

# Caracterizando a atividade de code review no github

## Introdução

A análise de code review em projetos open source no GitHub é fundamental para entender os fatores que influenciam na aceitação ou rejeição de contribuições. Este estudo foca em Pull Requests (PRs) nos 200 repositórios mais populares do GitHub, analisando como diferentes características dos PRs afetam seu resultado final e o processo de revisão.

Hipóteses Informais

Dimensão A: Feedback Final das Revisões (Status do PR)

### RQ01 - Tamanho dos PRs vs Feedback Final

Hipótese: PRs menores (com menos arquivos e menos linhas modificadas) têm maior probabilidade de serem aceitos (MERGED). PRs muito grandes podem ser rejeitados (CLOSED) por serem difíceis de revisar ou por introduzirem muitas mudanças de uma vez.

### RQ02 - Tempo de Análise vs Feedback Final

Hipótese: PRs que levam mais tempo para serem revisados têm maior chance de serem rejeitados, pois podem indicar discussões prolongadas ou problemas complexos. PRs aceitos podem ter um tempo médio de análise mais rápido.

### RQ03 - Descrição dos PRs vs Feedback Final

Hipótese: PRs com descrições mais detalhadas (mais caracteres) têm maior probabilidade de serem aceitos, pois facilitam o entendimento das mudanças propostas.

### RQ04 - Interações nos PRs vs Feedback Final

Hipótese: PRs com mais participantes e comentários têm maior chance de serem

rejeitados, pois podem indicar discordâncias ou problemas na implementação. PRs aceitos podem ter um número moderado de interações.

Dimensão B: Número de Revisões

RQ05 - Tamanho dos PRs vs Número de Revisões

Hipótese: PRs maiores exigem mais revisões, pois contêm mais mudanças a serem validadas.

RQ06 - Tempo de Análise vs Número de Revisões

Hipótese: PRs com mais revisões tendem a ter um tempo de análise maior, pois cada revisão adicional aumenta o tempo total.

RQ07 - Descrição dos PRs vs Número de Revisões

Hipótese: PRs com descrições mais claras (mais caracteres) podem exigir menos revisões, pois reduzem a necessidade de esclarecimentos adicionais.

RQ08 - Interações nos PRs vs Número de Revisões

Hipótese: PRs com mais participantes e comentários tendem a ter mais revisões, indicando discussões prolongadas sobre as mudanças.

---

## Metodologia

### Coleta de Dados

1. Seleção de Repositórios:
  - Os 200 repositórios mais populares do GitHub (com mais de 10.000 estrelas).
  - Repositórios com pelo menos 100 PRs (MERGED + CLOSED).
2. Filtragem de PRs:
  - PRs com status MERGED ou CLOSED.
  - Pelo menos uma revisão (não automática).
  - Tempo mínimo de revisão de 1 hora (para excluir revisões automatizadas).

### 3. Métricas Coletadas:

- **Tamanho:** Número de arquivos, linhas adicionadas (+), linhas removidas (-).
- **Tempo de Análise:** Intervalo entre criação e fechamento/merge do PR.
- **Descrição:** Número de caracteres no corpo do PR.
- **Interações:** Número de participantes e comentários.

### Análise Estatística

- Será utilizada a correlação de Spearman para avaliar as relações entre as métricas, pois não assumimos que os dados seguem uma distribuição normal.
- Os resultados serão sumarizados usando medianas (para reduzir o impacto de outliers).

---

## Resultados Preliminares

O dataset contém PRs válidos de 200 repositórios populares, com as seguintes métricas extraídas para cada PR:

- **changedFiles:** Número de arquivos modificados.
- **additions:** Linhas adicionadas.
- **deletions:** Linhas removidas.
- **Tempo de análise:** Duração entre criação e fechamento/merge.
- **Descrição:** Tamanho do texto de descrição (em caracteres).
- **Participantes:** Número de usuários envolvidos.
- **Comentários:** Número total de comentários.

Exemplo de Dados Coletados:

Relatório de Pull Requests Válidos								
Métricas extraídas de PRs com review após 1 hora								
Repositório	Título PR	Arquivos	+Linhas	-Linhas	Tempo de análise	Descrição (#chars)	Participantes	Comentários
decrafters-io/build-your-own-x	Update README.md	1	1	1	48 days, 22:47:44	34	3	0
	feat: added how to build a cli app with node 2024 guide	1	1	0	76 days, 5:18:15	143	2	1
	fix: fixed broken link from list	1	1	1	7 days, 0:48:47	13	6	2
	Added python commandline tool	1	2	0	2 days, 6:55:57	93	2	0
	new pull request	1	1	0	17:54:43	0	3	0
decrafters-io/build-your-own-x	Update README.md	1	3	3	469 days, 13:09:53	31	2	1
decrafters-io/build-your-own-x	Update README.md	1	1	0	5 days, 5:33:19	14	2	0
sindresorhus/awesome	Add CodeRabbit	1	1	0	46 days, 2:28:14	9542	6	13
sindresorhus/awesome	Add Large Language Model	1	1	0	9 days, 8:25:28	9022	7	4
sindresorhus/awesome	Add Free Apps	1	3	0	26 days, 6:47:07	9103	8	13