Predictive Maintenance

September 30, 2025

[23]:
```
import pandas as pd
import numpy as np
import seaborn as sea
import matplotlib.pyplot as plt

df = pd.read_csv('../data/ai4i2020.csv')

print("Data set loaded")
df.head()
```

```
Data set loaded
```

[23]:

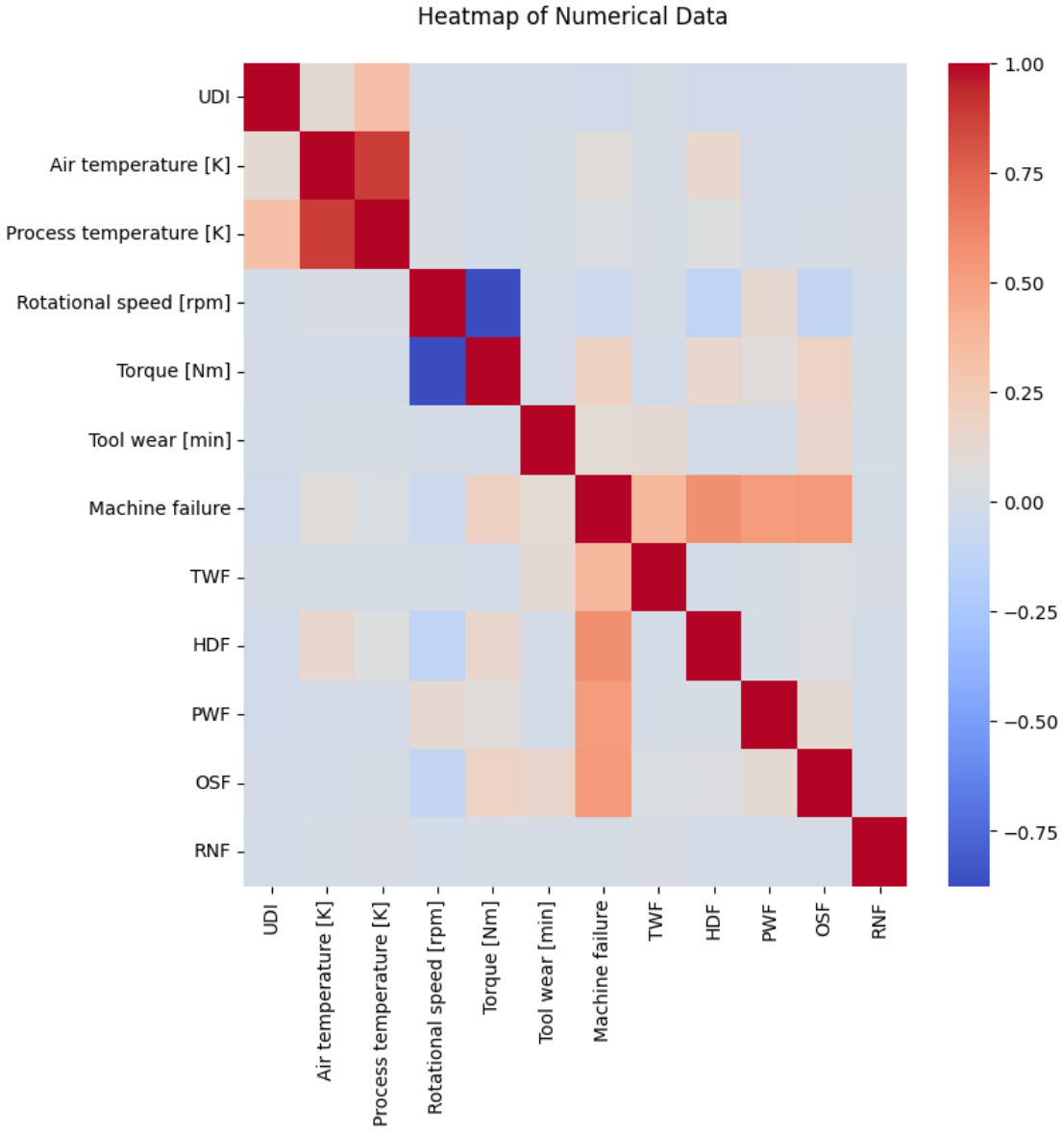
	UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	\
0	1	M14860	M	298.1	308.6	
1	2	L47181	L	298.2	308.7	
2	3	L47182	L	298.1	308.5	
3	4	L47183	L	298.2	308.6	
4	5	L47184	L	298.2	308.7	

	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	\
0	1551	42.8	0	0	0	
1	1408	46.3	3	0	0	
2	1498	49.4	5	0	0	
3	1433	39.5	7	0	0	
4	1408	40.0	9	0	0	

	HDF	PWF	OSF	RNF
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

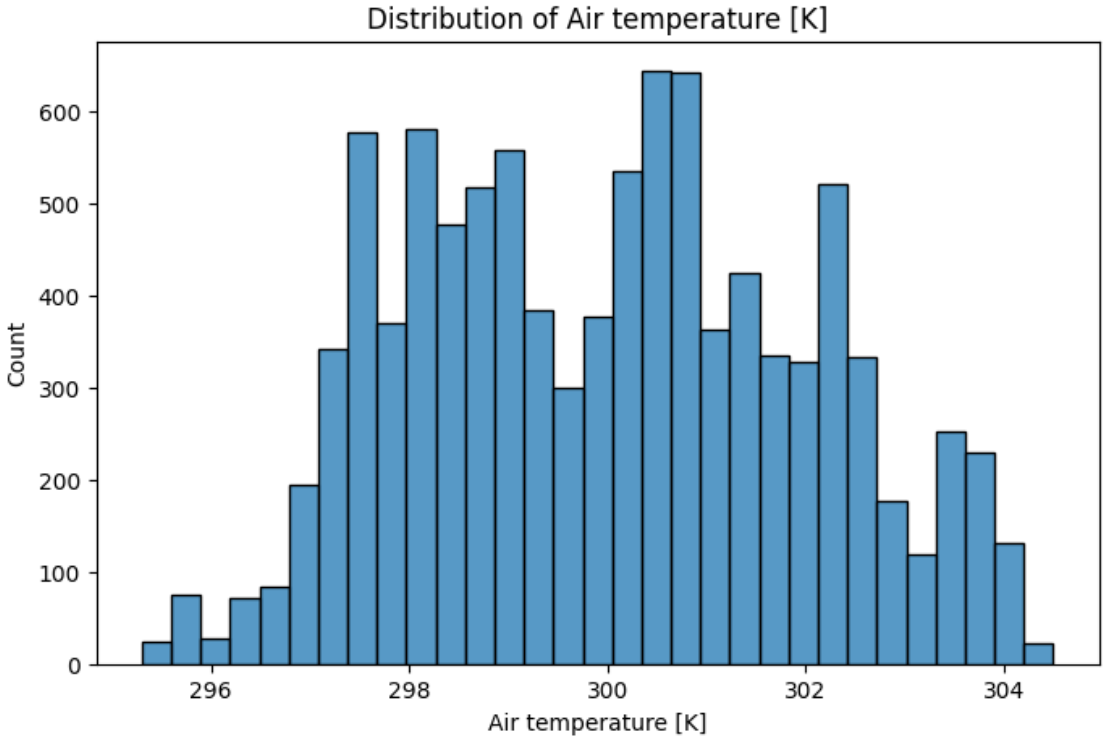
0.1 Correlation Heatmap

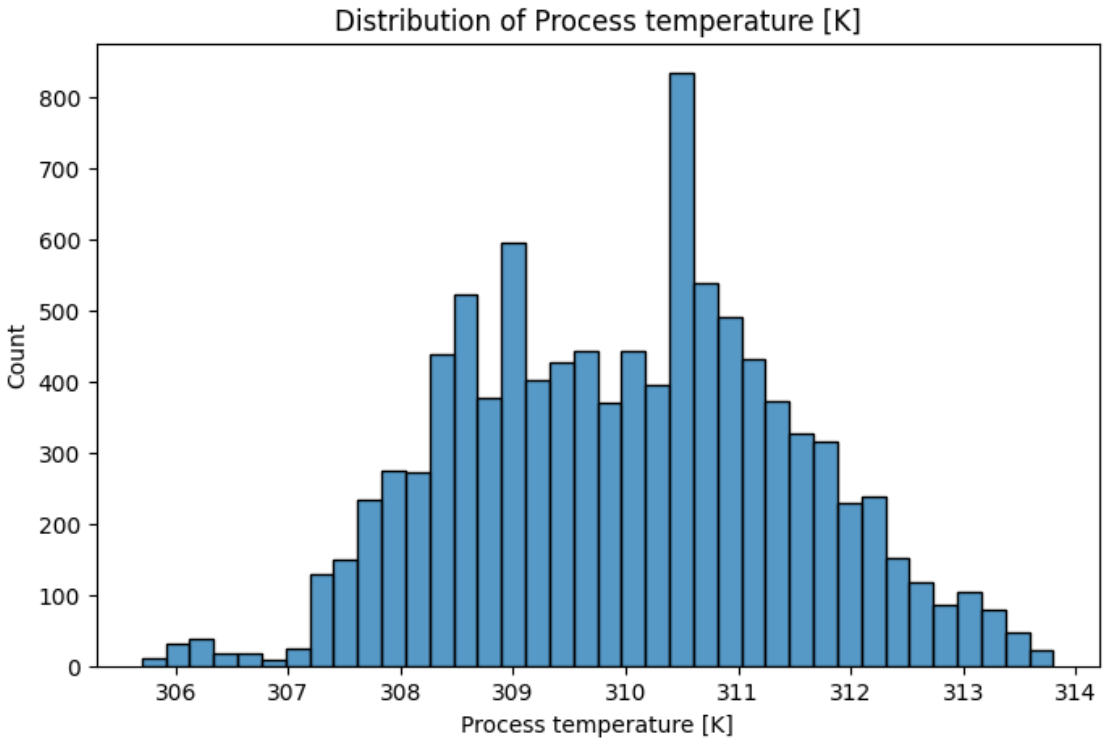
[27]:
```
df_numeric_only = df.drop(columns=['Product ID', 'Type'])

correlation_matrix = df_numeric_only.corr()

plt.figure(figsize=(8, 8))
sea.heatmap(correlation_matrix, annot=False, cmap='coolwarm')
plt.title(f'Heatmap of Numerical Data \n')
plt.show()
```

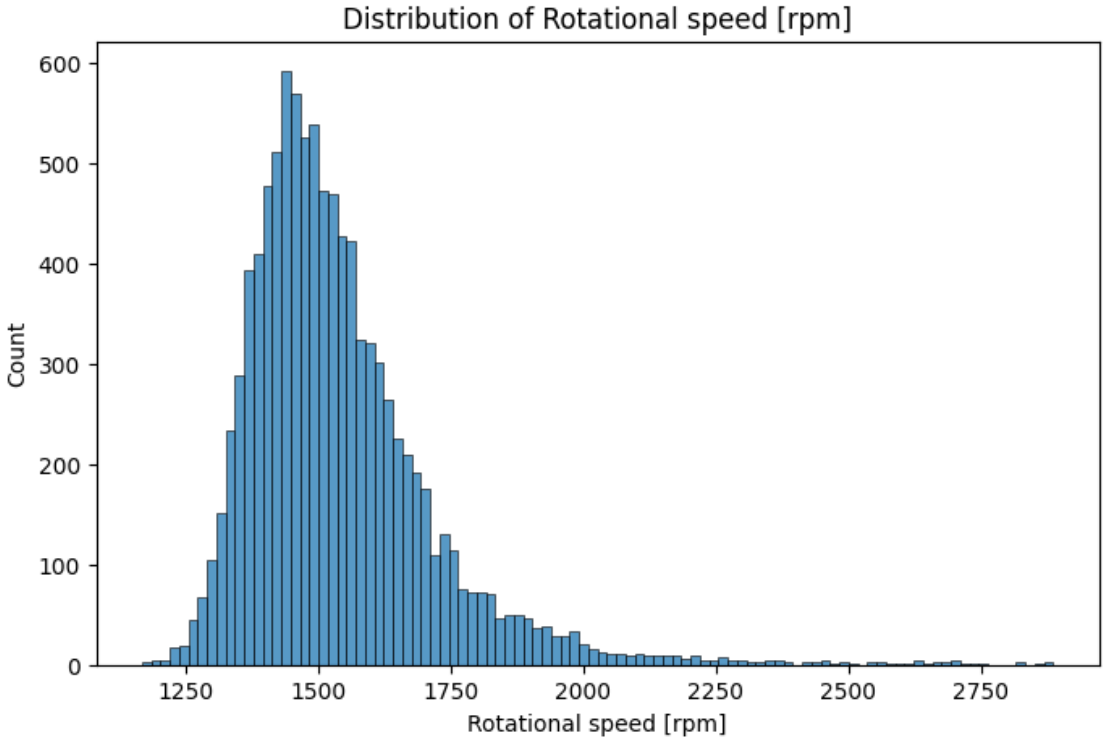


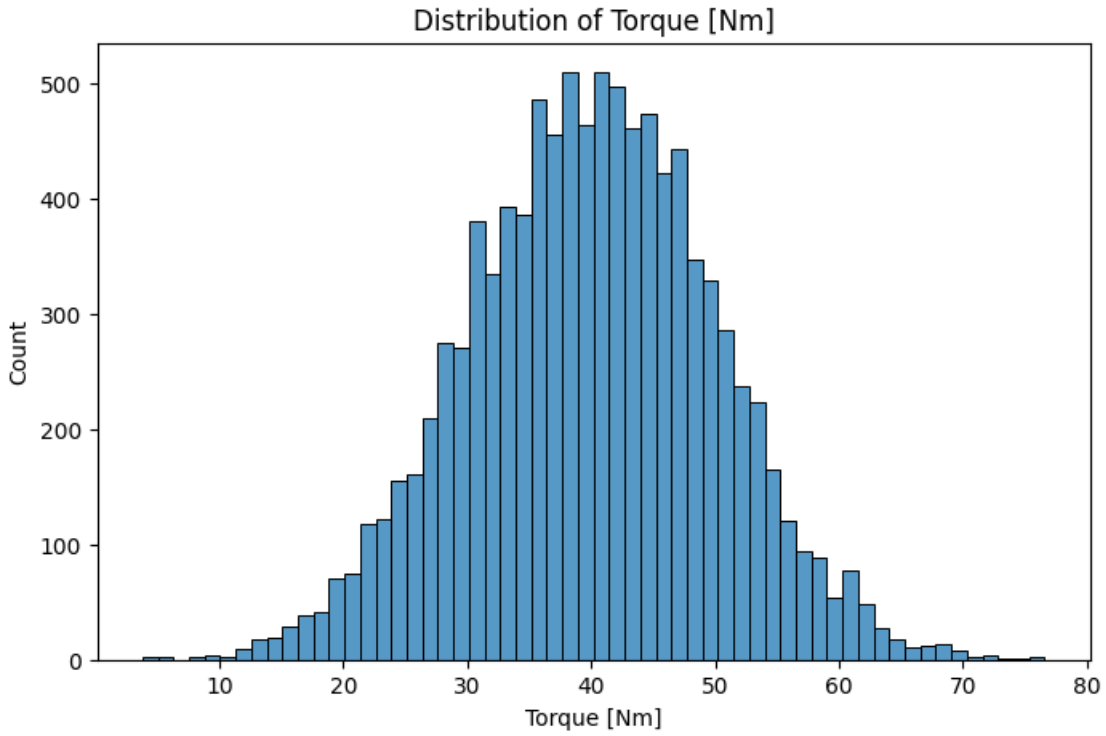
0.2 Histogram

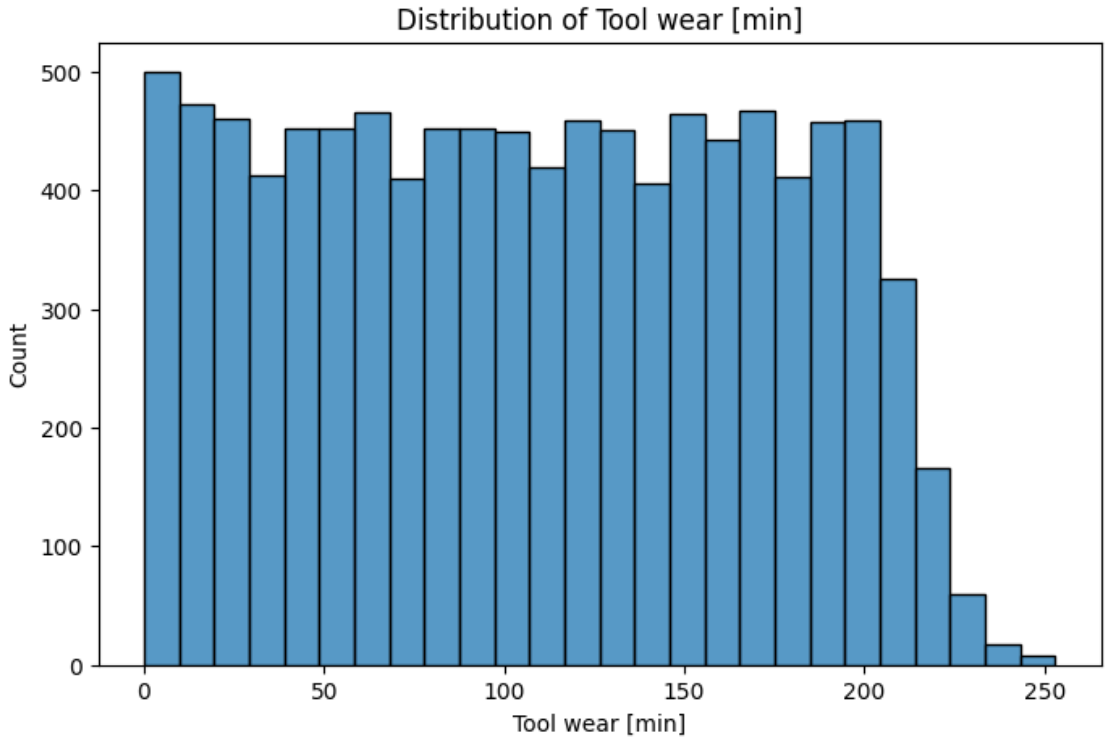
[25]:
```
hist_columns = [
    'Air temperature [K]',
    'Process temperature [K]',
    'Rotational speed [rpm]',
    'Torque [Nm]',
    'Tool wear [min]'
]

for column_name in hist_columns:
    plt.figure(figsize=(8, 5))
    sea.histplot(data=df, x=column_name)
    plt.title(f'Distribution of {column_name}')
    plt.show()
```




Distribution of Process temperature [K]
Count
Process temperature [K]
0
100
200
300
400
500
600
700
800
306
307
308
309
310
311
312
313
314

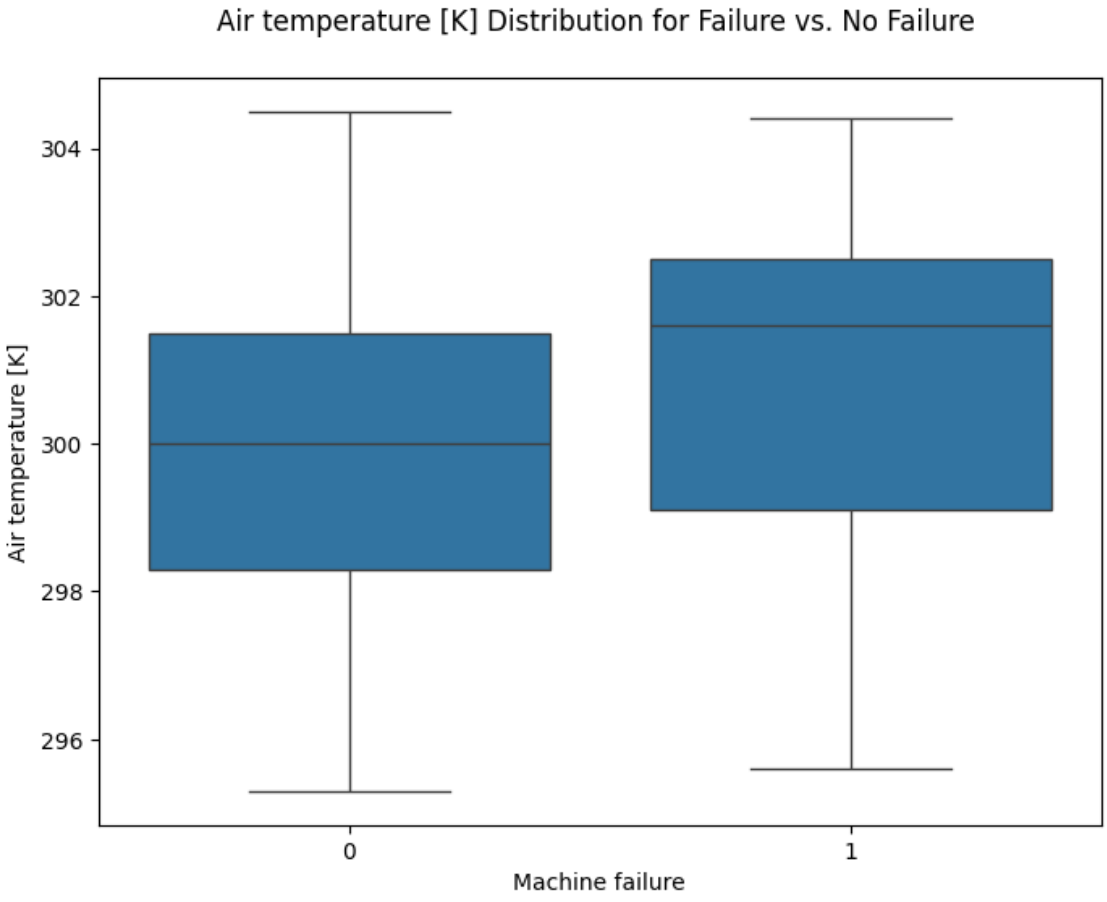

Distribution of Rotational speed [rpm]
Count
Rotational speed [rpm]
0
100
200
300
400
500
600
1250
1500
1750
2000
2250
2500
2750

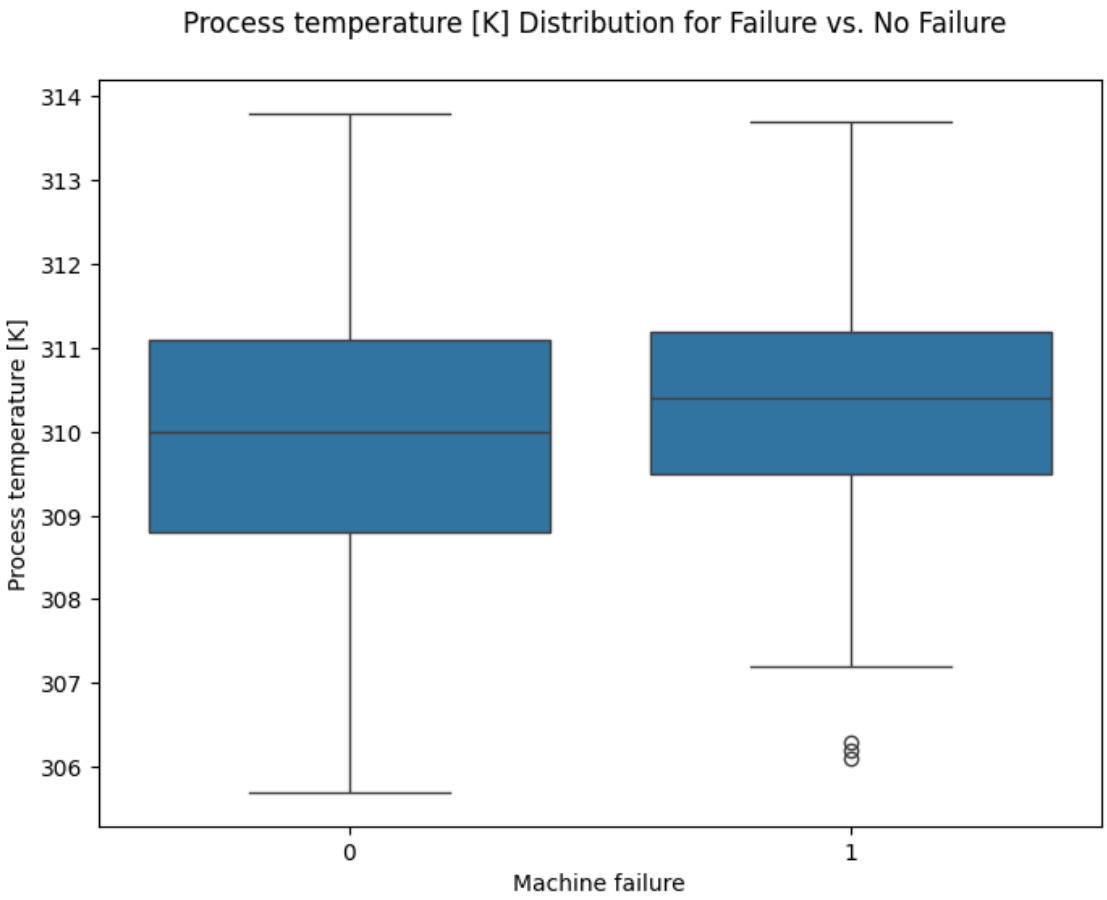

Distribution of Torque [Nm]
Count
Torque [Nm]
0
100
200
300
400
500
10
20
30
40
50
60
70
80

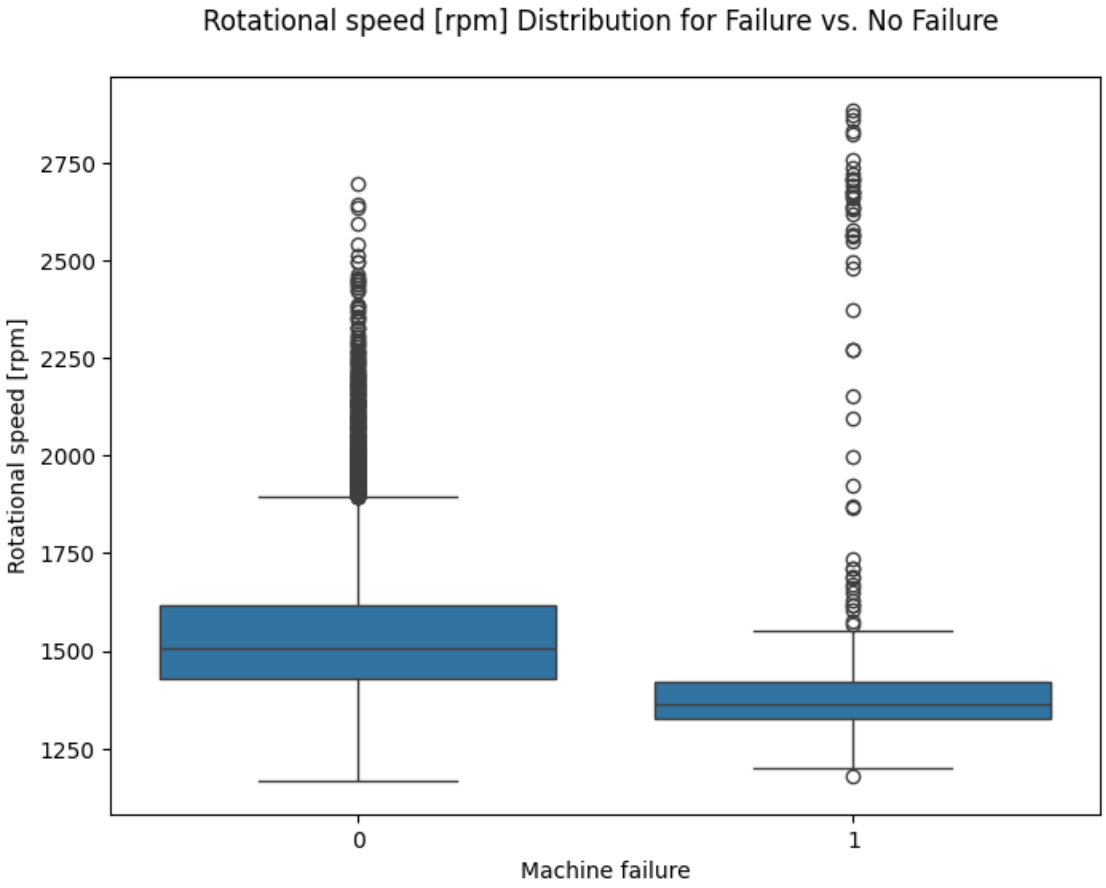

Distribution of Tool wear [min]
Count
Tool wear [min]
0
100
200
300
400
500
0
50
100
150
200
250

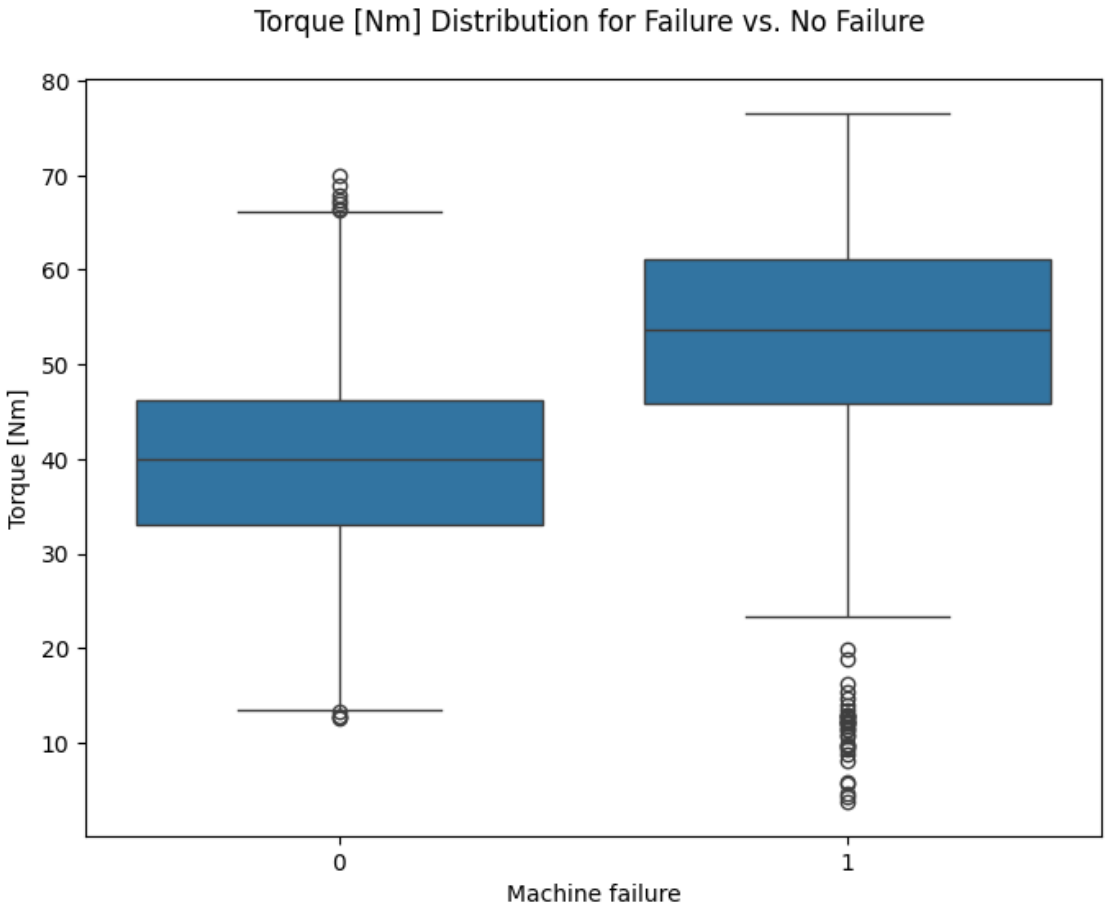
0.3 Box Plots

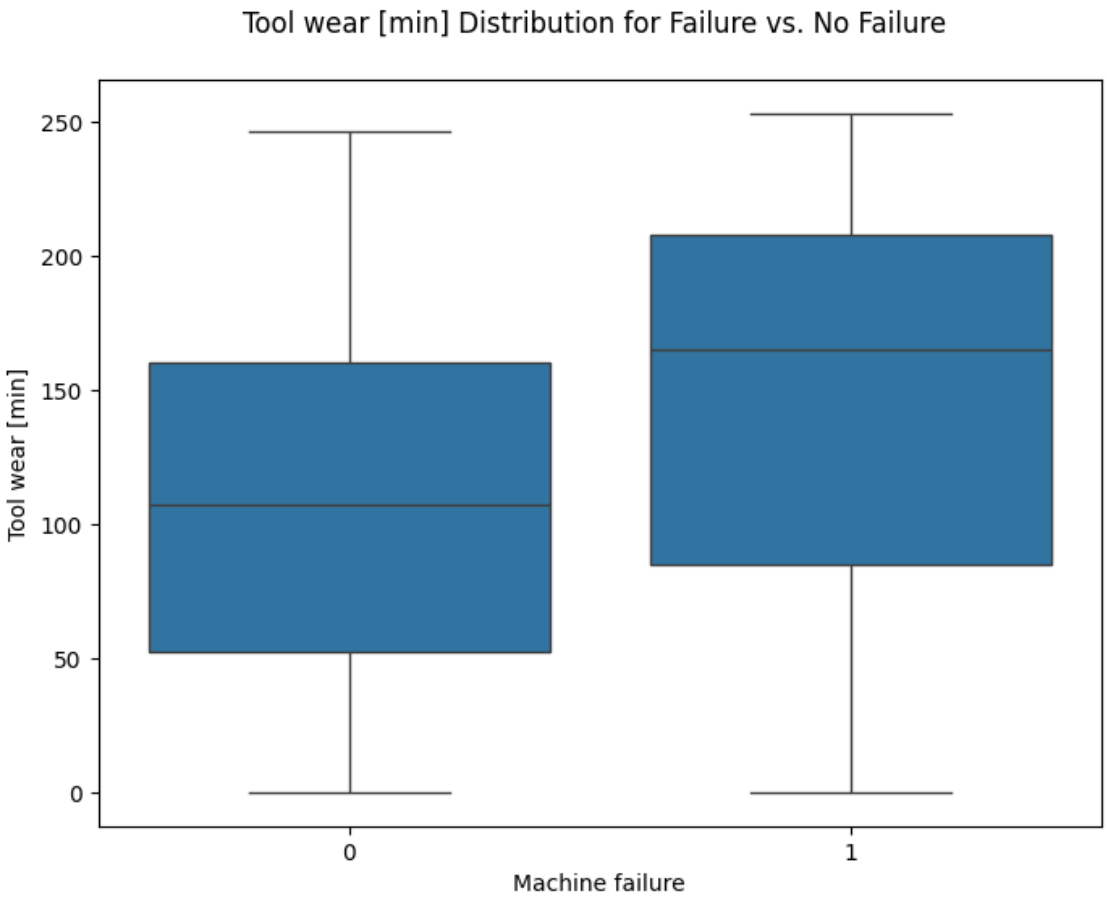
[26]:
```
sensor_columns = [
    'Air temperature [K]',
    'Process temperature [K]',
    'Rotational speed [rpm]',
    'Torque [Nm]',
    'Tool wear [min]'
]

for column in sensor_columns:
    plt.figure(figsize=(8, 6))
    sea.boxplot(x='Machine failure', y=column, data=df)
    plt.title(f'{column} Distribution for Failure vs. No Failure \n')
    plt.show()
```




Process temperature [K] Distribution for Failure vs. No Failure
Process temperature [K]
314
313
312
311
310
309
308
307
306
0
1
Machine failure


Rotational speed [rpm] Distribution for Failure vs. No Failure
Rotational speed [rpm]
2750
2500
2250
2000
1750
1500
1250
0
1
Machine failure


Torque [Nm] Distribution for Failure vs. No Failure
Torque [Nm]
80
70
60
50
40
30
20
10
0
1
Machine failure



0.4 Key Predictive Features

Based on the exploratory data analysis, the following features show the strongest correlation with machine failure and will be used for development:

- Torque [Nm]
- Rotational Speed [rpm]
- Tool Wear [min]