

Demonstração gráfica entre os dados das máquinas e seus dados de funcionamento versus os tipos de falhas apresentados.

*Fora descartada a apresentação das falhas randômicas por serem, justamente, randômicas (com 0.1% de ocorrência).

Inicialmente, decidi considerar os dados fornecidos na tabela "desafio_manutencao_preditiva_treino" para definir um padrão. Usando das informações fornecidas nas instruções do desafio e dos dados do arquivo csv extraídos pelos scripts em python, é como determinei como o caminho para realizar a previsão do tipo de falha. Para isso, recorri a explicações em páginas que tratam sobre o tema. Após, somei meu básico conhecimento em Python e comecei a planejar como manipular os dados para obter o resultado. Sem conhecimentos na área de dados e machine learning, recorri ao auxílio da inteligência artificial. De acordo com o que aprendi nas buscas e com o auxílio da inteligência artificial, obtive um código que primeiro pré-processa os dados, convertendo valores não numéricos em valores numéricos e preparando os dados para em sequência, se ter a previsão para o desafio. Naturalmente, houve a correção e adaptação de códigos para se enquadrarem nas características do desafio, como a retirada do valor "type" de consideração, pois seu cálculo de probabilidade esta fora do conjunto de valor considerado relevante (Apresentado no documento com dados gráficos). Estudando o tema, compreendi que o tipo de problema a ser resolvido é o de regressão, pois está sendo utilizado um conjunto de dados de entrada já analisados, para prever um conjunto de dados de uma situação não analisada. No presente momento, por ser iniciante na área, não fui capaz de determinar uma medida de performance para apresentar.

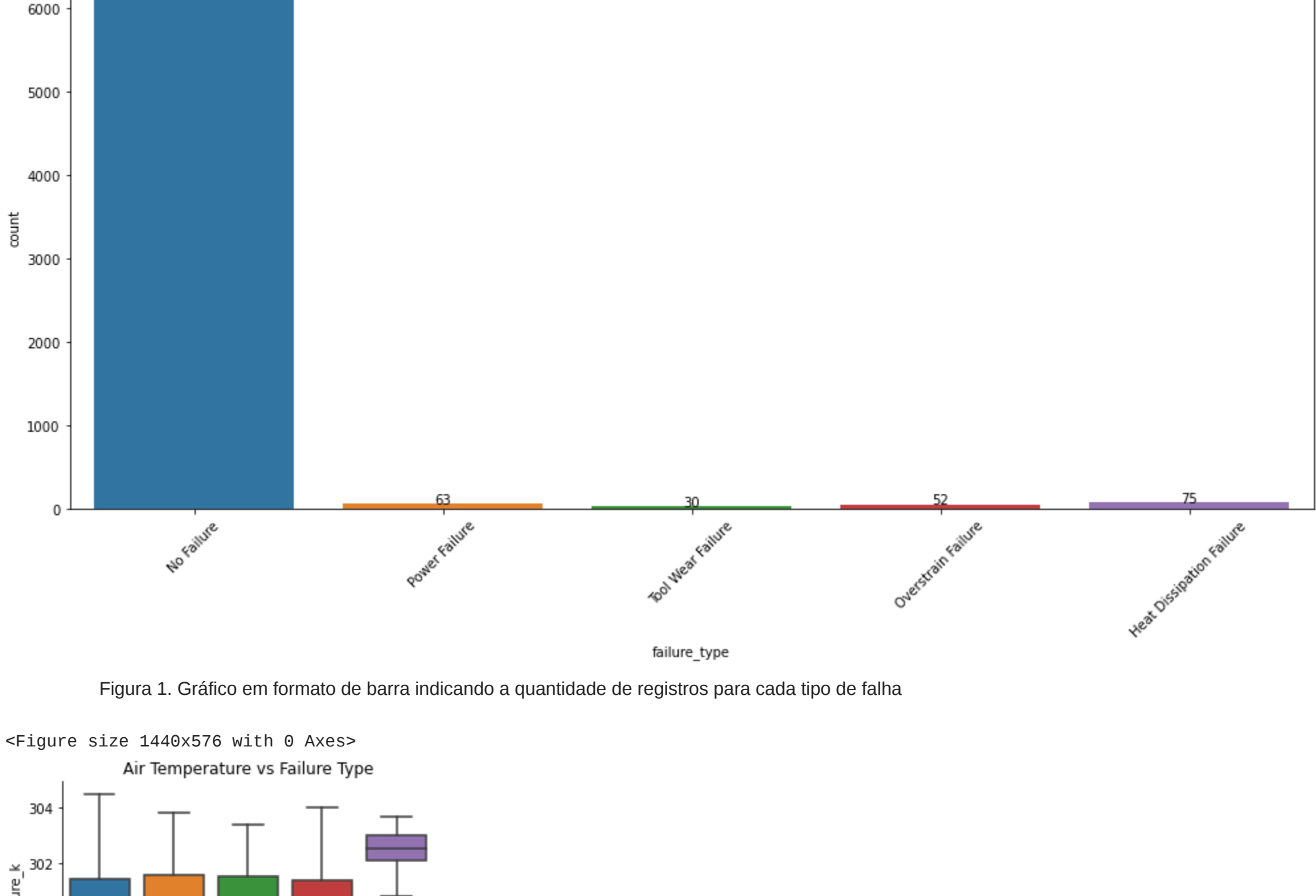


Figura 1. Gráfico em formato de barra indicando a quantidade de registros para cada tipo de falha

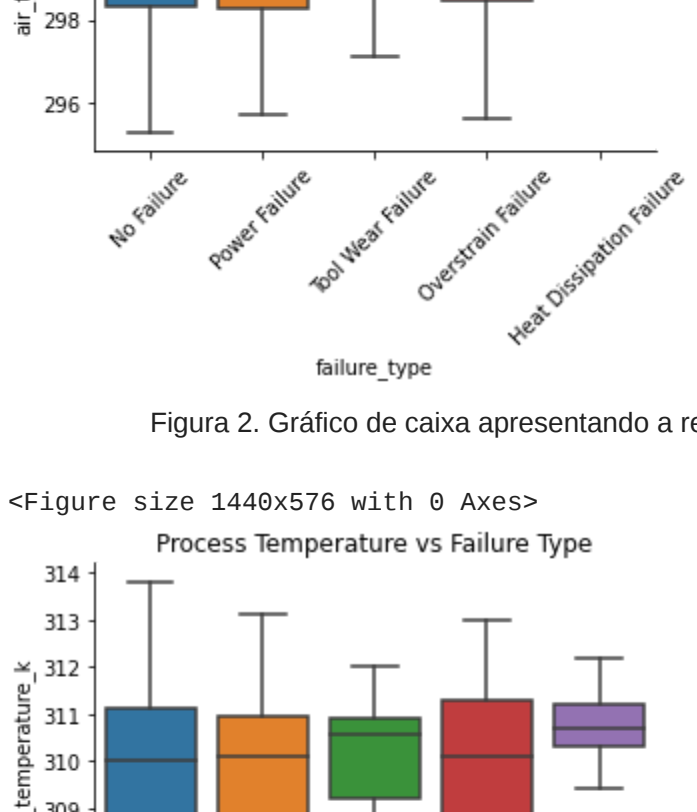


Figura 2. Gráfico de caixa apresentando a relação entre a temperatura do ar e os tipos de falhas.

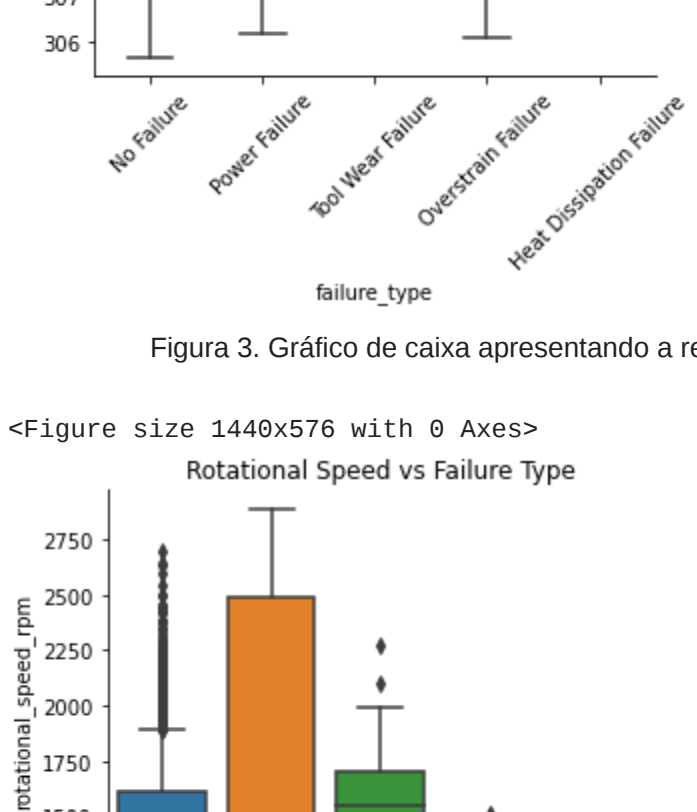


Figura 3. Gráfico de caixa apresentando a relação entre a temperatura do processo e os tipos de falhas.

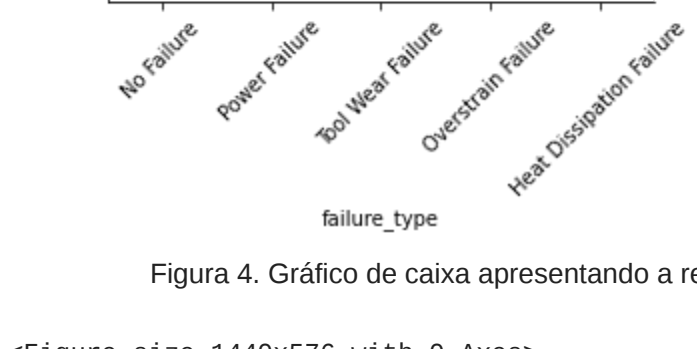


Figura 4. Gráfico de caixa apresentando a relação entre a velocidade rotacional em RPM e os tipos de falha

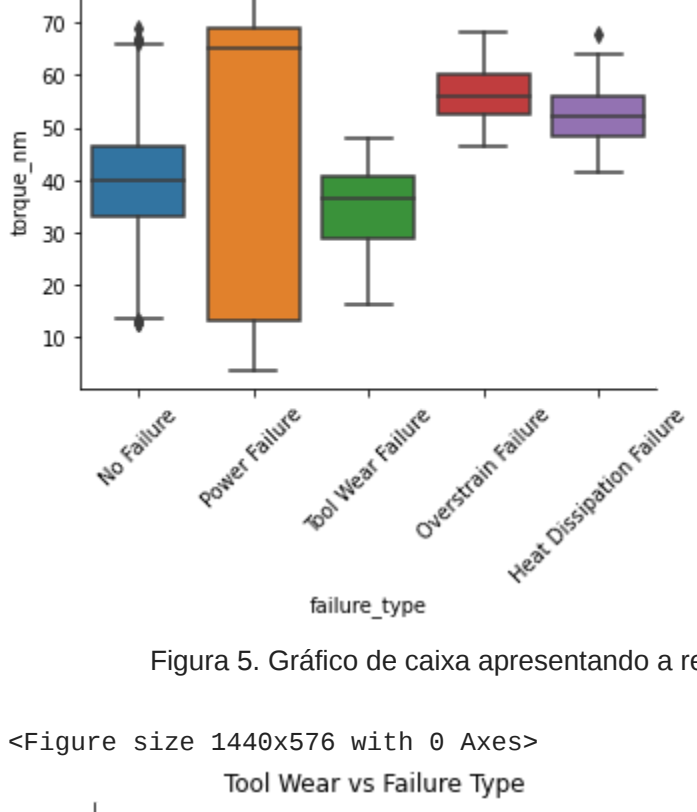


Figura 5. Gráfico de caixa apresentando a relação entre o torque e os tipos de falhas.

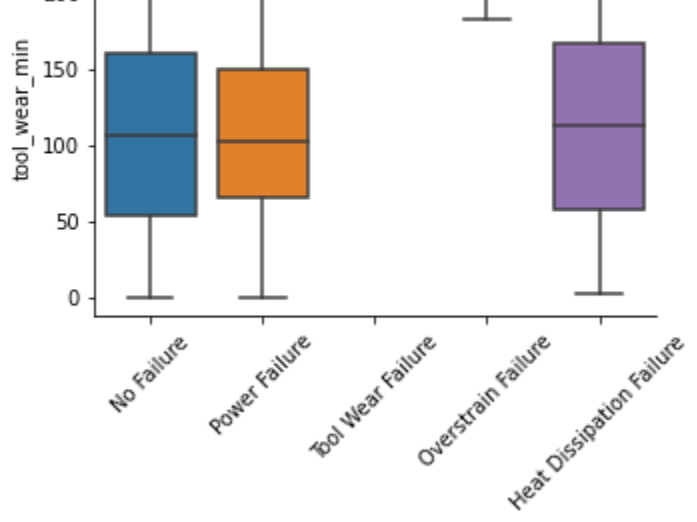


Figura 6. Gráfico de caixa apresentando a relação entre o desgaste do equipamento e os tipos de falhas.

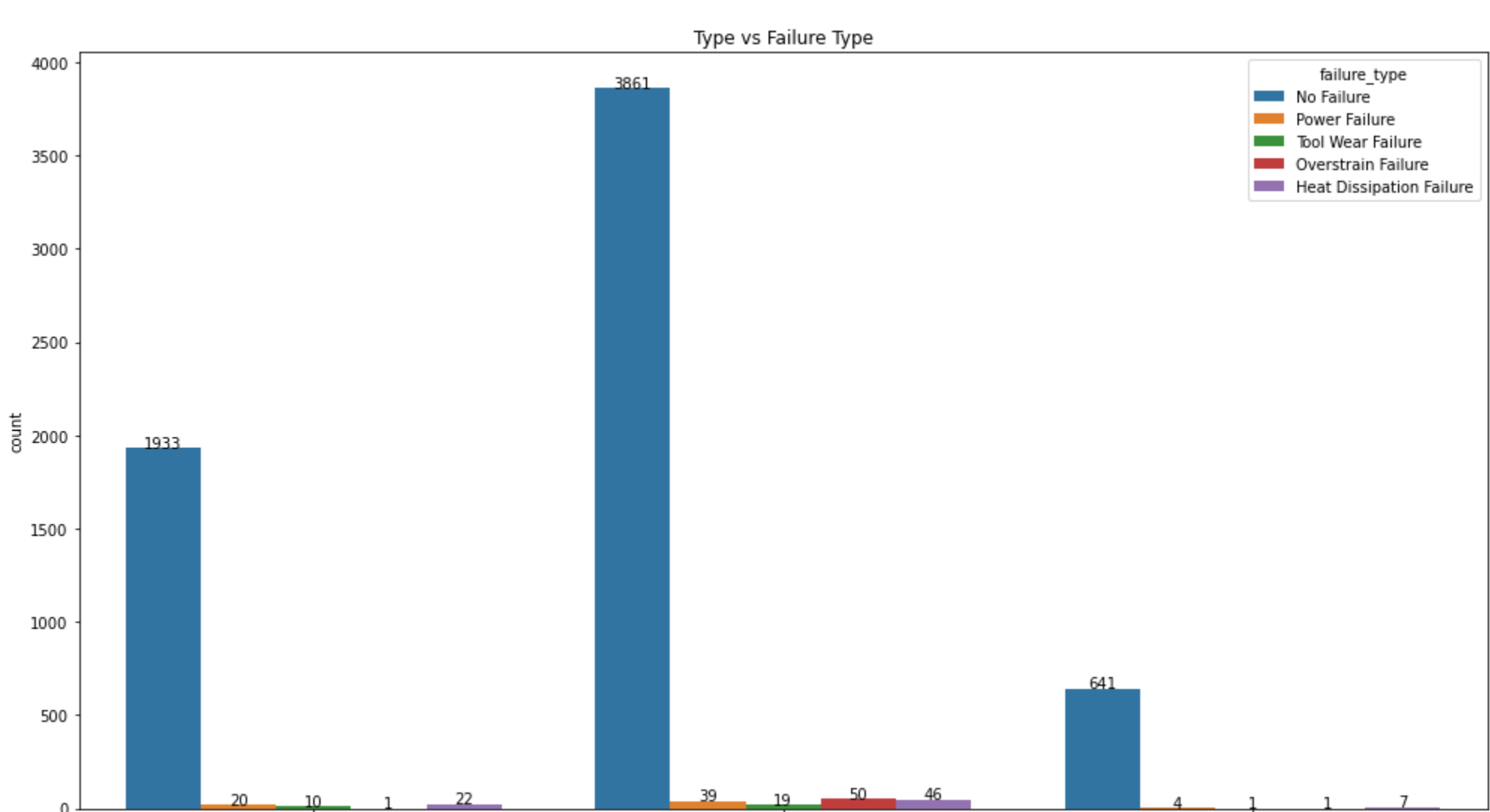


Figura 7. Gráfico de barras apresentando a relação entre os três níveis de qualidade do produto e os tipos de falhas.

As demonstrações gráficas acima foram escolhidas para se obter uma percepção visual da incidência dos diferentes tipos de falhas em números e em relação às condições de funcionamento.

Correlation between variables

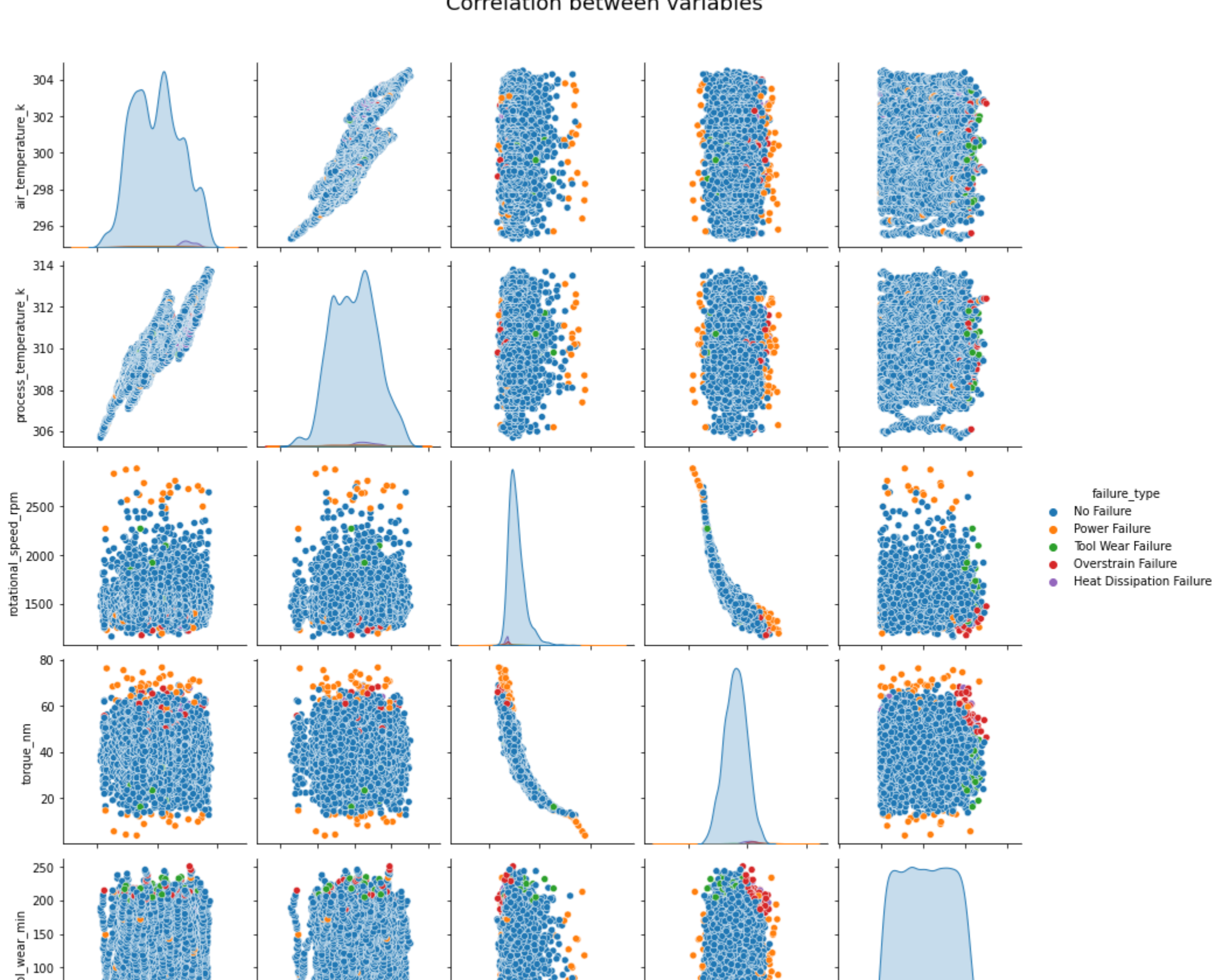


Figura 8. Apresenta uma matriz apresentando correlação entre as variáveis da tabela de treino.

Porcentual dos tipos de falhas:
No Failure 96.528174
Heat Dissipation Failure 1.124944
Power Failure 0.944953
Overstrain Failure 0.779961
Tool Wear Failure 0.448978
Random Failures 0.179993
Name: failure_type, dtype: float64

Temos então, os percentuais que representam os tipos de falhas dos dados de treino. Apresenta-se um total de aproximadamente 96.5% de máquinas que não apresentaram falhas e um total de aproximadamente 3.5% que apresentou algum tipo de falha.

A partir deste ponto, tendo adicionado ao arquivo "desafio_manutencao_preditiva_teste.csv" a coluna "failure_type" proveniente da previsão de falhas do código "Pre_Processamento e ML", se tem a comparação estatística dos dados. Abaixo as demonstrações gráficas extraídas a partir da previsão de falhas ("desafio_manutencao_preditiva_teste"), para fins de comparação com as extraídas do "treino".

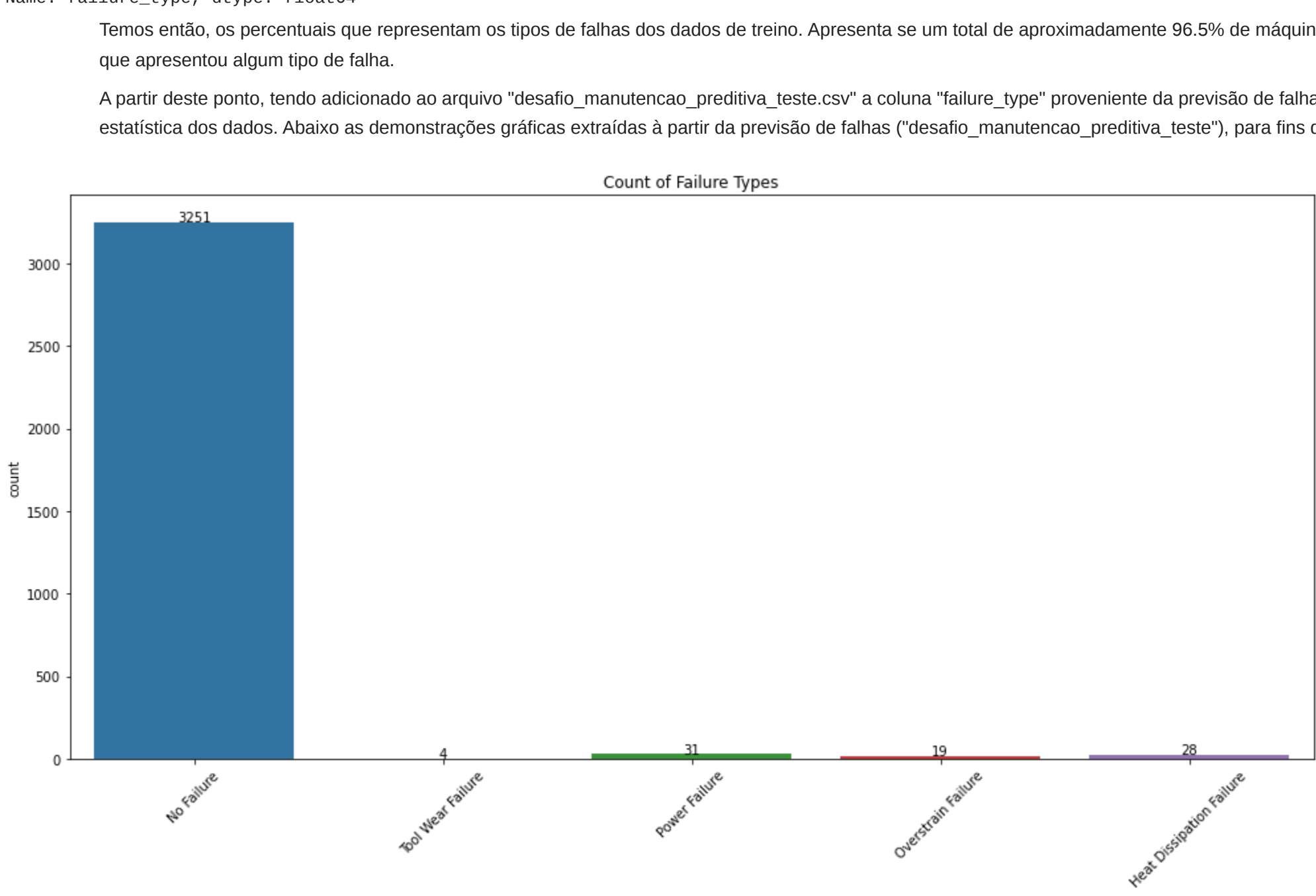


Figura 9. Gráfico em formato de barra indicando a quantidade de registros para cada tipo de falha

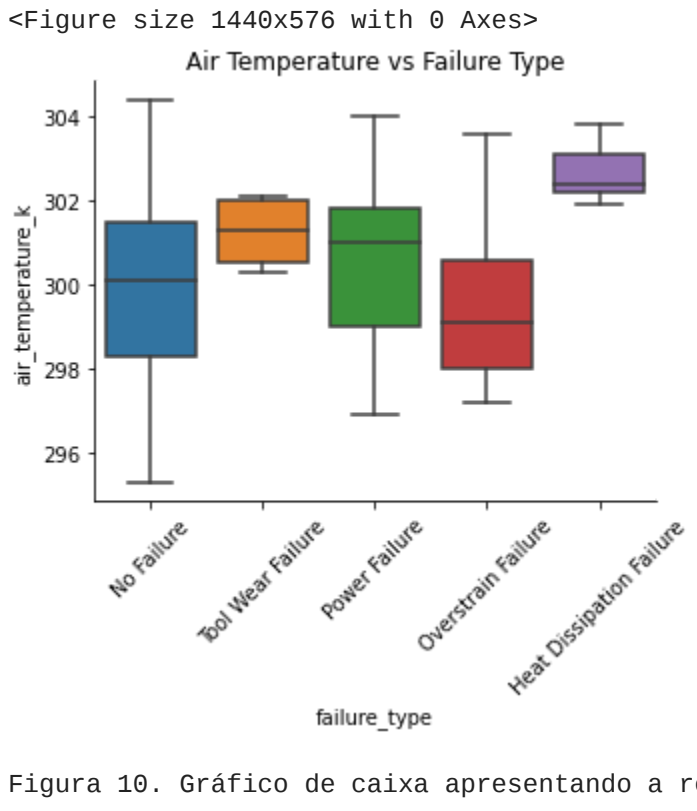


Figura 10. Gráfico de caixa apresentando a relação entre a temperatura do ar e os tipos de falhas.

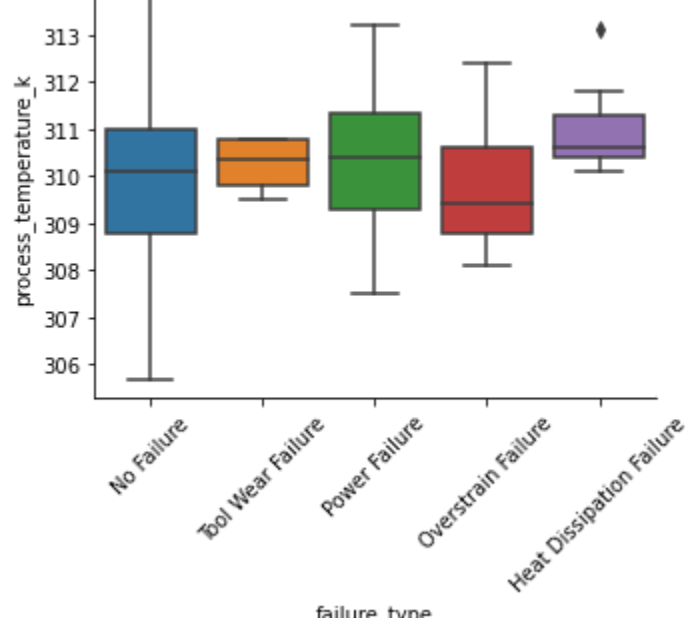


Figura 11. Gráfico de caixa apresentando a relação entre a temperatura do processo e os tipos de falhas.

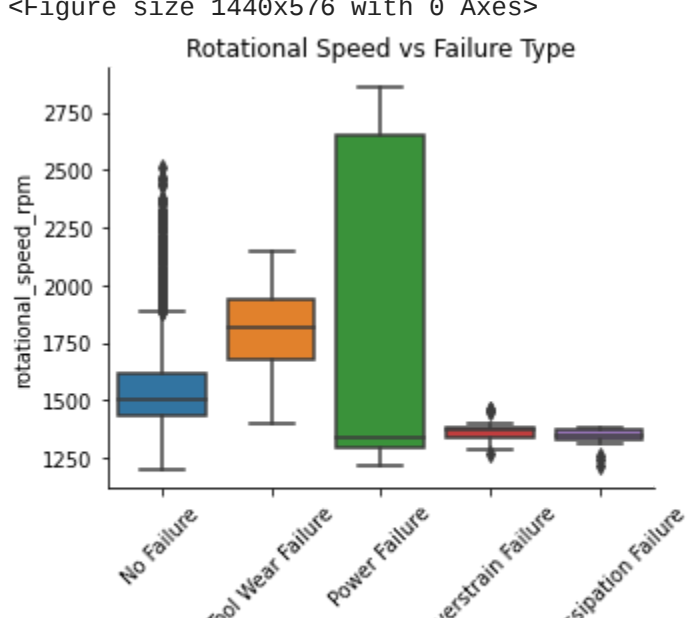


Figura 12. Gráfico de caixa apresentando a relação entre a velocidade rotacional em RPM e os tipos de falha

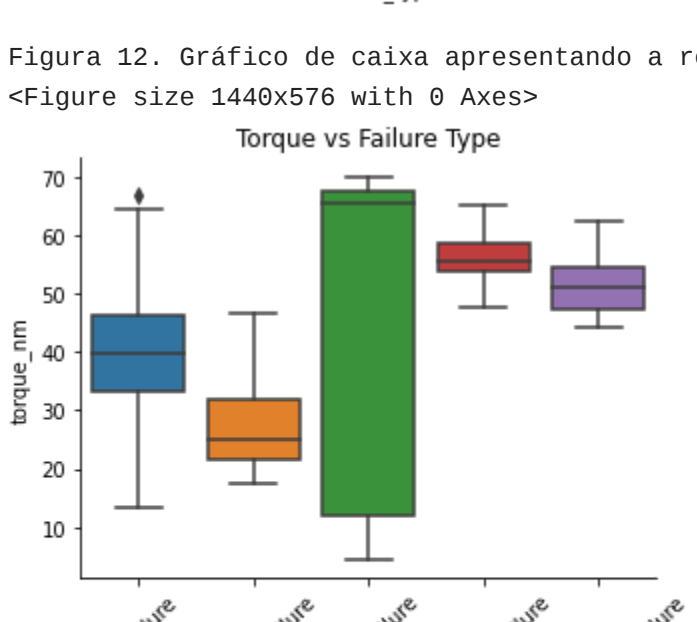


Figura 13. Gráfico de caixa apresentando a relação entre o torque e os tipos de falhas.

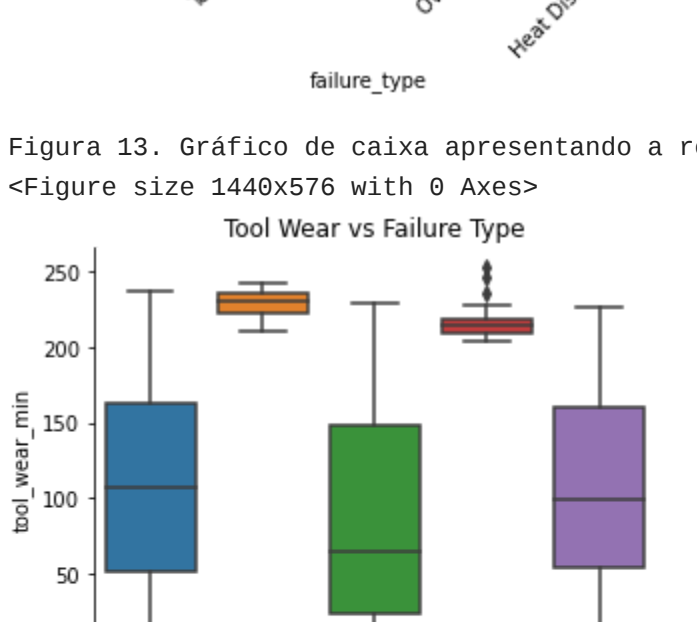


Figura 14. Gráfico de caixa apresentando a relação entre o desgaste do equipamento e os tipos de falhas.

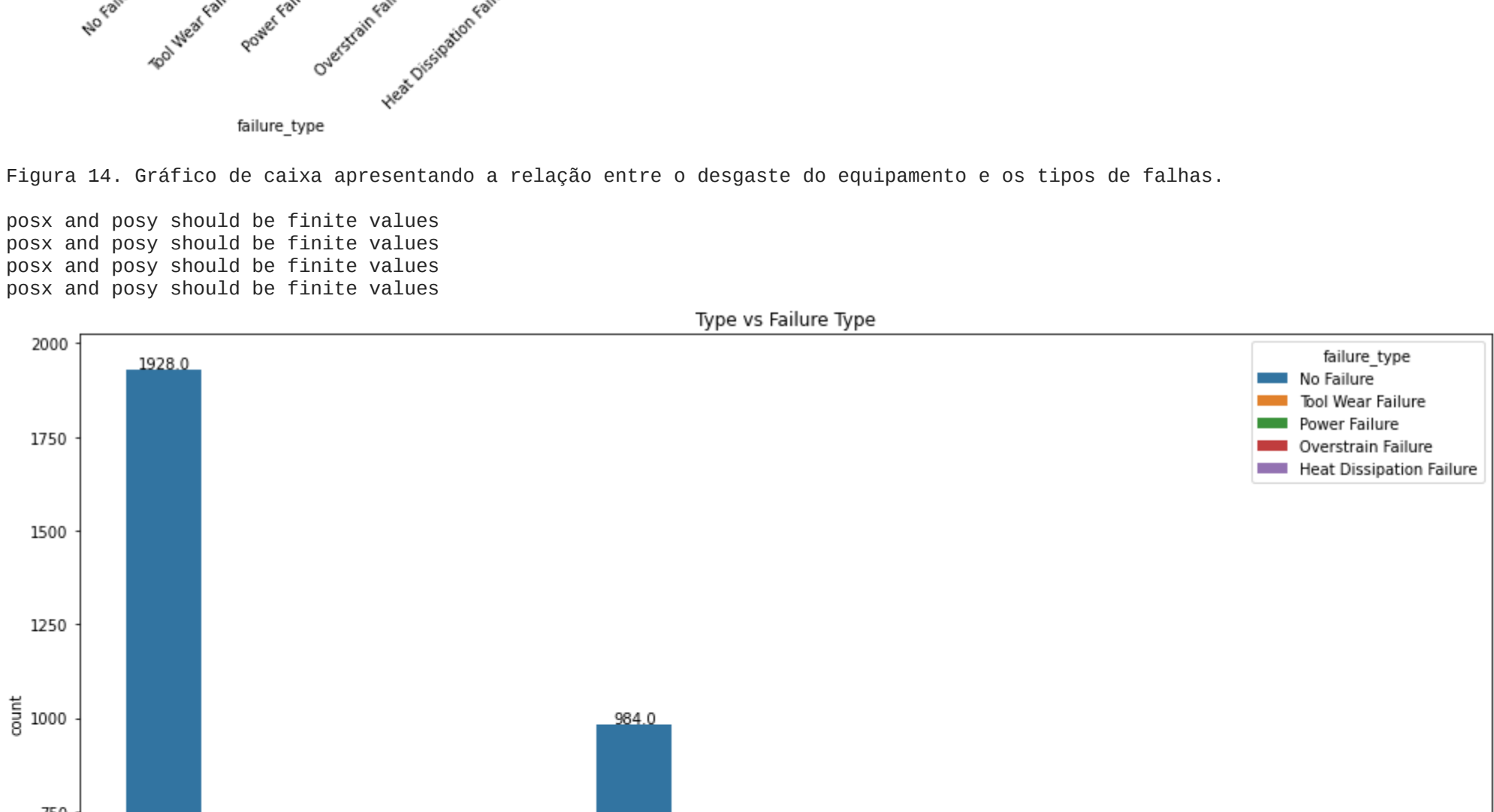


Figura 15. Gráfico de barras apresentando a relação entre os três níveis de qualidade do produto e os tipos de falhas.

Correlation between variables

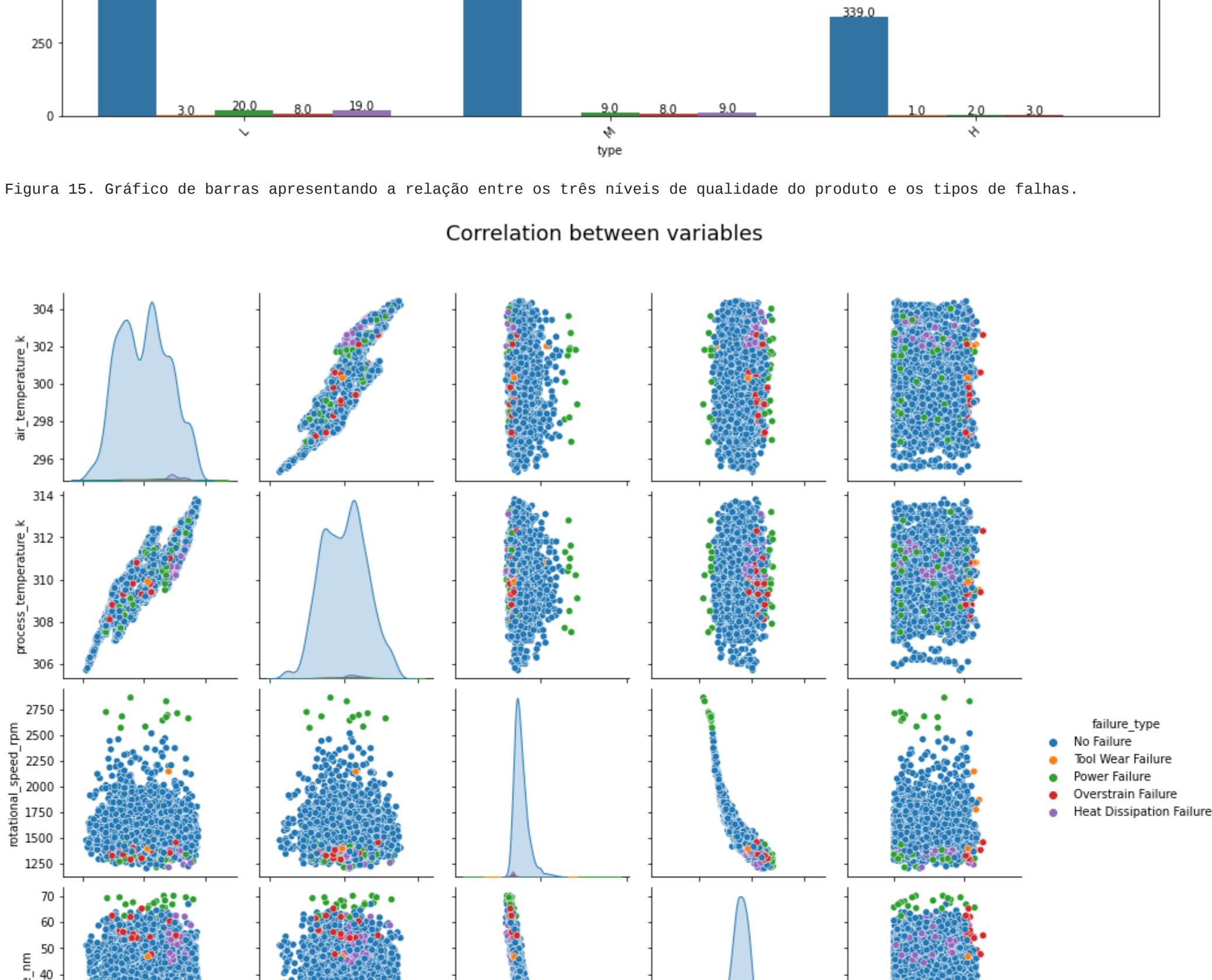


Figura 16. Apresenta uma matriz apresentando correlação entre as variáveis da tabela de teste.

Porcentual dos tipos de falhas:
No Failure 97.539764
Power Failure 0.938993
Heat Dissipation Failure 0.846538
Overstrain Failure 0.578987
Tool Wear Failure 0.120012
Name: failure_type, dtype: float64

Acima os percentuais que representam os tipos de falhas resultantes da previsão. Apresenta-se um total de aproximadamente 97.5% de máquinas que não apresentaram falhas e um total de aproximadamente 2.5% que apresentou algum tipo de falha.

Percentage Differences between the first and second CSV files for each failure type:

No Failure 49.478469
Overstrain Failure 66.461538
Power Failure 59.793553
Random Failures NaN
Tool Wear Failure 86.666667
Name: failure_type, dtype: float64

Acima, foi calculado a diferença percentual da ocorrência de falhas entre o arquivo de treino e de teste. Temos como resultado, um aumento percentual da ocorrência de falhas.

F-value: 1.1555818639565562
p-value: 6.314937659455715373

Acima, há o resultado do cálculo para considerar se há relevância do tipo da máquina (H.L ou M) no evento de falhas. De acordo com o valor p, que é apresentado como 0.31 ou 31%, e levando em consideração que o valor de probabilidade relevante deve ser igual ou menor que 0.05 ou 5%, se tem a presunção que o tipo da máquina não influencia significativamente no acontecimento de falhas.