



PUC Minas

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

**Graduação em Engenharia de Software – Instituto de Ciências Exatas e
Informática**

Renato Paganini Thurler e Marcos Paulo Freitas da Silva

Laboratório de Experimentação de Software

Belo Horizonte

2025

1: Introdução

Este estudo foi conduzido pelos alunos Renato Paganini Thurler e Marcos Paulo Freitas da Silva durante a disciplina de Laboratório de Medição e Experimentação de Software. Com o objetivo de examinar a prática de *code review* em repositórios populares do GitHub, buscando identificar fatores que influenciam a aprovação de um PR, serão analisadas métricas de software previamente definidas, para posterior análise sob a ótica dos desenvolvedores que submetem código aos repositórios selecionados.

2: Metodologia

A escolha dos repositórios que foram utilizados durante o estudo foi baseada na atividade de code review de repositórios open source desenvolvidos na plataforma GitHub, a partir da avaliação de contribuições submetidas por meio de Pull requests. Ou seja, para que um código integre a branch principal é necessária a avaliação e discussão por colaboradores do projeto.

Nesse contexto, foram realizadas análises das atividades de code review desenvolvidas em repositórios populares do GitHub, visando identificar variáveis que influenciam no merge de um PR. Uma vez identificados os repositórios, uma série de métricas previamente fornecidas foram analisadas para avaliar aspectos chave relacionados aos feedbacks e as avaliações fornecidas para os PR's. Entre as principais métricas que serão calculadas encontram-se o tamanho do PR, a quantidade de interações e a sua descrição. Por fim, com os resultados obtidos através das análises, foram levantadas hipóteses em relação ao assunto debatido, oferecendo uma visão clara sobre como essas características afetam na qualidade das revisões e posteriormente na qualidade do código da branch principal.

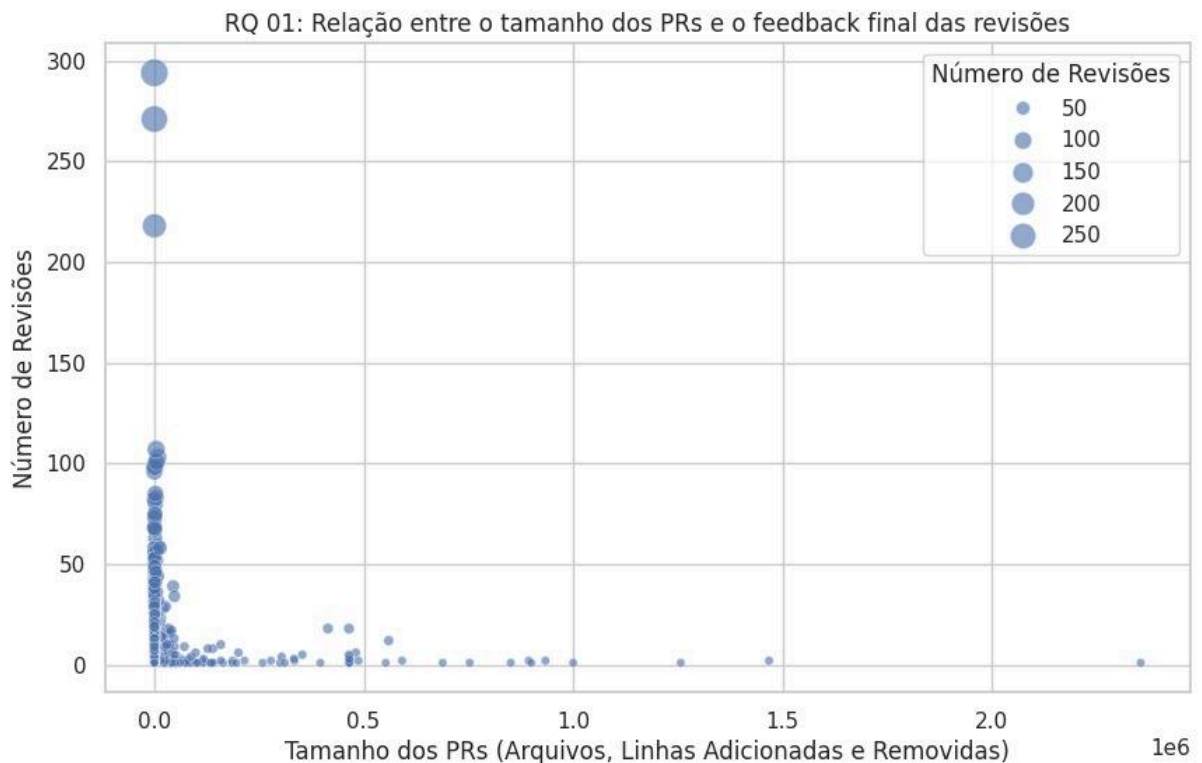
Além disso, para garantir a objetividade e precisão dos resultados, um programa desenvolvido em Python será responsável por automatizar o processo de extração e cálculo dessas métricas, permitindo a análise em larga escala dos dados. A escolha de Python como linguagem de implementação se dá pela sua flexibilidade e ampla adoção em projetos de análise de dados, o que facilitará o processamento das informações e a geração de relatórios que respaldam as conclusões deste estudo. Dessa forma, esperamos fornecer insights detalhados sobre as práticas e padrões predominantes em projetos Java open-source, contribuindo para uma compreensão mais profunda de como esses projetos são gerenciados e mantidos ao longo do tempo.

3: Resultados obtidos

RQ 01. Qual a relação entre o **tamanho** dos PRs e o feedback final das revisões?

PRs menores são mais observados e submetidos a mais revisões, provavelmente por serem mais simples de analisar e revisar minuciosamente.

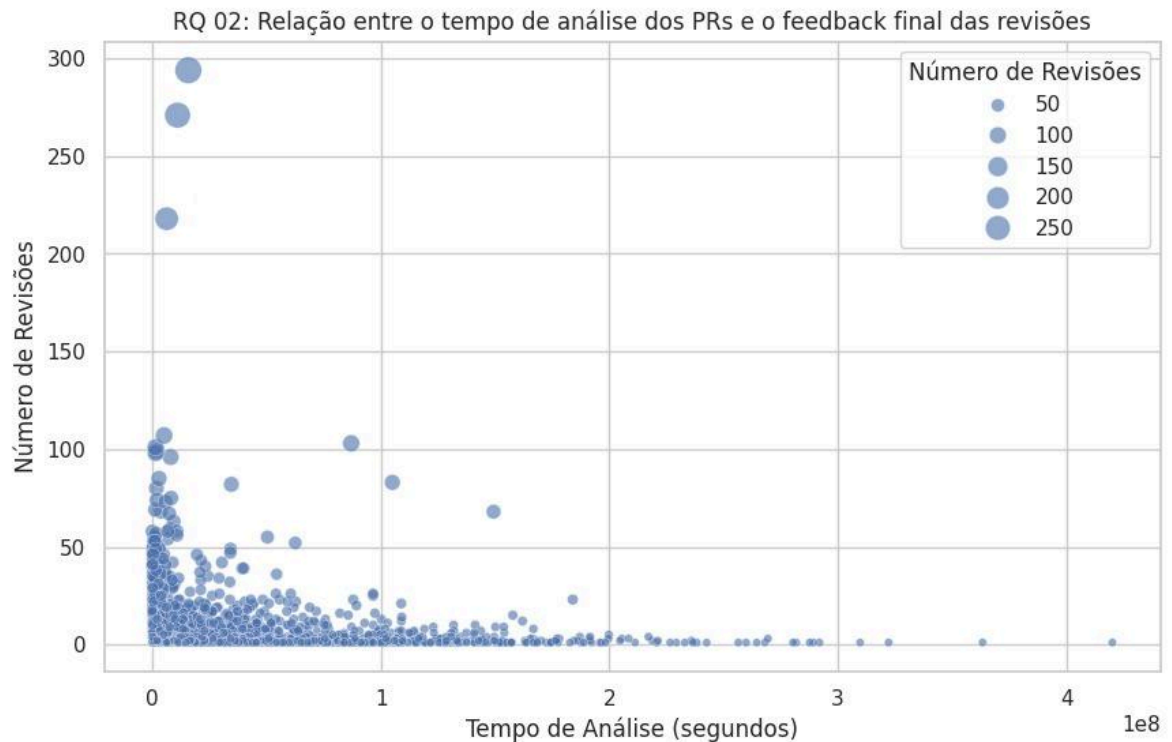
PRs mais extensos podem ser complicados de revisar totalmente, o que pode restringir a quantidade de revisões. Isso também pode indicar um método de revisão de alterações menores, ao invés de revisar grandes conjuntos de código de uma só vez.



RQ 02. Qual a relação entre o **tempo de análise** dos PRs e o feedback final das revisões?

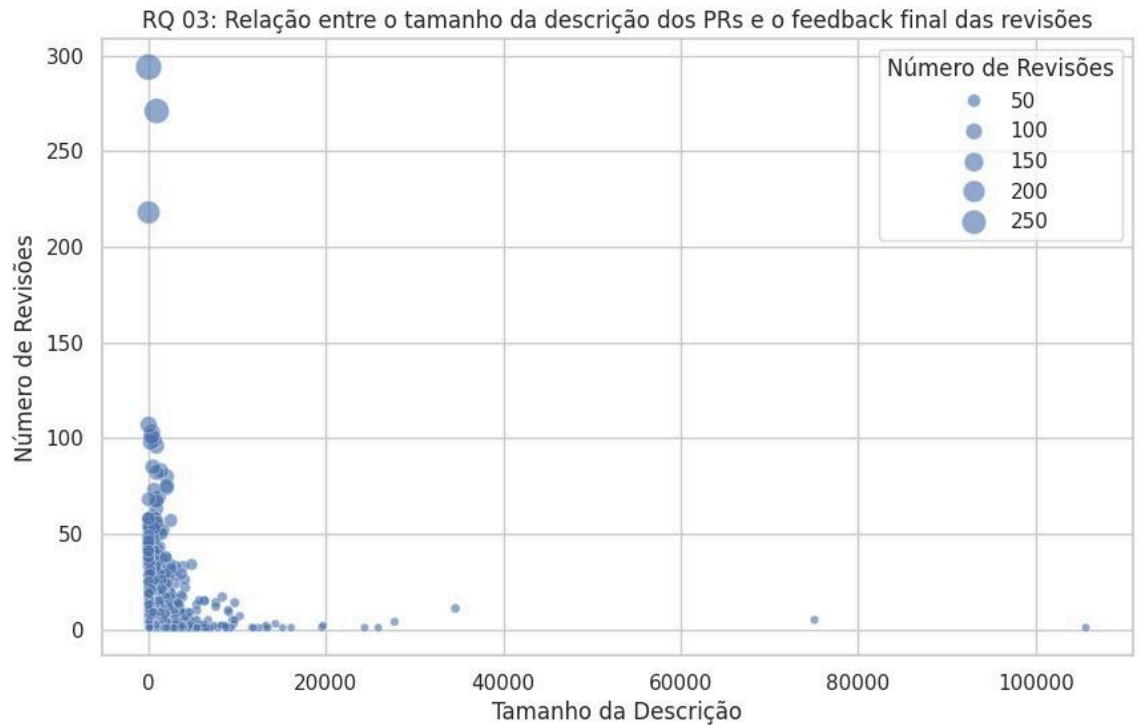
Pull requests que têm um tempo de análise reduzido são mais frequentes e recebem revisões com maior regularidade, podendo representar um método de revisão ágil e incessante para alterações menos significativas.

Pull requests com um tempo de revisão prolongado podem ser avaliados com menor frequência, possivelmente devido à sua maior complexidade ou detalhamento, o que exige uma análise mais cuidadosa e espaçada.



RQ 03. Qual a relação entre a **descrição** dos PRs e o feedback final das revisões?

O diagrama ilustra a conexão entre a extensão da descrição dos Pull Requests (PRs) e a quantidade de revisões que esses PRs receberam, evidenciada pelo retorno final das revisões. Parece haver uma predominância de revisões em PRs com descrições relativamente breves, o que pode sugerir que PRs com descrições mais breves tendem a receber mais avaliações. Desde um determinado ponto, o crescimento no tamanho da descrição não parece afetar significativamente a quantidade de avaliações, indicando que descrições muito extensas não elevam a quantidade de comentários.

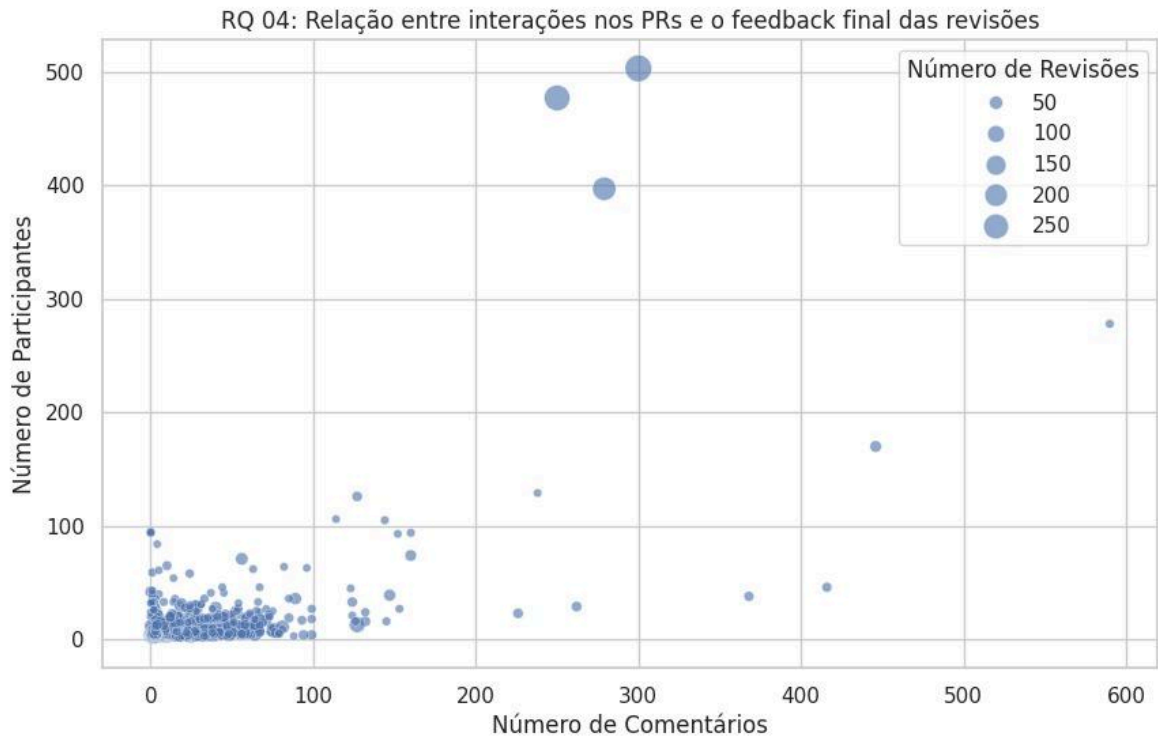


RQ 04. Qual a relação entre as **interações** nos PRs e o feedback final das revisões?

PRs com descrições curtas podem envolver discussões mais intensas ou revisões frequentes, talvez devido à falta de clareza na descrição ou à complexidade pouco clara.

PRs com descrições mais longas tendem a receber menos comentários, mas podem ser mais autoexplicativos, facilitando a compreensão dos revisores e reduzindo a necessidade de feedback adicional.

Estas observações podem ajudar a compreender a dinâmica e os detalhes da comunicação durante o processo de revisão.



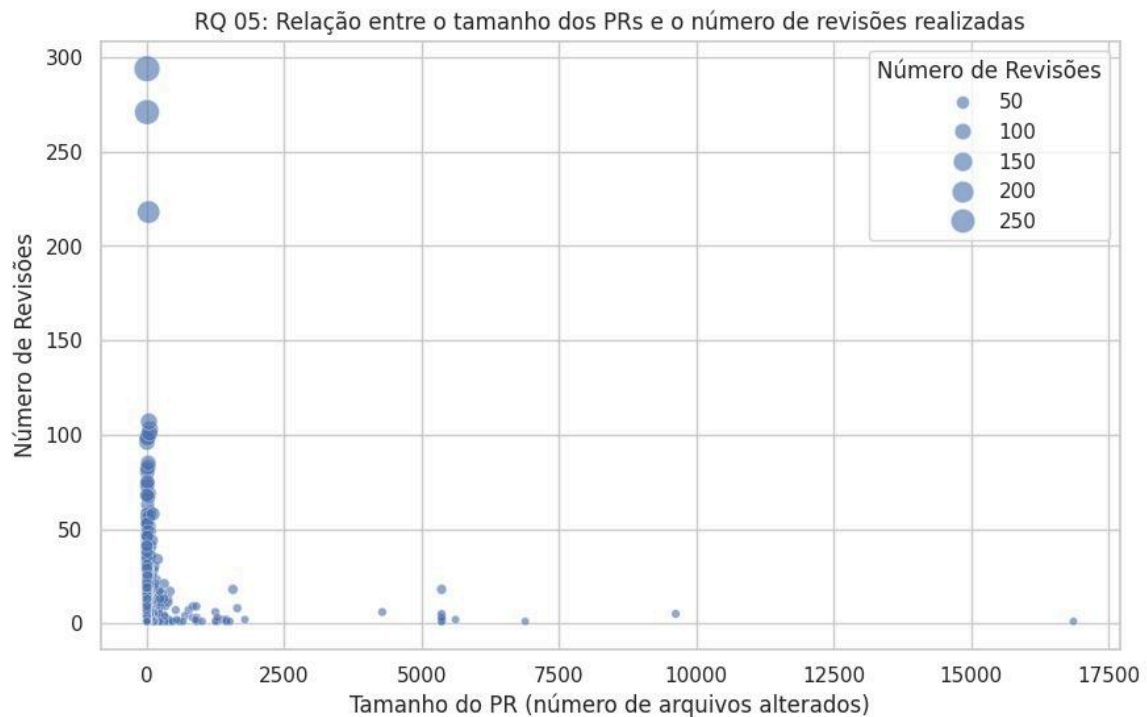
RQ 05. Qual a relação entre o **tamanho** dos PRs e o número de revisões realizadas?

Baixa densidade de interação: As áreas com maior concentração de dados são áreas com baixo número de comentários e participantes (menos de 100 comentários e até 100 participantes). Isto sugere que a maioria dos PRs tem pouca interação e envolve um número limitado de atores.

Valores discrepantes em comentários e colaboradores: Existem alguns valores discrepantes que representam PRs com mais de 200 comentários e mais de 400 colaboradores. Estes pontos podem indicar uma revisão ou PR altamente controversa que requer uma análise mais profunda, possivelmente devido a mudanças críticas ou grandes contribuições.

Número de comentários: O tamanho dos pontos indica que os PRs com maior número de comentários (marcados por círculos maiores) também geralmente têm mais colaboradores. Isto sugere uma correlação entre a revisão intensiva e a participação de mais pessoas, possivelmente indicando a complexidade ou importância destes PRs.

Frequência das revisões: Muito poucos BP conseguem um elevado número de revisões e estes casos tendem a ocorrer em locais isolados com um número razoável de comentários. Isto reforça a ideia de que PRs complexos exigem mais revisões e geram mais discussão.



RQ 06. Qual a relação entre o **tempo de análise** dos PRs e o número de revisões realizadas?

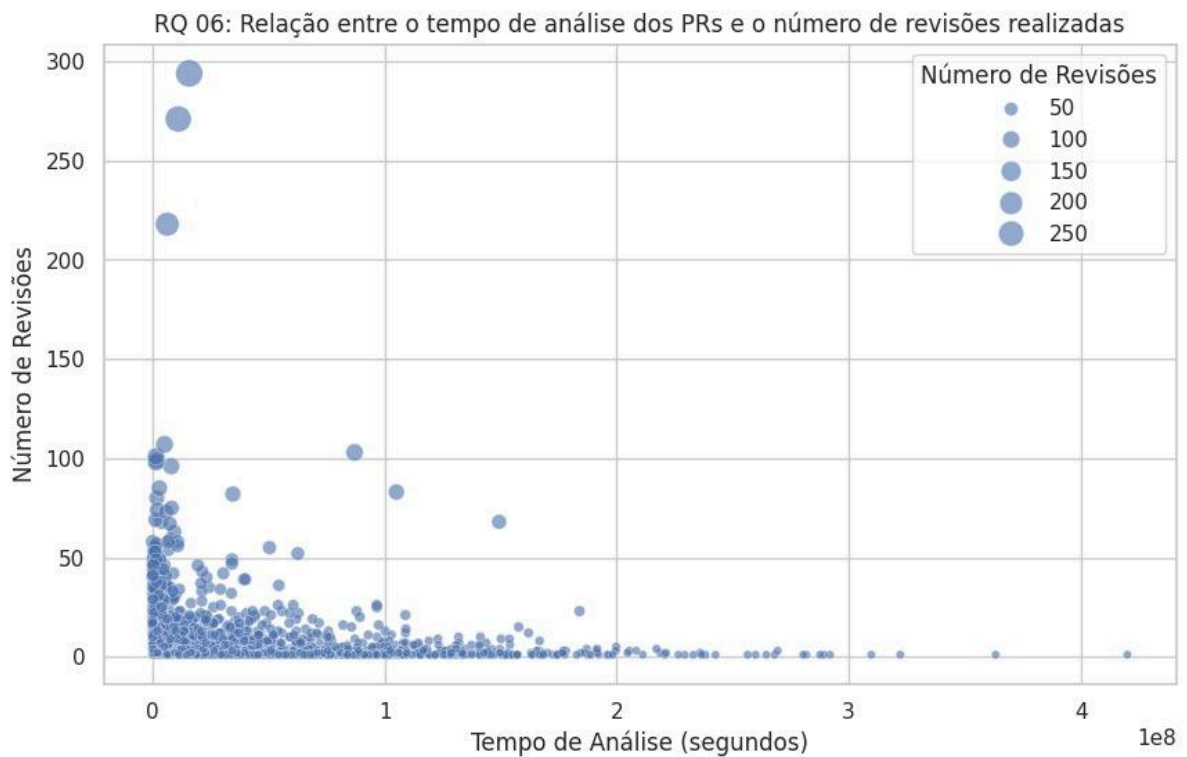
Alta densidade, baixo tempo de análise: a maioria dos PRs tem um tempo de análise muito curto (entre 0 e 1×10^8 segundos) e um pequeno número de comentários (menos de 50). Isto mostra que, em geral, a análise de RP é rápida e requer poucas modificações.

Valores discrepantes nos comentários: Existem alguns PRs com um número muito elevado de comentários (mais de 100), principalmente concentrados em áreas com tempos de comentários mais curtos. Esses PRs são destacados com pontos maiores, indicando que, apesar do elevado número de comentários, os

comentários não foram muito longos, indicando uma rápida interação entre os comentadores.

Baixa correlação entre o tempo de análise e o número de revisões: O gráfico não mostra uma relação direta entre o tempo de análise e o número de revisões. Mesmo que o tempo de análise aumente, o número de comentários não aumenta proporcionalmente.

Existem poucos casos de longos tempos de análise: há alguns pontos em que os tempos de análise são longos (2×10^8 segundos), mas o número de comentários é pequeno, sugerindo que esses PRs podem ter estado abertos por um longo tempo sem muito tempo. interação.



RQ 07. Qual a relação entre a **descrição** dos PRs e o número de revisões realizadas?

Descrições curtas são predominantes entre a maioria dos PRs, geralmente consistindo em menos de 20.000 caracteres e passando por menos de 50

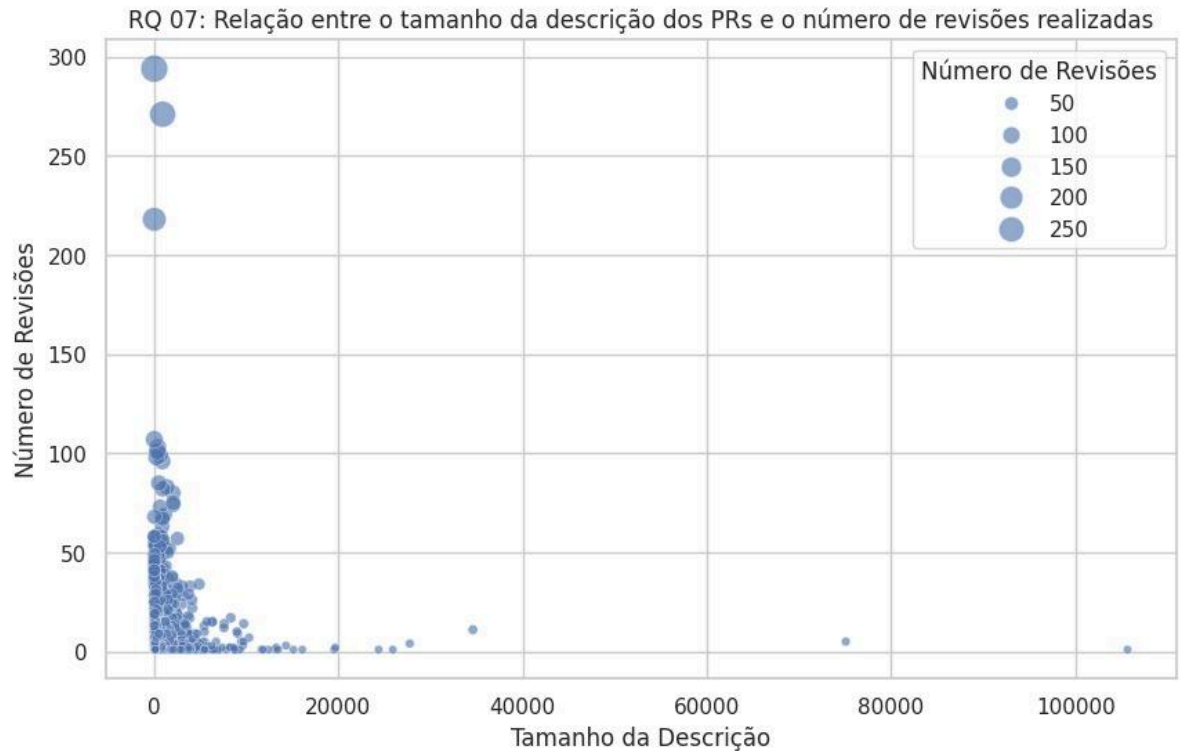
revisões. Isso sugere que tais descrições breves são encontradas com mais frequência e podem necessitar de menos revisão.

Altos Outliers de Revisão: Um pequeno número de PRs exibe uma contagem excepcionalmente alta de revisões (excedendo 100), particularmente encontrada em descrições breves. Os círculos maiores enfatizam essas instâncias, indicando que o comprimento da descrição não influencia inerentemente a frequência de revisões.

Fraca Relação entre Comprimento de Descrições e Contagem de Revisões: Existe uma distinta falta de correlação entre o comprimento das descrições e a frequência de revisões. Mesmo descrições que se estendem a 100.000 caracteres não mostram um número significativo de revisões, sugerindo que a complexidade de um PR pode não estar intimamente ligada ao comprimento de sua descrição.

A maioria das solicitações de pull (PRs) apresenta descrições breves, normalmente com menos de 20.000 caracteres, e exibe menos de 50 revisões. Isso indica que descrições concisas são mais prevalentes e geralmente necessitam de menos alterações.

Instâncias de descrições longas: Um punhado de PRs contém descrições muito longas, variando de 40.000 a 100.000 caracteres, mas exibem um número relativamente pequeno de revisões. Isso pode sugerir que, embora essas descrições sejam completas, elas passaram por menos revisões, possivelmente devido à sua clareza ou porque precisaram de menos ajustes.



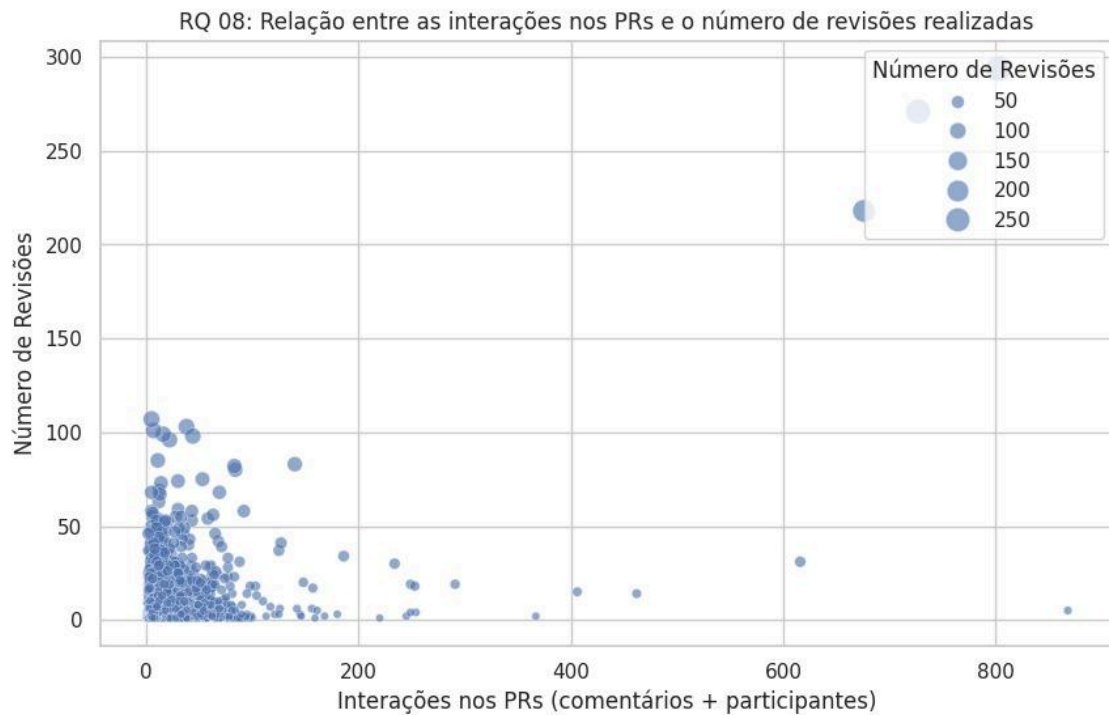
RQ 08. Qual a relação entre as **interações** nos PRs e o número de revisões realizadas?

Concentração em Interações e Revisões Reduzidas: A maior parte dos pontos está localizada na parte inferior esquerda, onde a quantidade de interações e revisões é bastante reduzida. Isso indica que, na maior parte das Propostas de Revisão, existem poucas interações e a quantidade de revisões também é reduzida.

Poucas Ocorrências com Grande Volume de Revisões: À medida que o número de interações cresce, nota-se que apenas algumas PRs exibem um volume considerável de revisões, mesmo que ainda sejam reduzidas em relação ao total.

Escala de Pontos (Bolhas): O número de pontos reflete o número de revisões em cada Procedimento de Revisão. Nota-se que os pontos mais significativos, que representam um maior número de revisões, também estão concentrados na área de interação reduzida, porém existe uma pequena variação.

Outliers: Existem em menor número alguns pontos entre 400 e 800, mas são casos isolados e não apresentam uma relação forte com o número de revisões



4: Conclusão

A primeira hipótese levantada é que quanto maior o tamanho dos PRs, pior será o feedback das revisões. Isso se baseia no fato de que o maior tamanho do PR causa com que o trabalho de feedback seja mais trabalhoso e complicado de ser detalhado.

Após análise dos dados obtidos a conclusão da hipótese 1 é que os revisores podem preferir trabalhar com PRs menores, onde a revisão é mais fácil e acessível, enquanto que PRs maiores podem ser mais complexos de revisar, levando a uma avaliação menos detalhada.

A segunda hipótese que será analisada no estudo contempla que quanto menor o tempo de análise, menos preciso será o feedback fornecido.

Em relação à segunda hipótese, realizadas as devidas análises obtém-se o seguinte resultado, as informações podem indicar uma inclinação para revisões ágeis e

repetidas em pull requests (PRs) simples, enquanto PRs que exigem uma análise mais aprofundada parecem solicitar uma abordagem mais cuidadosa e demorada, refletindo variações nas práticas de revisão de acordo com a complexidade do PR.

Como terceiro lugar, entendemos que quanto mais verboso (maior o tamanho da descrição) o PR, melhor e mais precisos serão os feedbacks finais. Isso se dá pelo fato do detalhamento mais completo, de maneira que o feedback saem mais precisos e focados no problema.

Utilizando-se dos dados para a apuração da terceira hipótese, a conclusão é que, em síntese, PRs com descrições mais breves tendem a receber mais revisões, ao passo que PRs com descrições extremamente extensas costumam receber menos avaliações. Isso pode sugerir que descrições claras simplificam a revisão.

Em quarto lugar, será levada em consideração a hipótese de que quanto mais interações presentes nas discussões do Pull Request, melhor o feedback final. A ideia por trás é que mais discussão traz um entendimento melhor do contexto da implementação e regra de negócio, de maneira que o PR tende a ser aprovado com maior confiabilidade.

Depois da devida apuração para a quarta hipótese, o resultado alcançado é que PRs com mais comentários e maior participação tendem a receber mais revisões, conforme indicado pelo tamanho dos círculos. Isso sugere que, quanto mais interações ocorrem em um PR (seja em forma de comentários ou participantes), mais revisões são realizadas, o que pode resultar em um feedback mais detalhado e potencialmente de maior qualidade.

Em seguida, a quinta hipótese é que um grande PR leva a mais revisões para garantir que todas as partes foram consideradas de forma correta. Nesse contexto, PRs maiores tendem a ser mais complexos e exigir mais revisão.

A averiguação da quinta hipótese retornou que a maioria dos RP tem interações relativamente limitadas em termos de comentários e participantes, com alguns casos envolvendo um grande número de comentários e participantes, possivelmente associados a RP complexos.

Após isso, foi levantado que o tempo de análise contribui para que sejam feitas mais revisões, Dessa forma possuem mais tempo para achar mais inconsistências e com isso são criadas mais revisões.

Nesse contexto, a sexta hipótese levantada é baseada na ideia de que quanto maior a descrição, menos revisões serão feitas. Dessa forma, quanto melhor a documentação, mais precisa será a análise.

Seguindo a sexta hipótese e seus testes de averiguação, a conclusão obtida é que a maioria dos RP são revistos rapidamente e com um pequeno número de revisões, enquanto poucos RP requerem um grande número de revisões. Além disso, o tempo de revisão geralmente não se correlaciona fortemente com o número de revisões, exceto em determinadas circunstâncias específicas.

Ainda nesse viés, como sétima hipótese consideramos que a descrição do PR permite encontrar inconsistências. Assim, os detalhes fazemos uma relação mais precisa do que tem que ser revisado e olhado com detalhe.

De acordo com a hipótese e a averiguação realizada através do experimento, o resultado obtido através da medição aponta que a maioria das solicitações de pull (PRs) apresenta descrições breves e revisões mínimas. Além disso, parece não haver correlação direta entre o comprimento da descrição e o número de revisões, pois tanto as PRs curtas quanto as longas exibem variações em suas contagens de revisão de forma independente.

A oitava e última hipótese conta que quanto mais interações no PR, mais revisões são realizadas. Com uma maior discussão acerca da implementação, mais revisões tornam-se necessárias à medida em que o contexto é melhor compreendido.

Por fim, como última análise chegamos à conclusão que de maneira geral a relação entre o número de interações e a quantidade de revisões aparenta ser limitada. O número de revisões e o número de interações não aparentam uma relação de causalidade onde o aumento das interações não necessariamente aumenta o número de revisões.

