Autor: Pablo Chamena
Comisión: 49140

Smart House: Electrical consumption

How does everyone's home use electricity?

In a world where energy efficiency becomes increasingly crucial, a smart home is the pinnacle of efficiency, control and savings.

Energy would not be wasted, every watt will be used efficiently. Lights will automatically adjust to natural light, appliances will turn on only when needed and the temperature will remain stable effortlessly. This is a goal to achieve and it may seem long but it is something that can become a reality thanks to artificial intelligence and data analysis.
At the heart of this smart home is a dataset that keeps the secrets of energy consumption.

Index

1) Context and audience
2) Hypothesis
3) Content
4) Exploratory analysis
5) Insights
6) Models and Results
7) Conclusions

1) Context and audience

The project focuses on optimizing the use of electrical energy, improving electrical load distribution, identifying potential energy quality problems, meeting sustainable objectives and helping to ensure compliance with regulations and standards.

To do this, an exhaustive analysis of a dataset with data on the electrical consumption of a house was carried out. The application of machine learning models to achieve prediction was also evaluated.

This analysis is focused on people within the energy field:

Industry sectors such as: manufacturing industry, energy sector, renewable energy sectors, energy service providers (Companies that supply electricity), etc. As well as people who are interested in knowing how energy consumption works.

Limitations: When analyzing the dataset, some conditions will be assumed within the theoretical framework:

The data set is located within an alternating system, in a steady state, it is a single-phase system, there is the absence of harmonics and the values were measured correctly.

Hypothesis

- The winter period is where higher energy costs are expected.
- With the variables included in the dataset, predictions can be made about home energy.
- The sector of the house that has the highest consumption is the 3rd submetering.

Questions

- Do the calculated (artificial) columns provide any explanation for the energy?
- Which of the variables present in the dataset is most important?
- Which sector of the house uses the most energy?
- Is energy expenditure the same in all time frames (time of year, days of the week or day and night.)

Content

The data set contains information about the energy consumption of the house, including:

- Active energy: The protagonist of our analysis represents the energy actually consumed by the house. kilowatts or kilowatts

- Reactive energy: The invisible energy that is not converted into useful work, but that generates losses in the system. kW

- Voltage: The force that drives electric current. Volts

- Current intensity: The amount of electron flow per unit of time. Amps

- Submeasurement 1: Corresponds to the kitchen, which mainly contains a dishwasher, oven and microwave. Watts/Hour

- Submeasurement 2: Corresponds to the laundry room, which contains a washing machine, dryer, refrigerator and light. Watts/hour

- Submeasurement 3: Corresponds to an electric water heater (hot water tank) and an air conditioner. Watts/hour

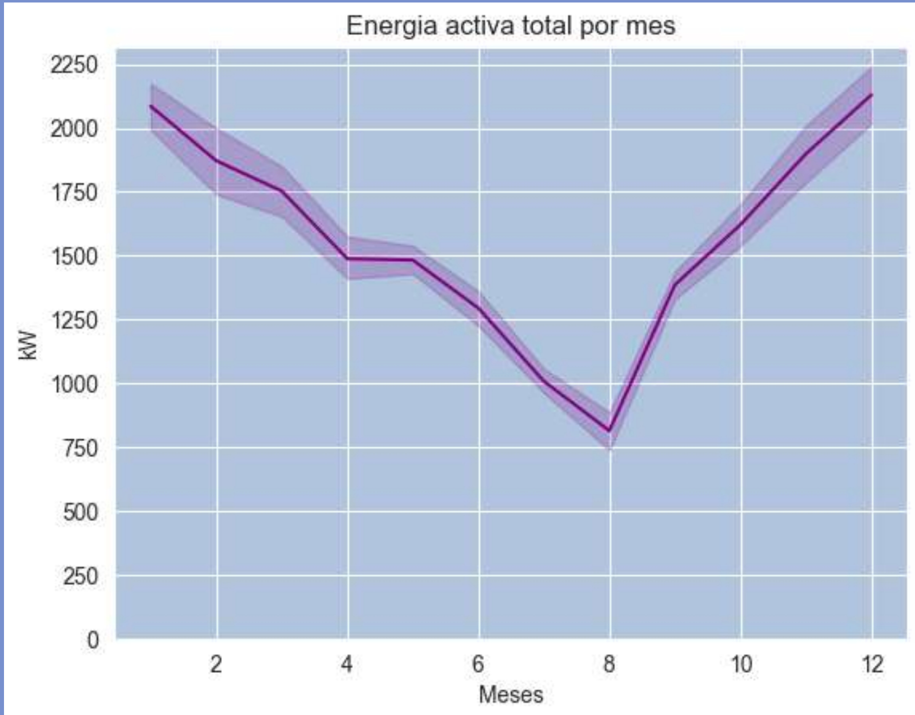
Exploratory analysis

Active energy is what is actually consumed and translates into the cost of the electricity bill. It is what is used to do useful work.

Furthermore, together with reactive energy, voltage and intensity, they are key indicators for the analysis of the quality of supply.

The start and end dates of the seasons of the year in France are:

-Winter: from December 21 to March 20 (1st quarter)
- Spring: from March 21 to June 20. (2nd quarter)
- Summer: from June 21 to September 20. (3rd quarter)
- Autumn: from September 21 to December 20. (4th quarter)

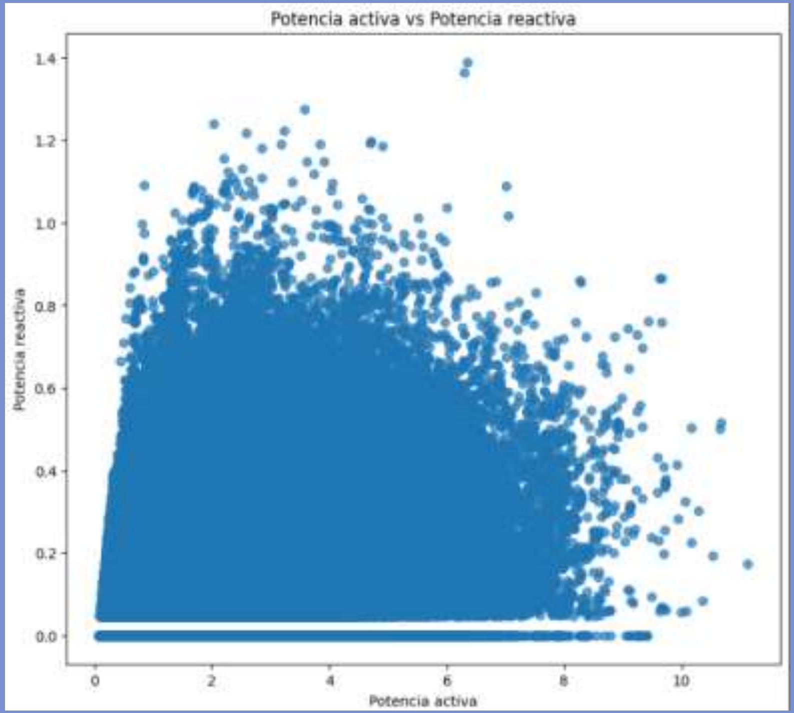


Exploratory analysis

Efficiency

This column corresponds to the time intervals of the instances, that is, to each minute, where the energy consumed is only active energy, therefore this column will have values of 1 when reactive energy is 0 (Q=0), and values of 0 when the reactive energy is unequal 0(Q!=0)

We see that 76.7% of the time is where the energy provided cannot be fully used, on the contrary 23.3% if it is managed to be used efficiently, that is, almost a quarter of the time.



What variable is the dataset analysis going to focus on?

Our analysis will focus on predicting and interpreting active energy on the one hand, because it is what is actually consumed and translates into the cost of the electricity bill, in addition to being key to energy efficiency, facility sizing, analysis of the quality of supply and the planning and management of the electrical network.

Our second analysis will focus on how much of the time we consume energy is actually used. For this purpose, the calculated efficiency column was created and it is the one that we will use to make a prediction about it and evaluate the energy efficiency.

This is relevant because it could be applied in some cases such as:

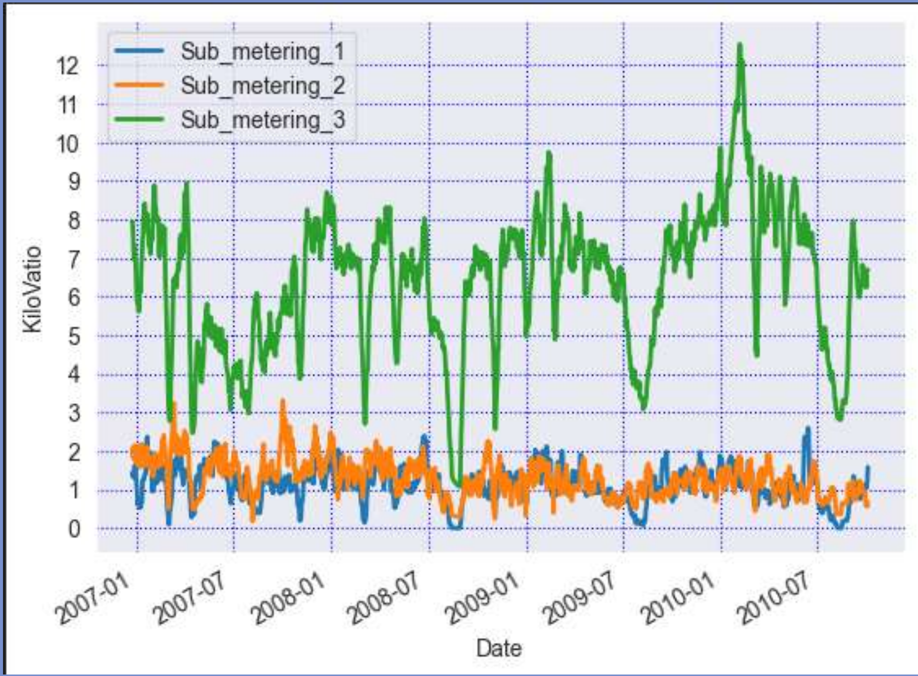
- An energy supplier can use active energy prediction to determine the amount of energy it needs to generate and distribute at any given time.

- A business can use active energy prediction to identify opportunities to reduce its energy consumption and save money on its electricity bills.

- A power grid operator can use active energy prediction to identify potential grid overloads and take measures to avoid them.

Which sector of the house has the highest consumption?

Submeter 3, which is made up of an electric hot water tank and an air conditioner, is the one with the highest expense.



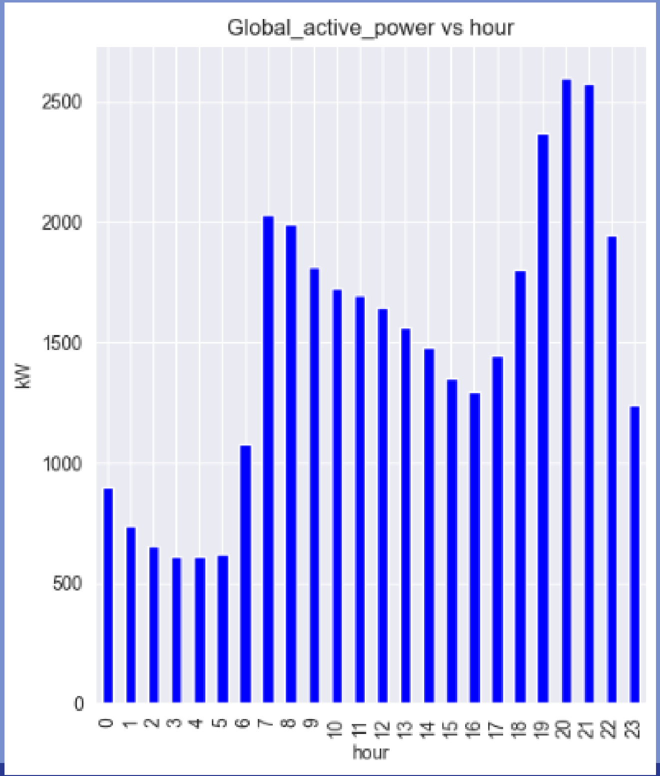
The type of load, whether inductive, capacitive or resistive, of the elements in question should be analyzed to give a better analysis.

Time frames: Active energy.

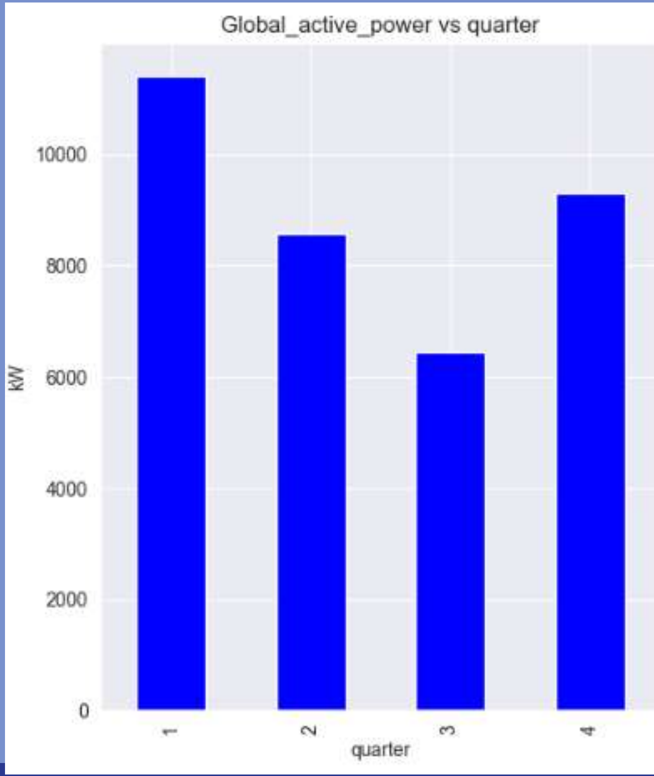
The highest energy consumption is found in the time period 7:00 p.m. to 9:00 p.m. and the lowest consumption is found in the first 6 hours of the day.

Regarding the quarters, it can be seen that the lowest expenditure is found in the 3rd quarter, this being the summer season of the house.

Per hour



By quarter



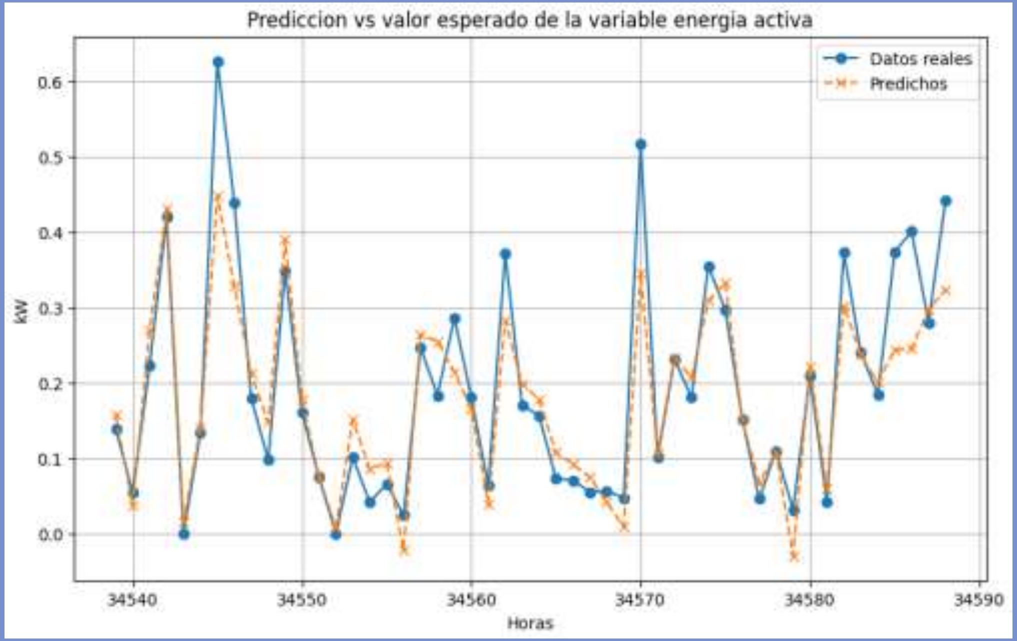
Insights

- The time of year with the highest consumption is in the first quarter or during the winter.
- The time of year with the lowest consumption is in the third quarter or during the summer.
- The afternoon/night time is when the greatest energy expenditure occurs.
- The third submeasurement is the one with the highest expense.
- Several synthetic variables explain, to a certain extent, natural variables.

Models and Results

For active energy, a multiple linear regression model was used where the following metrics were evaluated:

mean square error with 0.218
Root of mean absolute error: 0.3095
Determination coefficient (R^2): 0.8060



Models and Results

For the prediction of our variable of interest "Efficiency", a logistic regression was applied to analyze the influence of the other variables on it.

The prediction on class 1, in this case representing the cases where the reactive energy was 0, which indicate the energy efficiency, has a lower recall, so you could try to apply classification algorithms with greater resources.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.94    313354
           1       0.95      0.63      0.75    101698

    accuracy                           0.90    415052
   macro avg       0.92      0.81      0.85    415052
weighted avg       0.91      0.90      0.89    415052
```

Conclusions

Through the applied models we can conclude that:
It is possible to predict the energy with a certain margin of error, as well as the moments of time where the energy would be best used and to be able to classify them.
The combination of both models can provide a more complete view of the home's electrical energy consumption.

Metrics

Metrics obtained:

Regression:

R^2: 0.828 for instances per minute

R^2: 0.853 for instances per hour

Classification:

Recall: 1 = 0.63   0 = 0.99

F1-Score: 1 = 0.75   0 = 0.94

Accuracy: 0.9