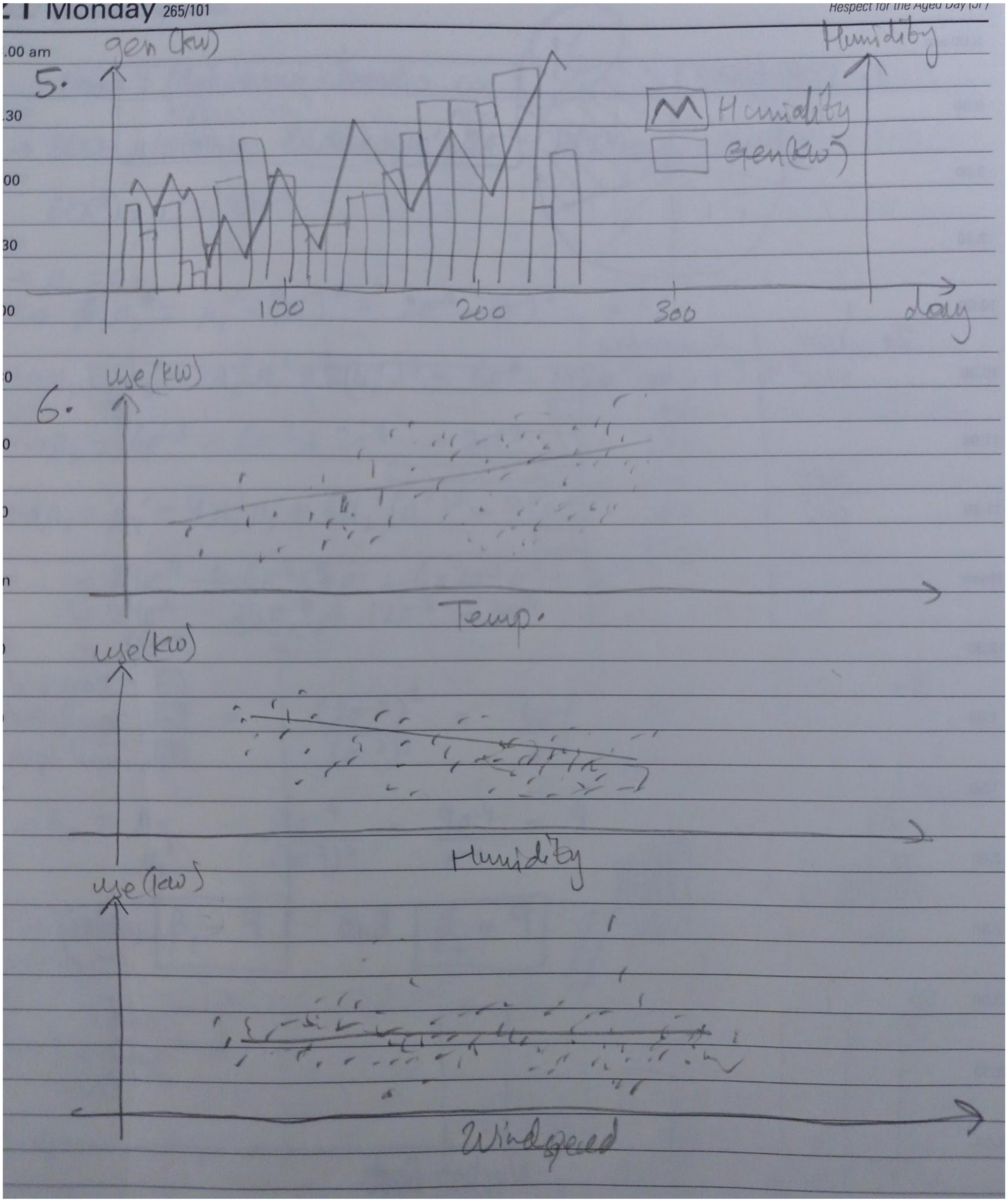
Step 2: (Visualization Chart for each corresponding query in excel file)

1. Pie chart
2. Highlight table
3. Histogram
4. Histogram
5. Dual Combination
6. Scatter plots
7. Histogram
8. Line charts
9. Histogram
10. Side-by-side histogram
11. Line charts

12. Line charts
13. Stacked bar charts

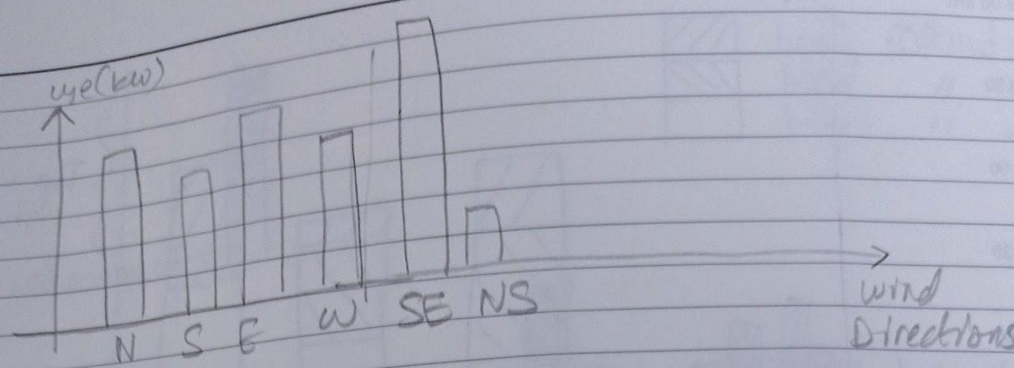
Paper Images for Step 2 (These images are randomly drawn charts - not based on data at this stage):


264/102 Sunday 20
1.
xxx electricity used
yyy " generated
abc weather condition
2. Icon
Summary
abc
def
ghi
3.
value
temperature
generated (kW)
wind speed
abc
def
weather conditions
4.
avg. generated (kW)
North
South
North-East

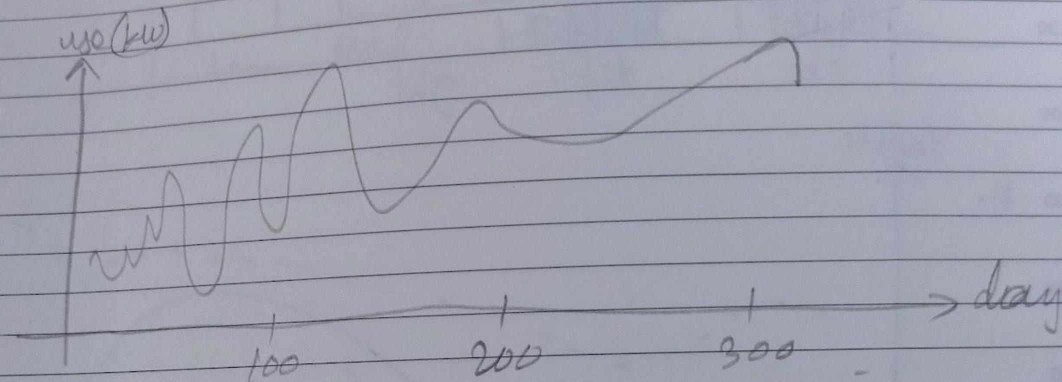

Monday 265/101
Respect for the Aged Day (JP)
5.
gen (kw)
Humidity
Humidity
Gen(kw)
100
200
300
day
6.
use(kw)
Temp.
use(kw)
Humidity
use(kw)
Windspeed

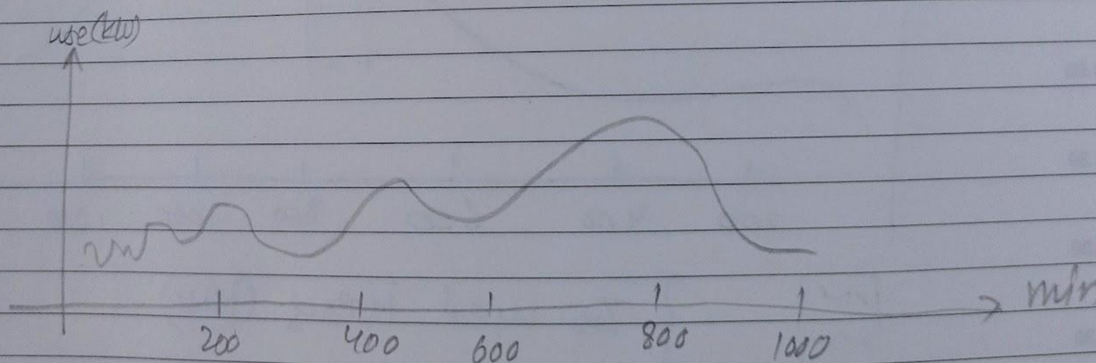
Autumnal Equinox Day (JP)

7.

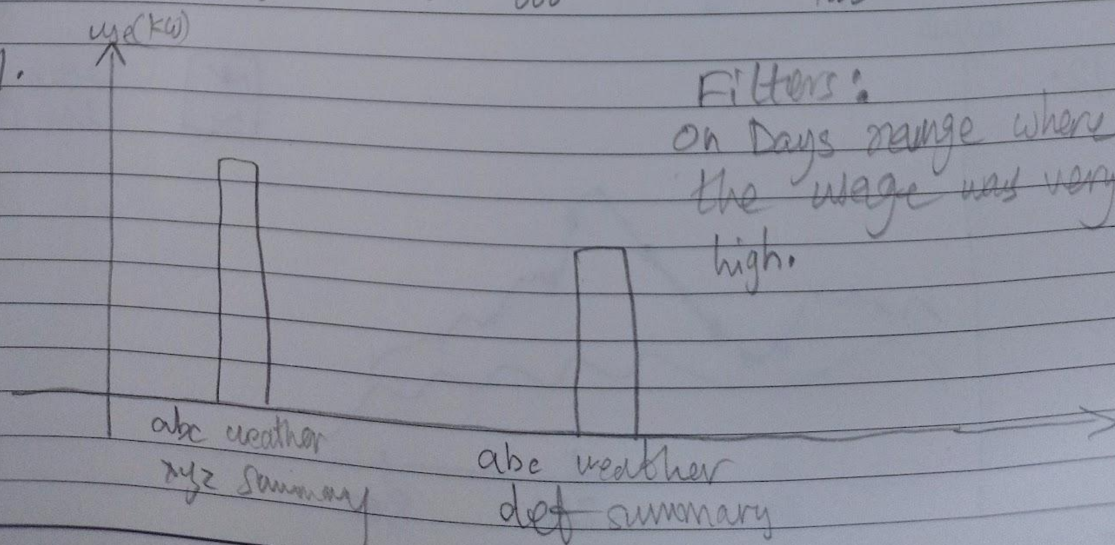


8.

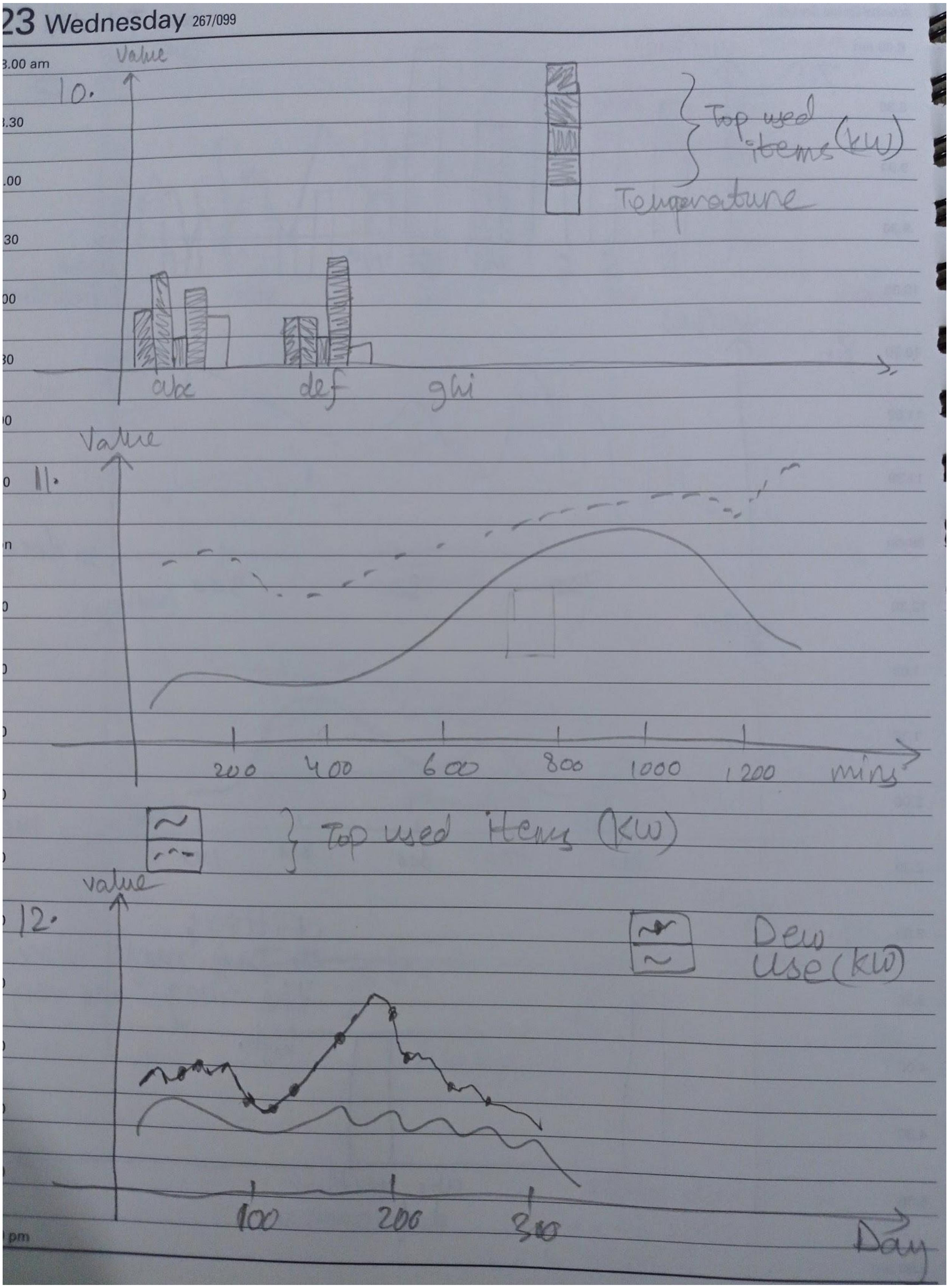


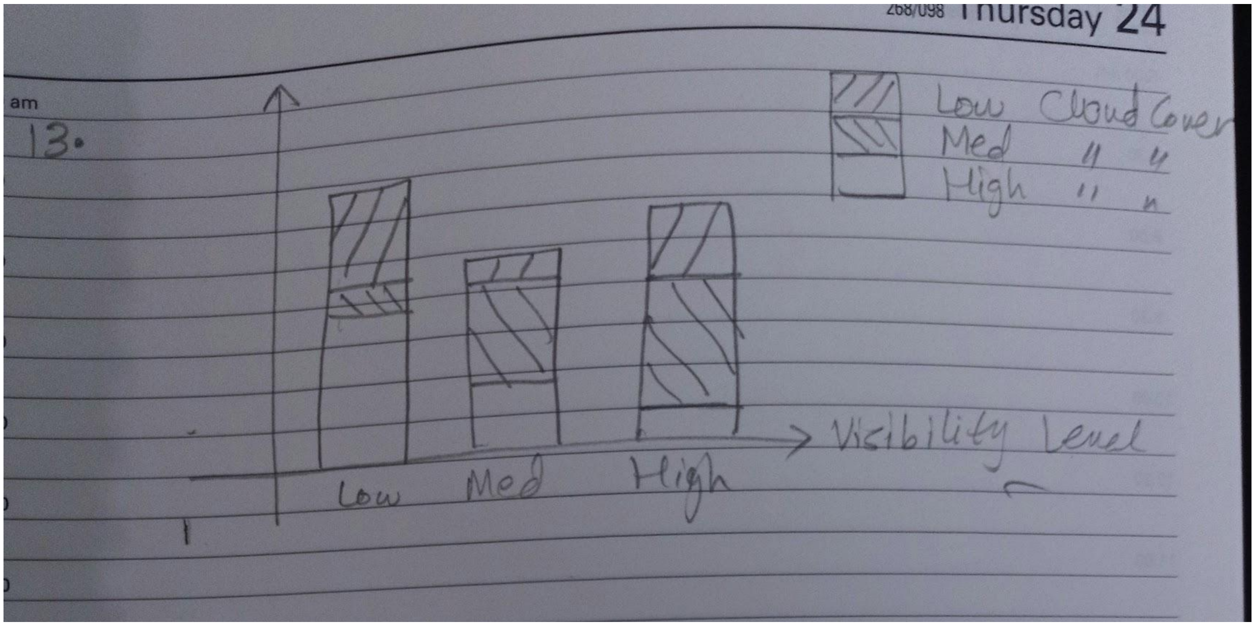


9.



Filters:
On Days range where the usage was very high.


23 Wednesday 267/099
Value
10.
Top used items (kW)
Temperature
abc
def
ghi
Value
11.
200
400
600
800
1000
1200
mins
Top used items (kW)
value
12.
Dew
Use (kW)
100
200
300
Day



Step 3: In search for a way to build up stories and dashboards

We can present two stories, one for each problem statement. We can show how each chart led to the idea of another chart with in-depth analysis to figure out the problem. Following is a hypothetical example showing a possible presentation for BI story related to the dataset's problem statement about energy generation:

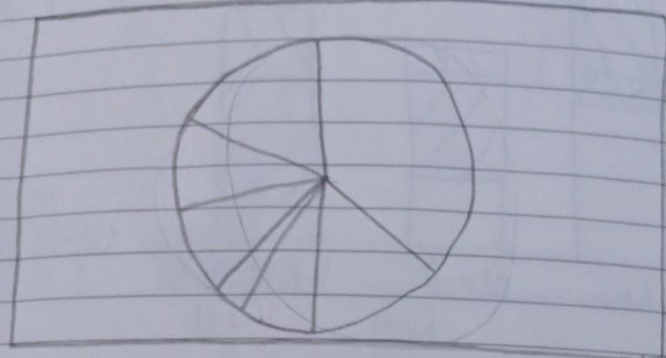
WEEK 39

25 Friday 269/097

BI Stories

→ An example of a possible story presentation for the problem of energy generation:

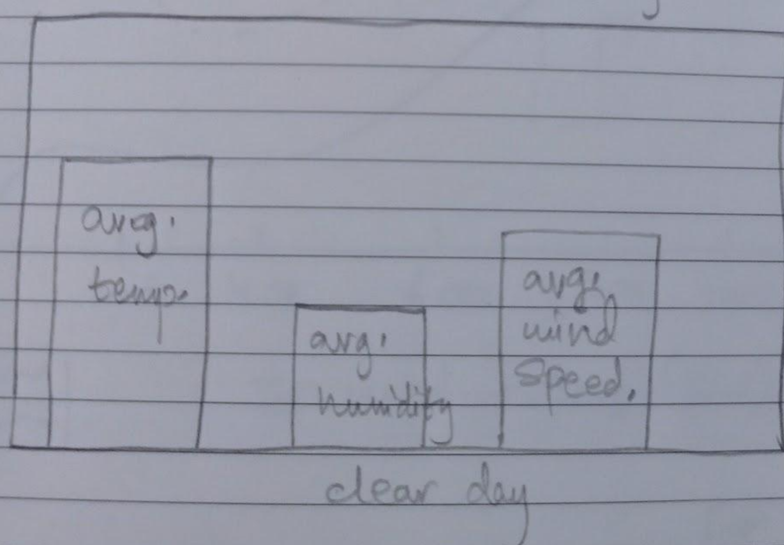
①



Here the energy generated in clear day is not what it should be.

We then dig into clear day's details.

②



Then we see why wind speed being low doesnot let the energy be generated efficiently.

And so on...

For dashboards we can create one for each problem, having main scorecards and bullet charts.

Dashboard# 1 for energy generation inefficiencies:

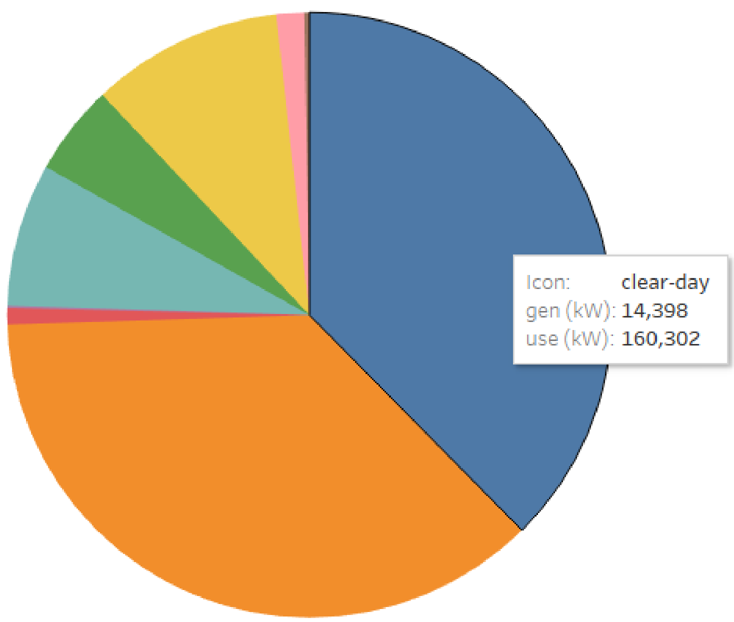
- Having main charts
- Average energy generated per minute in clear day and clear night scorecards
- Bullet chart comparing energy generated and used on average.

Dashboard# 2 for electricity usage analysis:

- Having main charts
- Correlations between usage and each of temperature, humidity, and wind speed
- Bullet chart comparing average furnace usage against overall usage for rainy weather

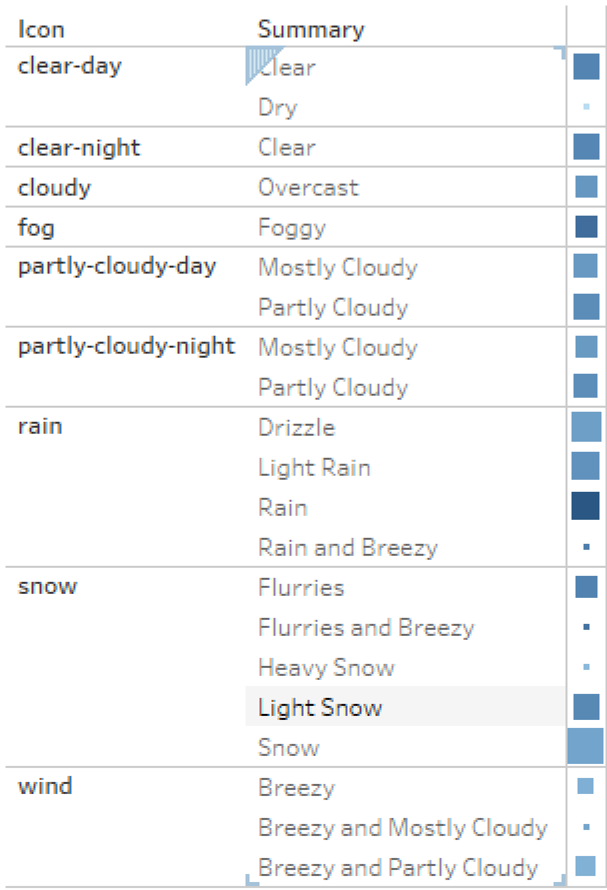
Step 4 (with analysis):

Chart # 1:



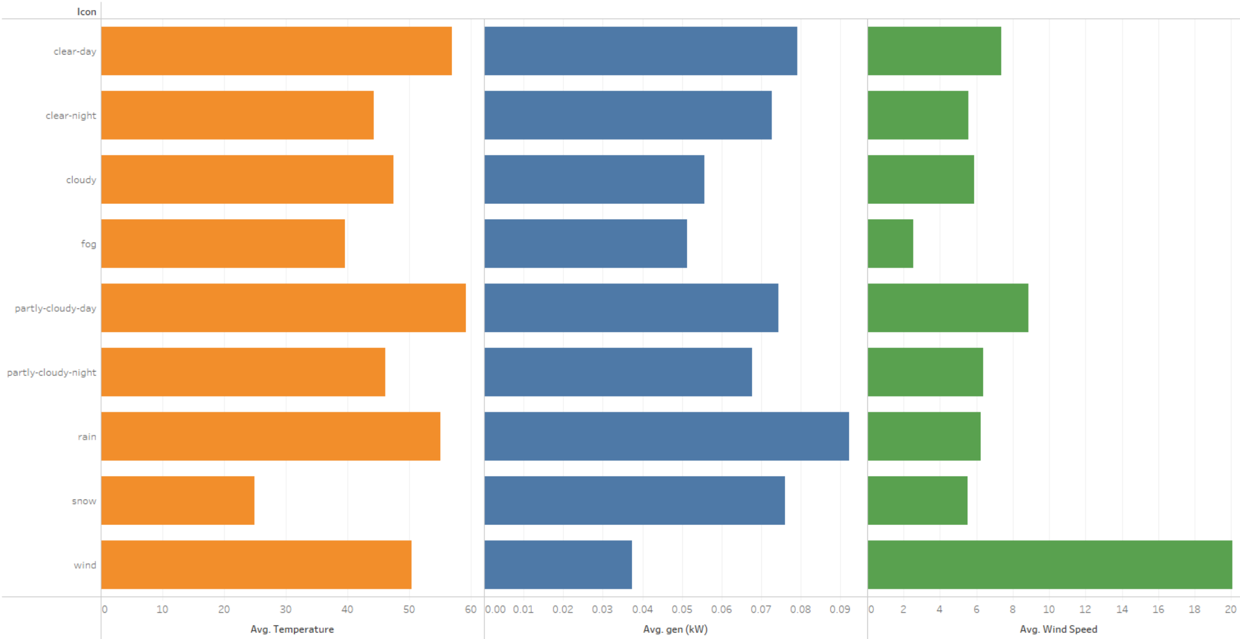
Shows electricity used and generated by each weather icon (clear-day, clear-night, rain, cloudy etc.) Normally, we get to see high energy generated in clear-day, but it was strange to find out that energy generated in clear-day was not more than generated in clear-night. *There's some problem with energy being generated efficiently*.

Chart# 2 and Chart# 3:



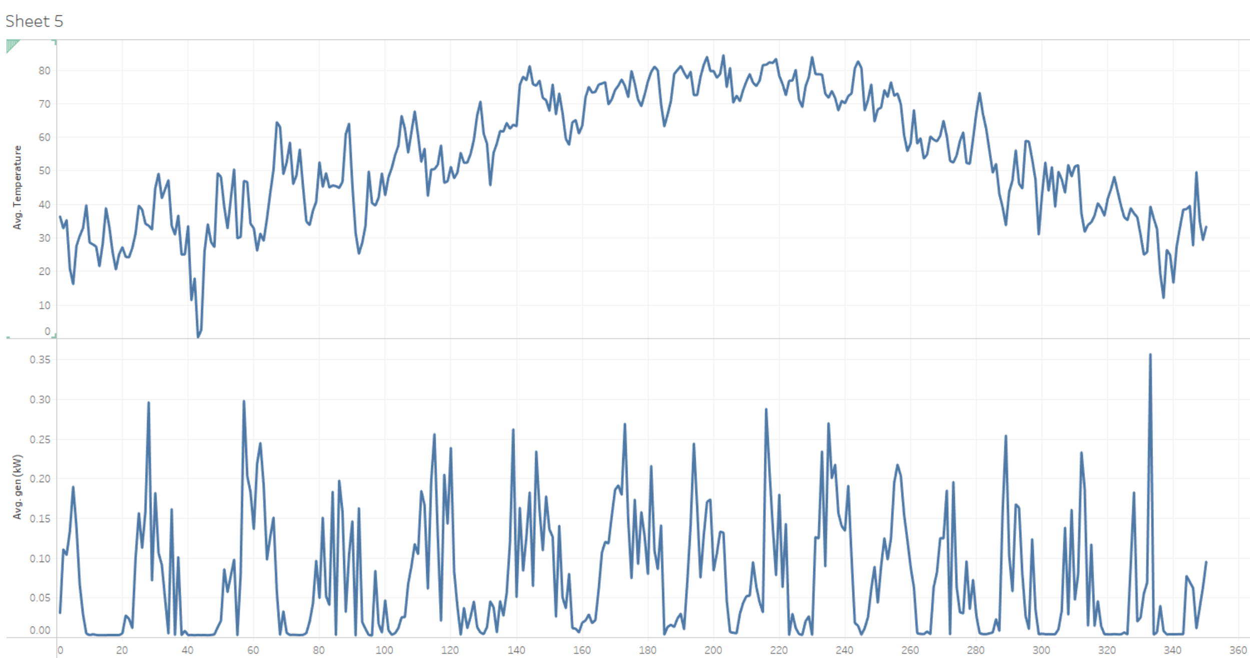
These are two different ways of visualizing the same information that as per our expectation from an average clear day, enough energy is not being generated. *There exists some inefficiency in this process.*

Chart# 4:



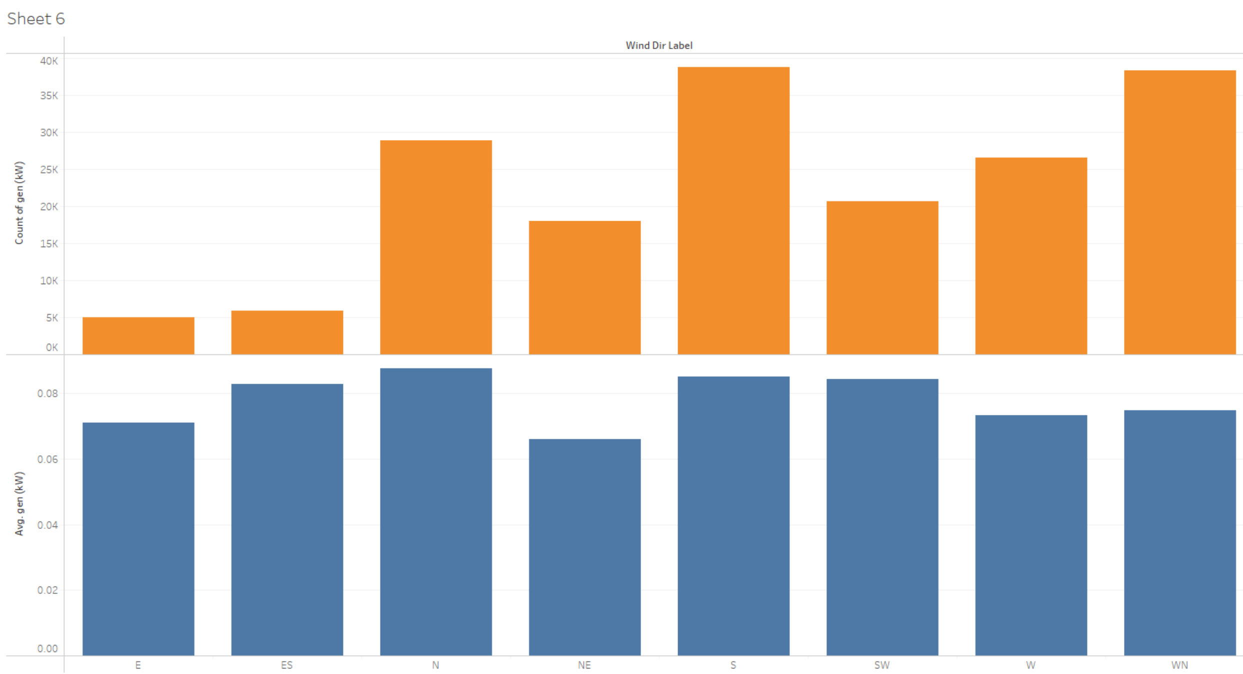
Compare for each weather icon, the average temperature, average energy generated per minute, and average wind speed. *The temperature and wind speed are good enough to generate energy on a clear day, but it's lacking behind.*

Chart# 5:



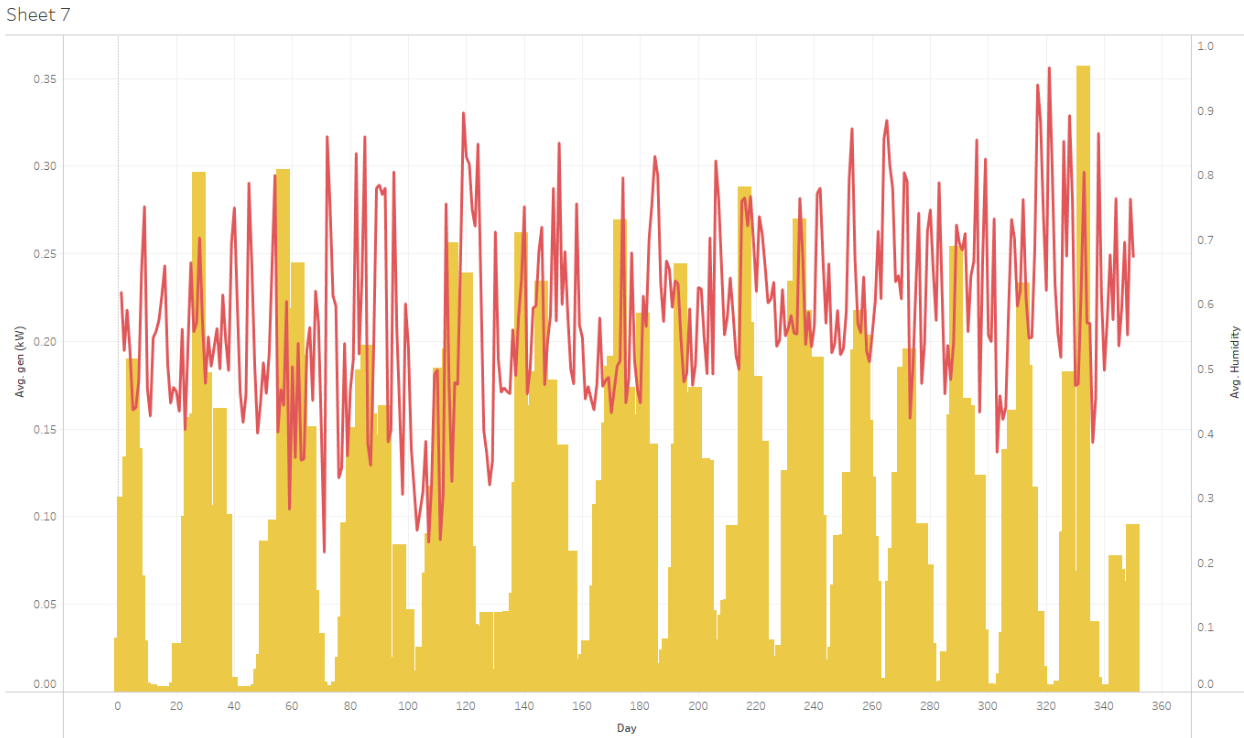
So we can see here that there are high fluctuations regardless of any changes in the overall daily average of temperature rises. *It therefore indicates an inefficiency in energy generation.*

Chart# 6:



As per the source, a southerly wind can increase the output of solar panels by up to 43%, but we see that *despite having a lot of southerly wind records, the average energy generation is not what it could have been.*

Chart# 7:



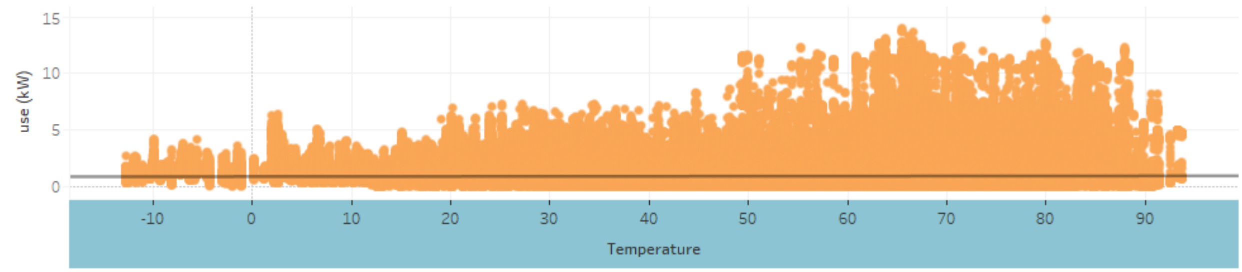
The histogram represents average energy generated per minute on a daily basis and the line represents average daily humidity. Keeping the source in mind that at low relative humidity, output current increases as well as power efficiency of solar modules hence solar electricity generation is high and vice versa, we see that spikes in humidity reduces the average energy generated many times in the observable days. But *there are some places where reduction in average humidity reduces the average energy generated.*

Conclusions on charts 1 to 7:

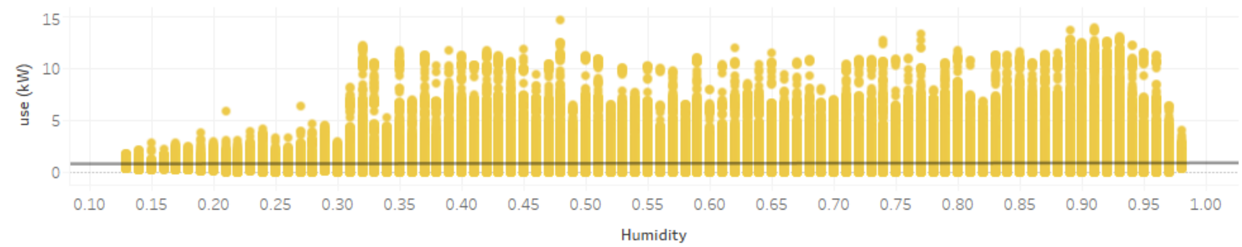
These were intended to figure out the problem in energy generation. And we've seen with visualizations from multiple dimensions (angles) that the energy generation is lacking. We most likely need to replace (or fix) the solar panels with effective ones.

Chart# 8, 9, and 10 combined:

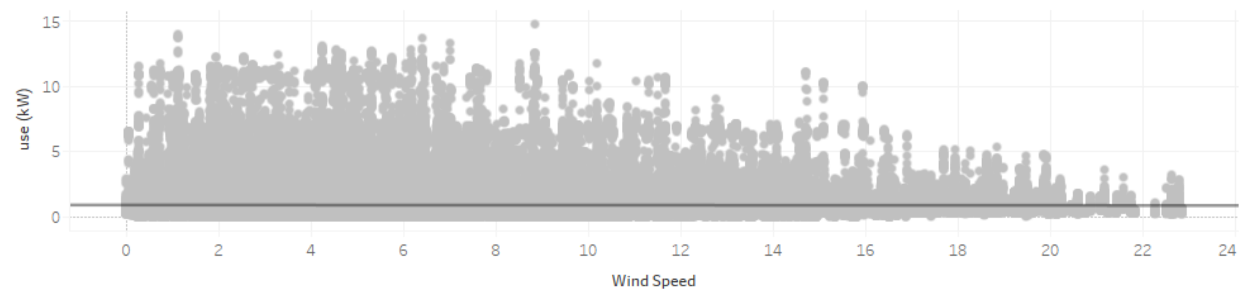
Sheet 8



Sheet 9

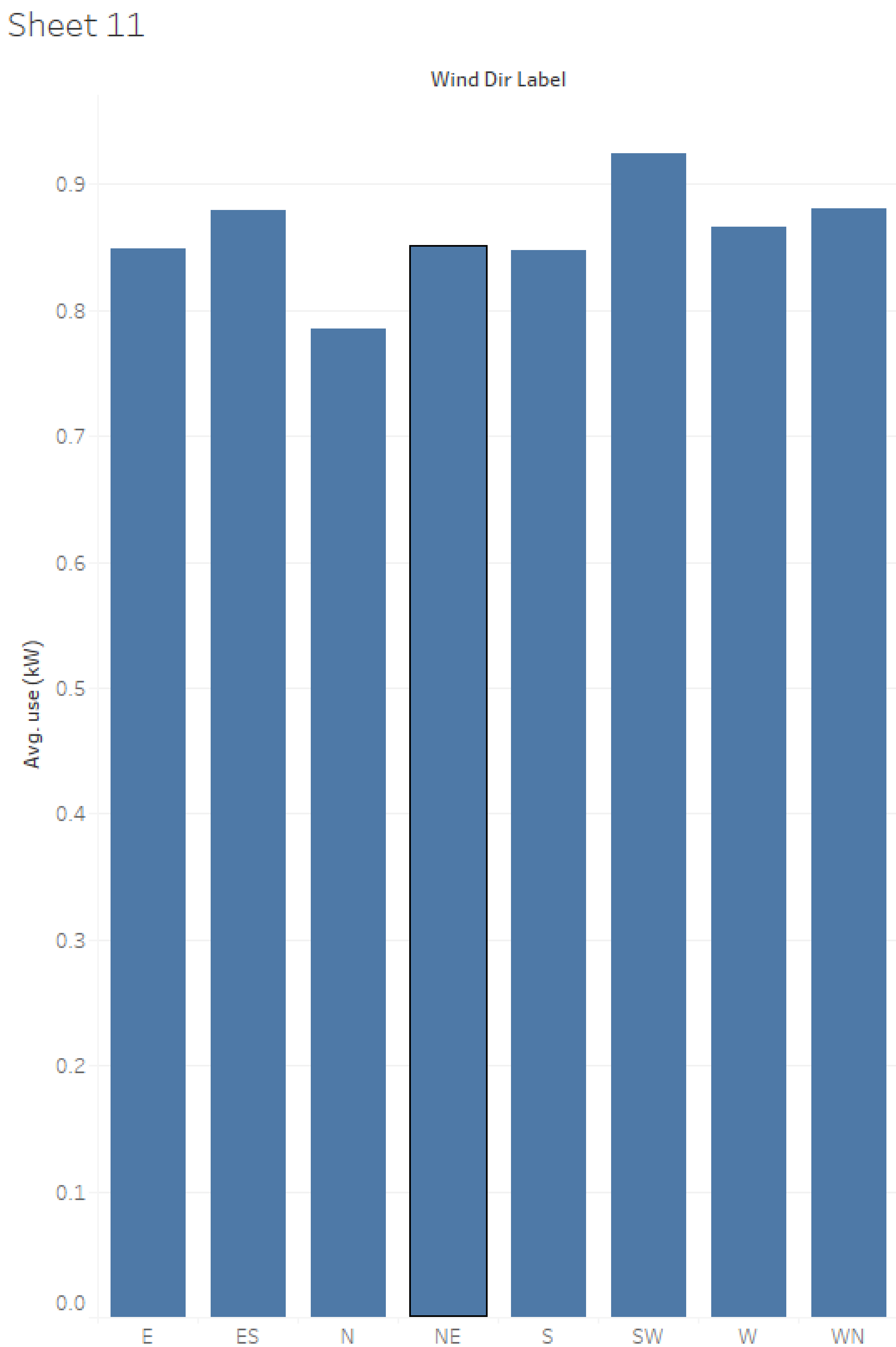


Sheet 10



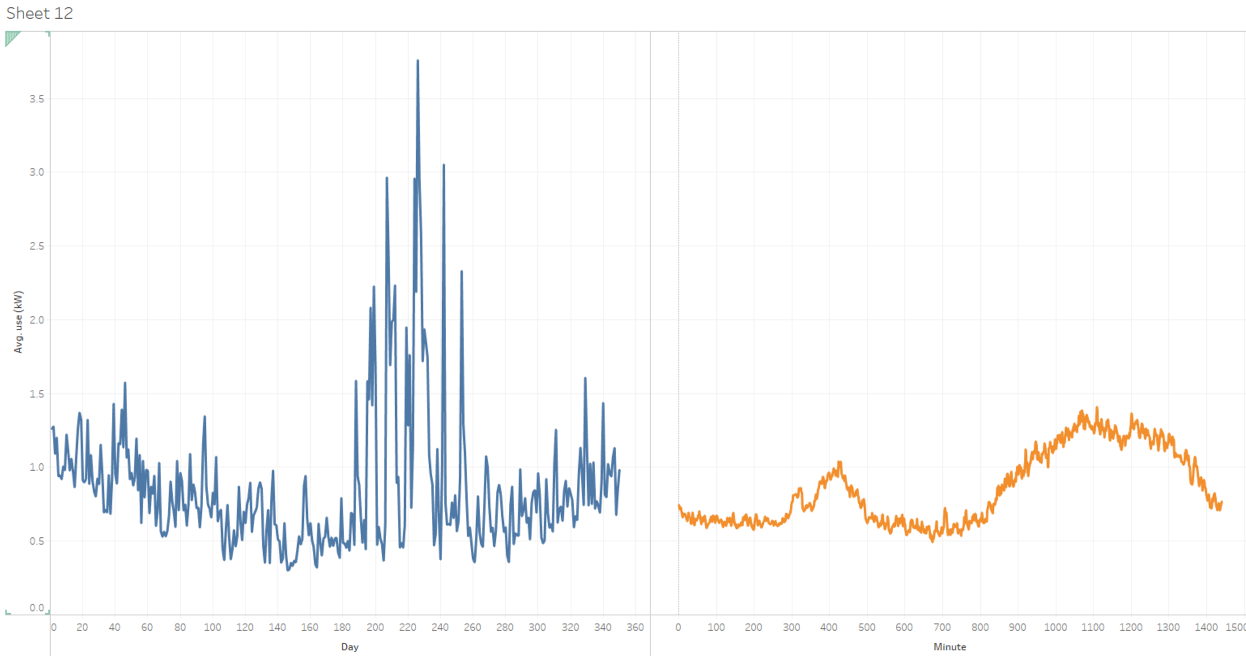
Shows that when wind speed, humidity, and temperature are independent variables, they've *no (or at least a very weak) correlation* with electricity usage.

Chart# 11:



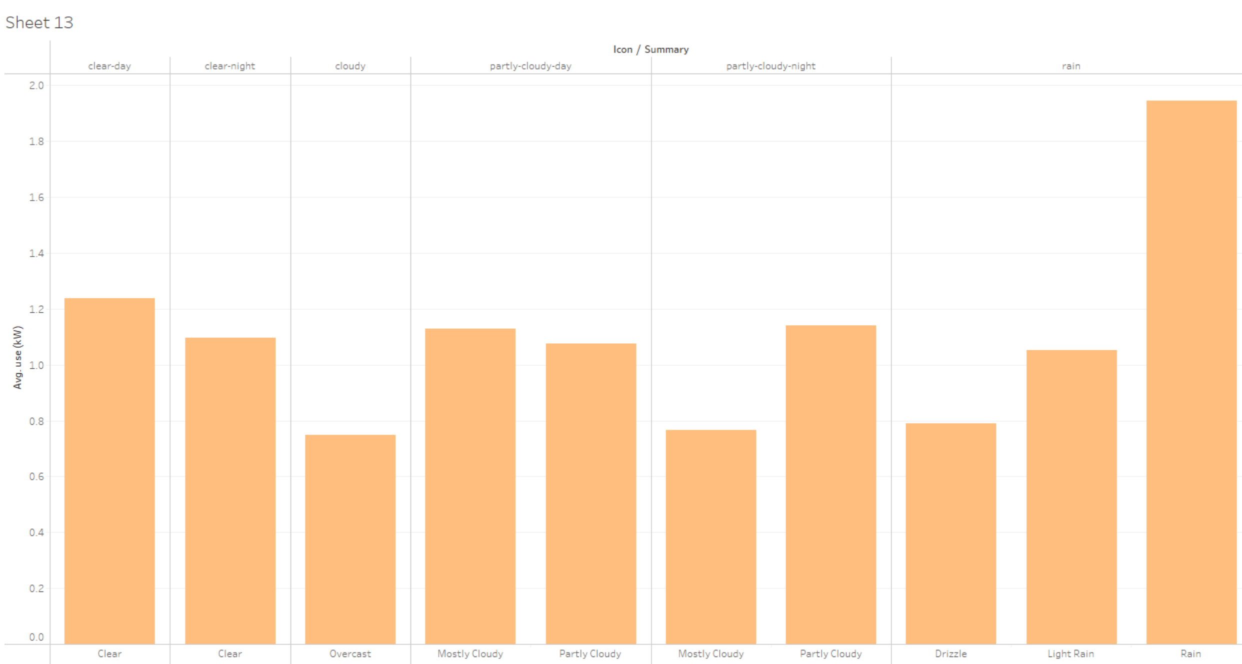
There's *no (or negligible) impact of wind direction on average usage* of electricity per minute.

Chart# 12:



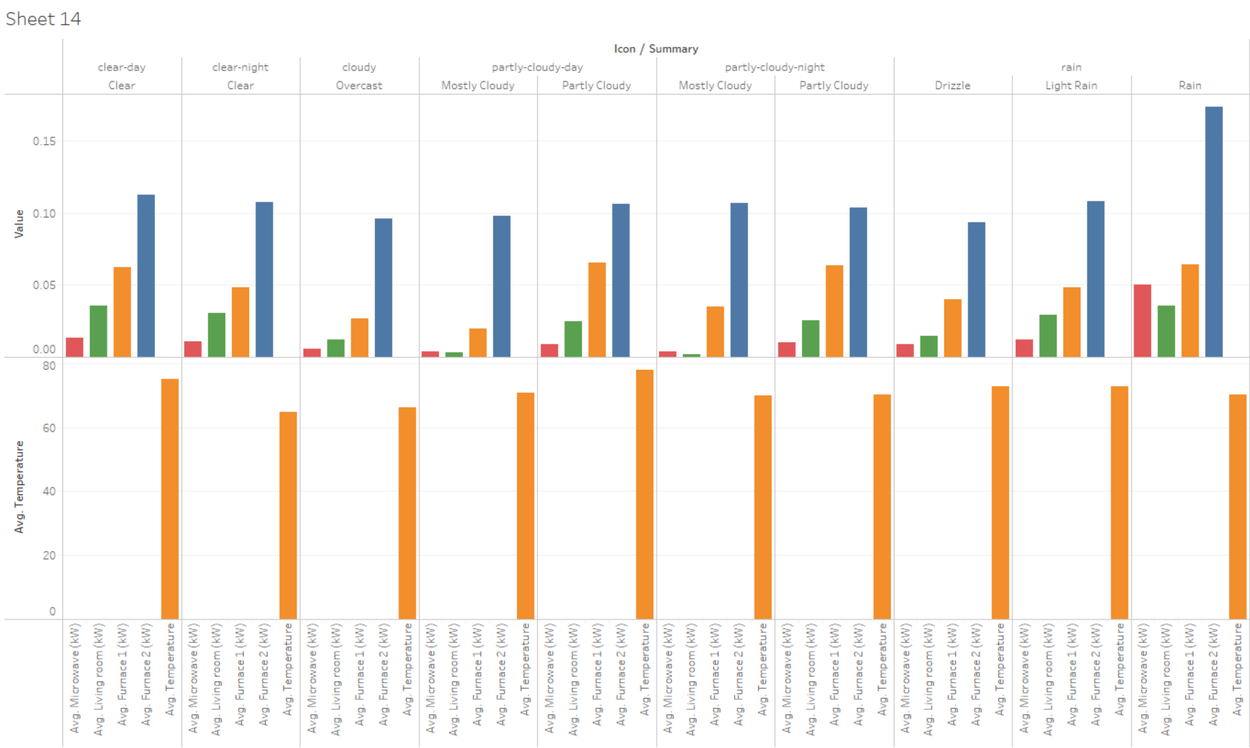
This overview look helped us find out that from *day# 180 to day# 260, there is a hype in average electricity usage*, and if we talk about any average day, then *from minute 800 (after 1300 hour) the average electricity usage starts to increase, and this lasts until evening.*

Chart# 13:



Shows a very *high average use of electricity when the weather was rainy.* For only those days, where there was hype of usage, according to what we found out in the previous chart.

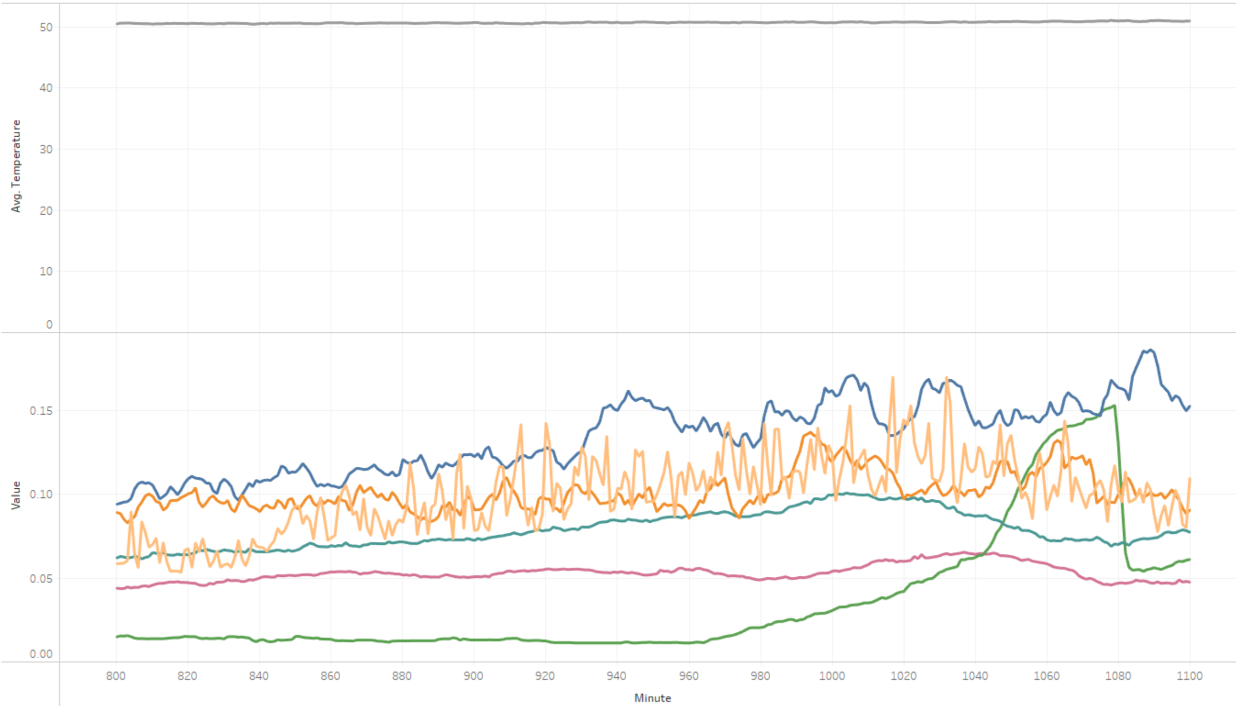
Chart# 14:



Finding out what appliances or use cases led to high average usage of electricity per minute, especially for rainy weather. *Found out that Furnace 2 was being highly used, even more than Furnace 1*. It is a zoomed in view for the previous chart, so it also applies to only those days where electricity usage was on hype in chart# 12.

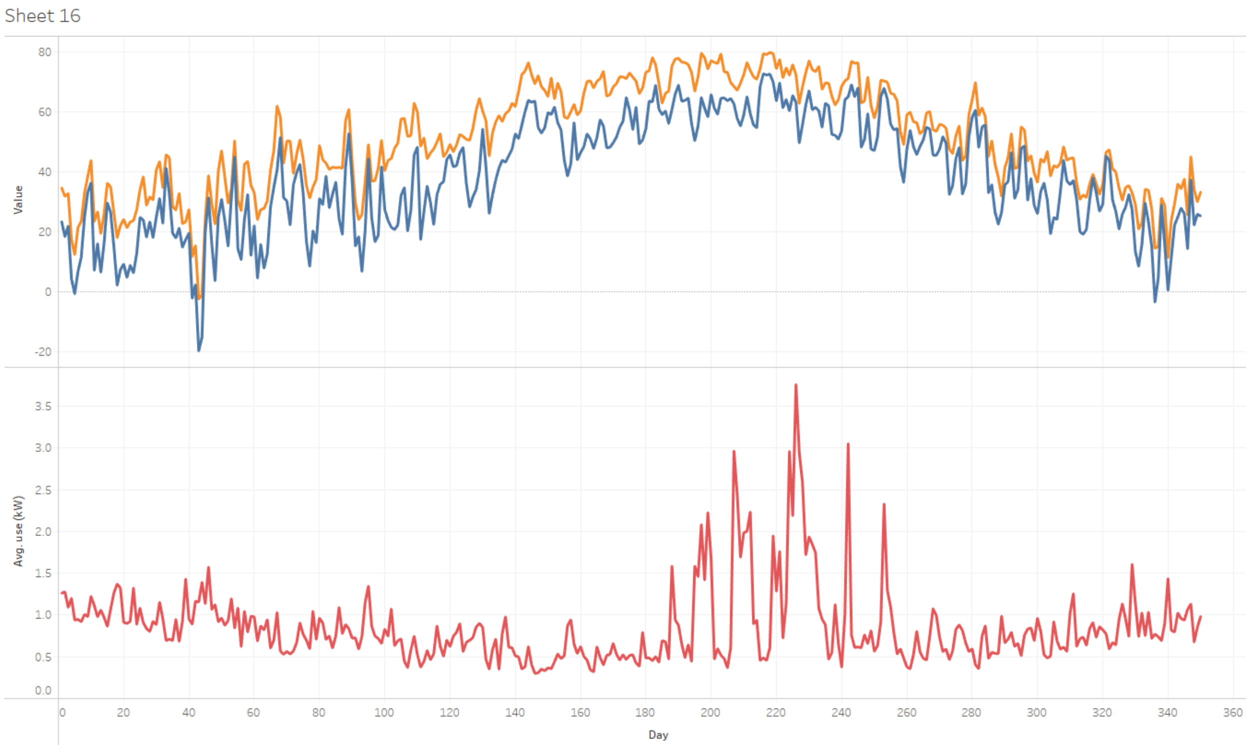
Chart# 15:

Sheet 15



A detailed look into average usage of top used items after 1300 hour till the end of the day. *And we found out that there was high electricity consumption in the living room as the end of day approached. Moreover furnace 2 was being highly used, despite the temperature being not too low.*

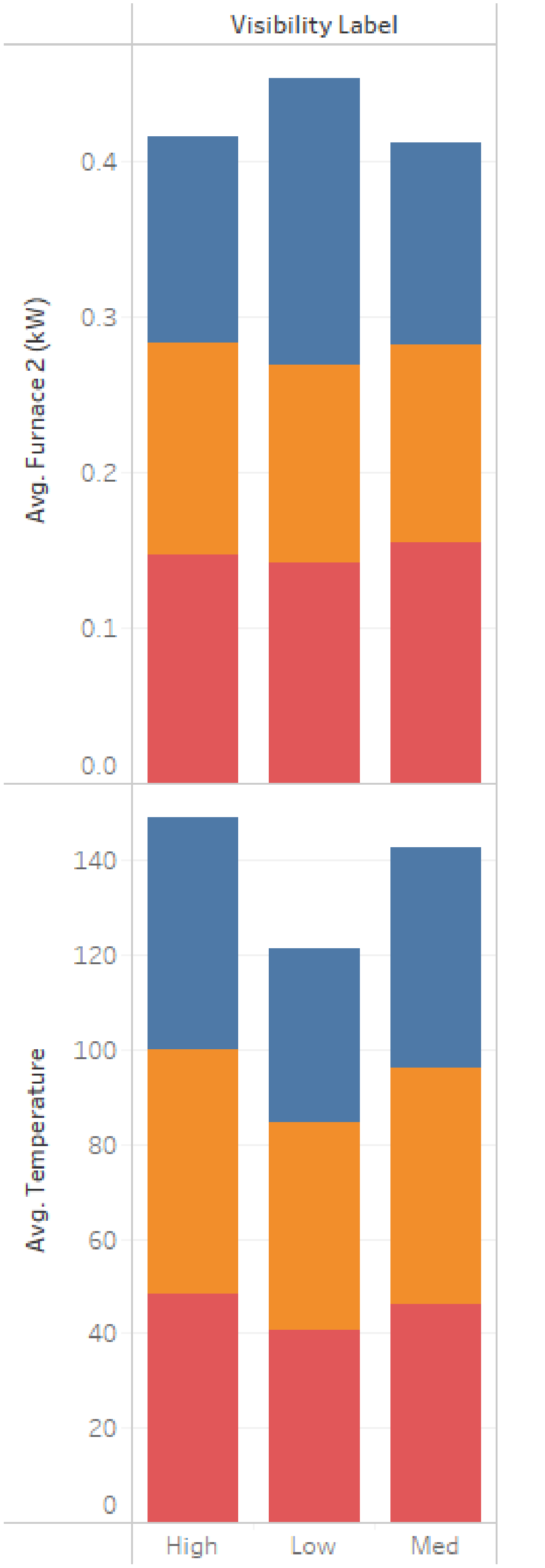
Chart# 16:



Compares the daily average usage of electricity against variations in dew point and temperature. The line in blue represents dew point temperature. There is an *overall trend that a rising dew point leads to rising usage of electricity*, which agrees with the source that a moderate rise in dew point temperature, electricity demand could increase up to 20 percent.

Chart# 17:

Sheet 17



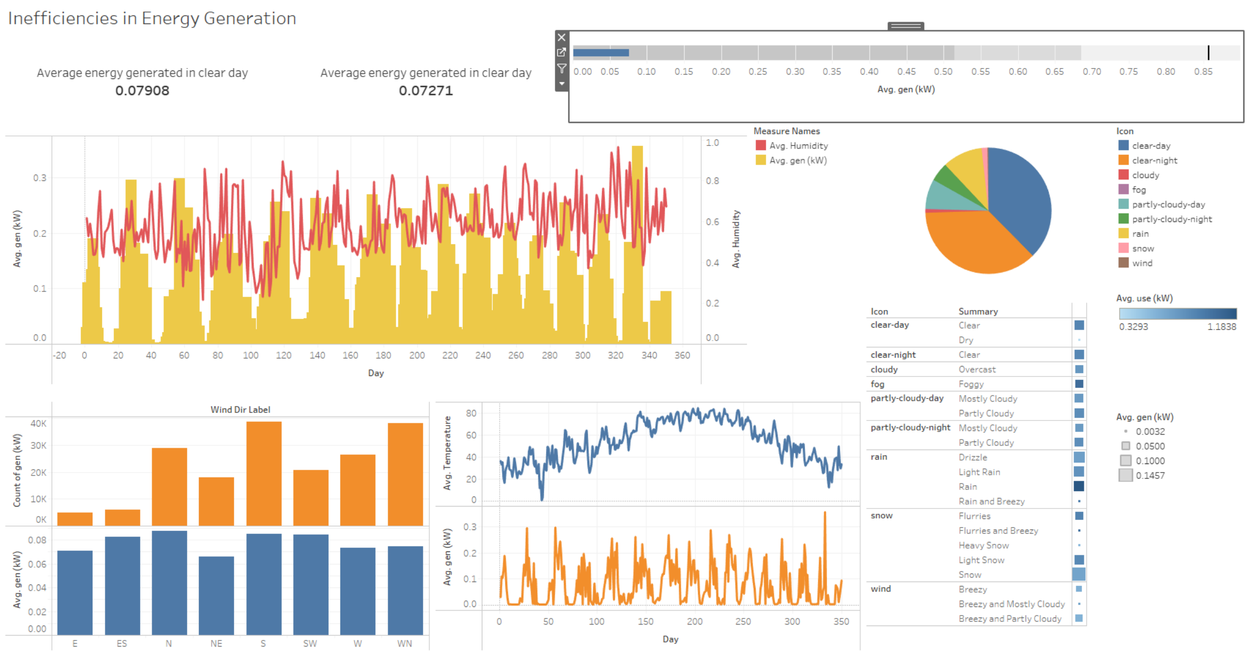
Visibility Label
Avg. Furnace 2 (kW)
0.4
0.3
0.2
0.1
0.0
Avg. Temperature
140
120
100
80
60
40
20
0
High
Low
Med

Shows the impact of different levels of cloud covers and visibility on average temperature and average usage of furnace 2 (since furnace 2 was one of the highly used and electricity consuming items). Blue represents high, yellow represents low, and red represents medium cloud cover level. For low visibility level, there is notably low temperature in Fahrenheit for each of the 3 categories of cloud cover levels. And the overall average usage of furnace 2 occurs when visibility is low. And the highest contribution in this usage is done by high cloud cover. So, *a high cloud cover and low visibility encourages use of furnaces.*

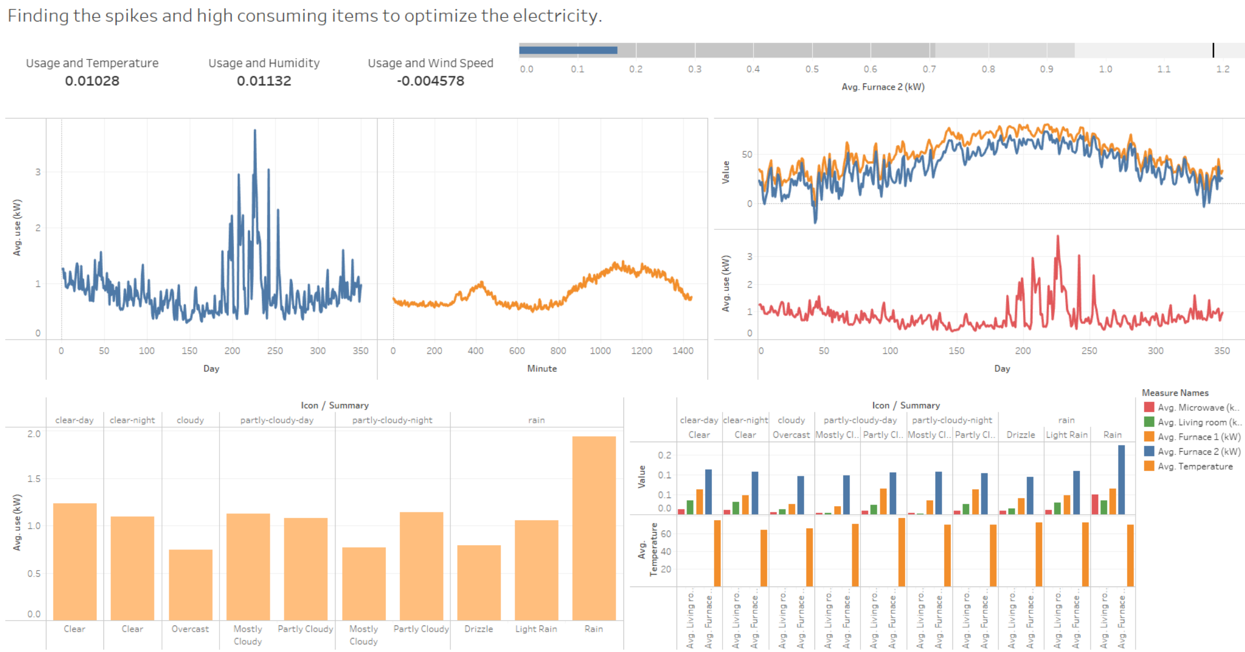
Conclusions on charts 8 to 17:

These were intended to figure out the problem in electricity usage. And we've seen with visualizations from multiple dimensions (angles) that there exist some inefficiencies in electricity usage. We need to encourage an alternative of furnaces. Dew point being one of the important factors in predicting demand of electricity, needs to be considered to decide electricity supply. More detailed analysis can be found at the end of the document.

Dashboards with summaries:



Based on the above dashboard, we have found out from multiple points of views that the energy being generated is not efficiently done. There's not much difference in energy generation in clear days and nights, hence clear days are not being leveraged to the extent they should be. According to the sources attached previously, the energy should be generated more in wind's southerly direction, but it didn't happen. On observation, there exist some areas where reduction in humidity reduces energy generation, which shouldn't be the case. And we can clearly see, despite a hike in temperature, energy generation still keeps fluctuating at the same rate. Hence no efficiency in this process. *The solar panels deployed need to be tested and fixed or replaced with the better versions.*



Based on the above dashboard, we see that electricity usage is not being impacted much by humidity, temperature, or wind speed. But it is dew point that gives a rise to electricity usage, which aligns with a research [source](#) that a moderate rise in dew point temperature, electricity demand could increase up to 20 percent. Thus, *dew point temperature needs to be taken into account to predict electricity demand for optimization in electricity supply.* Other than that, through the overlook of usage variations based on days and minutes, we found out from day# 180 to day# 260, there is a hype in average electricity usage, and if we talk about any average day, then from minute 800 (after 1300 hour) the average electricity usage starts to increase, and this lasts until evening. So, we dived in further detail to figure out what type of weather caused such an increase. We found out that electricity was used more in rainy weather. We then dived further to find out the items being used in such weather leading to an overall increase in electricity usage. We then found out the usage of furnaces. Furnaces seem to be inefficiently consuming the electricity, so we need to encourage users to move to better and efficient alternatives. Either, they can *use efficient furnaces or use boilers, heat pumps (source),* where boilers are highly efficient—energy efficiency ratings from 80 to an amazing 98 percent.

Note: Two stories have been created, one for each problem. They can be found in the tableau file. They contain the same sheets with sequence and explanatory captions, which will help understand the process of diving deeper into information to extract insightful analyses. They were not attached here, as they would make this document redundant and quite lengthy.