

Lab 03

Caracterizando a Atividade de Code Review no GitHub

Eduardo Bandeira de Melo Guimarães

Tecnologias Utilizadas

- Python
 - Bibliotecas: Requests; CSV; DateTime; Pandas; Matplotlib; Seaborn; ScyPy;
- API GraphQL do GitHub
- Visual Studio Code

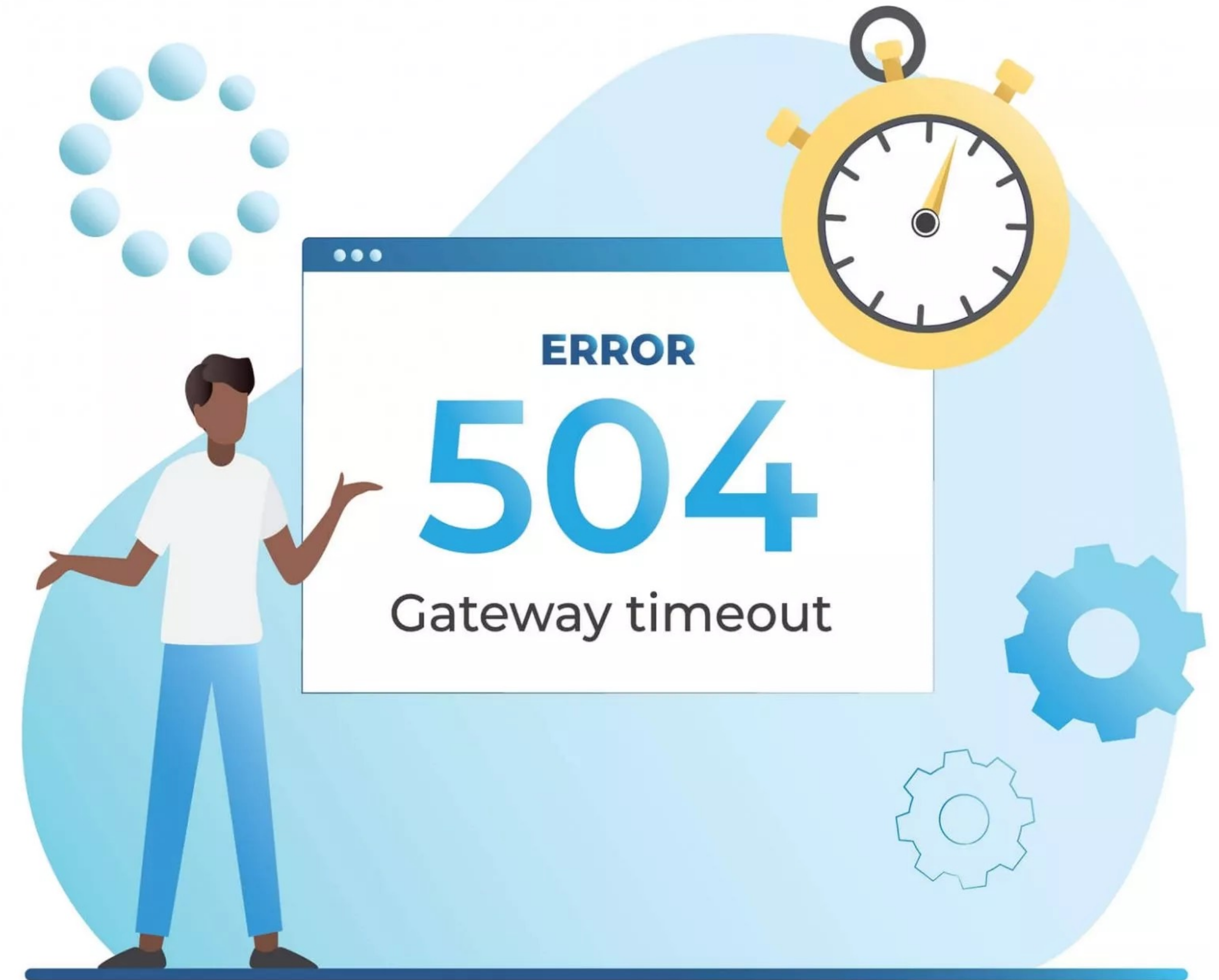


+



Diferenças em relação ao último Lab

- Necessidade de tornar o código mais robusto
- Esperar e repetir requisições em caso de erros
 - Timeout
 - Resposta nula
 - Resposta com erro (`{"errors": "..."}`)
 - Outros códigos de erro
- Necessidade de descartar alguns repositórios (8) por falhas persistentes



Questões de Pesquisa

- Qual é a relação entre o tamanho dos PR's e o Feedback final das revisões?
- Qual é a relação entre o tempo de análise dos PR's e o Feedback final das revisões?
- Qual é a relação entre a descrição dos PR's e o Feedback final das revisões?
- Qual é a relação entre as interações nos PR's e o Feedback final das revisões?

Métricas Utilizadas

- Tamanho dos PR's:
 - Número de arquivos e total de linhas modificadas (adicionadas + deletadas)
- Tempo de análise dos PR's:
 - Intervalo entre a criação e o fechamento do PR
- Descrição:
 - Número de caracteres do corpo da descrição do PR
- Interações:
 - Número de participantes e número de comentários

Questão de Pesquisa 01

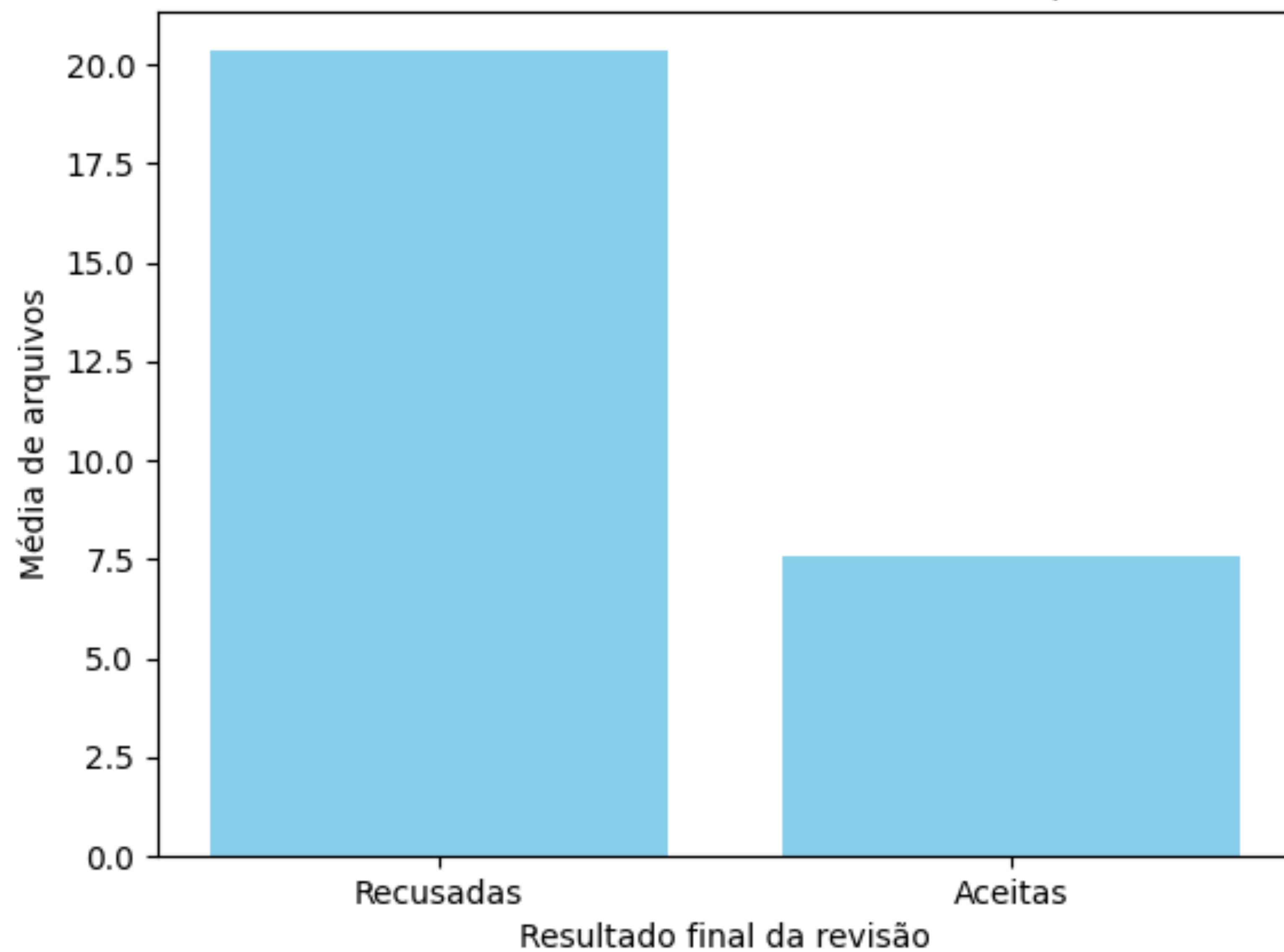
Hipótese:

Não existe relação entre o tamanho do PR e o resultado final

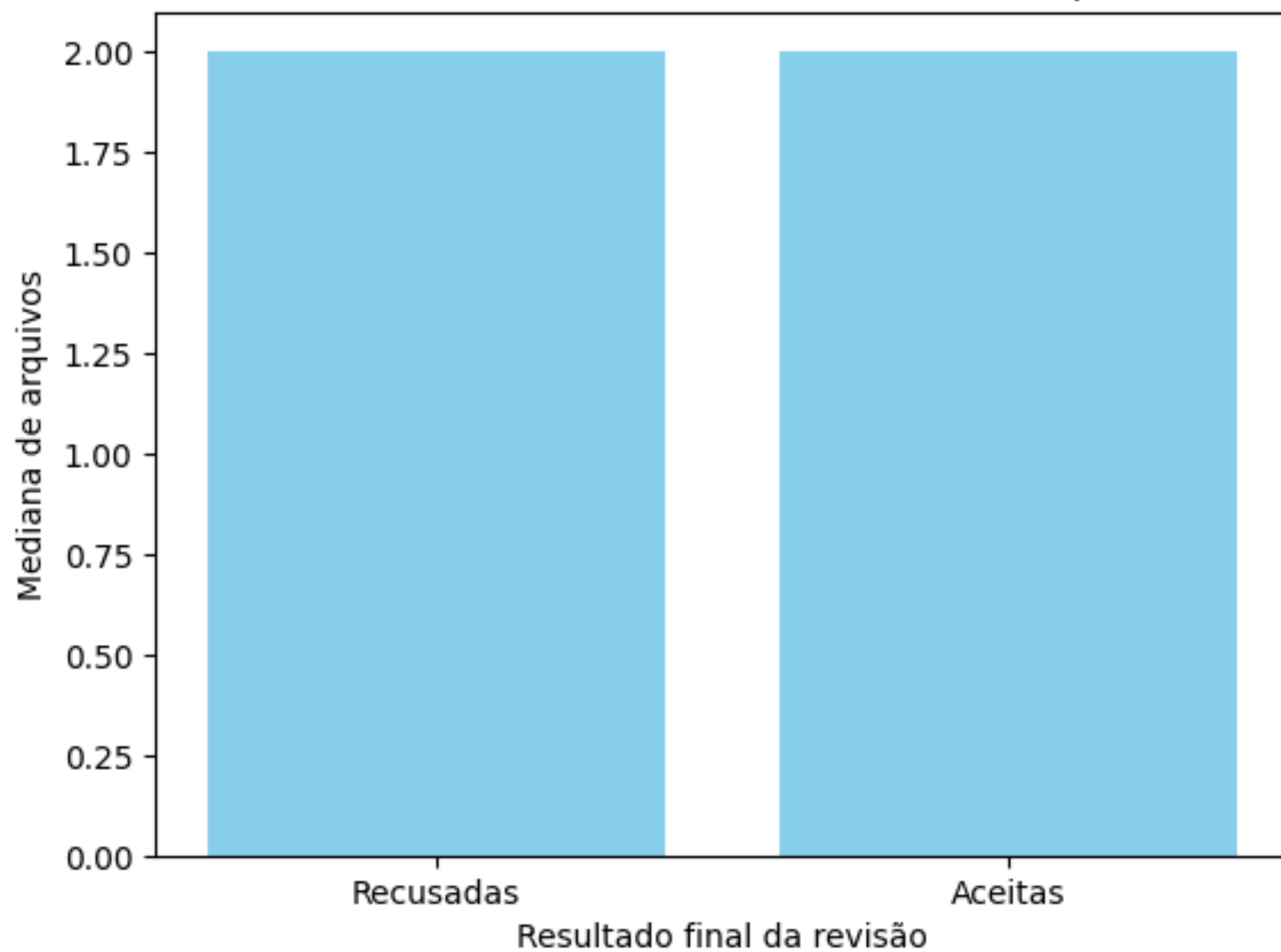
Justificativa:

Apenas os resultados das PR's são levadas em consideração
para aceitá-las ou rejeitá-las

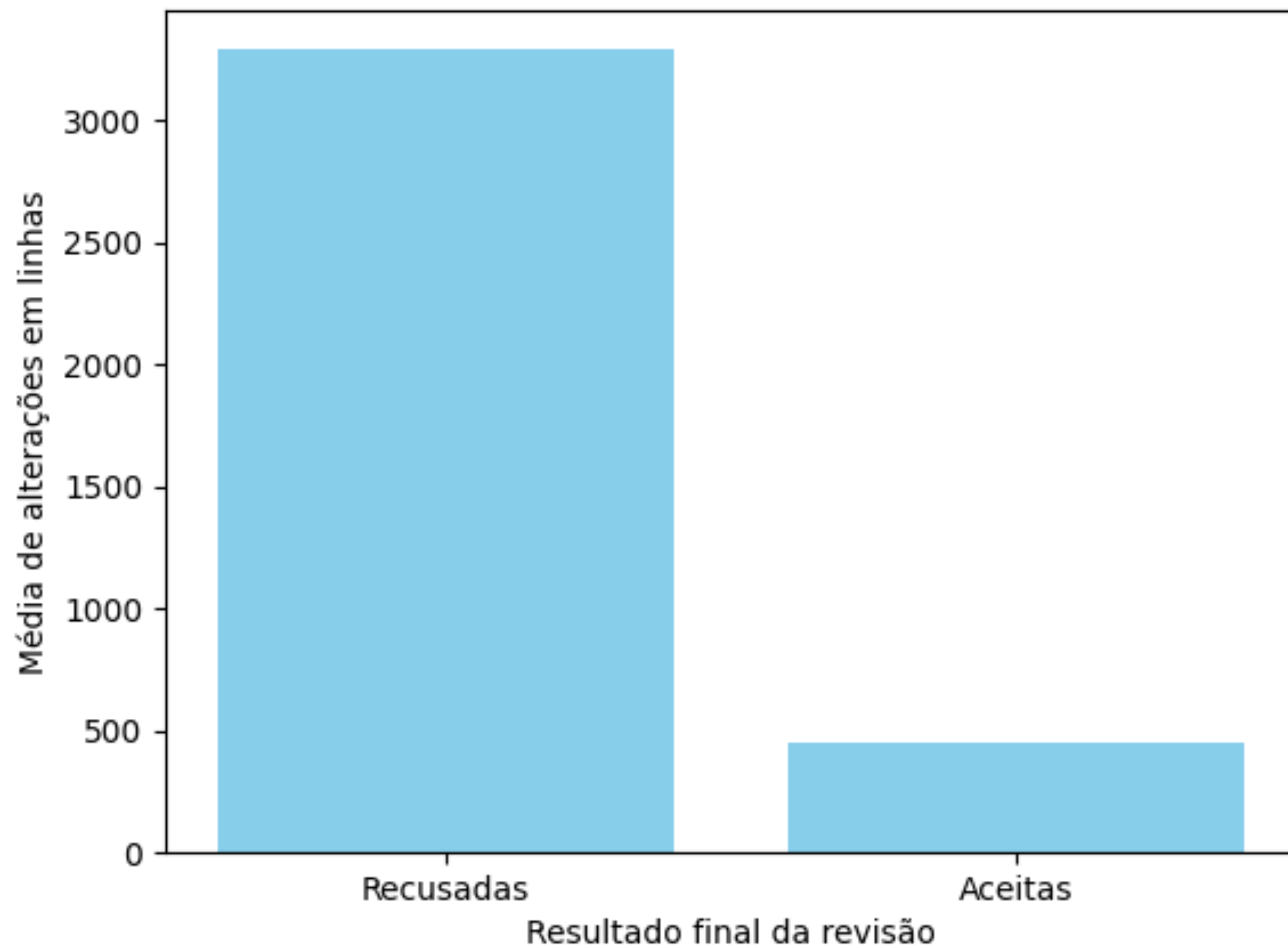
Resultado da revisão X Quantidade média de arquivos das PRs



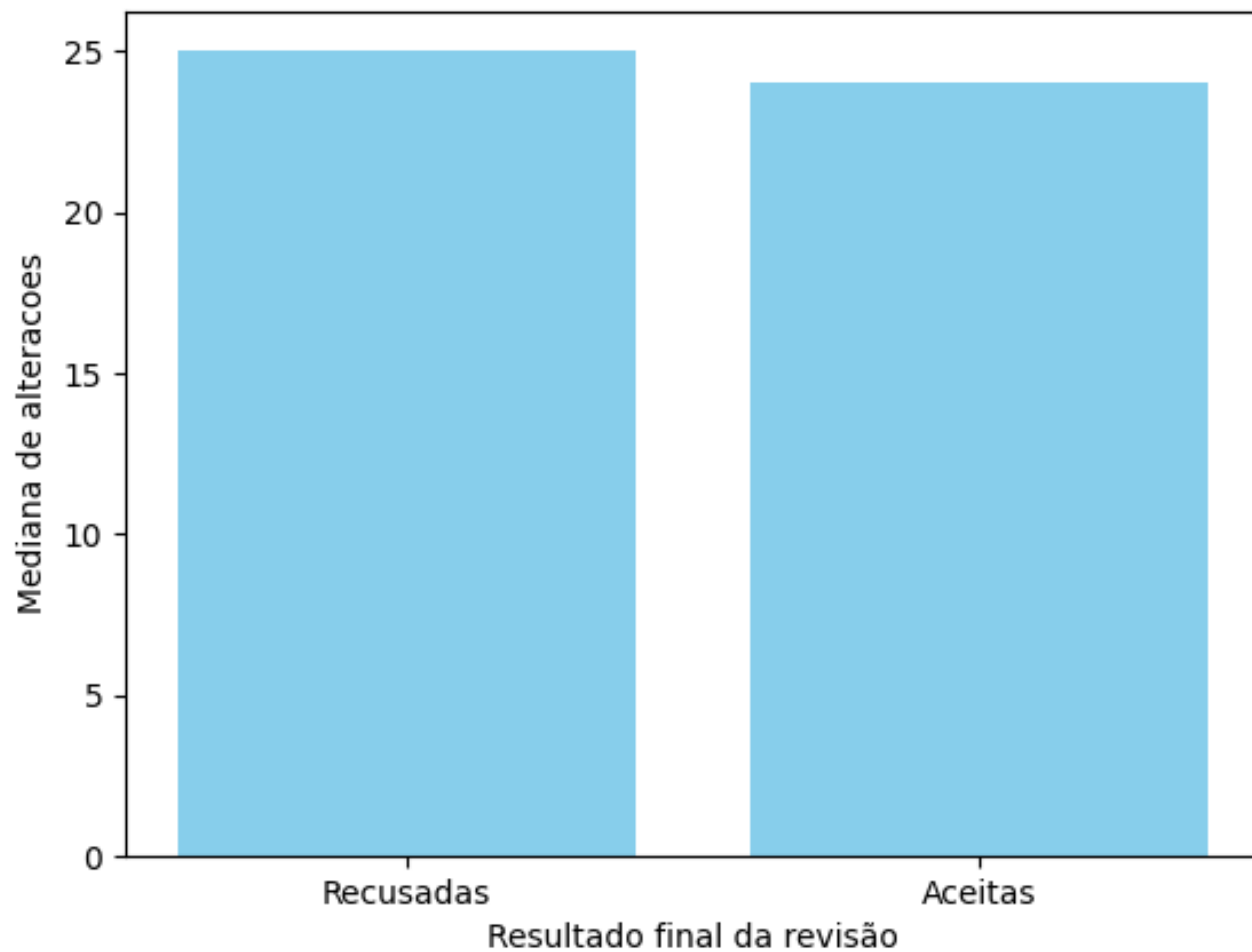
Resultado da revisão X Quantidade mediana de arquivos das PRs

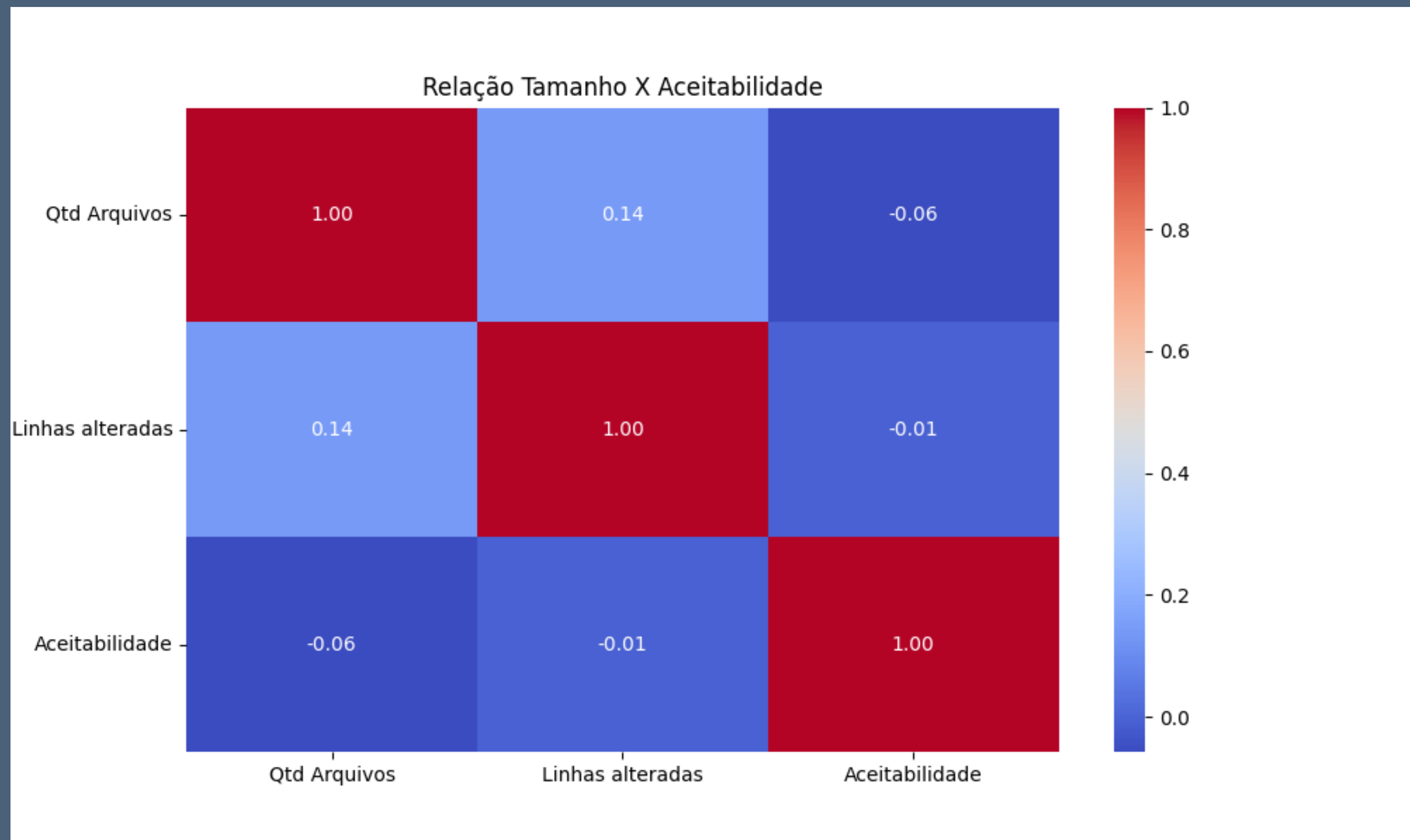


Resultado da revisão X Quantidade média de alteracoes em linhas das PRs



Resultado da revisão X Quantidade mediana de alteracoes em linhas das PR





Conclusão:

Hipótese confirmada.

Não existe relação entre o tamanho dos PR's e os resultados finais das revisões

Questão de Pesquisa 02

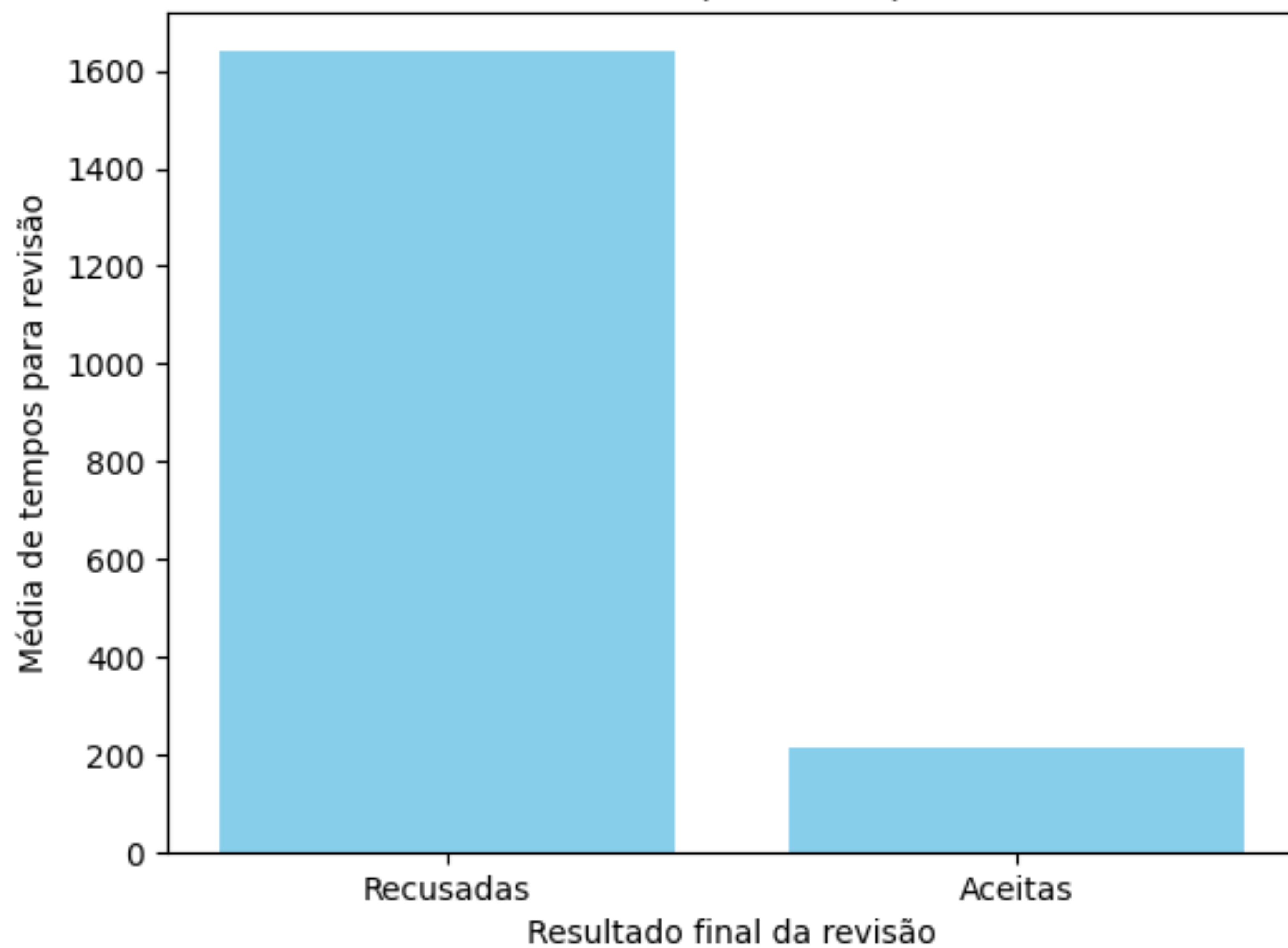
Hipótese:

Existe uma relação inversa entre o tempo de análise dos PR's e o Feedback final das revisões

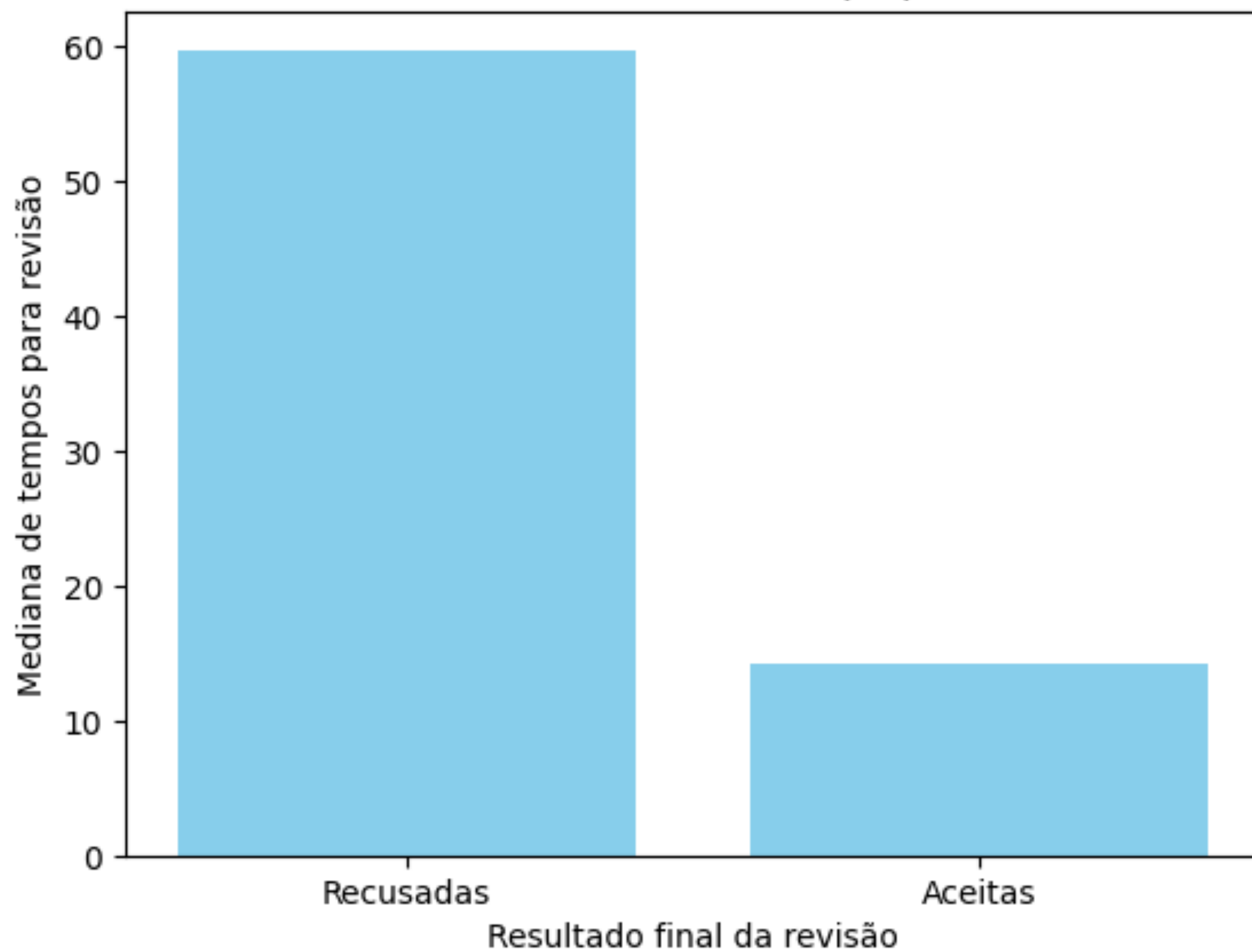
Justificativa

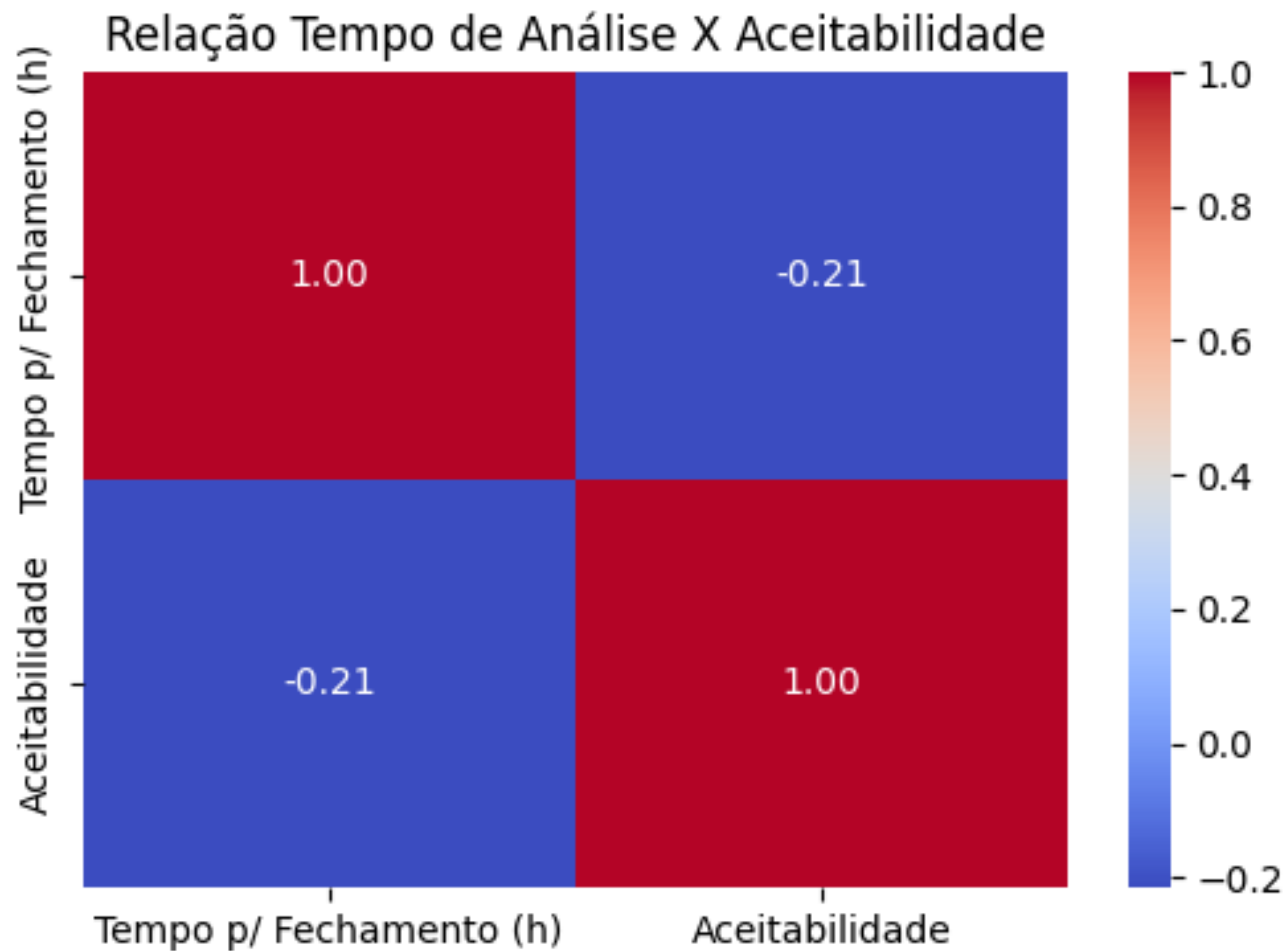
Revisões que duram mais tempo podem indicar código confuso e difícil de entender, o que pode aumentar o grau de rejeição.

Resultado da revisão X Tempo médio para revisão das PRs



Resultado da revisão X Mediana do tempo para revisão das PRs



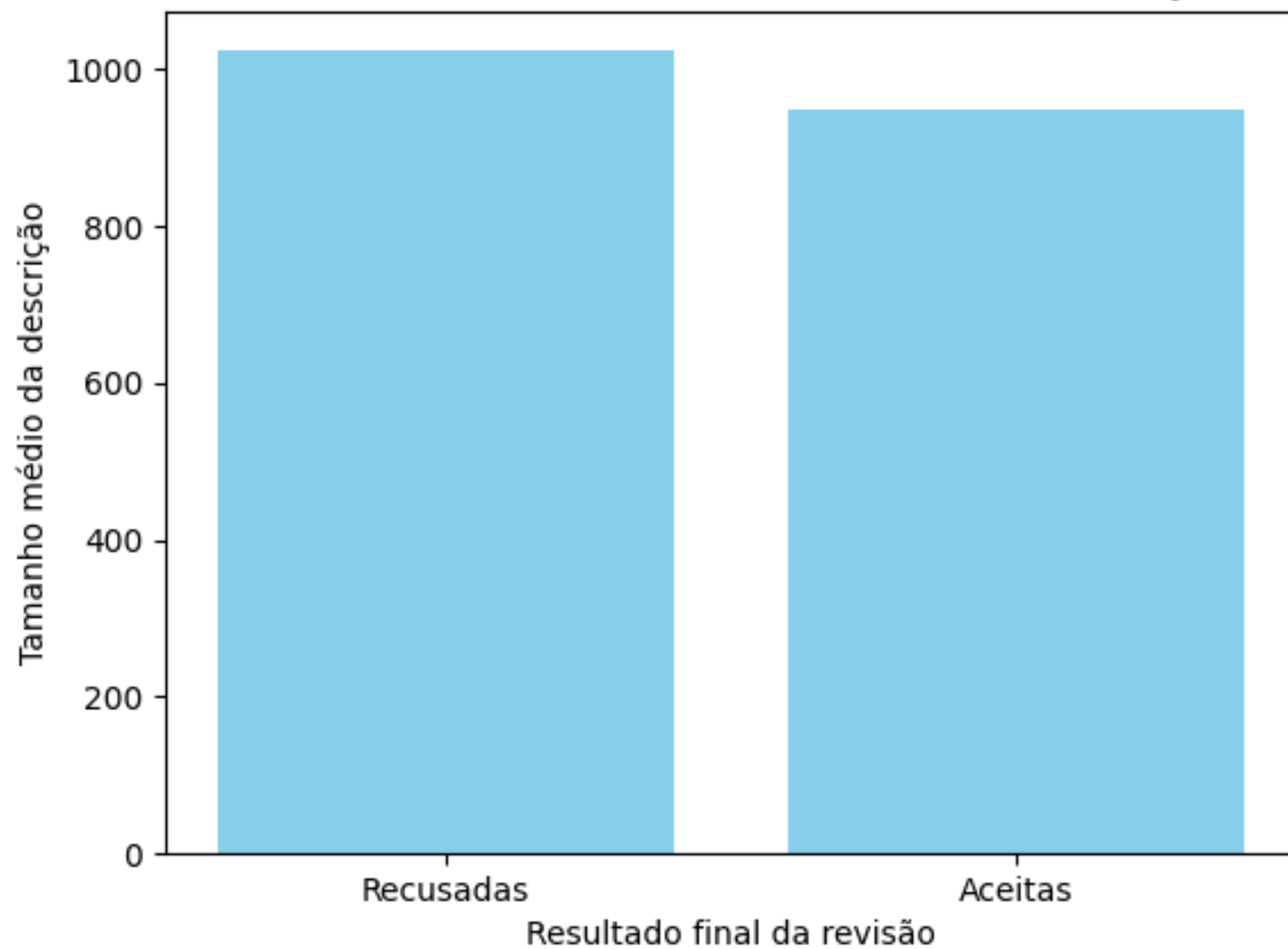


Conclusão:

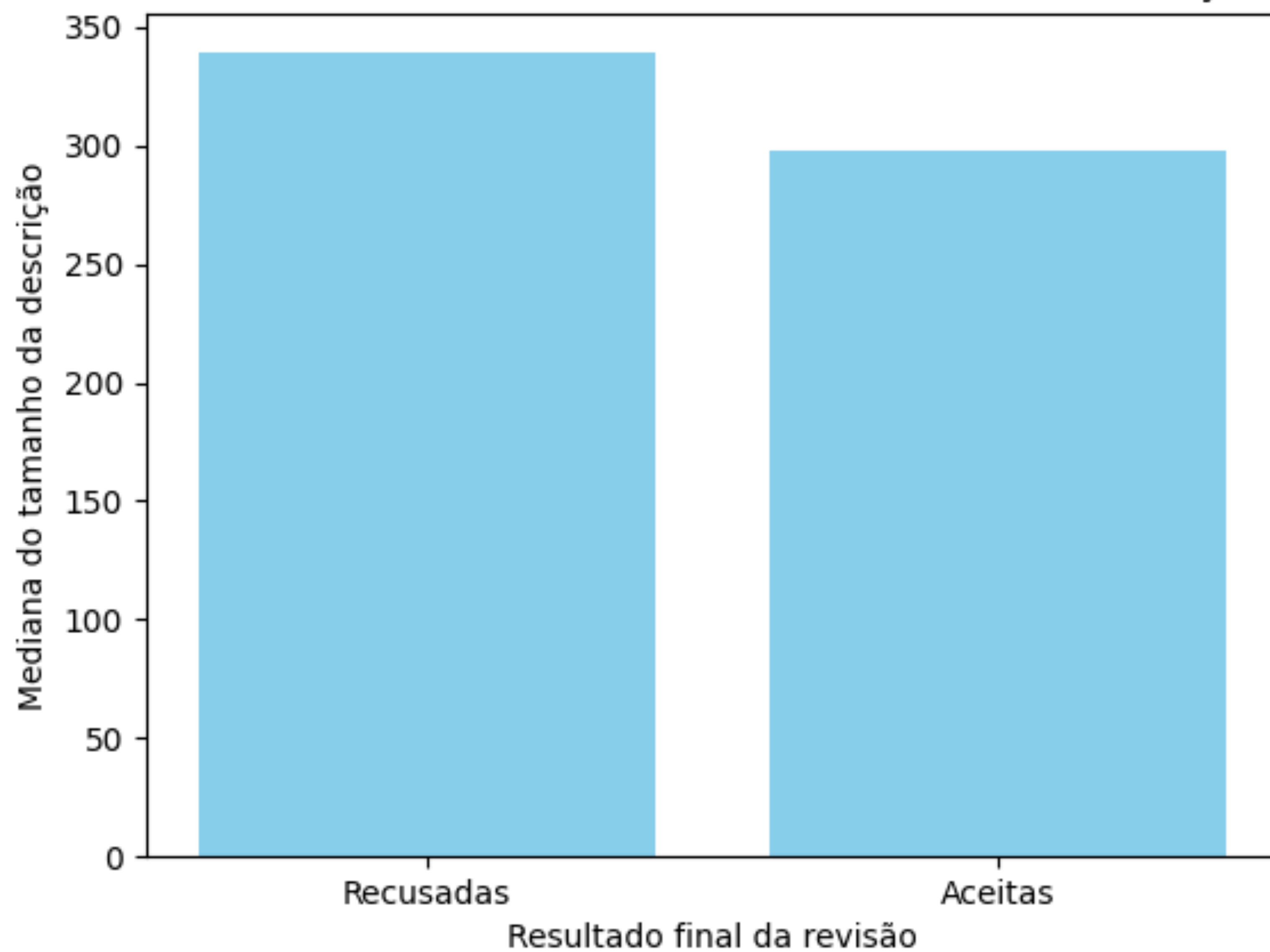
Hipótese confirmada (em partes).
De fato existe uma correlação fraca e negativa entre o tempo das revisões e o resultado final delas.

Questão de Pesquisa 03

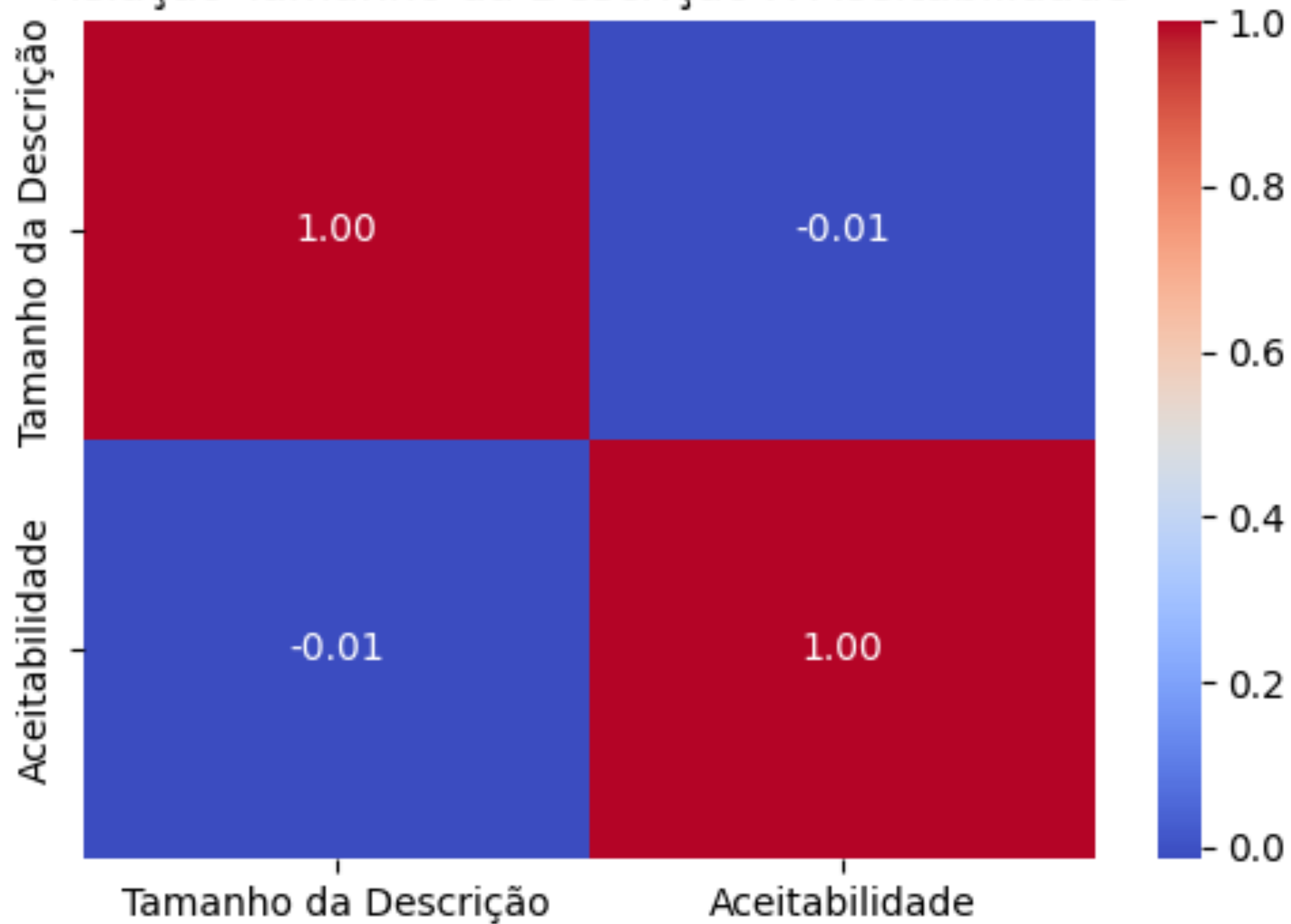
Resultado da revisão X Tamanho médio da descrição



Resultado da revisão X Mediana do tamanho da descrição



Relação Tamanho da Descrição X Aceitabilidade



Conclusão:

Hipótese confirmada.

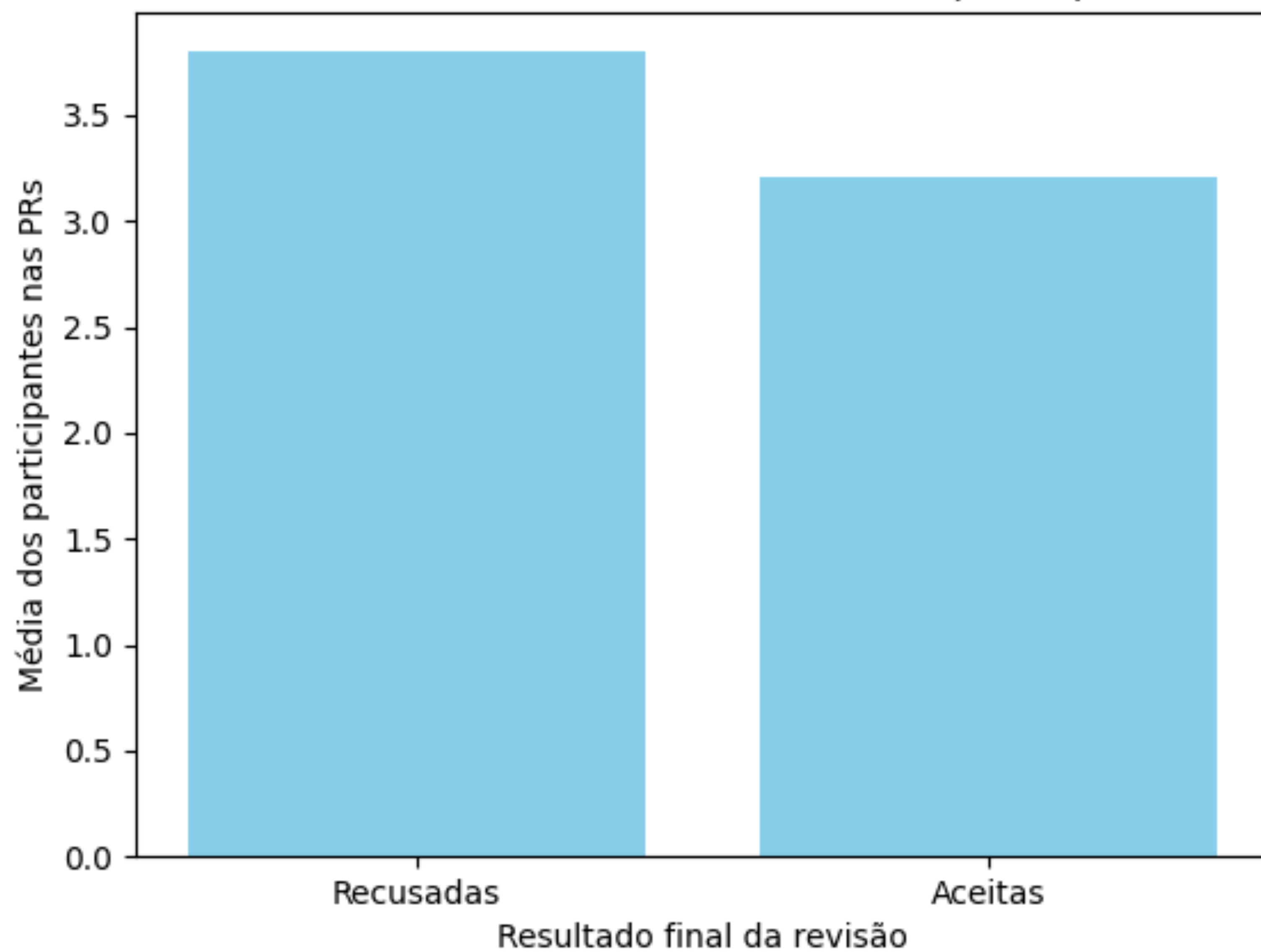
Não existe relação entre a descrição das PR's e o resultado final das revisões

Questão de Pesquisa 04

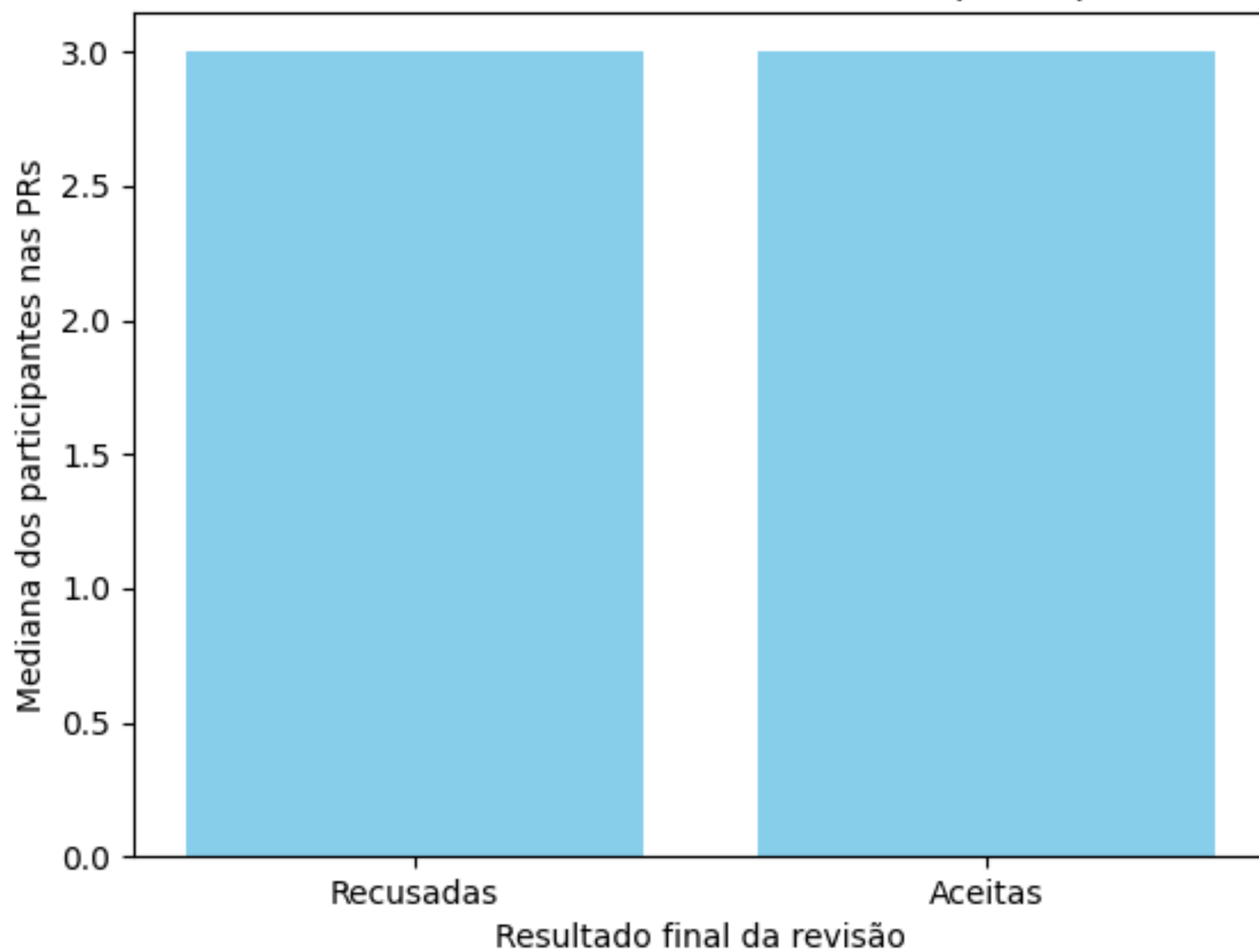
Hipóteses:

1. Não existe relação entre o número de participantes e o Feedback final das revisões
 - Justificativa: o número de participantes em uma alteração no código não determina, necessariamente, sua qualidade
2. Existe uma relação direta entre o número de comentários e o Feedback final das revisões
 - Justificativa: comentários tipicamente são feitos com o objetivo de fazer crítica ou apontar erros.

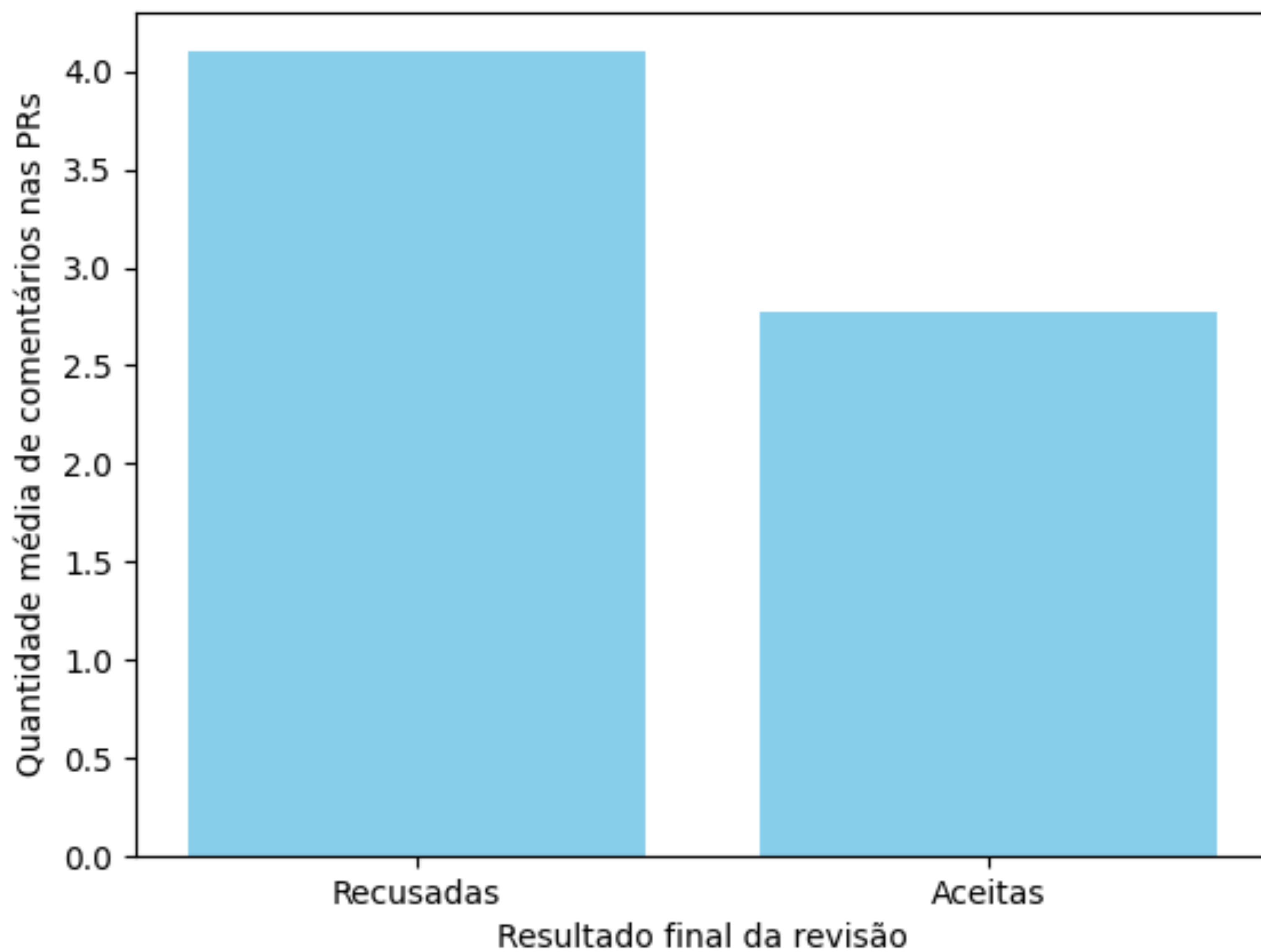
Resultado da revisão X Quantidade média de participantes na PR



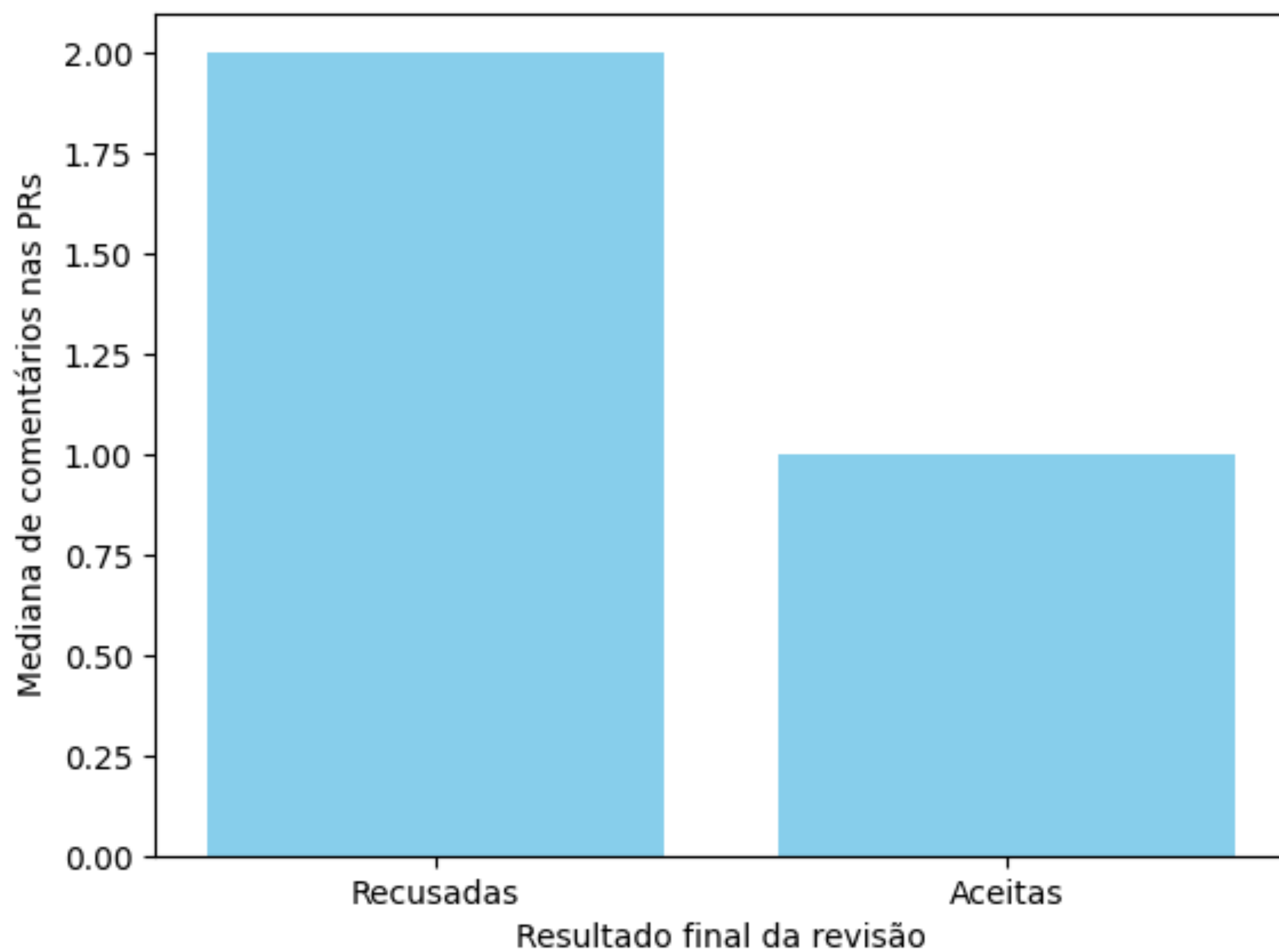
Resultado da revisão X Quantidade mediana de participantes nas PRs

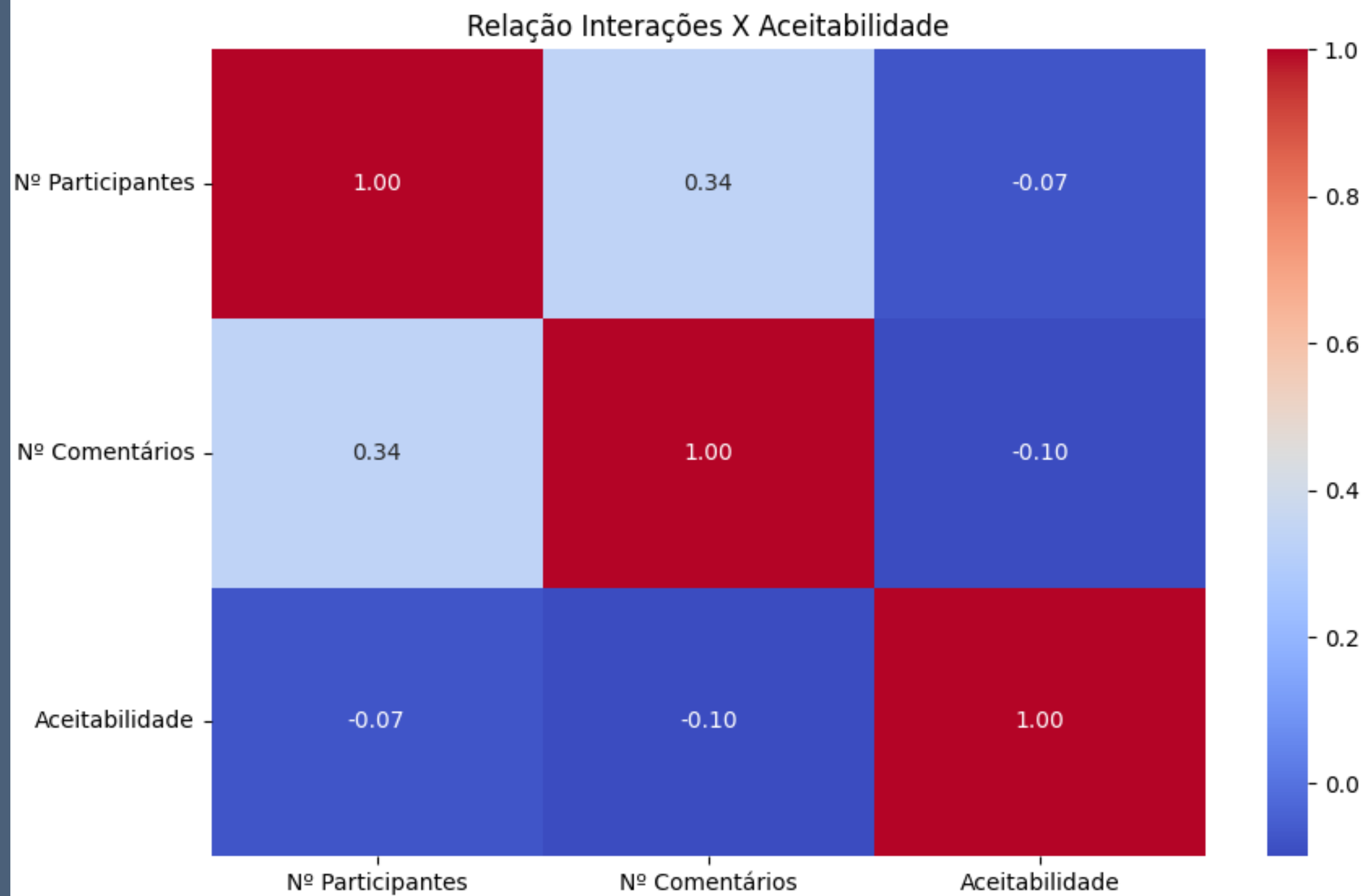


Resultado da revisão X Quantidade média de comentários nas PRs



Resultado da revisão X Quantidade mediana de comentários nas PRs





Conclusão:

1. Hipótese 1: Confirmada
2. Hipótese 2: Confirmada em partes (correlação negativa existente, mas muito fraca)

Obrigado!