

Caracterizando a Atividade de Code Review no GitHub

**Bruno Gomes Ferreira, João Pedro Mairinque de Azevedo,
Matheus Vieira dos Santos, Marcio Lucas Machado Pereira**

¹ Instituto de Ciências Exatas e Informática (ICEI)
Pontifícia Universidade Católica de Minas Gerais
Engenharia de Software
Belo Horizonte – MG – Brazil

1. Introdução

A prática de Code Review é uma constante nos processos de desenvolvimento ágeis. Envolve a interação entre desenvolvedores e revisores para inspecionar o código antes de integrá-lo à base principal. Isso garante a qualidade do código, respeito dos padrões pré-estabelecidos pelo projeto e evita a inclusão de defeitos. No contexto de sistemas open source, especialmente aqueles hospedados no GitHub, as atividades de Code Review ocorrem por meio de Pull Requests (PRs). Antes de um código ser integrado à branch principal, é necessário criar uma solicitação de pull, que será avaliada e discutida por um colaborador do projeto. Ao final desse processo, a solicitação de merge pode ser aprovada ou rejeitada pelo revisor. [Chen et al. 2021] Além disso, práticas como Continuous Integration (CI) e Continuous Delivery (CD) são amplamente adotadas na Engenharia de Software, elas orientam os desenvolvedores a automatizar testes, builds e deployment com o auxílio de ferramentas específicas. Essa abordagem contribui para a eficiência e qualidade do processo de desenvolvimento.

Ao conduzir uma revisão de código, é essencial garantir que o código atenda a padrões elevados de qualidade, manutenibilidade e funcionalidade. [Github 2024] De acordo com a documentação oficial do GitHub para Code Review, os desenvolvedores ao revisar um código devem verificar se o design geral do código está sendo mantido, verificando se a arquitetura está alinhada com os objetivos do projeto em termos de usabilidade e experiência do usuário. Além disso, deve-se atentar a existência de fatores essenciais como testes unitários do que está sendo proposto, nomenclatura clara das variáveis e uma complexidade coerente com o que está sendo proposto.

Para a caracterização do Code Review traçou-se duas áreas de estudo nas quais o grupo criou questões de pesquisa: Em relação ao feedback final das revisões, ou seja, o status da PR que foi aberta, se ela foi integrada a base principal de código ou não. E também perguntas de pesquisa em relação ao número de revisões feitas em uma PR.

Em relação ao feedback final das revisões, traçou-se as seguintes perguntas de pesquisa:

- RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?
- RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?
- RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?
- RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?

Para os autores, como hipóteses informais, os repositórios com PRs grandes tendem a ter um feedback negativo visto que serão revisões mais criteriosas, pois normalmente aqueles que irão revisar são pessoas que tem um conhecimento maior do código. O tempo de revisão é proporcional ao critério imposto para revisar aquela PR, Pull Requests mais críticas terão um tempo de revisão maior. A descrição da PR é um fator que facilita a revisão, por isso tende a ter melhores feedbacks. PRs que tem uma maior interação costumam ter um melhor feedback final das revisões já que tem uma maior discussão entre os desenvolvedores

Para as questões relacionadas ao número de revisões, foram traçadas as seguintes perguntas:

- RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?
- RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?
- RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?
- RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?

PRs grandes tendem a ter uma quantidade menor de revisões pois será revisado principalmente por quem tem uma maior familiaridade com o projeto. O tempo de análise de uma PR é inversamente proporcional ao número de revisões, pois PRs que demandam maior tempo de revisão normalmente tem um menor engajamento. PRs com descrição detalhada tendem a atrair mais revisões, pois o revisor terá um contexto do motivo daquela PR ser feita. Finalmente, PRs com maior número de interações terão uma quantidade maior de revisões

Com o intuito de responder as perguntas de pesquisa, foram utilizadas métricas de qualidade para caracterizar os Code Reviews realizados na plataforma do GitHub. Para este trabalho, foram selecionadas as seguintes métricas de qualidade:

- Tamanho: número de arquivos; total de linhas adicionadas e removidas.
- Tempo de Análise: intervalo entre a criação do PR e a última atividade (fechamento ou merge).
- Descrição: número de caracteres do corpo de descrição do PR (na versão markdown).
- Interações: número de participantes; número de comentários.

Essas métricas serão correlacionadas com as perguntas de pesquisa realizadas inicialmente a fim de traçar as características dos Pull Requests além disso será feito um cálculo utilizando a correlação de Spearman para cada um deles, essa correlação é utilizada para analisar a associação entre duas variáveis quantitativas. A relação de Spearman consegue avaliar relações monótonas, sejam elas lineares ou não, diferente da correlação de Pearson que tem resultados melhores quando a relação é linear. A relação de Spearman por se tratar de uma correlação pode ser interpretado através de valores que variam de -1 à 1 [Shimakura 2006]. A figura 1, corresponde ao quadro de interpretação que será utilizado nessa pesquisa para identificar as correlações.

Valor de ρ (+ ou -)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Figure 1. Interpretação da correlação para dado o valor numérico

Pode-se observar que a correlação é forte quando se aproxima de 1 e fraca quando se aproxima de 0. Além disso os números positivos e negativos indicam que aquela correlação é positiva ou negativa.

2. Metodologia

Com o objetivo de obter características de qualidade de Pull Requests, os autores seguiram a metodologia de pesquisas quantitativas onde através dos valores numéricos obtidos, será possível determinar as respostas das perguntas de pesquisa. Para a obtenção dos dados necessários foi utilizado as seguintes ferramentas: Python e GraphQL para obter os 200 repositórios mais populares. Para obter as métricas dos repositórios. O passo a passo da estratégia utilizada pode ser resumida pelo diagrama de fluxo (Figura 1)

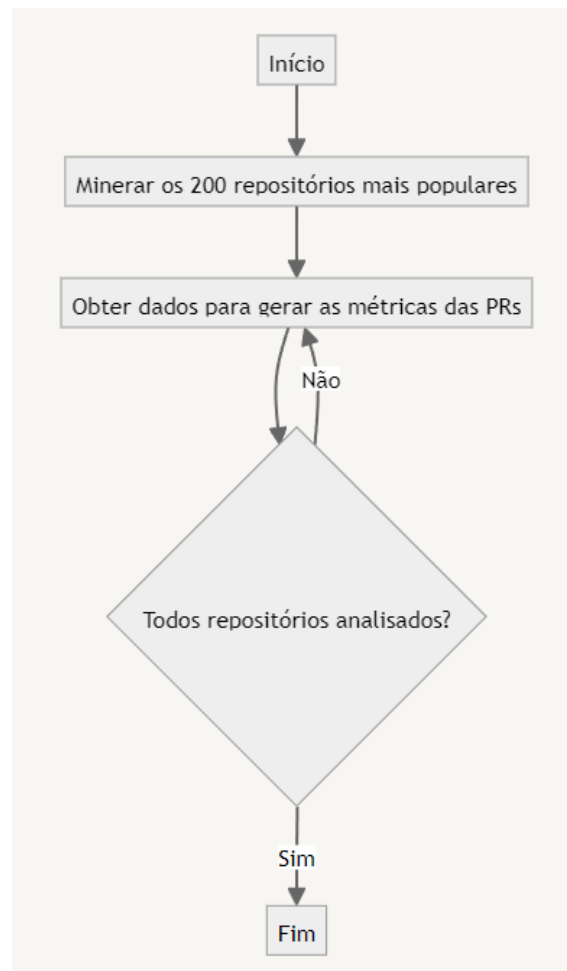


Figure 2. Diagrama de fluxo para obtenção dos resultados

Inicialmente, foram minerados os 200 repositórios mais populares no Github através do GraphQL. A linguagem de programação Python que é utilizada para desenvolver o script que fez a requisição da query na api do Github. Nessa query foi gerado um dataset em CSV contendo valores das métricas necessárias para a análise (tamanho, tempo de análise, descrição, interações) e suas relações com o status final do PR e o número total de revisões realizadas.

A requisição foi realizada de forma que nos assegurássemos que todos os 200 repositórios foram analisados, e que a obtenção de dados se repetisse até que a aquisição do objetivo fosse visível.

3. Resultados

Neste tópico será discutido os resultados obtidos através dos gráficos. Após obter os 200 repositórios, foi possível gerar gráficos para melhor visualização dos resultados e por meio disso avaliar as perguntas de pesquisa

3.1. Questão 1: Qual a relação entre o tamanho dos PRs e o feedback final das revisões?

Para essa questão traçou-se um gráfico do tipo Scatter Plot com o tamanho do PR por feedback final das revisões.

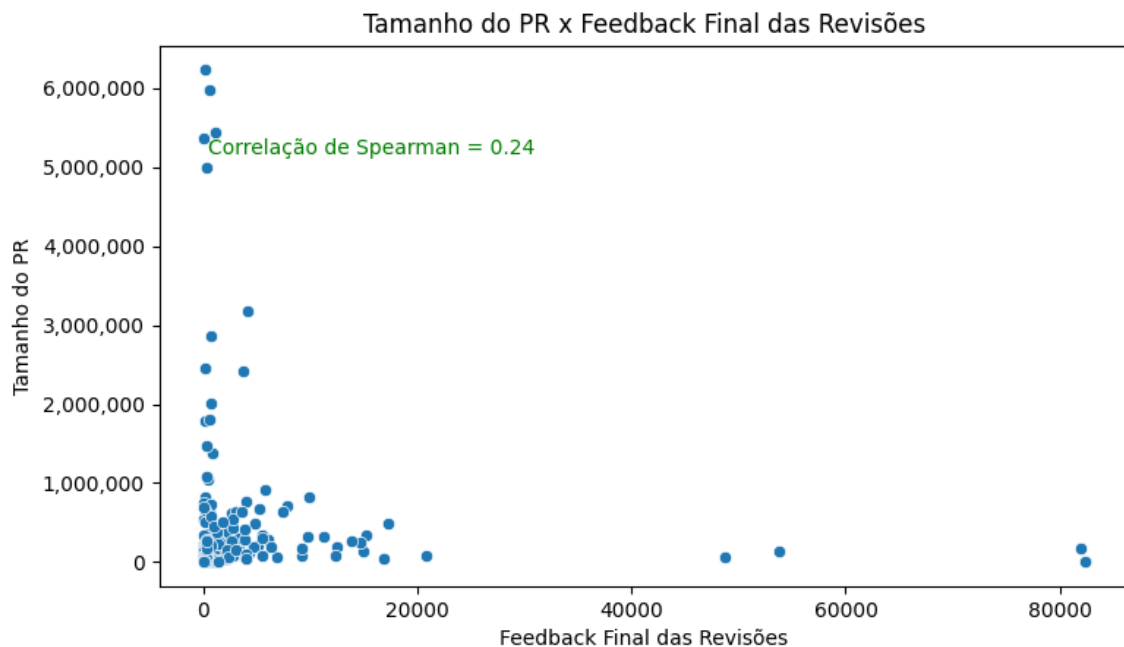


Figure 3. Tamanho do PR x Feedback Final das Revisões

Observa-se que a correlação de Spearman calculada para essa questão, deu cerca de 0,20 o que de acordo com o quadro 1, discutido no tópico de introdução, indica que é uma relação positiva porém baixa, a relação entre tamanho de PR com o Feedback existe porém baixa.

3.2. Questão 2: Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?

Para essa questão traçou-se um gráfico com o tamanho do PR por feedback final das revisões.

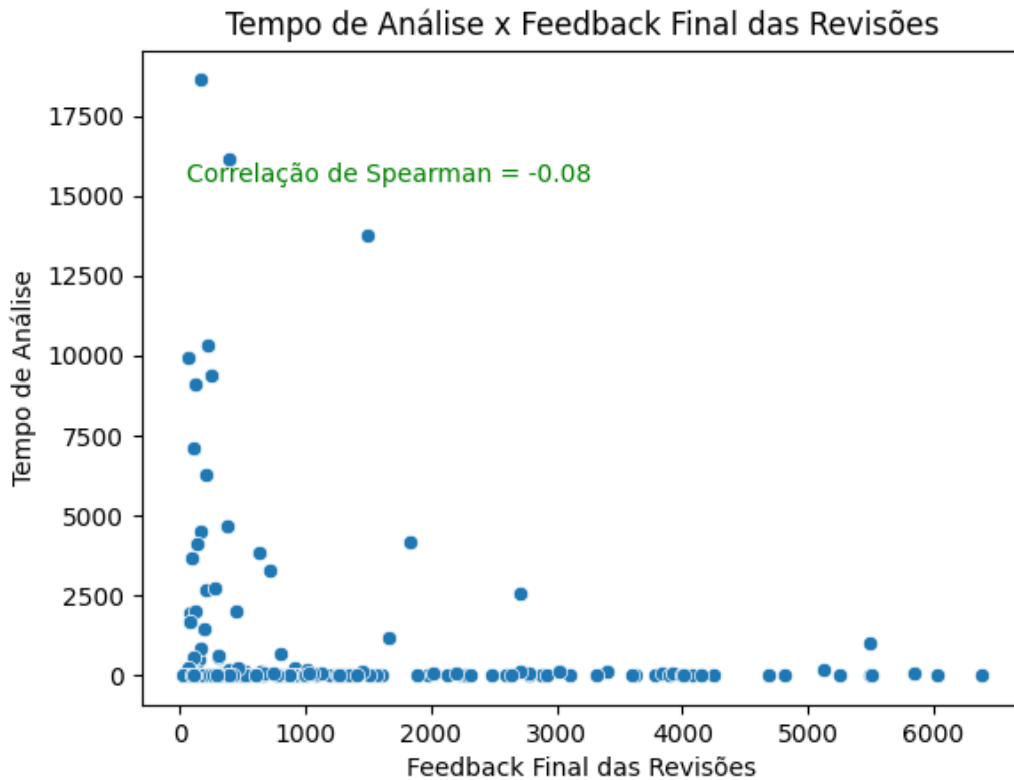


Figure 4. Tempo de análise x Feedback final das revisões

Nota-se que uma grande quantidade dos repositórios possui um tempo de análise de menos de 2500, apesar de que existe aproximadamente uma dezena de repositórios que superam essa métrica. Além disso, a quantidade de revisões realizadas é relativamente distribuída entre 0 e 6000 merges, porém é visível que a grande maioria dos repositórios se concentra na margem de 0 a 1000 menos de 2000 PRs concluídos. O coeficiente de Spearman é -0,08, indicando uma correlação negativa bem fraca entre essas duas variáveis.

3.3. Questão 3: Qual a relação entre a descrição dos PRs e o feedback final das revisões?

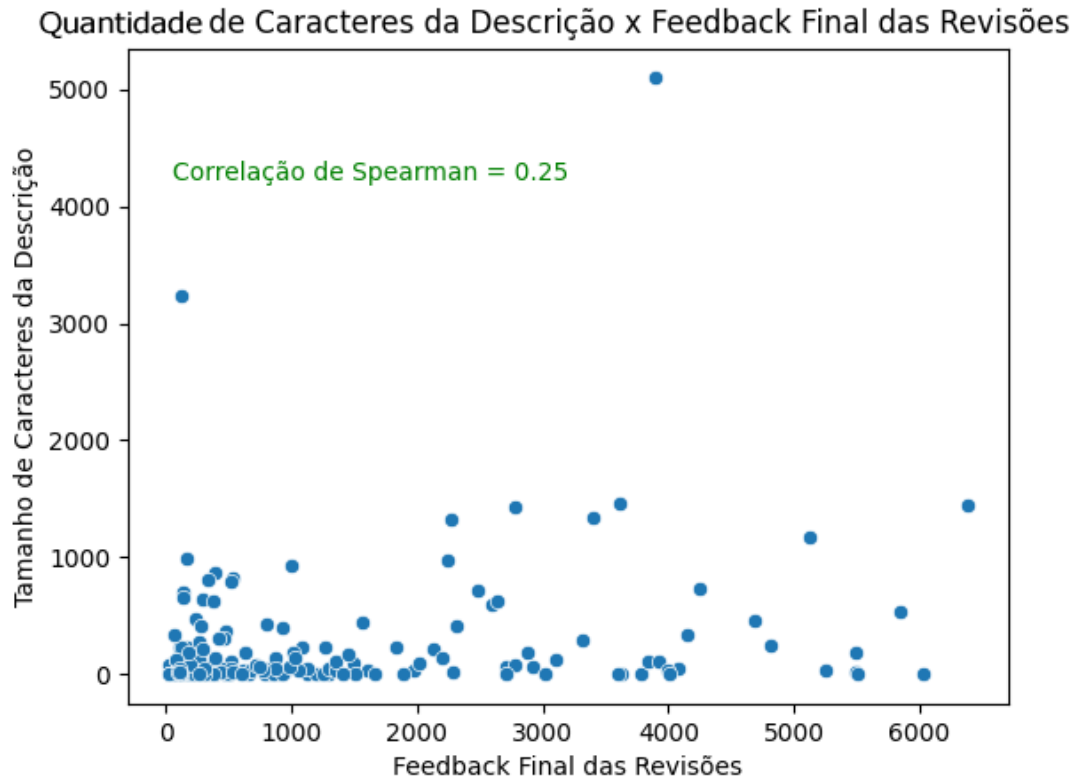


Figure 5. Quantidade de caracteres da descrição x Feedback final das revisões

Nota-se que a grande maioria dos repositórios compreende um tamanho de descrição com menos de 2000 caracteres, e possuem menos de 2000 PRs fechados. O coeficiente de Spearman calculado de 0,25 indica uma correlação positiva fraca entre essas duas variáveis.

3.4. Questão 4: Qual a relação entre as interações nos PRs e o feedback final das revisões?

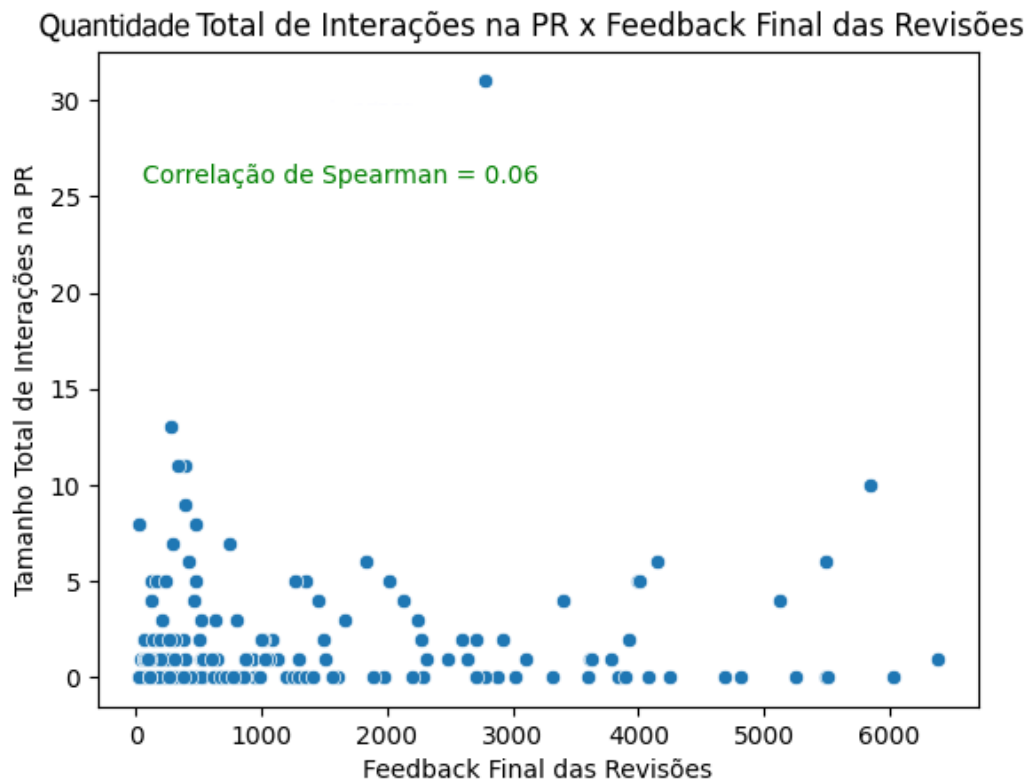


Figure 6. Quantidade total de interações x Feedback final das revisões

Nota-se que há uma concentração de repositórios com menos de 5 interações na PR e menos de 2000 feedbacks finais como mergeados. Apesar disso, uma quantidade considerável de repositórios está presente entre os 2000 e 6000 merges, e entre 3 e 15 interações na PR. O coeficiente de Spearman é 0,06 , indicando uma correlação positiva bem fraca entre essas duas variáveis.

3.5. Questão 5: Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?

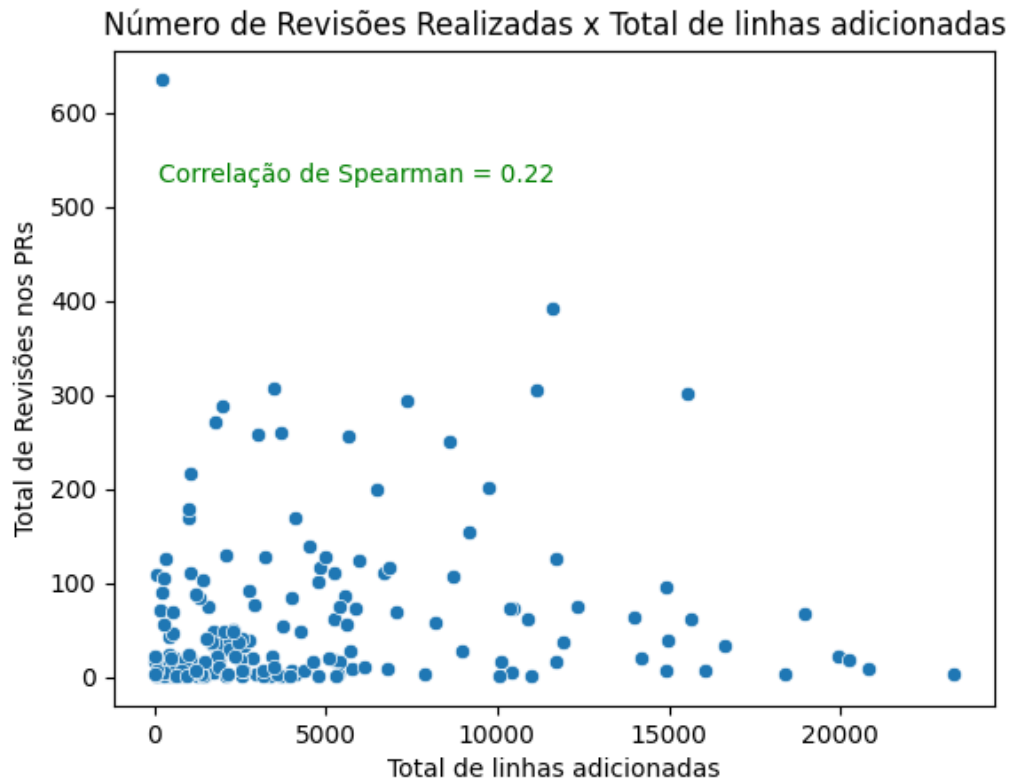


Figure 7. Número de revisões realizadas x Total de linhas adicionadas

Nota-se que uma grande quantidade dos repositórios possui um total de menos de 5000 linhas adicionadas, e menos de 100 revisões realizadas nos PRs. O coeficiente de Spearman é 0,22, indicando uma correlação positiva fraca entre essas duas variáveis.

3.6. Questão 6: Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?

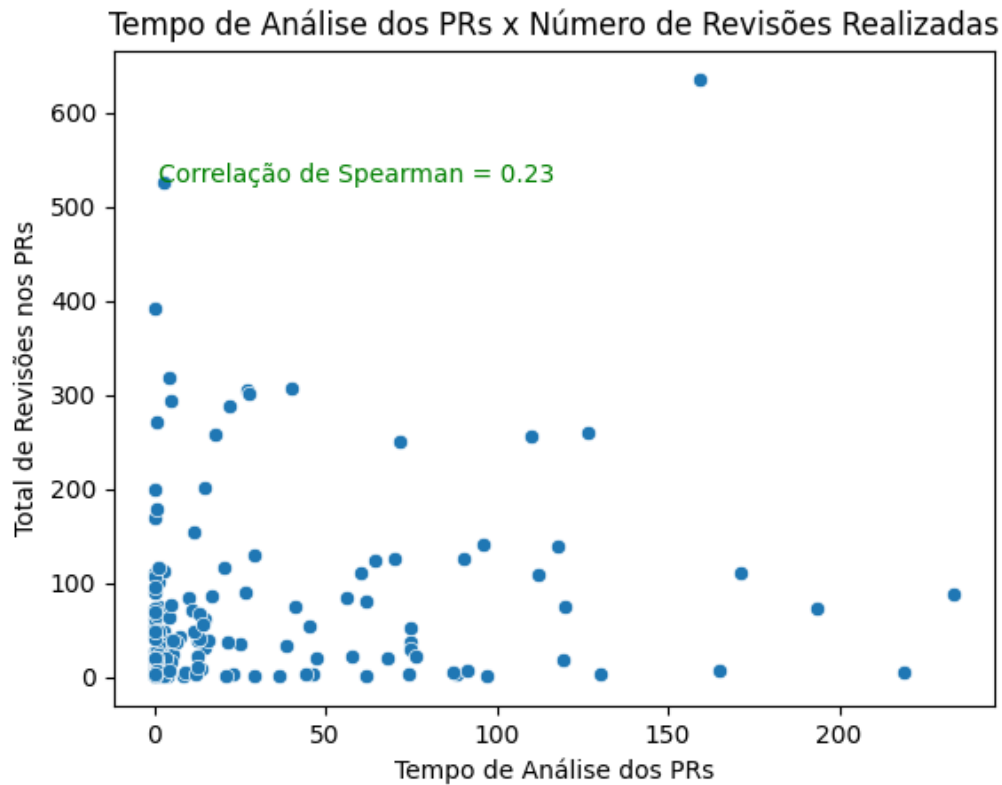


Figure 8. Tempo de análise dos PRs x Número de revisões realizadas

Nota-se que a grande maioria dos repositórios possui um total de revisões inferior a 200 e um tempo de análise inferior a 50. O coeficiente de Spearman calculado de 0,23 indica uma correlação positiva fraca entre essas duas variáveis.

3.7. Questão 7: Qual a relação entre a descrição dos PRs e o número de revisões realizadas?

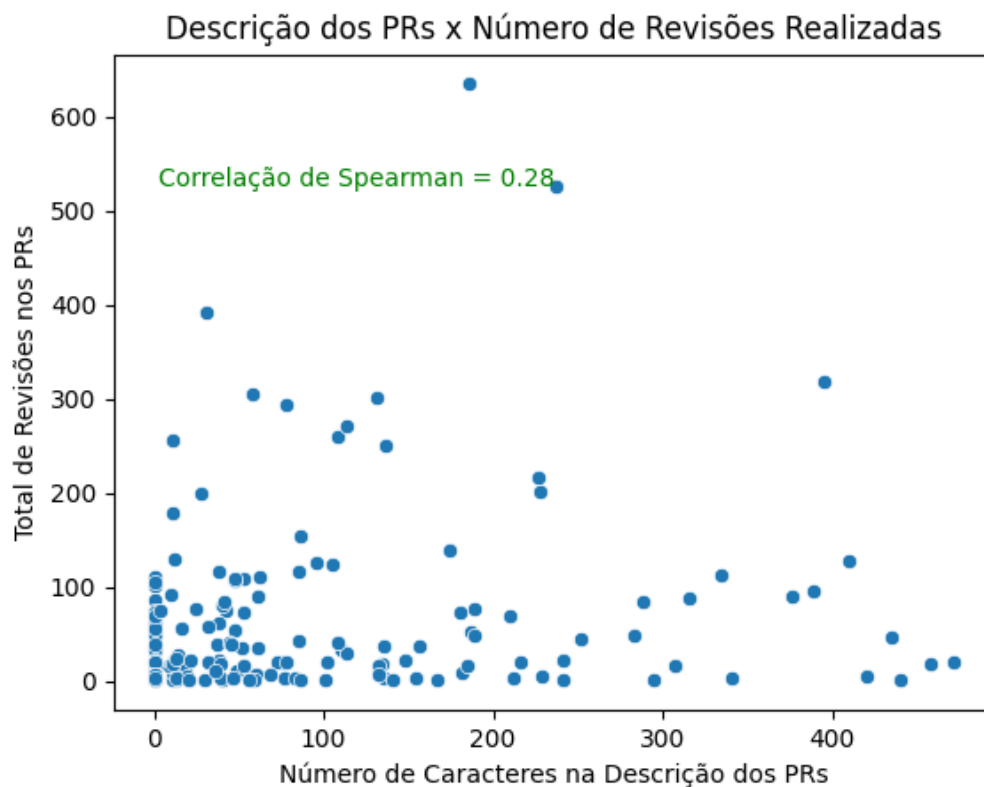
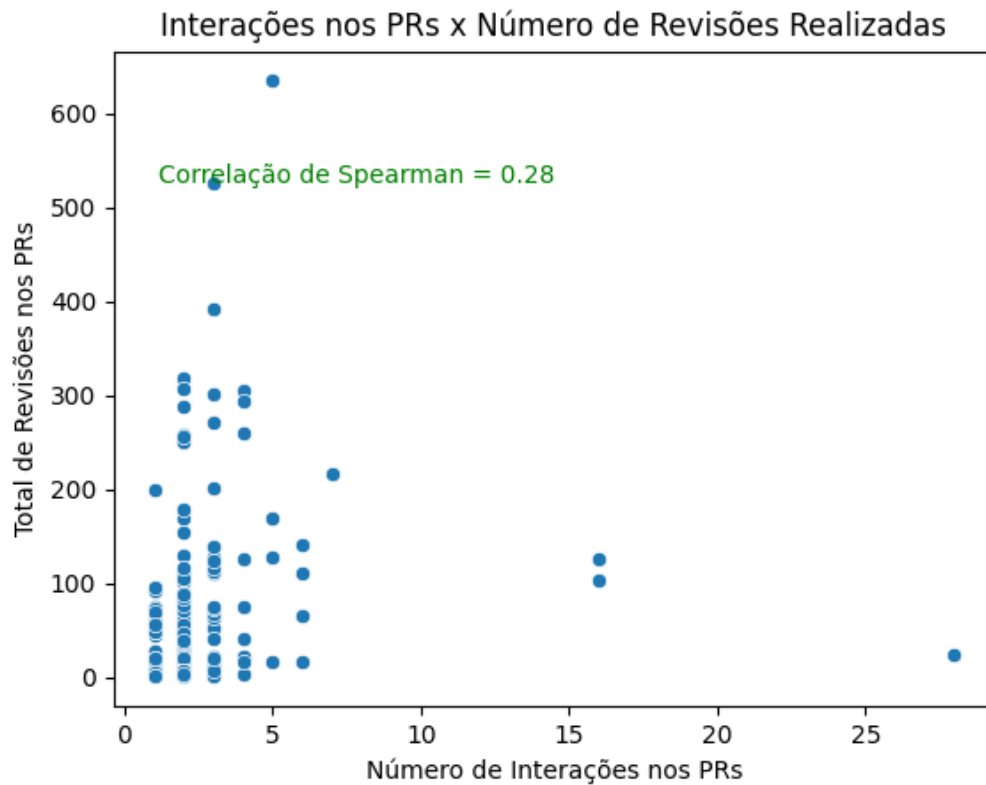


Figure 9. Descrição dos PRs x Número de revisões realizadas

Nota-se que uma grande quantidade dos repositórios possui um total de menos de 200 revisões, e menos de 100 caracteres nas descrições do PR (sendo que menos de 50 repositórios possuem um tamanho de texto maior que este). O coeficiente de Spearman é 0,28, indicando uma correlação positiva bem fraca entre essas duas variáveis.

3.8. Questão 8: Qual a relação entre as interações nos PRs e o número de revisões realizadas?



4. Discussão

Pode-se verificar que a maioria das hipóteses geradas no início deste relatório de fato ocorreram ao obter os dados dos 200 repositórios, porém baixa o suficiente ao ponto de não poder ser possível afirmar com total certeza que esse tipo de correlação é verídica. Neste tópico será discutido os resultados obtidos em comparação com o que foi inicialmente proposto pelos autores da pesquisa.

Iniciando pelo grupo de questões de pesquisa relacionados ao feedback das PRs, foi pressuposto que os sistemas populares teriam melhor feedback caso o tamanho das PRs fossem menores, o tempo de análise seria pequeno em PRs com um bom feedback, as descrições seriam detalhadas e interações constantes nesses PRs. Ao gerar o gráfico e consequentemente calcular o coeficiente de Spearman para verificar o quão forte eram as correlações, foram obtidos valores entre +0,20 e +0,39 o que indica que a correlação existe, mas é fraca. Isso não indica que esse tipo de correlação deve ser descartada, mas não é um valor expressivo ao ponto de se afirmar que as correlações possuem um alto grau de certeza.

Em relação ao segundo grupo de questões que era em relação ao número de revisões, foi observado que a distribuição do gráfico obtido foi bem similar ao que foi encontrado para o feedback final das PRs, o que também indica que há correlação existente, já que não foi um valor negativo, porém baixa.

5. Conclusão

Conclui-se que em todas as análises é visível uma correlação baixa positiva, e que o feedback final das revisões não possui tamanha dependência da descrição, tempo, interações, etc. Como a relação não é linear, o uso do cálculo da correlação de Spearman foi utilizada como a medida estatística que foi capaz de avaliar a relação entre as variáveis observadas. Dessa forma, a análise revelou que a prática de code review em repositórios populares são influenciadas pelas medidas analisadas, porém não é uma correlação forte.

References

- Chen, T., Zhang, Y., Chen, S., Wang, T., and Wu, Y. (2021). Let's supercharge the workflows: An empirical study of github actions. In *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, page 01–10.
- Github (2024). **What to look for in a code review.** <https://google.github.io/eng-practices/review/reviewer/looking-for.html>
Acesso em: 17/04/2024.
- Shimakura, S. E. (2006). **Interpretação do coeficiente de correlação.** <http://leg.ufpr.br/silvia/CE003/node74.html>
Acesso em: 22/04/2024.