

Laboratório de Experimentação de Software
Laboratório III - Caracterizando a Atividade de Code Review no GitHub
Grupo: Davi Santos; Rafael Duarte.

**Instituto de Informática e Ciências Exatas– Pontifícia Universidade de Minas Gerais
(PUC MINAS)**

Belo Horizonte – MG – Brasil

1. Introdução

Code review, é uma prática de desenvolvimento de software que consiste em ter outro desenvolvedor ou uma equipe de desenvolvedores analisando o código fonte escrito por um ou mais programadores. Sua finalidade é encontrar erros, bugs, falhas de segurança, otimizações e possíveis melhorias no código. A revisão de código pode ser realizada de diversas maneiras, incluindo revisão manual, automática ou uma combinação de ambas. Ao final, a solicitação de merge pode ser aprovada ou rejeitada pelo revisor. O objetivo deste laboratório é analisar a atividade de code review desenvolvida em repositórios populares do GitHub, identificando variáveis que influenciam no merge de um PR, sob a perspectiva de desenvolvedores que submetem código aos repositórios selecionados.

2. Hipóteses Informais

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

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

Pode-se existir uma correlação forte entre as variáveis. PRs com um tamanho menor tendem a ter um feedback mais rápido e positivo. PRs menores tendem a tornar a tarefa de revisão mais gerenciável, melhorando assim a qualidade do feedback final.

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

Pode-se encontrar uma correlação fraca entre as variáveis. Existem outros fatores como a complexidade do Pr em questão. Entretanto em algumas situações PRs que são analisados com um menor tempo, tendem a ter um feedback final mais positivo e uma taxa de aprovação mais alta.

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

Uma descrição clara e detalhada de um PR pode ajudar os revisores a entender melhor o objetivo e o contexto do código que está sendo enviado, o que, por sua vez, pode resultar em um feedback final positivo. Portanto existiria uma correlação forte entre a descrição e o feedback final.

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

Quanto mais interações ocorrerem durante o processo de revisão, maior é a probabilidade de o feedback final ser útil e preciso. As interações permitem que os desenvolvedores e revisores esclareçam dúvidas, discutam alternativas e expliquem suas sugestões de forma mais detalhada.

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

Identifica-se uma relação inversa entre o tamanho dos PRs e o número de reviews realizadas. PRs menores são geralmente mais fáceis de entender e revisar, permitindo que os revisores possam revisá-los de forma mais rápida e eficiente.

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

Existe uma correlação forte entre as variáveis, pois quanto mais um PRs fica aberto para revisão, maior é o número de revisões realizadas. À medida que o tempo passa, mais revisores têm a oportunidade de examinar o PRs e fornecer feedback.

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

Existe relação entre a descrição dos PRs e o número de reviews realizadas. Uma descrição clara e detalhada do PR pode ajudar a atrair mais revisores e aumentar o número de reviews realizadas. Permitindo que os revisores entendam o contexto das alterações e forneçam feedback mais específico e preciso.

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

Há uma relação forte entre as interações e o número de reviews realizadas. PRs com mais participantes e mais comentários tendem a receber mais revisões, há uma maior chance de que diferentes perspectivas sejam apresentadas e mais erros ou problemas sejam identificados.

3. Metodologia

Este estudo analisou os repositórios mais populares do GitHub, observando aqueles com pelo menos 100 estrelas. A partir deste critério foram selecionados 200, de modo que cada um apresentasse pelo menos 100 PRs, sendo a soma dos status *Merged* ou *Closed*. Outra restrição adotada é que estes PRs possuam ao menos uma revisão. Para remover do *dataset* os PRs que foram revisados de forma automática, foram filtrados aqueles que a revisão durou pelo menos uma hora.

Para a coleta de dados, foi utilizada a *API GraphQL* do *GitHub* além de *scripts* Python, que permitiram a obtenção de informações para a geração do *dataset*. Esses dados foram salvos em um arquivo CSV para posterior análise.

Em relação aos dados dos PRs, para a construção dos gráficos: tamanho por *state*, tempo por *state*, descrição por *state* e interação por *state*, criou-se *boxplot* com t-test. Esta é uma técnica estatística que pode ser utilizada para analisar um conjunto de dados que contém duas variáveis distintas, neste caso o *state* é definido por *Merged* ou *Closed*. O t-test é executado para determinar se as diferenças entre as médias dos grupos são estatisticamente significativas.

Para geração dos gráficos: tamanho por número de reviews, tempo por número de reviews, descrição por número de reviews e interação por número de reviews, os dados foram submetidos a uma correlação simples.

Os resultados finais foram utilizados para gerar *scatterplots*, *heatmap* e *t-teste*, utilizando um script Python que usou as bibliotecas *Pandas*, *Matplotlib* e *Scipy*.

4. Apresentação dos resultados

Este tópico descreve os resultados dos gráficos obtidos através da metodologia apresentada.

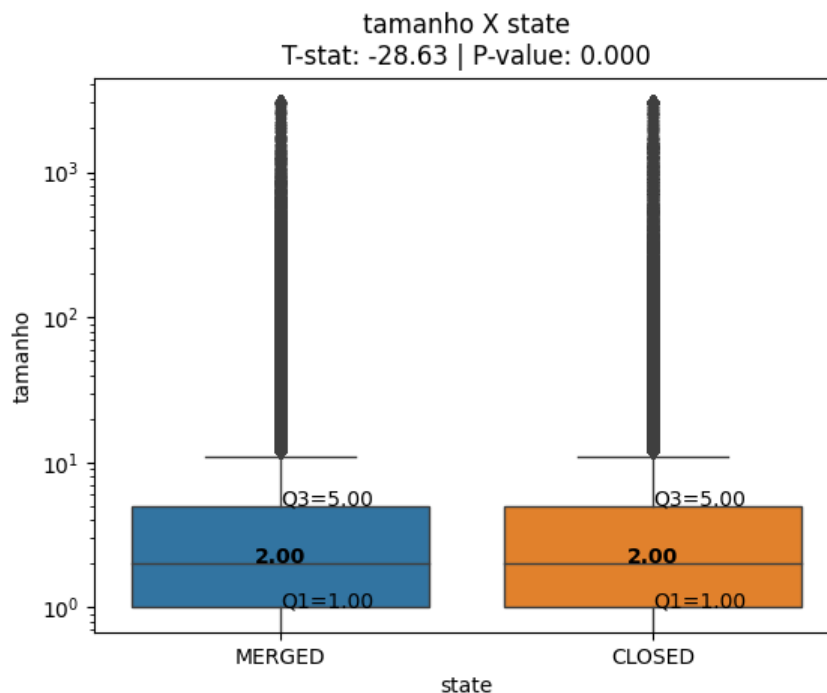


Figura 1 – Tamanho dos PRs e a distribuição por *Merged* e *Closed*

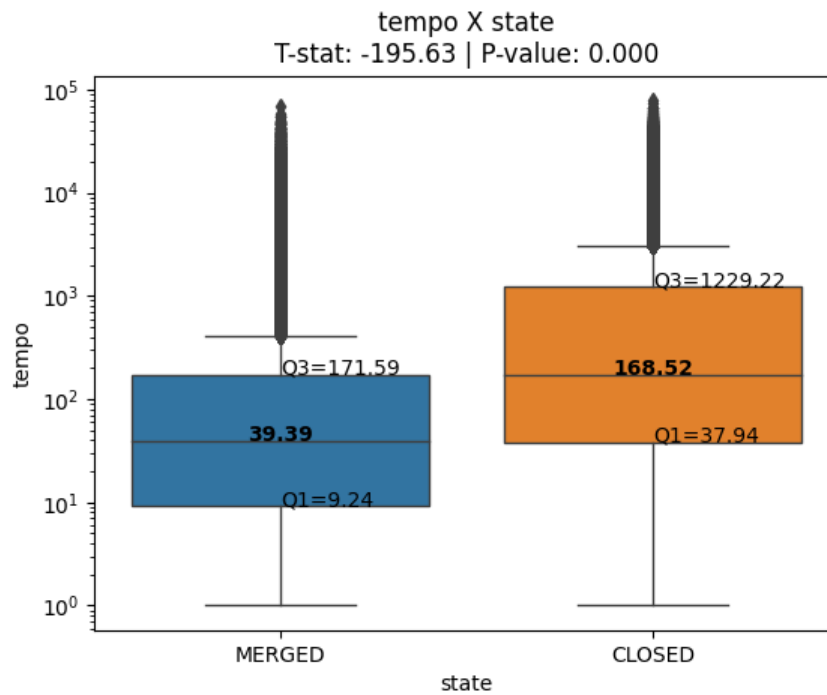


Figura 2 – Tempo dos PRs e a distribuição por *Merged* e *Closed*

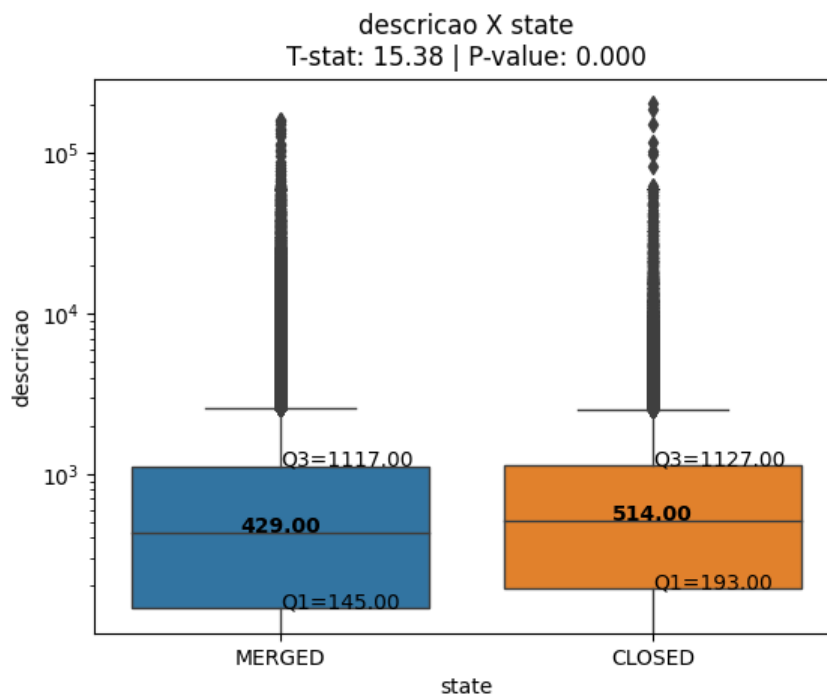


Figura 3 – Descrição dos PRs e a distribuição por *Merged* e *Closed*

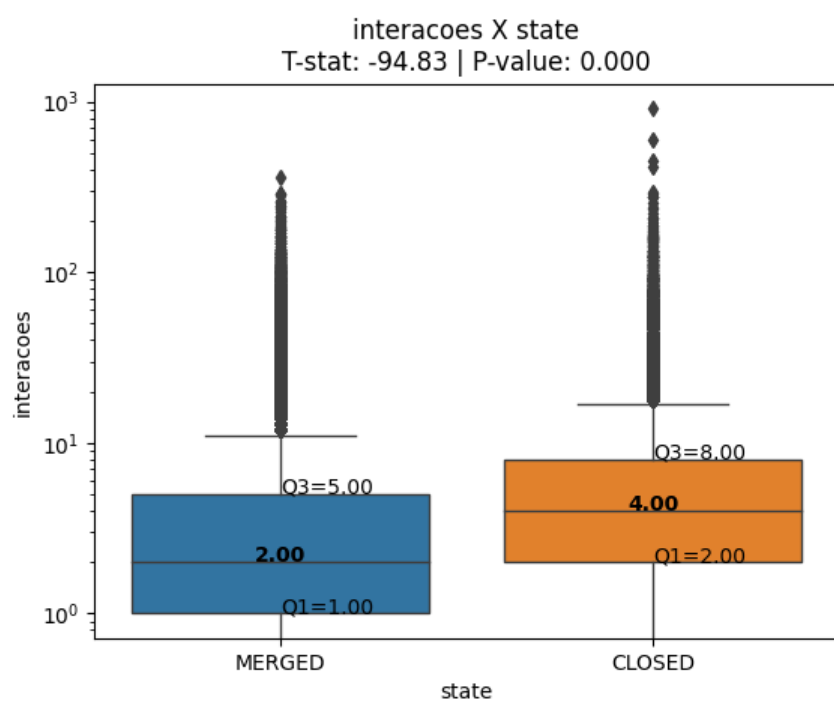


Figura 4 – Interações dos PRs e a distribuição por *Merged* e *Closed*

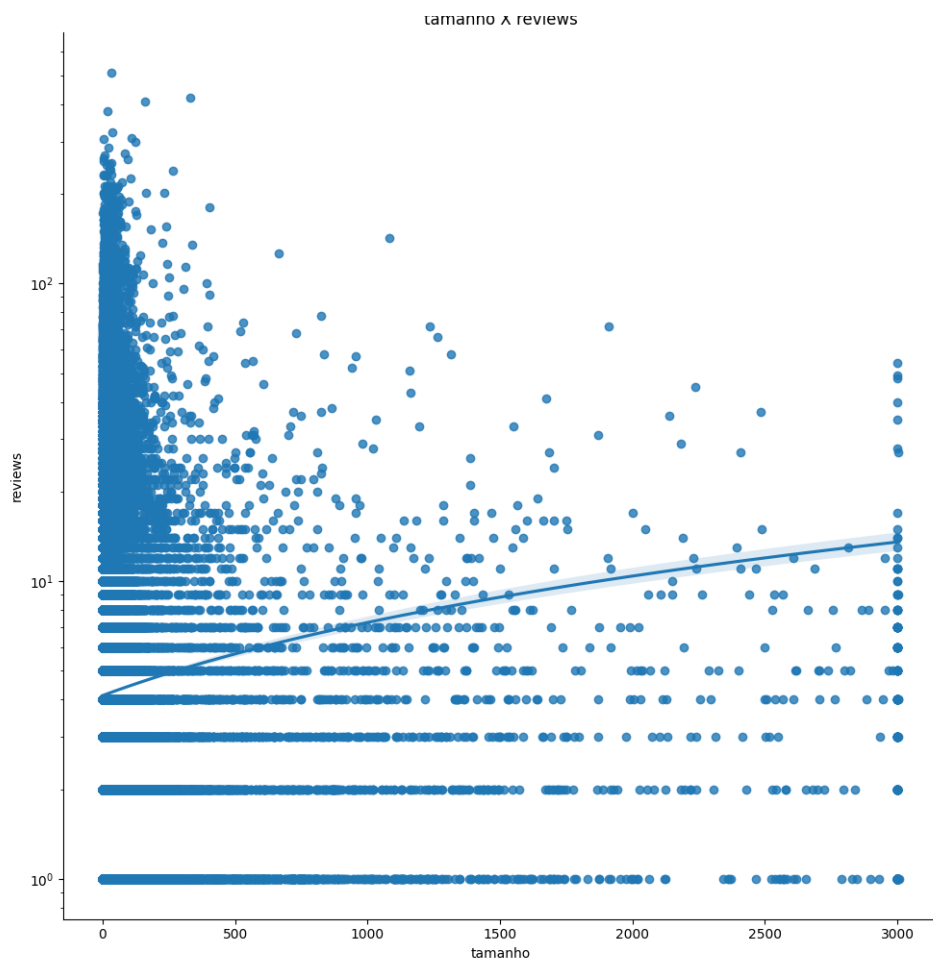


Figura 5 – Tamanho dos PRs correlacionando com o número de reviews

Na figura 5, pode-se observar uma fraca correlação entre o tamanho dos PRs e o número de reviews. O coeficiente de correlação é igual a 0,29. A partir dos dados analisados, é possível concluir que o tamanho dos PRs não influenciaram nos números de reviews.

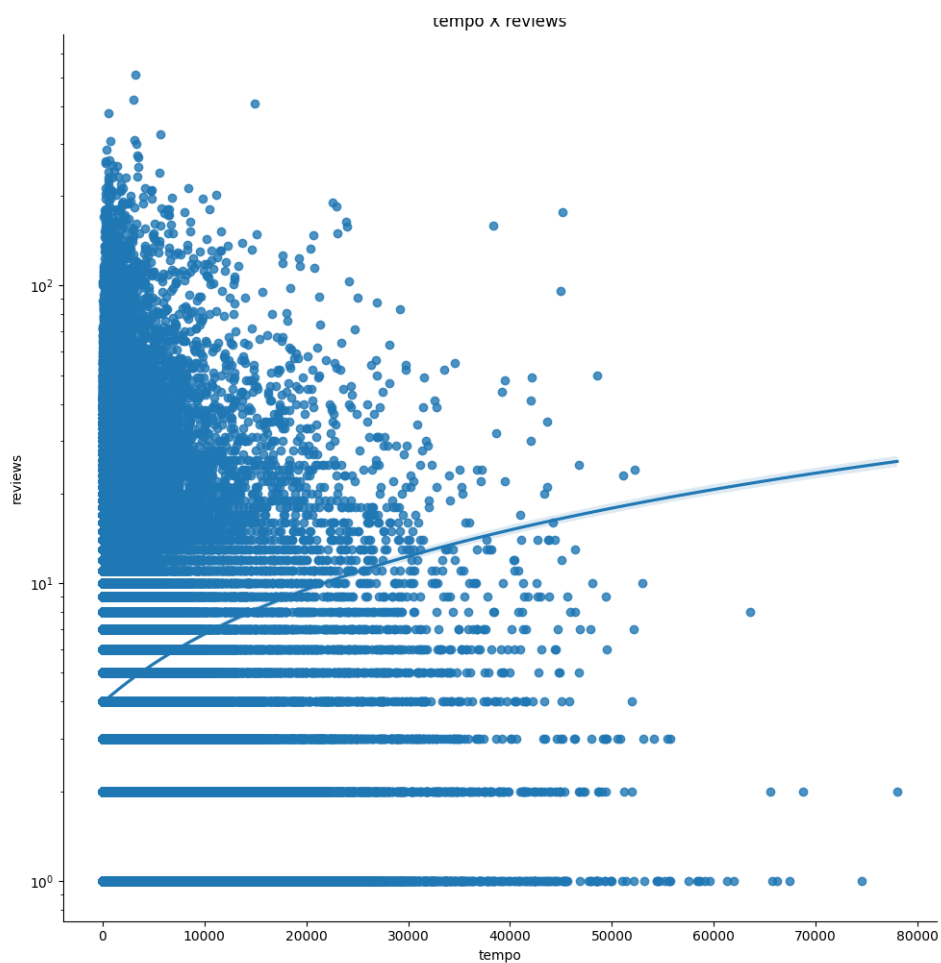


Figura 6 – Tempo dos PRs correlacionando com o número de reviews

Na figura 6, notou-se uma fraca correlação entre o tempo de análise dos PRs e o número de reviews. O coeficiente de correlação é igual a 0,34. A partir dos dados analisados, é possível concluir que o tempo de análise dos PRs não influenciaram nos números de reviews.

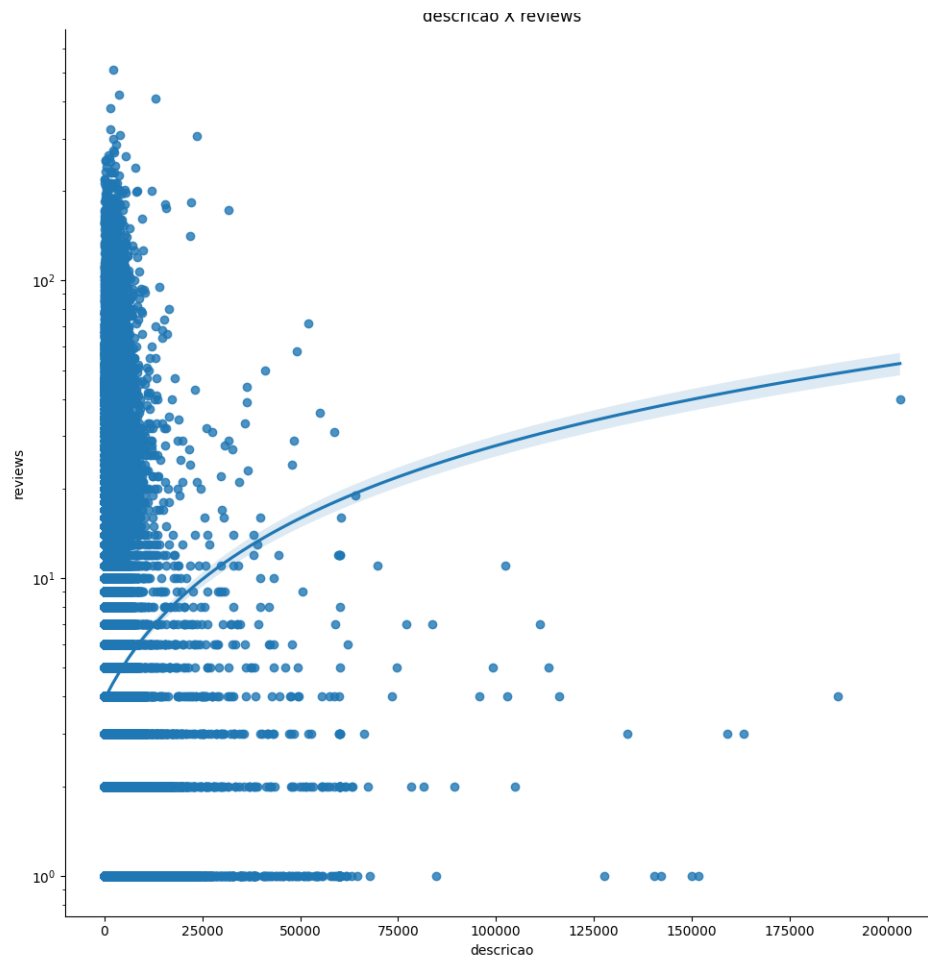


Figura 7 – Descrição dos PRs correlacionando com o número de reviews

Na figura 7, notou-se uma correlação moderada entre as descrições dos PRs e o número de reviews. O coeficiente de correlação é igual a 0,15. A partir dos dados analisados, é possível concluir que descrição dos PRs podem influenciar nos números de reviews.

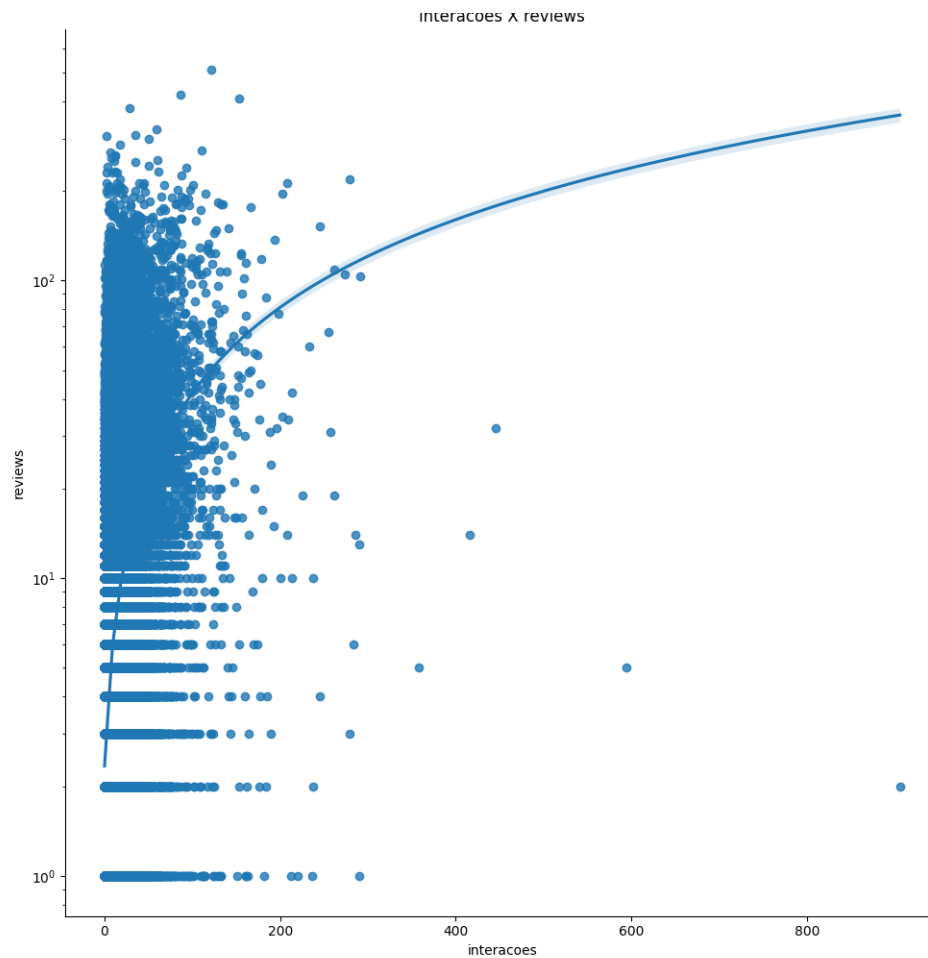


Figura 8 – Interações dos PRs correlacionando com o número de reviews.

Na figura 8, pode-se observar uma fraca correlação entre as interações dos PRs e o número de reviews. O coeficiente de correlação é igual a 0,38. A partir dos dados analisados, é possível concluir que o tamanho dos PRs não influenciaram nos números de reviews.

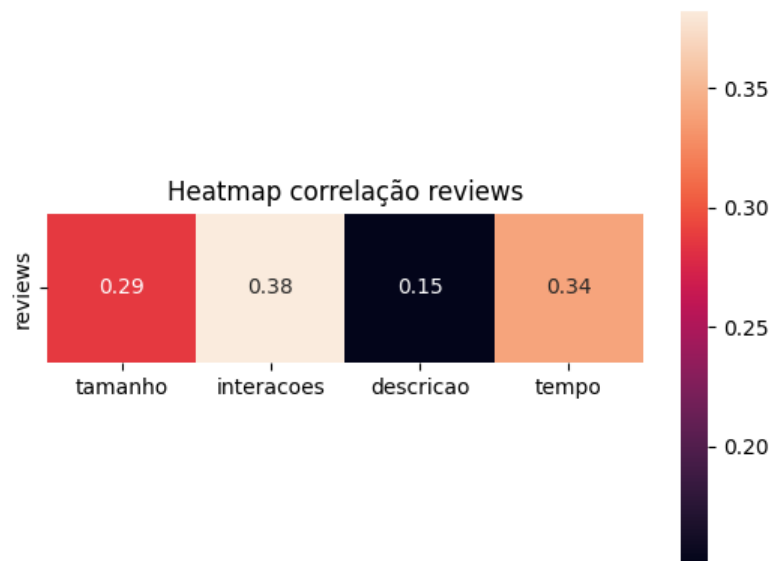


Figura 9 – Heatmap correlacionando número de reviews com tamanho, interações, descrição e tempo.

5. Discussão

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

Através dos dados analisados, observou-se que existe uma relação entre o tamanho do PR e o feedback final. Nos gráficos obtidos, nota-se uma mediana igual a 2 para as duas variáveis (Merged e Closed). Outra informação importante é o valor t de $-28,63$ e valor p de $0,000$. Com estes dados pode-se concluir que um valor negativo t, indica que a média da amostra "merged" é menor do que a média da amostra "closed", ou seja, PRs menores são mais propensas a serem fundidas do que PRs maiores, ou que PRs maiores são mais propensas a serem fechadas.

Em relação a hipótese, os dados confirmaram a ideia inicial onde um PRs menores tendem a ter um feedback mais rápido e positivo, tornando a tarefa mais gerenciável.

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

A partir da análise dos dados, identificou-se relação entre o tempo de análise das PRs e o feedback final das revisões. Contrariando a hipótese inicial, onde apontou uma correlação fraca entre as variáveis.

PRs merged tiveram mediana de $39,39$, enquanto PRs closed tiveram mediana de $168,52$. O teste t foi igual a $-195,63$, e o valor p igual a $0,000$. Com estes valores, pode-se

concluir que o tempo de análise das PRs está relacionado ao feedback final em repositórios populares do GitHub. A diferença estatisticamente significativa no tempo médio de análise das PRs entre as duas amostras (*merged* e *closed*) sugere que as PRs que são fechadas levam significativamente mais tempo para serem analisadas do que as PRs que são fundidas.

Isso pode indicar que PRs que levam mais tempo para serem analisadas são mais propensas a serem fechadas, possivelmente devido à complexidade do código ou ao número de problemas identificados durante a análise. Por outro lado, PRs que são analisadas rapidamente e fundidas indicam que o código é mais simples e menos problemático.

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

Com base nos resultados obtidos, pode-se observar que existe uma forte relação entre a descrição do PR e o feedback final da revisão. PRs *merged* tiveram mediana de 429 caracteres, enquanto PRs *closed* tiveram mediana de 514 caracteres. Além disso, o teste t mostrou que há diferença estatisticamente significativa entre as descrições de PRs *merged* e *closed*, com valor t de 15,38 e valor p de 0,000, indicando que os valores são pouco prováveis terem sido obtidos por acaso.

Embora se possa especular que uma diferença na descrição possa afetar o feedback final das revisões, deve-se enfatizar que existem outros fatores que podem estar envolvidos nesse processo. Por exemplo, a qualidade do código, a relevância da *feature* ou correção desenvolvida, a experiência dos revisores e outros fatores podem desempenhar um papel mais influente no feedback final.

Portanto é necessário considerar outros fatores envolvidos no processo de code review além de simplesmente o tamanho. Além disso, é importante destacar que a relação encontrada não implica necessariamente que os PRs com maior descrição tendem a ter mais aprovações que os com menor descrição conforme foi formulado na hipótese inicial, mas sim que há uma diferença estatisticamente significativa entre as descrições dos PRs com status *merged* e *closed*.

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

Os resultados obtidos indicam que existe uma relação considerável entre as interações nos PRs e o feedback final das revisões. Constatou-se que os PRs com state *merged* apresentam uma mediana menor de interações em comparação aos PRs com state *closed*, sendo a mediana dos PR *closed* o dobro dos PR *merged*. O valor negativo do valor t indica que essa relação é estatisticamente significativa.

Esses resultados não são capazes de provar a hipótese inicial levantada de que a interação pode levar a um feedback mais útil e preciso. Uma vez de que, embora os PR *closed* possuam uma mediana maior de interações em relação aos PRs *merged* e os testes estatísticos mostrarem uma relação entre o número de interações e o status do PR, não é possível estabelecer qual tipo de relação é essa, se mais interações fazem com que um *Pull Request* seja *mergeado* ou fechado. Além disso, também não se leva em consideração a quantidade de alterações que foram feitas no intervalo de tempo entre a abertura do PR e seu fechamento.

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

Durante as hipóteses informais, levantou-se a possibilidade de uma relação inversa entre o tamanho dos PRs e o número de reviews realizadas., onde PRs menores são geralmente mais fáceis de entender e revisar, de forma que os revisores iriam revisá-los rapidamente. Entretanto não foi encontrado correlação entre os dados analisados.

Pode-se concluir que o tamanho de uma PRs não é necessariamente um indicador confiável do número de revisões que ela recebe. Isso sugere que outros fatores podem influenciar a quantidade de revisões que uma PRs recebe, como a complexidade do código, a qualidade do código e a experiência dos revisores.

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

Inicialmente, levantou-se a hipótese de existir uma correlação forte entre as variáveis, imagina-se que quanto mais um PRs fica aberto para revisão, maior é o número de revisões realizadas. Durante análise dos dados foi encontrado uma fraca correlação entre as variáveis.

Conclui-se que o tempo de análise de uma PRs não é necessariamente um indicador confiável do número de revisões que ela recebe. Outros fatores podem influenciar a quantidade de revisões que uma PRs recebe, como a complexidade do código, a qualidade do código e a experiência dos revisores.

Destaca-se que a fraca correlação identificada pode ter implicações na prática de revisão de código em equipes de desenvolvimento. Se o tempo de análise da PRs não for um fator determinante na quantidade de revisões necessárias, os revisores devem estar atentos a outros aspectos do código e do processo de desenvolvimento ao decidir a quantidade de revisões necessárias para garantir a qualidade do código. Isso pode incluir avaliar a complexidade do código, a qualidade da documentação e testes, bem como a experiência dos desenvolvedores que criaram a PRs.

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

Durante análise dos resultados e geração dos gráficos identificou-se uma correlação moderada entre a descrição dos PRs e o número de reviews realizadas. Reforçando que uma boa descrição da PRs pode ajudar a atrair a atenção dos revisores e aumentar a probabilidade de que a PRs seja revisada com mais atenção e com mais revisões.

Destaca-se que uma descrição clara e precisa, pode fornecer informações importantes sobre a finalidade da PR e sobre as mudanças realizadas no código, o que pode facilitar a compreensão e a revisão do código pelos revisores. Por outro lado, uma descrição incompleta ou pouco clara pode levar a mal-entendidos e dificultar a revisão da PRs.

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

Através dos dados analisados, observou-se que as interações em uma *Pull Request* que podem ser definidas como comentários e revisões adicionais solicitadas, não são necessariamente um indicador confiável do número de revisões que ela recebe.

Contrariando as hipóteses iniciais, onde acreditava-se que PRs com mais participantes e mais comentários tendiam a receber mais revisões, pois teriam há uma maior chance de que diferentes perspectivas sejam apresentadas e mais erros ou problemas sejam identificados.

As interações na PRs não devem ser usadas como a única métrica para determinar a qualidade da revisão de código e pode haver outros fatores importantes que devem ser considerados pelos revisores, como por exemplo a complexidade do código. No entanto, as interações ainda são uma parte importante do processo de revisão e podem ajudar a melhorar a qualidade do código e a comunicação entre os desenvolvedores e revisores.