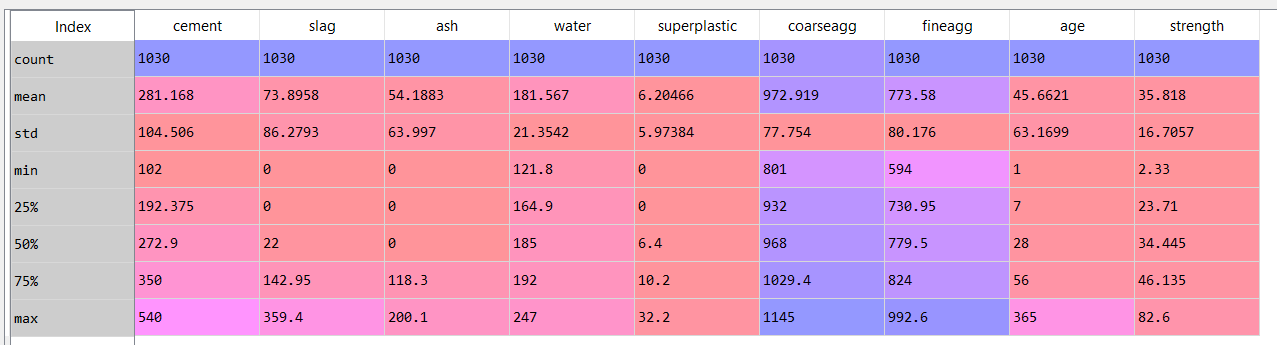
**Case Study of Concrete data**

**Problem Statement:**

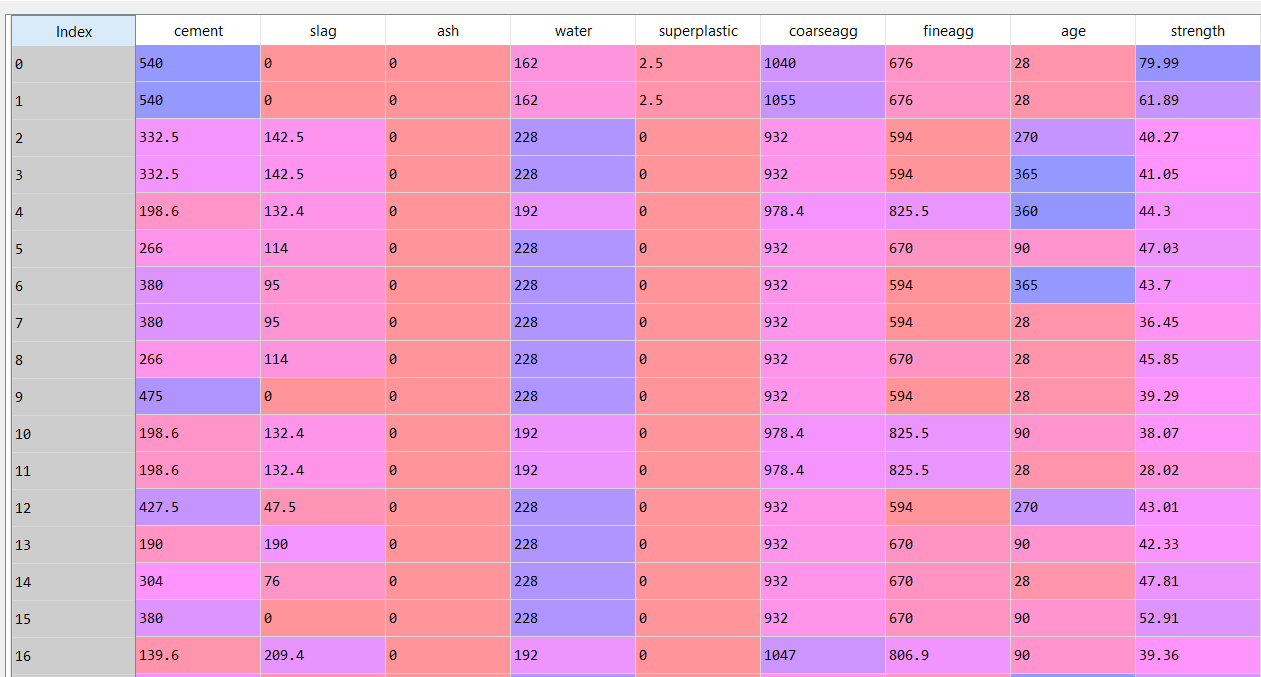
Prepare a prediction model for profit of concrete data. Do transformations for getting better predictions of profit

**Exploratory data analysis:**

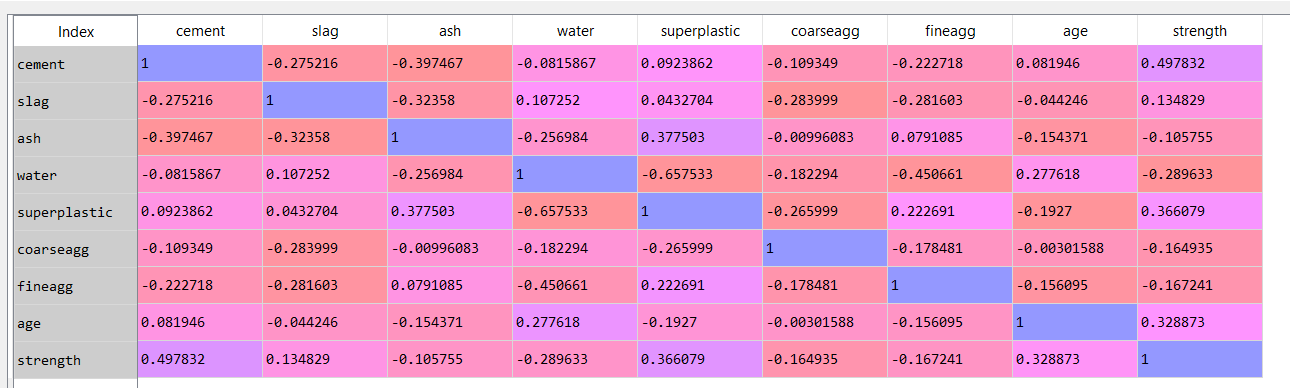


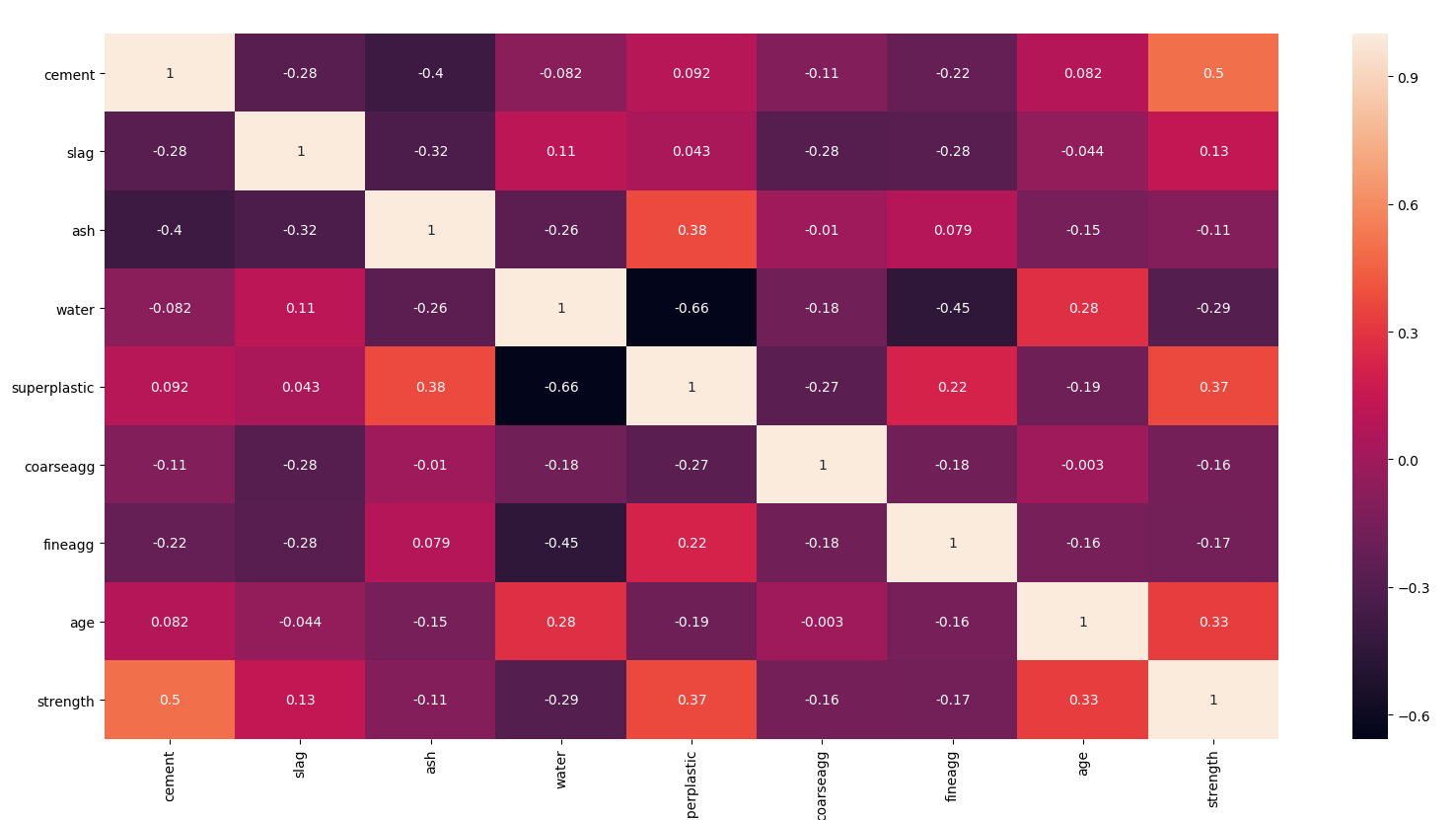
The above is the descriptive statistics of the concrete dataset. There are 1030 observations and 9 variables in the data frame.

The dataset looks like the following:



The heat map of the correlation between different variables and the correlation matrix are as follows:

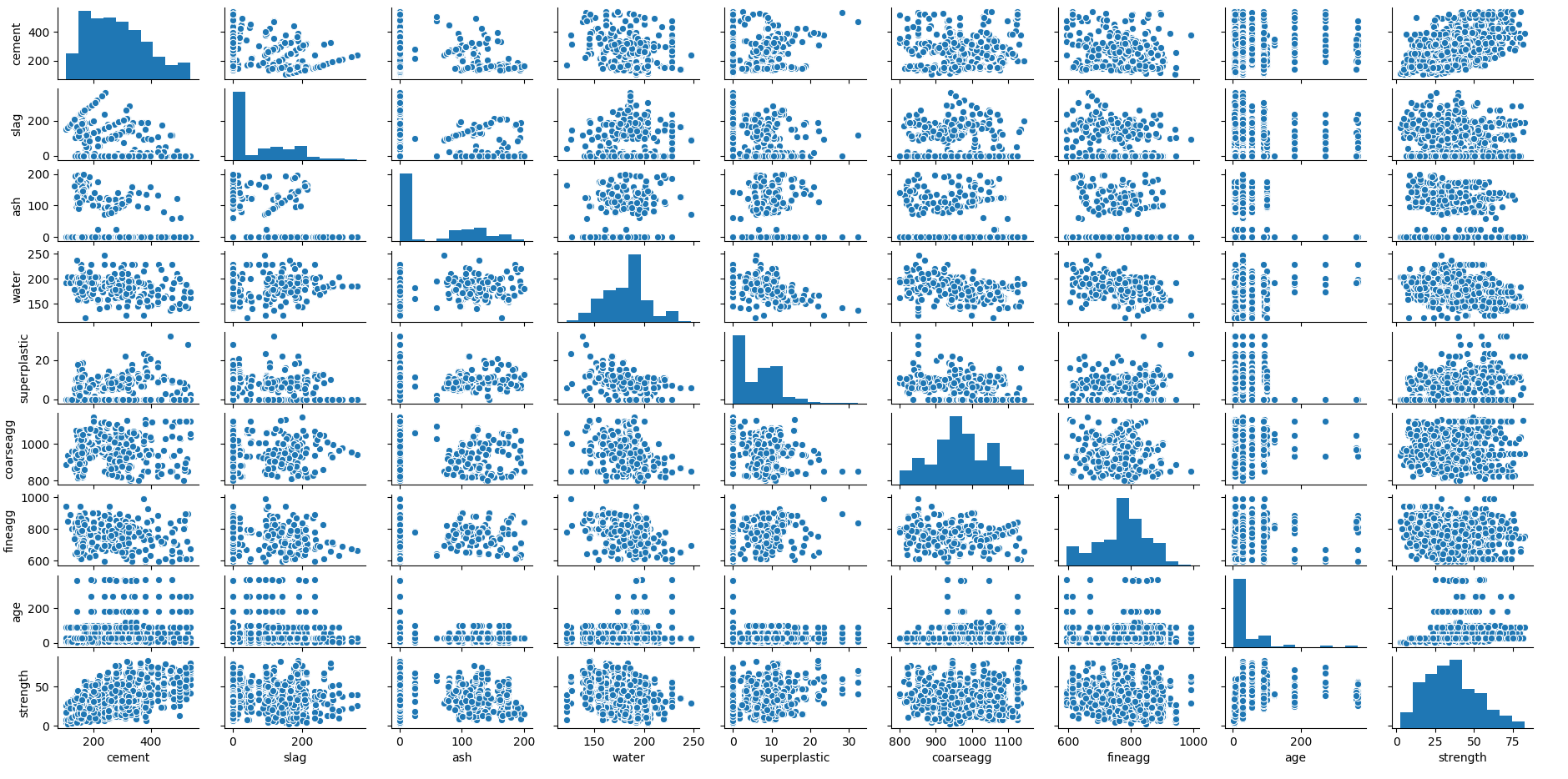




From the above heat map we can estimate the following:

* There is no high negative/positive correlation the variables
* There is slightly high correlation between the independent variables of water and superplastic that is about -0.66, which is a negative correlation but it is not high enough to cause any co-linearity problems.

The following is the pair plot of the dataframe:



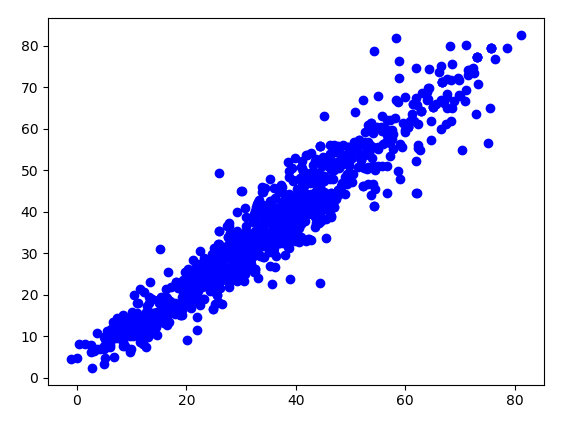
Building the sequential neural network model:

The following variables are set as the predictors:



And the target variable is set to “strength”.

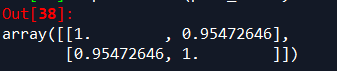
We get the prediction values of the strength variable and we evaluate the model by looking at the following scatter plot:



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.

The RMSE value for the above model is 5.055211577532011. This means that the error in the model is minimal.