

Pontifícia Universidade Católica de Minas Gerais

Laboratório de Experimentação de Software

Pesquisa: Características de Repositórios populares

Ian Asenjo Dominguez Cunha

Jully Ketely Alves da Silva

Warley Leandro dos Anjos

OBS.: A análise da base de dados foi feita no dia 05 de Março, ou seja, resultados podem estar desatualizados até a data atual.

Foram analisados os 1000 repositórios mais populares do Github, com o objetivo de analisar características e responder algumas questões, através de resultados analíticos. A linguagem utilizada foi Python para mineração, e GraphQL para requisições de API.

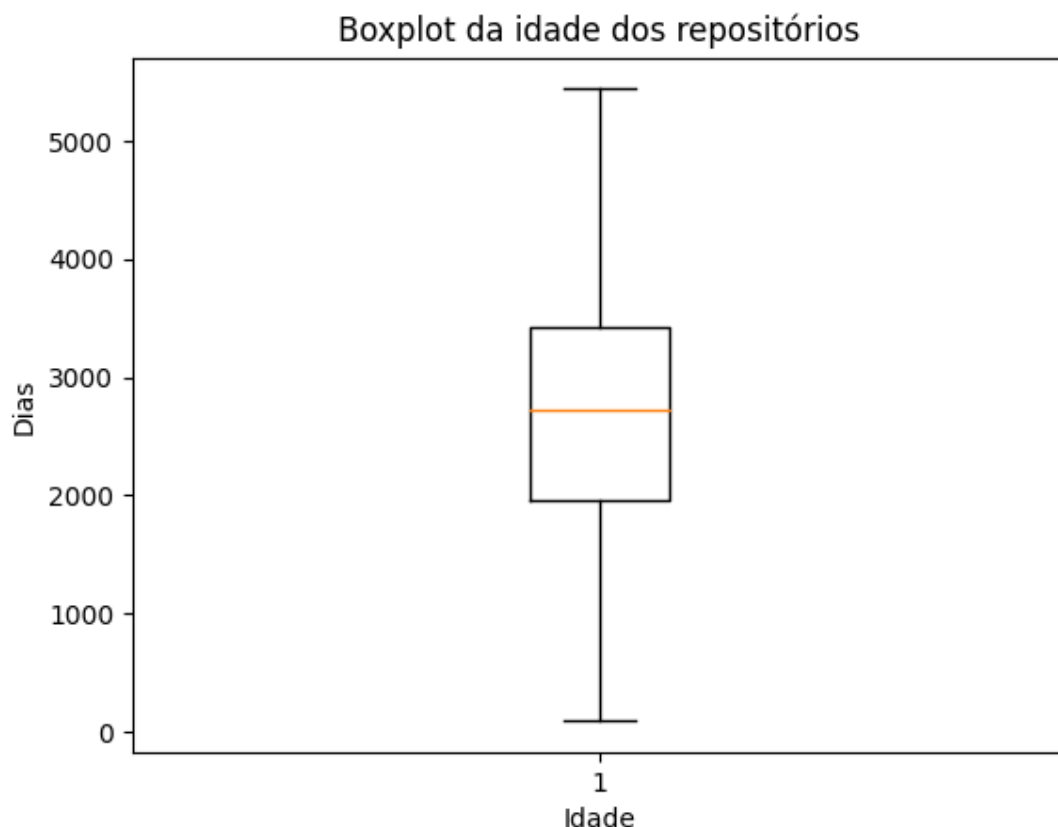
O sistema gera um conjunto de dados, que são armazenados em um arquivo .csv (planilha). Com base, nesta planilha, o script gera os resultados e os gráficos para prosseguir com a análise e consequente chegar a um resultado coerente.

QUESTÕES

RQ 01. Sistemas populares são maduros/antigos?

Hipótese: Sim, pois estes sistemas antigos tem, na maioria das vezes, o maior período de contribuições e evolução do repositório.

Resultado: A mediana da idade do repositório é 2728.5 dias, ou seja, por volta de 7 anos e 5 meses e 1 semana. Podemos chegar a conclusão que, sim, os repositórios mais populares são maduros.

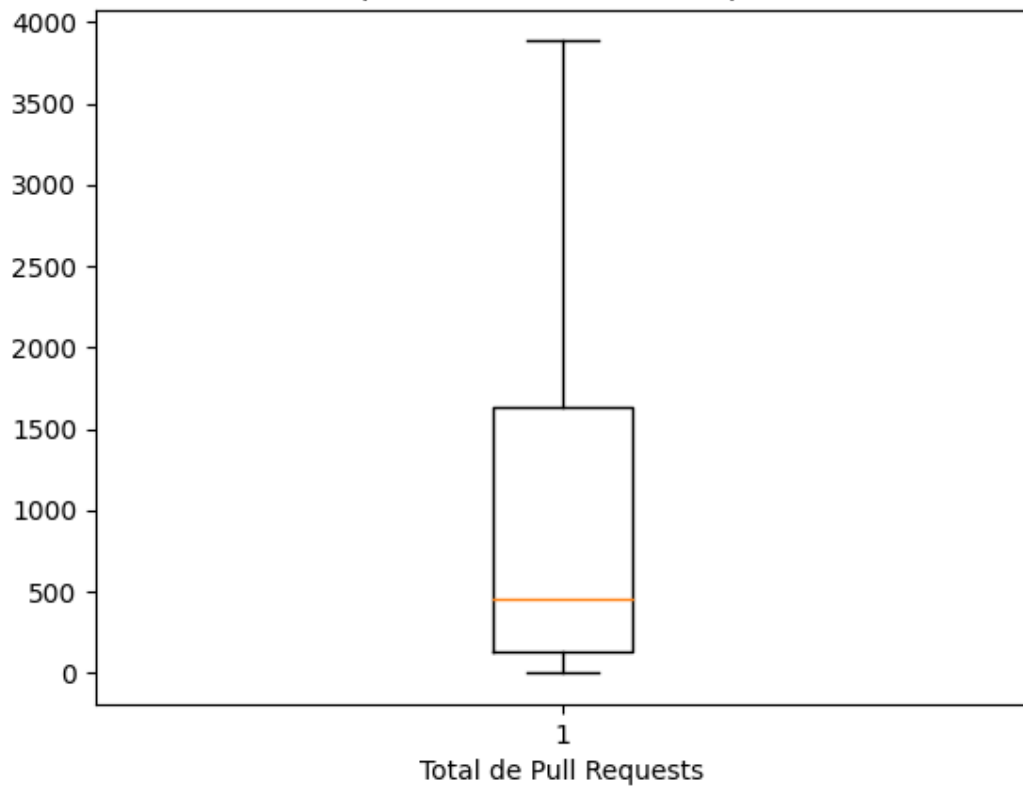


RQ 02. Sistemas populares recebem muita contribuição externa?

Hipótese: Não, pois sistemas populares costumam receber um padrão e performance de código, e alguns colaboradores ficam receosos em expor suas opiniões na comunidade open source.

Resultado: A mediana da total de pull requests que os repositórios aceitam é 459. Concluimos que, não, os repositórios não costumam aceitar/receber tanta contribuição externa. A mediana está bem abaixo da média dos "pull requests" feitas em repositórios *open source*.

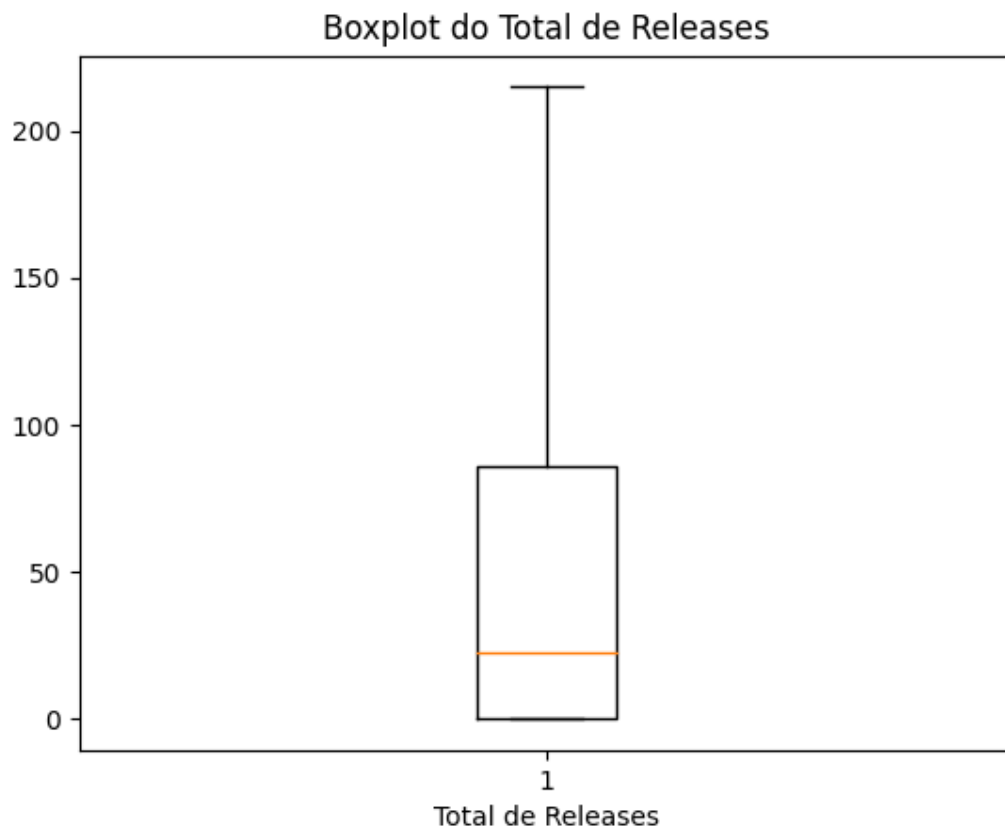
Boxplot do Total de Pull Requests



RQ 03. Sistemas populares lançam releases com frequência?

Hipótese: Sim, sistemas populares tendem a manter seus repositórios atualizados sempre que possível para manter.

Resultado: A mediana da total de Releases é 22.5. A média de releases é 73,64 portanto não, os repositórios populares não costumam lançar releases com frequência e isto pode ser um problema, ainda por cima por ser popular e *open source*.



RQ 04. Sistemas populares são atualizados com frequência?

Hipótese: Sim. Pois, são vários contribuidores ajudando na evolução e desenvolvimento do sistema, principalmente em repositórios *open source*.

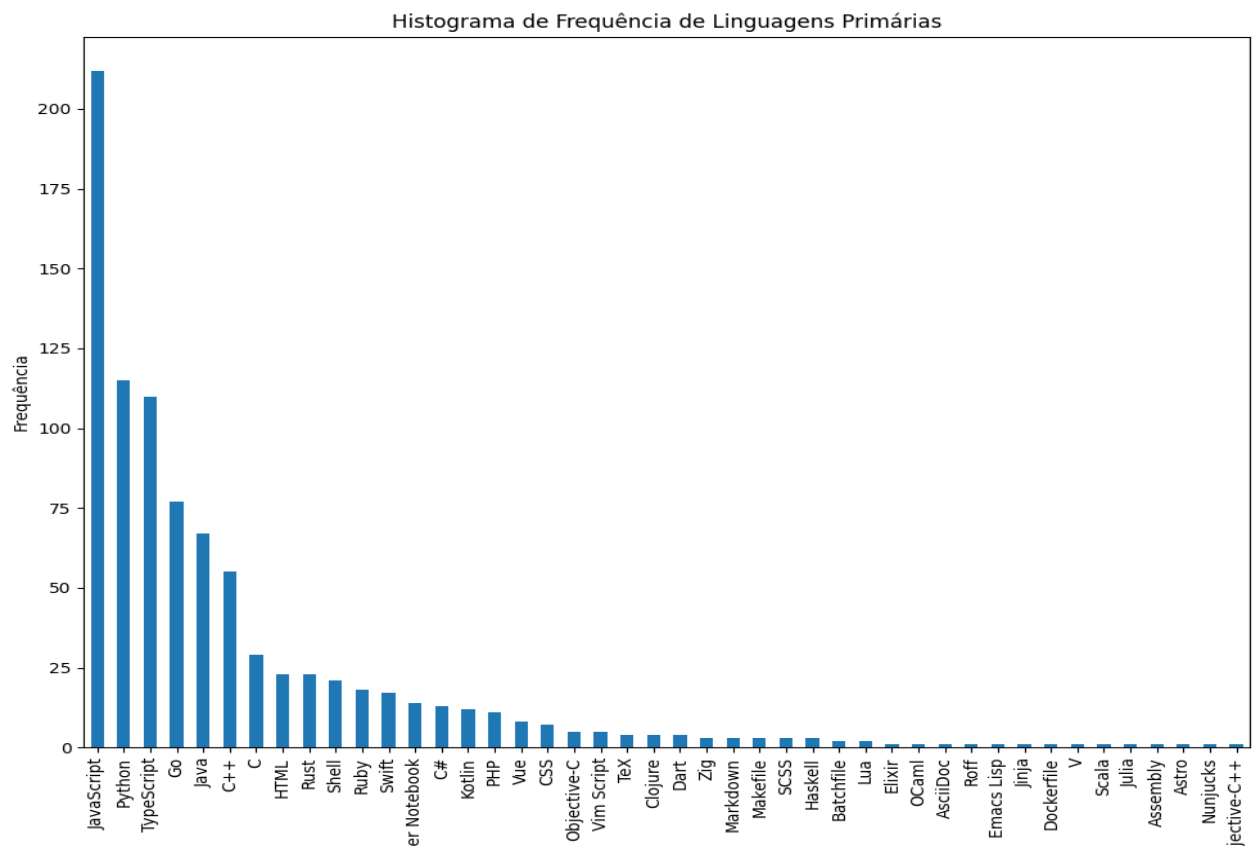
Resultado: A mediana do tempo de atualização é de 8 dias. Podemos concluir que, apesar da quantidade de contribuições, os repositórios populares costumam ser atualizados a cada 8 dias, e isso pode ser um problema, dependendo do objetivo do projeto e como funciona o processo interno dele, devido a quantidade de contribuidores e por ser um sistema *open source*.



RQ 05. Sistemas populares são escritos nas [linguagens mais populares](#) [Links to an external site?](#)

Hipótese: De acordo com o site da *octoverse* o top 5 das linguagens são JavaScript, Python, Java, TypeScript e C#. Acredito que a linguagem mais popular seja o C# seja o C# por está mais tempo no mercado e ser bastante usado.

Resultado: O Top 5 de acordo com o resultado da mineração: JavaScript; Python; TypeScript; Go; e Java. Os repositórios mais populares estão sendo feitos em linguagem JavaScript e Python. E interessante é Python se tornou uma linguagem popular há pouco tempo e isso mostra o engajamento dos desenvolvedores em relação a esta linguagem voltada para Data Science entre outros.



RQ 06. Os sistemas populares possuem um alto percentual de issues fechadas?

Hipótese: Sim. Pensando na quantidade de contribuições feitas ao repositório, que é *open source* e popular.

Resultado: A mediana do total de Issues fechadas é 87.32%. Portanto, sim, com base no gráfico podemos ver que a razão de issues fechados em relação ao total de issues ultrapassar a média. Isso diz, que repositórios populares possuem muitas issues fechadas.

