U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Master in Industrial Engineering and Management

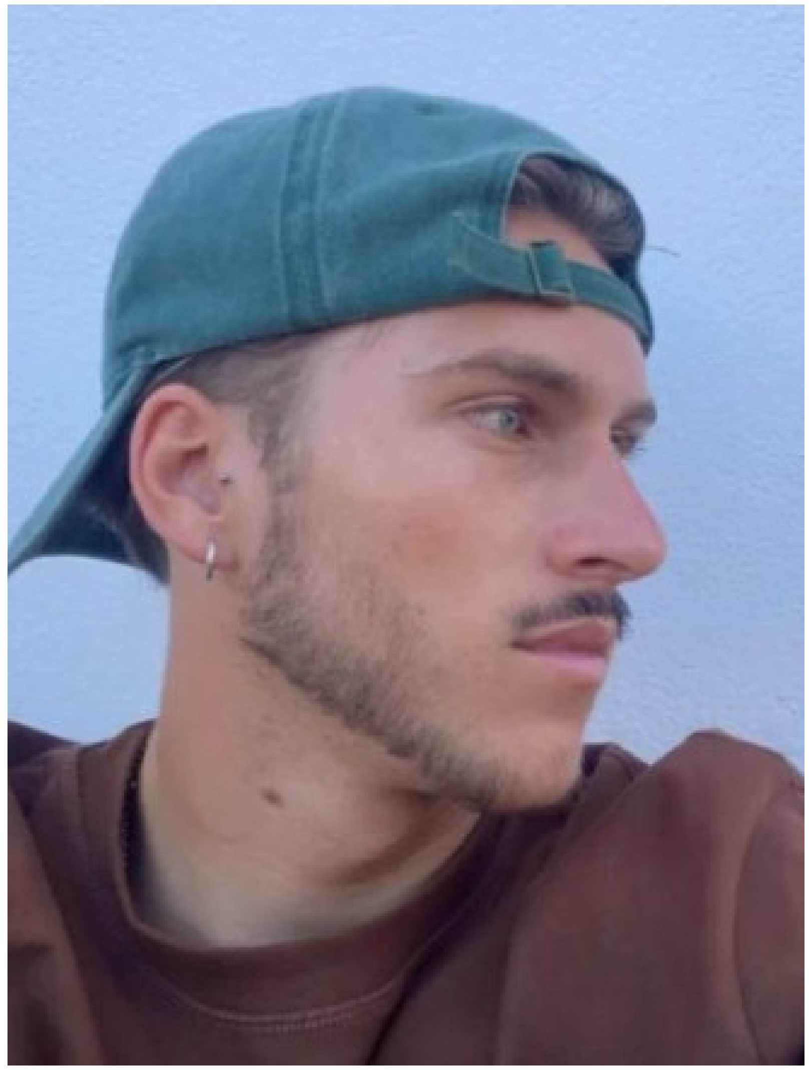
Business Analytics

Predictive Maintenance

Authors: António Santos, Dinis Bichança, Francisco Barros Leonor Silva, Miguel Lopes, Rodrigo Costa

2023/2024

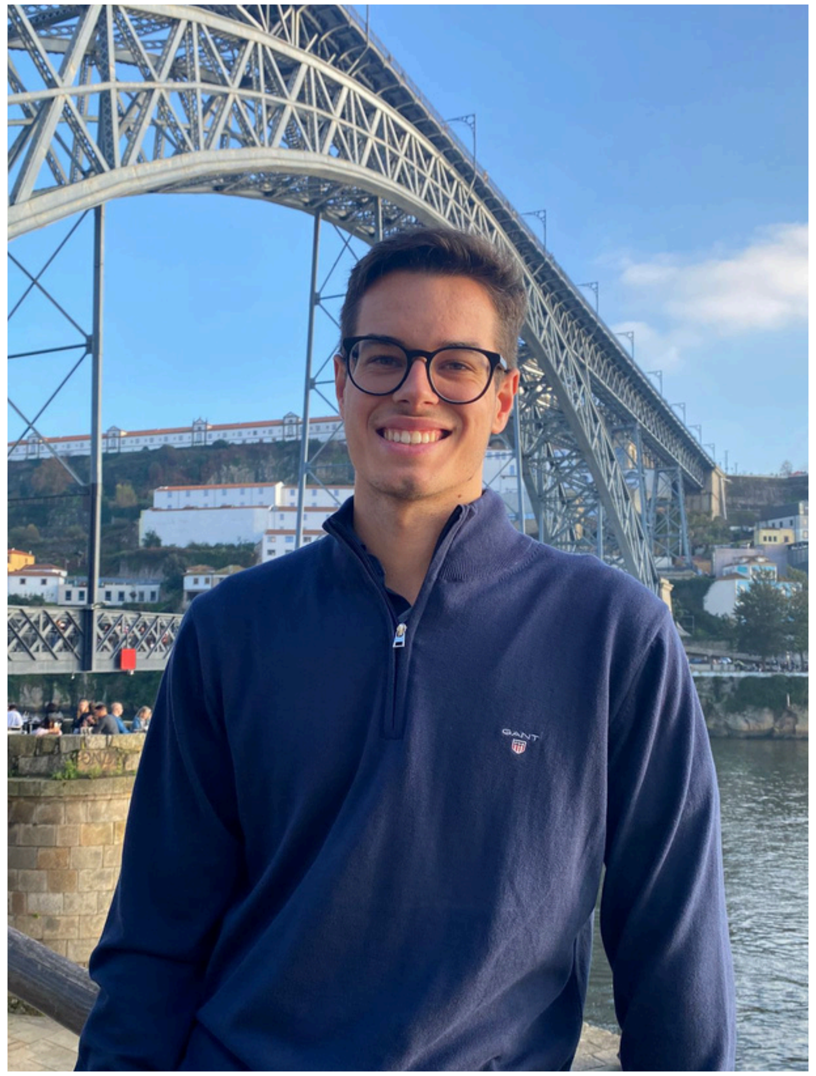
Meet the team



Miguel Lopes

Both students at Faculdade de Engenharia da Universidade do Porto

Both of us are from the Azores Islands, Miguel is from São Miguel, Rodrigo is from Terceira



Rodrigo Costa

Agenda

01. Methodology

02. Business Understanding

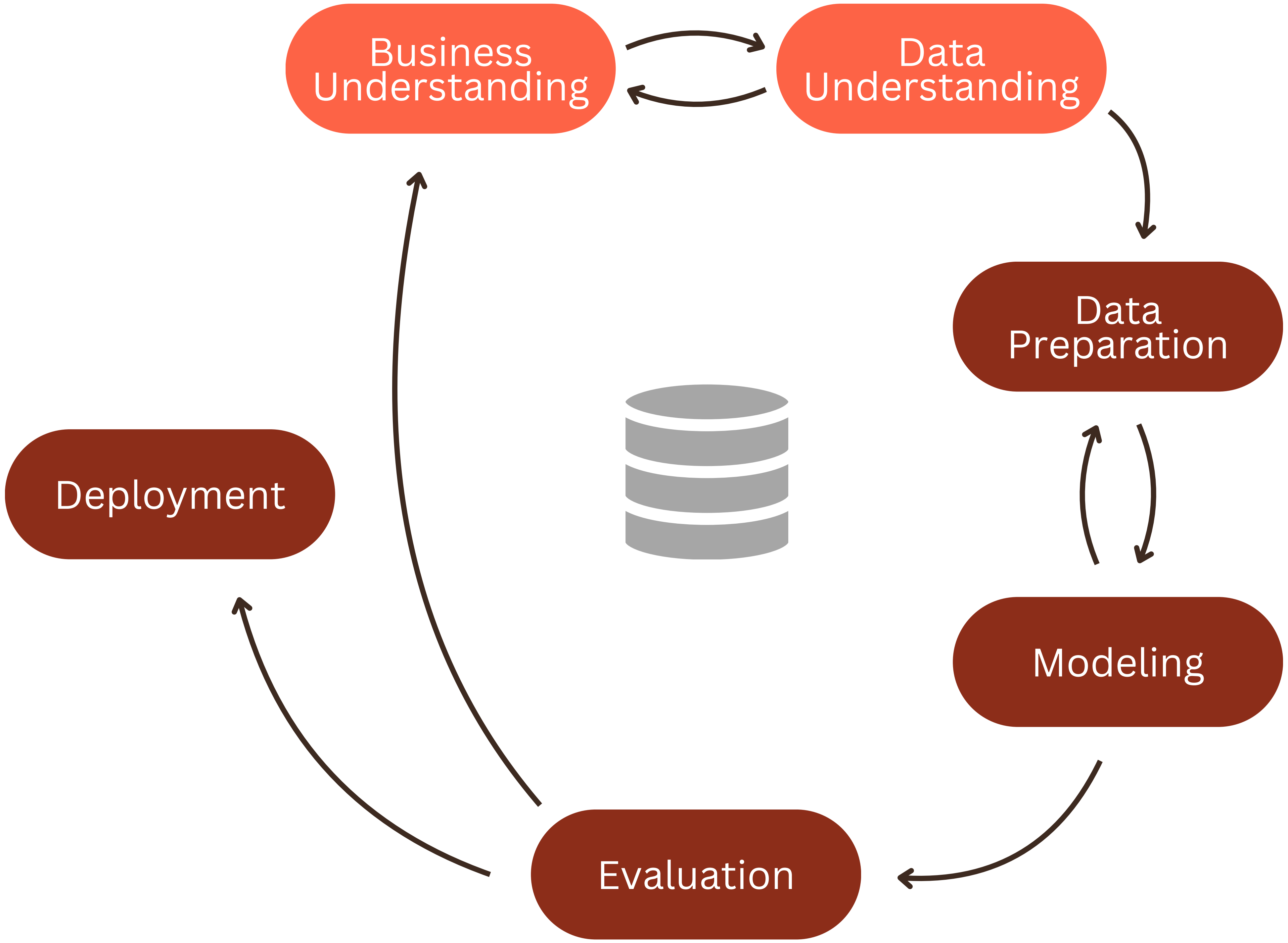
03. Data Understanding

04. Initial Data Analysis

01. Methodology

To support the problem solving, a **CRISP-DM** approach will be used.

For now, we will only focus on the Business and Data Understanding phases.



Agenda

01. Methodology

02. Business Understanding

03. Data Understanding

04. Initial Data Analysis

02. Business Understanding

Data-driven, proactive maintenance method designed to analyse the condition of equipment on an ongoing basis and **foresee potential failures**.

To monitor the condition of equipment and alert technicians to upcoming failures, it has **four main phases**:

- Sensing and gathering data using predictive maintenance technologies
- Transmitting the data in real time between machines and analytics systems.
- Applying **intelligent technologies** like AI and machine learning analytics
- Establishing the **maintenance and response protocols** required

02. Business Understanding

Key Technologies and Tools

- Internet of Things (IoT)
- Cloud Computing
- Data Analytics Software
- Machine Learning and AI

Industries

- Manufacturing
- Energy and Utilities
- Heathcare

...

Advantages and **disadvantages** of predictive maintenance:

↑ Operating costs reduction

↑ Downtime minimization

↑ Performance improvement

↑ Safety and lifespan increase

↓ High investement

↓ High level of knowledge and expertise required

Agenda

01. Methodology

02. Business Understanding

03. Data Understanding

04. Initial Data Analysis

03. Data Understanding

1000 observations | 10 attributes

Numerical

Attribute	Description	Type
ID	Unique identifier ranging from 1 to 10000	Discrete
ToolWear	The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool and a 'machine failure' label that indicates whether the machine has failed in this data point for any of the following failure modes are true	Discrete
RotSpeed	Calculated from power of 2860 W, overlaid with a normally distributed noise	Discrete
AirT	Generated using a random walk process later normalized to a standard deviation of 2 K around 300 K	Continuous
ProcessT	Generated using a random walk process normalized to a standard deviation of 1 K, added to the air temperature plus 10 K	Continuous
Torque	Torque values are normally distributed around 40 Nm with an Ïf = 10 Nm and no negative values	Continuous

03. Data Understanding

1000 observations | 10 attributes

Categorical

Product	Consisting of a letter L, M or H for low (50% of all products), medium (30%), and high (20%) as product quality variants and a variant-specific serial number	Nominal
FailType	Type of Failure Heat Dissipation/Overstrain Failure/Power Failure/Random Failure/Tool Wear Failure/No Failure	Ordinal
Target	Failure or Not	
Type	Consisting of a letter L, M or H for low (50% of all products), medium (30%), and high (20%) as product quality variants	

Agenda

01. Methodology

02. Business Understanding

03. Data Understanding

04. Initial Data Analysis

04. Initial Data Analysis

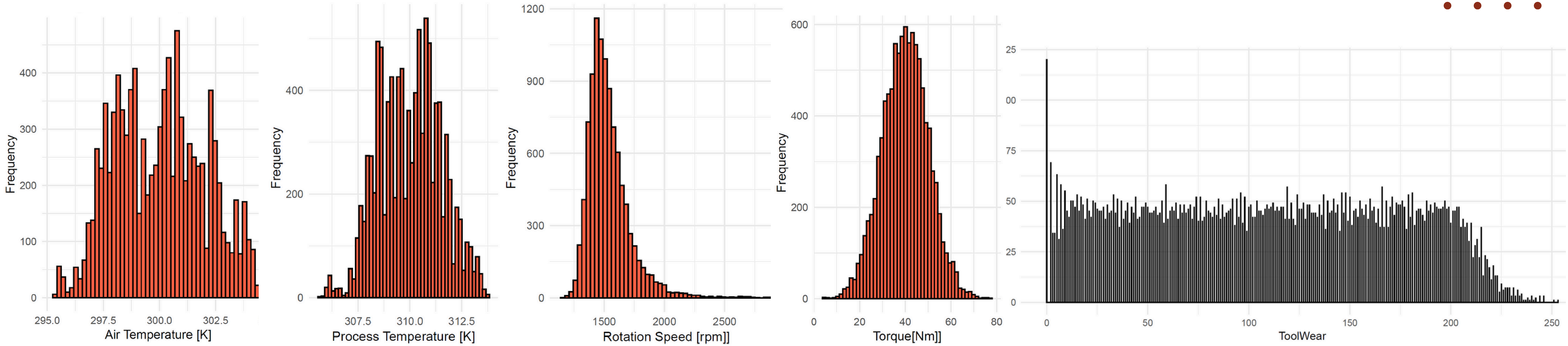


Figure 1 - Histograms of the variables Air Temperature, Process Temperature, Rotation Speed, Torque and ToolWear

Looking at the histograms (Figure 1) all the variables, with the exception of ToolWear, which follows a uniform distribution, seem to follow a normal distribution. However, using the Kolmogorov-Smirnov test (Figure 2), it can be concluded that only the variable Torque follows a normal distribution.

	variable	test	TS	Pval
1	AirT	Kolmogorov-Smirnov	0.066707	0
2	ProcessT	Kolmogorov-Smirnov	0.049147	0
3	RotSpeed	Kolmogorov-Smirnov	0.10407	0
4	Torque	Kolmogorov-Smirnov	0.0090335	0.38812
5	ToolWear	Kolmogorov-Smirnov	0.060058	0
6	Target	Kolmogorov-Smirnov	0.54039	0

Figure 2 - Kolmogorov-Smirnov tests

04. Initial Data Analysis

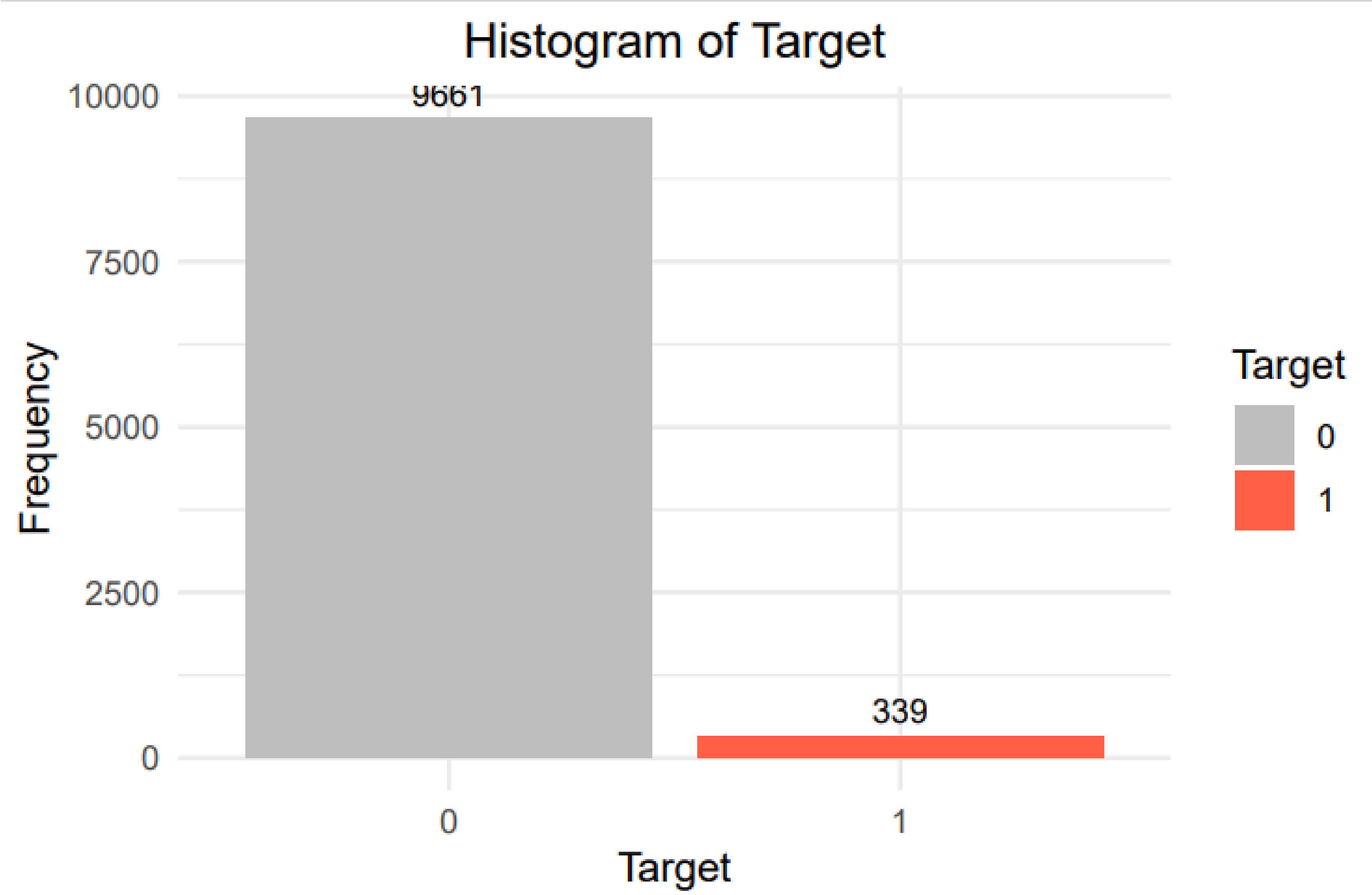
The variables **Process Temperature** and **Air Temperature** are highly (directly) correlated with a correlation coeficient of **0,88**.

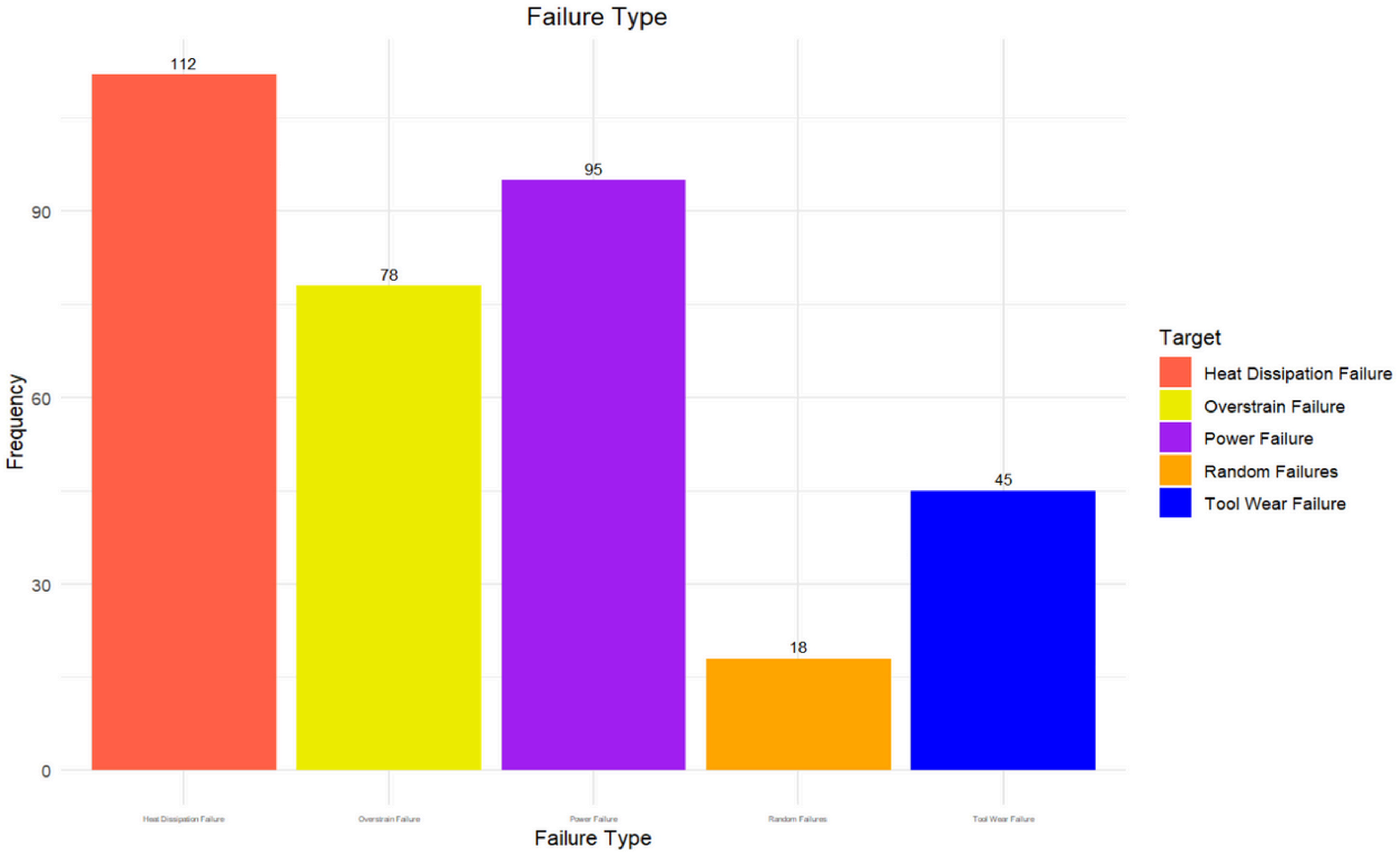
The variables **Torque** and **Rotation Speed** are highly (inversely) correlated with a correlation coeficient of **-0,88**.

There is no other pair of variables significantly correlated, once that the rest of correlation coeficients are between -0,04 and 0,19.

Figure 3 - Heatmap representing the correlation between variables

04. Initial Data Analysis





- Rpresents 3,39% of Failure in the given data, meaning this data set is imbalance.
- Heat Dissipation Failure and Power Failure are responsible for more than half of the failures

04. Initial Data Analysis

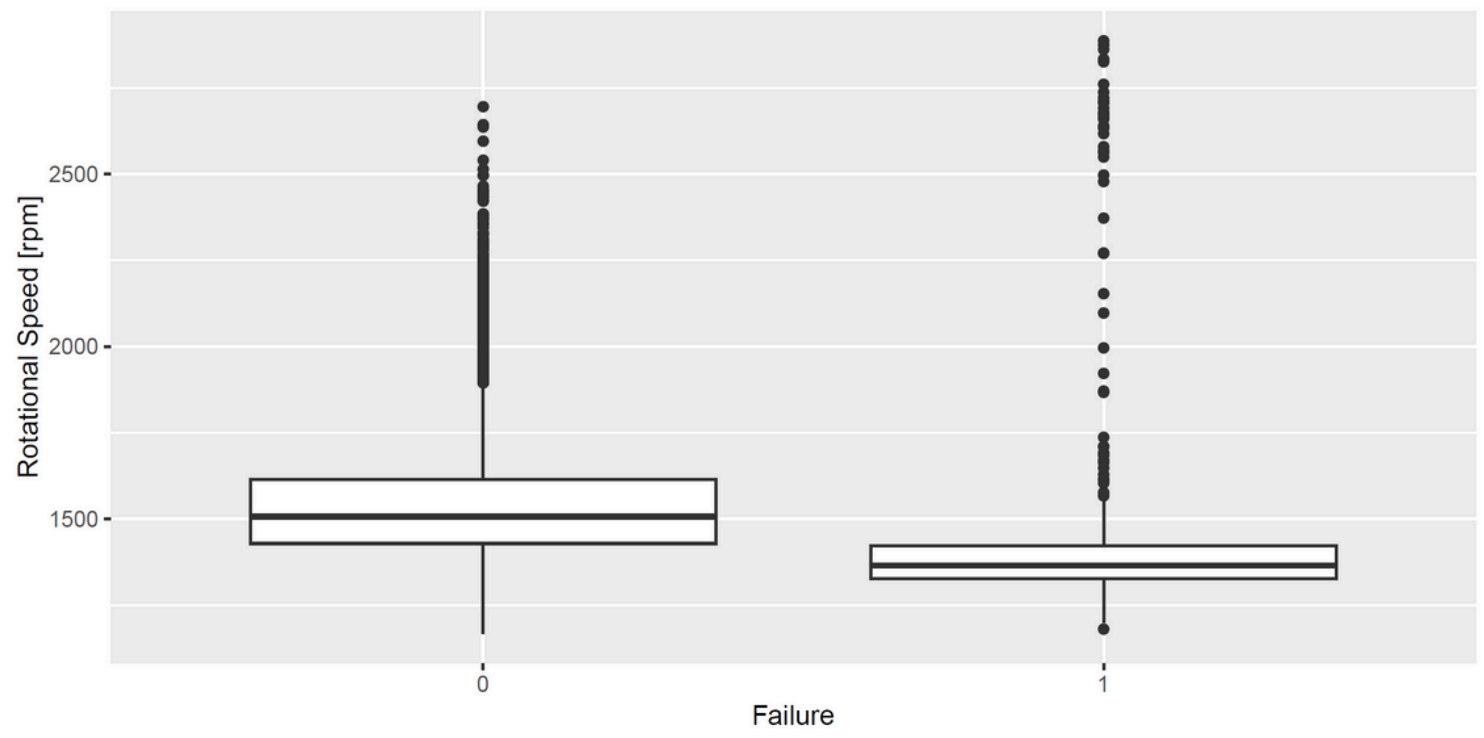


Figure 6 - Boxplot of the target variable Failure against Rotational Speed

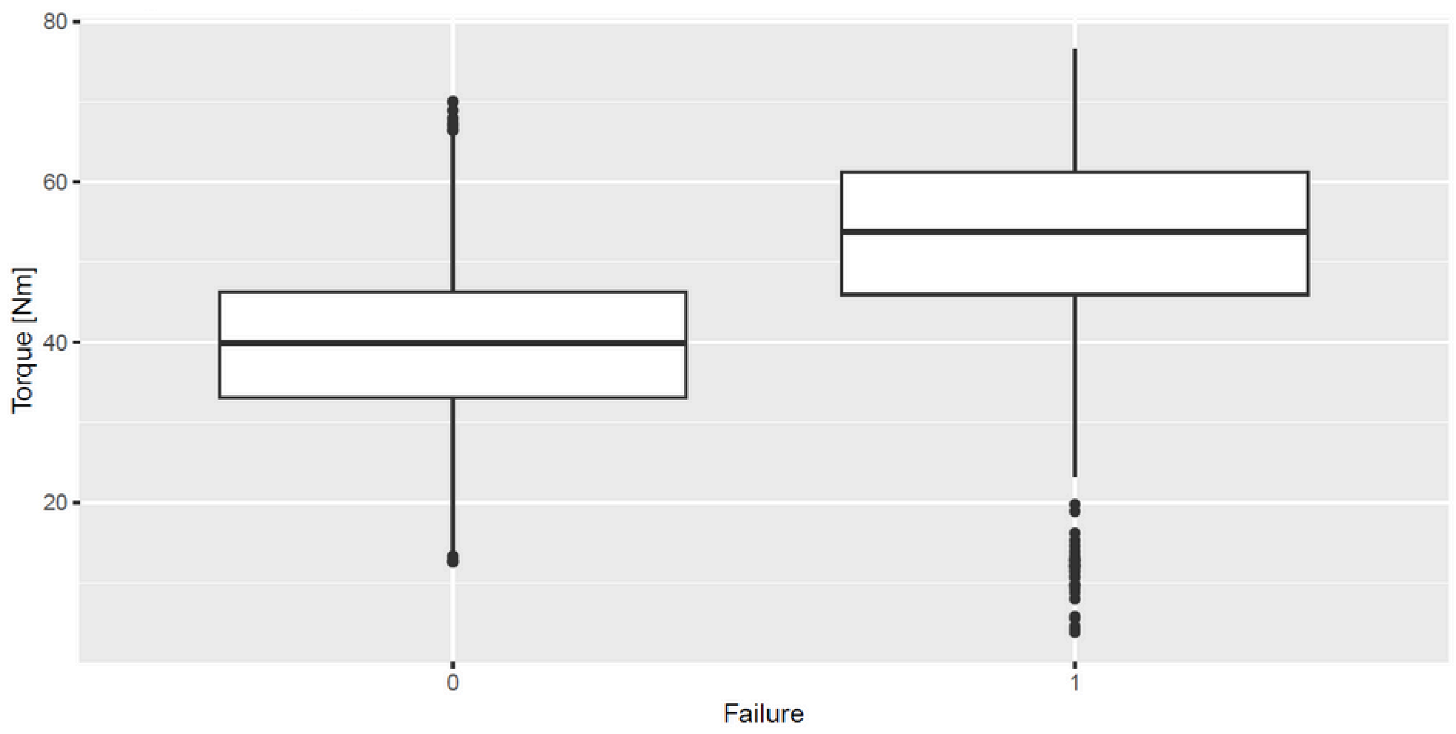


Figure 7 - Boxplot of the target variable Failure against Torque

On average, failure occurs more often for low values of Rotational Speed, when compared to non-failure.

On average, failure occurs more often for higher values of Torque, when compared to non-failure.

04. Initial Data Analysis

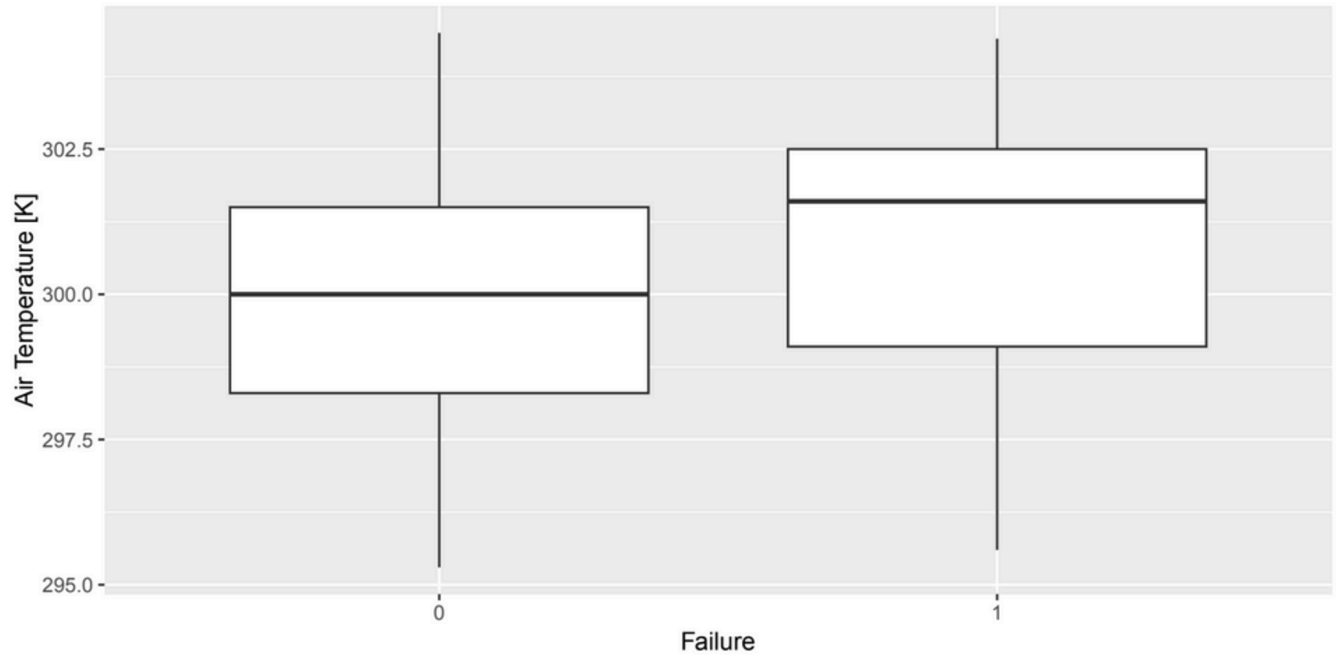


Figure 4 - Boxplot of the target variable Failure against Air Temperature

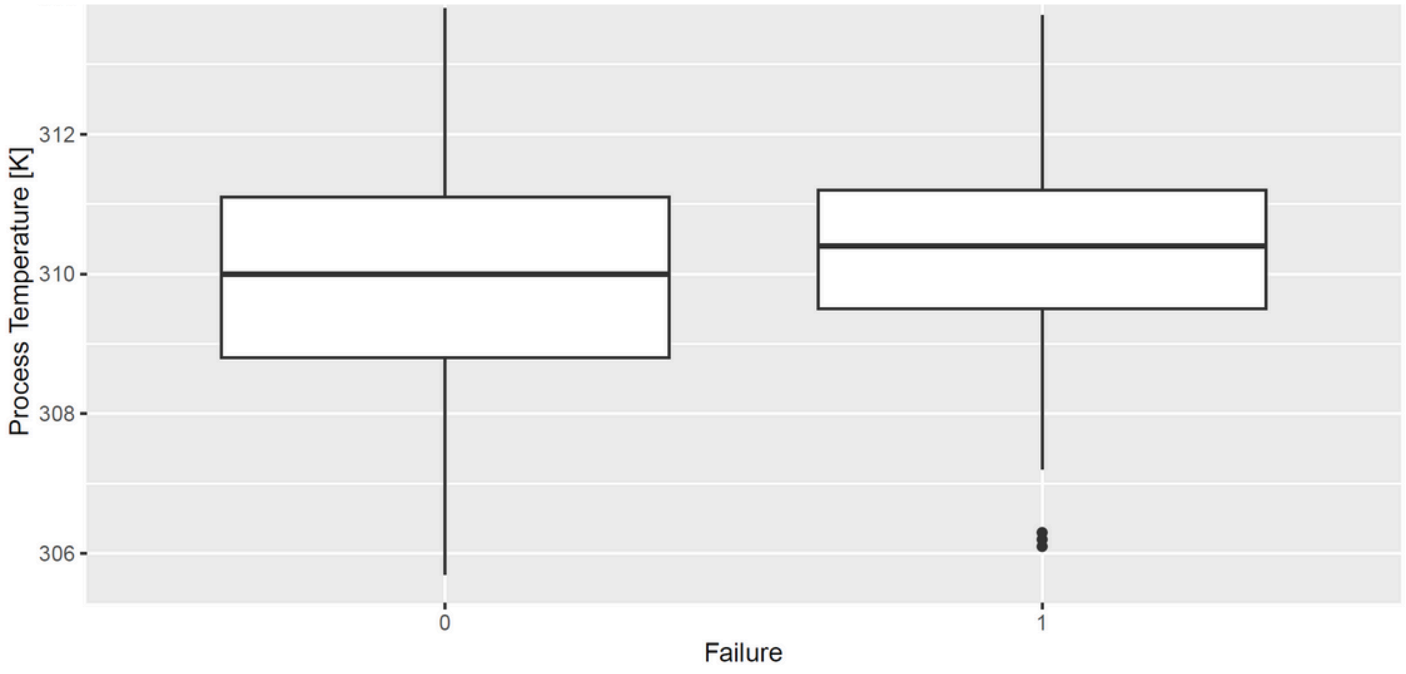


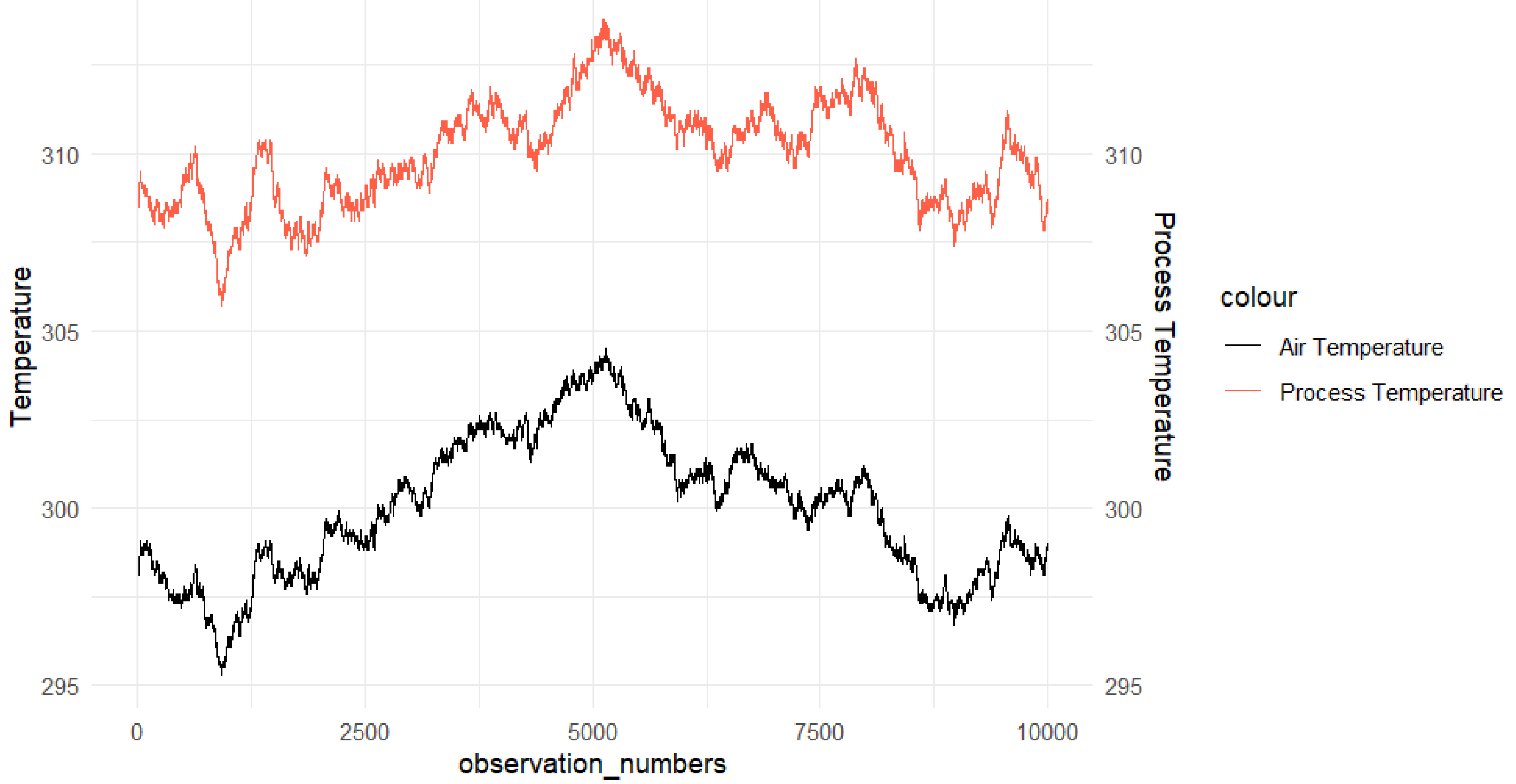
Figure 5 - Boxplot of the target variable Failure against Process Temperature

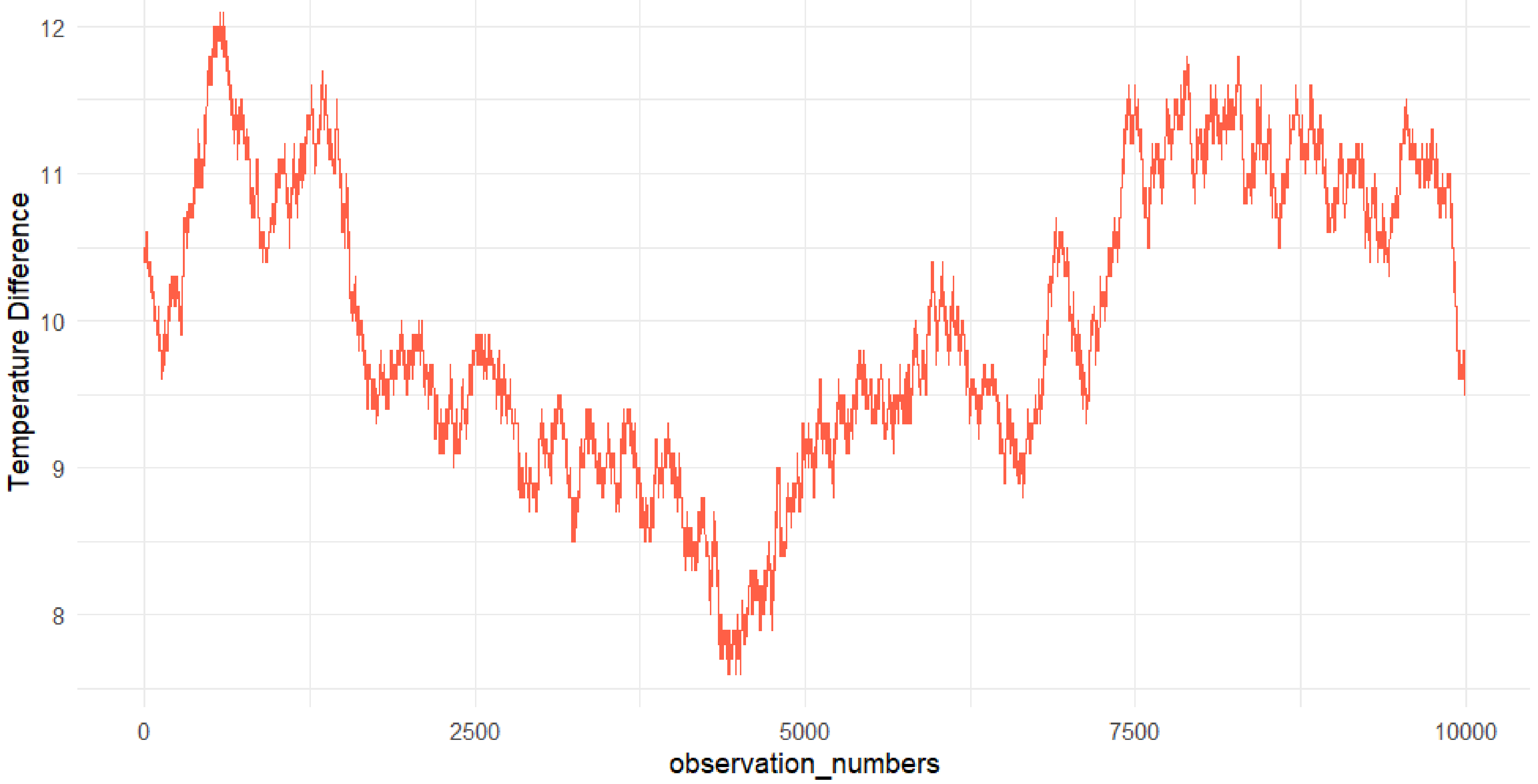
On average, failure occurs more often for higher values of Air Temperature, when compared to non-failure.

On average, failure occurs more often for higher values of Process Temperature, when compared to non-failure.

04. Initial Data Analysis

Temperature difference analysis



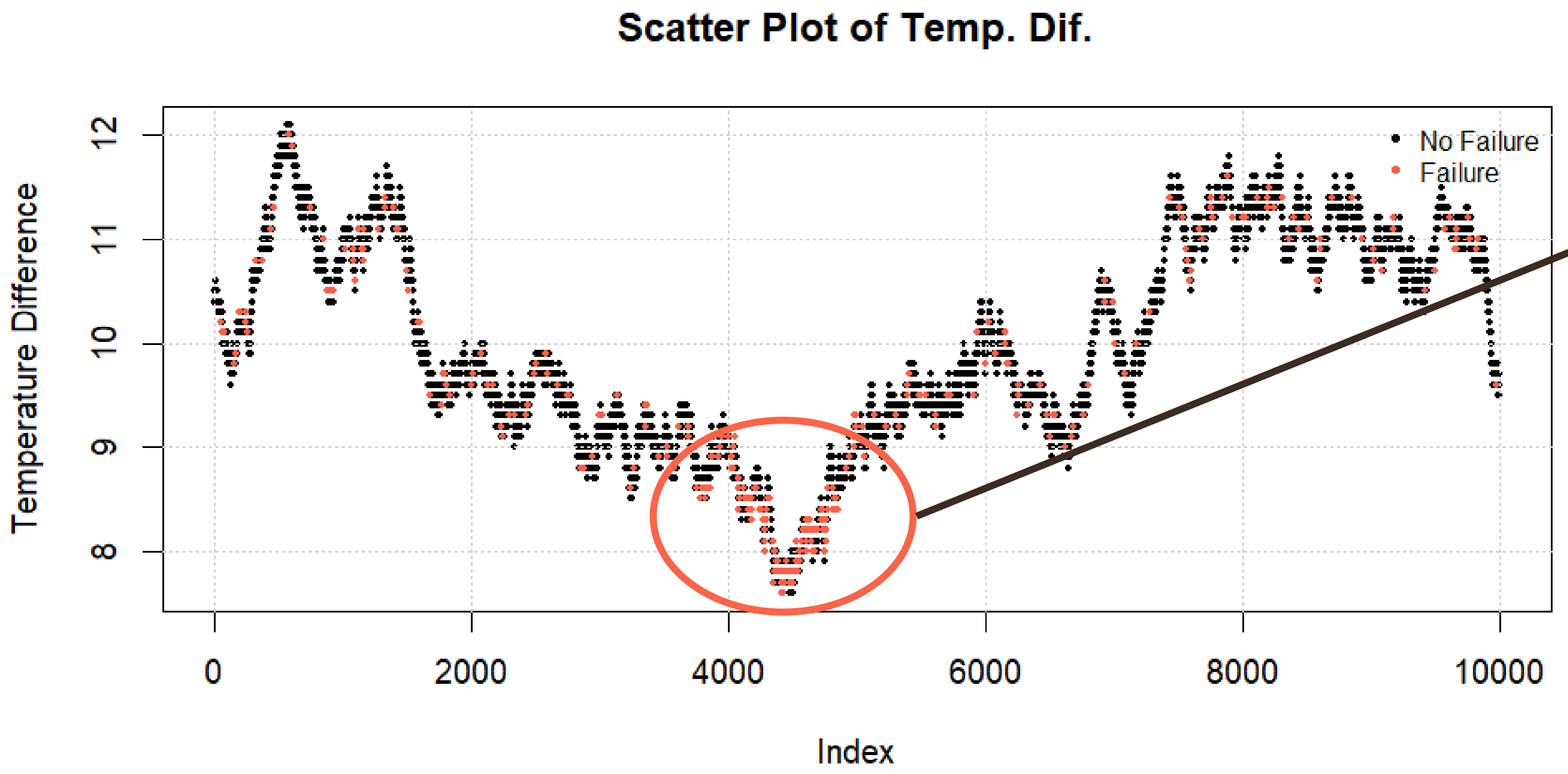


In what comes to the temperature related variables, it is relevant to analyze how they behave, and the impact of the difference between them

It now stands clear that for **higher values of air, and process temperature**, the **difference between them is smaller**, and the opposite also happens.

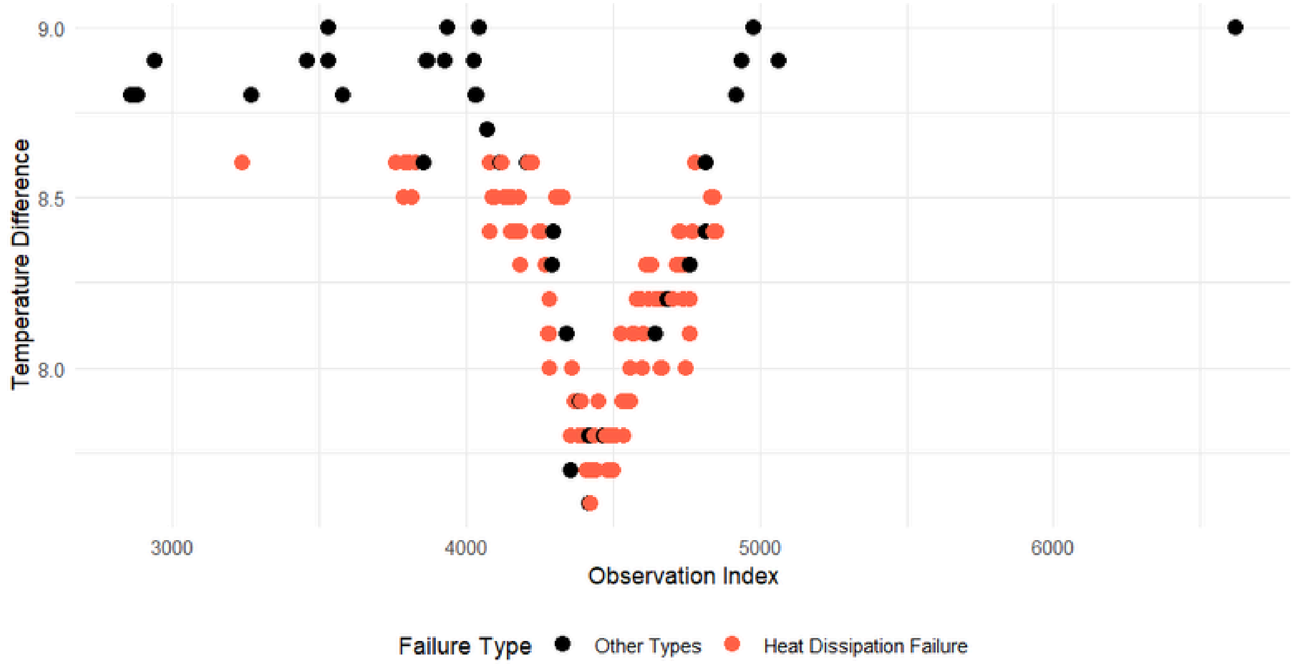
04. Initial Data Analysis

Temperature difference analysis



It now stands clear that for **higher values of air, and process temperature**, the **difference between them is smaller**, and the opposite also happens.

After analyzing the type of failure that occur in this region, it was concluded that there is a **high concentration of Heat Dissipation Failures** happening when the temperature difference reaches smaller values (**below 9 K**) - **73.68%**.



The 112 observations, plotted in red, account for 100% of the failures induced by Heat Dissipation Failure.

04. Initial Data Analysis

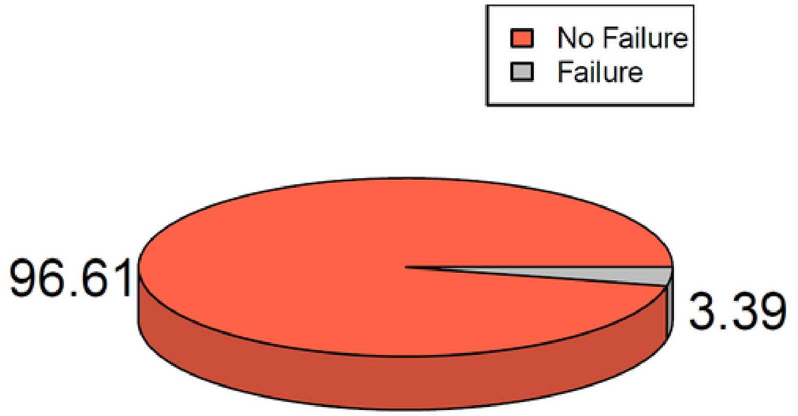


Figure 8 - Pie chart showing the percentage of failures and no failures

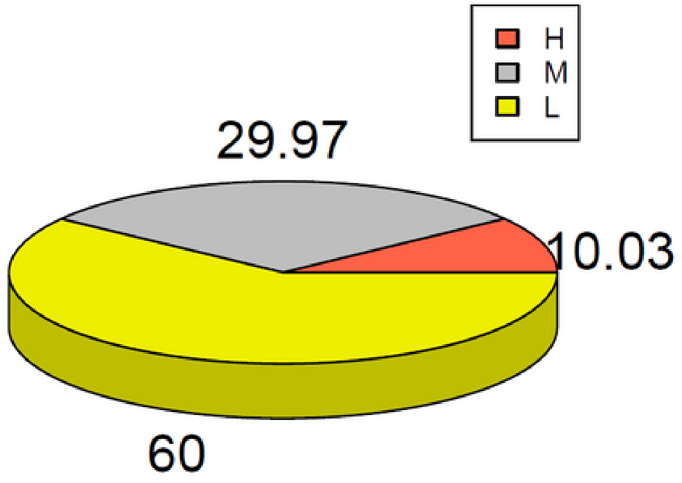


Figure 9 - Pie chart showing the percentage of product types

- The most abundant product in the dataset is type L (60%), followed by type M (29.97%) (Figure 9)
- In the whole dataset 3,39% of products have reached failure (Figure 8)
- The product with the greatest tendency to failure is type L, where around 3.92% of products fail (Figure 10). The product with the fewest failures is type H (Figure 12)

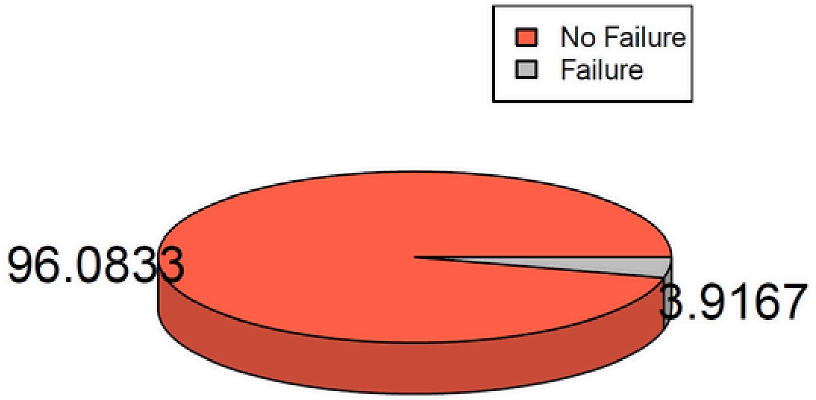


Figure 10 - Pie chart showing the percentage of failures and no failures for **type L**

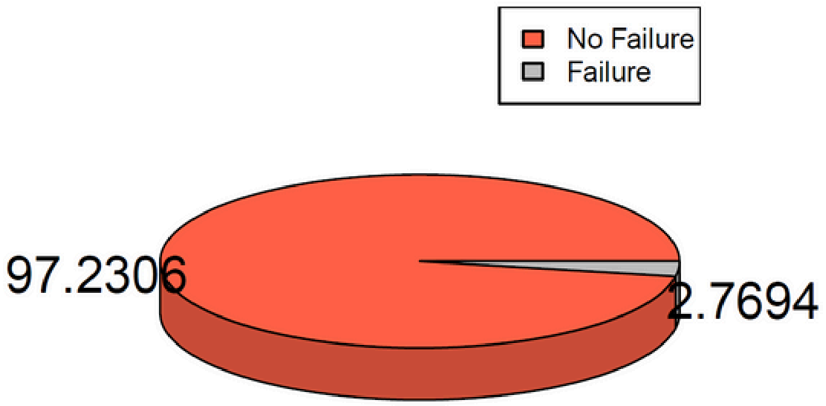


Figure 11 - Pie chart showing the percentage of failures and no failures for **type M**

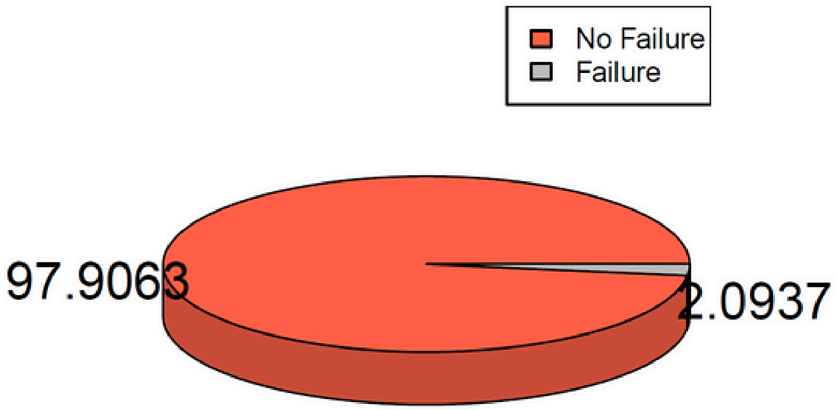


Figure 12 - Pie chart showing the percentage of failures and no failures for **type H**

04. Initial Data Analysis

Data inconsistencies

UDI	ID	Type	AirT	ProcessT	RotSpeed	Torque	ToolWear	Target	FailType
1222	M16081	M	297	308,3	1399	46,4	132	0	Random Failures
1303	L48482	L	298,6	309,8	1505	45,7	144	0	Random Failures
1438	H30851	H	298,8	309,9	1439	45,2	40	1	No Failure
1749	H31162	H	298,4	307,7	1626	31,1	166	0	Random Failures
2073	L49252	L	299,6	309,5	1570	35,5	189	0	Random Failures
2560	L49739	L	299,3	309	1447	50,4	140	0	Random Failures
2750	M17609	M	299,7	309,2	1685	28,9	179	1	No Failure
3066	M17925	M	300,1	309,2	1687	27,7	95	0	Random Failures
3453	H32866	H	301,6	310,5	1602	32,3	2	0	Random Failures
4045	M18904	M	301,9	310,9	1419	47,7	20	1	No Failure
4685	M19544	M	303,6	311,8	1421	44,8	101	1	No Failure
5472	L52651	L	302,7	312,3	1346	61,2	170	0	Random Failures
5490	L52669	L	302,6	312,1	1499	35	215	0	Random Failures
5496	H34909	H	302,9	312,5	1357	55	12	0	Random Failures
5510	L52689	L	302,8	312,2	1509	36,5	52	0	Random Failures
5537	M20396	M	302,3	311,8	1363	54	119	1	No Failure
5554	L52733	L	302,5	311,9	1306	59,7	172	0	Random Failures
5640	L52819	L	302,6	312,1	1668	28,7	180	0	Random Failures
5942	L53121	L	300,6	310,7	1438	48,5	78	1	No Failure
6092	L53271	L	300,9	310,7	1412	57,5	16	0	Random Failures
6479	L53658	L	300,5	309,8	1663	29,1	145	1	No Failure
6914	L54093	L	300,8	311,2	1481	38,5	181	0	Random Failures
6961	L54140	L	300,7	311	1413	52	91	0	Random Failures
7489	L54668	L	300,3	311,7	1545	43,5	160	0	Random Failures
7869	H37282	H	300,4	311,9	1438	46,7	41	0	Random Failures
8507	L55686	L	298,4	309,6	1710	27,3	163	1	No Failure
9016	L56195	L	297,2	308,1	1431	49,7	210	1	No Failure

Inconsistencies were found

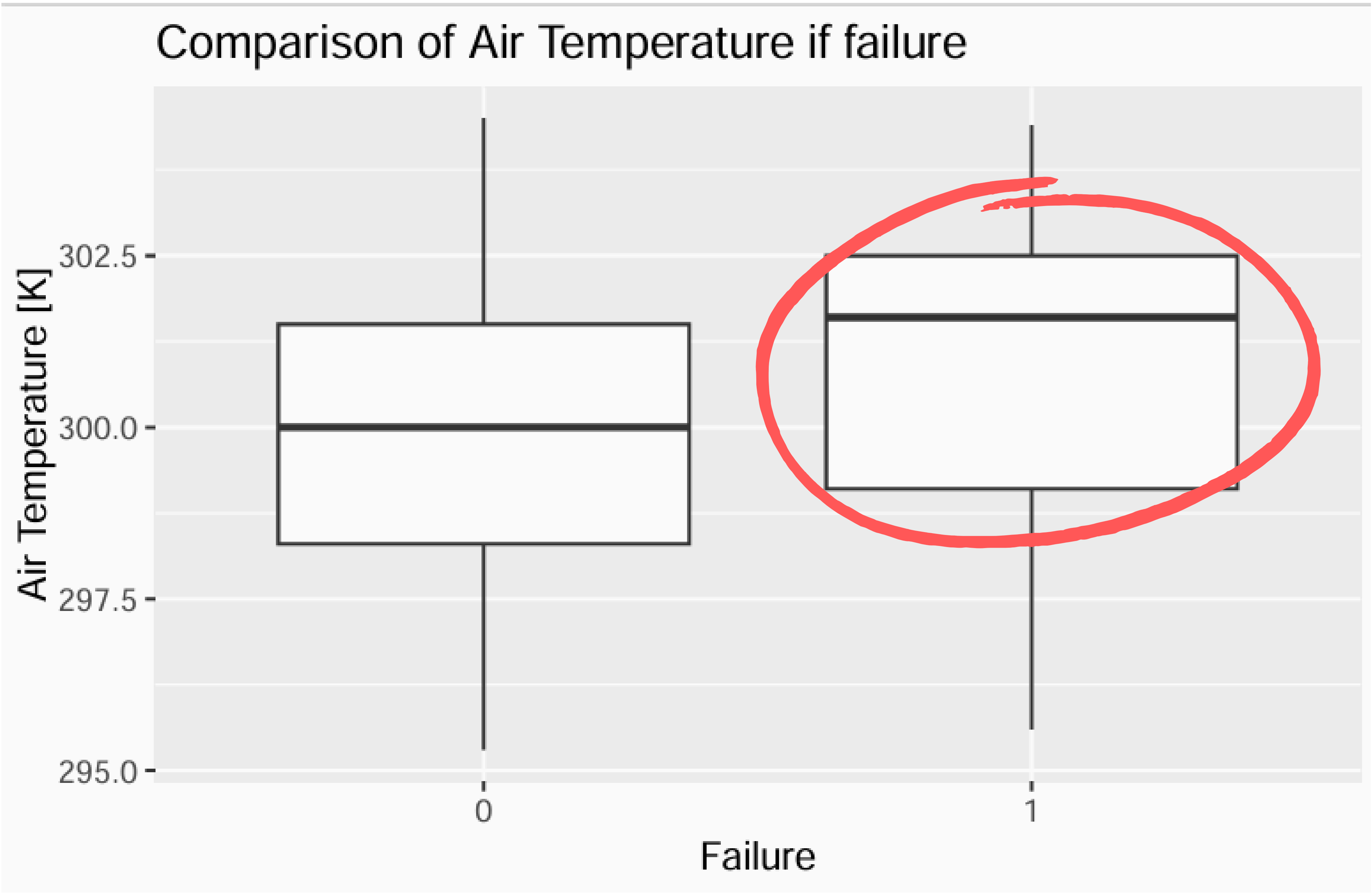
0.27% of the data was considered **inconsistent**	"**Random Failures**" failure type is assigned to a target value of **0 —** Eighteen observations	Nine observations with **target 1** and "**No Failure**" assigned on the failure type

IF

They all represent real failures	Only 0-"Random Failures" are true	Only 1-"No Failure" are true
7.96%	**2.65%**	**5.31%**

Value added in information related to failures occured

Thank You!


Comparison of Air Temperature if failure
Air Temperature [K]
302.5
300.0
297.5
295.0
0
1
Failure