Capstone Project - 2

Appliance Energy Prediction

Team Member

1. Pratik D. Jori
2. Pritesh Tambat

Introduction

- In this new era ability to predict energy consumption of House appliances not only save the money for consumer but also help in generating money for the user by giving excess energy back to Grid(in solar panel usage). In this case regression analysis will be used to predict Appliance Energy Consumption based on data collected from the sensors.

- The energy prediction will be come under supervised machine learning. Aiming to Appliance Energy consumption for house based on factor like temperature, humidity, pressure. Gradient descent algorithm and linear regression applied to predict the energy consumption.

Problem Statement

- The energy consume by home appliances are significant amount of the owner's budget. Consumers are looking to improve the energy efficiency and reduce the costs.
- Demand of the electricity is increasing significantly as the world population has growth over 8 billion.
- The objective is to build a predictive model which will help to predict the total consumption of energy from home appliances.
- Appliance energy consumption based on factor like temperature, humidity, pressure, many techniques gradient descent algorithm, and linear regression have been applied to predict the energy consumption.
- Predict the electricity consumption heating and cooling appliances in household based on internal and external temperature and other weather condition.

Points for Discussion

- Data Summary
- Mean Energy Consumption Per Day Of Week
- Mean Energy Consumption Per Hour Of The Day
- Energy Consume Consumption Distribution
- Heatmap & Correlation Matrix
- Feature Importance
- Fitting Models
- Future Work scope
- Conclusion

Data Summary

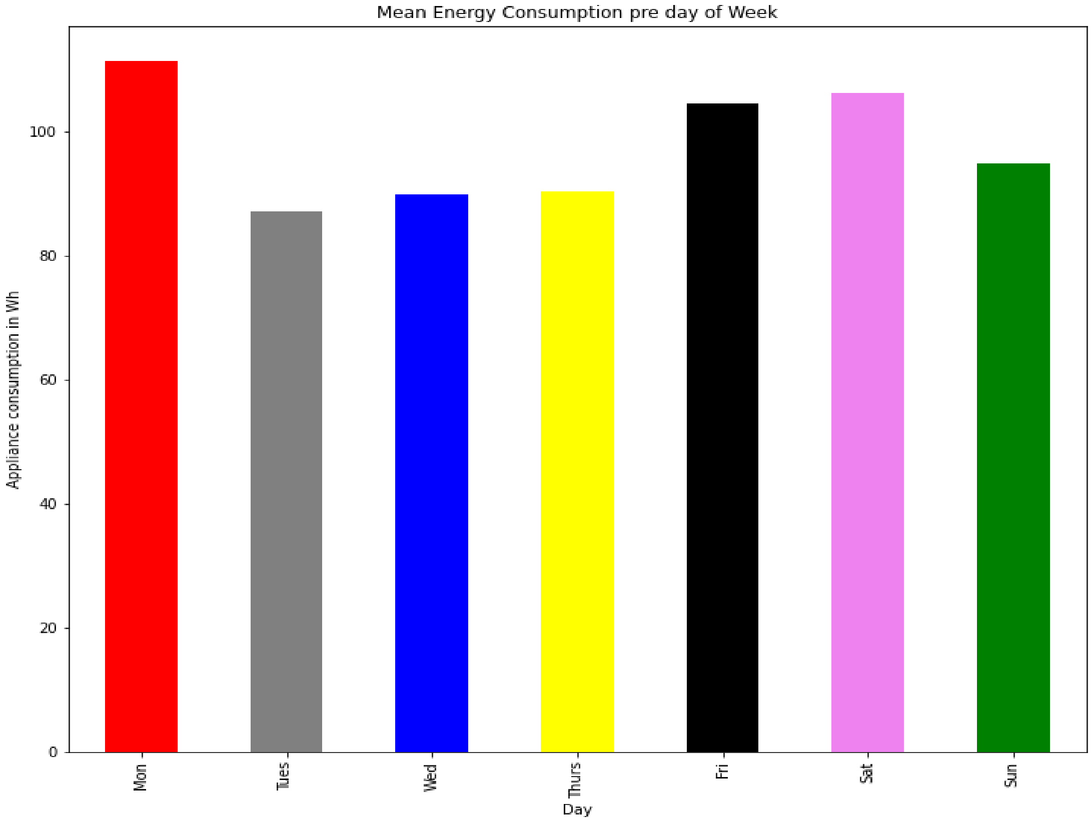
- date :- time year-month-day hour:minute:second
- lights :- energy use of light fixtures in the house in Wh
- T1 :- Temperature in kitchen area, in Celsius
- T2 : Temperature in living room area, in Celsius
- T3 : Temperature in laundry room area
- T4 : Temperature in office room, in Celsius
- T5 : Temperature in bathroom, in Celsius
- T6 : Temperature outside the building (north side), in Celsius
- T7 : Temperature in ironing room, in Celsius

- T8 : Temperature in teenager room 2, in Celsius
- T9 : Temperature in parents room, in Celsius
- T_out : Temperature outside (from Chievres weather station), in Celsius
- Tdewpoint : (from Chievres weather station), Â°C
- RH_1 : Humidity in kitchen area, in %
- RH_2 : Humidity in living room area, in %
- RH_3 : Humidity in laundry room area, in %
- RH_4 : Humidity in office room, in %
- RH_5 : Humidity in bathroom, in %
- RH_6 : Humidity outside the building (north side), in %

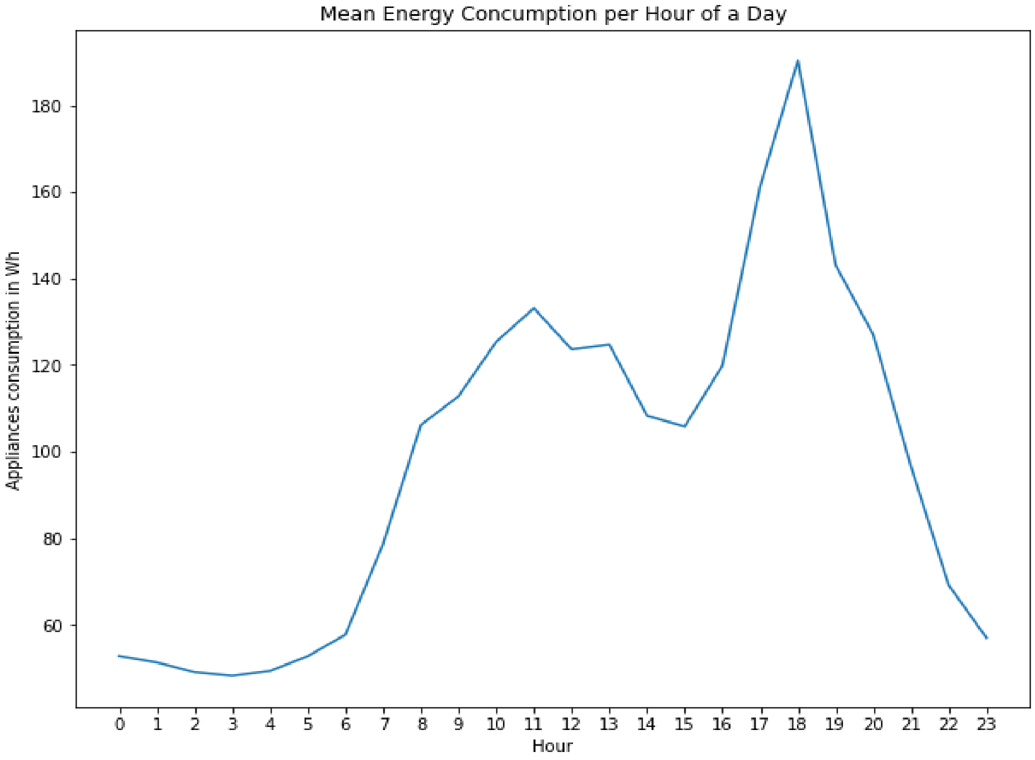
- RH_7 : Humidity in ironing room, in %
- RH_8 : Humidity in teenager room 2, in %
- RH_9 : Humidity in parents' room, in %
- RH_out : Humidity outside (from Chievres weather station), in %
- Press_mm_hg : (from Chievres weather station), in mm Hg
- Windspeed: (from Chievres weather station), in m/s
- Visibility :(from Chievres weather station), in km
- rv1 :Random variable 1, non-dimensional
- rv2 :Random variable 2, non-dimensional
- Appliances : Total energy used by appliances, in Wh

EDA

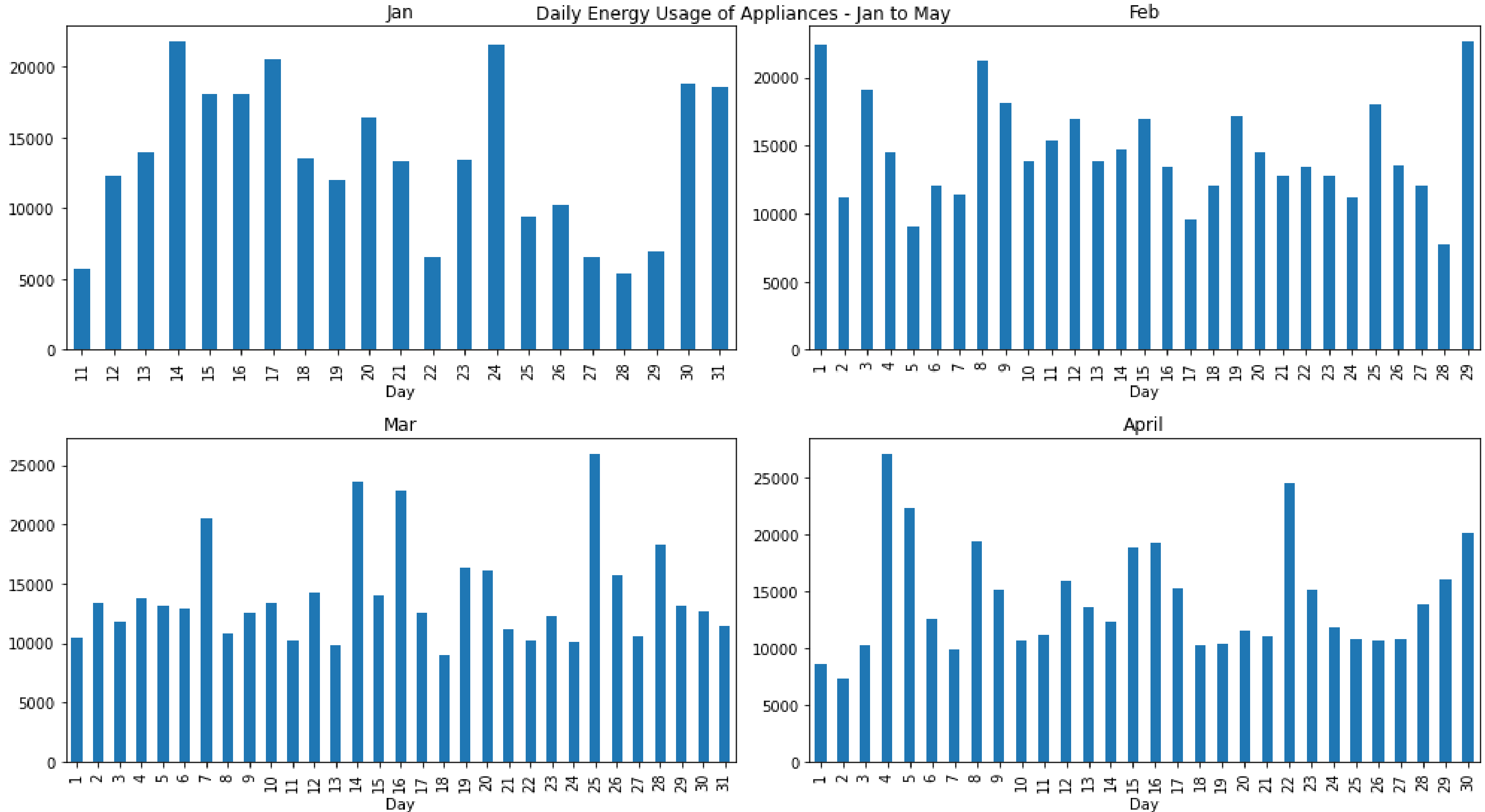
Mean Energy Consumption Per Day Of Week

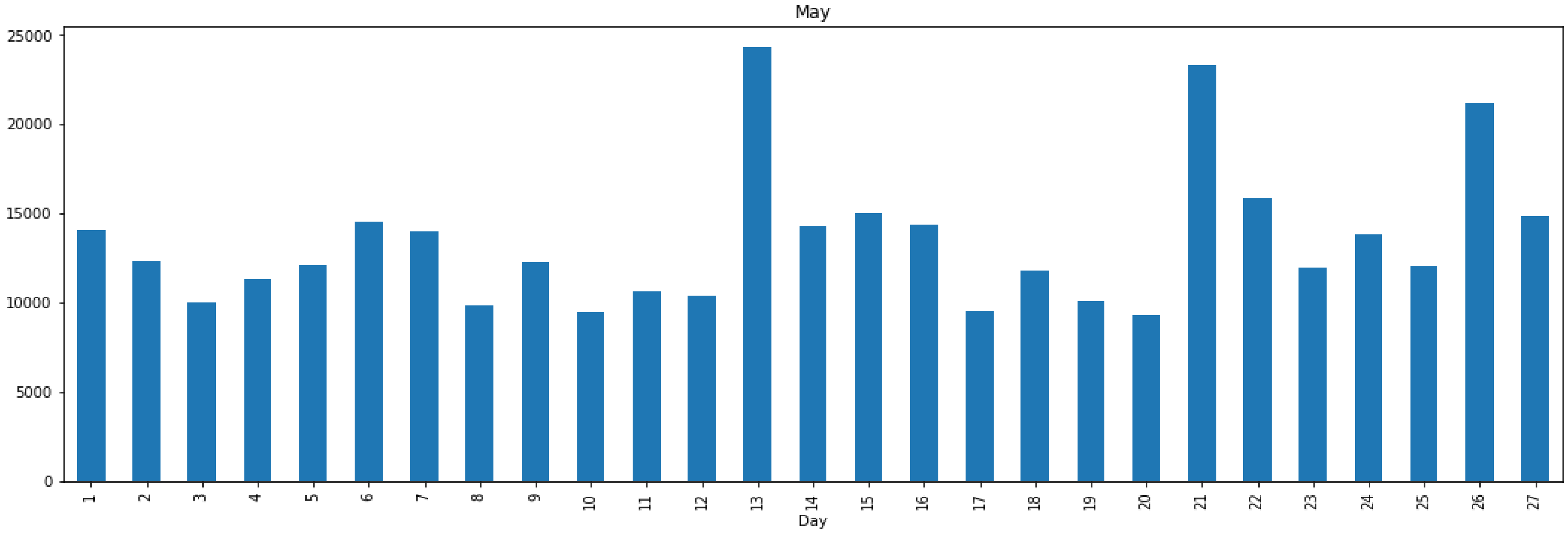


Mean Energy Consumption Per Hour Of The Day

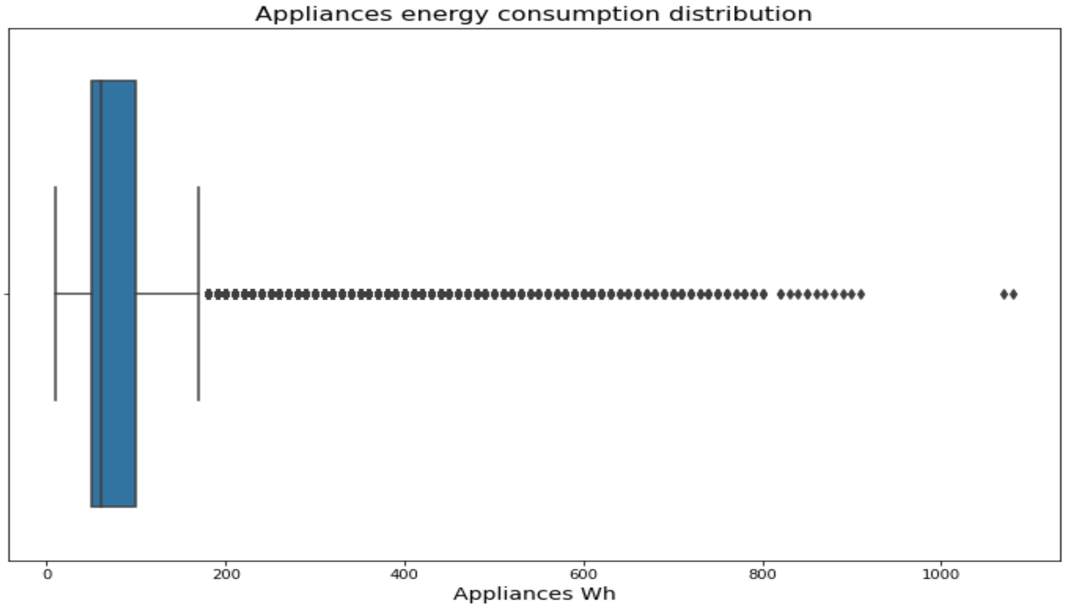


Daily Appliance Energy Uses From Jan To May



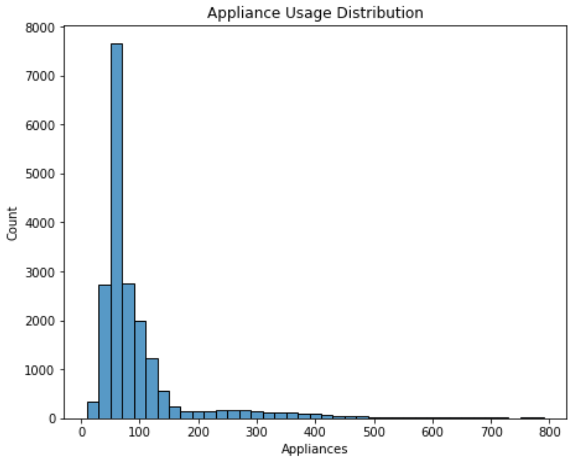

May
25000
20000
15000
10000
5000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
Day

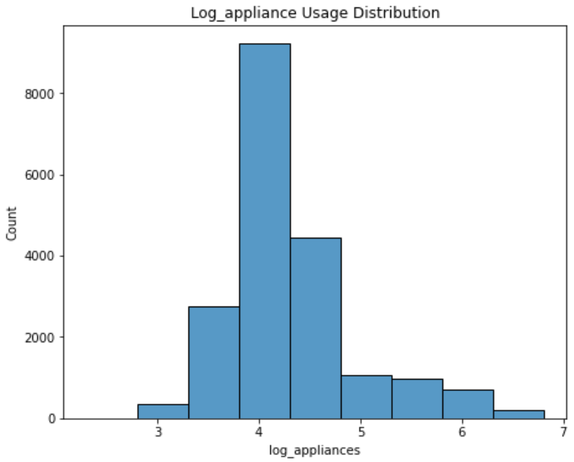
Outlier For Appliance Energy Usage



The number of the top values of appliance load is 19 and they have power load higher than 790Wh. So, we remove the instances above 790Wh.

Appliance Energy Consumption Distribution





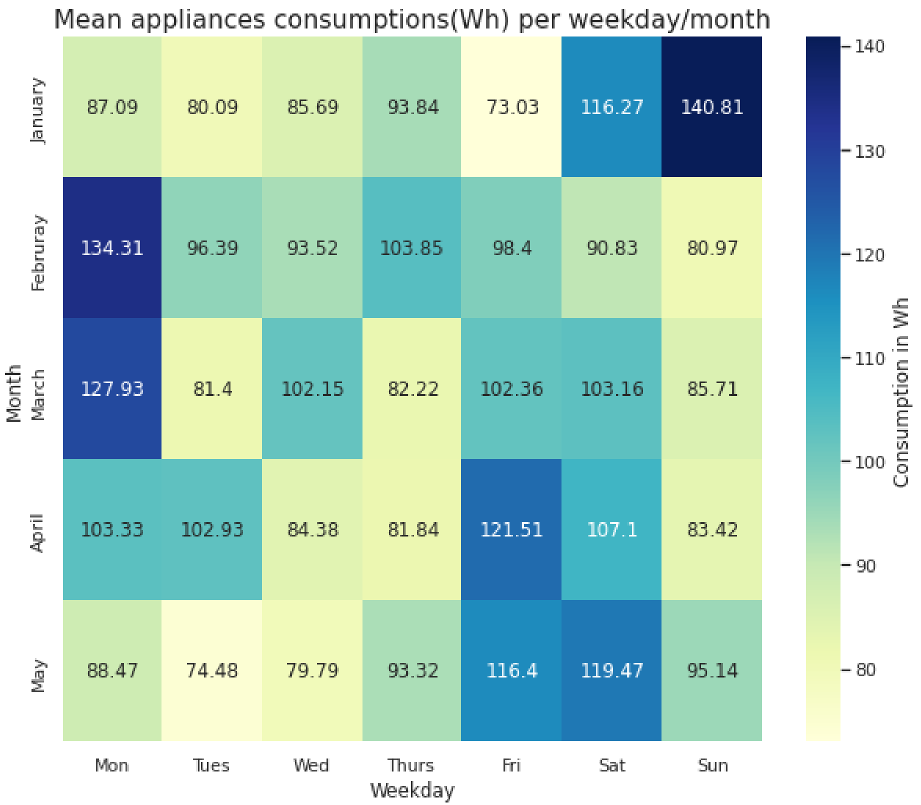
Appliance hist plot of time series, it shown seasonality component is decreasing it may suggest an exponential decrease from season to season. A log transform can be used to flatten out exponential changes to linear relationship.

We are looking at the mean and variance, we are assume that the data conform Gaussian(bell or normal curve). Appliance log histogram plot from the time series the ball curve-like shape of the Gaussian distribution, perhaps with longer right tail

Heatmap

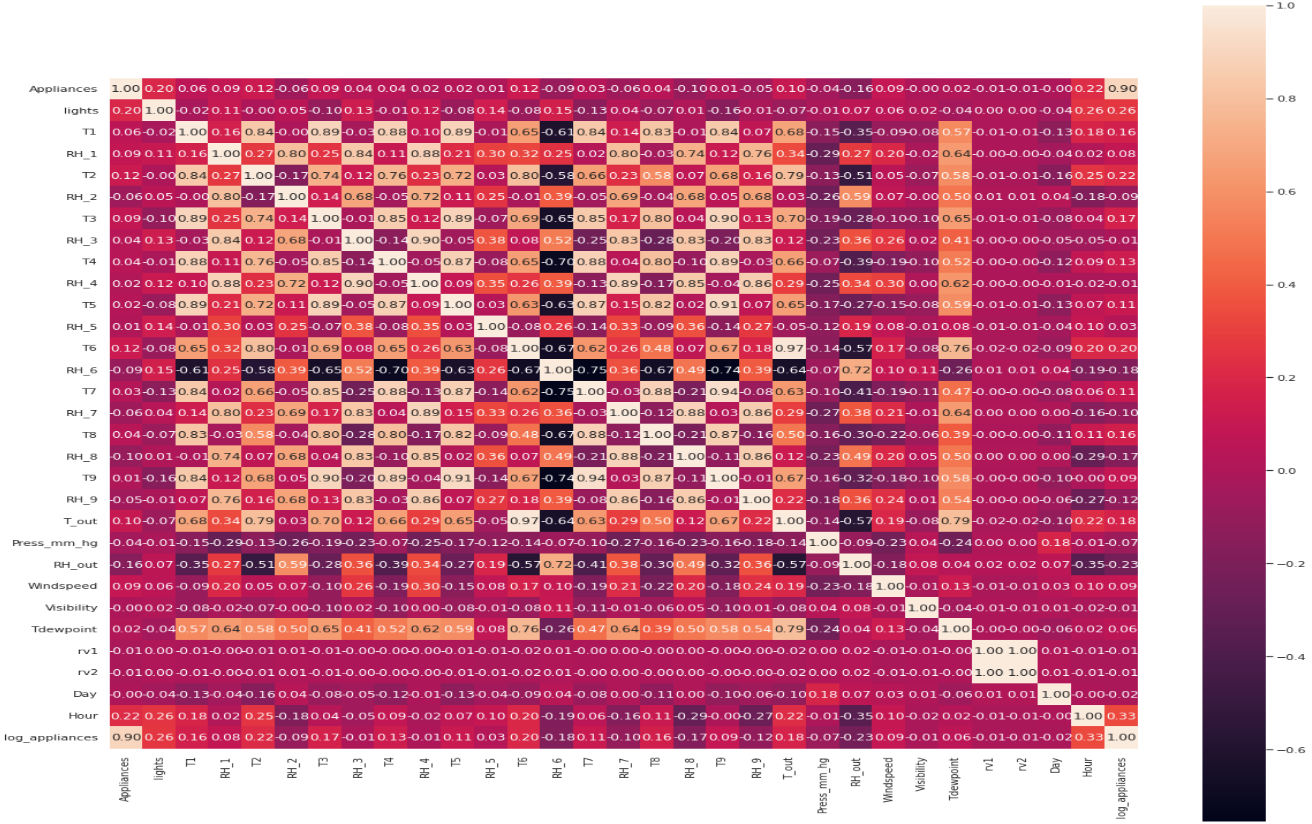
The more power is consumed on Monday, Friday and Sunday is valid for each month.

In this dataset we have only 4.5 months and from that we can't use months as feature for our model.

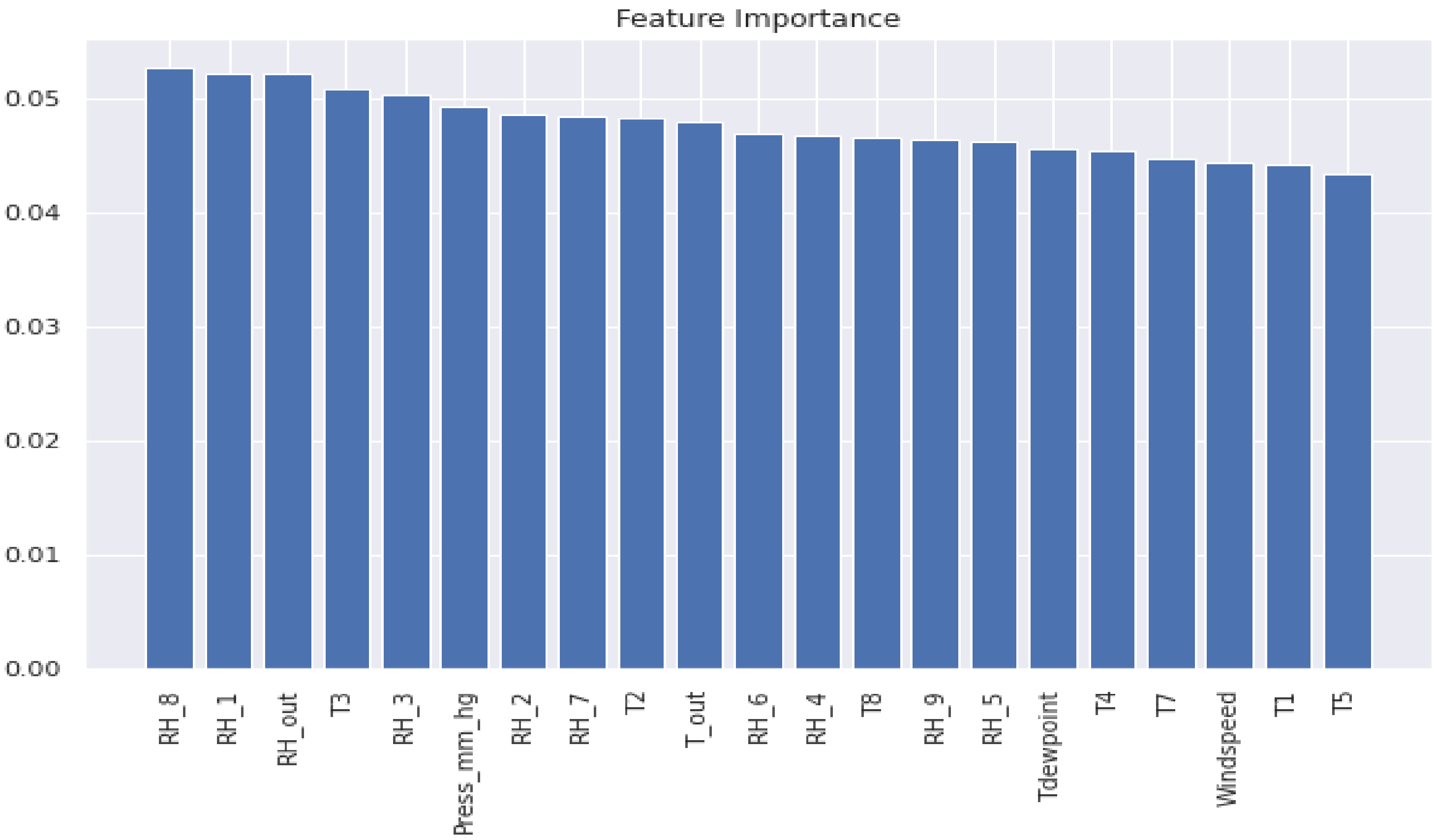


Correlation Matrix

From this we examine linear dependence among some basic feature in this dataset. In linear regression problem only linear independent variables can be used as feature to explain consumption in other way we will have multicollinearity issue.



Feature Importance



Machine Learning Models

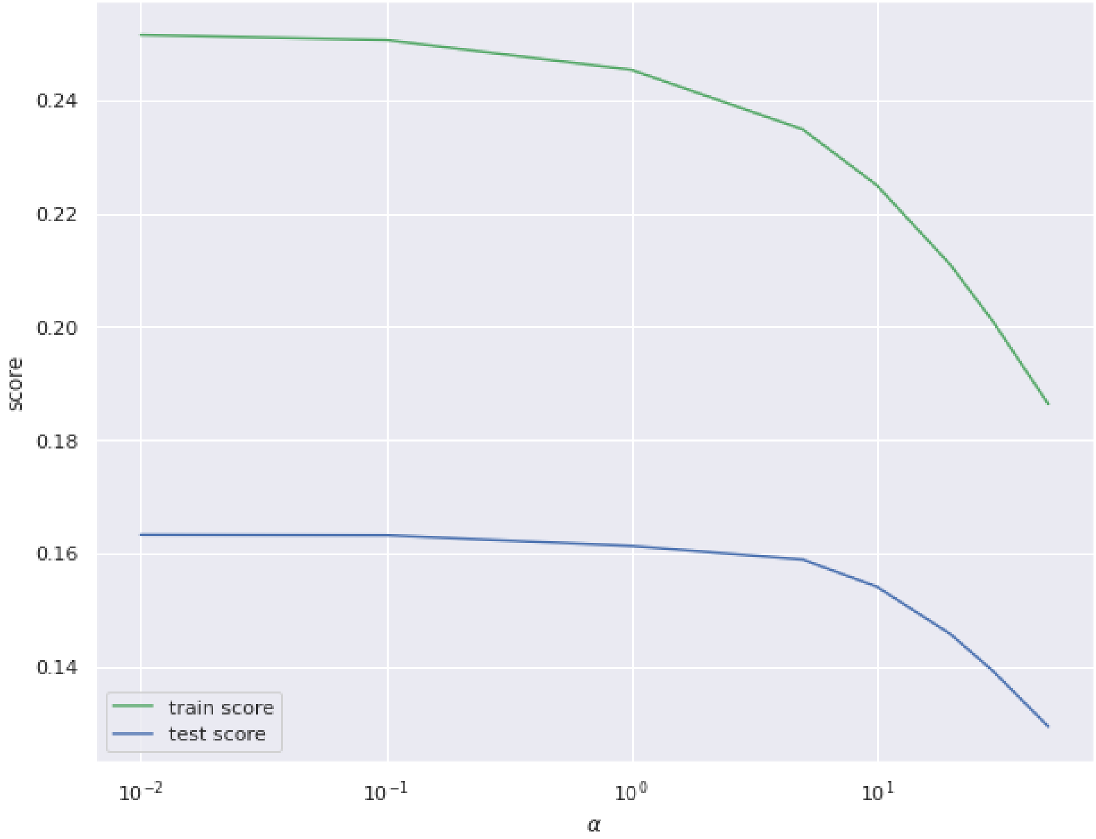
Linear Regression

- Train score for linear regression model without using kfold validation is 0.248513.
- Train score and Test score for linear regression using 10fold cv is 0.1631 and 0.01484 respectively.
- RMSE for linear regression is -12934.4967

Ridge Regression

- Train score and Test score for Ridge regression using 10fold cv is 0.22562 and 0.15188 respectively.

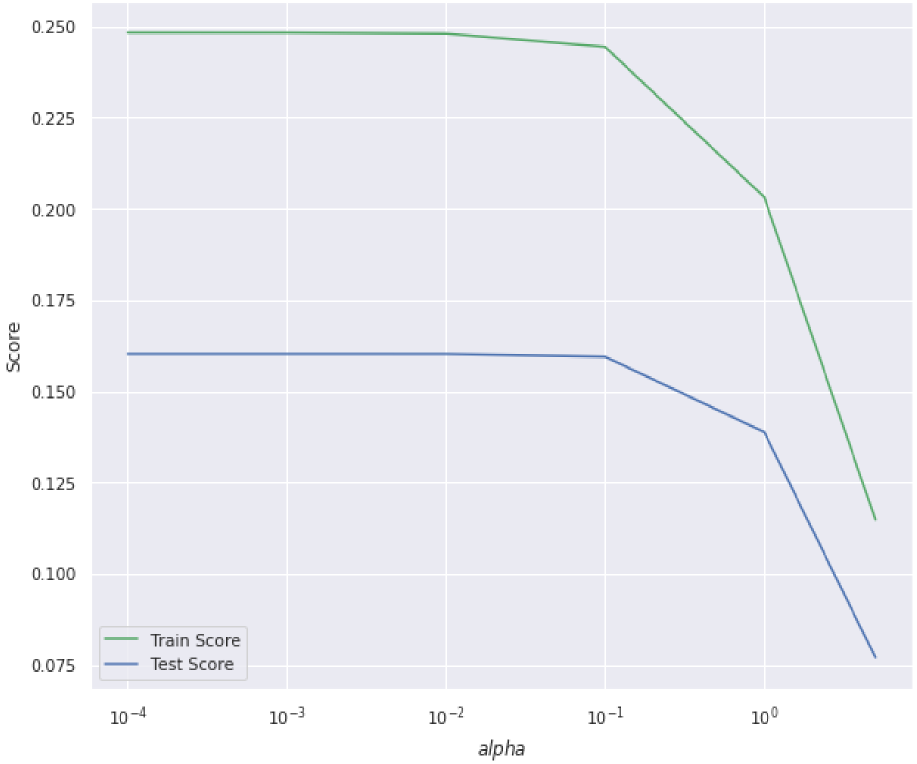
- RMSE of Ridge regression for train is -12934.0826 and for test is -16290.2937.



Lasso Regression

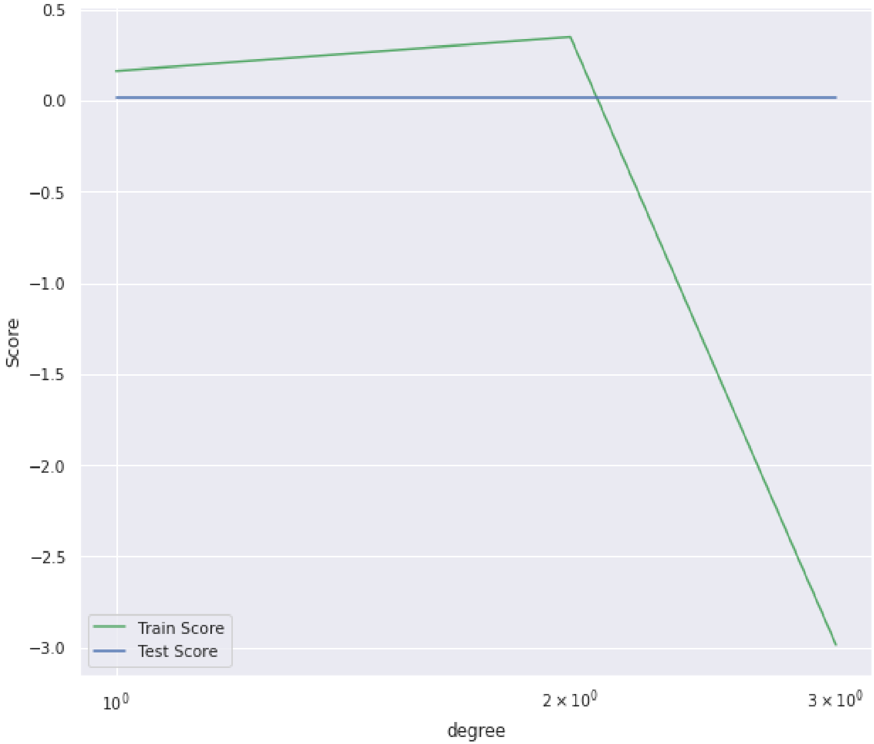
- Train score and Test score for Lasso regression using 10fold cv is 0.21785 and 0.14266 respectively.

- RMSE of Ridge regression for train set is -12986.4135 and for test is -16415.7130



Polynomial Regression

- Train score and Test score for Polynomial regression using 10fold cv is -0.82425 and 0.01484 respectively.
- RMSE of Ridge regression for train is -9285.0456 and for test is -16302.6514.



Decision Tree Regression

- Train score and Test score for Decision Tree regression using 10fold cv is 0.60574 and 0.14859 respectively.

- RMSE of Ridge regression for train is -11765.2068 and for test is -21090.0106.

Result

	Regression	Train_Score	Test_Score	RSME
0	Linear_Regression	-2.985711	0.014843	-12934.496708
1	Ridge_Regression	0.601659	0.405025	-16290.293754
2	Lasso_Regression	0.435703	0.285328	-16415.713084
3	Polynomial_Regression	-0.824257	0.014843	-16302.651455
4	Decision_Tree_Regression	0.605747	0.148599	-21090.010683

Future Work Scope

- We can do hyperparameter tunning to deploy the best algorithm such as Support Vector Regressor, Gradient Boosting, Neural Networks to get lower RMSE value.

- Also we can approach to "Time Series Analysis" to forecast the energy consumption of home appliances.

- Training other machine models can further boost the predictivity capacity, energy consumption is vast domain and have lots of scope in future.

Conclusion

- The best feature yields in the feature importance graph, where we see that the 'RH_8' is the most important feature and 'T5' is least important feature.
- The top three important features are 'Humidity' attributes, which is lead to the conclusion that the 'Humidity' affects the power more than the Temperature.
- 'Windspeed' and 'Temperature' features are least important as wind does not affect the power consumption inside the house.
- So the controlling the 'Humidity' inside house may lead to save Energy consumptions.
- 'Decision Tree Regression' found to be the best performing model
- This Dataset has the date time component so the best performance can be achieved by using 'Time Series Analysis' concept.

Thank You