

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
Bacharelado em Engenharia de Software - Instituto de Ciências Exatas e Informática Unidade  
Educacional Praça da Liberdade

Daniel Henrique Vargas

**Relatório Final**  
**Análise de Repositórios Populares**  
**Laboratório de Experimentação de Software**

Belo Horizonte  
2018

## Sumário

1. Introdução	1
2. Metodologia	1
2.1. Hipóteses	2
3. Apresentação e Discussão dos Resultados	2
3.1. Sistemas populares são maduros/antigos?	2
3.2. Sistemas populares recebem muita contribuição externa?	3
3.3. Sistemas populares lançam releases com frequência?	3
3.4. Sistemas populares são atualizados com frequência?	4
3.5. Sistemas populares são escritos nas linguagens mais populares?	5
3.6. Sistemas populares possuem um alto percentual de issues fechadas?	5

## 1. Introdução

Durante o processo de desenvolvimento de um software, é muitas vezes necessário fazermos o uso de ferramentas de controle de versão do código que está sendo desenvolvido, como forma de organizá-lo e disponibilizá-lo.

Uma dessas ferramentas, e possivelmente a mais popular, é o Git, que possui uma plataforma de disponibilização de código chamada GitHub, plataforma na qual foi realizada a mineração e análise dos 1000 repositórios open-source mais populares da plataforma, com o objetivo de estudar as principais características de sistemas populares open-source.

Para alcançar este objetivo, esse trabalho visa responder às algumas questões de pesquisas, utilizando as seguintes métricas:

Objetivos	Questões de Pesquisa	Métricas
<b>Estudar as principais características de sistemas populares open-source</b>	Sistemas populares são maduros/antigos?	Idade do repositório (calculado a partir da data de sua criação)
	Sistemas populares recebem muita contribuição externa?	Total de pull requests aceitas
	Sistemas populares lançam releases com frequência?	Total de releases
	Sistemas populares são atualizados com frequência?	Tempo até a última atualização (calculado a partir da data de última atualização)
	Sistemas populares são escritos nas linguagens mais populares?	Linguagem primária de cada um desses repositórios
	Sistemas populares possuem um alto percentual de issues fechadas?	Razão entre número de issues fechadas pelo total de issues

## 2. Metodologia

A mineração dos dados foi feita por meio de uma requisição em forma de query realizada em uma API de GraphQL, disponibilizada pelo próprio GitHub, para obter informações sobre os 1000 repositórios com mais de 100 estrelas da plataforma.

Entretanto, para poder obter os dados necessários de todos os 1000 repositórios, visto que o volume de informações era muito alto para a API obter de uma vez só, foi preciso fazer o uso de recursos como paginação e um sistema para repetir a requisição caso está falhasse.

Logo após realizar a consulta dos 1000 repositórios, os dados foram enviados para um arquivo CSV, onde foram organizados, tratados e analisados.

## 2.1. Hipóteses

As seguintes hipóteses foram formuladas priori a análise dos resultados:

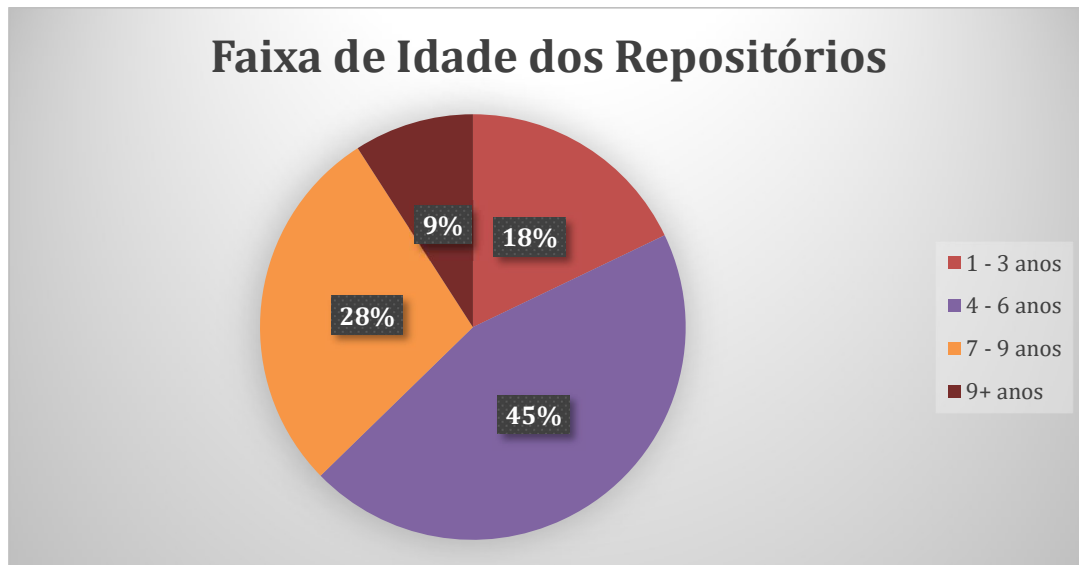
- **Hipótese 1** – Os sistemas mais populares são mais antigos/maduros, devido ao fato de que a popularidade de um sistema tende a crescer conforme o tempo avança, uma vez que sua estabilidade e suas funcionalidades também aumentam com o passar do tempo.
- **Hipótese 2** – Os sistemas mais populares não recebem muita contribuição externa, uma vez que uma certa curadoria do código, mesmo que este seja open-source, é necessária para manter um padrão de qualidade mais elevado, evitando que muitas contribuições externas ocorram.
- **Hipótese 3** – Os sistemas mais populares lançam releases com frequência, uma vez que muitas pessoas provavelmente utilizam esses sistemas e precisam deles cada vez melhores e mais estáveis.
- **Hipótese 4** – Os sistemas mais populares são atualizados com frequência, como forma de manter o sistema sempre em dia com o mercado e aprimorá-lo cada vez mais.
- **Hipótese 5** – Os sistemas mais populares são escritos nas linguagens mais populares, visto que a popularidade de um sistema pode ser baseada na popularidade da linguagem primária do mesmo, tornando os sistemas escritos nas linguagens populares mais procurados e movimentados do que outros.
- **Hipótese 6** – Os sistemas mais populares possuem um alto percentual de issues fechadas, uma vez que sendo populares tendem a possuírem um suporte mais dedicado e um grande volume de usuários experientes que podem resolver os mais diversos tipos de problemas relacionados ao sistema e fechar um alto número de issues.

## 3. Apresentação e Discussão dos Resultados

Após a organização, tratamento e análise dos dados, os seguintes resultados foram obtidos e comparados com as hipóteses originalmente formuladas.

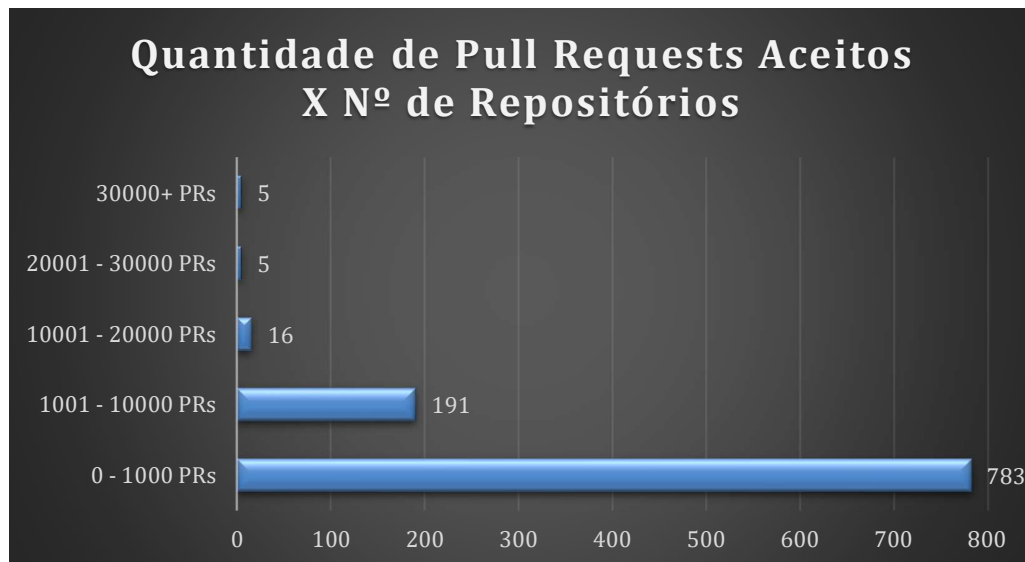
### 3.1. Sistemas populares são maduros/antigos?

Com base nos resultados obtidos, podemos perceber que a maioria dos repositórios populares não é antiga, e sim recente, com 45% dos repositórios estando entre a faixa de 4 a 6 anos de idade. Desse modo, a hipótese original é refutada, uma vez que 28% dos repositórios se encontra entre a faixa de 7 a 9 anos de idade, e apenas 9% estão na faixa de mais de 9 anos de idade. Segue abaixo um gráfico com os resultados relativos à essa questão:



### 3.2. Sistemas populares recebem muita contribuição externa?

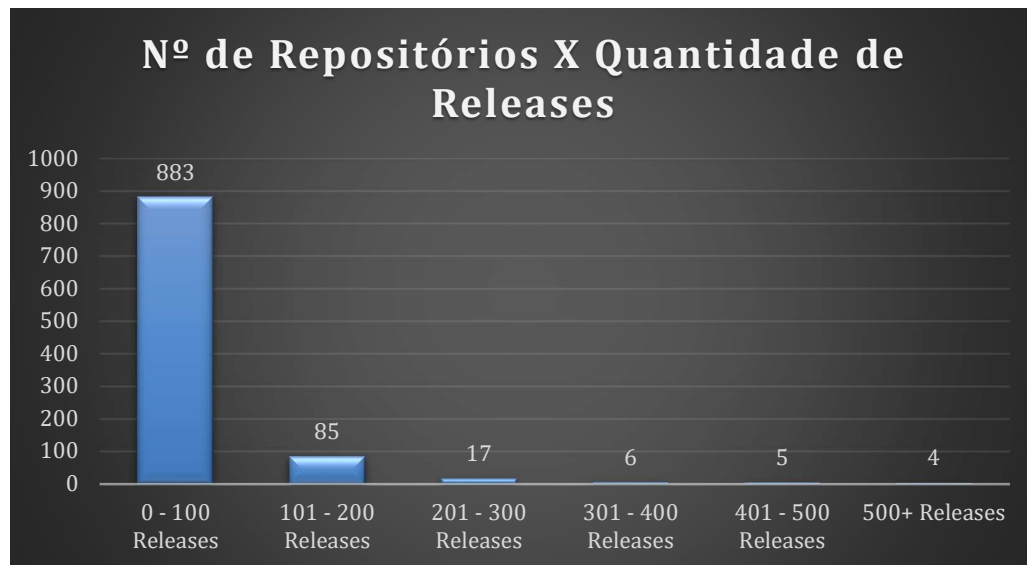
Com base nos resultados obtidos, podemos perceber que de fato a maioria dos repositórios populares não recebem muita contribuição externa, com uma quantidade esmagadora dos repositórios, cerca de 78,30%, tendo entre 0 e 1000 Pull Requests aceitos apenas. Esse resultado corrobora com a hipótese originalmente levantada sobre essa questão. Segue abaixo um gráfico com os resultados obtidos:



### 3.3. Sistemas populares lançam releases com frequência?

Com base nos resultados obtidos, podemos perceber que a maioria dos repositórios populares não lançam Releases com muita frequência, de modo que cerca de 88,30% dos repositórios possuem apenas entre 0 e 100 Releases lançadas. Dessa forma, a hipótese

originalmente apresentada para essa questão é refutada. Segue abaixo um gráfico com os resultados obtidos:



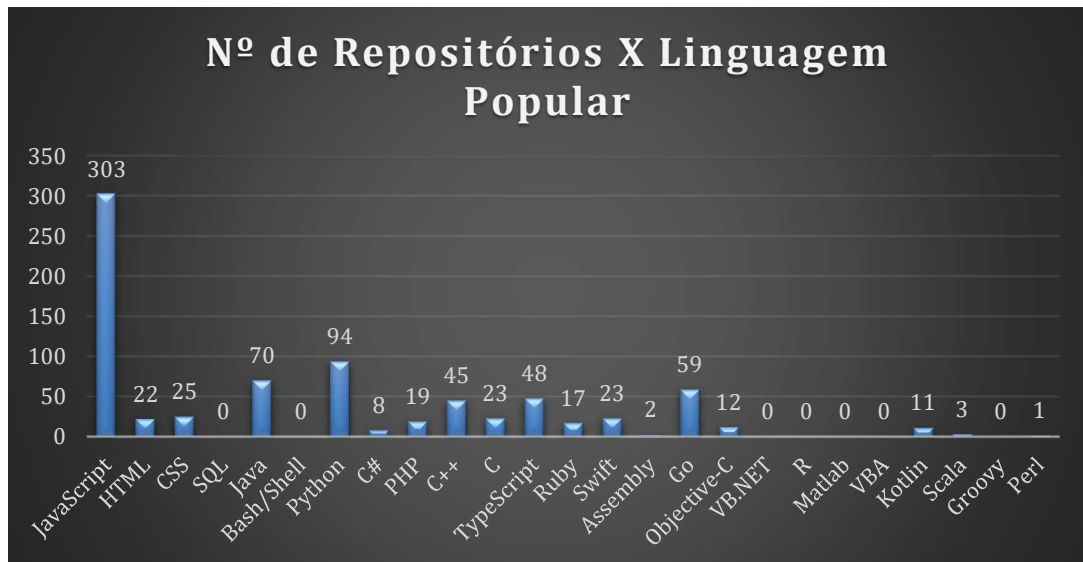
#### 3.4. Sistemas populares são atualizados com frequência?

Com base nos resultados obtidos, podemos perceber que a maioria dos repositórios populares são atualizados com muita frequência, uma vez que, usando como base a data de obtenção dos dados (04/03/2020), um total de 63,70% dos repositórios tinha cerca de algumas horas desde a última atualização, e cerca de 35,10% tinham apenas 1 dia desde a última atualização. Esse resultado corrobora com a hipótese originalmente levantada sobre essa questão, uma vez que menos de 2% dos repositórios possuía 2 dias ou mais desde a última atualização. Segue abaixo um gráfico com os resultados relativos a essa questão:



### 3.5. Sistemas populares são escritos nas linguagens mais populares?

Com base nos resultados obtidos, podemos perceber que a maioria dos repositórios populares de fato são escritos nas linguagens mais populares, uma vez que cerca de 78,50% dos repositórios são escritos em algumas das 25 linguagens mais populares, corroborando para a hipótese originalmente levantada sobre essa questão. Segue abaixo um gráfico com os resultados obtidos:



### 3.6. Sistemas populares possuem um alto percentual de issues fechadas?

Com base nos resultados obtidos, podemos perceber que os repositórios populares realmente possuem um alto percentual de issues fechadas, com cerca de 86% de issues fechadas e apenas 14% de issues não fechadas. Esse resultado corrobora com a hipótese originalmente levantada sobre essa questão. Segue abaixo um gráfico com os resultados obtidos:

