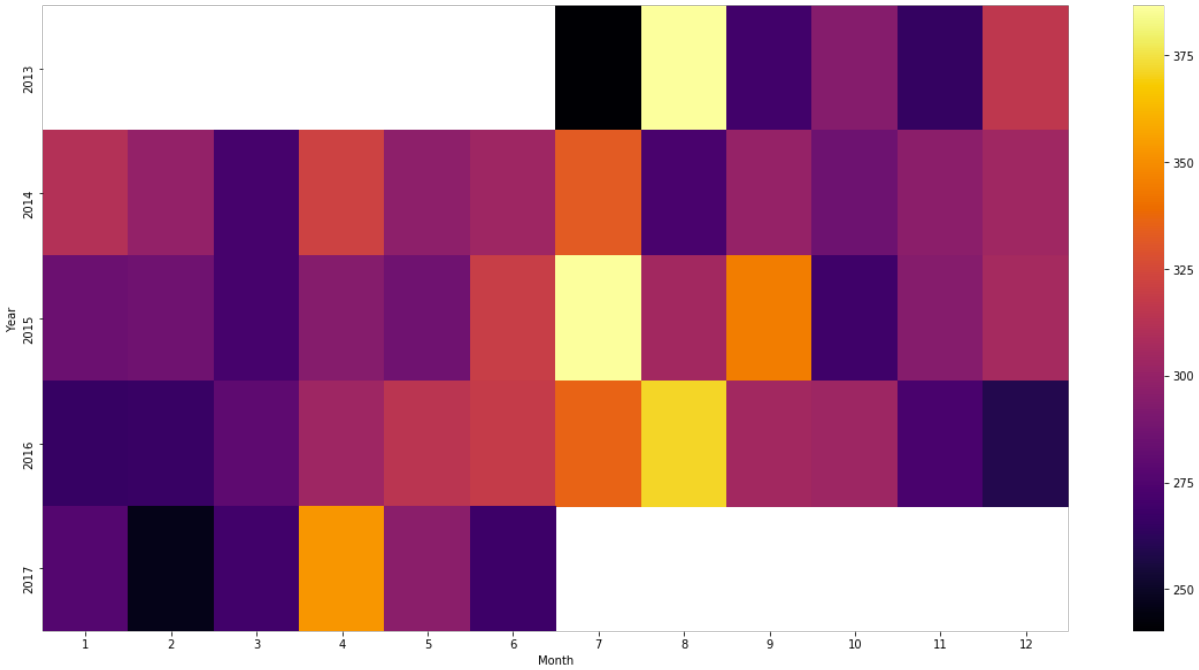
Analytics Vidhya Hiring Hackathon Code Rationale

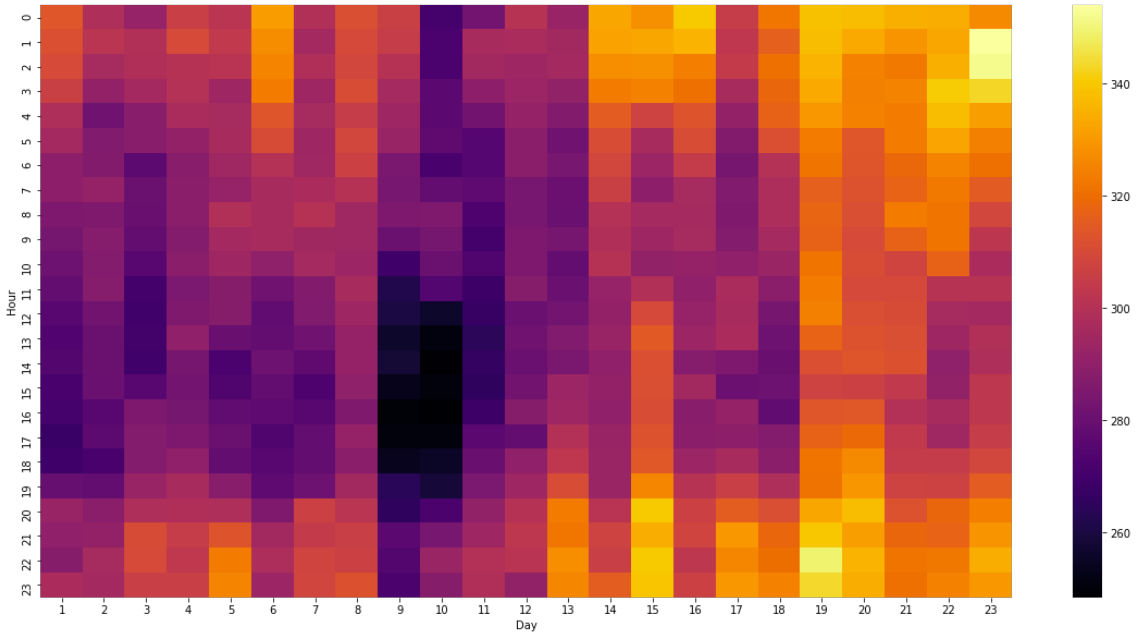
"The document contains minimalistic explanation to the problem. For the EDA code, please look at the notebook file in the folder"

Feature Engineering Date Time Features (1/3)

The Heat map from the EDA clearly shows the variations of the electricity consumption the Year and Month wise

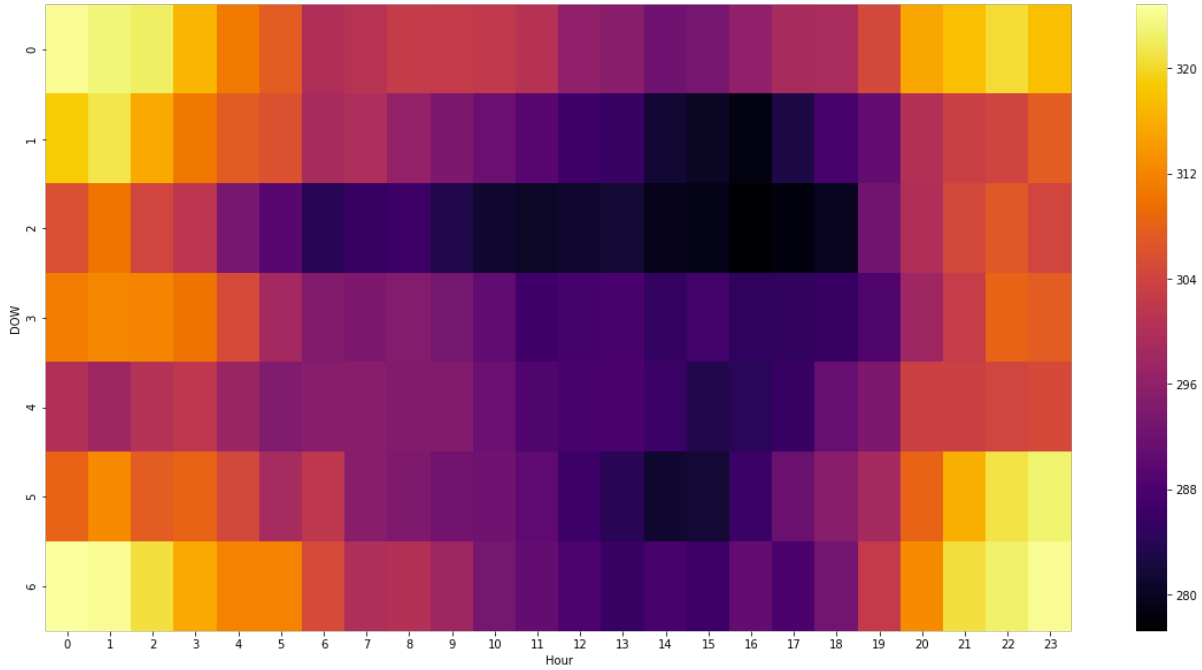


The similar heat map between the Hour and the Day reveals that the consumption peaks in the end of the month



The Hour and Day of Week (DOW) reveals another interesting information on the consumption at the weekends and holidays.

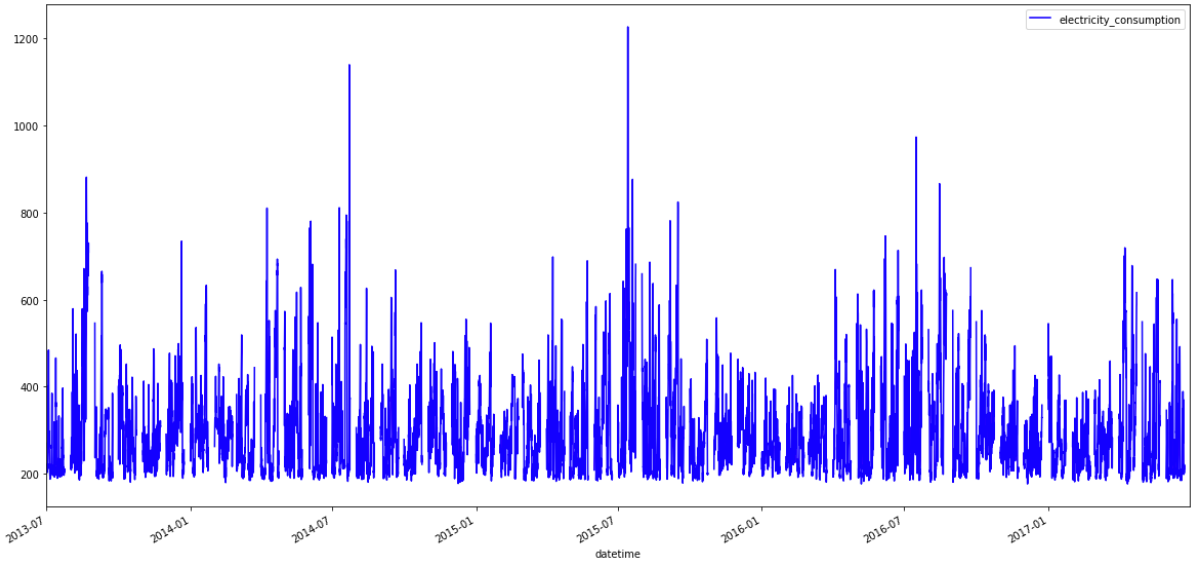
It would be awesome to use the holiday data along this, but per rules we are to avoid any external data



Surely, it safe to claim these features as importance factors in predicting the electricity consumption.

Feature Engineering: Group by Mean, Min, Sum and Max of the factor variables (2/3)

From the rolling mean of the electrical consumption. Leaving out the date time factors there are other complex features may have effect on the consumption. So, the Group by of various numerical features in terms of mean, min and sum.

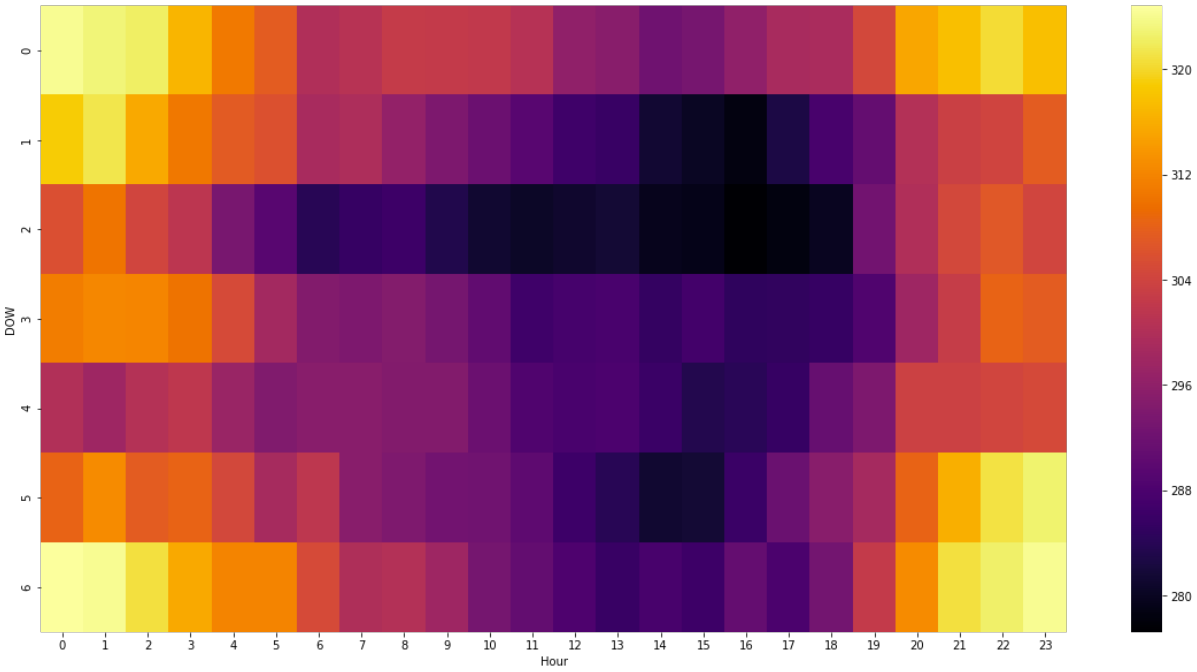


Please do refer the code script on the code part of this group by features

Feature Engineering: Time of the Day, Weekend Proximity, Holiday Flag, Relations with var1 (3/3)

Time of the day sounds like important factor too from the heat maps above. Let's add a feature that indicates the early, morning, afternoon and late-night feature to help our model understand the consumption better!

Also, let's add weekend – Saturday and Sunday holiday flag along the weekend proximity.



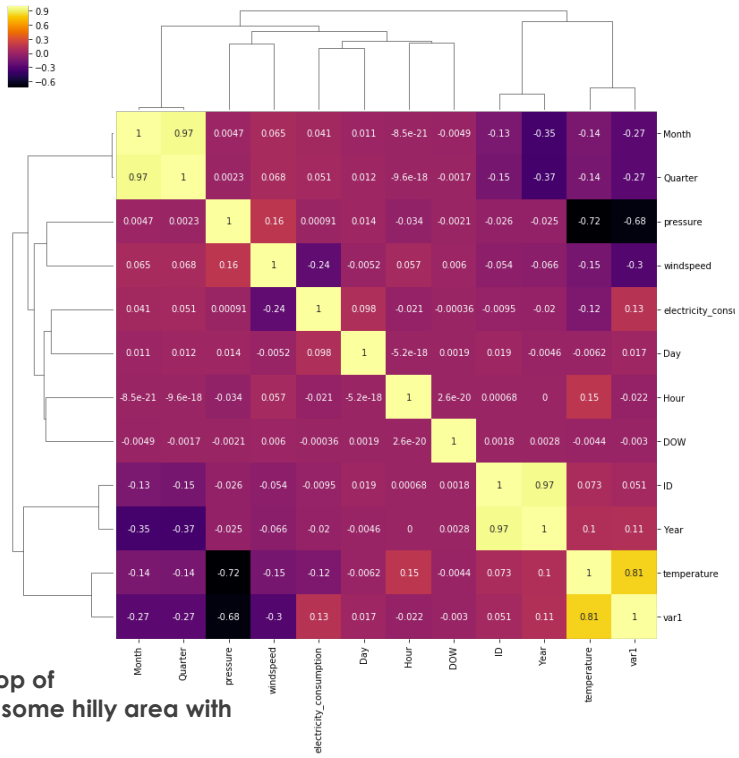

```python
# Functions for Daytimes and Is Weekend Feature
def DayTimes(Hour):
    if Hour in range(0,5):
        return "MidNight"
    elif Hour in range(5,7):
        return "Morning"
    elif Hour in range(7,17):
        return "Afternoon"
    elif Hour in range(17,24):
        return "Night"

def Weekend(DOW):
    if DOW in [0,5,6]:
        return 1
    else:
        return 0
```

Relations with var1

The first thing to notice in this beautiful color cluster map is the square grids of correlations. Clearly the **var1 and the **temperature** has direct correlation of 81%.**

As this is an anonymized variable, we cannot draw direct conclusion but a right question and explain this anonymized thingy!



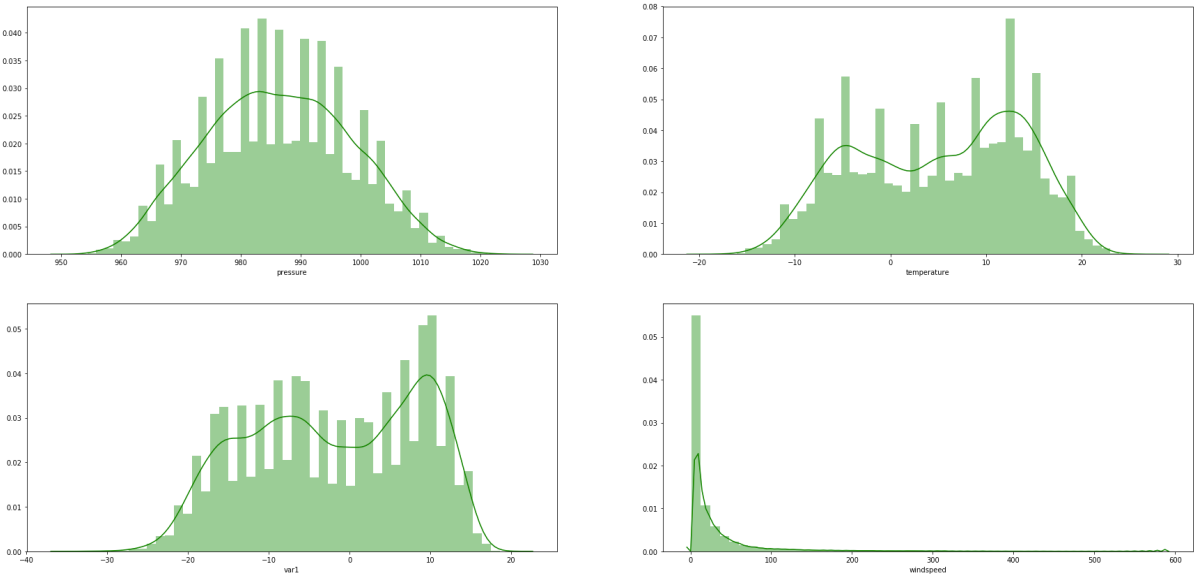
What Physical/Electrical properties are directly proportional to the temperature?

Things to take from the plot-

1. Remember I being physic professor in the last graph of some relation with var1 and temperature. Here we can safely conclude that it isn't **resistance, or voltage as it varies from - to + values. It can be some very complex - resistivity or sensitivity related too.**

2. The varying -10 to +20 of both var1 and temperature are the clear signs of them being some temperature values in Celsius scale and 15 degrees being max makes it more obvious.

3. Wind speeds are moderate, so we are not on the top of Himalayas here. So, this might be some hill station or some hilly area with normal pressure distribution



Hence, the solution is included with possible relations of various direct and inverse var1 features.

The CV strategy:

The LOCALTEST () function includes a common with forward time split with log1p conversion of the target variable for better score. The scores were positive as model RMSE score was improving as we move forward. This was extensively covered while modelling of the problem

Thanks AV Team for hosting and conducting such awesome Hackathons!

~Shaz13