

# Caracterizando a atividade de code review no github

Alfredo Luis Vieira, Vinicius Salles

15 de outubro de 2025

## Introdução

A prática de revisão de código (code review) é um pilar fundamental nos processos de desenvolvimento de software modernos, especialmente em metodologias ágeis. Ela consiste na inspeção do código-fonte por outros membros da equipe antes de sua integração à base principal do projeto, garantindo a qualidade do que é produzido e prevenindo a introdução de novos defeitos.

No ecossistema de desenvolvimento de software de código aberto (open source), plataformizado em grande parte pelo GitHub, a atividade de revisão de código se materializa por meio das Pull Requests (PRs). Desenvolvedores submetem suas contribuições através de PRs, que são então avaliadas e debatidas por outros colaboradores do projeto. Ao final desse processo iterativo, a contribuição pode ser aceita (merged) e integrada ao código principal, ou rejeitada (closed).

Este estudo tem como objetivo principal analisar e caracterizar a atividade de code review em repositórios populares do GitHub. Buscamos identificar as variáveis que influenciam tanto o resultado final de uma Pull Request (seja ela aceita ou rejeitada) quanto a intensidade do processo de revisão, medido pelo número de revisões que uma PR recebe. A análise será conduzida sob a perspectiva de desenvolvedores que submetem suas contribuições, investigando como fatores como o tamanho da contribuição, o tempo de análise, a qualidade da descrição e o nível de interação podem se correlacionar com o desfecho do processo de revisão.

## Metodologia

A construção do dataset foi realizada através da funcionalidade de busca avançada do GitHub. Primeiramente, elaboramos uma consulta estruturada com filtros para selecionar Pull Requests (PRs) que atendiam aos critérios da pesquisa, como status (merged ou closed) e popularidade do repositório, esta estratégia se difere da utilização da API GraphQL do Github, nesse caso, foi utilizado uma query de busca com alguns filtros (status merged ou closed, e que possuía uma revisão), para que a pesquisa seja realizada. Devido a massa de dados ser muito grande, foi necessário deixar uma máquina executando a busca particionada durante o período da noite.

Com os dados obtidos, foram implementados scripts para analisar estes dados, estes scripts foram implementados utilizando python como linguagem de programação, esta análise está voltada a dados acerca das reviews da pr, como comentários, descrição e etc.

Foi utilizado a biblioteca Pandas do python para realizar a criação de gráficos para melhor visualização dos dados obtidos.

Questões de pesquisa

Métricas utilizadas

1 Considerações gerais

2 Resultados obtidos

3 Considerações finais