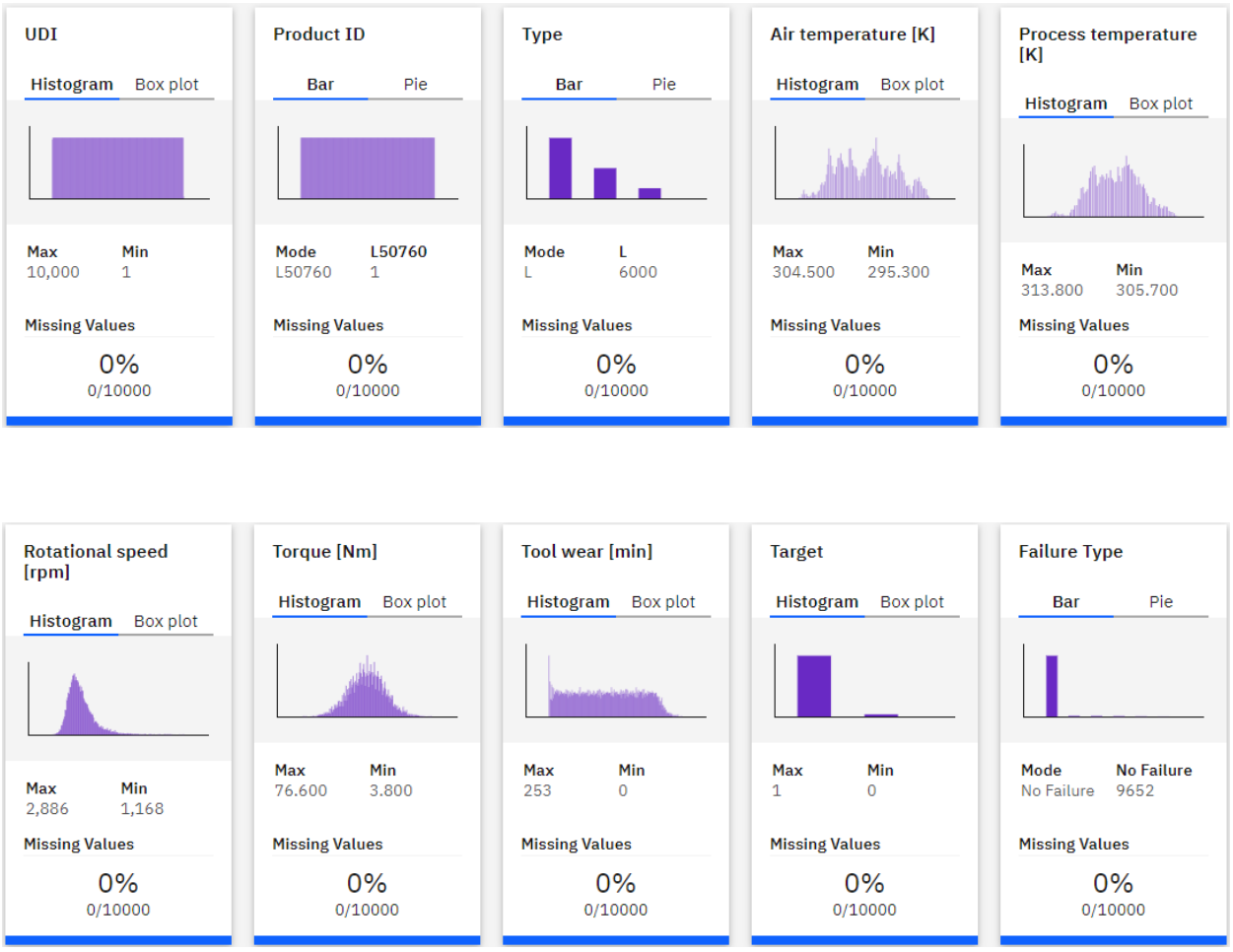
Data Quality Report

Machine Predictive Maintenance Classification

➢ SPSS Audit Report:



View Output: Data Audit

Compare

Audit | Quality | Search Columns

Name	Graph	Measurement	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
UDI		Continuous	Long	1	10,000	5000.500	2886.896	0	0	10000
Air temperature [K]		Continuous	Double	295.300	304.500	300.005	2.000	0.114	0	10000
Process temperature [K]		Continuous	Double	305.700	313.800	310.006	1.484	0.015	0	10000
Rotational speed [rpm]		Continuous	Long	1,168	2,886	1538.776	179.284	1.993	0	10000
Torque [Nm]		Continuous	Double	3.800	76.600	39.987	9.969	-0.010	0	10000
Tool wear [min]		Continuous	Long	0	253	107.951	63.654	0.027	0	10000
Target		Continuous	Long	0	1	0.034	0.181	5.150	0	10000

Audit | Quality | Search Columns

Name	Measurement	Type	Outliers	Extremes	% Complete	Valid	Null Value	Empty String	White Space	Blank Value
UDI	Continuous	Long	0	0	1	10000	0	0	0	0
Air temperature [K]	Continuous	Double	0	0	1	10000	0	0	0	0
Process temperature [K]	Continuous	Double	0	0	1	10000	0	0	0	0
Rotational speed [rpm]	Continuous	Long	418	106	1	10000	0	0	0	0
Torque [Nm]	Continuous	Double	69	0	1	10000	0	0	0	0
Tool wear [min]	Continuous	Long	0	0	1	10000	0	0	0	0
Target	Continuous	Long	339	339	1	10000	0	0	0	0

➢ Missing Attributes and Blank Fields

None of the attributes have missing or blanks values.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 10 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   UDI                     10000 non-null  int64
 1   Product ID              10000 non-null  object
 2   Type                    10000 non-null  object
 3   Air temperature [K]     10000 non-null  float64
 4   Process temperature [K] 10000 non-null  float64
 5   Rotational speed [rpm]  10000 non-null  int64
 6   Torque [Nm]             10000 non-null  float64
 7   Tool wear [min]         10000 non-null  int64
 8   Target                  10000 non-null  int64
 9   Failure Type            10000 non-null  object
dtypes: float64(3), int64(4), object(3)
memory usage: 781.4+ KB
```

```
df.isnull().sum()

UDI                        0
Product ID                 0
Type                       0
Air temperature [K]        0
Process temperature [K]    0
Rotational speed [rpm]     0
Torque [Nm]                0
Tool wear [min]            0
Target                     0
Failure Type               0
dtype: int64
```

➢ Spelling Inconsistencies

None of the attributes have spelling inconsistencies.

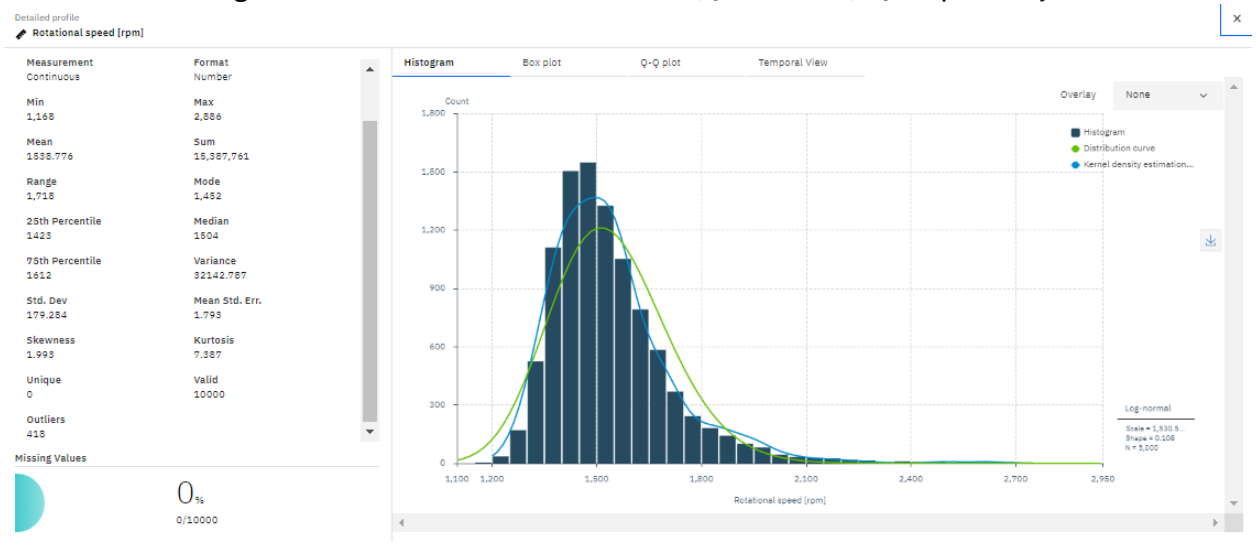
```
[23]   [i for i in df["Failure Type"].unique()]

['No Failure',
 'Power Failure',
 'Tool Wear Failure',
 'Overstrain Failure',
 'Random Failures',
 'Heat Dissipation Failure']
```

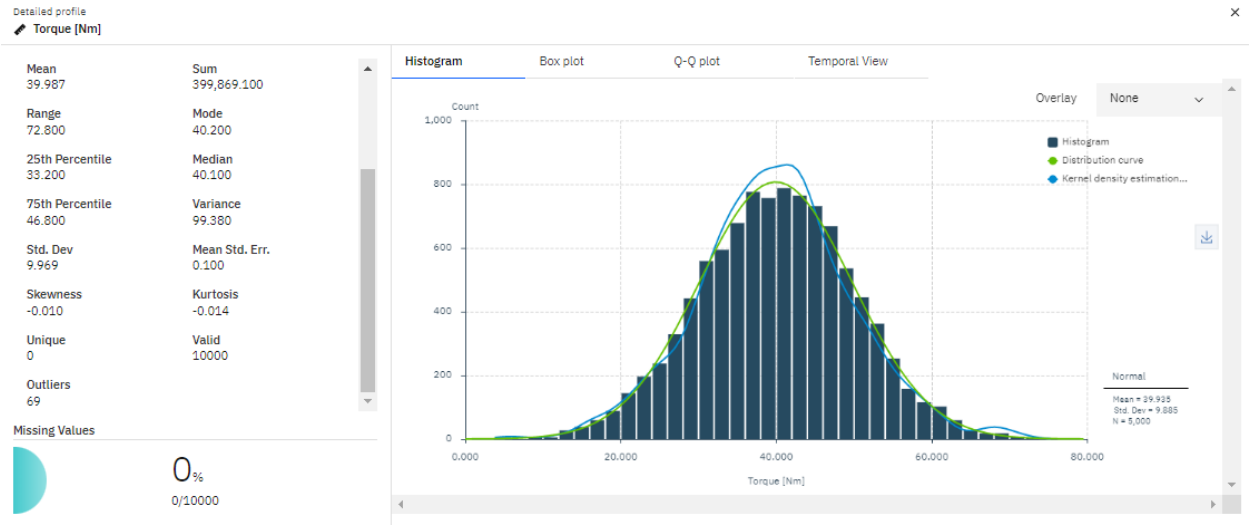
no spelling inconsistencies

➢ Noise Analysis

We can observe that there is a 415 outlier in ***Rotational Speed*** . But could be due to heat and differences in efficiency of model. And skews data at a ***Rotational Speed*** of about 1.993. We can also see in the image that the means and the median is 1538.77 and 1503 respectively.



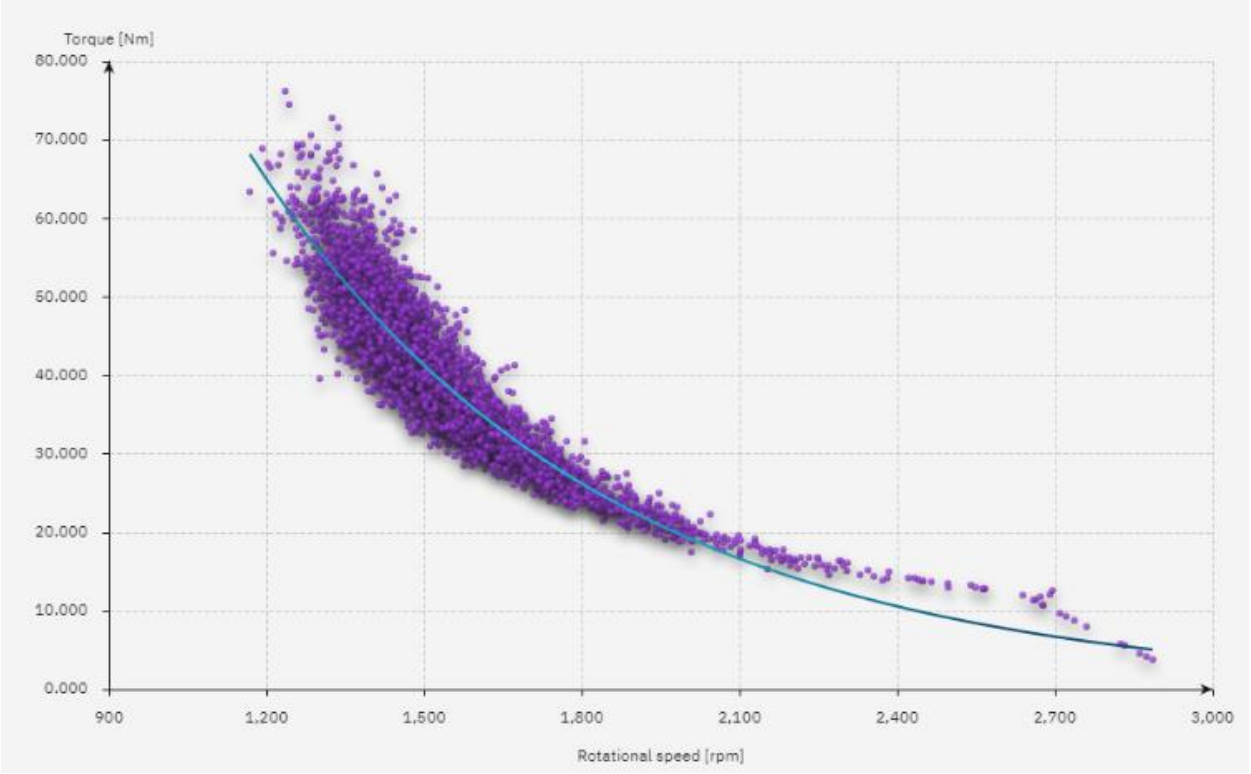
In Torque, we can also observe that there are outliers of about 69. It has a mean of 39.9, a median is 40.1 which is slightly left-skewed from the mean. But it will not affect the result.



➢ Plausibility check

There was no plausibility for any of the attributes.

There was chance for ***Torque*** vs ***Rotational Speed,*** but we can see that it was inversely proportional to each other.



Relational Scatter Graph

➢ Effect Of Excluding data/Attributes

We can exclude the Attributes from UID or Product ID as they are not impact on hypothesis.

➢ Data Format/Delimiter

Data stored in CSV format.

All the delimiters are consistent among the files.

Each Records contains the same number of fields.