

Correlation Matrix Heatmap

Dr Amr Amin

Ahmed Alqassas

Autumn 2025-2026

Introduction

Correlation matrices and heatmaps are powerful visual tools in data visualization, especially when **analyzing numeric features** in machine learning datasets.

They reveal hidden relationships, detect redundancy, and support feature selection.

What Is a Correlation Matrix?

A correlation matrix is a structured table showing the correlation coefficients between pairs of variables.

	Feature A	Feature B	Feature C
Feature A	1.00	0.78	-0.22
Feature B	0.78	1.00	-0.34
Feature C	-0.22	-0.34	1.00

Usage

1. Measures strength and direction of relationships.
2. Identifies multicollinearity (overlapping data).
3. Guides feature selection for machine learning.

Value Range	Meaning
0.90 to 1 or -0.90 to -1	Very strong
0.70 to 0.89 or -0.70 to -0.89	Strong
0.40 to 0.69 or -0.40 to -0.69	Moderate
0.10 to 0.39 or -0.10 to -0.39	Weak
0.00 to ± 0.09	No correlation

Heatmaps

A correlation heatmap uses a color spectrum to represent **correlation strength and direction.**

Brighter or darker colors often indicate stronger positive or negative relationships.

Benefit	Explanation
Quick pattern recognition	A glance reveals strong/weak relationships
Better dimensional understanding	Shows how features relate collectively
Outlier detection	Strange correlation blocks may indicate unusual behavior
Supports ML pipelines	Helps drop redundant variables

Practical Insights in Machine Learning

Feature Selection

- Remove features with very high correlation with each other ($\geq \pm 0.9$).
- Prevents overfitting and reduces computation.

Practical Insights in Machine Learning

Model Stability

- Algorithms like Linear Regression, Logistic Regression are sensitive to multicollinearity.
- Use heatmaps to avoid unstable coefficient estimation.

When Should You Avoid Heatmaps?

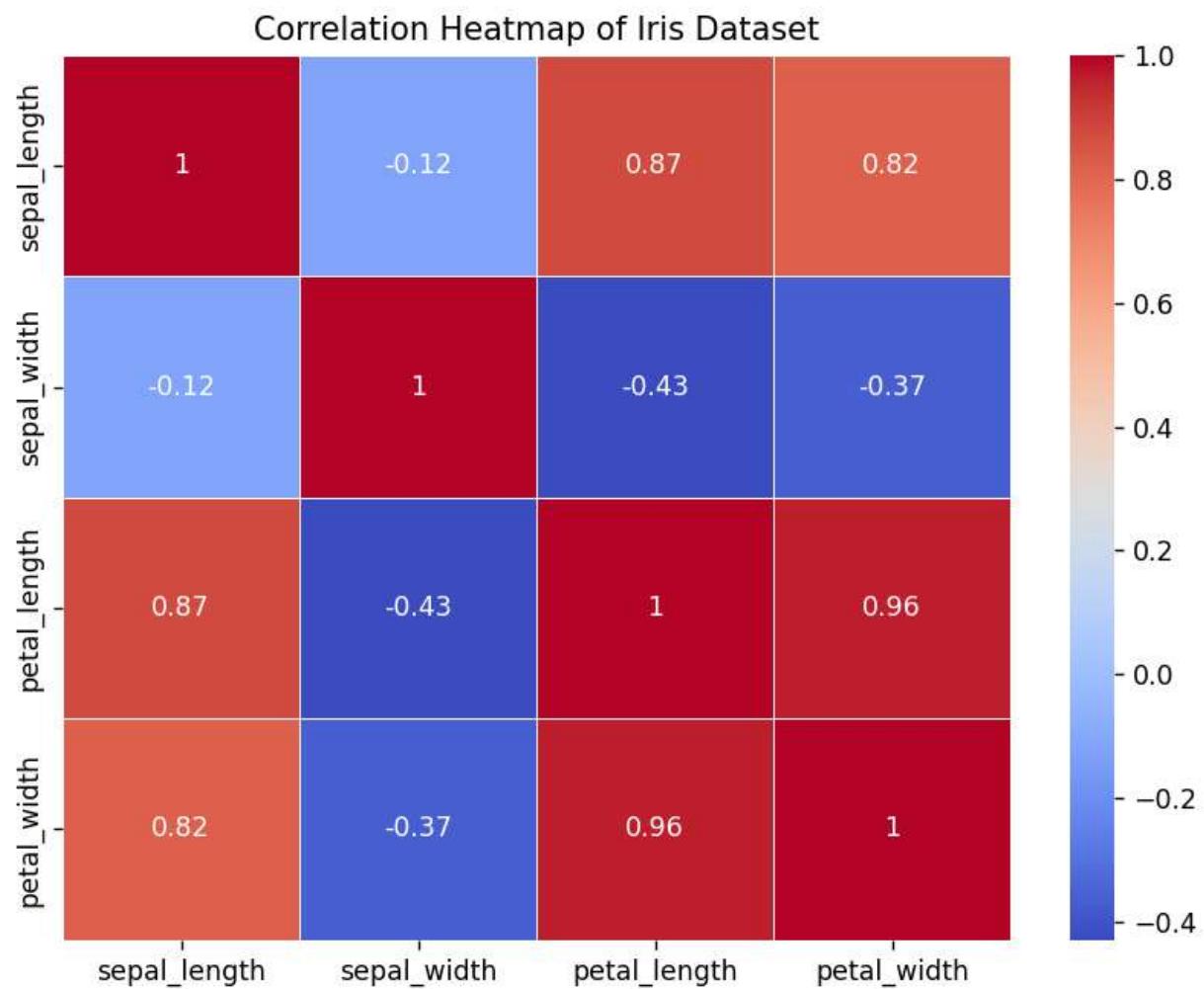
Situation	Reason
Many (>50) variables	Becomes unreadable
Non-numeric features	Correlation invalid without encoding
Very small datasets	Results are statistically weak

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Sample dataset
df = sns.load_dataset("iris")

# Compute correlation matrix
corr = df.corr(numeric_only=True)

# Plot heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(corr, annot=True, cmap="coolwarm", linewidths=0.5)
plt.title("Correlation Heatmap of Iris Dataset")
plt.show()
```



Summary

Aspect	Correlation Matrix	Heatmap
Format	Numeric table	Color-coded visualization
Best For	Quantitative analysis	Quick pattern recognition
Usage	Calculations, modeling	Presentations, exploration
Visibility	Harder with large sets	Intuitive at any scale