

# 01\_exploration

October 25, 2025

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
from pathlib import Path

# Load dataset
df = pd.read_csv("../data/leadership_toxicity_kpis.csv")

# Overview
display(df.head())

# Basic descriptive stats
display(df.describe()[["supervisor_toxicity", "motivation", "absenteeism_days",
    ↪ "performance_score"]])

# Correlation heatmap (simple version)
corr = ↵
    ↪ df[["supervisor_toxicity", "absenteeism_days", "performance_score", "motivation"]].
    ↪ corr()
print("\nCorrelation matrix (Pearson):")
print(corr.round(2))

# Scatterplot Toxicity vs Performance
plt.figure()
plt.scatter(df["supervisor_toxicity"], df["performance_score"], alpha=0.5)
plt.title("Toxicity vs Performance")
plt.xlabel("Supervisor Toxicity (0-10)")
plt.ylabel("Performance Score (0-100)")
plt.grid(True)
plt.show()

# Boxplot of Absenteeism by Team
plt.figure()
df.boxplot(column="absenteeism_days", by="team", grid=False)
plt.title("Absenteeism by Team")
plt.suptitle("") # remove automatic title
plt.ylabel("Days per Quarter")
plt.show()
```

	employee_id	team	gender	supervisor_toxicity	boundary_violations	\
0	1	Tech	M	1.63	1	
1	2	Sales	M	3.32	2	
2	3	Tech	M	2.95	2	
3	4	Support	M	5.16	1	
4	5	Ops	M	1.17	0	

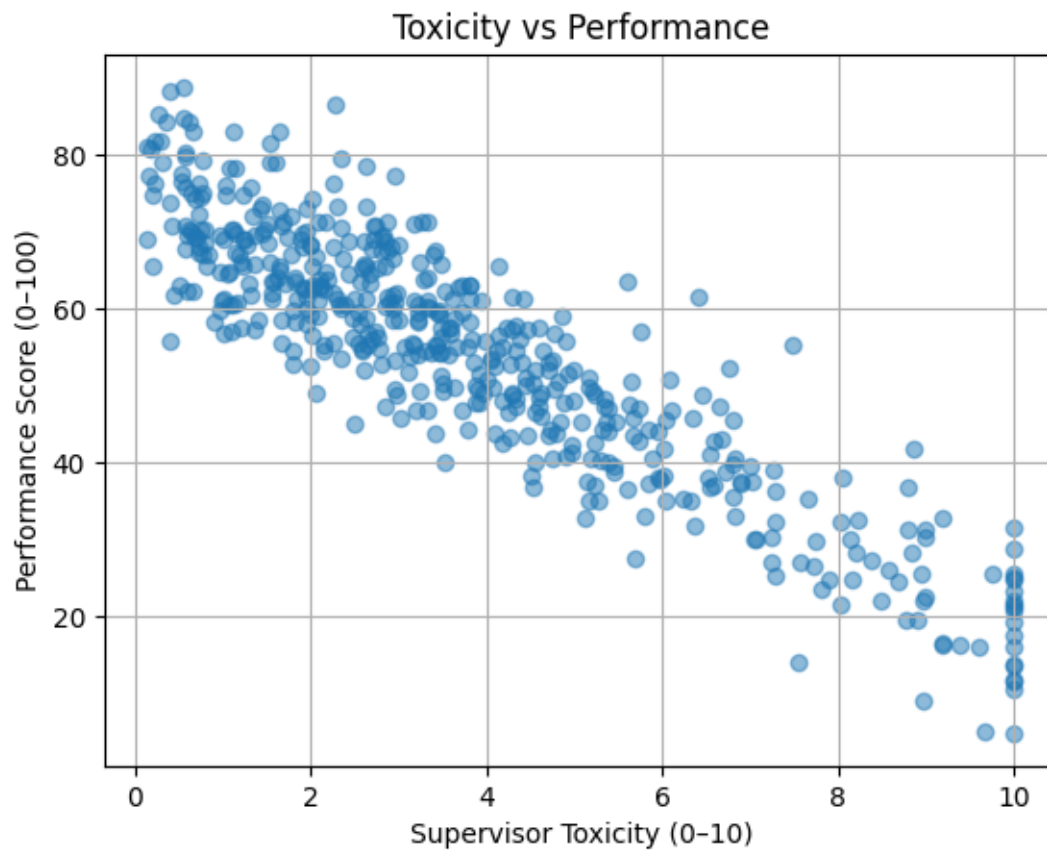
	boundary_blur_flag	motivation	absenteeism_days	performance_score
0	0	65.1	5.7	64.0
1	0	55.2	4.8	46.9
2	0	61.1	4.6	77.3
3	0	55.6	8.7	35.1
4	0	72.5	5.4	60.7

	supervisor_toxicity	motivation	absenteeism_days	performance_score
count	500.000000	500.000000	500.000000	500.000000
mean	3.826340	55.223800	6.67180	54.209000
std	2.557704	17.902543	3.21741	16.283665
min	0.150000	8.200000	0.30000	4.900000
25%	1.897500	43.400000	4.20000	44.400000
50%	3.320000	57.400000	6.20000	56.700000
75%	5.222500	68.700000	8.50000	65.800000
max	10.000000	96.600000	15.50000	88.700000

Correlation matrix (Pearson):

	supervisor_toxicity	absenteeism_days	performance_score	\
supervisor_toxicity	1.00	0.92	-0.90	
absenteeism_days	0.92	1.00	-0.85	
performance_score	-0.90	-0.85	1.00	
motivation	-0.90	-0.86	0.83	

	motivation
supervisor_toxicity	-0.90
absenteeism_days	-0.86
performance_score	0.83
motivation	1.00



<Figure size 640x480 with 0 Axes>

