

Análise e evolução de sistemas open-source baseados em microserviços: Investigando issues e commits

João Vitor Staub Castanho

Sergio Alvarez da Silva Junior

Filtrando o dataset

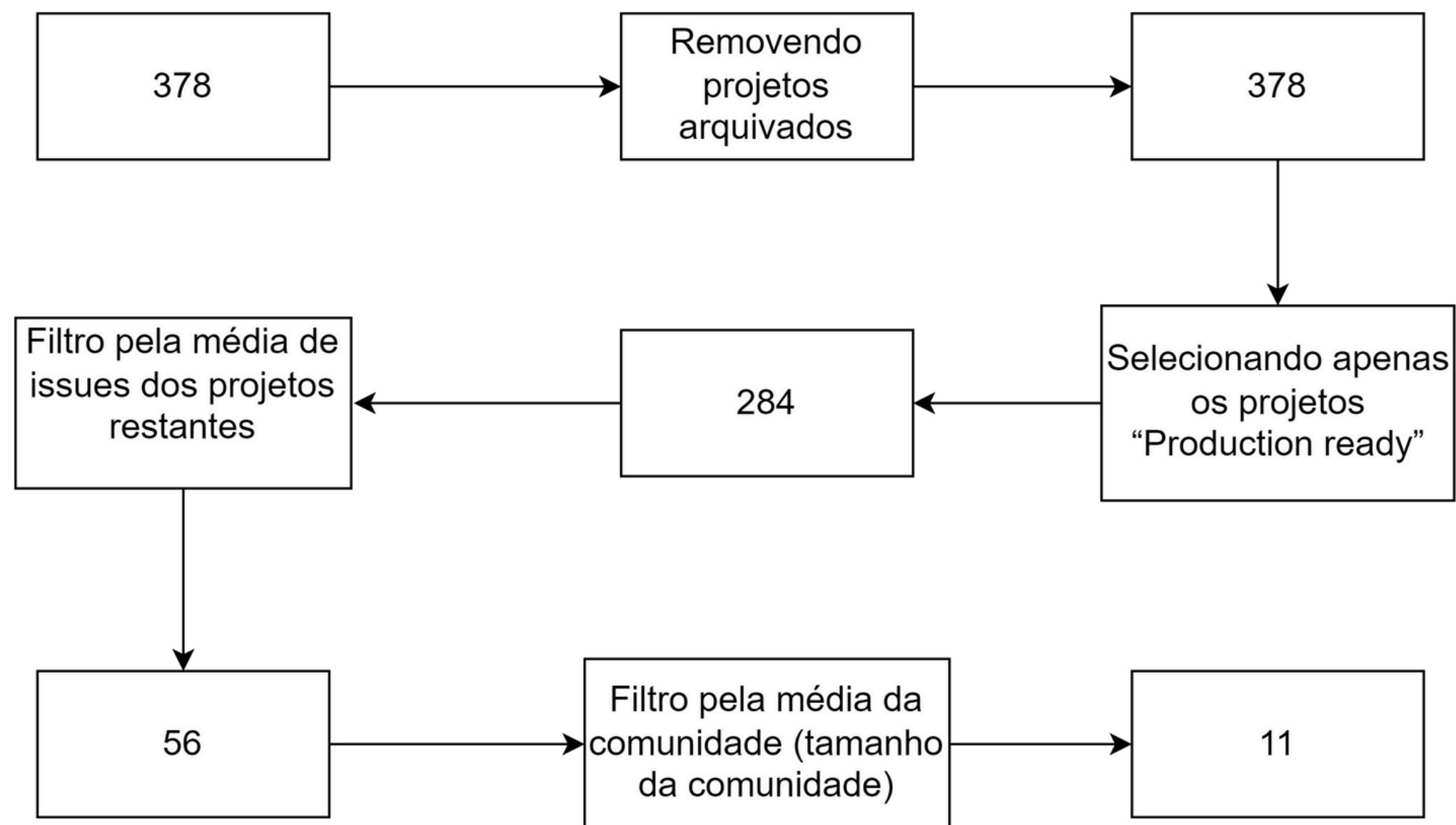


Figura 1 - Fonte: autoria própria.

Resultado do scraping

1. github.com/mirumee/saleor
2. github.com/cfpb/hmda-platform
3. github.com/golastmile/rasa_nlu
4. github.com/samvera/hyrax
5. github.com/mesosphere/marathon
6. github.com/openstates/openstates-scrapers
7. github.com/letsencrypt/boulder
8. github.com/deviantony/docker-elk
9. github.com/akka/alpakka
10. github.com/andryyy/mailcow-dockerized
11. github.com/nextcloud/docker



Resultado do scraping

- Commits:
 - Média: 11556.181818181818
 - Mediana: 6969
 - Desvio Padrão: 10105.038761114989
- Issues:
 - Média: 6399.181818181818
 - Mediana: 5670
 - Desvio Padrão: 4210.69706386441

Linguagem utilizadas dos repositórios.

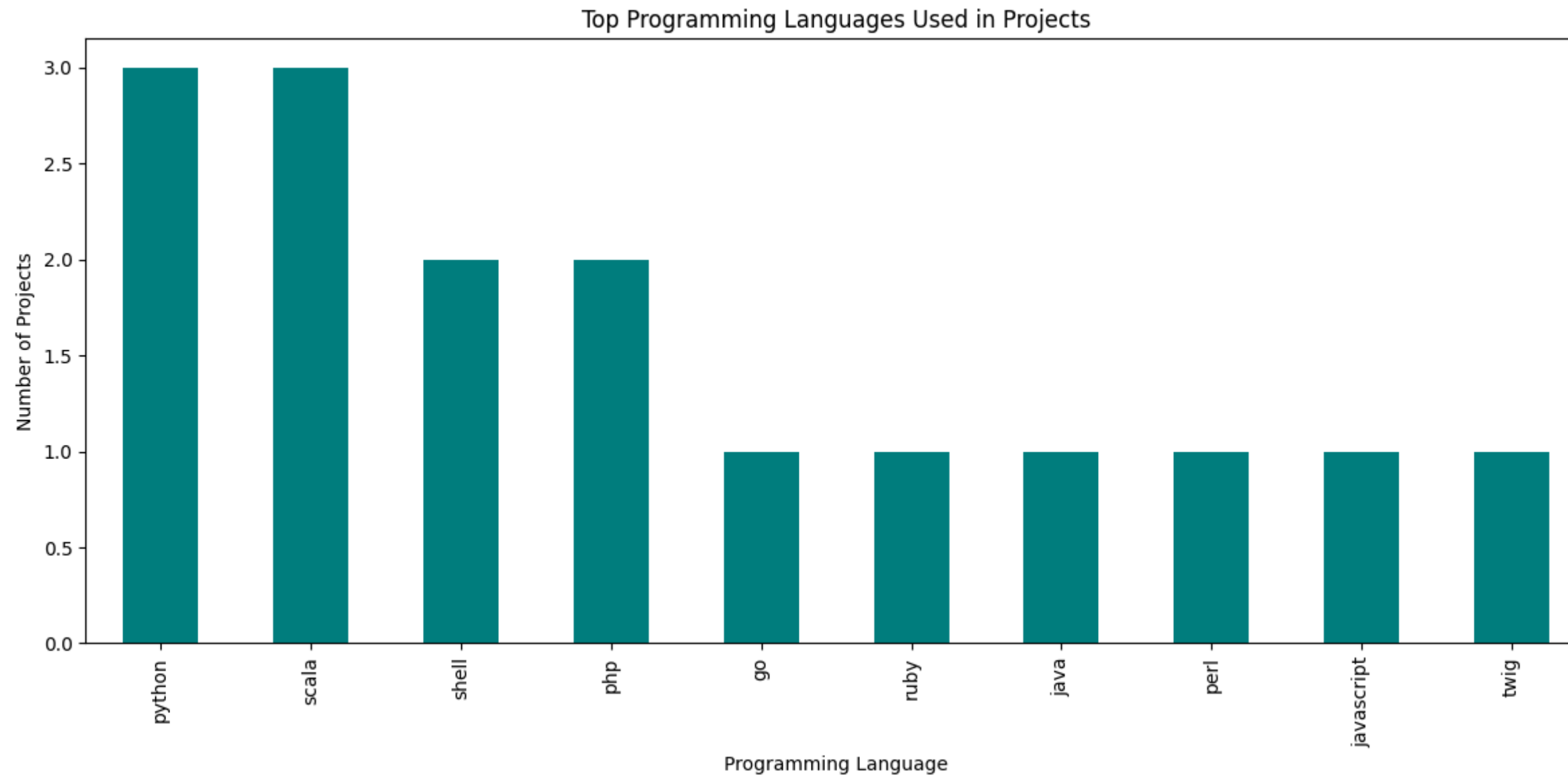


Figura 2 - Fonte: autoria própria.

Research Questions

RQ1. Qual é o status atual e a evolução dos repositórios de microserviços abertos?

Este conjunto de perguntas explora as tendências gerais e padrões na adoção, desenvolvimento e engajamento da comunidade de microserviços.

RQ1.1 - Ao analisar o número de commits, o engajamento das comunidades em repositórios Open Source está aumentando?

RQ1.2: Existem autores proeminentes na comunidade de microserviços?

RQ1.3: Os autores contribuem para diversos projetos?



Research Questions

RQ2. Como os dados das PRs fechadas podem ser utilizados para avaliar o tempo de manutenção de repositórios abertos de microsserviços

Este conjunto de perguntas foca em entender os esforços de desenvolvimento e manutenção em andamento em repositórios de microsserviços com base no histórico de issues fechadas.

RQ2.1. Qual é o tempo médio de resolução das issues no repositório?

RQ2.2. Qual o impacto do tamanho do pull request no tempo de resolução de issues?

RQ2.3. Qual o impacto do tamanho da comunidade no tempo de resolução de issues?

RQ2.4. Qual é a proporção de issues que tem PRs (pull requests) no repositório?

RQ2.5. Como o o tempo de resolução das PRs está relacionado ao tamanho do repositório?

Divisão dos repositórios

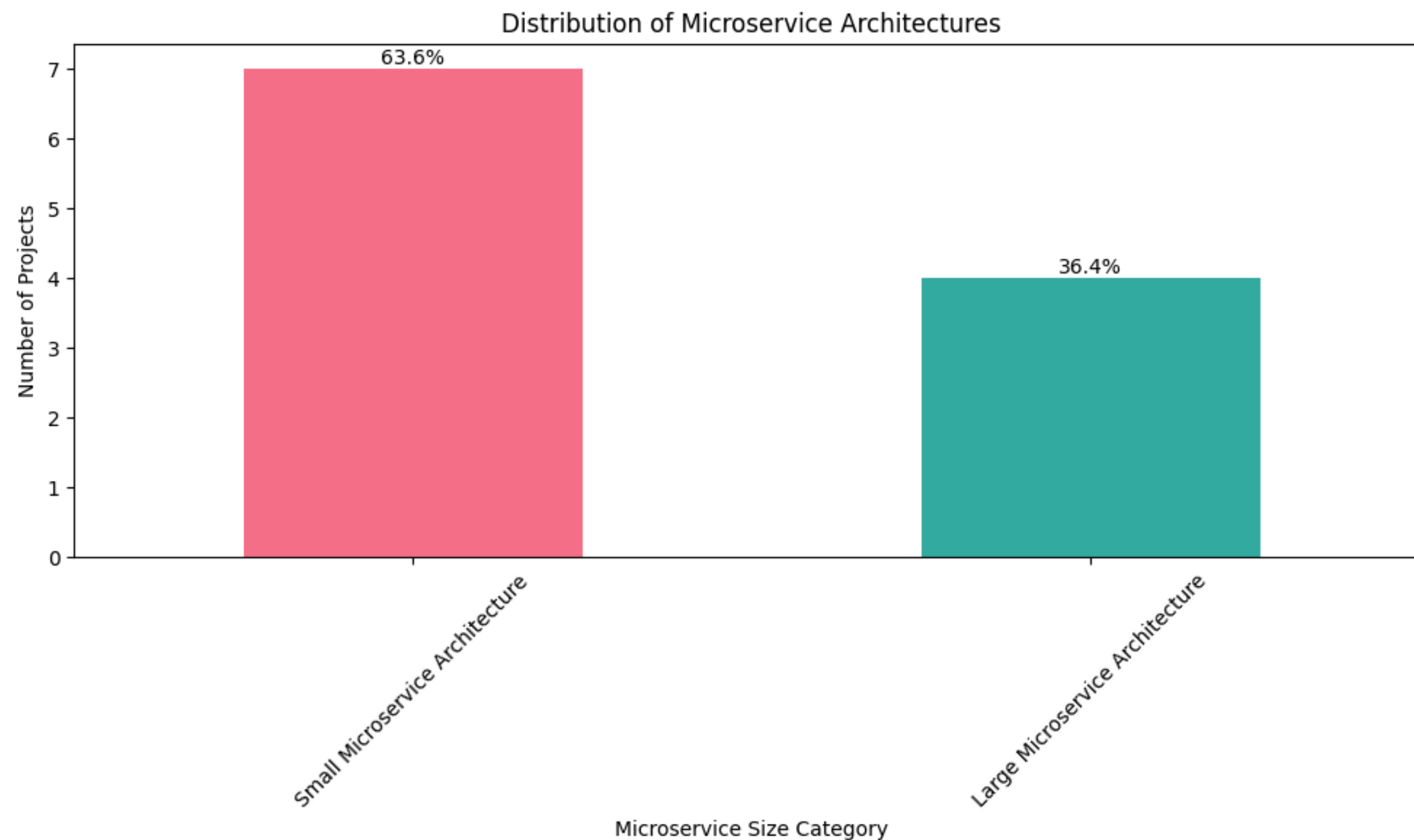


Figura 6 - Fonte: autoria própria.

RQ1.1 - Ao analisar o número de commits, o engajamento das comunidades em repositórios Open Source está aumentando?

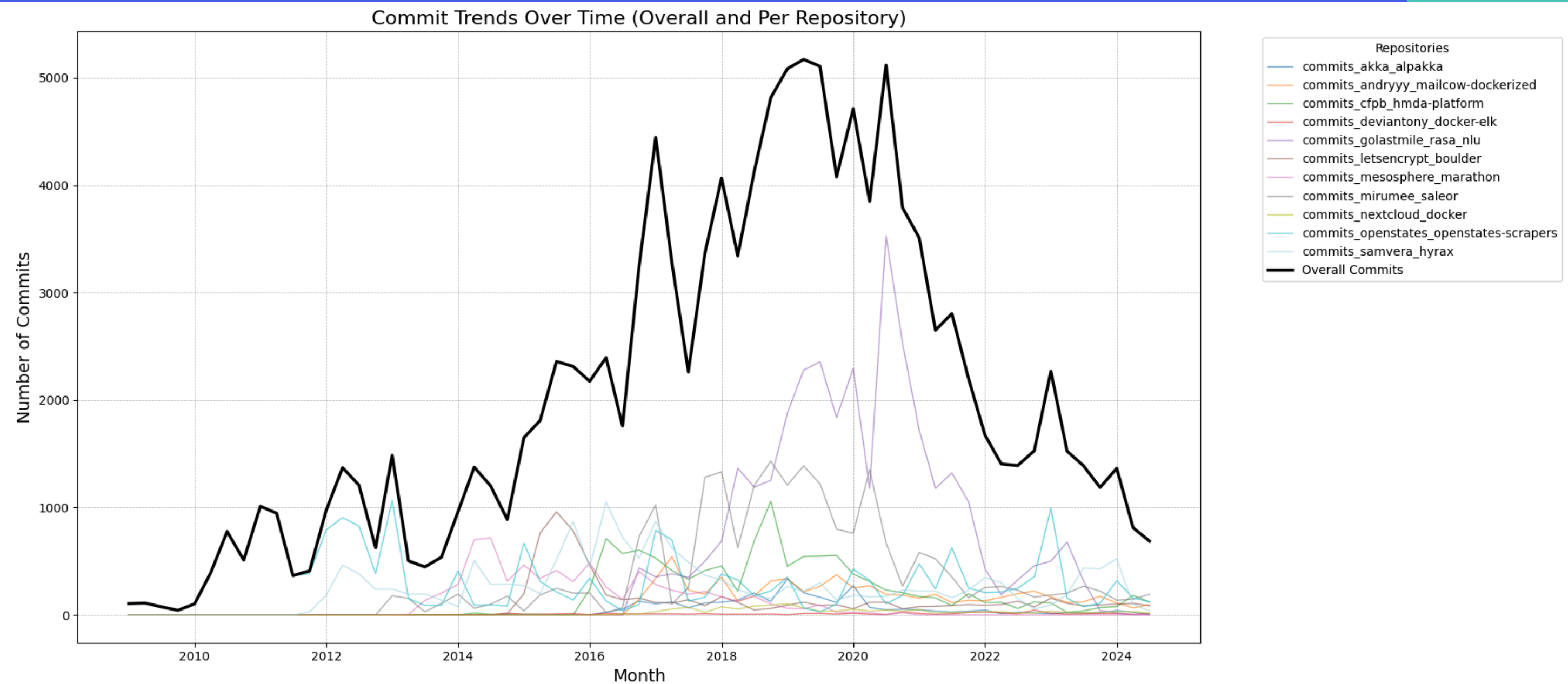


Figura 3 - Fonte: autoria própria.

RQ1.2: Existem autores proeminentes na comunidade de microserviços?

- Diversos projetos apresentam autores proeminentes que contém a maior parte dos commits realizados no projeto.

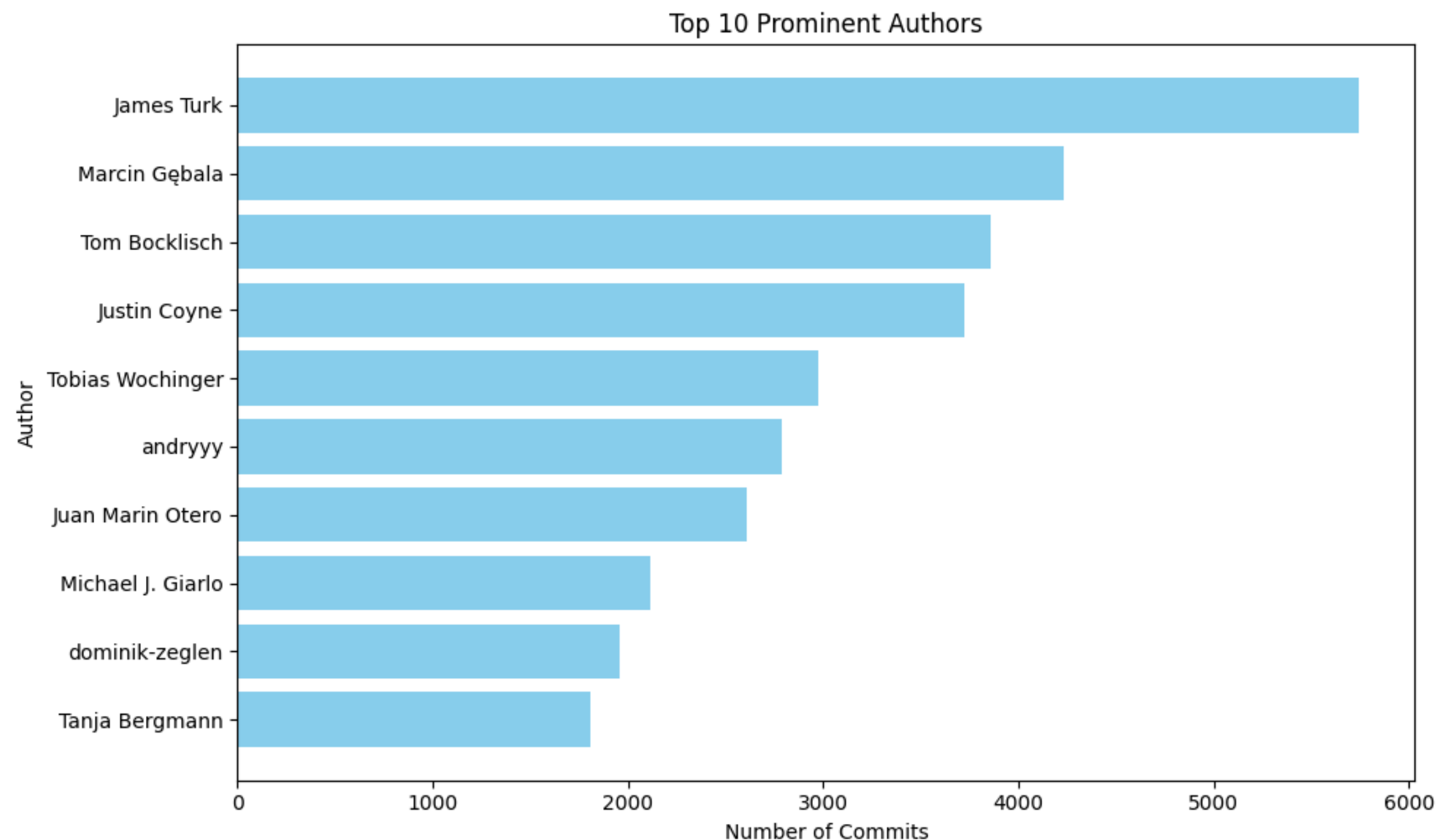


Figura 4 - Fonte: autoria própria.

RQ1.3: Os autores contribuem para diversos projetos?

- Dentre os repositórios mineirados, a minoria dos autores contribui para mais de um projeto.

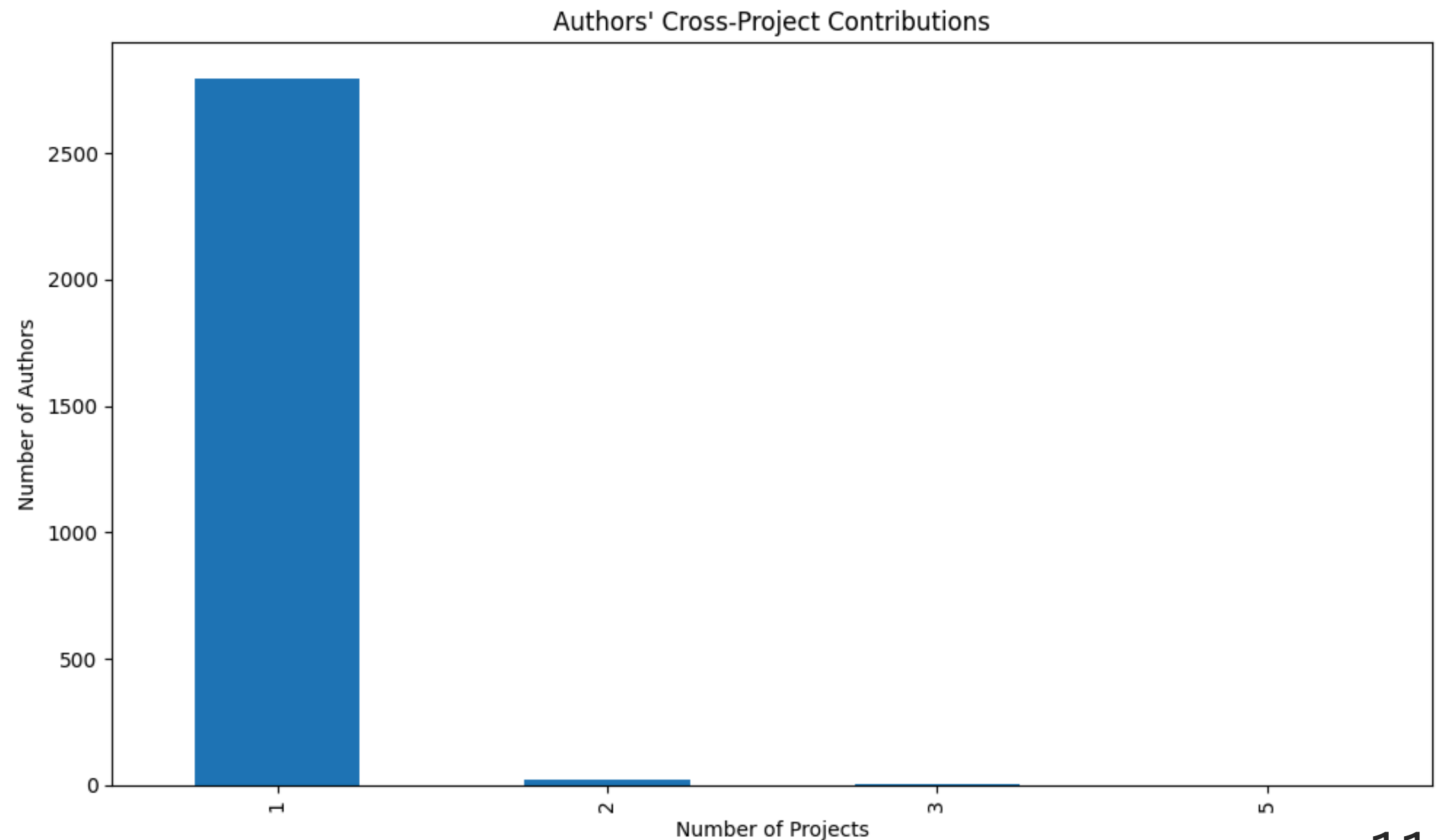


Figura 5 - Fonte: autoria própria.

RQ2.1. Qual é o tempo médio de resolução das issues no repositório?

- O tempo médio de resolução de issues é de 10,66 dias.
- Quando os outliers são removidos, a média diminui para 3,47 dias.

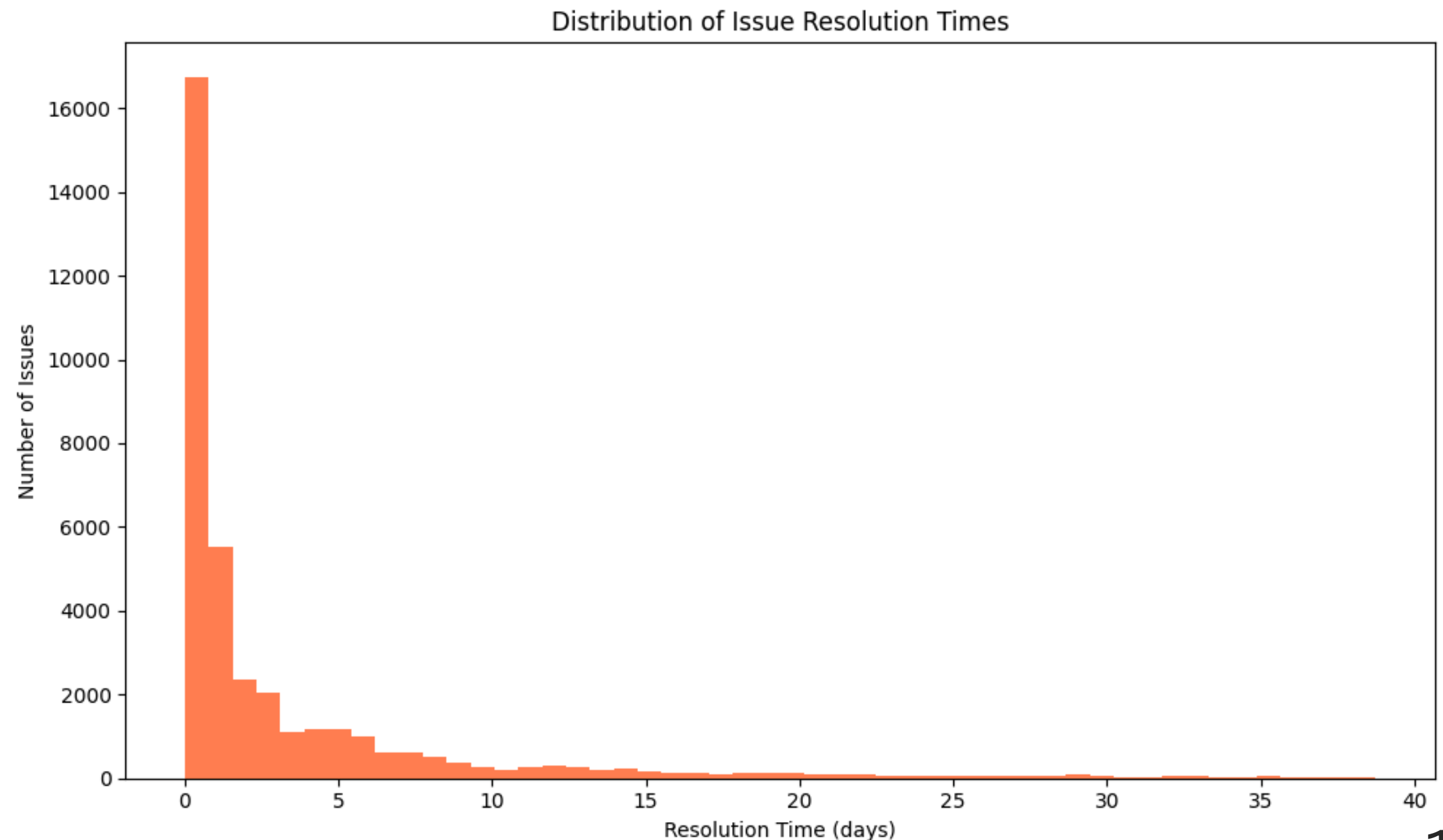


Figura 7 - Fonte: autoria própria.

RQ2.2. Qual o impacto do tamanho do pull request no tempo de resolução de issues?

- Não é possível determinar uma correlação a partir do gráfico.
- Análise de Pearson apresentou valor de 0.16

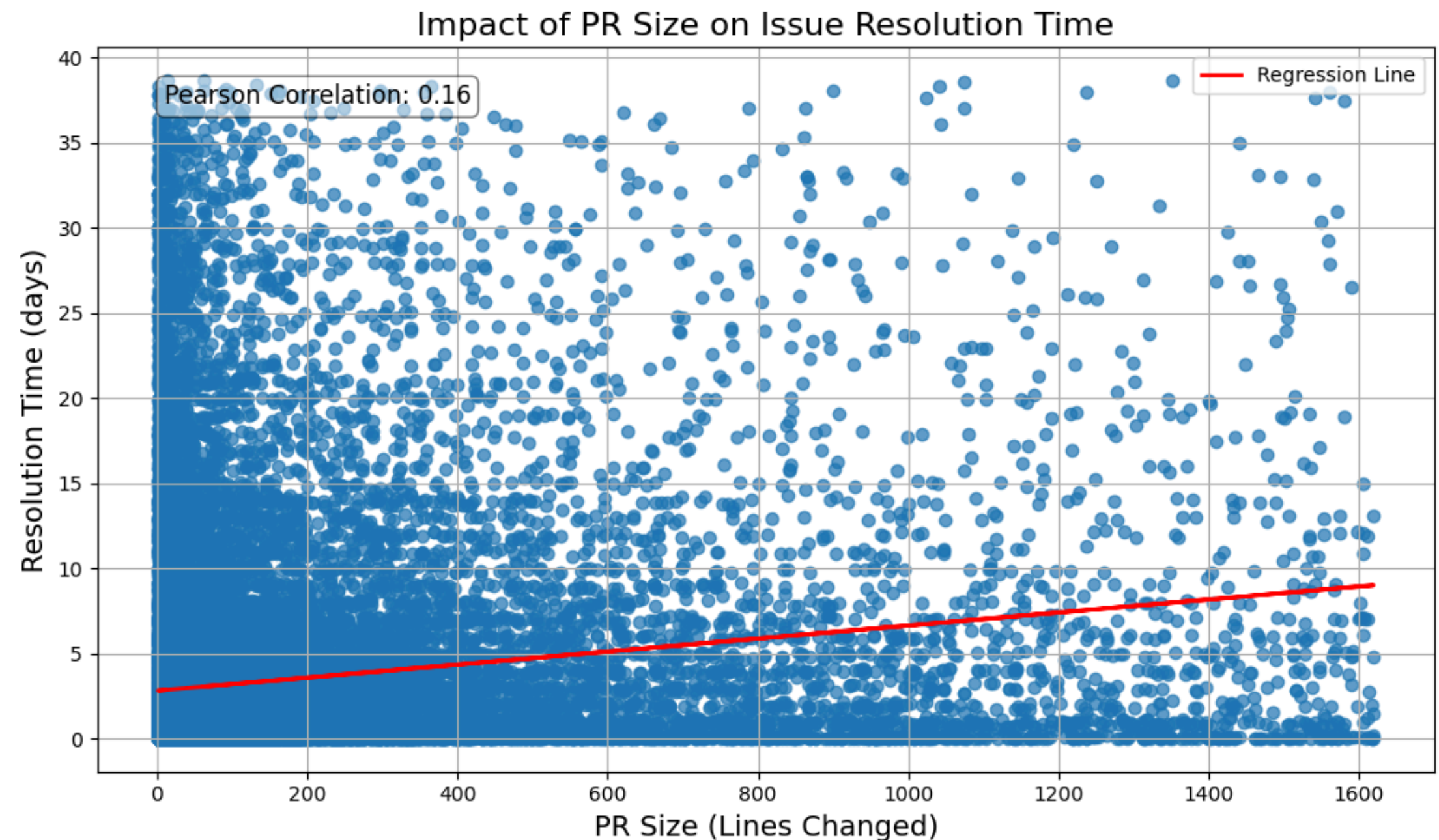


Figura 8 - Fonte: autoria própria.

RQ2.3. Qual o impacto do tamanho da comunidade no tempo de resolução de issues?

- Não é possível determinar uma correlação a partir do gráfico.
- Análise de Pearson apresentou valor de 0.17



Figura 9 - Fonte: autoria própria.

RQ2.4. Qual é a proporção de issues que tem PRs (pull requests) no repositório?

- Ao analisar as issues, foram encontradas apenas **58,71%** com PR(s), isso mostra que muitas issues apesar de finalizadas nenhuma modificação foi feita no código.

RQ2.5. Como o tempo de resolução das PRs está relacionado ao tamanho do repositório?

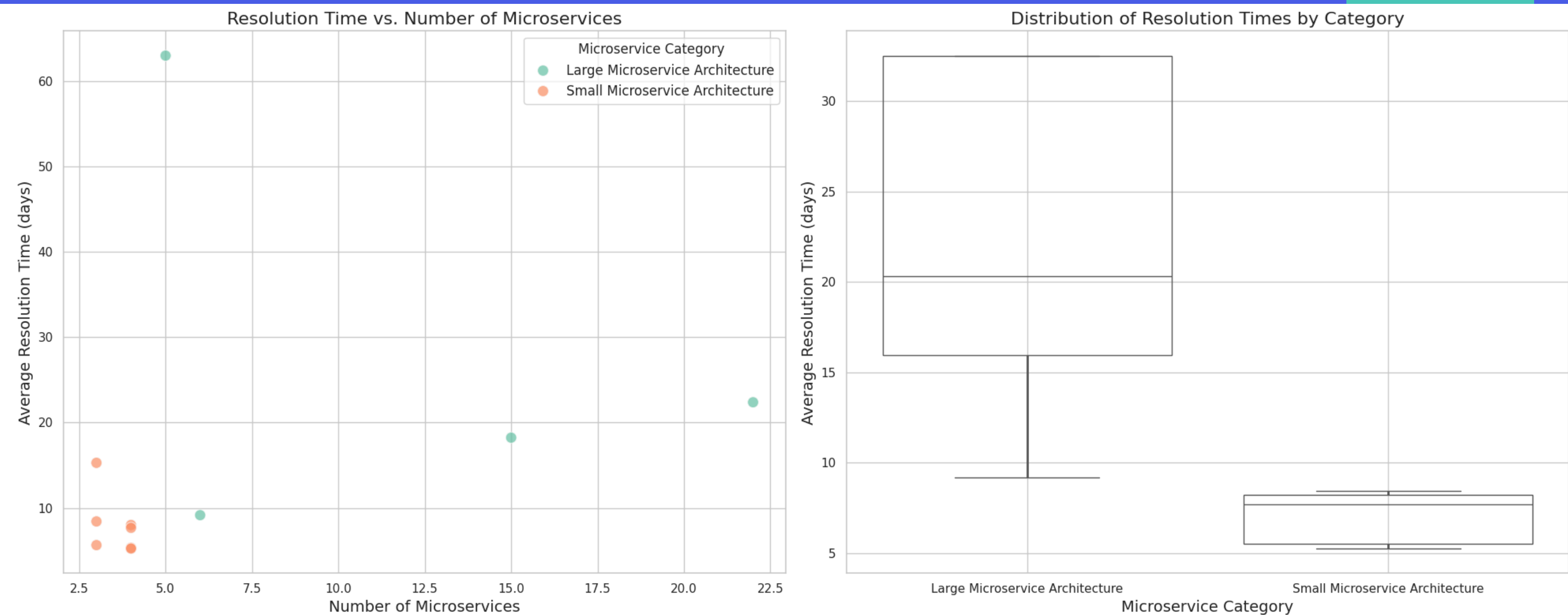


Figura 10 - Fonte: autoria própria.

RQ2.5. Como o tempo de resolução das PRs está relacionado ao tamanho do repositório?

- A análise de Kruskal-wallis mostrou uma diferença estatisticamente significativa nos tempos de resolução de issues entre repositórios com arquiteturas de microserviços grandes e pequenas ($p = 0.014$).
- Especificamente, arquiteturas de microserviços menores tendem a ter uma resolução de issues mais rápida em comparação com as maiores.



Fontes

- **1. Lewis, J.; Fowler, M. Microservices: A Definition of This New Architectural Term. Disponível em: <https://www.martinfowler.com/articles/microservices.html>. Acesso em: 31 ago. 2024.**
- **2. BLINOWSKI, Grzegorz; OJDOWSKA, Anna; PRZYBYŁEK, Adam. Monolithic vs. microservice architecture: A performance and scalability evaluation. IEEE Access, v. 10, p. 20357-20374, 2022.**
- **3. AMOROSO D'ARAGONA, Dario et al. A Dataset of Microservices-based Open-Source Projects. In: PROCEEDINGS of the 21st International Conference on Mining Software Repositories. 2024. p. 504-509**



Obrigado!