

# Relatório de Análise de Repositórios Open-Source Populares

## 1. Introdução

O objetivo deste estudo é analisar as principais características de sistemas open-source populares, definidos como aqueles com maior número de estrelas no GitHub.

A partir da coleta de dados dos 1.000 repositórios mais estrelados, buscamos responder às seguintes questões de pesquisa: idade dos sistemas, contribuição externa, frequência de releases, frequência de atualização, linguagens utilizadas e percentual de issues.

## 2. Metodologia

Para responder às questões de pesquisa, utilizamos a API GraphQL do GitHub.

Desenvolvemos um script em C# que realiza requisições automáticas à API, aplicando paginação até coletar dados de 1.000 repositórios ordenados por número de estrelas.

Foram coletados os seguintes atributos: nome, descrição, URL, número de estrelas, número de forks, data de criação, data da última atualização, linguagem primária, número total de issues, número de issues fechadas, total de pull requests e total de releases.

Os dados foram armazenados em um arquivo CSV, permitindo análise posterior em ferramentas de manipulação de dados.

Para análise estatística, utilizamos principalmente valores medianos, que representam melhor a tendência central dos dados e reduzem o impacto de outliers.

## 3. Resultados

Cálculo com base nos Resultados obtidos	
Mediana da idade dos repositórios (em anos)	8
Mediana do número de pull requests	1147
Mediana do número de releases	35.5
Contagem das linguagens mais usadas	Python
Mediana da razão issues fechadas / issues totais	1.14

## 4. Discussão

Os resultados confirmam parcialmente as hipóteses formuladas:

- **Idade dos repositórios:** A mediana de 8 anos indica que muitos projetos populares são maduros, embora também existam repositórios mais recentes que obtiveram grande crescimento rapidamente.
- **Contribuições externas:** A mediana de 1147 pull requests sugere que, embora muitos repositórios recebam contribuições externas significativas, nem todo projeto popular necessariamente tem grande volume de colaboração.
- **Releases:** Com mediana de 35,5 releases, observa-se certa regularidade na publicação de versões, mas também a presença de projetos que não adotam formalmente o mecanismo de releases.
- **Atualizações:** Embora não tenhamos os dados de tempo exato até a última atualização, a manutenção está ativa entre os repositórios populares.
- **Linguagens utilizadas:** Python foi a linguagem predominante, confirmando que linguagens populares concentram muitos dos repositórios mais estrelados no GitHub.
- **Issues fechadas:** A mediana da razão de issues fechadas de 1,14 indica que há esforço considerável na resolução de problemas e manutenção contínua desses projetos.