AC Power Consumption

September 19, 2021

1 Problem Statement

There are 18 ACs installed in one of the hotel in Gurgaon. We need to find out the pattern and insights from the data and use the outside temperature of Gurgaon to forecast the power consumed.

2 Dataset Info

2.1 Data Granularity

The collected data has one minute granularity. Moreover, this data was provided from different sensors installed in the hotel. The reason for concluding that individual sensors were used in collecting the data of the different ACs. That's why there are the points in the data where the value in NaN for one and for other ACs the power consumption value is there.

2.2 Shape of data

The dataset has 87840 rows and 19 columns.

2.3 Missing values

At least 50% of the power consumption data was missed for all of the ACs.

3 Approach

3.1 Data Cleaning

The provided data has timestamp column named as `0` and the other columns have names like `AC 1`. For dealing with such things and improving the productivity while writing the code the timestamp column has been named as `timestamp` and the power consumed by the AC columns have been renamed as `ac_1`.

After performing the above steps the data is dumped into the **data/cleaned** directory with name as `ac_pwr_cleaned.csv`.

3.2 Data Preparation

Since the collected is of minute granularity and has missing values the following approaches have been used for data preparation:

3.2.1 Imputing missing value

The missing values have been filled by linear method of interpolation. The imputed data has been dumped under `data/cleaned` directory with name as `ac_pwr_na_filled.csv`.

3.2.2 Data Resampling

The data has been resampled to hourly, daily, and monthly for going forward with the analysis.

Hourly While performing the hourly resampling the mean of the values of the individual AC power has been taken and these individual values for every hour by different ACs have been summed up to calculate the power consumed by the hotel on an hourly basis.

Daily The hourly power consumed by ACs is then resampled into daily by sum.

Monthly For monthly resampling the daily resampled data is summed up.

Final Power Consumed The final power consumed dataframe has been created by summing up the monthly power consumed by the ACs.

The resampled data have been dumped into the `data/resampled` directory with the following names:

- **hourly resampled** dumped as `ac_pwr_hourly.csv`
- **daily resampled** dumped as `ac_pwr_daily.csv`
- **monthly resampled** dumped as `ac_pwr_monthly.csv`
- **final power consumed** dumped `ac_pwr_total_consumed.csv`

3.3 EDA

3.3.1 Which AC was used the most/least?

For answering this question the `ac_pwr_total_consumed.csv` was used. `ac_5` is used the least and `ac_18` is used the most.

3.3.2 Are all the ACs of same capacity?

After observing the summary statistics of the hourly resampled data one observation has been made that the max value of power consumed by the ACs i.e. for `ac_5`, `ac_10`, and `ac_17` was way less than the max value of power consumed by the other ACs. This gave a sign that the ACs are not of equal capacity.

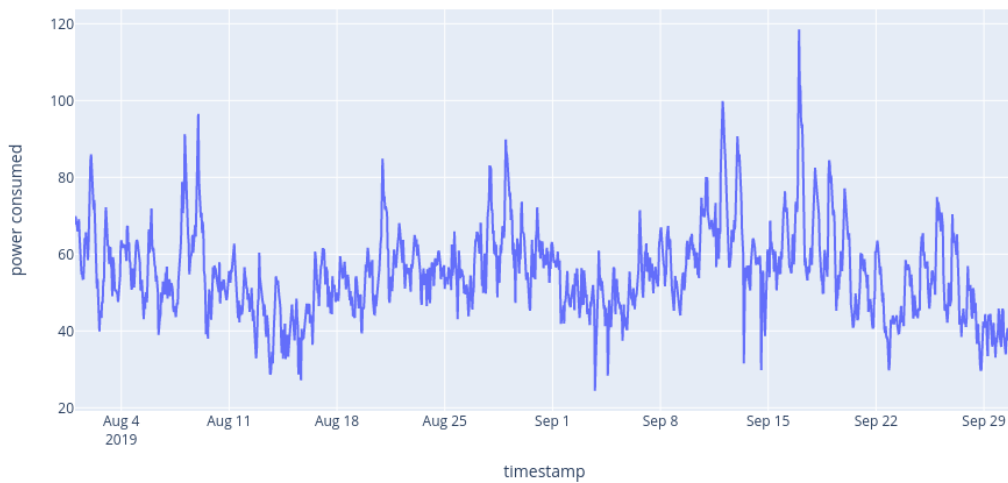**Comparing hourly power consumed by the ACs**
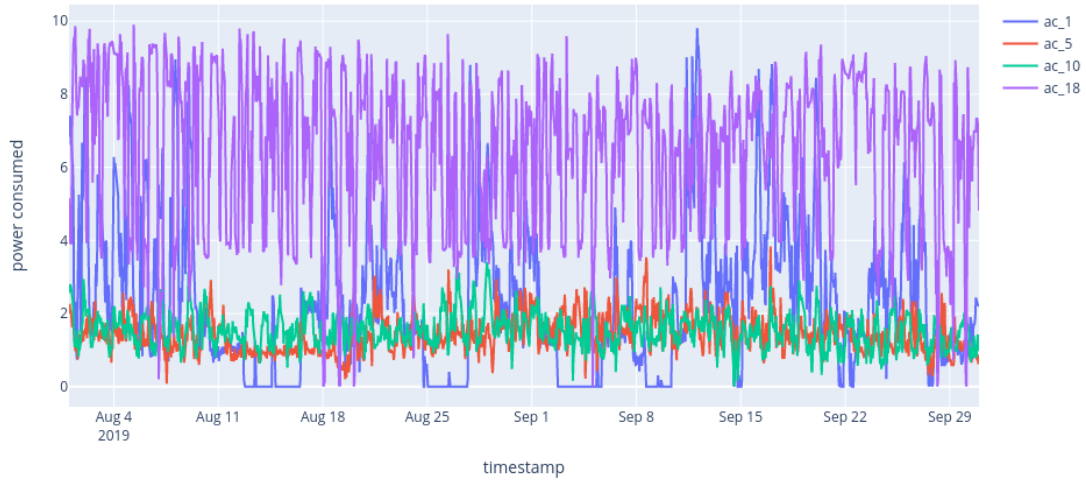
The above plot clearly showed us that the ACs **#5**, **#10**, and **#17** having less capacity than the other ACs.

This the reason for **ac__5** being used the least.

Moreover, it also looks like the AC **#18** is installed at the hotel reception. Because the power consumption for it reaches 0 a few times.

3.3.3 Power consumed by the hotel

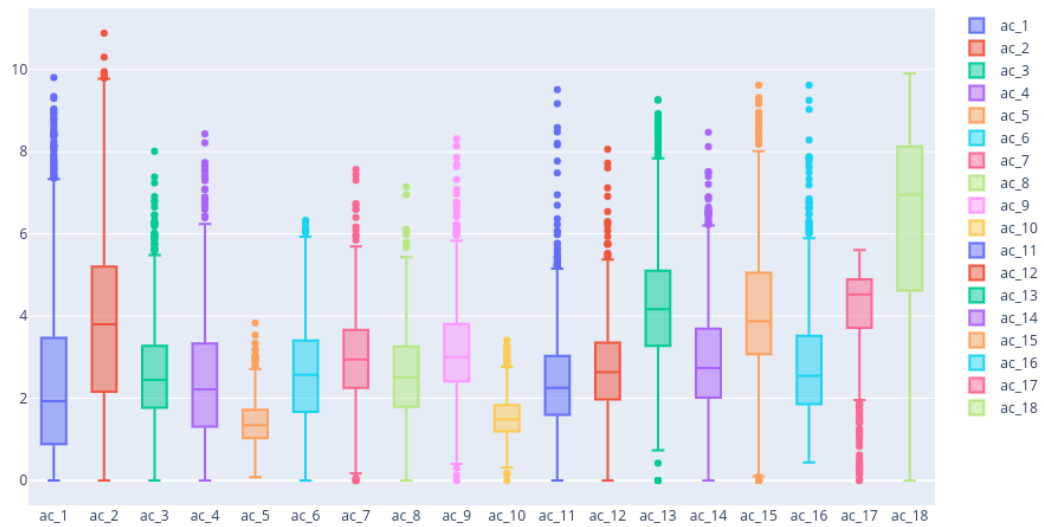**Hourly power consumed by the hotel**

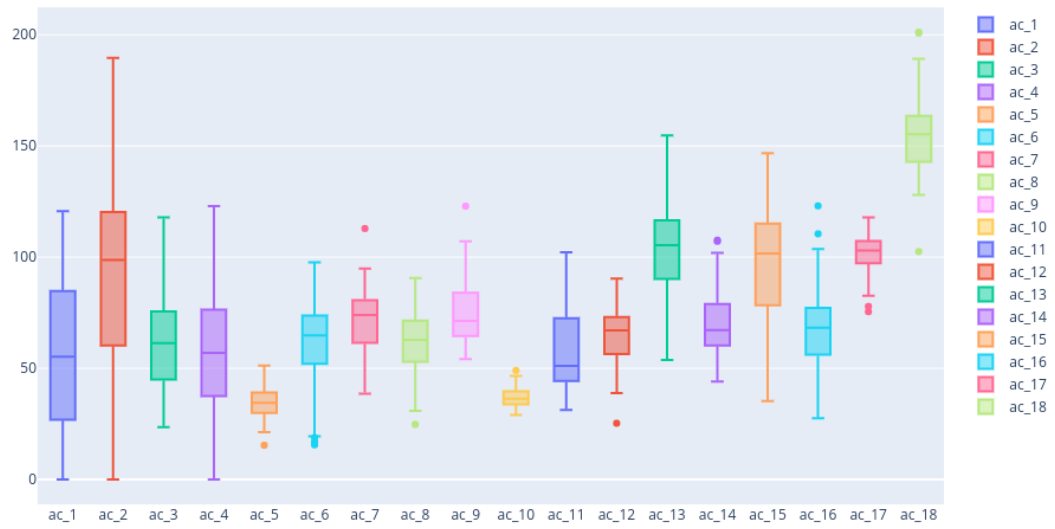**Daily power consumed by the hotel**

Looks like on holidays i.e. during Aug 12-Aug 15, 2019 the power consumed is less.

3.3.4 Analyzing distribution of power consumed

Boxplot of the hourly power consumed by ACs
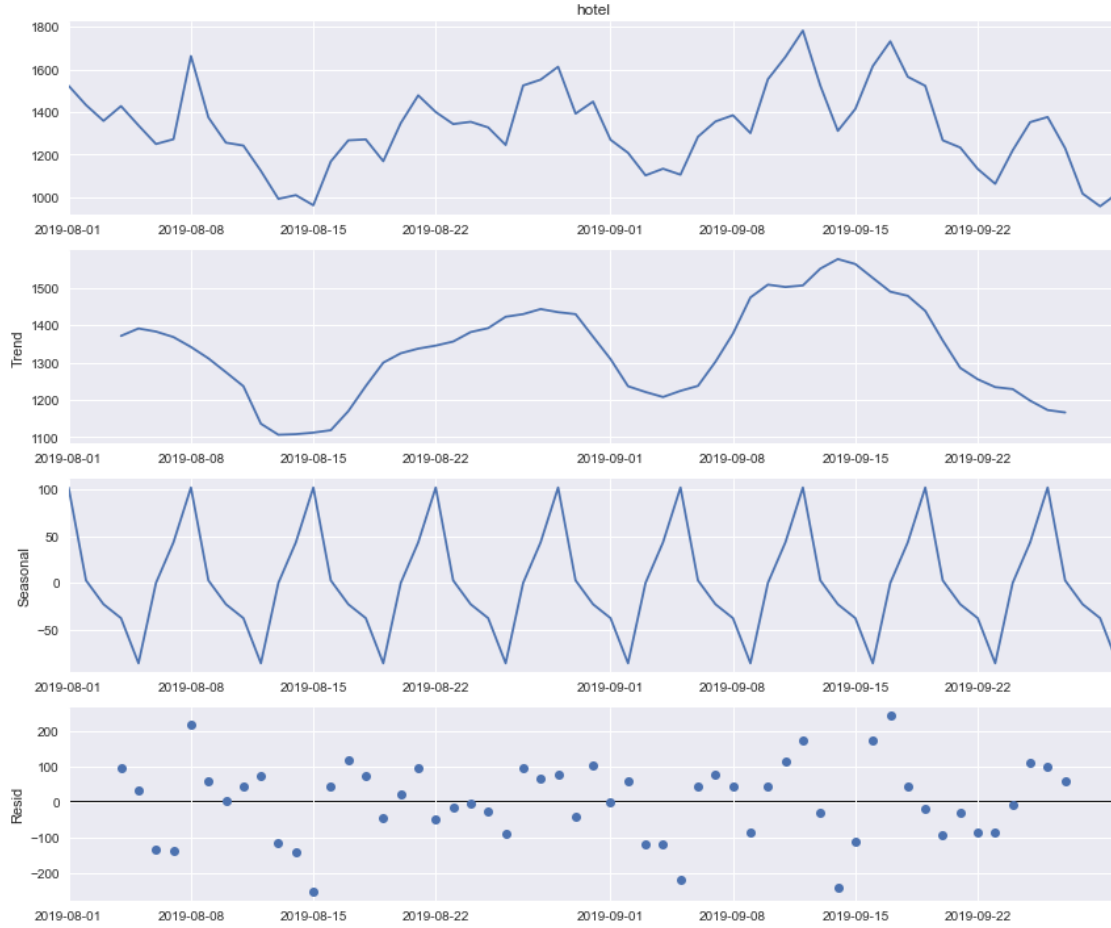
Boxplot of the daily power consumed by ACs

For AC **#18** the minimum daily power consumed is nearly 100.

3.3.5   Seasonal Decomposition

Hourly

Daily Both the hourly and daily power consumed data of hotel has strong seasonality.

3.3.6 Feature Engineering

The following features have been created using the hourly resampled data:

- lag values of 1 and 2

- rolling window with periods as 1 and 2 using mean, min and max as stats

3.3.7 Model Building

Since the data is available from Aug 1, 2019 - Sep 30, 2019. Will try to **forecast** the *power consumed* by the hotel for the next **7 days on hourly granularity**. For forecasting the simple linear regression model was used with input features as standard scaled.

The training data is from Aug 1, 2019 to Sep 23, 2019 and test data is from Sep 24, 2019 to Sep 30, 2019.

Evaluation Metrics

- **MAE (mean absolute error):** 6.142342324191227e-15
- **MSE (mean squared error):** 6.364111591999873e-29