Manish's Plan to Stop Machine Breakdowns

- **Here's the Deal**: It's 6:00 PM IST on Friday, June 13, 2025, and I'm Manish, the maintenance supervisor at a busy factory in India. Our machines keep breaking down, and it's slowing us down big time. I've got a dataset of 10,000 machine runs with info on temperatures, speed, torque, tool wear, and failures. My job? Find the patterns to stop these breakdowns. Let's dive in!

My Quick Checks: Finding the Problem

Manish's Plan to Stop Machine Breakdowns

- **Here's the Deal**: It's 6:00 PM IST on Friday, June 13, 2025, and I'm Manish, the maintenance supervisor at a busy factory in India. Our machines keep breaking down, and it's slowing us down big time. I've got a dataset of 10,000 machine runs with info on temperatures, speed, torque, tool wear, and failures. My job? Find the patterns to stop these breakdowns. Let's dive in!

My Quick Checks: Finding the Problem

- **Check 1: How Often Do Machines Break?**
  First, I need the basics. What's the distribution of the `Machine failure` label in the dataset? How many machines failed, and how many didn't? This will show if breakdowns are a rare issue or a big headache.
- **Check 2: What Products Are We Making?**
  What is the distribution of the 'productID' variable in the dataset? How many instances are of low, medium, and high quality variants?
- **Check 3: How Are Machines Running? Any Weird Numbers?**
  Let's look at machine conditions. What's the range of values for `Air temperature`, `Process temperature`, `Rotational speed`, `Torque`, and `Tool wear`? Are there any outliers in the dataset ?
- **Check 4: What Conditions Cause Failures?**
  I need to spot red flags. Is there any correlation between the continuous variables and the `Machine failure` label? For example, does higher tool wear lead to more machine failures, or is something else the culprit?
- **Check 5: Does Product Type Change How Machines Run?**
  Do product types affect machine conditions? Is there any correlation between the `Product ID` (`Type`) and the continuous variables? For example, is the `Rotational speed` higher for high-quality products than low-quality ones, or do some products stress the machines more?
- **Check 6: Are There Sneaky Patterns to Catch?**
  Let's look deeper. Are there any interactions or non-linear relationships between the variables that matter for predictive maintenance? For example, does the torque shoot up fast with rotational speed, or are there other patterns to help us predict breakdowns?

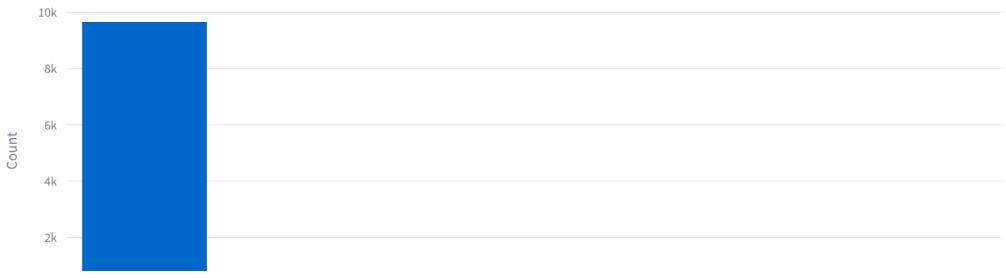
✅ Check 1: How Often Do Machines Break?

What's the distribution of the `Machine failure` label in the dataset?

- The success rate of the machine is **96.61%**

-The highest type of failure is HDF(Heat Dissipation Failure) with 1.15% failure rate.

Failure Type Distribution

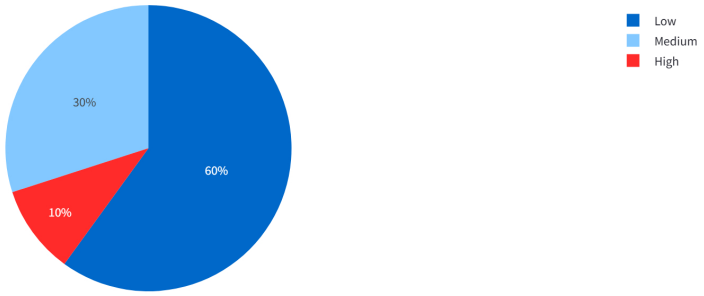


✅ Check 2: What Products Are We Making?

Distribution of the 'productID' or `Type` variable:

- Low: **60.0%**
- Medium: **30.0%**
- High: **10.0%**

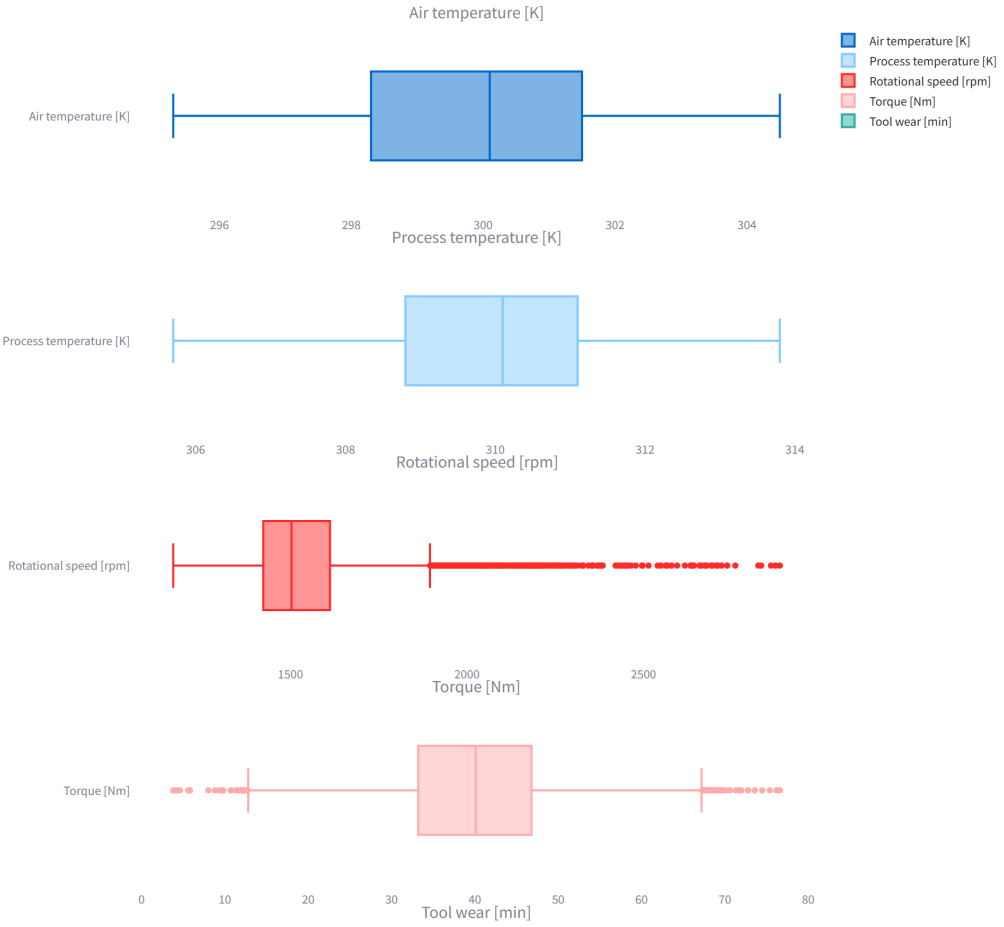
Product Type Distribution

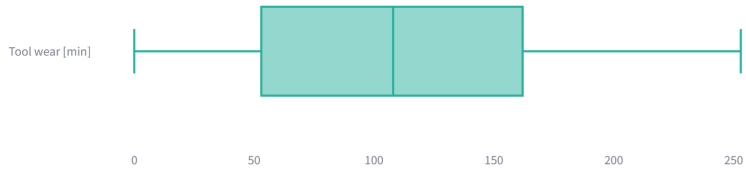


✅ Check 3: How Are Machines Running? Any Weird Numbers?

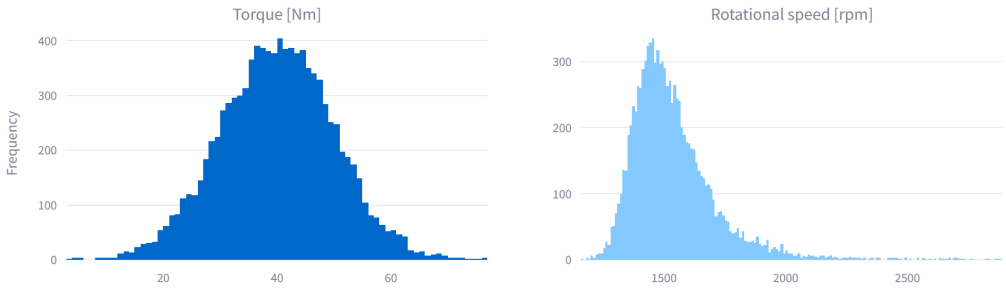
Let's look at machine conditions. Are there any outliers?

Box Plots of Continuous Variables





Histograms of Torque and Speed

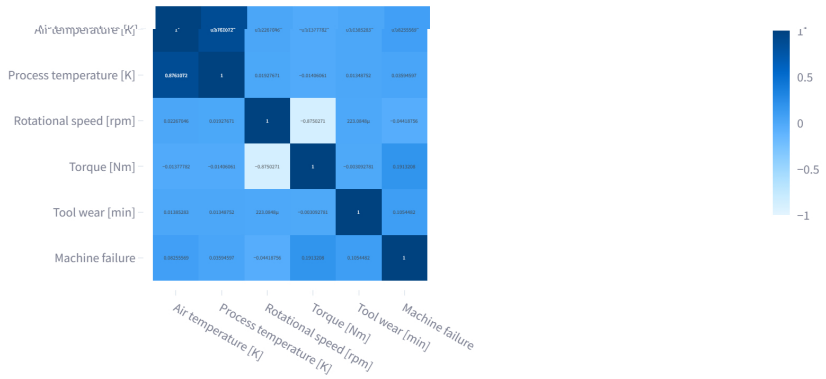


Rotational speed may or may not be actual outliers, therefore we'll keep them in the dataset for now. (same for torque)

✅ Check 4: What Conditions Cause Failures?

I need to spot red flags. Is there any correlation between the continuous variables and the `Machine failure` label? For example, does higher tool wear lead to more machine failures, or is something else the culprit?

Correlation Heatmap



Let's test the hypothesis that continuous variables influence failures using statistical tests.

Null Hypothesis: There is no signifcant relationship between the different columns and Machine Failure.

Alternate Hypothesis: There is a significant relationship between the different columns and the machine failure label.

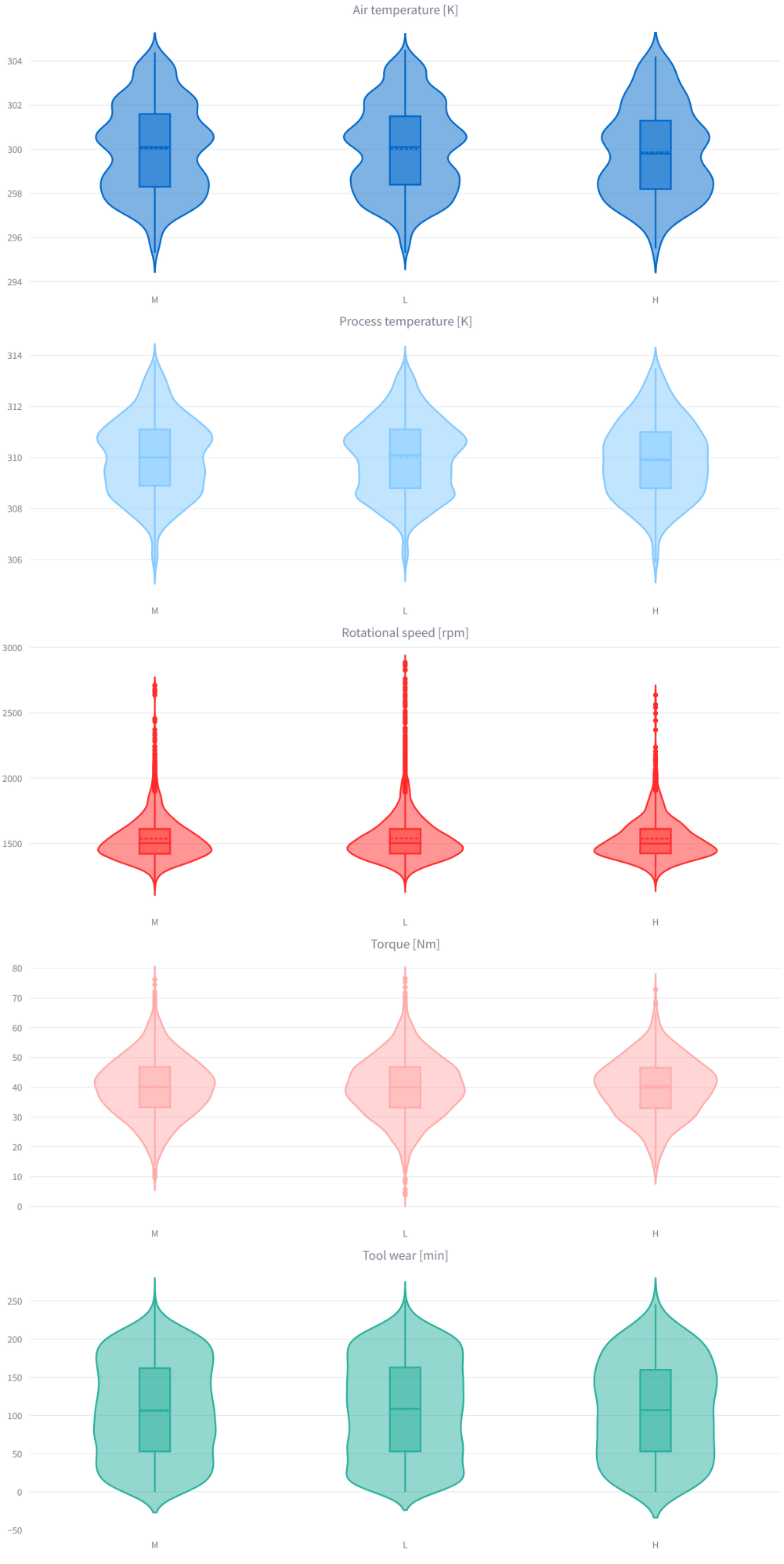
- **Air temperature [K]**: p-value = 0.0000 ✅ Significant
- **Process temperature [K]**: p-value = 0.0003 ✅ Significant
- **Rotational speed [rpm]**: p-value = 0.0000 ✅ Significant
- **Torque [Nm]**: p-value = 0.0000 ✅ Significant
- **Tool wear [min]**: p-value = 0.0000 ✅ Significant

Conclusion: test confirmed that `Air temperature` , `Process temperature` , `Rotational speed` , `Torque` , and `Tool wear` have a strong link to machine failures.

✅ Check 5: Does Product Type Change How Machines Run?

Do product types affect machine conditions? Is there any correlation between the `Product ID` (`Type`) and the continuous variables? For example, is the `Rotational speed` higher for high-quality products than low-quality ones, or do some products stress the machines more??

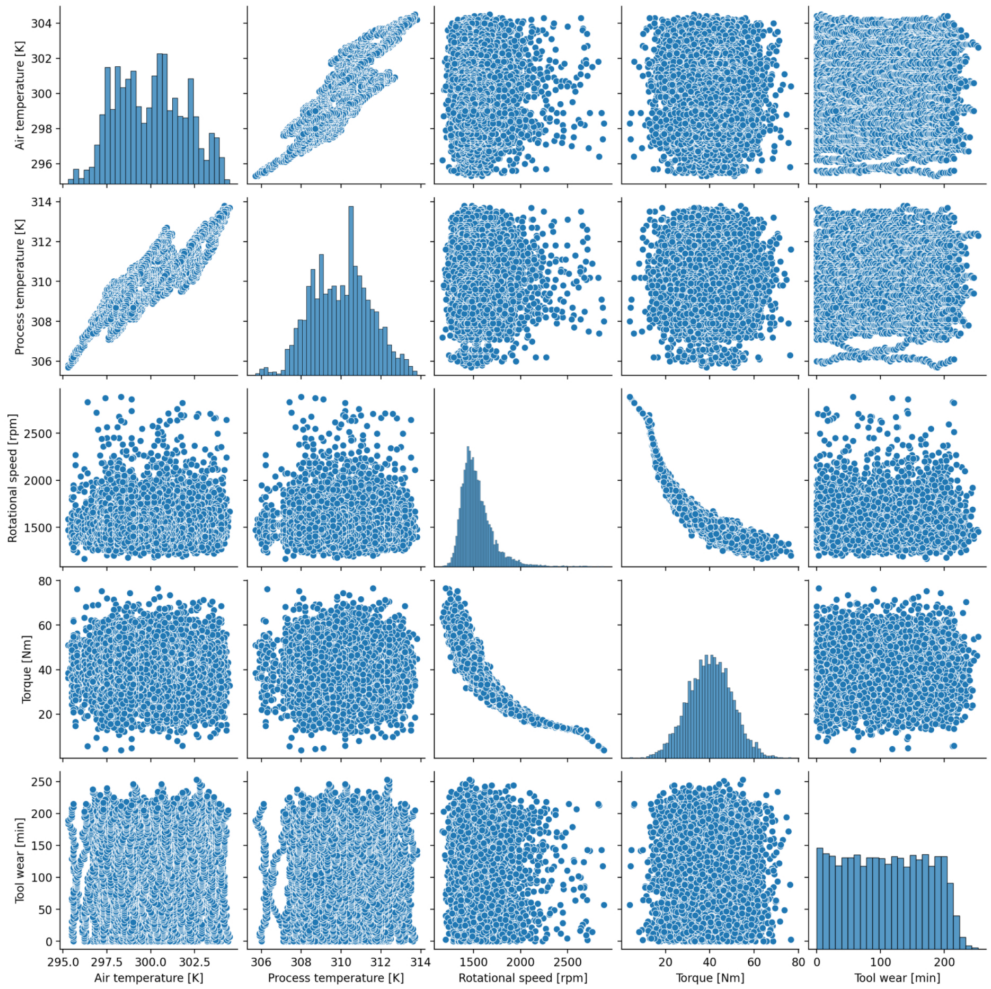
Distribution of Numeric Features by Product Type


Air temperature [K]
304
302
300
298
296
294
M
L
H
Process temperature [K]
314
312
310
308
306
M
L
H
Rotational speed [rpm]
3000
2500
2000
1500
M
L
H
Torque [Nm]
80
70
60
50
40
30
20
10
0
M
L
H
Tool wear [min]
250
200
150
100
50
0
−50
M
L
H

Conclusion: low-quality products may stress machines more, and key conditions like tool wear significantly impact failures

✅ Check 6: Are There Sneaky Patterns to Catch?

Let's look deeper. Are there any interactions or non-linear relationships between the variables that matter for predictive maintenance? For example, does the torque shoot up fast with rotational speed, or are there other patterns to help us predict breakdowns?



Among all possible combinations of continuous variables, Rotational Speed vs Torque have a negative correlation and process temperature vs air temperature have a positive correlation.

🛠️ Final Conclusion

Manish's analysis reveals that:

- Machine failures occur in **3.48%** of cases
- Low-quality products cause higher stress on machines
- Key features like **tool wear** and **torque** significantly influence failures

He plans to monitor low-quality products closely and build a predictive system to minimize breakdowns and boost efficiency.