

1. Introdução

A prática de **code review** consolidou-se como parte essencial dos processos de desenvolvimento ágil, garantindo qualidade, legibilidade e confiabilidade no software entregue. Dentro do ecossistema do **GitHub**, os **Pull Requests (PRs)** são a principal forma de integrar código à branch principal, dependendo da revisão de outros desenvolvedores.

Este relatório tem como objetivo analisar **quais variáveis influenciam a aceitação (merge) ou rejeição (close) de PRs**, bem como a quantidade de revisões que cada solicitação recebe. A investigação é fundamentada em um **dataset construído a partir dos 200 repositórios mais populares do GitHub**, considerando apenas PRs revisados manualmente.

Hipóteses gerais

- PRs maiores tendem a ter maior dificuldade de aceitação.
 - O tempo de análise é proporcional à complexidade, podendo afetar negativamente o merge.
 - PRs com descrições mais completas têm maior chance de aprovação.
 - Interações (comentários e participantes) estão relacionadas a uma revisão mais rigorosa, podendo aumentar tanto o tempo de análise quanto as chances de aprovação.
-

2. Metodologia

1. Criação do dataset:

- Repositórios mais populares do GitHub (200).
- Apenas PRs com status **MERGED** ou **CLOSED**, que tenham recebido pelo menos **uma revisão** e com tempo de revisão **superior a 1 hora**.

2. Métricas utilizadas:

- **Tamanho do PR:** número de arquivos alterados, linhas adicionadas e linhas removidas.
- **Tempo de análise:** intervalo entre criação e última atividade (merge/close).

- **Descrição:** tamanho em caracteres da descrição.
- **Interações:** número de participantes e comentários no PR.

3. Análise estatística:

- Para identificar relações, utilizaremos **teste de correlação de Spearman**, por não assumir normalidade dos dados.
 - Serão construídas distribuições (valores medianos) para cada métrica e sua relação com **status final (MERGED/CLOSED)** e **número de revisões**.
-

3. Questões de Pesquisa e Hipóteses

Dimensão A – Feedback Final das Revisões

RQ01. Relação entre tamanho dos PRs e feedback final

- H0: Não existe relação significativa entre o tamanho do PR e a probabilidade de merge.
- H1: PRs menores têm maior chance de merge do que PRs maiores.

RQ02. Relação entre tempo de análise e feedback final

- H0: O tempo de análise não influencia no resultado (merge/close).
- H1: PRs analisados em menos tempo têm maior chance de merge.

RQ03. Relação entre descrição dos PRs e feedback final

- H0: O tamanho da descrição do PR não influencia no resultado.
- H1: PRs com descrições mais completas têm maior chance de merge.

RQ04. Relação entre interações nos PRs e feedback final

- H0: O número de interações não afeta a probabilidade de merge.
- H1: Maior número de interações aumenta a chance de merge.

Dimensão B – Número de Revisões

RQ05. Relação entre tamanho dos PRs e número de revisões

- H0: O tamanho dos PRs não influencia no número de revisões.
- H1: PRs maiores recebem mais revisões.

RQ06. Relação entre tempo de análise e número de revisões

- H0: O tempo de análise não influencia no número de revisões.
- H1: PRs com maior tempo de análise recebem mais revisões.

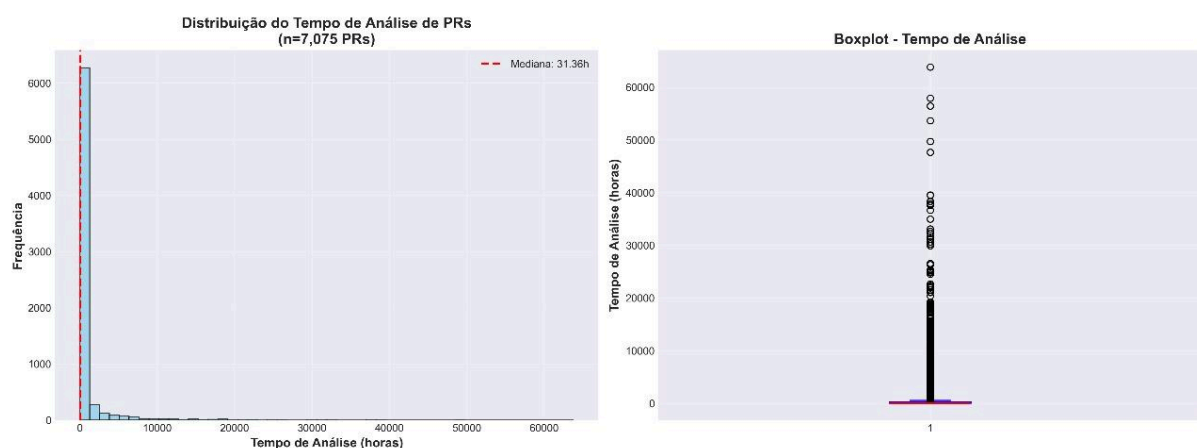
RQ07. Relação entre descrição dos PRs e número de revisões

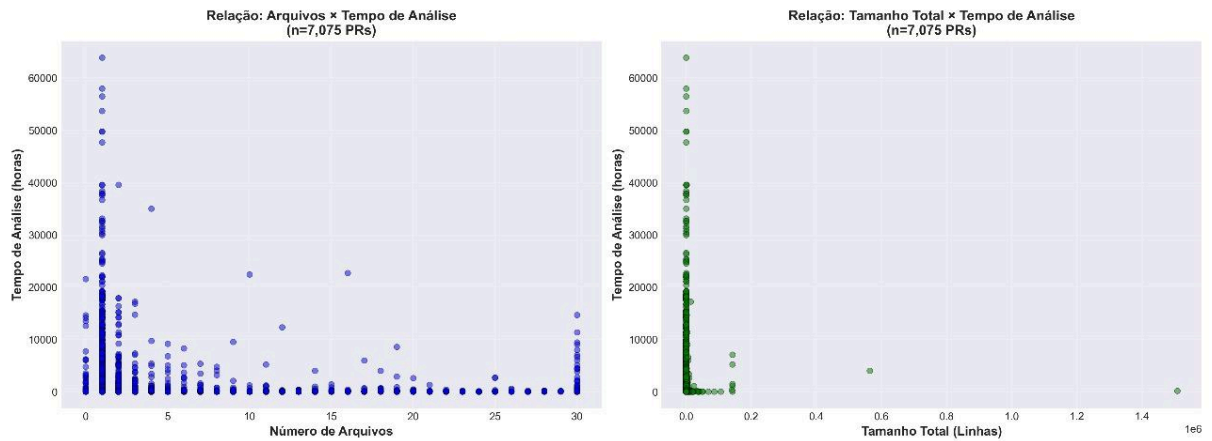
- H0: O tamanho da descrição não influencia no número de revisões.
- H1: PRs com descrições maiores recebem mais revisões.

RQ08. Relação entre interações nos PRs e número de revisões

- H0: O número de interações não influencia no número de revisões.
 - H1: PRs com maior número de interações recebem mais revisões.
-

4. Resultados Preliminares





DASHBOARD - ANÁLISE DE PULL REQUESTS

ESTATÍSTICAS GERAIS DO DATASET - MINERAÇÃO DE PULL REQUESTS DO GITHUB

📦 Volume de Dados: 7,075 Pull Requests | 199 Repositórios | 4,061 Autores Únicos

🕒 Tempo de Análise: Média = 870.53h | Mediana = 31.36h | Máximo = 63835.35h

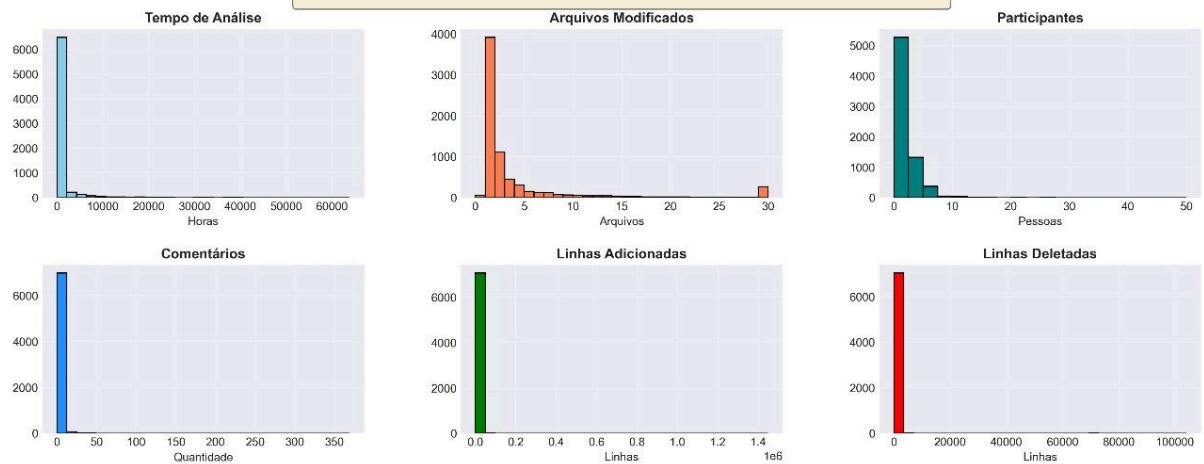
📁 Arquivos: Total = 26,908 arquivos | Média por PR = 3.80

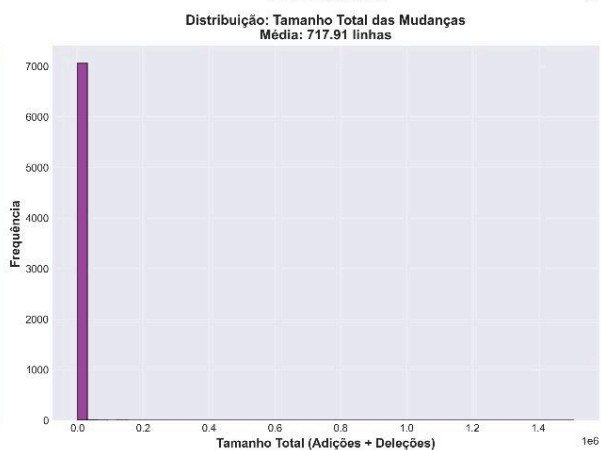
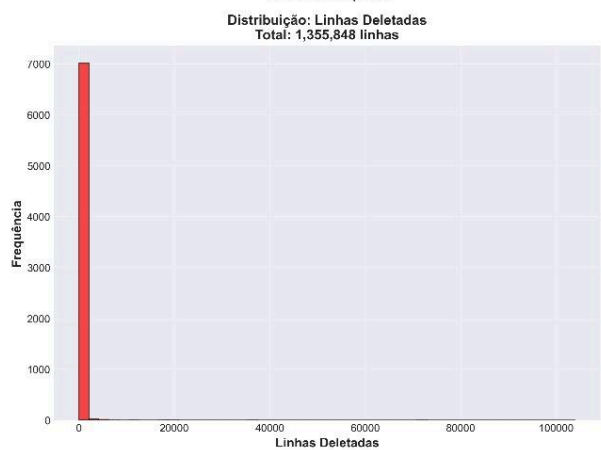
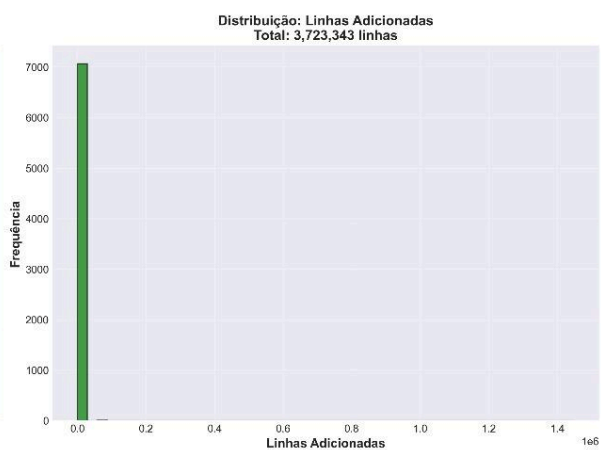
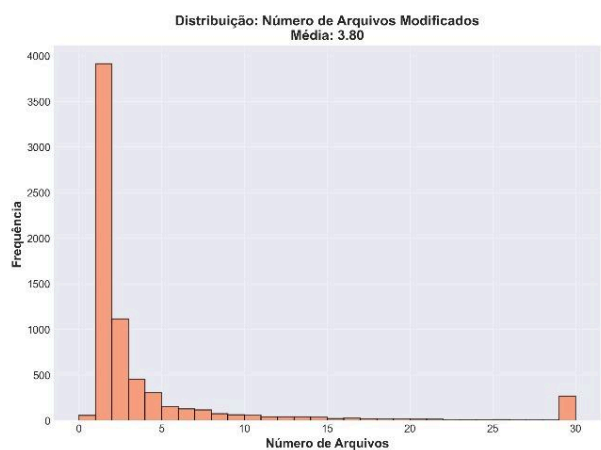
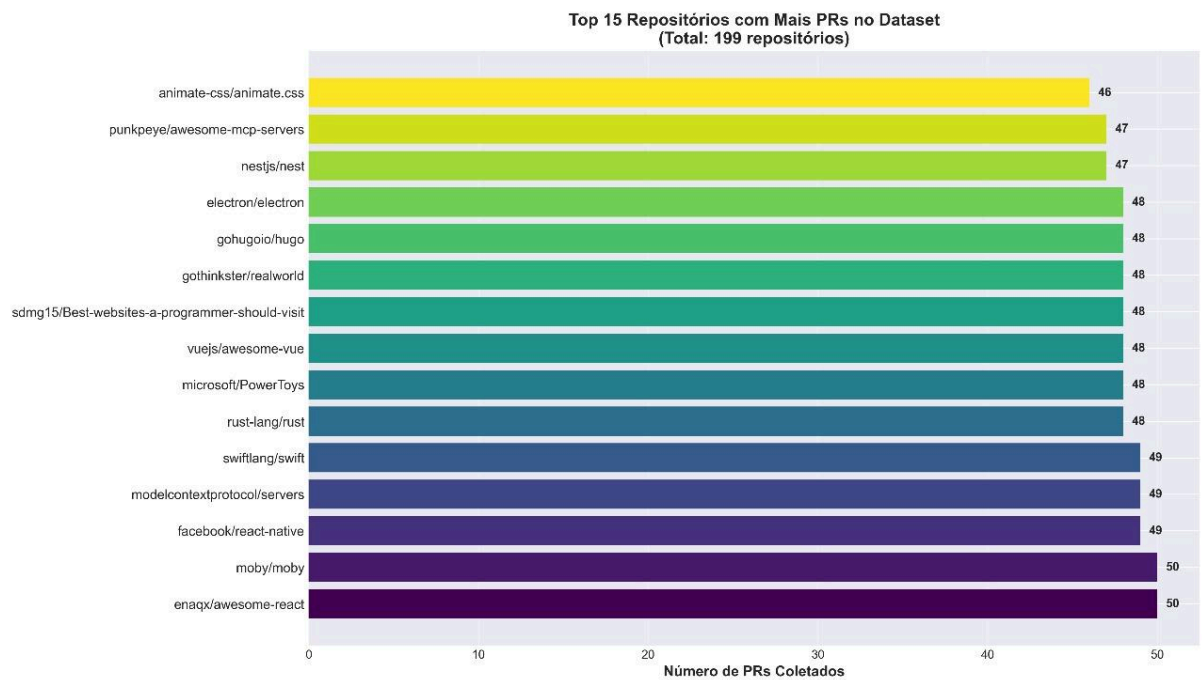
➕ Linhas Adicionadas: 3,723,343 linhas | Média por PR = 526.27

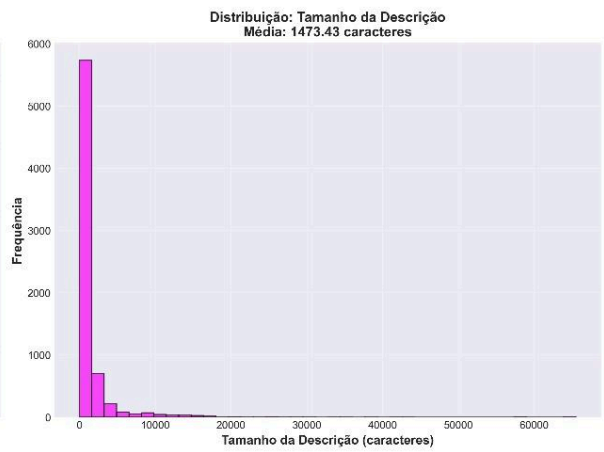
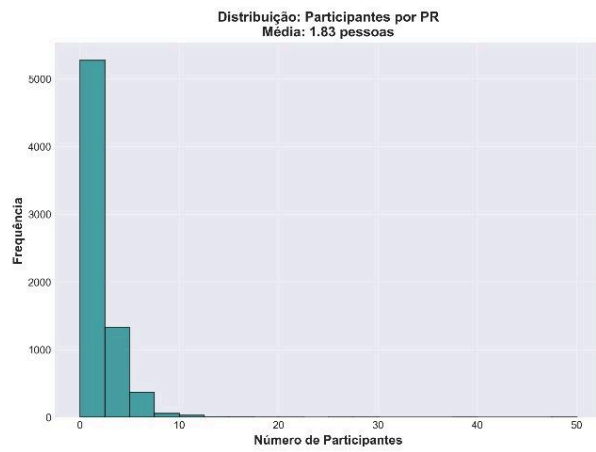
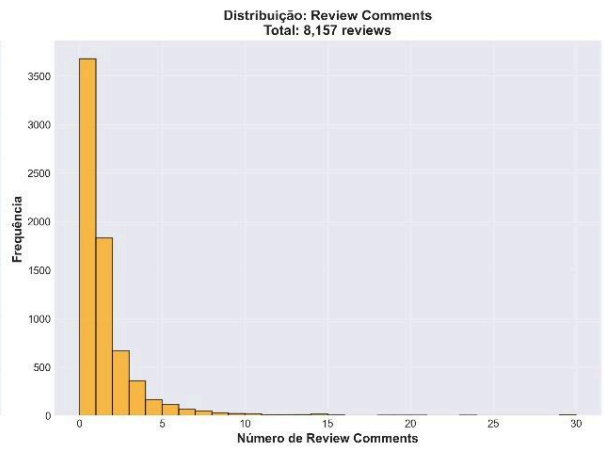
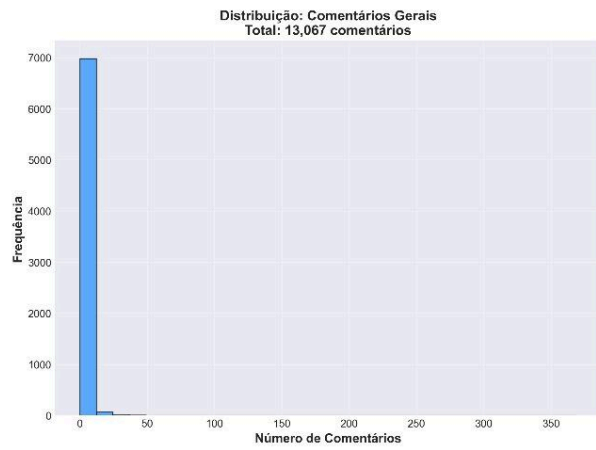
➖ Linhas Deletadas: 1,355,848 linhas | Média por PR = 191.64

💬 Comentários: 13,067 comentários gerais | 8,157 review comments

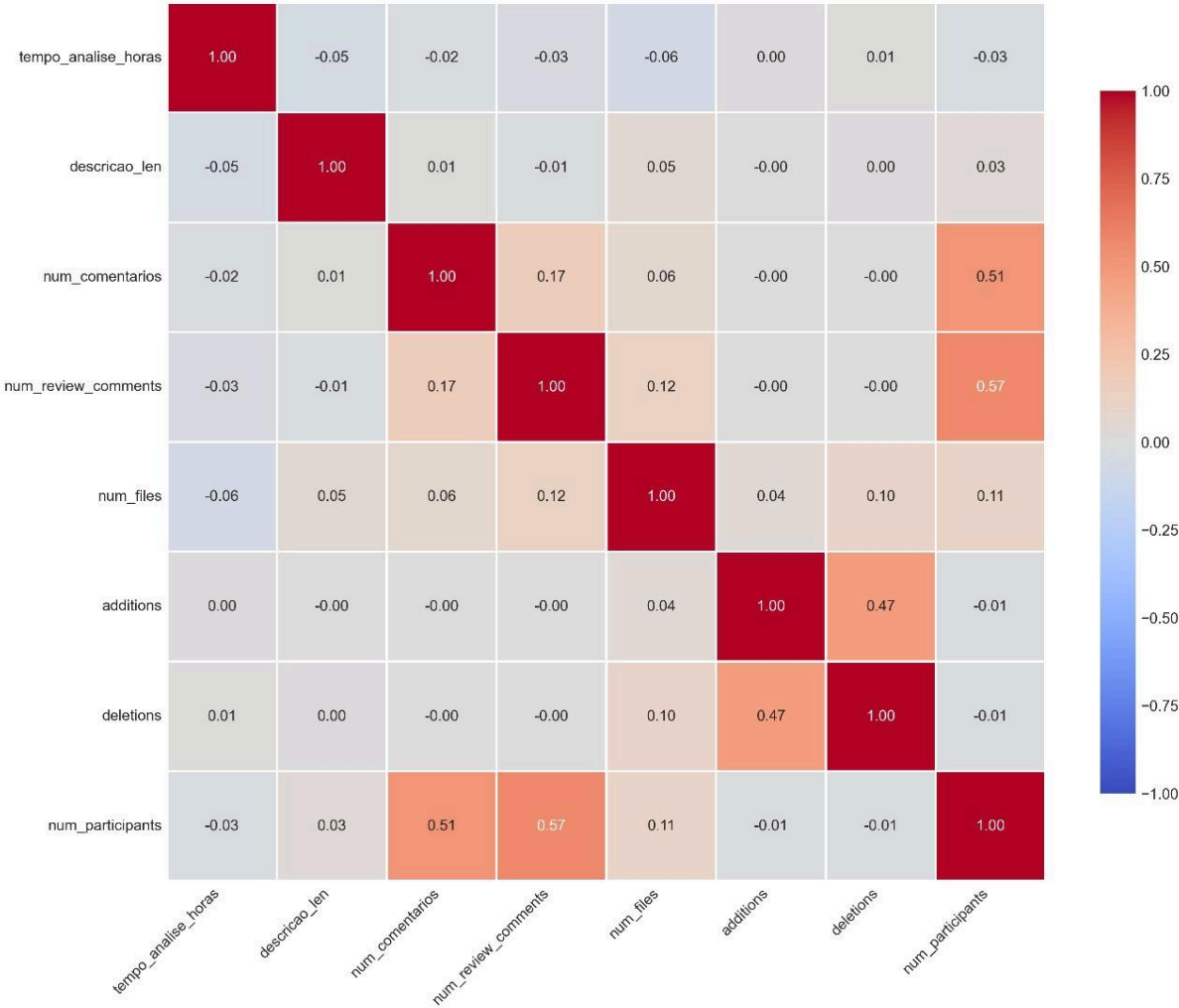
👥 Participação: Média de 1.83 participantes por PR

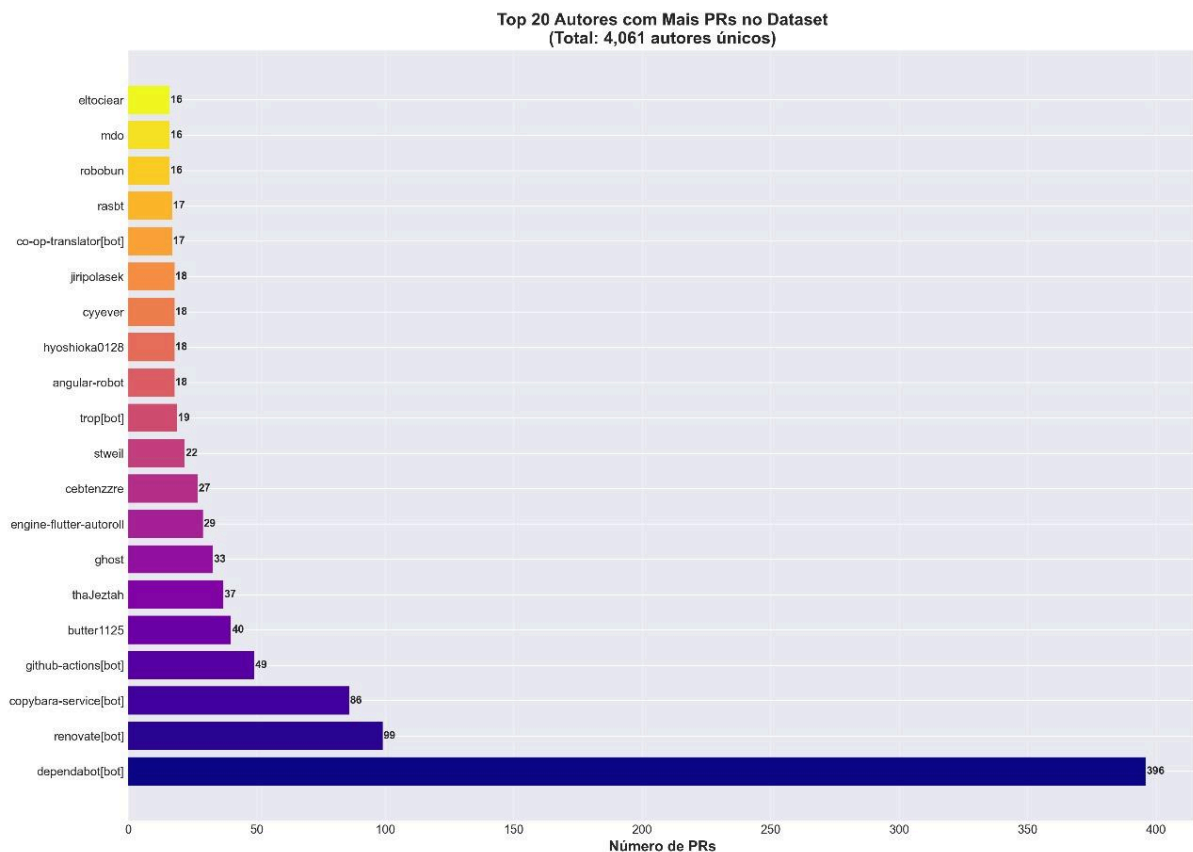






Matriz de Correlação entre Métricas de PRs
(n=7,075 Pull Requests)





5. Discussão

- **RQ01 e RQ05 (Tamanho dos PRs):** PRs grandes tendem a ser mais arriscados, com maior probabilidade de rejeição, mas também podem receber mais revisões.
- **RQ02 e RQ06 (Tempo de análise):** PRs com tempo elevado podem indicar problemas de complexidade ou falta de clareza, afetando o merge.
- **RQ03 e RQ07 (Descrição):** descrições detalhadas ajudam revisores, favorecendo aceitação e estimulando revisões mais aprofundadas.
- **RQ04 e RQ08 (Interações):** PRs que geram discussão ativa sugerem maior engajamento, podendo resultar em maior qualidade e probabilidade de merge.

6. Conclusão

A análise permite observar que múltiplas variáveis impactam o processo de code review em repositórios open source. Em especial, a **clareza da descrição** e o **tamanho dos PRs** surgem como fatores determinantes tanto para aceitação quanto para o número de revisões.

Com isso, recomenda-se que contribuidores submetam PRs menores, bem documentados e abertos a interações, aumentando as chances de aprovação no fluxo de desenvolvimento colaborativo.