For this dataset, there are 3 insights I found after exploring some features and also target variables:

1. Imbalance data can be found between binary classes on target variable. After calculation, there are 8404 labelled “0” and 5443 labelled “1” target variable for total of 13847 observations.

2. Using *mutual_info_classif()* method on scikit-learn to evaluating the Information gain of each variable in the context of the target variable or in other words, measures the dependency between the features and the target variable.

   CHP2Temp1(Deg C) 0.5509484000930598

   CHP2Temp2(Deg C) 0.550259620067963

   CHP2Vib1(mm/s) 0.4625149128689099

   CHP2Vib2(mm/s) 0.45978536822018623

   CHP1Vib2(mm/s) 0.19567828846226898

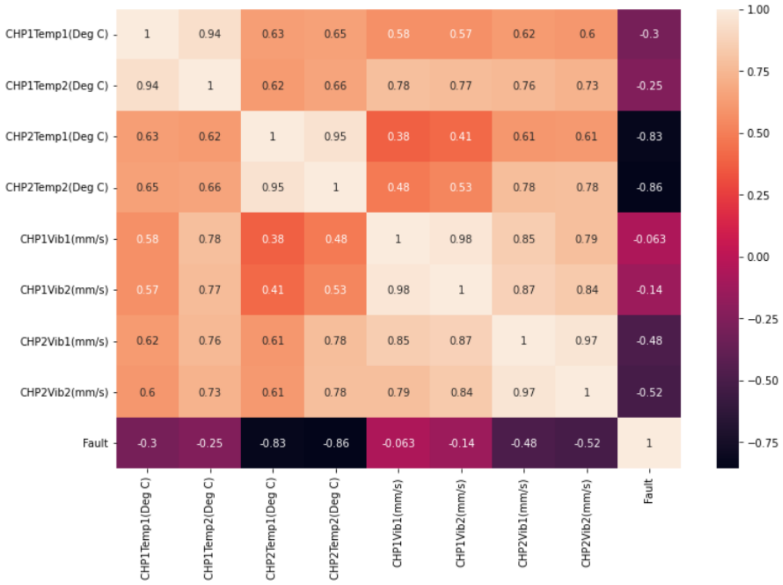
   CHP1Vib1(mm/s) 0.19285954457513266

   CHP1Temp1(Deg C) 0.1784938851598714

   CHP1Temp2(Deg C) 0.07303022517117852

   As we can see, using this method, target variable (“Fault” label) has higher dependency to Chiller pump CHP2 parameters than CHP1

3. Using Pearson Correlation,



	CHP1Temp1(Deg C)	CHP1Temp2(Deg C)	CHP2Temp1(Deg C)	CHP2Temp2(Deg C)	CHP1Vib1(mm/s)	CHP1Vib2(mm/s)	CHP2Vib1(mm/s)	CHP2Vib2(mm/s)	Fault
CHP1Temp1(Deg C)	1	0.94	0.63	0.65	0.58	0.57	0.62	0.6	-0.3
CHP1Temp2(Deg C)	0.94	1	0.62	0.66	0.78	0.77	0.76	0.73	-0.25
CHP2Temp1(Deg C)	0.63	0.62	1	0.95	0.38	0.41	0.61	0.61	-0.83
CHP2Temp2(Deg C)	0.65	0.66	0.95	1	0.48	0.53	0.78	0.78	-0.86
CHP1Vib1(mm/s)	0.58	0.78	0.38	0.48	1	0.98	0.85	0.79	-0.063
CHP1Vib2(mm/s)	0.57	0.77	0.41	0.53	0.98	1	0.87	0.84	-0.14
CHP2Vib1(mm/s)	0.62	0.76	0.61	0.78	0.85	0.87	1	0.97	-0.48
CHP2Vib2(mm/s)	0.6	0.73	0.61	0.78	0.79	0.84	0.97	1	-0.52
Fault	-0.3	-0.25	-0.83	-0.86	-0.063	-0.14	-0.48	-0.52	1

We can see that both sensor with same parameter on the same Chiller Pump has very high correlation. Also, Temperature sensor on both Chiller pump has high correlation, same case found on Vibration sensor.