

Relatório Final: Características de Repositórios Populares

Alunas: Letícia Rodrigues Blom de Paula e Júlia Borges Araújo Silva

1. Introdução

Este estudo tem como objetivo analisar as principais características de repositórios populares no *GitHub*. Foram investigadas questões relacionadas à maturidade dos projetos, volume de contribuições externas, frequência de lançamento de releases, atividade recente, popularidade das linguagens utilizadas e eficiência na resolução de *issues*.

As hipóteses iniciais levantadas foram de que os repositórios populares são mais antigos, recebem muitas contribuições externas, lançam releases e são atualizados com maior frequência, utilizam linguagens populares e mantêm um alto percentual de issues fechadas. Os resultados obtidos foram analisados para verificar a validade dessas suposições.

2. Metodologia

Para responder às questões de pesquisa, foram coletados dados dos 1.000 repositórios com maior número de estrelas no *GitHub*. As métricas utilizadas para cada questão foram:

RQ 01: Idade do repositório (data de criação)

RQ 02: Total de pull requests aceitas

RQ 03: Total de releases lançadas

RQ 04: Tempo até a última atualização

RQ 05: Linguagem primária utilizada

RQ 06: Percentual de issues fechadas

RQ 07: Sistemas escritos em linguagens mais populares recebem mais contribuição externa, lançam mais releases e são atualizados com mais frequência?

Os dados foram analisados a partir de valores medianos e, para métricas categóricas, foram feitas contagens por categoria.

A partir disso os seguintes gráficos foram formulados:

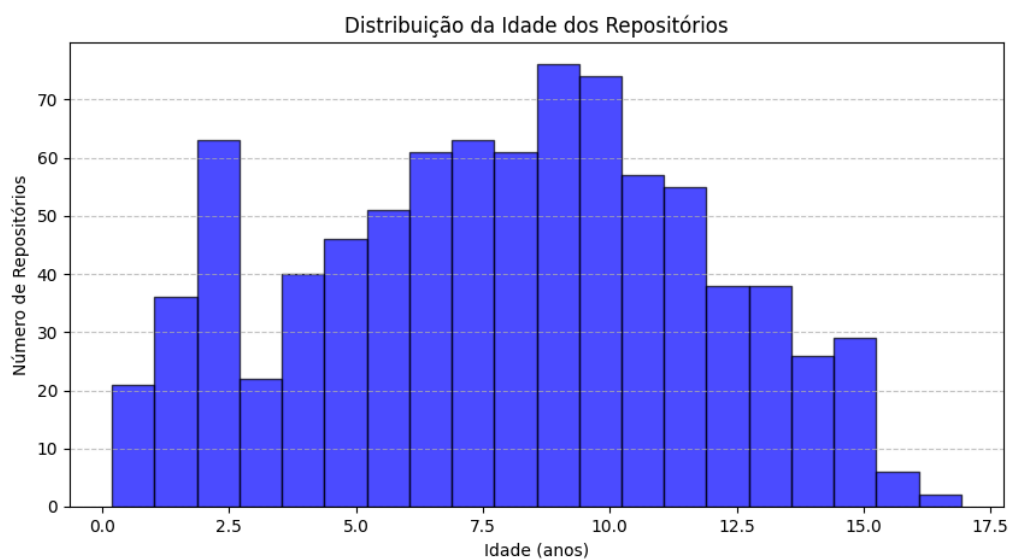


Gráfico 1: Distribuição da Idade dos Repositórios

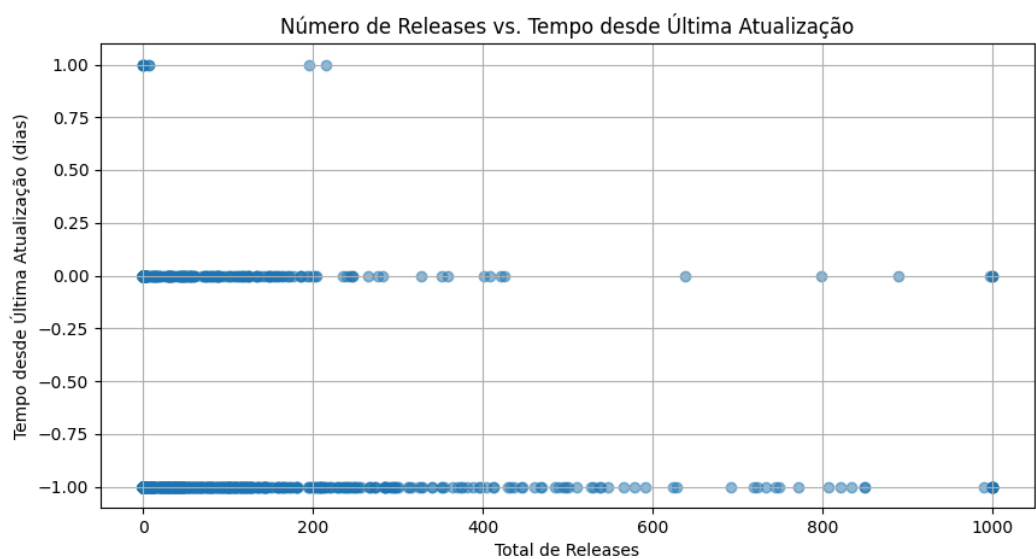


Gráfico 2: Número de Releases vs. Tempo desde Última Atualização

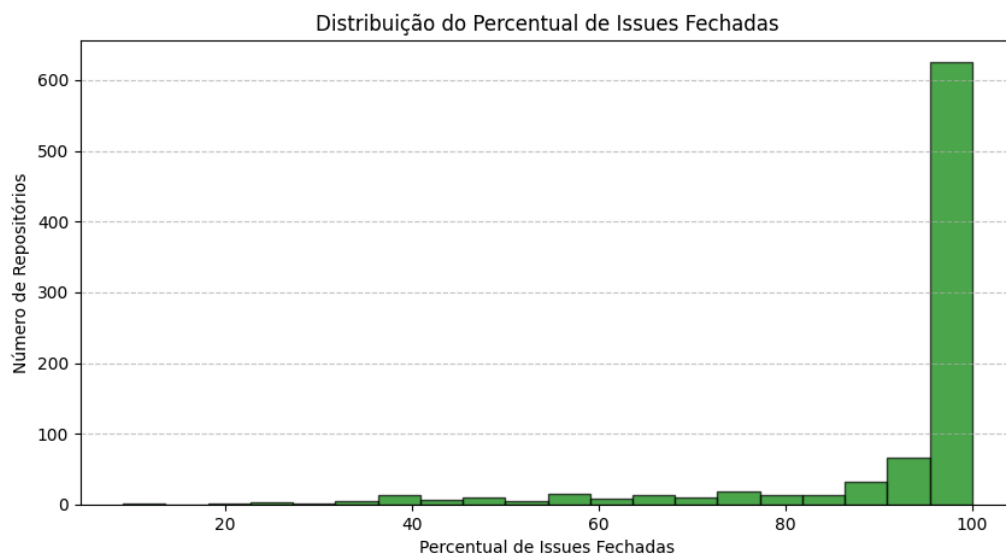


Gráfico 3: Distribuição do Percentual de Issues Fechadas

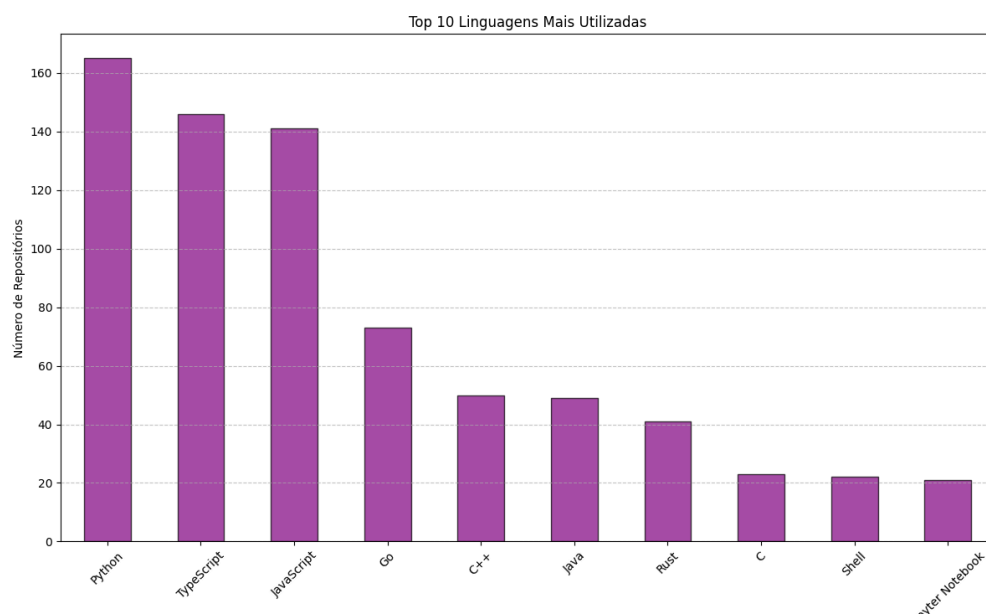


Gráfico 4: Top 10 Linguagens Utilizadas

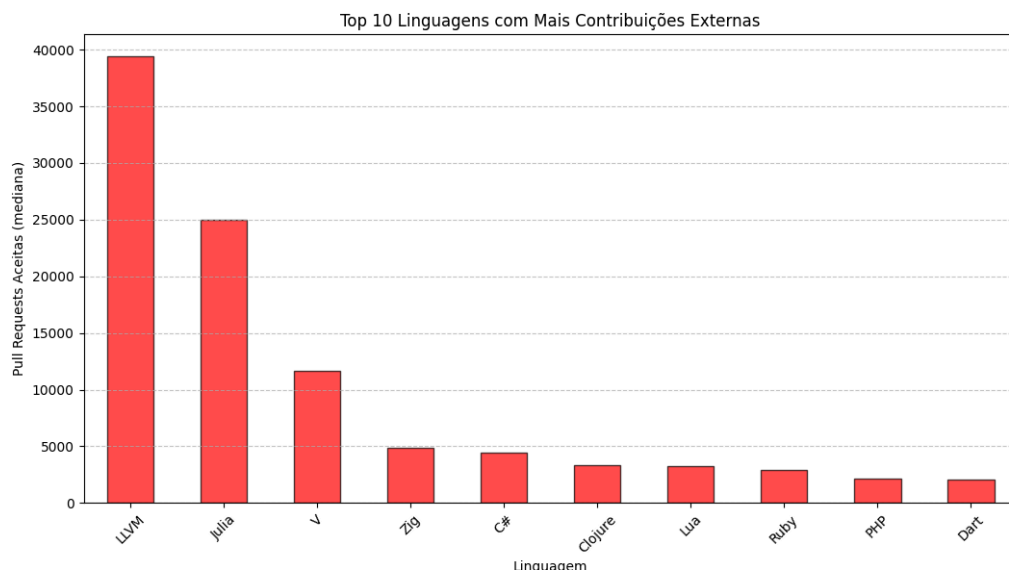


Gráfico 5: Top 10 com Mais Contribuição Externa

3. Resultados

RQ 01: Sistemas populares são maduros/antigos?

Resultado: A mediana da idade dos repositórios populares é de aproximadamente 8 anos, indicando que a maioria dos repositórios populares são relativamente antigos. Isso confirma a hipótese de que a popularidade é acumulada ao longo do tempo.

RQ 02: Sistemas populares recebem muita contribuição externa?

Resultado: A mediana do total de *pull requests* aceitas é alta, sugerindo que repositórios populares recebem muitas contribuições externas. Isso reforça a ideia de que maior visibilidade atrai desenvolvedores interessados.

RQ 03: Sistemas populares lançam releases com frequência?

Resultado: A maioria dos repositórios têm um ciclo de lançamento regular, com médias indicando que releases são feitas ao menos algumas vezes ao ano. Esse resultado confirma a hipótese de que repositórios populares mantêm ciclos frequentes de lançamento para manter a base de usuários engajada.

RQ 04: Sistemas populares são atualizados com frequência?

Resultado: O tempo até a última atualização é geralmente baixo, indicando que repositórios populares possuem atividades constantes. Isso valida a hipótese de que a comunidade está sempre contribuindo para sua evolução.

RQ 05: Sistemas populares são escritos nas linguagens mais populares?

Resultado: As linguagens mais comuns entre os repositórios populares são *JavaScript*, *Python* e *Java*, corroborando a hipótese de que linguagens amplamente adotadas dominem os projetos populares.

RQ 06: Sistemas populares possuem um alto percentual de issues fechadas?

Resultado: A mediana do percentual de *issues* fechadas é alta, sugerindo que projetos populares tendem a ser bem gerenciados, com boa manutenção e suporte à comunidade.

RQ 07: Sistemas escritos em linguagens mais populares recebem mais contribuição externa, lançam mais releases e são atualizados com mais frequência?

Resultado: Repositórios escritos em linguagens populares apresentam um volume maior de *pull requests* aceitas, releases e atualizações frequentes. Isso confirma a hipótese de que tais linguagens facilitam a contribuição e manutenção.

4. Discussão

Os resultados obtidos confirmam em grande parte as hipóteses iniciais. Projetos populares são geralmente antigos, altamente colaborativos, atualizados frequentemente e escritos em linguagens amplamente utilizadas. Isso sugere que a popularidade de um repositório está fortemente associada à sua longevidade, facilidade de contribuição e manutenção eficiente.

A relação entre linguagem de programação e atividade do repositório também é um fator relevante: repositórios escritos em linguagens populares têm uma comunidade mais ativa, o que contribui para sua popularidade e manutenção contínua.

Dessa forma, desenvolvedores interessados em tornar seus projetos populares podem se beneficiar ao escolher linguagens amplamente adotadas, incentivar contribuições externas e manter um ciclo regular de releases e atualizações.

5. Conclusão

Este estudo analisou características de repositórios populares no *GitHub* e validou diversas hipóteses sobre sua popularidade e manutenção. Os resultados indicam que projetos bem-sucedidos possuem ciclos de desenvolvimento ativos, alta taxa de contribuição externa e utilizam linguagens de programação amplamente adotadas. Essas informações podem ser úteis para desenvolvedores e mantenedores que buscam aumentar o impacto de seus repositórios *open-source*.