Temporal Processing &
RNN – Assignment 5
ID: eo9232
Name: Md Reza
IE7860 – Winter 2022

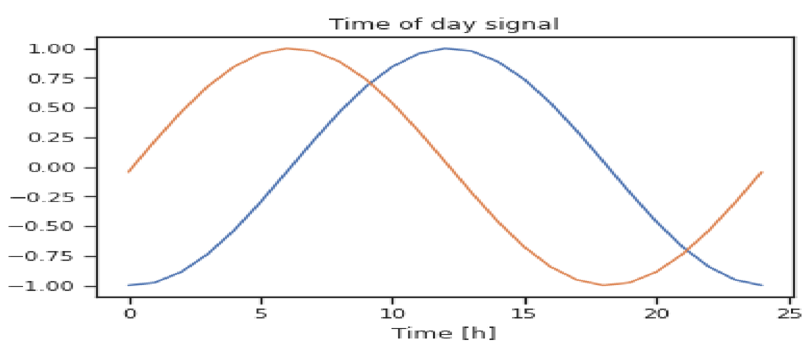
Project Report

Background and Methods:

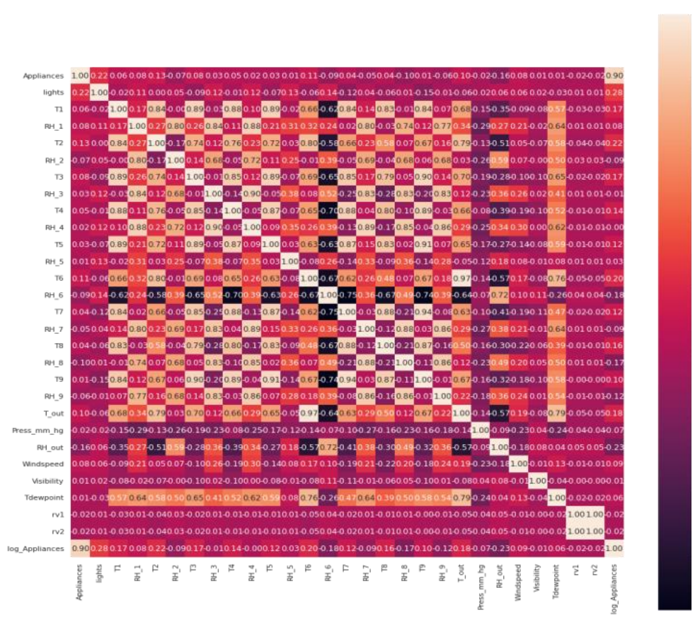
Appliances Energy Prediction Data: This data set records the energy consumption at 10 min resolution for about 4.5 months. In addition, a ZigBee wireless sensor network monitored the house temperature and humidity conditions. The temperature and humidity conditions were averaged and logged for 10 minutes with m-bus energy meters. This study proposes time series forecast to predict energy consumption using the time-lagged and recurrent neural networks time-series modeling in the presence of covariates.

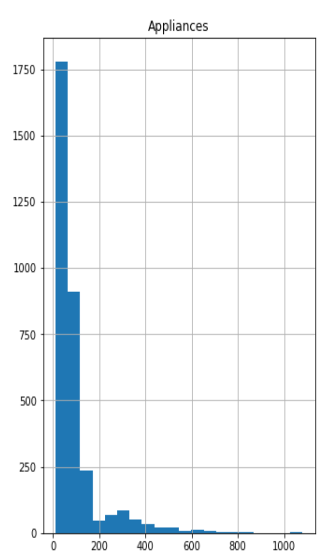
Pre-processing:

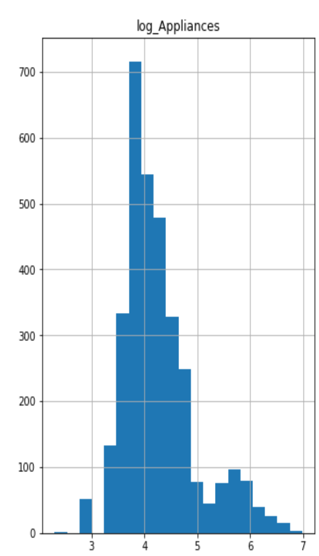
Appliances Energy Prediction	**1.** Creates the log of Appliances *(i.e., the target variable)* for the symmetric distribution of frequencies and improves model performance. **2.** Scaled and normalized the dataset. **3.** Inspect and clean the data. **4.** Leverage *sin* and *cos* to convert the time to "Time of the day" and "Time of the year" signals. This gives the model access to the most critical frequency features. Demonstrated in the figure below:



Correlation and partial-correlation plot to understand the dataset dynamics:





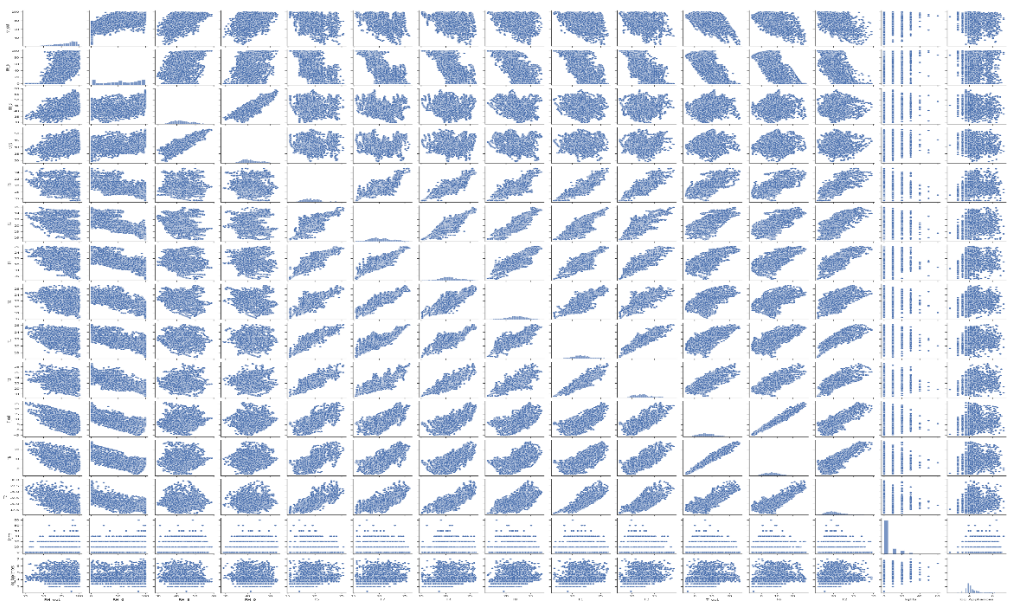


From the confusion matrix and histogram above, it turns out that the ***Appliances*** distribution is asymmetry. Therefore, to improve model performance, we will be using the ***Log of Appliances*** that is more symmetric in distribution.

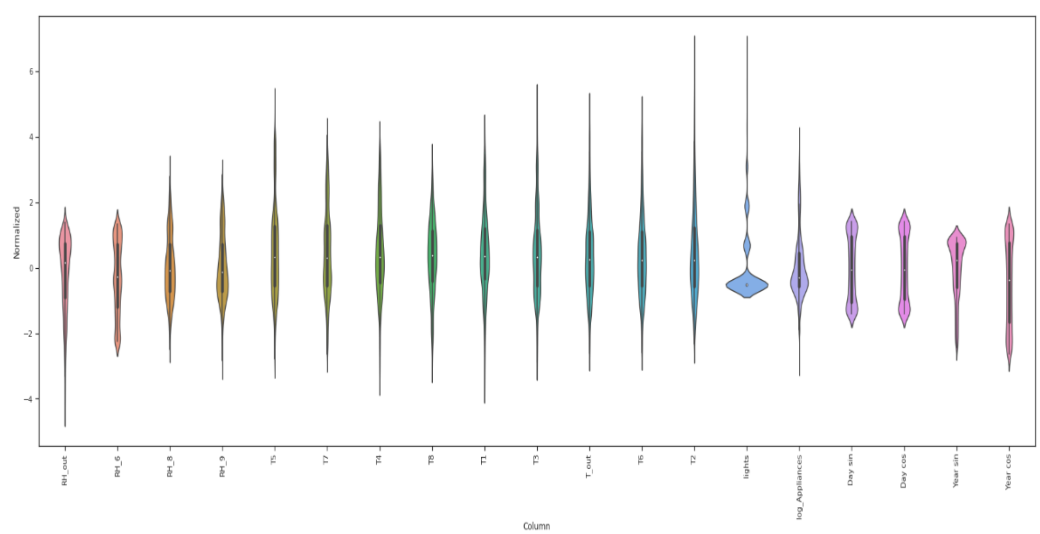
Multi collinearity check & features selections steps:

- Defined function to get diagonal and lower triangular pairs of the correlation matrix
- Defined function to select top features with absolutes correlations
- Leveraged the above functions to remove features with less than 10% (0.1) correlations with the target variable.
- Pairs plots with selected features have more than a 10% correlation to understanding the dynamics

The Pair Plot below shows that the extracted features carry some weight to build better models.



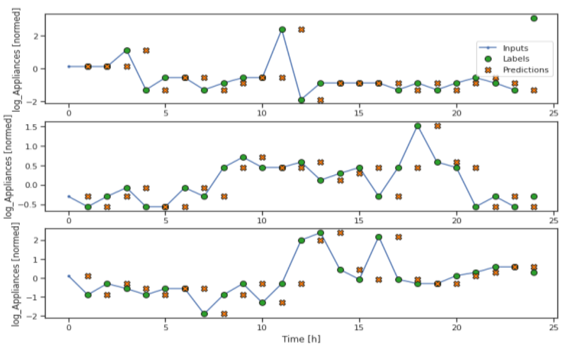
Now check the distribution of the features in the violin plot below. Though some features have long tails , there are no apparent errors with unexpected values.



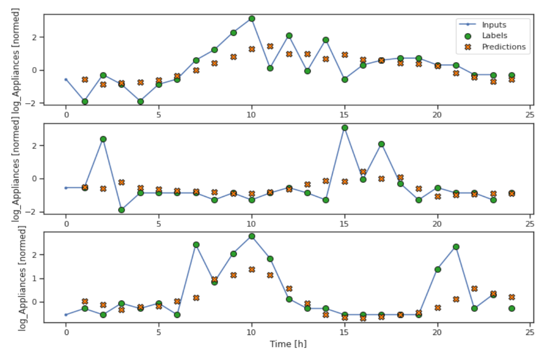
Design and evaluation of the networks:
One way to measure time-series models' performance is to compare the loss and mean absolute error derived from model predictions based on inputs, targets, and other required parameters. To evaluate model performance, the mean absolute error of each model from single-step and multi-steps are compared and plotted over the average of all output. For details, please refer to the model-based observations plot below. In addition, a function was created to visualize model performance with inputs and labels that were converted to time windows using the *split window* function. Please refer to the **MLP Model Summary Table** in the appendix below for details on model performance.

Models' performance based on the single-step window to predict energy consumption for 1 hour in the future, given 24 hours past:

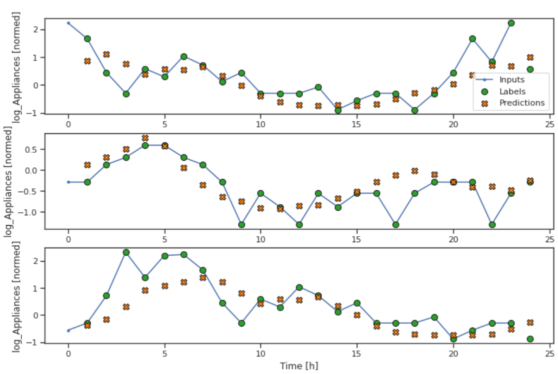
Baseline



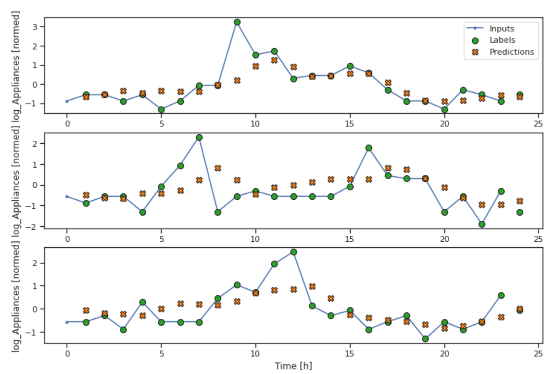
LSTM



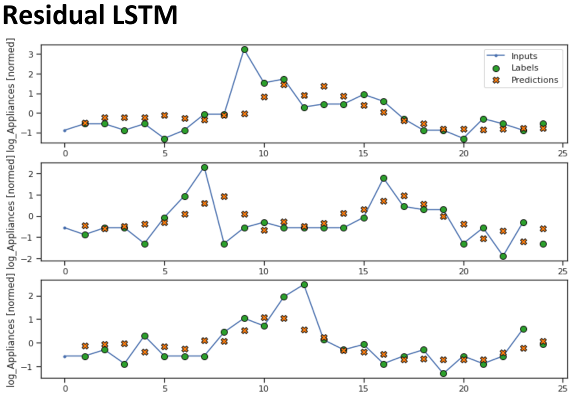
GRU



RNN



Residual LSTM

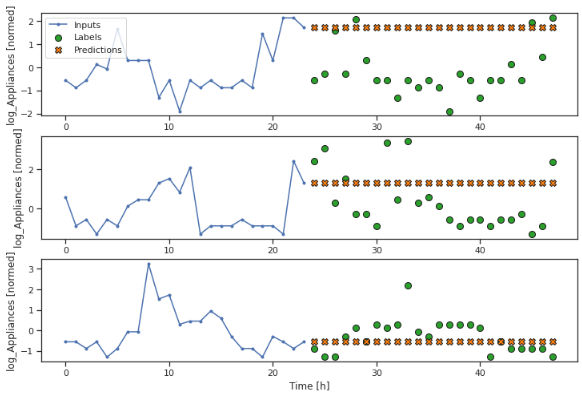


The plot from the single-step model runs over 24 hours to predict one hour in the future, as demonstrated above, where:

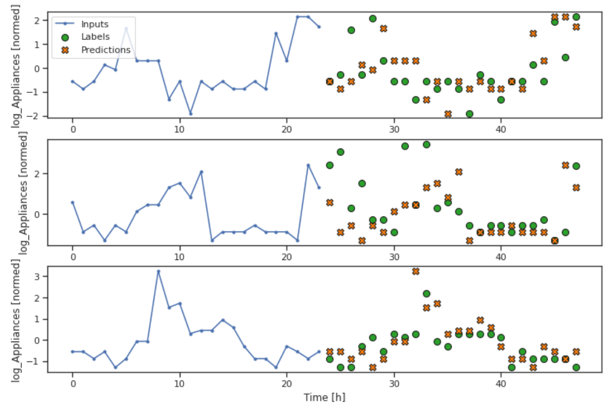
- The blue Inputs line shows the input value at each time step.
- The green Labels dots show the target prediction value.
- The orange Predictions crosses are the model's predictions for each output time step.

Models' performance based on the multi-step window to predict energy consumption for 24 hours in the future, given 24 hours past:

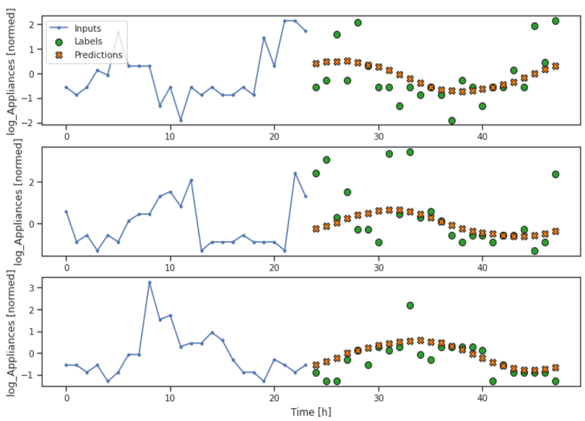
Baseline Last



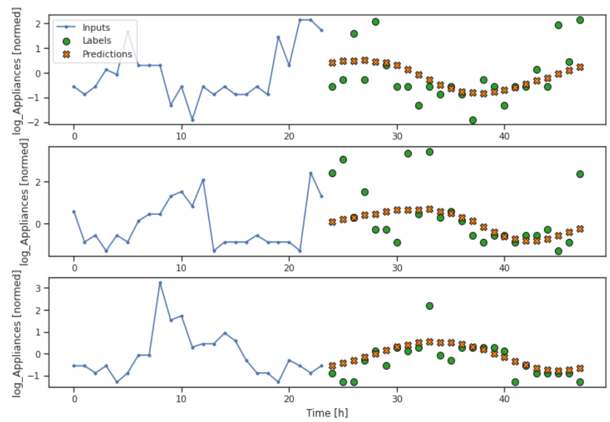
Baseline Repeat



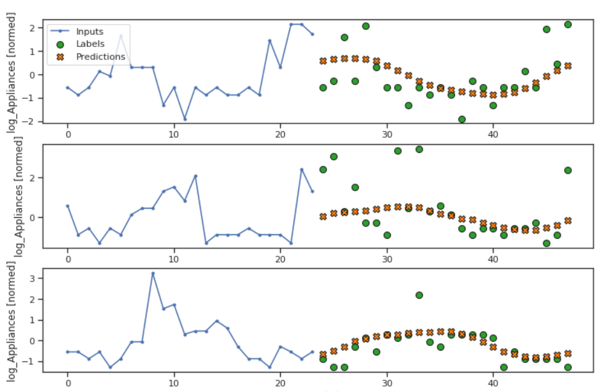
LSTM



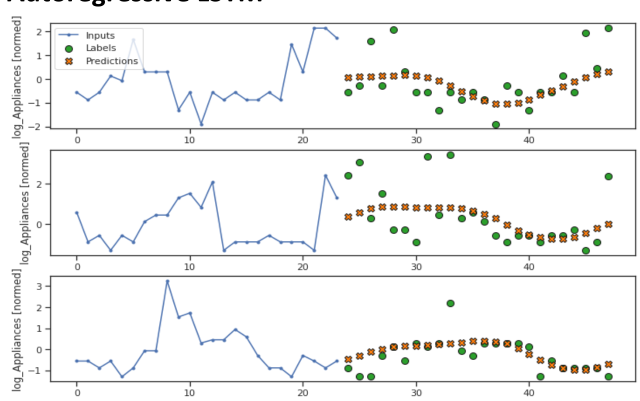
GRU



RNN



Autoregressive LSTM

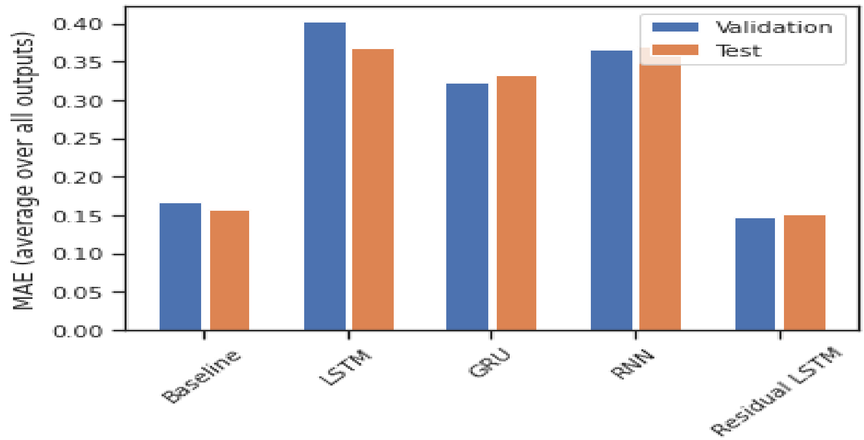


The plot from the multi-step model runs over 24 hours to predict 24 hours in the future, as demonstrated above, where:

- The blue Inputs line shows the input value at each time step.
- The green Labels dots show the target prediction value.
- The orange Predictions crosses are the model's predictions for each output time step.

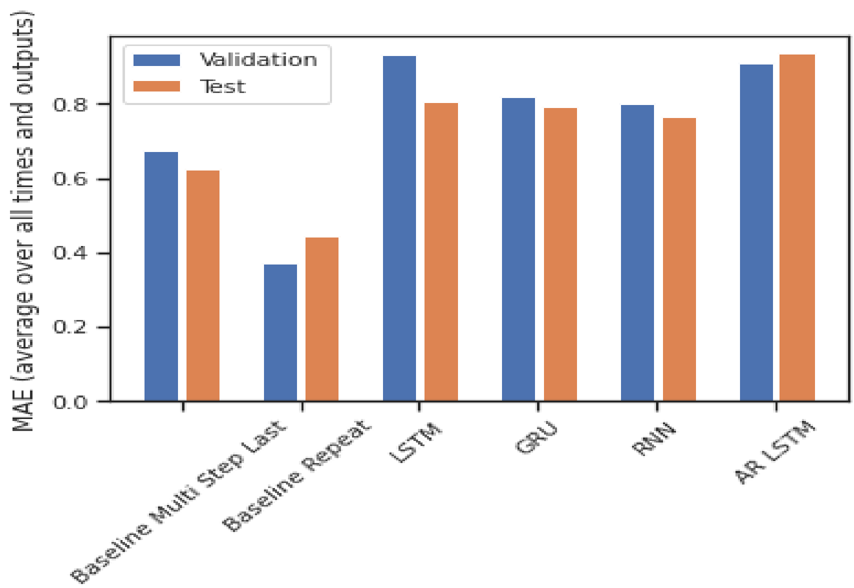
Model-based performance with MAE over the average of all output (Single-Step):

With single step, it turns out that the gains from LSTM, GRU, and RNN performance are almost the same, but the Residual LSTM performed well.



Model-based performance with MAE over the average of all output (multi-Step):

With multi-step, it turns out that the gains from LSTM, GRU, and RNN are only a few percent (if any), and the autoregressive model performed the worst. Of course, these complex approaches may not be feasible for this problem, but one could not comprehend that without trying complicated models like these.



Compare your results to Candanedo et al. (2017):

The residual LSTM performed well with a lower loss and mean-absolute-error rate with single-step than LSTM, RNN, and GRU. Unlike **Candanedo et al.'s (2017)** research, all the models" ranked time as the most important in predicting the appliances' energy consumption. The exploratory data analysis showed that data from the kitchen, laundry room, living room, outside temperature, teenage room, parents' room, and bathrooms had the highest contributions in predicting appliance energy consumption.**[1]**

Finally, per **Candanedo et al.'s (2017)** research and the outcome of this **Temporal Neural Network & RNN** project, the prediction of appliance consumption with data from the wireless network indicates that it can help locate locations of the building where the main appliances contribute the highest consumption of energy within the building.

Reference:

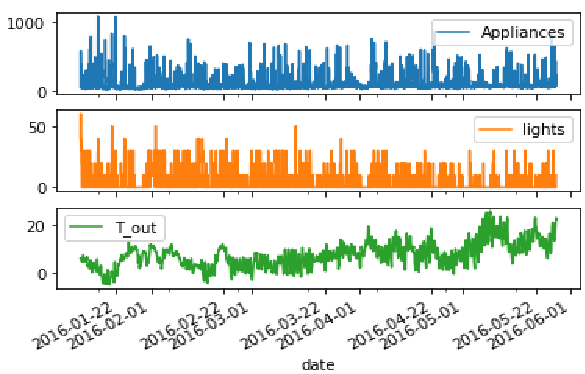
[1]. Candanedo Ibarra, Luis & Feldheim, Veronique & Deramaix, Dominique. (2017). Data driven prediction models of energy use of appliances in a low-energy house. Energy and Buildings. 140. 10.1016/j.enbuild.2017.01.083.

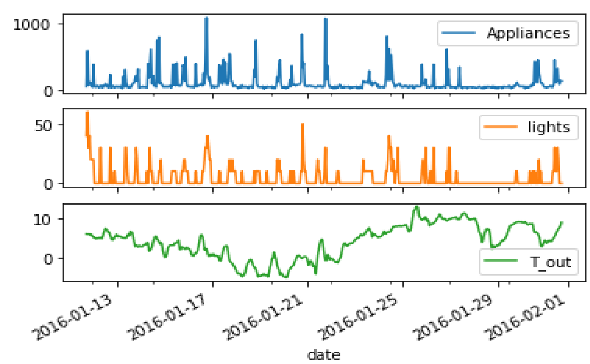
APPENDIX:

Appliances Energy Prediction Model Summary:

Model	**Model Performance Single-Step (MAE)**	**Model Performance Multi-Step (MAE)**
Baseline	**0.1587**	**N/A**
Baseline Last	**N/A**	**0.6249**
Baseline Repeat	**N/A**	**0.4439**
LSTM	**0.3678**	**0.8049**
GRU	**0.3333**	**0.7937**
RNN	**0.3700**	**0.7659**
Residual LSTM	**0.1511**	**N/A**
AR LSTM	**N/A**	**0.9361**

The plot below demonstrates how features like Appliances, Lights, & Temperature evolve over time:





Fast furious transform *(i.e., FFT)* of the appliances' energy consumption over time.