

Preprocessing dataset

Two csv files containing household power consumption data of 2017 and 2018 are given. Each dataset contains 30 columns. Some preprocessing has been done on the raw datasets to rearrange the Dataframe format and extract couple of helpful features and merged to a single one.

All the codes, Notebooks and Tableau packages used in the analysis are available in this repository (<https://github.com/orionfoysal/fujitsuHouseholdPowerConsumption>)

Relationship between maximum temperature and daily consumption

Let's examine the distribution of daily usage at different maximum temperature of the day.

max_temp vs daily_usage

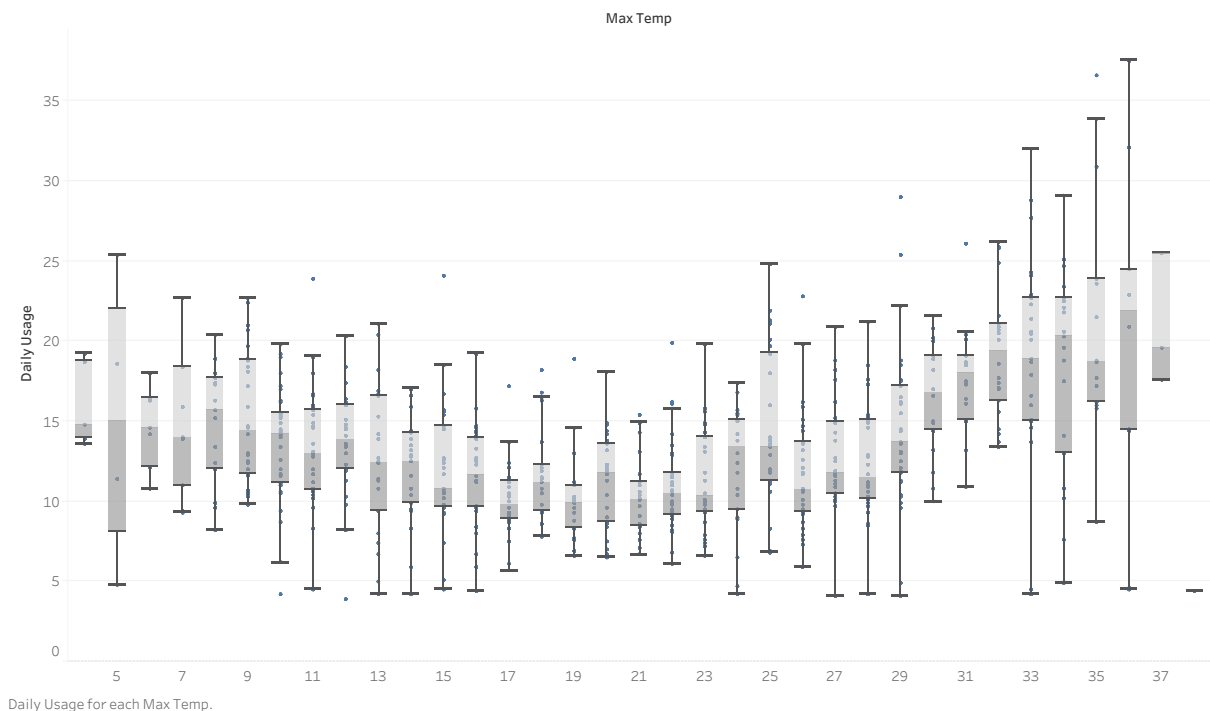


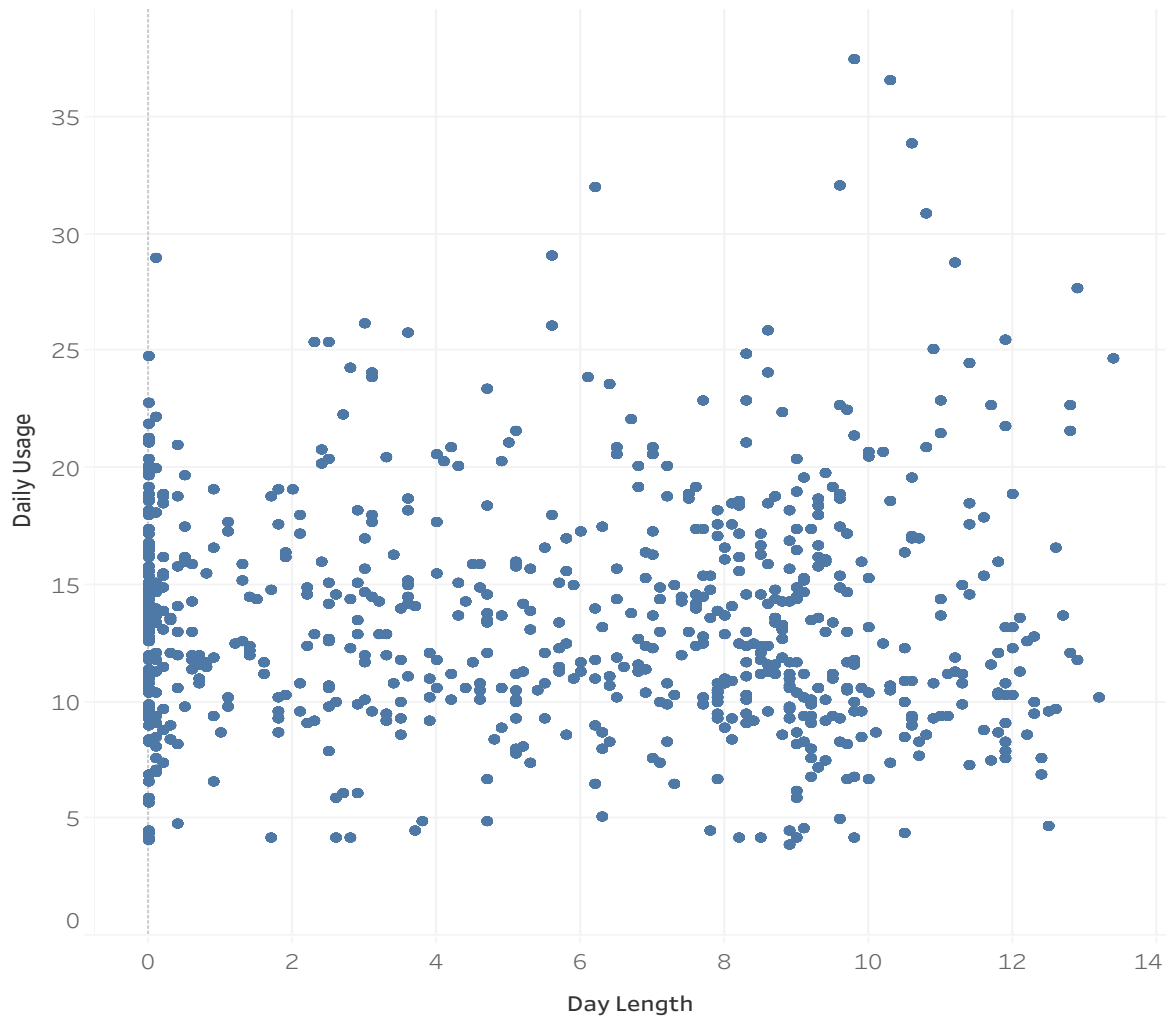
Fig: Daily usage vs Max temperature of the day

We see that average consumption is lower in ambient temperature. On the other hand, power consumption increases with the increase of maximum daily temperature.

Relationship between daylight hours and daily power consumption

Let's generate a scatterplot of daily power consumption data at different day lengths to find the relationship between these two parameters.

day_length - daily_usage



Day Length vs. Daily Usage.

Fig: Daily Usage vs Daylight hours

From the above figure we see there is no clear relationship between daylight hours and daily power consumption data.

Power Consumptions at different weekdays

Let's check power consumption at different day of the week.

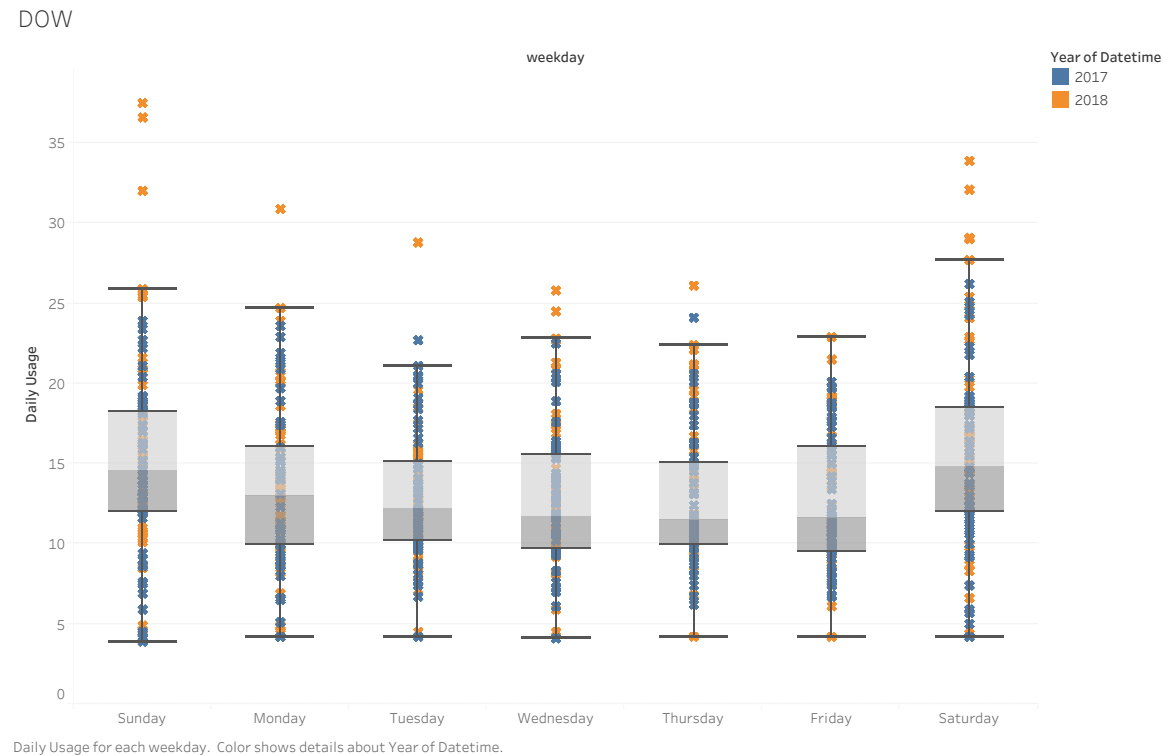


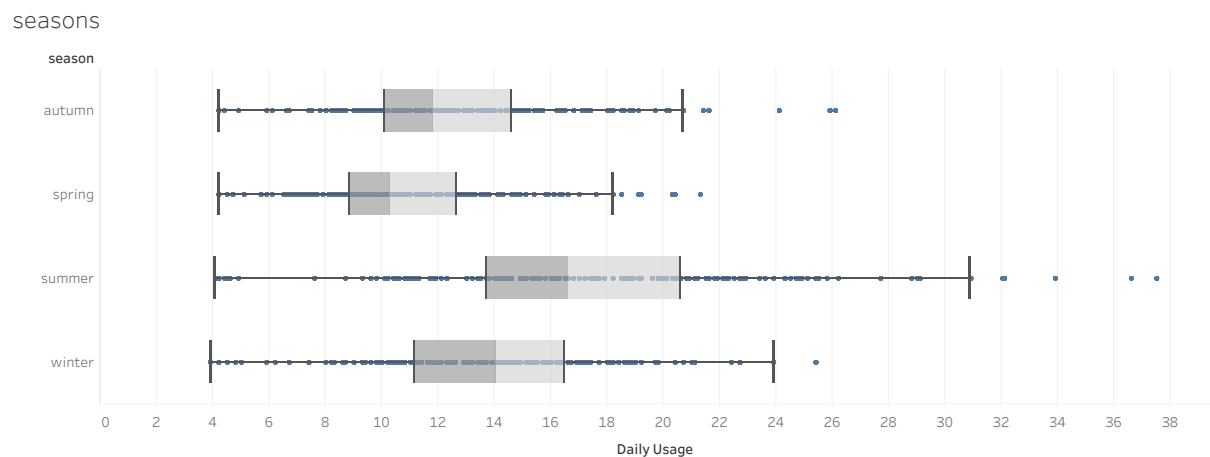
Fig: Daily Power consumption in different weekdays

From the above boxplot we can see that daily power consumption is higher in Saturday and Sunday. As these are weekly holidays in Japan, consumption is higher in households. For the other weekdays this distribution is quite similar. We can also see some extreme outliers in weekends. This could be because of the visit of guests during those holidays.

Effect of Seasons and Weather in Power Consumption

The given dataset doesn't contain any season column. But we can easily extract season information using the month of the year information.

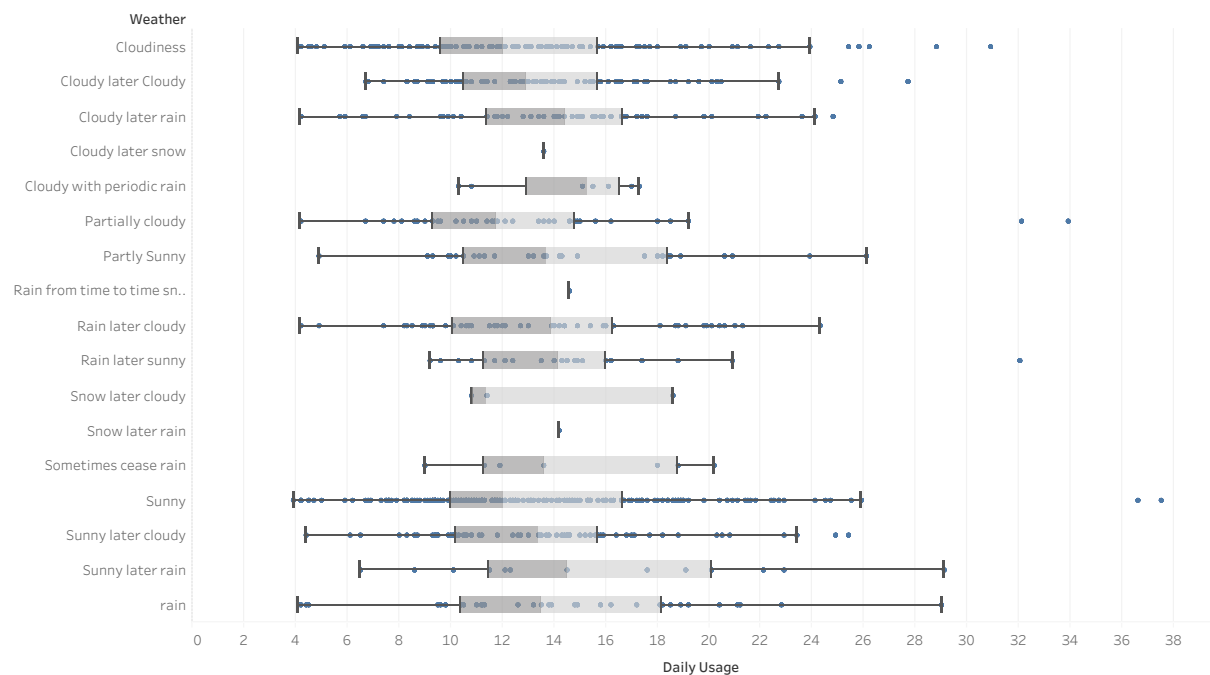
Now let's examine the distribution of daily power usage in different seasons with the help of boxplot.



From the above boxplot it is clear that power consumption is highest during summer and lowest during spring. Consumption is second highest during winter. One of the main reason behind this might be the higher use of air-conditioner in summer and use of heater in winter.

Now let's check the distribution of daily usage in different weather conditions.

weather

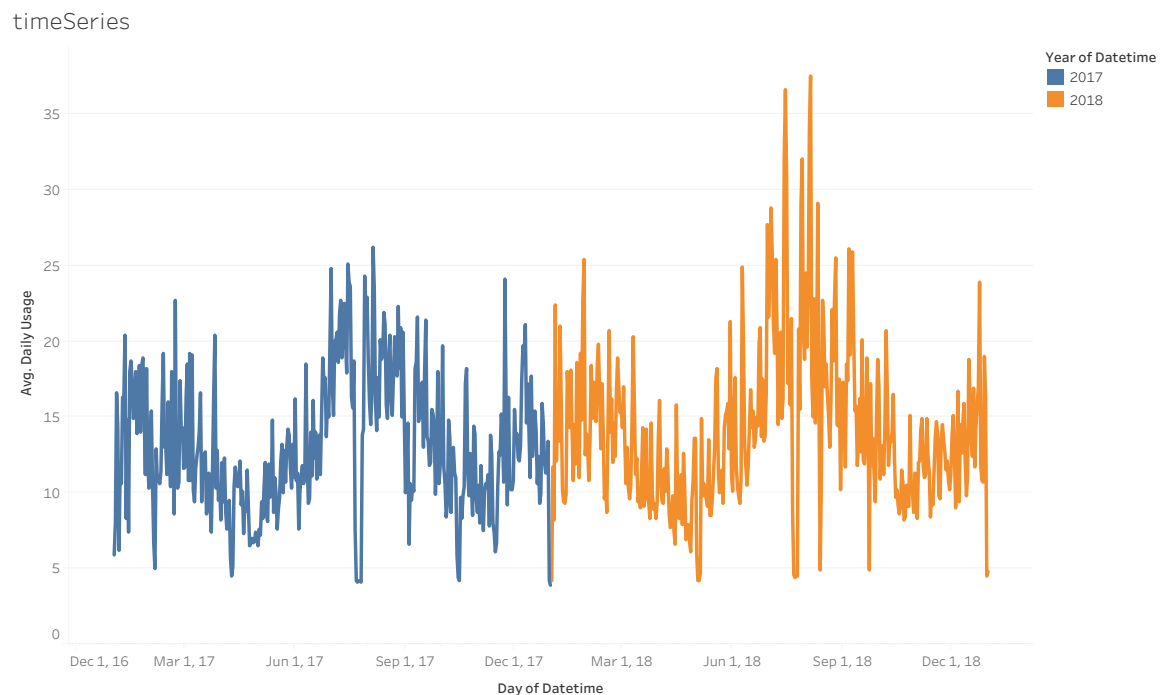


Daily Usage for each Weather.

As for weather condition there are 17 distinct conditions available in the dataset. In this case the spreads of the distributions are larger. The median seems higher during the weather conditions where rain was involved.

Periodicity

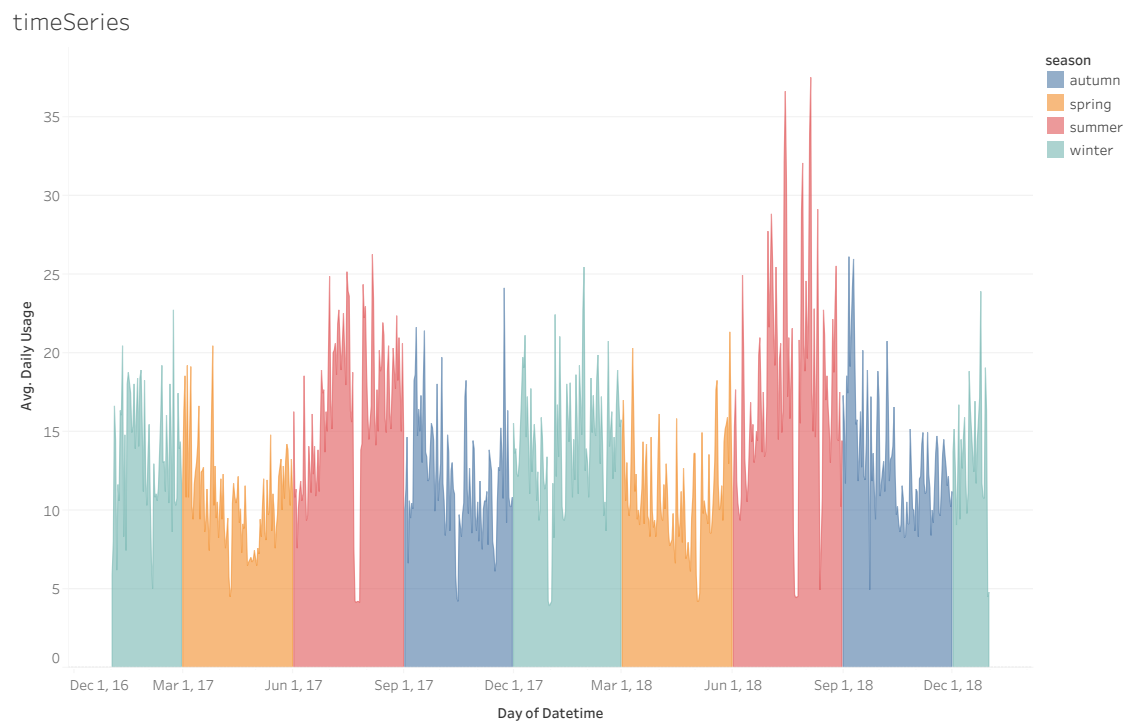
Let's plot the timeseries of day on day power consumption data. We can see a very clear yearly periodicity in 2017 and 2018 data.



The trend of average of Daily Usage for Datetime Day. Color shows details about Datetime Year.

Fig: Yearly periodicity in daily power consumption data

Seasonality



The plot of average of Daily Usage for Datetime Day. Color shows details about season.

Fig: seasonal effect on day on day power consumption

In the last figure clear seasonal effect is visible in the data.

Have there been any changes in the use of electricity in this house in the 2 years with the data?

Let's compare month on month total power consumption of 2017 and 2018.

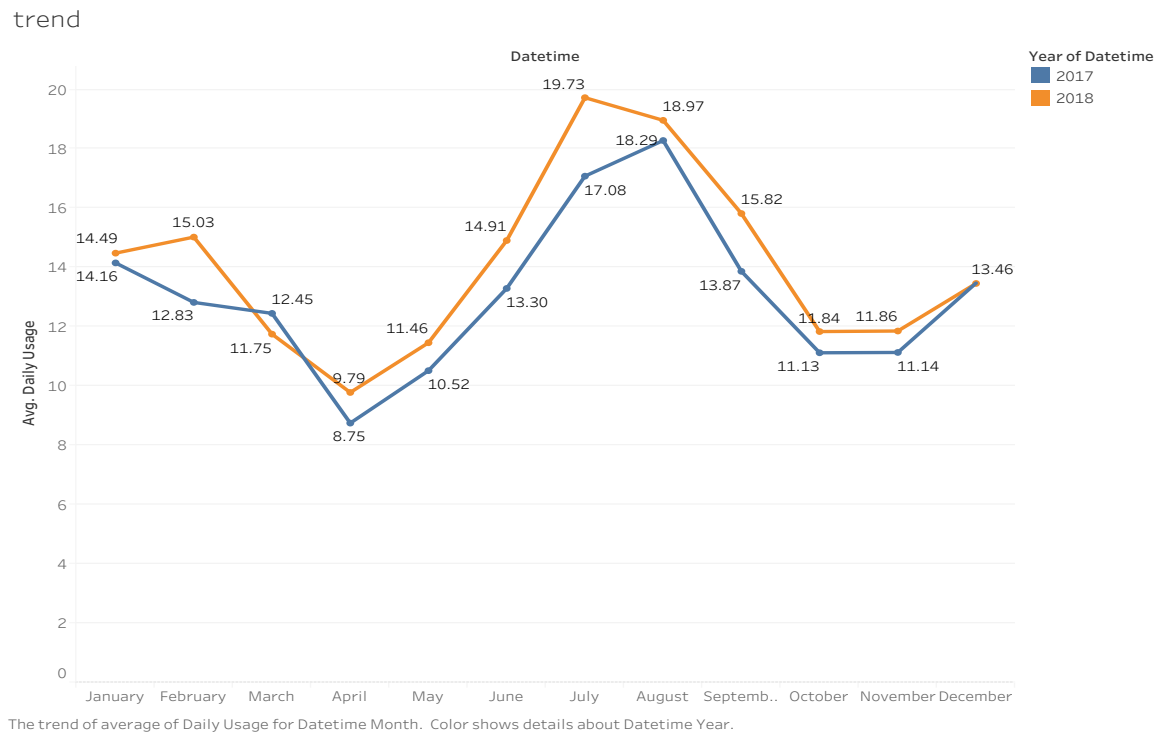


Fig: Comparison of monthly total power consumption in 2017 and 2018

From the above figure we see that total power consumption had increased in every month (except March) of 2018 compared to previous year.

One noticeable thing is the increment is **higher during summer season and lower in winter**. Could this be the effect of Global Warming?? We can't say for sure from the given data. But it can be an interesting study!

Predicting the electricity usage of this house in the future

We can forecast future electricity usage using technique like *ARIMA*. Here a forecast of next one year is generated using '**PROPHET**'¹, a library developed by facebook for time series forecasting. Prophet works best with time series that have strong seasonal effects and several seasons of historical data.

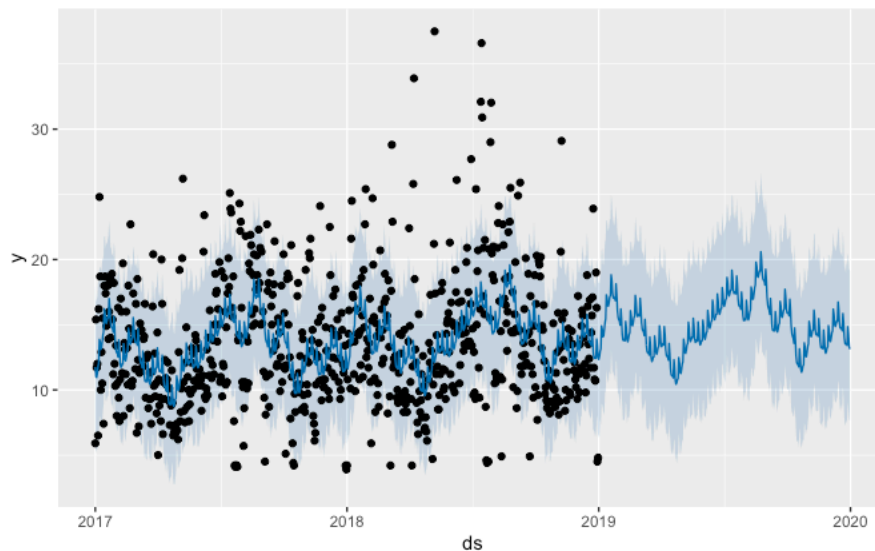


Fig: Forecast of 2019 electricity usage data based using PROPHET

We can also check for the trend and seasonality of the data in the forecasted model generated with prophet.

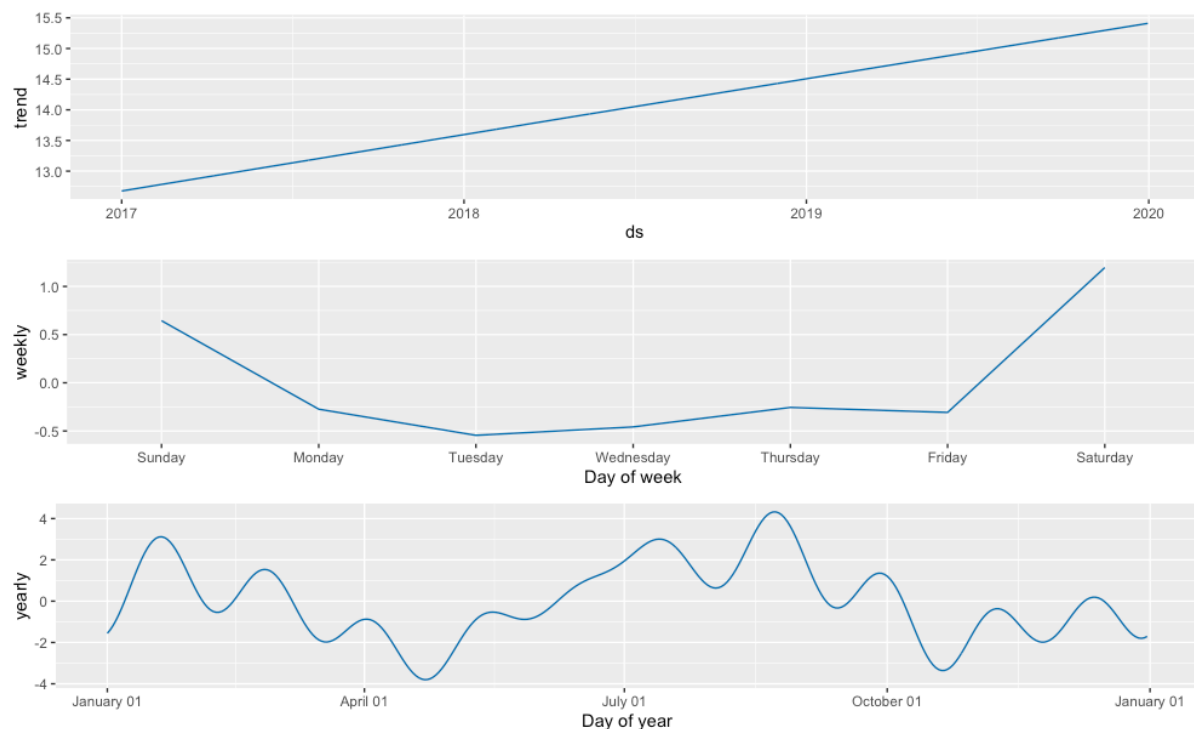
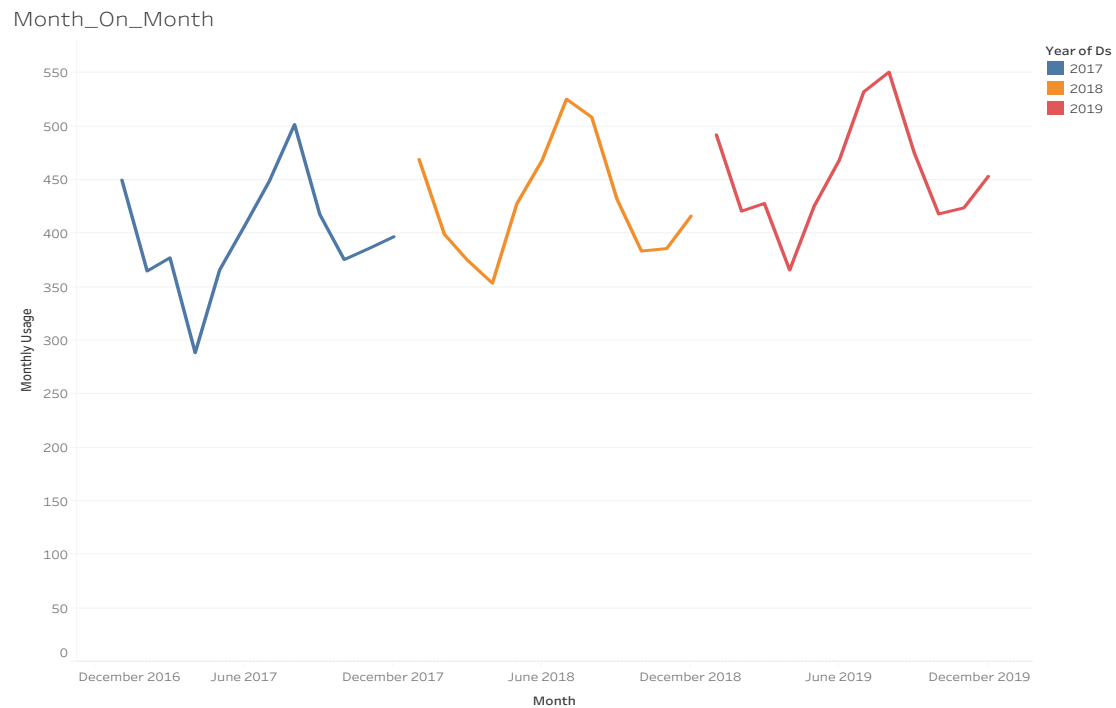


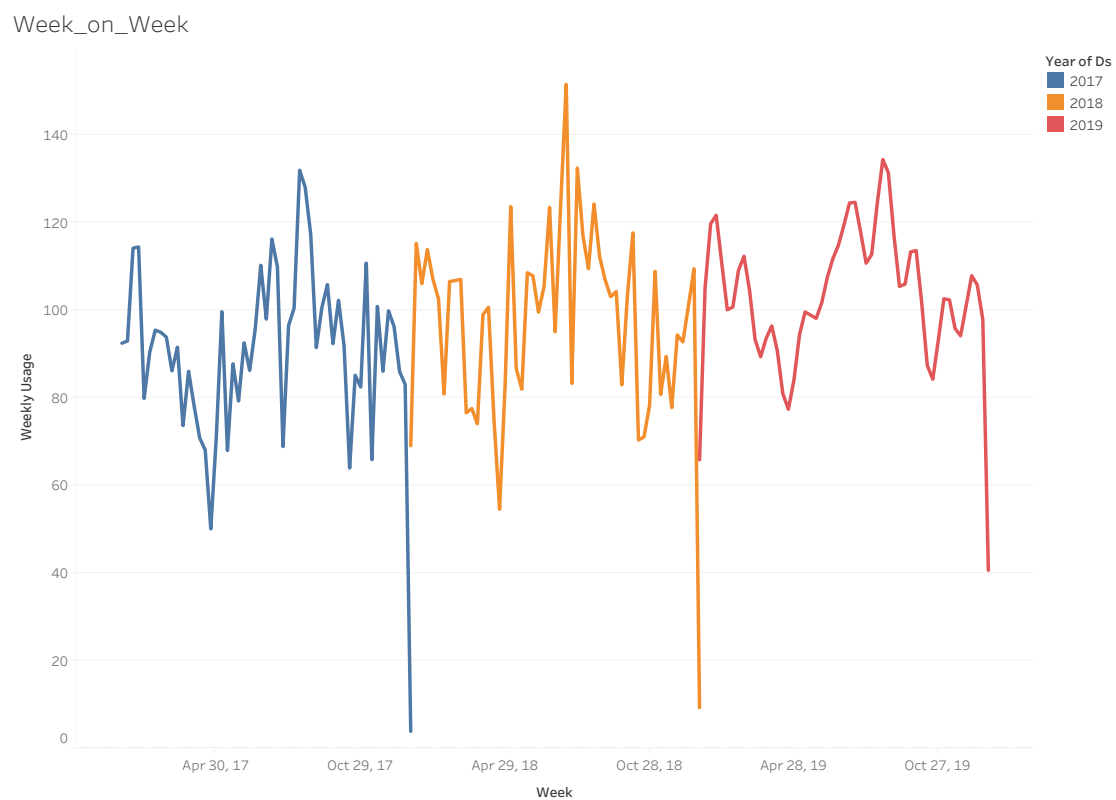
Fig: Trend, weekly and yearly seasonality in the forecasted model

We can also forecast monthly or weekly power consumption data. Here is a quick forecast of monthly and weekly total power consumption in 2019 generated using **Tableau software**.



The trend of sum of Y for Ds Month. Color shows details about Ds Year.

Fig: Monthly total power consumption forecast in 2019



The trend of sum of Y for Ds Week. Color shows details about Ds Year.

Fig: Weekly total power consumption forecast in 2019

Predictive Modelling

Another approach to predict the power consumption in future can be predictive modelling of the given dataset. We can use the available and extracted categorical and numerical features and train it to predict hour level or day level power consumption. Multiple regression, SVM, Tree based model or more advanced ensemble models like Random Forest/XGBoost/LGBM can be used for this. But as there is only two years of data available it might be bit difficult to fit this data using these algorithms.

Number of persons in the house

In 2013 per capita household electricity consumption in Japan was around 2245 KWh/year ². Let's assume that this hadn't been changed much since then to 2017 and 2018.

The average electricity consumption of the lowest-income class is 272.68 kWh per month, while that of the highest-income class is 571.67 kWh per month ³.

Total consumption in the house in 2017 = 4781 KWh

Approximate number of persons in the house = $4781/2245 = 2.12$

The average monthly usage suggests this as a middle to high income class household. So, the average consumption of electricity will be higher than our approximation. Hence, we can assume the number of people in that house is 2

What Job do they do?

Let's check the distribution of hourly data excluding the holidays and weekends.

hourly_usage_distribution

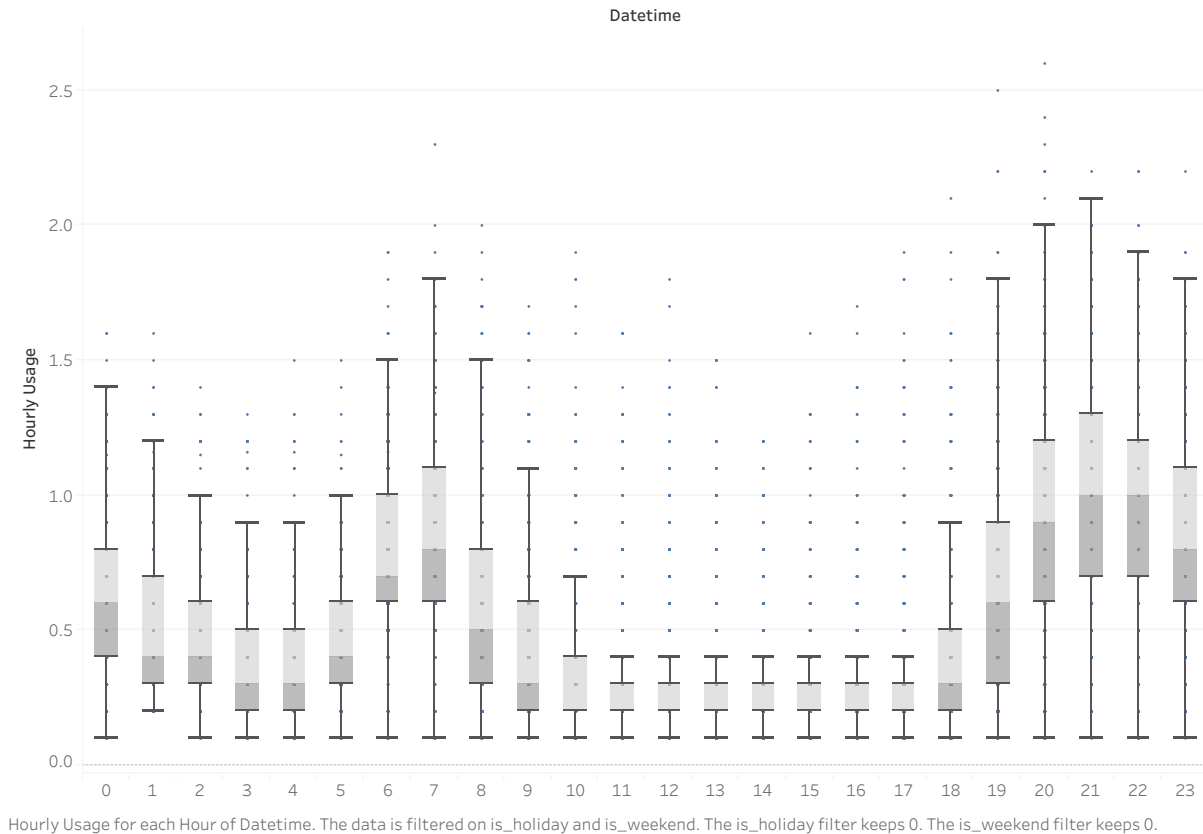


Fig: Distribution of Hourly power consumption at different hours

We see the very clear trend in the distribution. In morning and night there is a high demand. On the other hand, during office hours (from 9 AM to 6 PM) the demand is very low. Which means both persons living in the house remain out of the home during office hours. Both of them might be office worker.

References:

1. <https://github.com/facebook/prophet>
2. <https://www.statista.com/statistics/597901/household-consumption-of-electricity-per-capita-in-japan/>
3. <http://www.waseda.jp/prj-ricem/dp/dp1803.pdf>

前処理データセット

2017 年と 2018 年の家庭用電力消費データを含む 2 つの csv ファイルが提供されています。各データセットには 30 列が含まれています。生のデータセットでいくつかの前処理が行われ、Dataframe 形式が再配置され、いくつかの有用な機能が抽出され、単一の機能にマージされました。

分析で使用するすべてのコード、ノートブック、Tableau パッケージは、このリポジトリで利用可能です

[<https://github.com/orionfoysal/fujitsuHouseholdPowerConsumption>]

最高気温と毎日の消費量の関係

異なる最高温度での 1 日の使用量の分布を調べてみましょう。

max_temp vs daily_usage

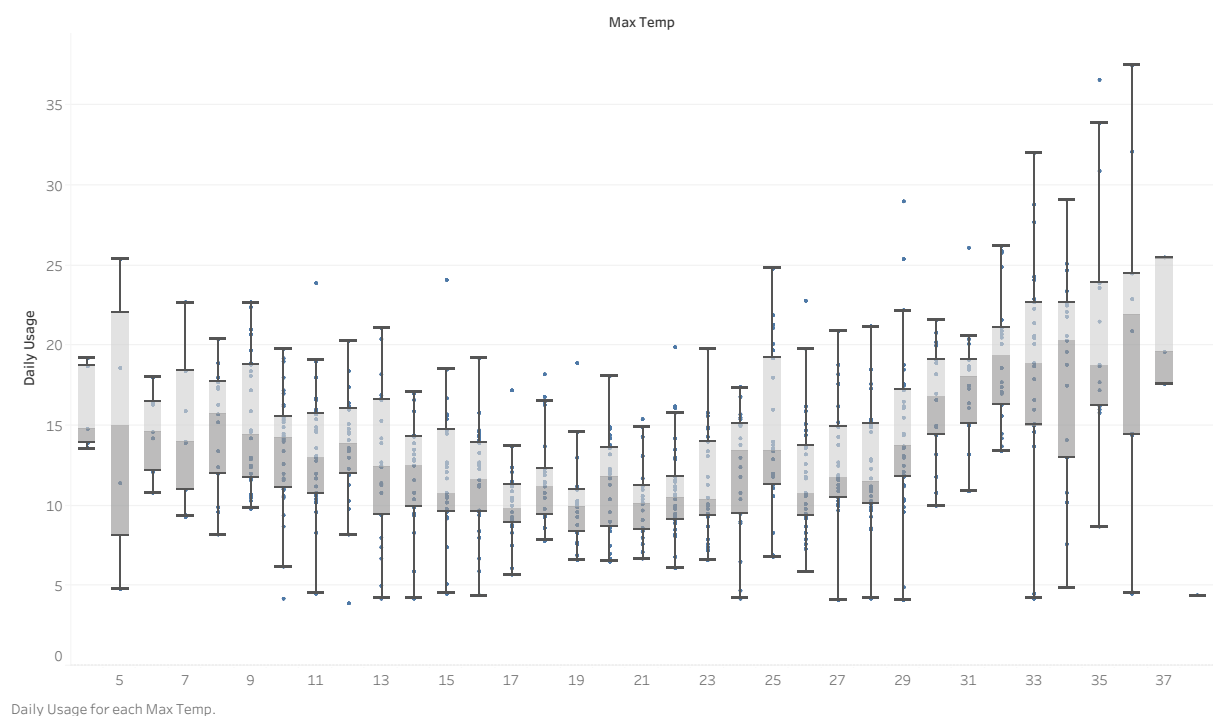


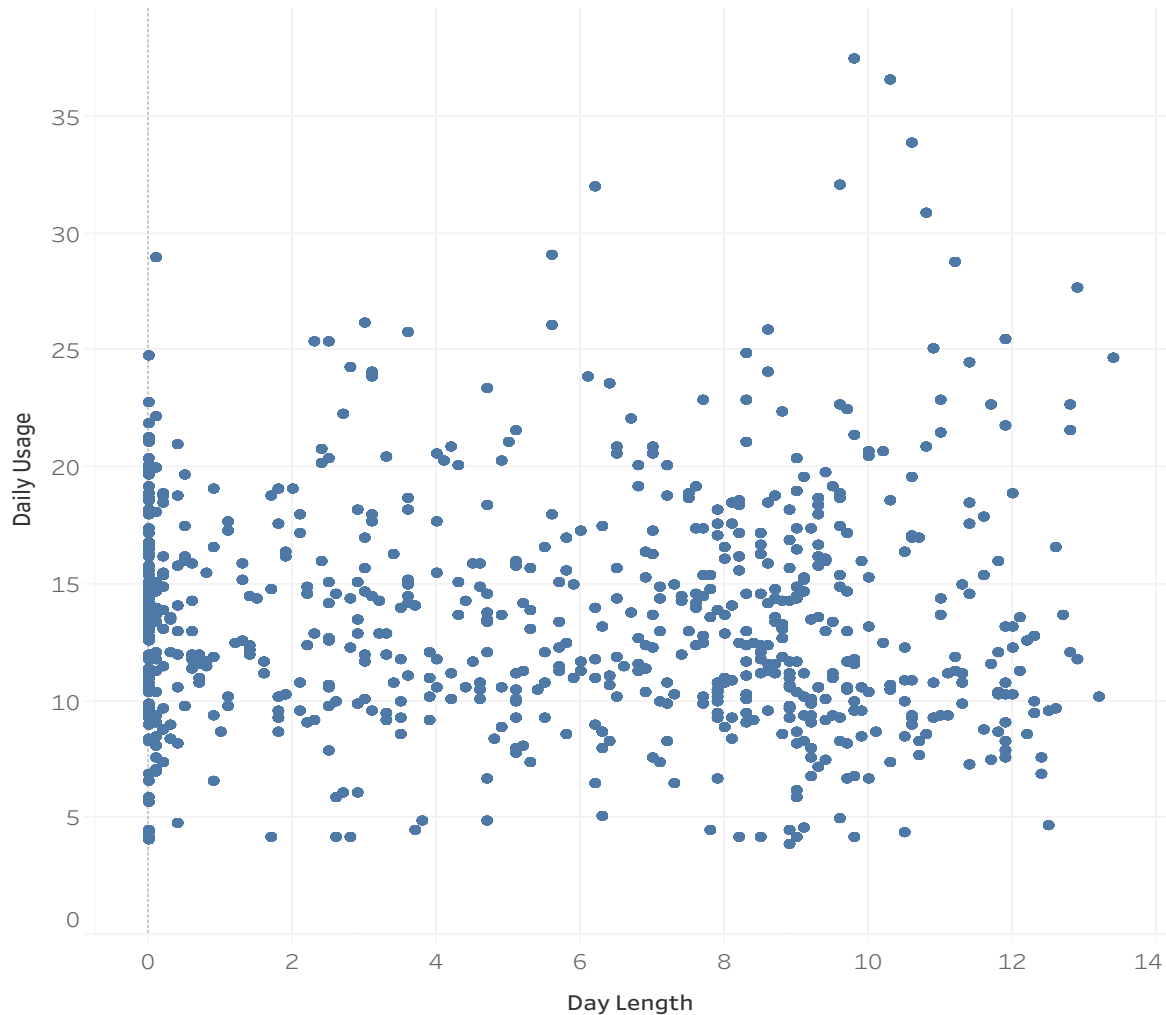
Fig: 1 日の使用量とその日の最高気温

周囲温度では平均消費量が少ないことがわかります。一方、消費電力は、毎日の最高気温の増加とともに増加します。

日中の時間と毎日の消費電力の関係

これら 2 つのパラメーターの関係を見つけるために、異なる日長での毎日の電力消費データの散布図を生成してみましょう。

day_length - daily_usage



Day Length vs. Daily Usage.

Fig: 毎日の使用量と夏時間

上の図から、日中の時間と毎日の消費電力データとの間に明確な関係はありません。

異なる平日の電力消費

別の曜日の消費電力を確認しましょう。

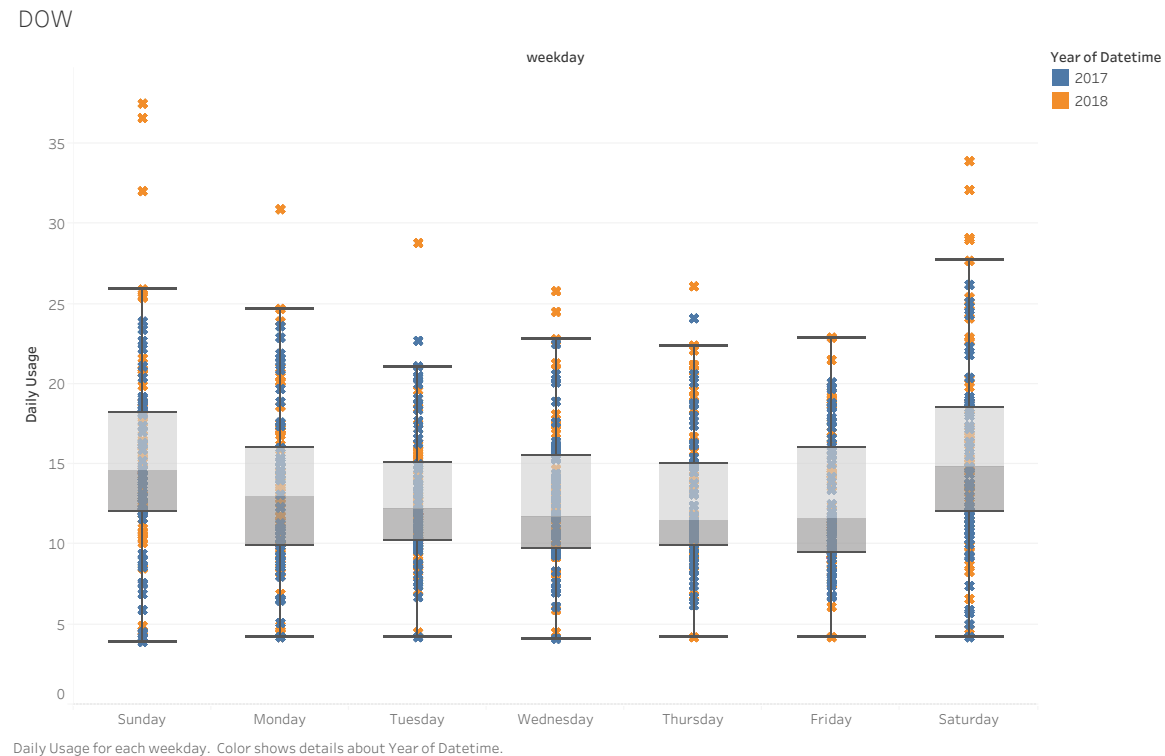


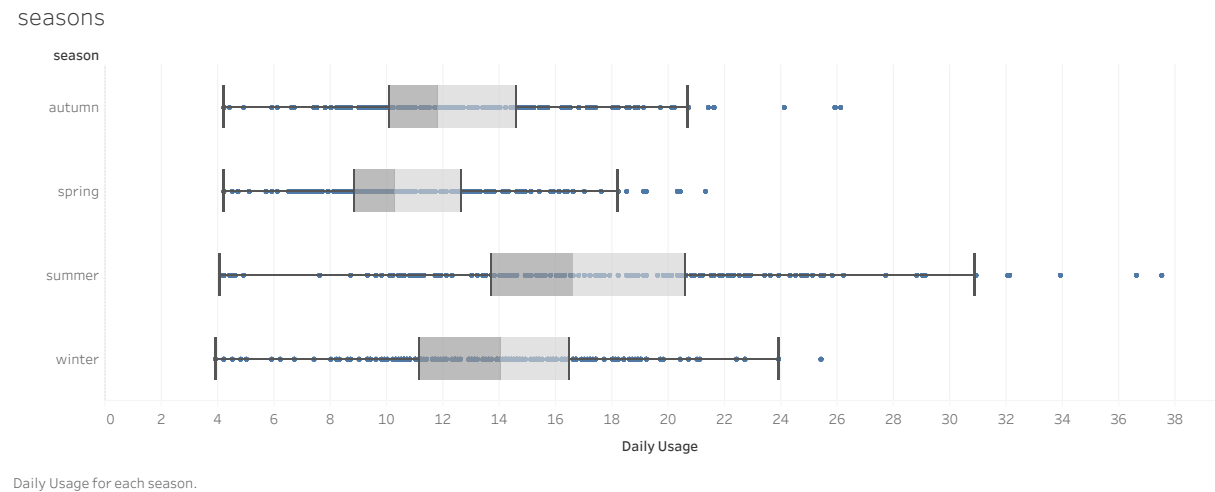
Fig: 異なる平日の毎日の電力消費

上記のボックスプロットから、土曜日と日曜日の1日の電力消費量が多いことがわかります。これらは日本では毎週の祝日であるため、家庭では消費量が多くなります。他の平日では、この分布は非常に似ています。週末には極端な異常値も見られます。これは、それらの休暇中のゲストの訪問が原因である可能性があります。

電力消費における季節と天気の影響

指定されたデータセットにはシーズン列が含まれていません。ただし、年の月情報を使用して季節情報を簡単に抽出できます。

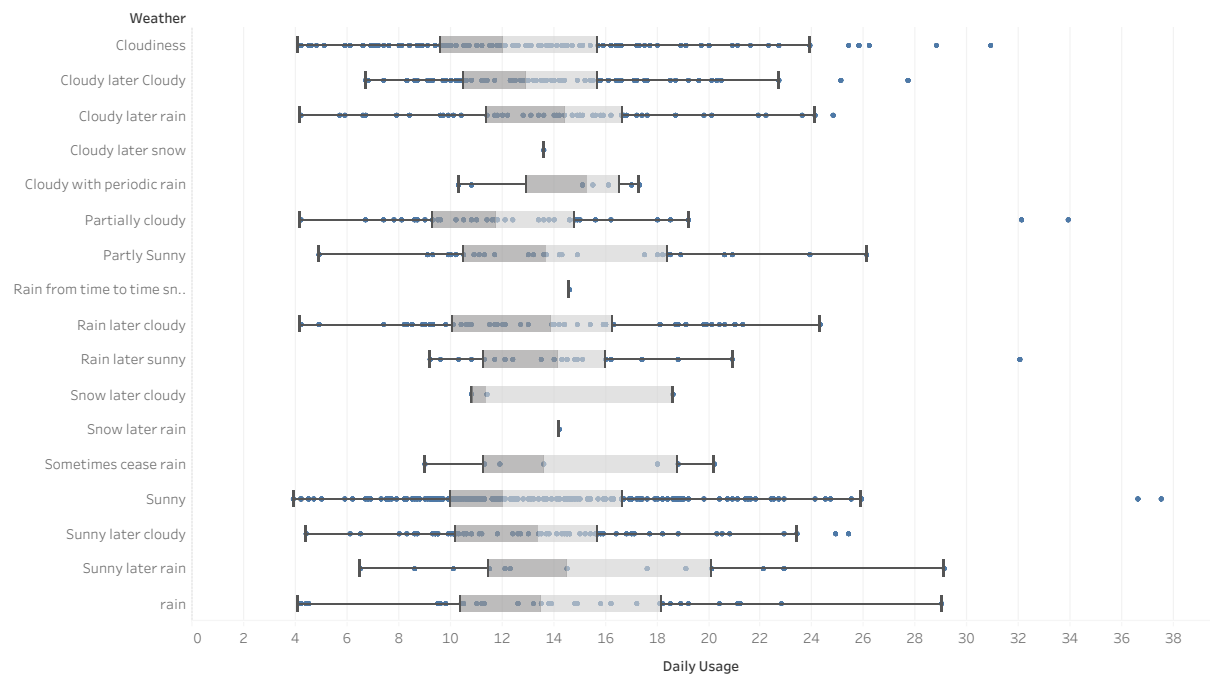
次に、箱ひげ図を使用して、季節ごとの1日の電力使用量の分布を調べてみましょう。



上記の箱ひげ図から、消費電力は夏に最も高く、春に最も低いことが明らかです。冬の消費量は2番目に高くなっています。この背後にある主な理由の1つは、夏にエアコンを使用し、冬にヒーターを使用することです。

次に、さまざまな気象条件での毎日の使用量の分布を確認しましょう。

weather

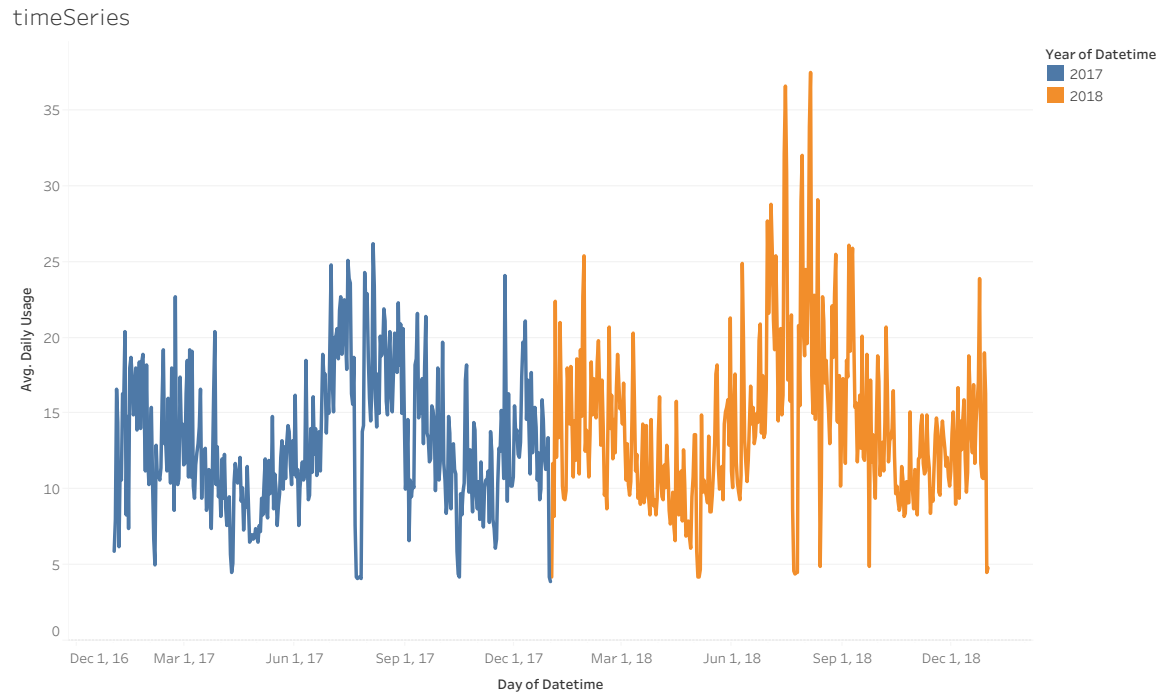


Daily Usage for each Weather.

気象条件に関しては、データセットで利用可能な 17 の異なる条件があります。この場合、分布の広がりは大きくなります。雨が含まれる気象条件では、中央値が高くなっています。

周期性

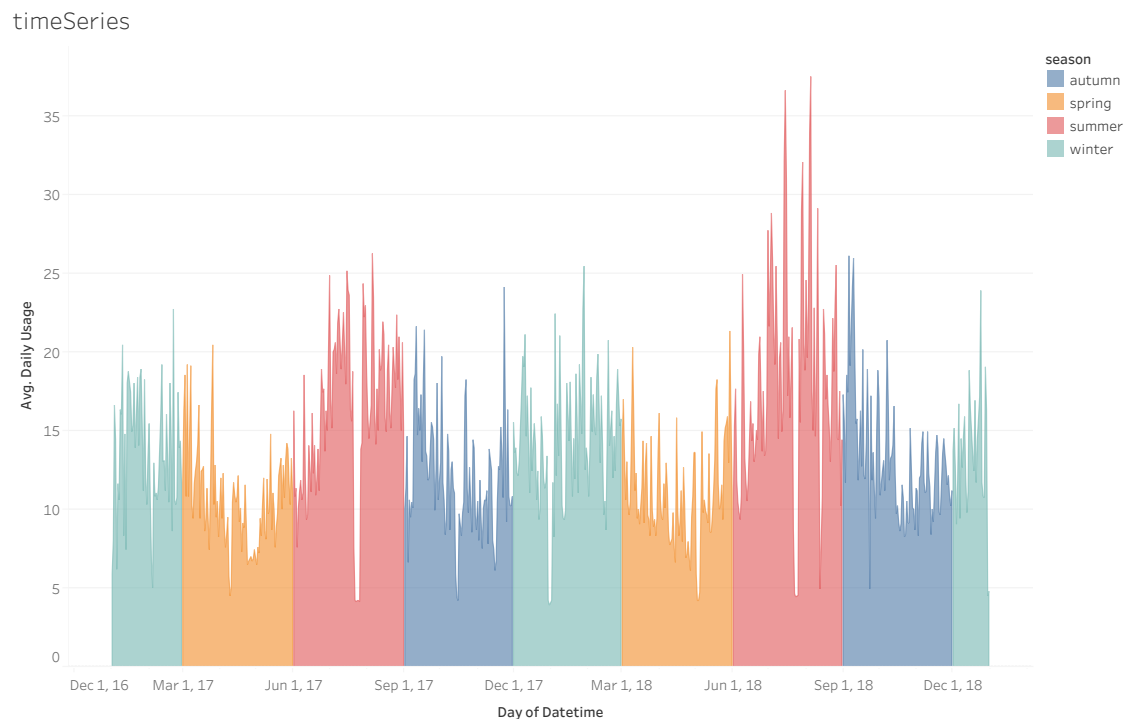
1日の時系列を1日の電力消費データにプロットしてみましょう。2017年と2018年のデータでは、非常に明確な年次周期を確認できます。



The trend of average of Daily Usage for Datetime Day. Color shows details about Datetime Year.

Fig: 毎日の電力消費データの年間周期

Seasonality



The plot of average of Daily Usage for Datetime Day. Color shows details about season.

Fig: 1日の電力消費に対する季節の影響

最後の図では、明確な季節効果がデータに表示されています。

2年間でこの家の電気の使用に何か変化がありましたかデータ？

2017年と2018年の月ごとの総消費電力を比較しましょう。

trend

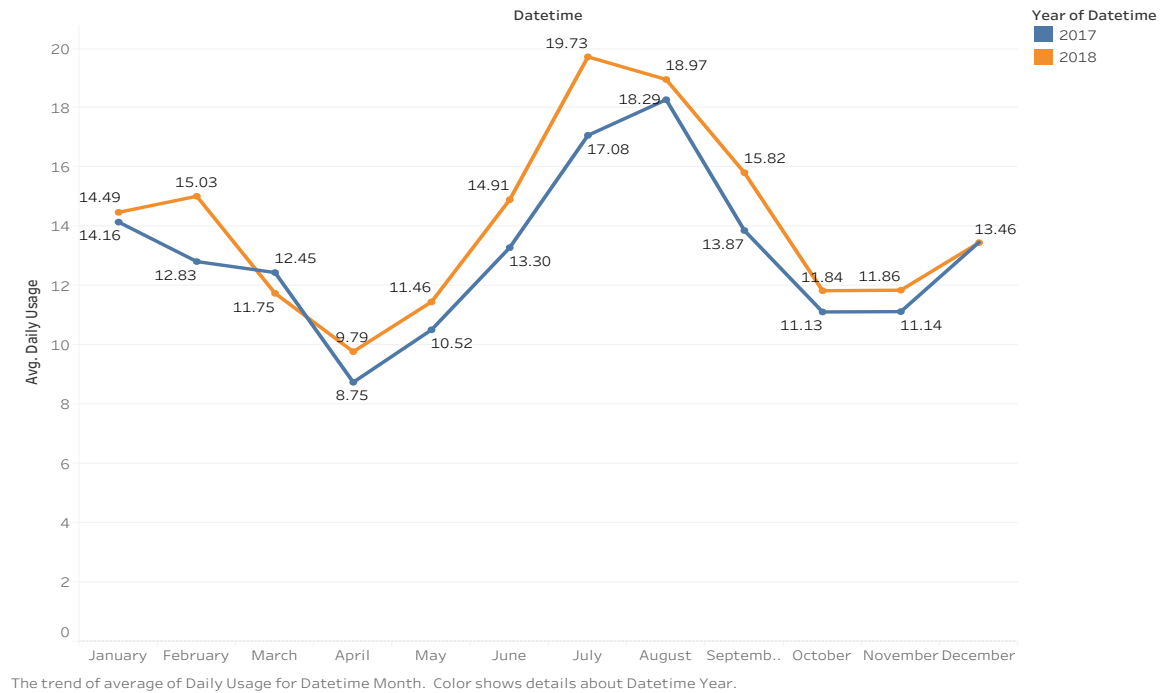


Fig: 2017年と2018年の月間総消費電力の比較

上記の図から、2018年の毎月（3月を除く）に総電力消費量が前年に比べて増加していることがわかります。

注目すべき点の1つは、増分が夏のシーズン中に高くなり、冬の期間に低くなることです。これは地球温暖化の影響でしょうか?? 与えられたデータから確実に言うことはできません。しかし、それは興味深い研究になる可能性があります！

将来のこの家の電気使用量の予測

ARIMA のような手法を使用して、将来の電力使用量を予測できます。ここでは、Facebook が時系列予測用に開発したライブラリ「PROPHET」¹を使用して、今後 1 年間の予測が生成されます。預言者は、季節効果が強く、いくつかの季節の履歴データがある時系列で最適に機能します。

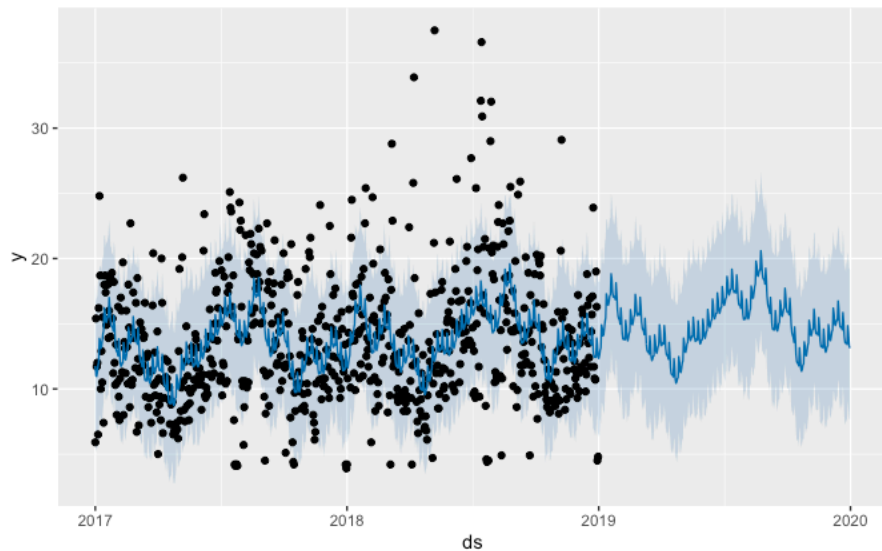


Fig: PROPHET を使用した 2019 年の電力使用量データの予測...

また、預言者で生成された予測モデルのデータの傾向と季節性を確認できます。

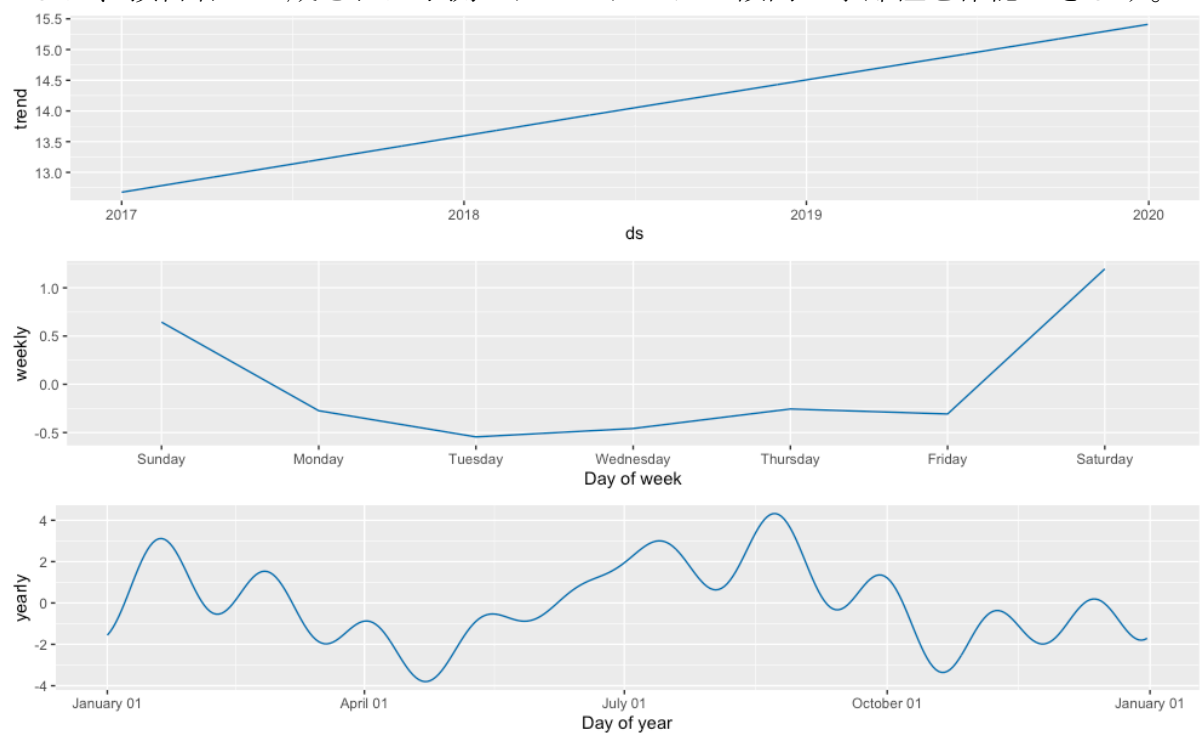
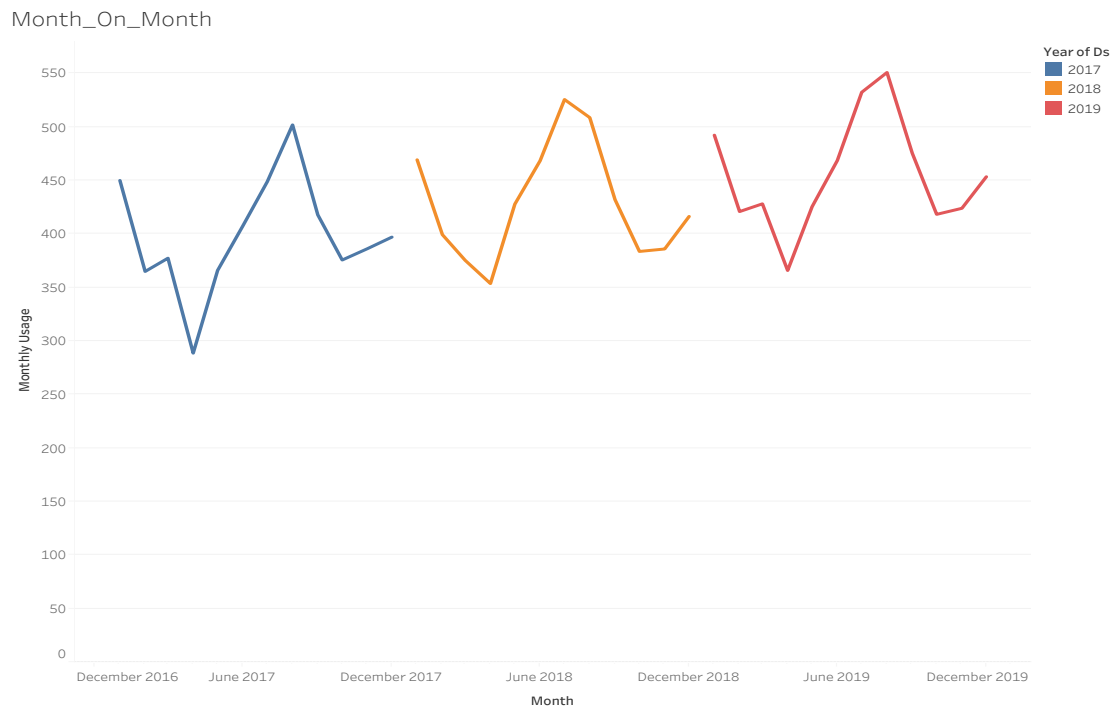


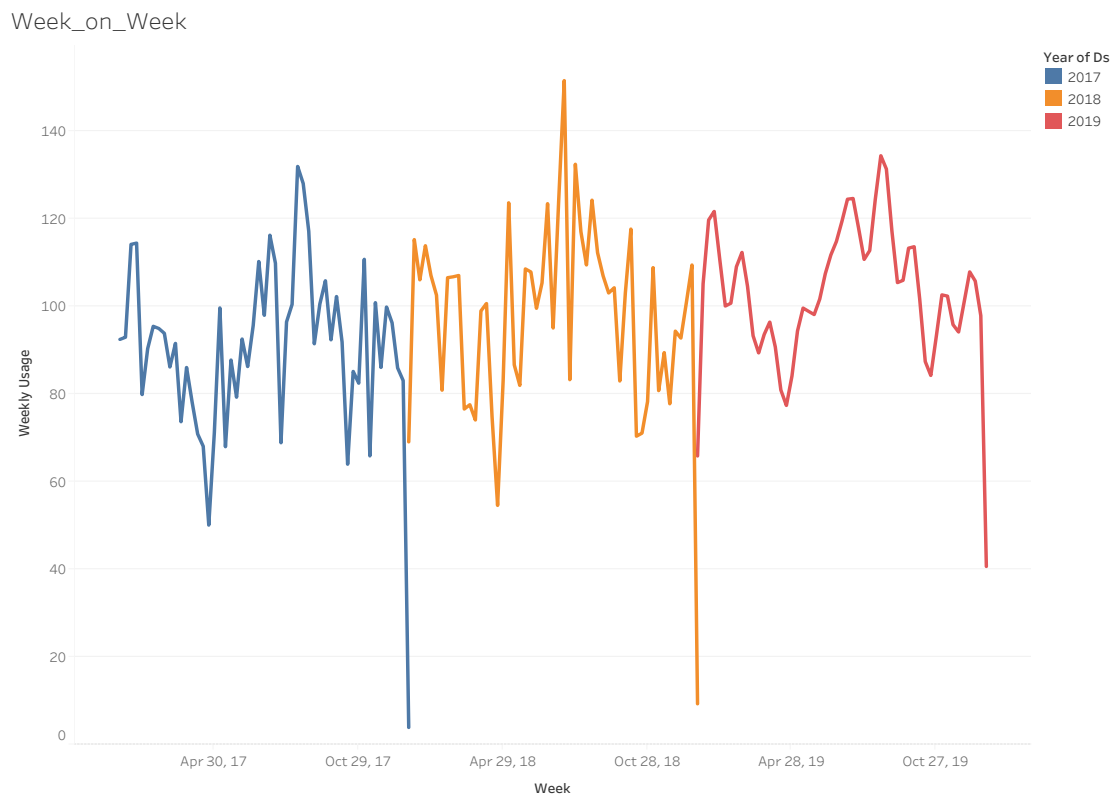
Fig: 予測モデルの傾向、

週次および年次の季節性月単位または週単位の消費電力データを予測することもできます。以下は、Tableau ソフトウェアを使用して生成された 2019 年の月間および週ごとの総消費電力の簡単な予測です。



The trend of sum of Y for Ds Month. Color shows details about Ds Year.

Fig: 2019 年の週ごとの総消費電力予測...



The trend of sum of Y for Ds Week. Color shows details about Ds Year.

Fig: 2019 年の週ごとの総消費電力予測

予測モデリング

将来の消費電力を予測する別のアプローチは、特定のデータセットの予測モデリングです。利用可能な抽出されたカテゴリおよび数値の機能を使用して、時間レベルまたは日レベルの消費電力を予測するためにトレーニングすることができます。これには、重回帰、SVM、ツリーベースのモデル、またはランダムフォレスト/XGBoost / LGBM などのより高度なアンサンブルモデルを使用できます。ただし、利用可能なデータは 2 年しかないため、これらのアルゴリズムを使用してこのデータを適合させるのは少し難しいかもしれません。

家の人数

2013 年の日本の 1 人当たりの家庭での電力消費量は約 2245 kWh /年 2 でした。それ以降、2017 年と 2018 年まではほとんど変わらなかったと仮定しましょう。

最低所得層の平均電力消費量は 1 か月あたり 272.68 kWh です。

一方、最高所得層のそれは 1 か月あたり 571.67 kWh です 3。

2017 年の家の総消費量= 4781 kWh

家のおよその人数= $4781/2245 = 2.12$

平均月間使用量は、これが中所得層から高所得層の世帯であることを示唆しています。したがって、電気の平均消費量は概算よりも高くなります。したがって、その家の人数は 2 人と想定できます。

彼らはどんな仕事をしていますか？

休日と週末を除く時間ごとのデータの分布を確認しましょう。

hourly_usage_distribution

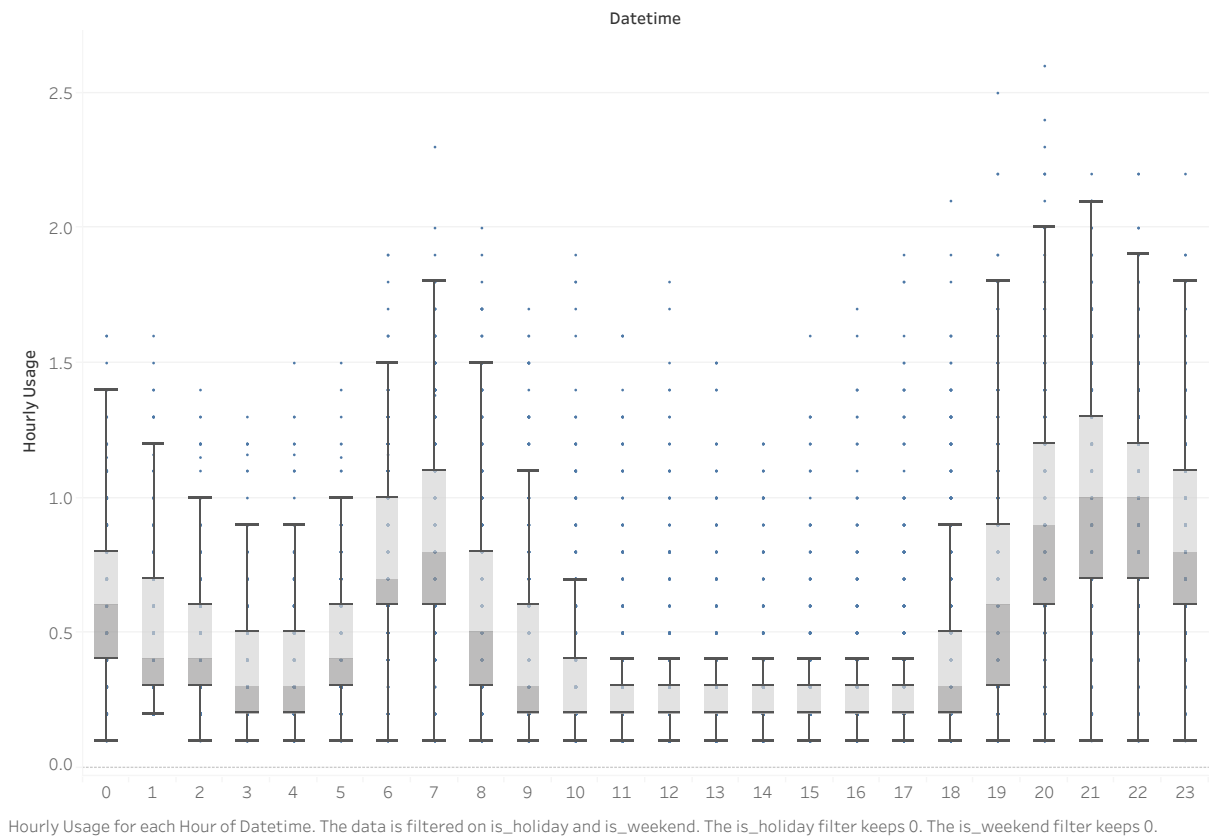


Fig: 異なる時間での時間ごとの電力消費の分布

分布には非常に明確な傾向が見られます。朝と夜には高い需要があります。一方、営業時間中（午前9時から午後6時まで）の需要は非常に低くなっています。つまり、家に住んでいる両方の方は、営業時間中は家にいないままです。両方とも会社員かもしれません。

参照資料:

1. <https://github.com/facebook/prophet>
2. <https://www.statista.com/statistics/597901/household-consumption-of-electricity-per-capita-in-japan/>
3. <http://www.waseda.jp/prj-ricem/dp/dp1803.pdf>

