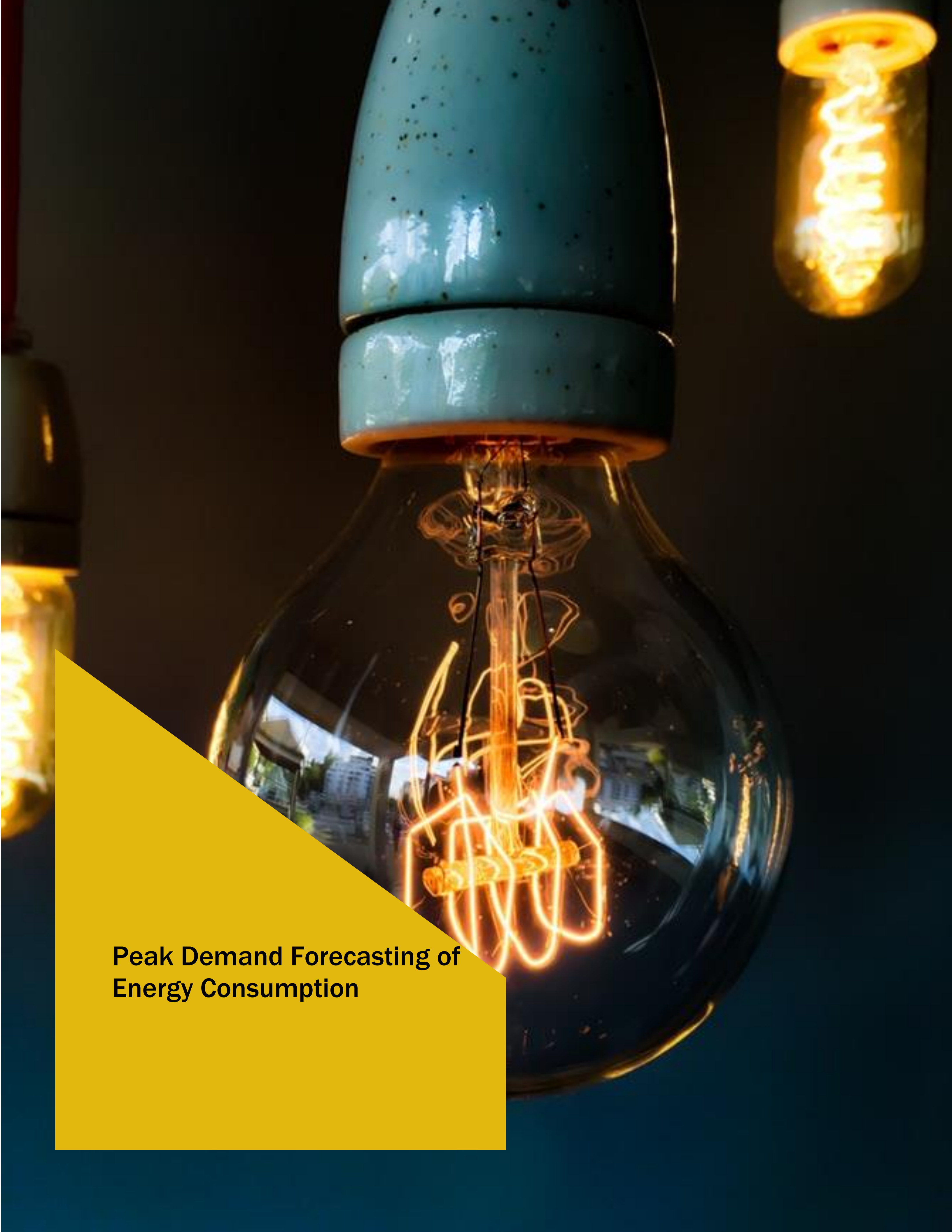


Peak Demand Forecasting of Energy Consumption

Summary

The main objective of the data challenge is to perform predictive analysis on energy demand to for better decision-making process. In this challenge, we defined the problem in demand management, and collected data from open sources for understanding and analysis. Important features are then selected for building our energy demand model. Decision making for future new business opportunities can be enhanced by leveraging the forecasting result from our model.

1. Problem and Motivation:

Like most countries in the world, nuclear power constitutes the largest portion of French energy production. With the Act on energy transition for green growth, France committed fully to energy transformation, with an objective of national renewable energy share of 35% in 2030. With wind energy being one of the highest renewable energy sources in national energy mixes, it is important for energy producers and operators to predict customer demand for flexible production and better demand management, and more importantly, to avoid any non-compliance on the changing regulations.

In this data challenge, we identify the problem faced by energy operators –How can we forecast energy consumption, and manage production and demand efficiently? By using open data, we built a model to predict energy demand in one year, with peak and low period of demand identified. Based on the result from the model, we will discuss how data analysis can be used for creating future business opportunities.

2. Approach:

We identified the business problem of identifying the peak periods during the electricity production so that excess electricity could be stored and used at later times such as maintenance or high demands. For this we went through the many datasets at opendatasoft.com and analyzed what could be the best datasets for our approach. These datasets are named and explained in the later sections of this report. We did a lot of brainstorming and created a task sheet to accomplish this solution after breaking down in subtasks.

The approach for forecasting the peak period in electricity production:

I. Perform descriptive analytics and study current regulations - Descriptive analytics is a preliminary stage of data processing that creates a summary of historical data to yield useful information and possibly prepare the data for further analysis.

II. Identifying factors for forecast of peak period electricity production - Based on the study of certain articles and understanding the datasets, we figured out following factors play a favorable role in the identification of the peak period forecast

   A. Temperature

B. Humidity
C. Wind Speed
D. Time
E. Production data
F. Consumption data

III. Correlation Analysis - We used python to eliminate the highly correlated variables (factors), before finalizing the above-mentioned factors.

IV. Building the prediction model using Python

3. Datasets

The opendatasoft datasets were used to perform exploratory data analysis and build a predictive model. Following datasets are used:

- Energy Consumed by different industries like rail transport, supply chain industries, Steel and chemical industries.
- Energy Consumed by different segments like residential, commercial, small and medium enterprises, large industries.
- Energy consumed on an hourly basis, daily and annual basis.
- Temperature and Energy Consumed: the daily consumption of energy along with reference temperature values and average temperature values is used as input to model.

4. Tools & Analytics

Before building up our model, we used Tableau, python and Dataiku to understand the data collected.

- Tableau is used to study the insights and patterns in the energy consumption across different segments, sectors.
- Python is used for data cleaning, preparation along with model building and evaluation
- Dataiku is used for building the baseline model and comparison of different models such as random forest, neural network, decision trees, etc.

First, we identified the energy consumption by segment, energy consumption by industry, and peak and trough distribution using the past data. Of all elements, there is a highest correlation between energy consumption and outdoor temperature. **A piecewise regression** is used, the independent variable i.e. energy consumption is divided into intervals and separate segments are fitted for each interval. In this case the energy consumption was divided into two major segments i.e. winter part and summer part. The significant change in temperature in winter leads to increase consumption of energy whereas in Summer, even though the temperature increases the consumption of energy remains constant

By understanding the data, we included the most important features in the model for a more accurate prediction.

After data understanding, we started to build our model for demand prediction using the following algorithm with Python.

1. **XGBoost:** Extreme Gradient Boosting is a very popular algorithm which is based on the ensemble methodology of trees. The objective function is a combination of training loss and regularization. It is a CART (Classification and Regression Trees) algorithm i.e. used for classification and prediction purposes as well.
   Using external environmental factors, the relation between temperature and energy consumption is studied using regression and correlation analysis. Using past patterns in energy consumption along with external factors like temperature and other features generated the algorithm is applied to predict the energy consumption for future days.

2. **Anomaly Detection Methods:** In order to identify the higher demand and lower demand of energy, the anomaly detection methods are used to determine the lower peaks and higher peaks. In this case, the QuantileAD [1] method is used to determine the peaks. QuantileAD compares each time series with its historic quantile and identifies the values which are above specified values. For determining the higher peaks, we have set the value for high as 96%. i.e. when consumption of energy increases above 96 % percentile as compared to data from 2012 to 2019, it is considered as peak.
   The forecast of temperature is added as input to the algorithm, in order to generate the forecast of energy consumption. The anomaly detection methods are applied on top of entire data including forecast to identify the peak period of energy consumption.

5. Results

The past data was spitted into training and test sets. Models are evaluated based on accuracy metrics such as Mean Absolute Percentage error (MAPE). MAPE of 3.07 was observed while forecasting the energy consumption on the test dataset (2019).
In order to determine the accuracy of peak forecast, the actual peak days (obtained from RTE website) are compared with the predicted results and the algorithm can predict for 75% cases (i.e. 9 out of 12 days for year 2019).

6. Contributions and Uniqueness

The key insights gained from the data analysis are:

- Residential segment has the highest energy demand. (Refer Fig 2)
- Among the commercial industry, operational or supply chain has the highest energy consumption.
- There is a high correlation between outdoor temperature and energy demand, where the peak period is in Jan, and low period is in August summertime. (Refer Fig 3)
- Good prediction on peak period demand, with an accuracy rate of 75%.

7. Appendix: Visualization / analytics summary

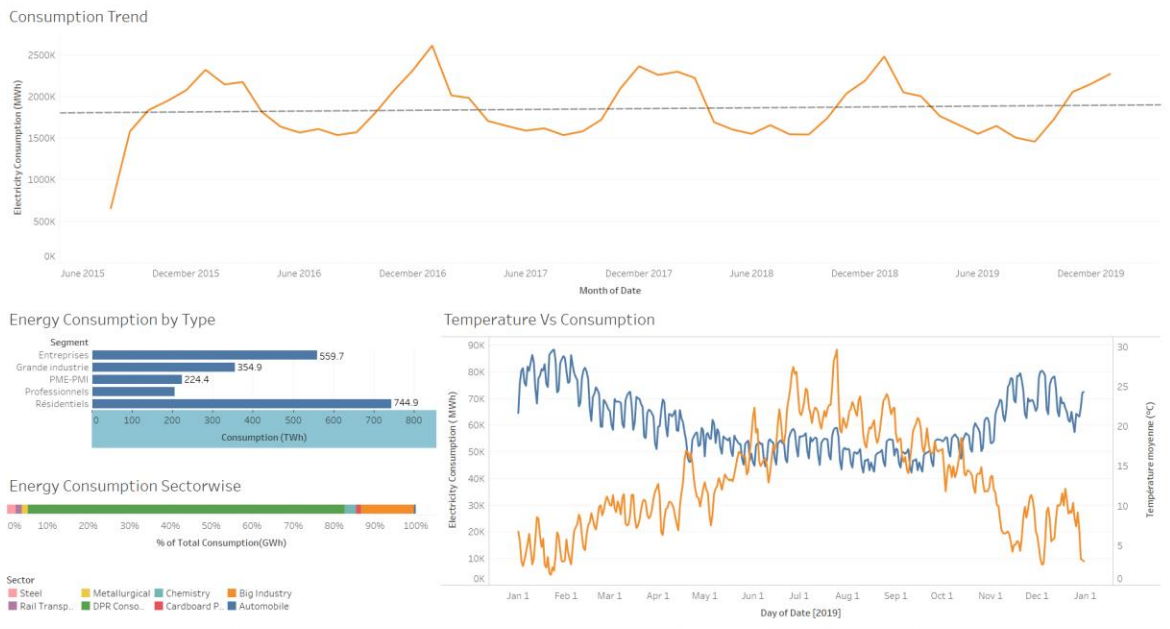


Fig 2: Dashboard showing insights about energy consumed in different sectors

- The dashboard above shows a trend of energy over the period. It can be seen annually the residential sector consumes 744.9 TWh which is highest amongst all other sectors.
- 77% of energy is consumed by operational industries such as supply chain industries.
- By analyzing the graph of Temperature vs Consumption, it can be observed that as the temperature increase the energy consumption decreases. Furthermore, the demand of energy is lowest during the summer month August and highest during winter period in month of January and February

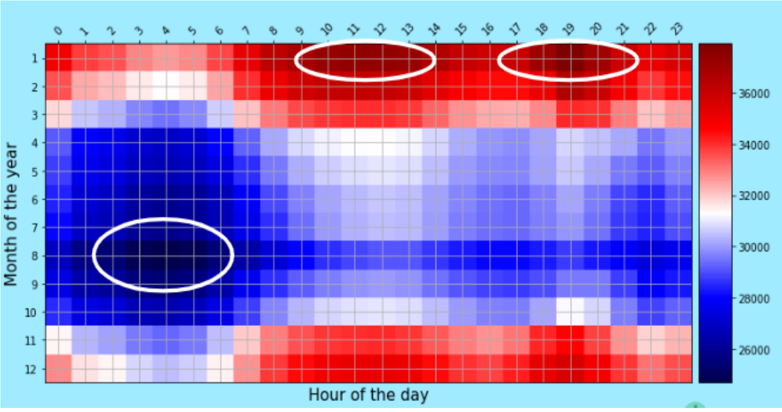
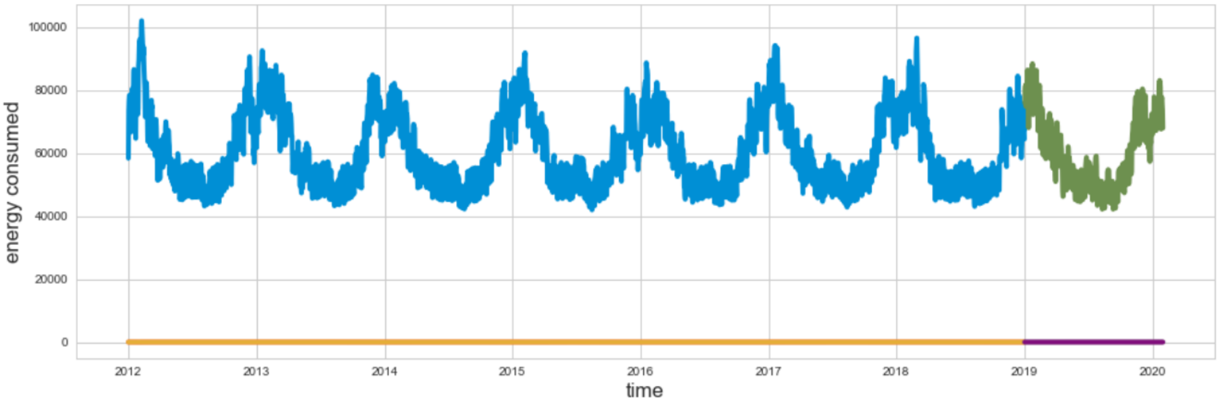
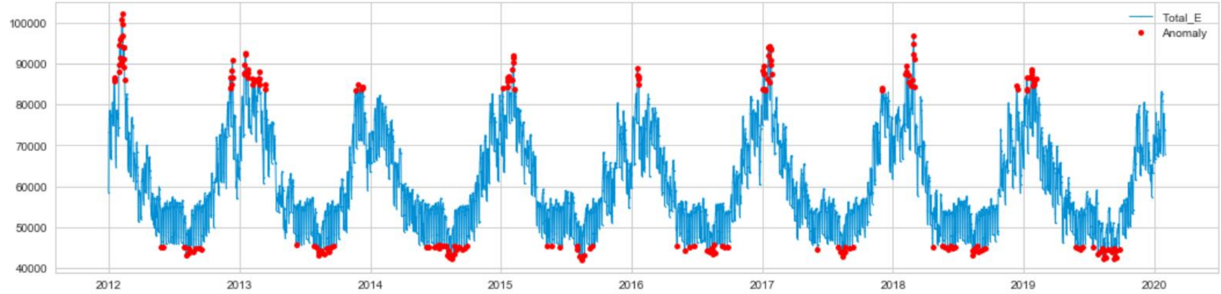


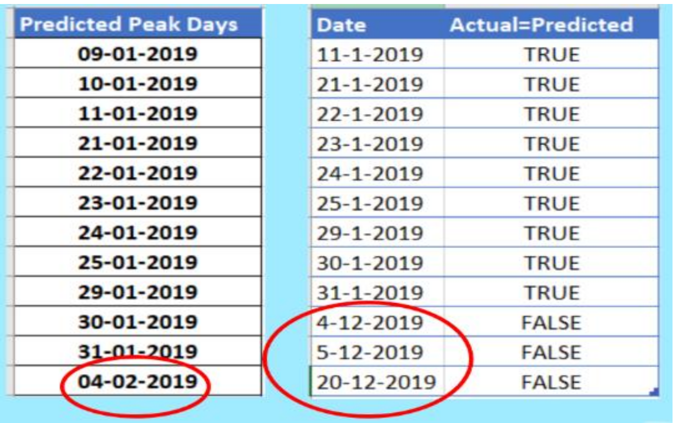
Fig 3: Heatmap showing energy consumption during month of year and hour of day

From data analyzed from 2012 to 2019, it can be observed that the Highest Energy Consumption is in January is in Morning from 9:30 to 13:30 and in Evening from 18:30 to 19:30. The Lowest Energy Consumption is in month of August from 2:30 to 5:30 am





The above viz demonstrates the forecast of energy consumption (in green) for future periods along with peaks detected by the algorithm. From above graph the algorithm can not only help understand the high peaks, but also predict the low peaks to suffice the low demand of energy.



Predicted Peak Days
09-01-2019
10-01-2019
11-01-2019
21-01-2019
22-01-2019
23-01-2019
24-01-2019
25-01-2019
29-01-2019
30-01-2019
31-01-2019
04-02-2019

Date	Actual=Predicted
11-1-2019	TRUE
21-1-2019	TRUE
22-1-2019	TRUE
23-1-2019	TRUE
24-1-2019	TRUE
25-1-2019	TRUE
29-1-2019	TRUE
30-1-2019	TRUE
31-1-2019	TRUE
4-12-2019	FALSE
5-12-2019	FALSE
20-12-2019	FALSE

The above viz shows comparison of actual and predicted peak days from the algorithm. Out of 12 days the algorithm has accurately predicted 9 days of peak and missed 3 days of peak.