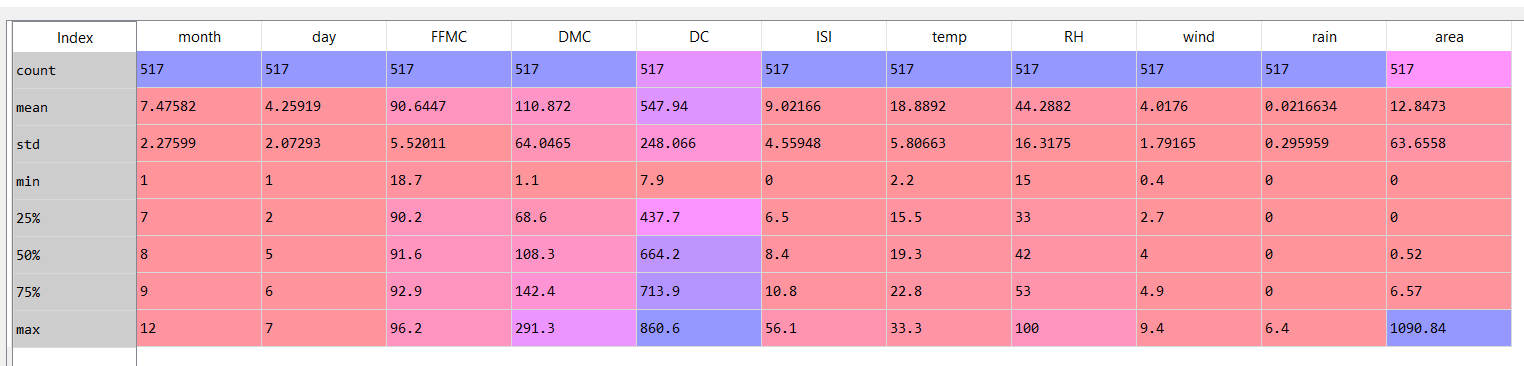
**Case Study of forestfires**

**Problem Statement:**

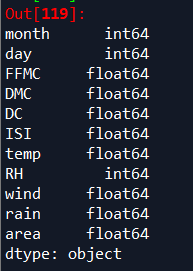
Prepare a prediction model for profit of forestfires data. Do transformations for getting better predictions of profit

**Exploratory data analysis:**

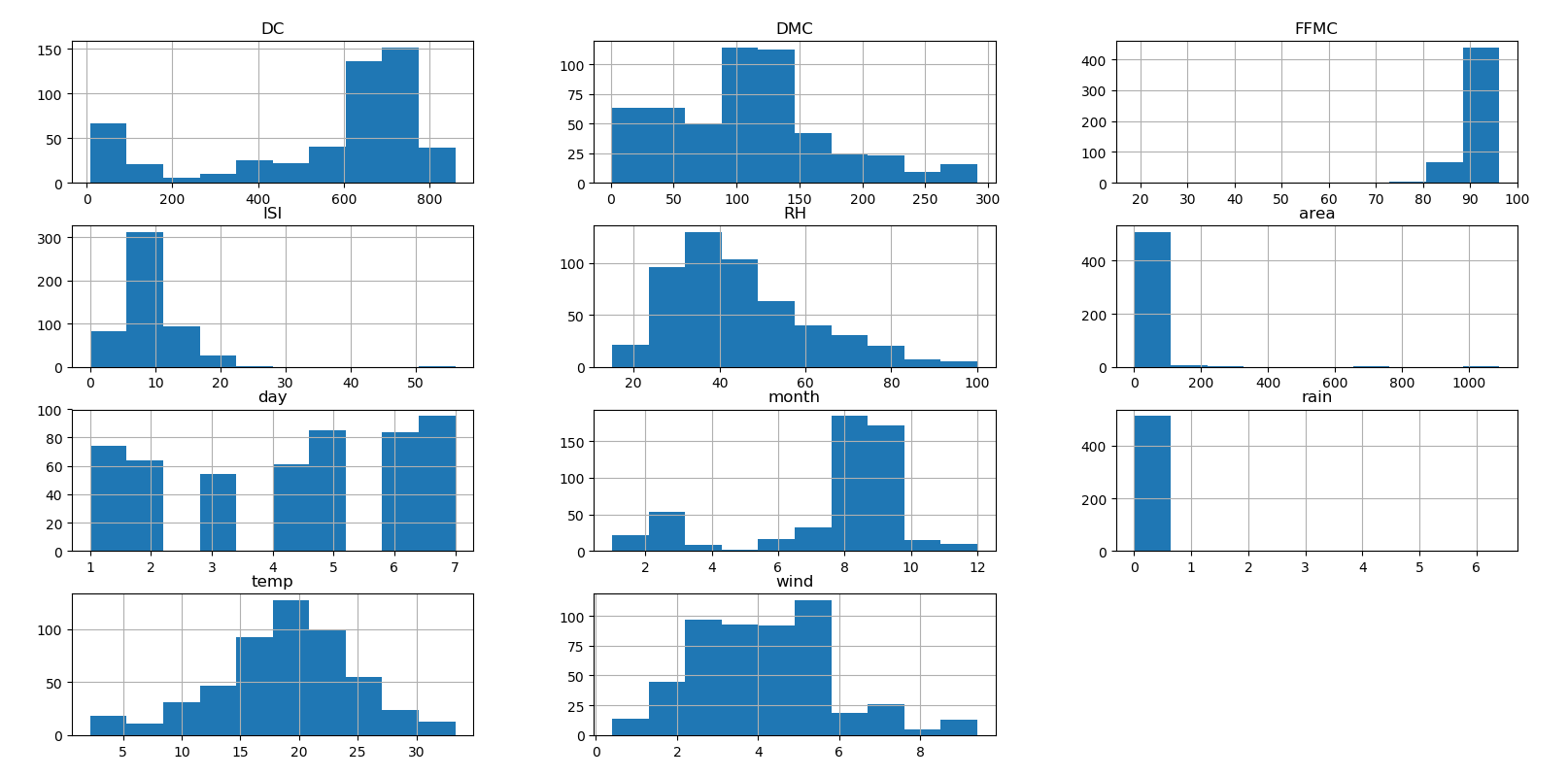


The above is the descriptive statistics of the forestfire dataset. There are 517 observations and 11 variables in the data frame. The missing variable in the above descriptive analysis is the “State” variables. Which is a categorical variable.

The following are the variable types:

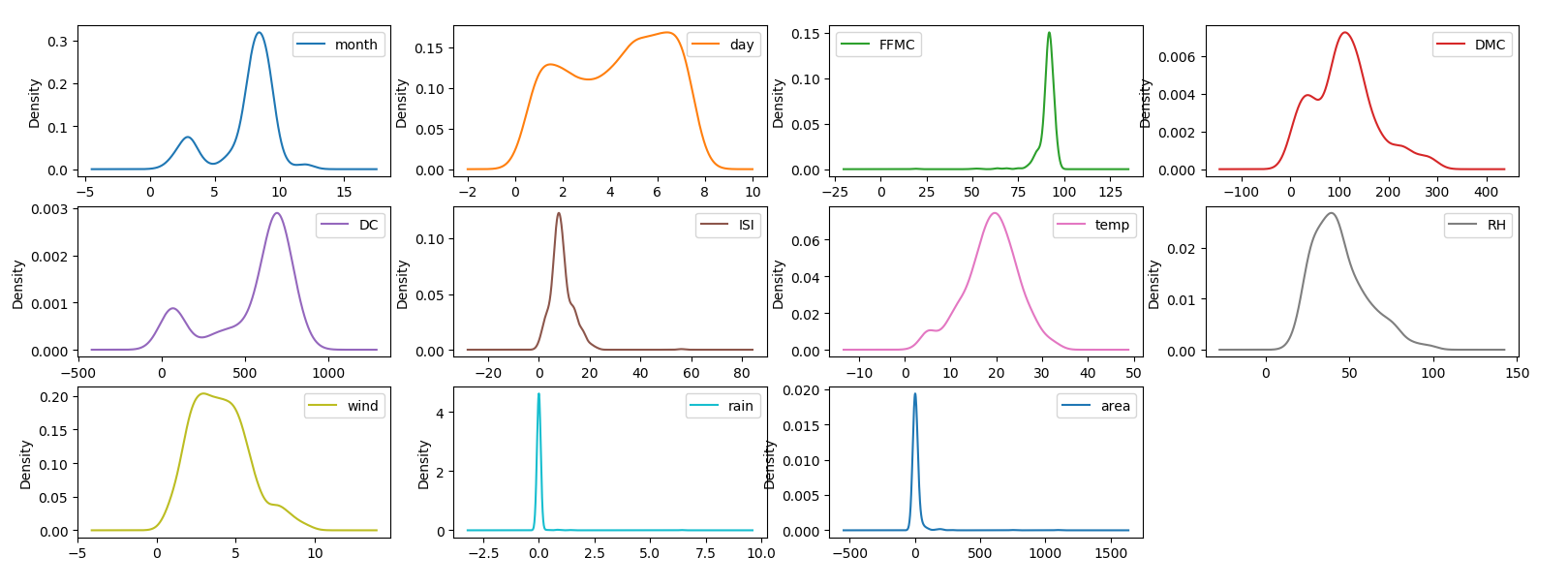


The following are the histograms of the dataset:

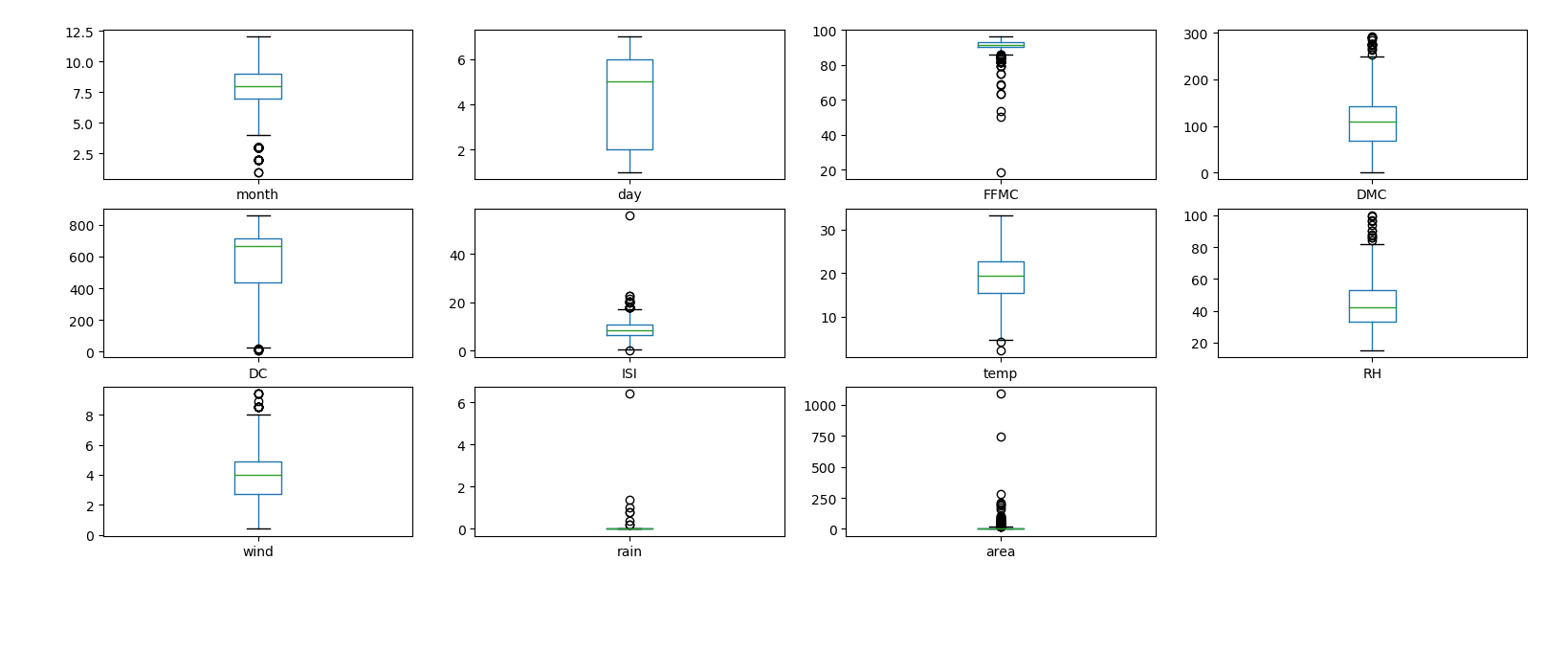


The variables ‘temp’ has near Gaussian distribution or near normal distribution. The rest of the histograms are either positively or negatively skewed.

The following is the continuous density function:

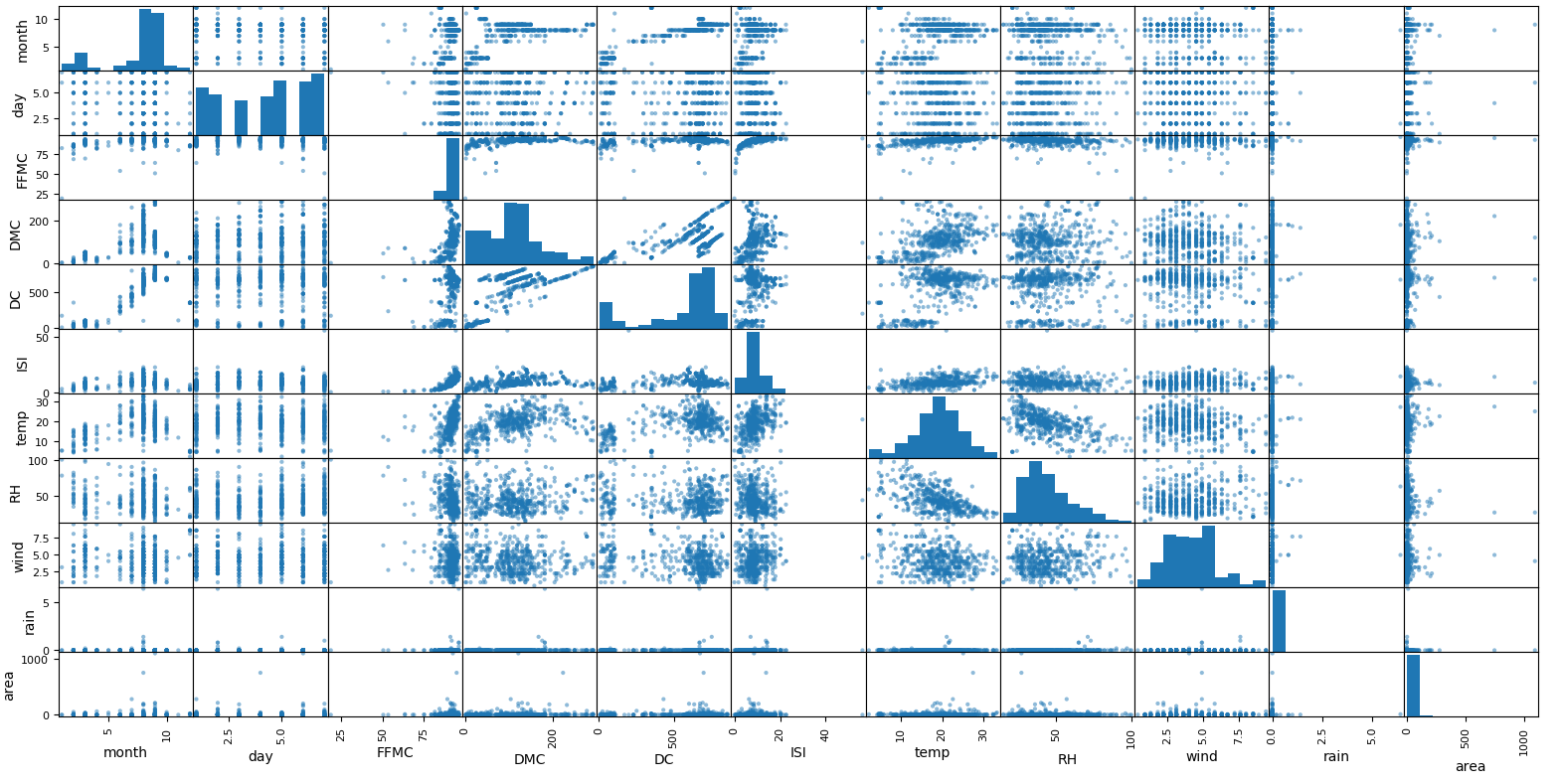


The following are the boxplots of the variables

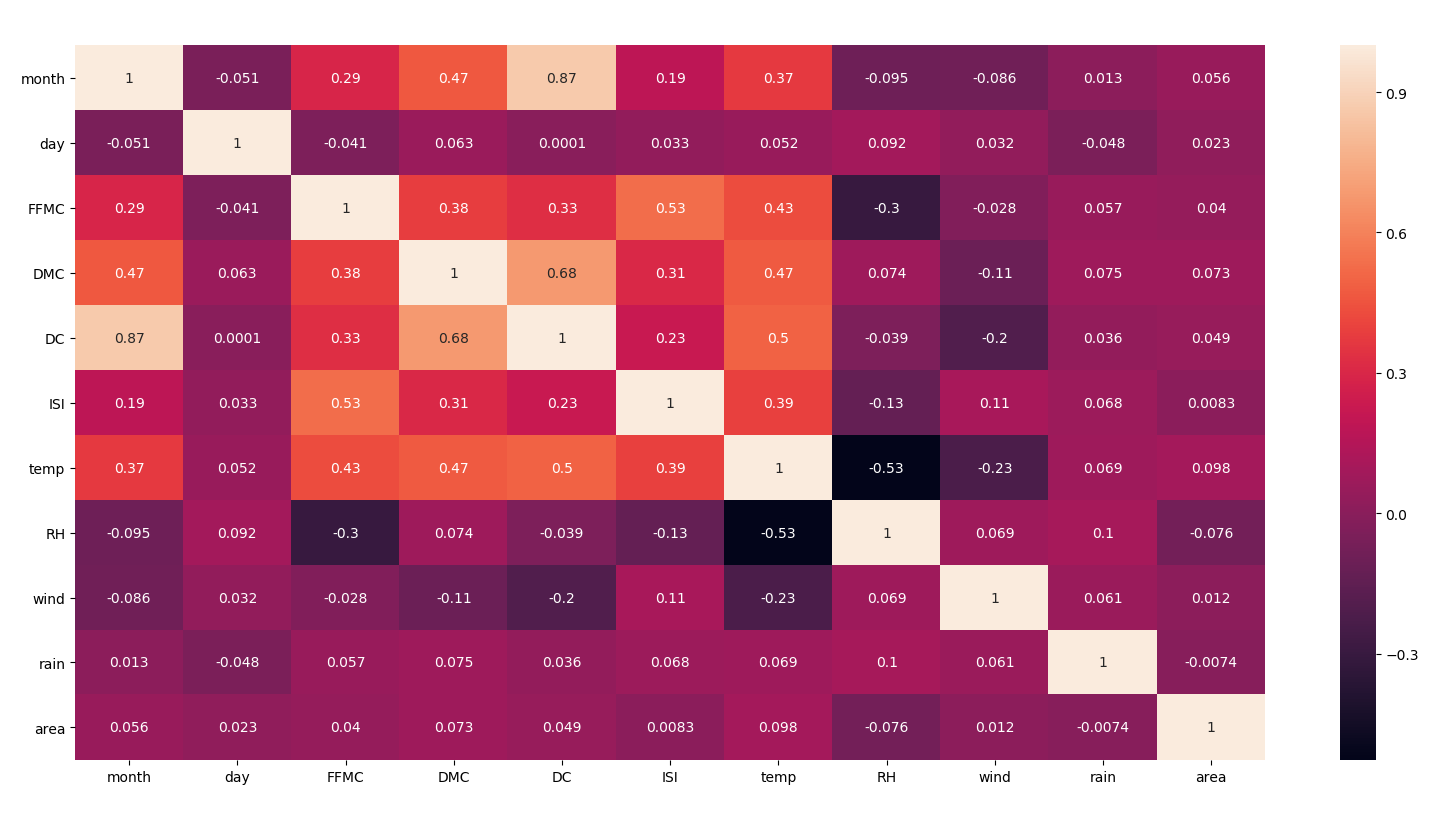


The variables FFMC, month, DMC, DC, ISI, RH, wind, rain and area have statistical outliers

The following is the pair plot or the scatter matrix of the variables:



The heat map of the correlation between different variables

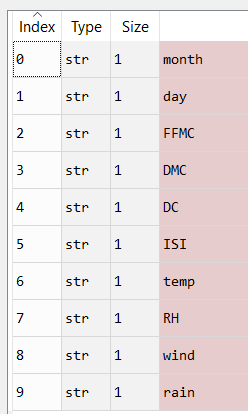


From the above heat map we can estimate the following:

* High negative correlation between month and DC
* The correlation between rest of the independent variables are not significant enough to consider Multi Co linearity problem in the given dataframe.

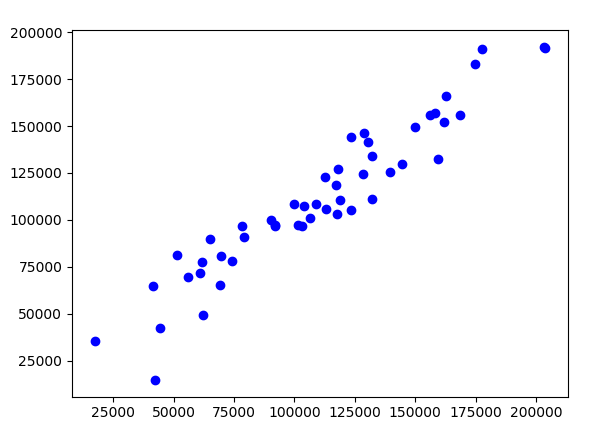
Building the sequential neural network model:

The following variables are set as the predictors:



And the target variable is set to “Area”.

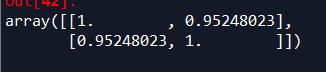
We get the prediction values of the Area variable and we evaluate the model by looking at the following scatter plot:



The above scatter plot is between the predicted values and the existing values of the variable Profits.

* There is an upward trend which means that it is positively co-related
* The points are tightly packed which means that there is less difference between the predicted and existing values
* The spread of the points is very narrow

The following is the co-relation value:



The co-relation value is 0.95, this means that the model is 95% accurate.