

## I. Introdução

Esse experimento tem como objetivo analisar sistemas populares open-source, explorando aspectos sobre o seu desenvolvimento, frequência das contribuições externas, lançamento de releases, entre outras características.

Inicialmente, formulamos algumas hipóteses acerca dos repositórios mais populares no GitHub, a saber:

- Possuem um volume elevado de contribuições, ou seja, o número de Pull Requests aceitas seria superior ao observado em repositórios menos populares.
- Empregam as linguagens de programação mais populares.
- Apresentam um maior grau de maturidade.
- São atualizados com maior frequência.
- Lançam releases de forma regular.
- Resolvem issues de maneira mais ágil.

Para validar essas suposições, coletamos dados dos 1000 repositórios mais bem avaliados - medidos pelo número de estrelas - por meio de consultas GraphQL. Este estudo, portanto, proporciona uma análise robusta e abrangente dos repositórios mais populares do GitHub, oferecendo insights valiosos sobre suas práticas e tendências predominantes.

## II. Metodologia

A metodologia deste estudo foi dividida em quatro etapas principais:

**Coleta de Dados:** Utilizamos a GitHub GraphQL API para reunir as informações necessárias para responder às questões propostas e atingir o objetivo do estudo.

**Filtragem e Paginação:** Devido ao volume substancial de informações e objetos consultados, foi necessário empregar a funcionalidade de paginação da API durante a coleta de dados. Este processo, que envolveu a gestão de múltiplas solicitações e um alto volume de dados, teve uma duração média de cinco minutos.

**Tratamento de Dados:** Os dados coletados foram sintetizados, determinando-se os valores medianos. Para os dados descritivos, realizou-se uma contagem de itens por categoria.

**Ordenação dos Repositórios:** Os repositórios foram classificados em ordem decrescente com base no número de estrelas obtidas.

Essa abordagem metódica permitiu uma análise rigorosa e sistemática dos repositórios mais populares do GitHub, proporcionando resultados confiáveis e precisos.

### III. Resultados

#### RQ 01. Sistemas populares são maduros/antigos?

**Métrica:** Idade do repositório (Calculado a partir da data de sua criação).

**Resultado:** Média de 2971 dias (8 anos).

Com uma média de 8 anos de vida, podemos dizer que os repositórios mais populares são mais maduros.

#### RQ 02. Sistemas populares recebem muita contribuição externa?

**Métrica:** Total de pull requests aceitas.

**Resultado:** Média de 580.

Pensando na média de idade dos repositórios, esse valor de PRs aceitas acabaria resultando em 6 merges por mês, uma quantidade baixa considerando a popularidade desses repositórios.

**Cálculo:** 8 anos são 96 meses. 580 pull requests dentro de 96 meses é convertido em aproximadamente 6 por mês.

#### RQ 03. Sistemas populares lançam releases com frequência?

**Métrica:** Total de releases.

**Resultado:** Média de 30,5 releases.

Visto a média de 30,5 releases podemos dizer que os sistemas populares lançam releases com frequência.

#### RQ 04. Sistemas populares são atualizados com frequência?

**Métrica:** Tempo desde a última atualização (Calculado a partir da data de última alteração de arquivos do repositório)

**Resultado:** Média de 4 dias. Os 6 repositórios com mais estrelas foram atualizados no dia que os dados foram consultados.

Então os repositórios mais populares são atualizados com frequência.

#### RQ 05. Sistemas populares são escritos nas linguagens mais populares?

**Métrica:** Linguagem primária de cada um desses repositórios.

**Resultado:** As linguagens que mais apareceram nos dados coletados foram:

1. Python
2. Javascript
3. TypeScript

De acordo com o Stack Overflow em “Comparing tag trends with our Most Loved programming languages” de Janeiro de 2023, as linguagens mais populares são Python, Java e Javascript. Aproximadamente 38% dos repositórios tem como linguagem principal uma dessas 3.

Além disso, o TypeScript é um superconjunto de JavaScript e está em ascensão, então pode-se considerar que ele estaria também dentro dessa lista. Concluindo que os repositórios foram escritos nas linguagens mais populares.

#### **RQ 06. Sistemas populares possuem um alto percentual de issues fechadas?**

**Métrica:** Razão entre número de issues fechadas pelo total de issues.

**Resultado:** Média de 86% de issues fechadas.

Com uma média de 1490 issues por repositório, temos que aproximadamente 208 issues continuam abertas, o que não é um bom número. Porém, analisando os 5 repositórios com mais estrelas, temos uma média de 96% de issues fechadas, o que mostra um alto percentual.

#### **IV. Conclusão**

A presente pesquisa confirmou, em sua maioria, as hipóteses iniciais que se alinhavam com o senso comum, somente os dados que se relacionam às contribuições externas não foram como o esperado. Este resultado pode ser atribuído à natureza de alguns dos repositórios mais populares, que não são softwares, mas projetos voltados para a disseminação de conteúdo relacionado a área de desenvolvimento de sistemas.

Desta forma, sugere-se como potencial aprimoramento para futuros estudos semelhantes, a implementação de um filtro mais refinado nas consultas, de modo a refletir com maior precisão o objeto de análise pretendido.

Adicionalmente, foi possível observar uma notável discrepância entre os dados dos repositórios situados no topo da lista, como os 10 primeiros, em comparação com os demais repositórios. Essa divergência foi particularmente evidente no que se refere ao percentual de issues completas e aos dias desde a última atualização.

Portanto, este estudo oferece insights valiosos sobre a dinâmica dos repositórios populares do Github, fornecendo uma base sólida para futuras investigações nesta área.