What Makes InnerSource Communities Healthy?

Uncovering Collaboration Patterns In 1,000+ GitHub Repositories

Ana Gamito · **AllWomen** Data Analytics Bootcamp · Nov 2025

The Challenge

Teams rely on vanity metrics like **stars** and **forks**, missing deeper signals of responsiveness and sustainability.

But are these the **right signals**? Do they predict community success?

Key Questions

→ How do we know if a community is truly healthy?

→ What metrics matter most?

→ How do we scale collaboration sustainably?

Dataset & Methodology

1,052 → **972**

Repositories — Valid Observations

Stars & Forks
Social signals of adoption

Issues & PRs
Indicators of activity

Contributors
Measure of collaboration

Language
Technical context

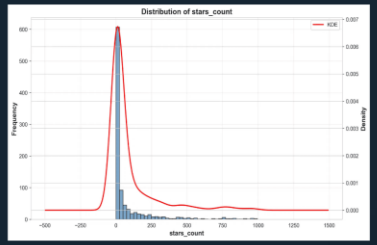
Data Quality
Cleaned & validated

Rigor
Removed 80 duplicates

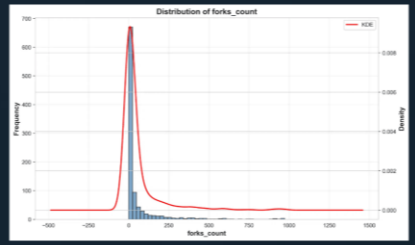
After rigorous cleaning—removing duplicates and handling missing data—we achieved **972 valid, high-quality observations** for comprehensive analysis.

Metrics Overview: The Power Law

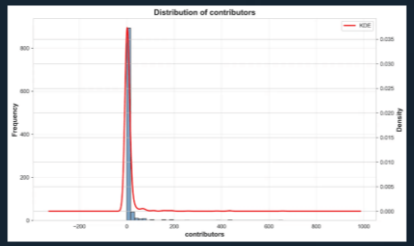
Stars Distribution



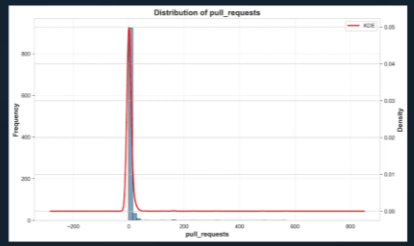
Forks Distribution



Contributors Distribution



Pull Requests Distribution



Most repositories have **modest engagement**, but a few "superstars" skew the average.

- **Right-skewed distribution** is typical in open source—most projects have concentrated **core teams** of **5-20 contributors.**
- **Scale doesn't predict health.** A 100-star (*small*) project with 3 dedicated contributors can be healthier than a 10,000-star abandoned (*large*) project.
- **Understanding this baseline matters** because it tells us we need normalized, composite metrics to compare across projects fairly.

Understanding Responsiveness

Definition

Responsiveness is the speed and effectiveness with which a community responds to and manages interactions and contributions. It measures how actively a project engages with issues, pull requests, and community members.

Issues Count

Number of open issues being managed and discussed.

- Reflects active problem-solving engagement
- Higher activity = more community involvement

Pull Requests

Number of contributions being reviewed and integrated.

- Indicates review speed and integration capacity
- Shows willingness to accept community contributions

✓ High Responsiveness

- ✓ Issues actively discussed and resolved
- ✓ PRs reviewed and merged in reasonable time
- ✓ Constant interaction between maintainers & contributors
- ✓ Community feels "alive" and engaged

⚠ Low Responsiveness

- ✗ Issues left unanswered for extended periods
- ✗ Pull requests accumulate without review
- ✗ Minimal feedback or interaction
- ✗ Project appears abandoned or inactive

💡 Why This Matters for Our Analysis

Responsiveness is a key indicator of community health. In Q2, we investigate whether more contributors automatically lead to higher responsiveness, or if team efficiency and processes matter more. This metric helps us understand what truly drives sustainable, healthy InnerSource communities.

Feature Engineering: Derived Metrics

Health Score

popularity adoption maintenance

$$\frac{\text{(Stars + Forks) - Open Issues}}{\text{Contributors}}$$

team size

Measures: Popularity and maintenance relative to team size.

Interpretation

High: Popular, well-maintained, active community

Low: Low adoption OR high unresolved issues

Activity Ratio

Work merged

$$\frac{\text{Pull Requests}}{\text{Open Issues}}$$

Work requested

Measures: Community responsiveness and workflow efficiency.

Interpretation

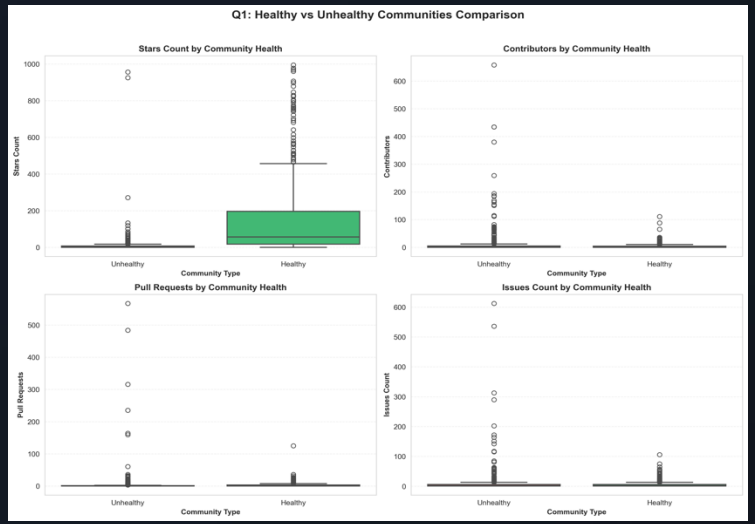
> 1: More PRs merged than issues opened = responsive

< 1: Backlog accumulation = less responsive

*These formulas are inspired by **CHAOSS and GitHub Insights** industry standards —grounded in how professionals measure community health.*

Q1: What Distinguishes Healthy Communities?

Four distinct community profiles based on Health Score & Activity Ratio



The Goal

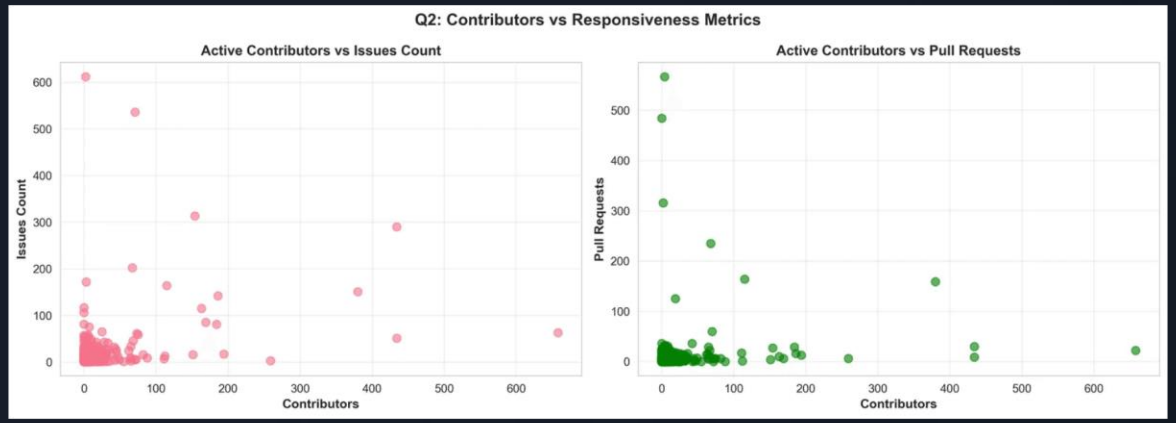
Move projects toward **Thriving**: popular, responsive, and well-maintained communities.

Watch Out For

Projects trending toward **At Risk** show declining responsiveness and accumulating backlog. Most projects face sustainability challenges—only 18% are thriving.

Q2: Contributors & Responsiveness

Does team size predict community responsiveness?



Key Takeaway:

A team of 5 organized contributors can outpace a team of 20 with poor communication. Hiring more developers won't automatically solve responsiveness problems—you need clarity on roles, communication channels, and merge processes.

Key Question

The Finding

Moderate positive correlation exists between contributors and issues. More people often bring more issues—that makes sense.

The Critical Insight

The correlation is **not strong enough to be predictive**. Responsiveness depends on **team efficiency and processes**, not just headcount.

Correlation Analysis: Variable Relationships

Strong Correlations

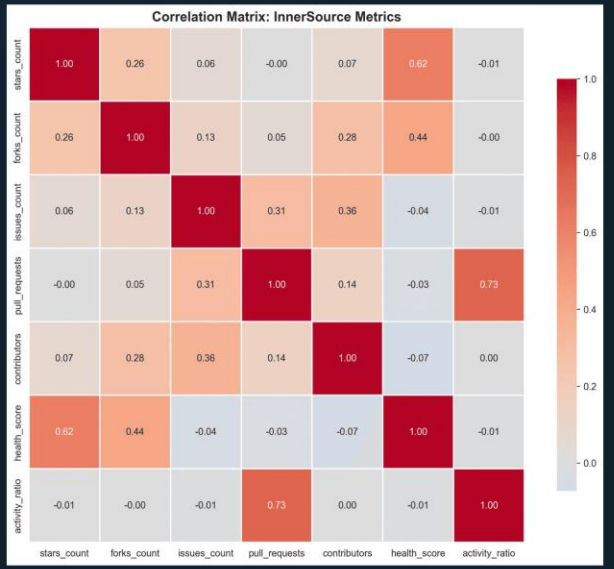
- **Stars ↔ Forks**: High correlation (popularity metrics move together)
- **Contributors ↔ Issues**: Moderate correlation (more people = more reported issues)
- **Contributors ↔ PRs**: Moderate correlation (more contributors = more work)

Weak Correlations

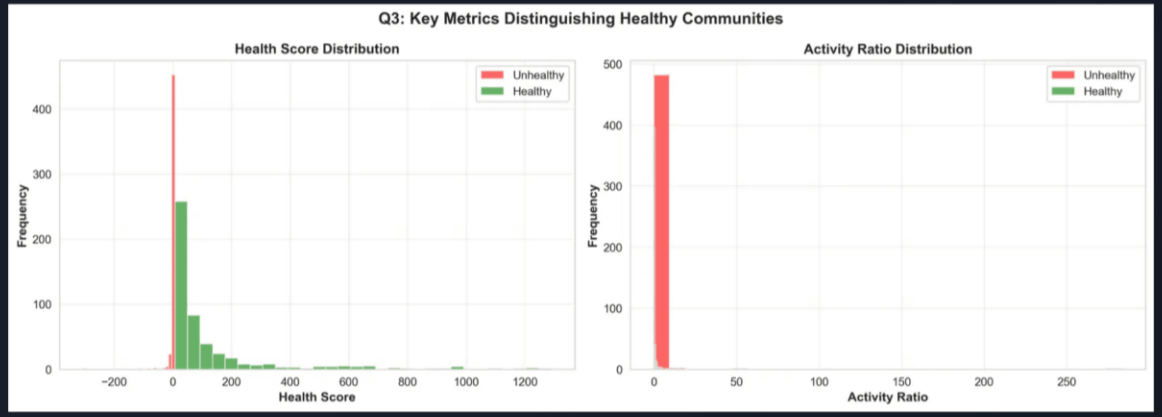
- **Stars ↔ Contributors**: Weak correlation (popularity ≠ team size)
- **Stars ↔ Issues**: Weak correlation (popularity ≠ maintenance burden)

What This Means

Variables don't move in lockstep. This is why **composite metrics** (*Health Score, Activity Ratio*) are essential—they capture the nuanced relationships between different aspects of community health.



Q3: Actionable Patterns



Early Warning Signals & Intervention

→ **Activity Ratio dropping:** More issues than PRs being merged

→ **Issues-per-contributor rising:** Team is stretched thin

→ **Intervention:** Clarify contribution guidelines, establish SLAs

Healthy Communities

- Higher Health Scores and Activity Ratios
- More distributed contributors (scaled beyond single-person teams)
- Regular Pull Request activity (consistent workflow)
- Issues resolved faster than they accumulate

At-Risk Communities

- Backlog accumulation (issues grow faster than PRs)
- Contributor burnout signals (issues per contributor spikes)
- Long periods of inactivity
- Declining responsiveness and engagement

Key Insights from the Data

1 Health ≠ Popularity

High stars don't guarantee sustainability. A 10,000-star project with 2 contributors is fragile. A 100-star project with 20 engaged contributors is sustainable.

2 Responsiveness is Critical

Communities that stay healthy are the ones where issues get addressed, pull requests get reviewed, and contributors feel heard. Responsiveness trumps size.

3 Collaboration is a Practice

Processes, clarity, and communication scale better than headcount. A well-organized 5-person team outpaces a chaotic 20-person team.

4 Metrics Enable Action

By monitoring health scores and activity ratios, you get visibility and early warning signals. You can intervene before communities collapse.

These insights are backed by **data from 972 real repositories**. They're not opinions—they're patterns in how successful communities actually work.

Recommendations

For organizations building or scaling InnerSource communities

1. Establish Baseline Metrics

Start measuring your InnerSource communities today.

- Calculate Health Score & Activity Ratio
- Segment projects into quadrants
- Document baseline for future comparison

2. Monitor & Track Progress

Make metrics part of your regular cadence.

- Review quarterly (or monthly for at-risk projects)
- Track trends and early warning signals
- Share results with project leads

3. Invest in Practices, Not Just Headcount

Build processes that scale responsiveness.

- Establish code review SLAs
- Create contributor onboarding guides
- Improve communication channels

4. Support Communities Strategically

Tailor interventions to community health stage.

- Celebrate thriving communities
- Provide resources to emerging projects
- Intervene early in at-risk projects

Implementation Timeline

Month 1
Set up metrics collection and establish baseline for all projects

Month 2–3
Implement process improvements and communication channels

Month 4+
Monitor quarterly, adjust strategies, celebrate progress

> **The best way to predict the future is to build it together. You can only build together if you have the visibility and practices to make it work.**

This analysis shows that **healthy InnerSource communities aren't accidents**—they're the result of intentional measurement, clear processes, and genuine commitment to collaboration.
By understanding the patterns in 972 real repositories, we can build communities that are not just popular, but **sustainable, responsive, and truly collaborative**.

Thank you

Questions?

I'm happy to dive deeper into any part of the analysis. The full notebook, visualizations, and methodology are available for review. Let's discuss how to apply these insights to **your** InnerSource communities.

Ana Gamito · **AllWomen** Data Analytics Bootcamp · Nov 2025