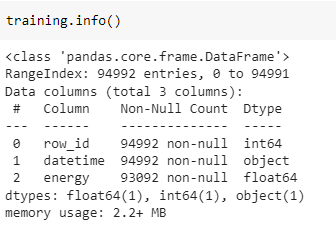
# **Energy Demand Forecasting**

**Objective**  
  
build a machine learning/deep learning approach to forecast the total energy demand on an hourly basis for the next 3 years based on past trends.

**Approach**

**Import data set 🡪 Perform EDA 🡪 Data Pre-processing 🡪 Model Building 🡪 Evaluation 🡪 Final model 🡪 prediction/Forecasting for the required time span.**

**EDA**



From the training data set provided it can be observed that there are a total of 94992 data points provided. with 3 feature name-row\_id , datetime , energy.

In order to extract more information based on the date or time column , The feature datetime should be converted to datetime 64 data type instead of object type.

*Checking for missing values:*

Chart

Description automatically generatedTable

Description automatically generated

Here from the data set we can observe 1900 missing values belonging to energy column. So since it is a time series analysis We have to implement imputation technique instead of dropping the missing values.

So the next step is to decide the imputation method here in the case I have implemented last observation carried forward method which is best suited to problem statement as the energy demand of the next day is closely connected to previous day’s demand.

*Checking for Duplicated values:*

Graphical user interface, application, Word

Description automatically generated

We have observed that there are no duplicated values in the dataset.

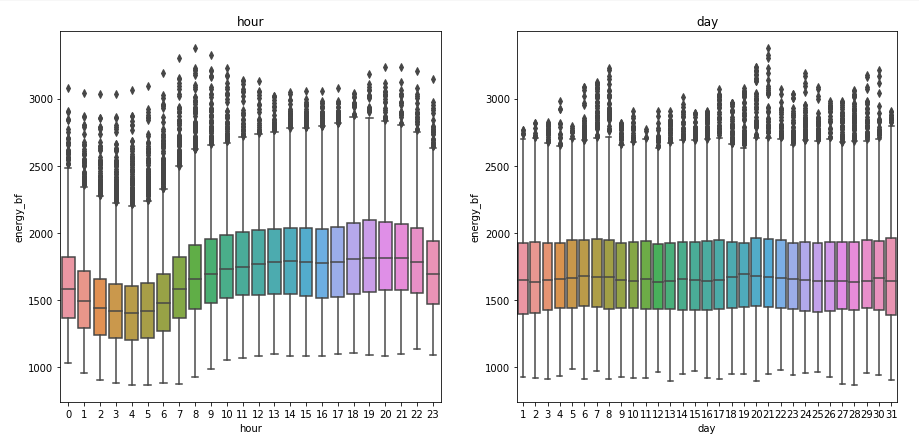
Plots:

Chart

Description automatically generated

Line plot which show the energy demand against the time .

Box plot

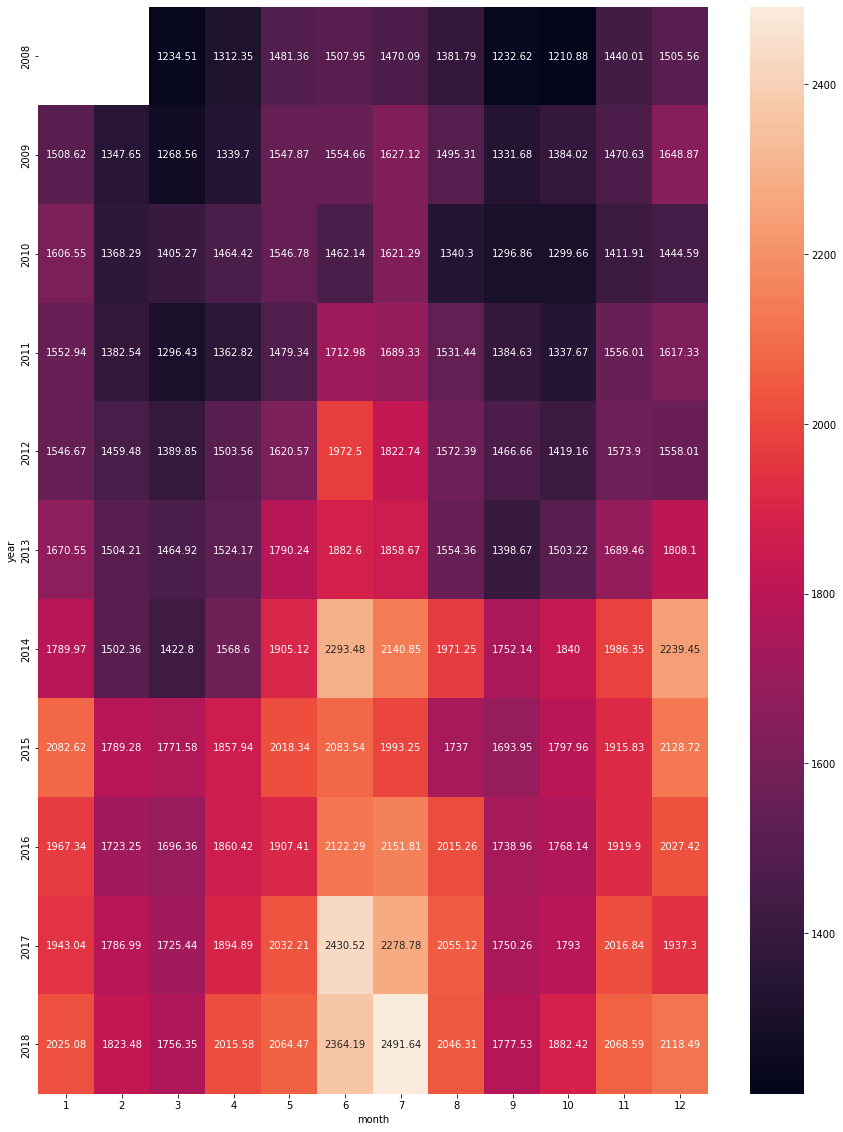


Chart, box and whisker chart

Description automatically generated

From the above plots we can infer that the highest demand for energy is observed : hours basis = 8th hour day basis = 21st day month basis = 7th month year basis = 2018

From the year basis plot, we can see that there is an overall increase in the demand. We can also observe less demand during the 2,3,9,10 month of the year. We can observe almost constant median value when looking at the day wise consumption. Also, lesser consumption during the initial 7 hour of the day.



The highest demand we observed during the 7th month of 2018

Data -pre-processing:

Table

Description automatically generated

* Performed feature transformation’s like taking square of the time parameter(row\_id) for building the quadratic model.
* Extracted month , year, day for additive seasonality model
* These feature extraction helped to get inference/insights on the demand of energy.

Stationary :

Graphical user interface, text, application

Description automatically generated

Performed a  [Augmented Dickey Fuller Test to check whether the data is stationary or not.](https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/)

Model Building:

Model Implemented:

* Linear model
* Exponential model
* Quadratic model
* Additive seasonality model
* additive seasonality quadratic
* Multiplicative seasonality
* Simple exponential smoothing
* Holt winter method
  + Additive seasonality
  + Multiplicative seasonality
* ARIMA

Model evaluation :

Graphical user interface, text, application

Description automatically generated

The evaluation metrics which was taken is Root mean squared error.

The final model is selected on the least RMSE value.

In the Table above we can observe that linear model is giving the least RMSE value but when we look at the plot it is giving a straight line. Where in Arima model a slight trend, seasonality can be observed .