OWASP 2025
GLOBAL
AppSec

BARCELONA

CENTRE DE
CONVENCIONS
INTERNACIONAL DE
BARCELONA

OWASP 2025 GLOBAL AppSec | BARCELONA MAY 26-30

METRICS THAT MATTER - DRIVING APPSEC SUCCESS WITH DATA-DRIVEN INSIGHTS

DAVID ANDERSSON

Agenda 🗂️

What we'll cover today

📏 1. Why Metrics Matter
The origins of measurement in management and the pitfalls of poorly chosen metrics

🔍 2. Understanding AppSec Metrics
Categories, types and examples of meaningful appsec metrics

⚙️ 3. Gathering the Right Data
Explore data sources from SAST, DAST, Pentest, Bug Bounties, etc.

🧮 4. Calculating Risks & Scorecards
How to combine multiple metrics into one weighted score

📊 5. Visualising Metrics
Make metrics useful for engineers and executives with effective dashboards

🚀 6. Driving Program Impact
Make metrics actionable - training, coverage, and more

🗂️ Who am i?

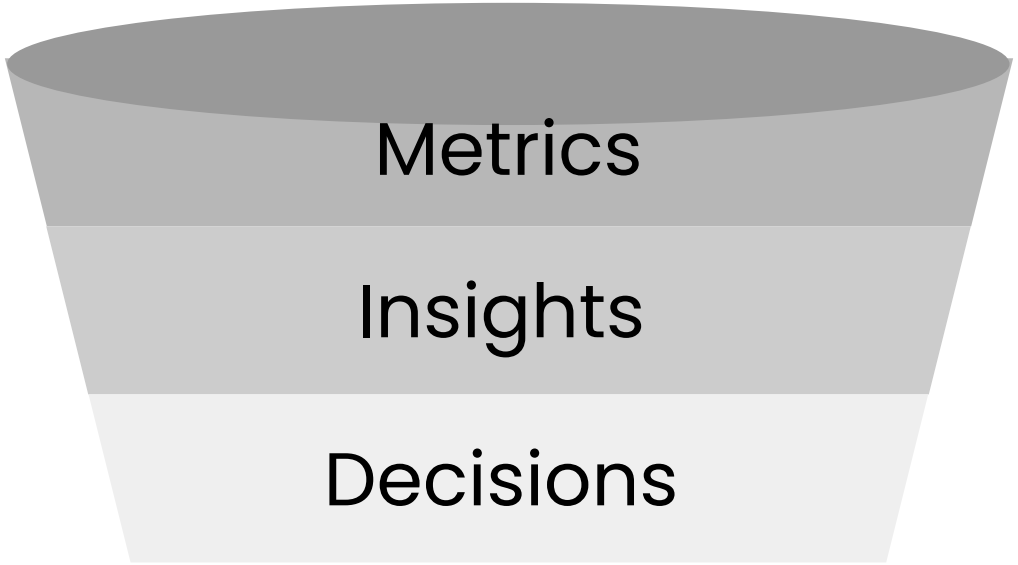
Short introduction

David Andersson

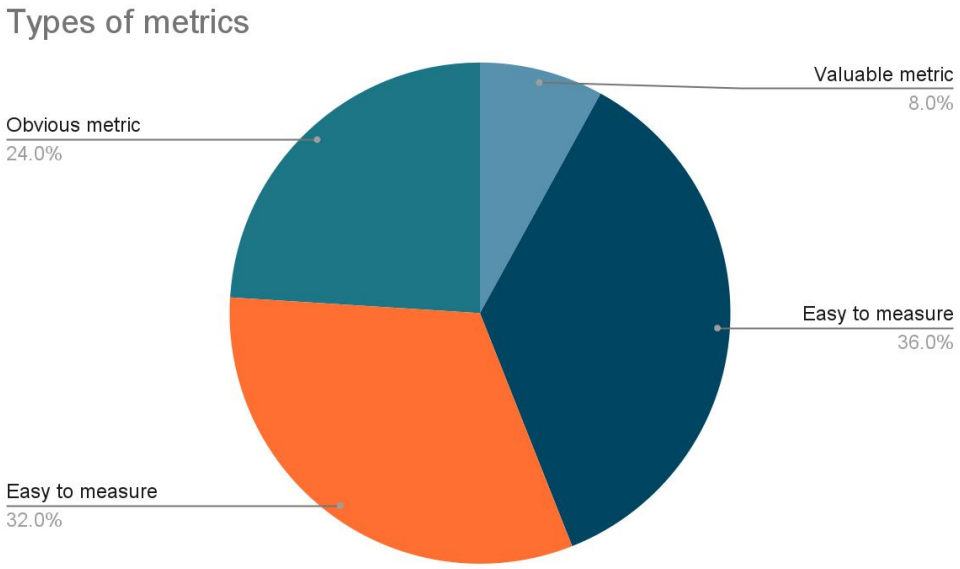
- Lives in Sweden
- Worked in the security industry since 2005
- AppSec leader since 2017
- Security Engineering @ Grafana Labs since 2024

Why Metrics Matter 📏

Linking metrics to security impact

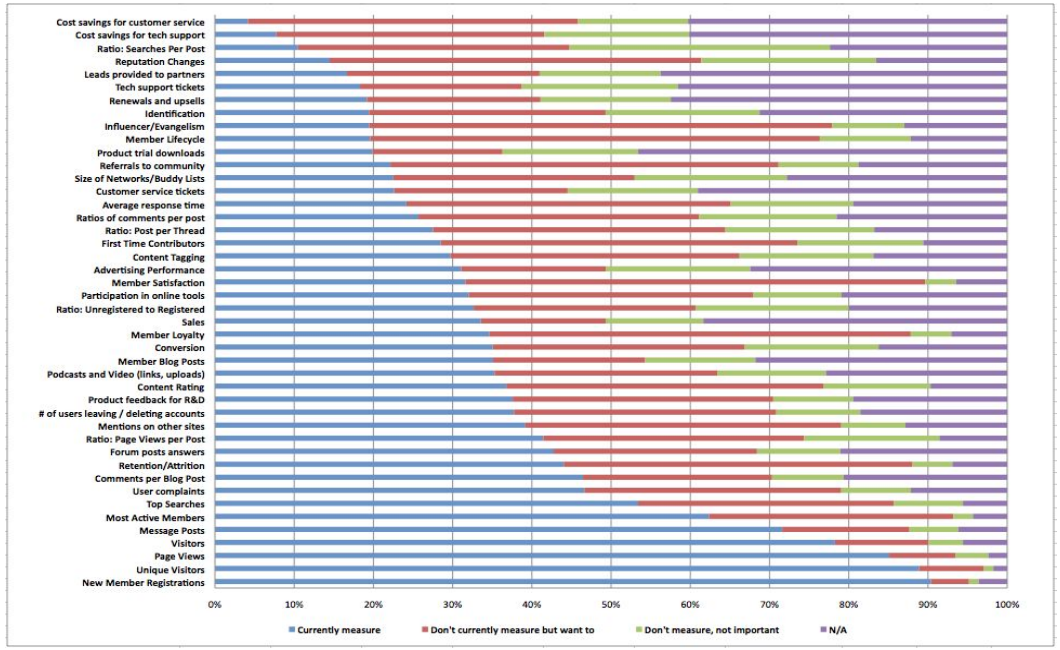



"Civil Engineering Training" by Georgia National Guard CC BY 2.0.


Types of metrics
Valuable metric
8.0%
Obvious metric
24.0%
Easy to measure
36.0%
Easy to measure
32.0%

The metric maze

When more data means less clarity



"*Online Community Metrics - OCRN Aug 2009*" by redplasticmonkey CC BY 2.0

Goodhart's Law

"When a measure becomes a target, it ceases to be a good measure"

Standards give us direction - not rules, but reason

"*Wollongong Breakwater Lighthouse*" by Bernard Spragg CC0 1.0

Types of AppSec Metrics 📐

Operational vs. Technical / Quantitative vs. Qualitative

	Operational	Technical
Quantitative	Count or percentage of applications have completed penetration testing in the last “n” months	Count or percentage of vulnerabilities by weakness
Qualitative	Maturity of AppSec practices	Security Culture in Engineering Teams

What Makes a Good Metric? 🧩

Five traits for trust and actionability

✅ **Relevant**

⏱️ **Timely**

📏 **Measurable**

🔁 **Repeatable**

🧠 **Understandable**

From Trait to Metric 🎯

Examples of metrics that get it right

MTTD/MTTR

Open Vulnerabilities per Severity

Number of vulnerable dependencies

Average bug bounty reports per day

Number of security trained engineers

Sources of truth

Where your data comes from

Burp Suite

Semgrep

INTIGRITI

hackerone

Cobalt

docebo®

Acunetix

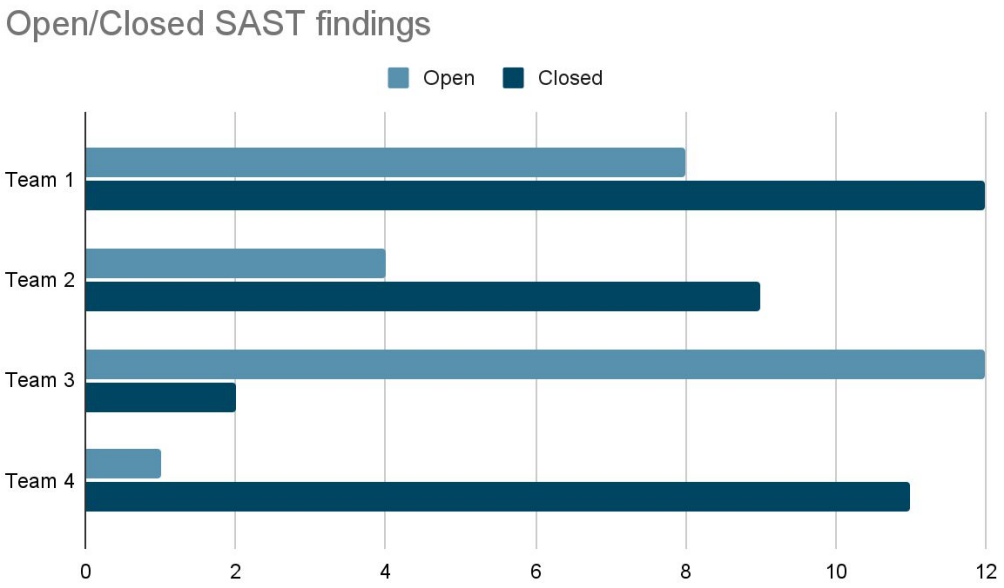
Practical Metrics 📊

Choosing the right signals

Data Source	Metric	What it tells us
LMS	Training Coverage	Developer Maturity
Ticketing System	MTTR	Responsiveness
Scanners	MTTD	Detection capabilities
Engagements	Last Date	How to plan engagements

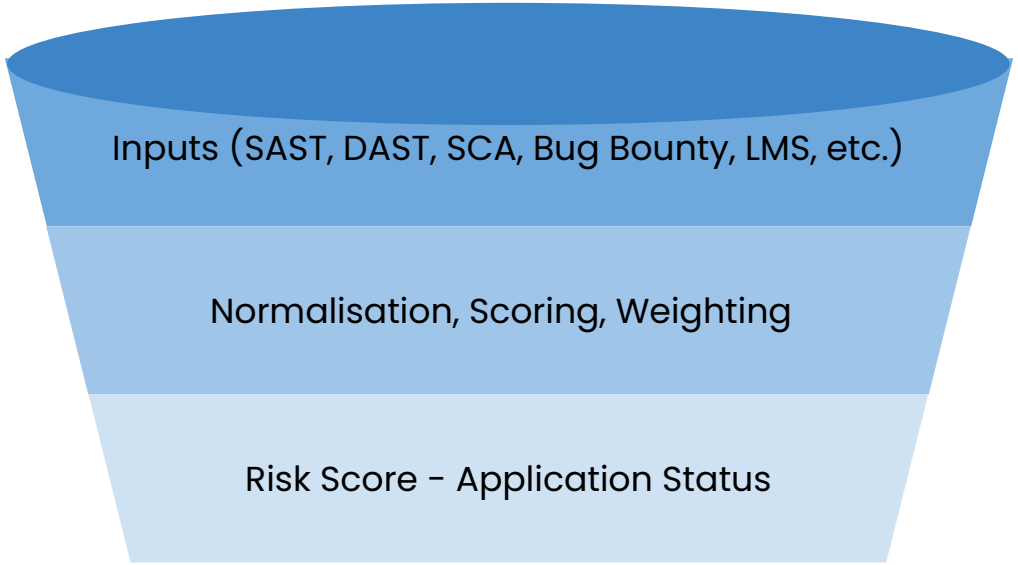
Getting Insights from Details 🔍

Interpreting MTTR/MTTD & open vs. closed trends



From Metrics to Meaning 🧩

Building the insight pipeline



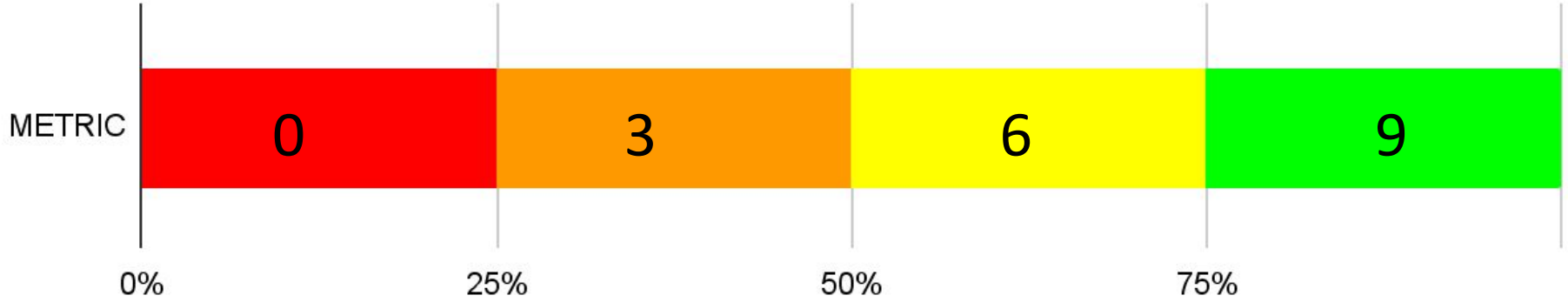
Normalizing the Noise 🔢

Turning messy raw signal into comparable signal

Raw Metric	Normalized Metric
SAST findings	SAST Severity Score per KLoC
SCA findings	SCA Severity Score per KLoC
DAST findings	DAST Severity Score per scanned URL
Days to Close High Bug Bounty Issues	Bug Bounty Responsiveness
Recency of Last Pentest	Pentest Freshness Score
Count of Open Critical Vulns	Open Critical Risk Density

Normalizing the Noise 🔢

Defining bands



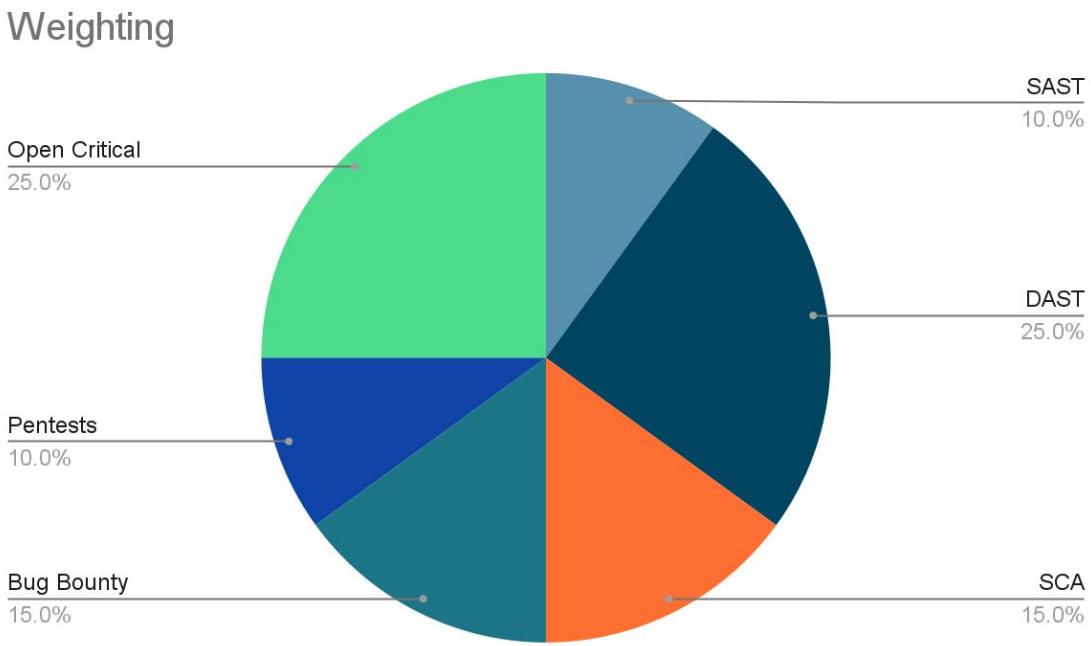

```python
# Normalisation helpers
def score_sast(row):
    raw_score = (100 * row["critical"] + 50 * row["high"] +
                 10 * row["medium"] + row["low"]) / max(row["loc"], 1000) * 1000
    if raw_score <= 10:
        return 9
    elif 10 < raw_score <= 25:
        return 6
    elif 25 < raw_score <= 50:
        return 3
    return 0
```

```python
def score_bug_bounty(row):
    raw_score = row["mttr_high"]

    # Normalized Bug Bounty score
    if raw_score <= 7:
        return 9
    elif 7 < raw_score <= 14:
        return 6
    elif 14 < raw_score <= 30:
        return 3
    elif raw_score > 30:
        return 0
```

Scoring Strategy Overview ⚖️

Weighting what matters most



Risk Score Formula 🧮

Combining normalized scores into a unified rating

$$Risk = \sum(Normalized_i \times Weight_i)$$

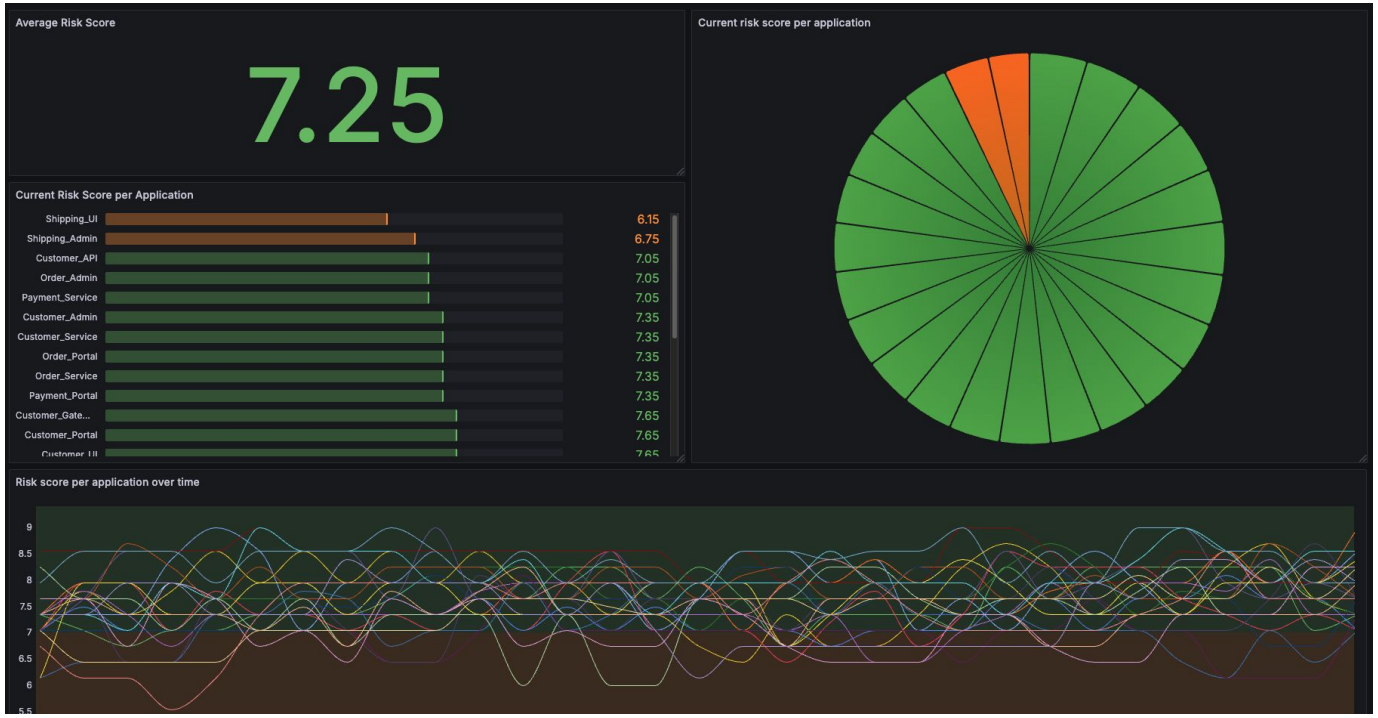
Scorecard Example 🔢

Sample risk calculation for one application

Source	Metric	Normalised score	Weight factor	Score	Status
SAST	SAST Severity Score per KLoC	6	10%	0.6	🟠 Amber
SCA	SCA Severity Score per KLoC	3	15%	0.45	🔴 Red
DAST	DAST Severity Score per scanned URL	6	25%	1.5	🟠 Amber
Bug Bounty	Bug Bounty Responsiveness	9	15%	1.35	🟢 Green
Penetration Test	Pentest Freshness Score	0	10%	0	🔴 Red
Any Source	Open Critical Risk Density	6	25%	1.5	🟠 Amber
Total				**5.4**	🟠 **Amber**

Visualizing Risk 📈

What could it look like?

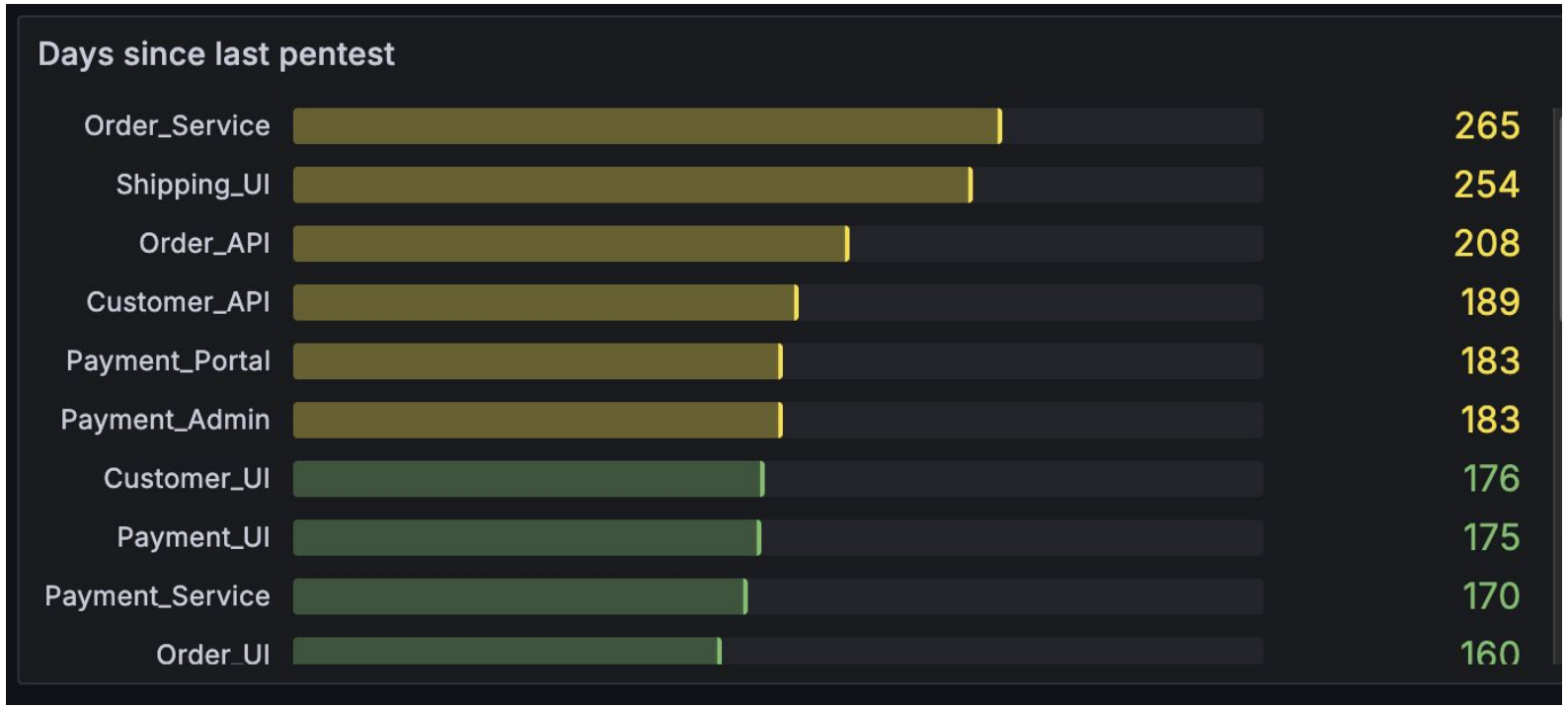


Mapping Metrics to Actions

CWE mapping to targeted training

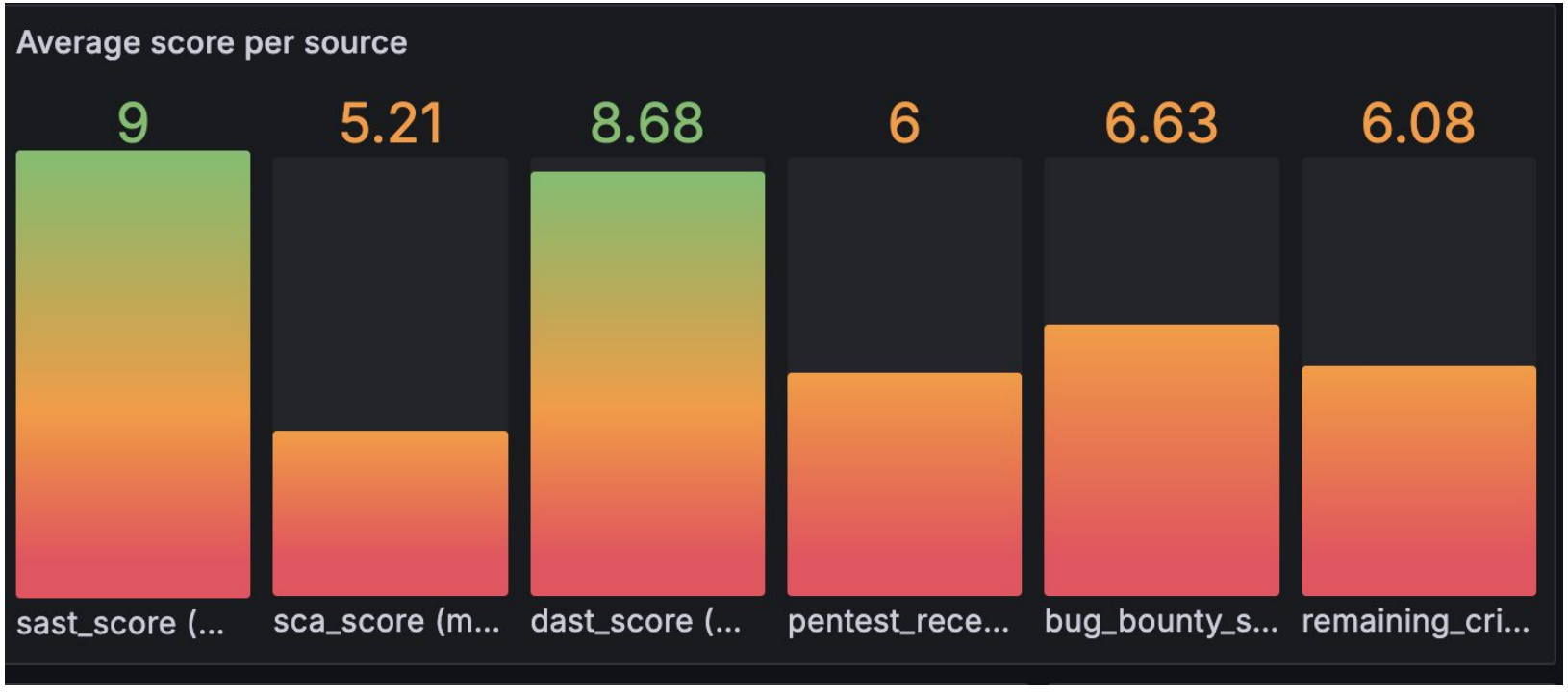
Mapping Metrics to Actions

Help in planning



Mapping Metrics to Actions

Signal per score



Metrics as Dialogue 💬

Start conversations, not end them

“Metrics should start conversations, not end them”

Recap & Key Takeaways 📌

What to remember

Measure with purpose

Normalize intelligently

Make metrics actionable

Thank You ✨

Stay in touch!

davidandersson-se

davidandersson.se

Info (a) davidandersson.se


OWASP 2025
GLOBAL
AppSec
BARCELONA
THANK YOU!