Trility

● ● ●

Dose versus Quality

Modeling

| Data Clean	Exploration	ML
Null values	Statistics	Feature selection
Outliers	Visualisation	Model Selection
Rounding	Correlations	

# Visualisation - Averages, Correlations


```
#visualisation

fig, axes = plt.subplots(ncols=4, figsize=(24,8))
sns.heatmap(Plant_Average_Dose.T, cmap='Reds', ax=axes[0])

sns.heatmap(Plant_input_average.T, cmap='Reds', ax=axes[1])
sns.heatmap(corrAll, mask=np.zeros_like(corr, dtype=np.bool), cmap=sns.diverging_palette(220, 10, as_cmap=True),
            square=True, ax=axes[2])
sns.heatmap(corr4, mask=np.zeros_like(corr, dtype=np.bool), cmap=sns.diverging_palette(220, 10, as_cmap=True),
            square=True, ax=axes[3])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa15c86ce48>
```
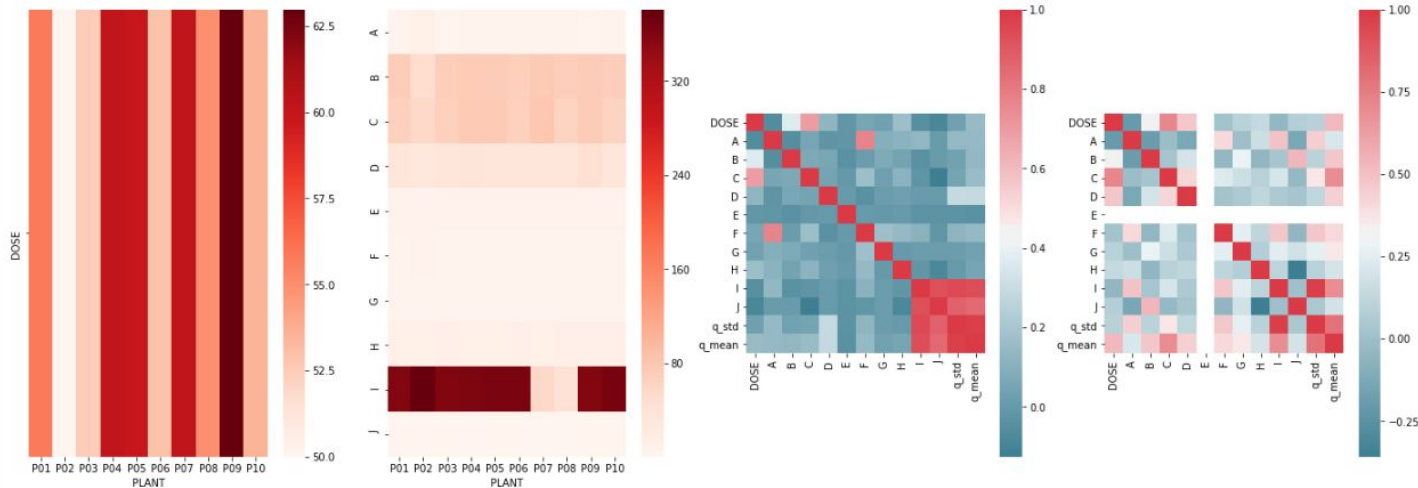
Visualisation - Pairplots of Time series of Dose and Quality
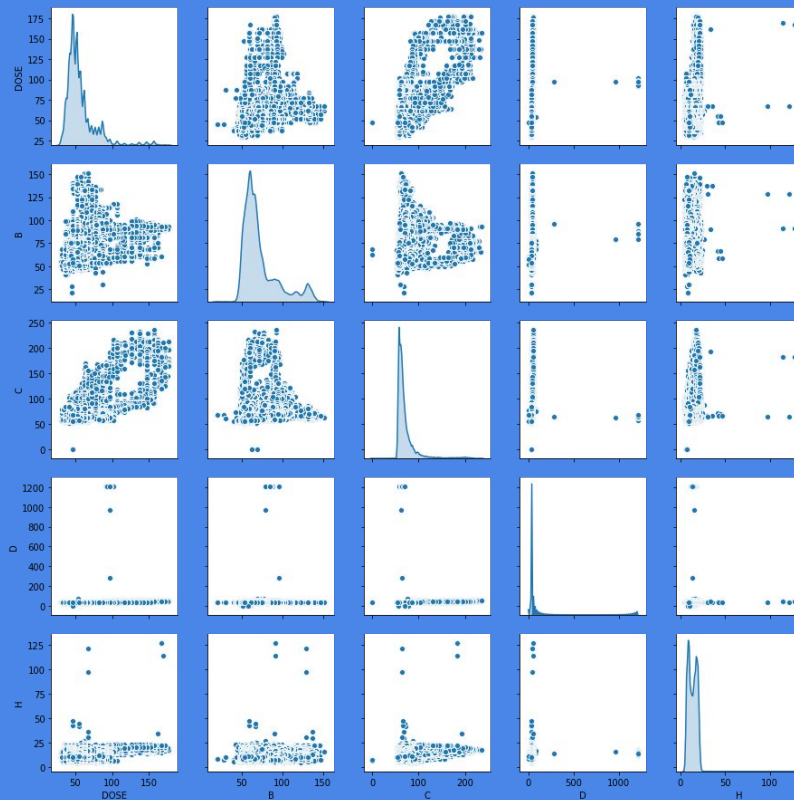
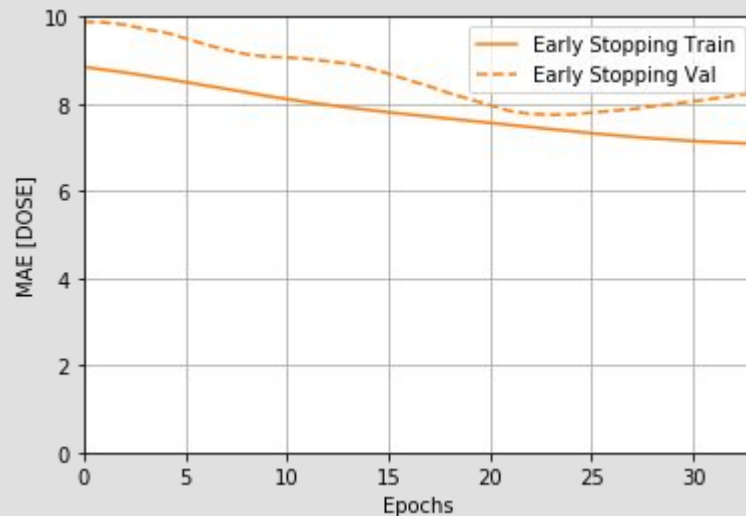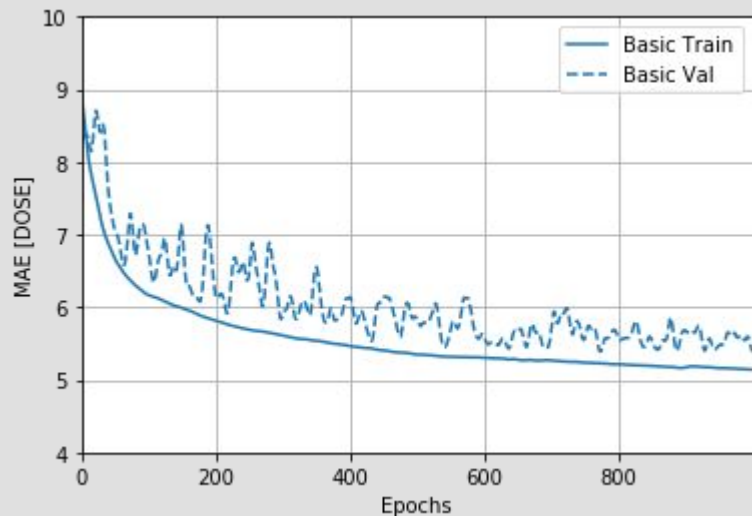# Methods of outlier detection and other data cleaning steps

Remove Null Values

Remove Floats (round the numbers)

Standard data types

___

Model selection and validation process:

Features.

Features that didn't appear to correlate were cut out.

Reasons:

Likely to make model better

Also fast

Would normalize and create new features in future.

Model Selected

Keras Regression model with RMS optimizer.

Reasons:
- RMS oscillations in the vertical direction - which would help with the spread in this data set.
- Quick and dirty does the job.

Selection of accuracy metric(s) and prediction accuracy achieved

MSE measures the average of the squares of the errors.

MAE is the average of the absolute difference between the predicted values and observed value.

These are great for regression.

MAE was (ABS): 17.25 units.

MSE was 329.8207 units.

Room for improvement with another 2 hours :)

How to implement / monitor / analyse

Given continuous time series data from a plant, one would adjust the dosage over time to lower the cost and improve the performance of a water treatment plant.

Client Front end application to describe what changes are needed to end consumer ; possibly automatically adjust dosage as required.

Trillity - Data Science team constantly monitors the models for worse performance, adding new data sets and creating new features as required.