

Relatório Final - Características de repositórios populares

INTRODUÇÃO E HIPÓTESES INICIAIS:

Os repositórios open-source mais populares no GitHub possuem características distintas que podem indicar padrões de desenvolvimento, manutenção e engajamento da comunidade. Neste estudo, buscamos analisar 1.000 dos repositórios mais bem avaliados na plataforma, investigando métricas como idade do repositório, contribuição externa, frequência de lançamentos, tempo de atualização, linguagem utilizada e taxa de fechamento de issues.

O objetivo é entender se repositórios populares tendem a ser mais antigos, receber mais contribuições externas, serem frequentemente atualizados e escritos nas linguagens mais utilizadas. Além disso, será analisado se a popularidade de uma linguagem influencia o nível de contribuição, o número de releases e a frequência de atualizações.

Com base na intuição sobre projetos open-source populares, formulamos as seguintes hipóteses:

- **H1:** Repositórios populares tendem a ser mais antigos, com uma idade média superior a **5 anos**.
- **H2:** Projetos populares recebem uma quantidade significativa de contribuições externas, com pelo menos 1.000 pull requests aceitas em média.
- **H3:** Repositórios populares lançam releases com alta frequência, tendo um total médio de mais de 10 releases ao longo de sua existência.
- **H4:** Projetos amplamente utilizados são frequentemente atualizados, com um tempo médio desde a última atualização inferior a **30 dias**.
- **H5:** Repositórios populares tendem a ser escritos nas linguagens de programação mais utilizadas, como JavaScript, Python e TypeScript, representando pelo menos 50% do total.
- **H6:** Projetos populares possuem um alto percentual de issues fechadas, com pelo menos 70% das issues resolvidas.
- **H7 (bônus):** Repositórios escritos em linguagens populares recebem 20% mais pull requests, lançam 15% mais releases e são atualizados com maior frequência do que aqueles escritos em linguagens menos comuns.

Essas hipóteses serão avaliadas por meio da coleta e análise dos dados extraídos da API GraphQL do GitHub, permitindo a identificação de padrões e tendências entre os projetos mais populares da plataforma.

METODOLOGIA

Coleta de Dados: Utilizamos a API GraphQL do GitHub para obter informações sobre os 1.000 repositórios com mais estrelas. A consulta incluiu dados como data de criação, pull requests aceitas, releases, última atualização, linguagem principal e issues fechadas.

Armazenamento: Os dados foram extraídos e salvos em um arquivo `.csv` para facilitar a análise.

Processamento: Realizamos uma análise estatística, utilizando valores medianos para evitar distorções por outliers.

Classificação: Para métricas categóricas (ex.: linguagem), realizamos uma contagem por categoria.

Análise Comparativa: Investigamos a relação entre a popularidade da linguagem e as métricas de contribuição, releases e frequência de atualizações

RESULTADOS

RQ 01. Sistemas populares são maduros/antigos?

Métrica: idade do repositório (calculado a partir da data de sua criação)

- Idade média dos repositórios:
- Mediana da idade:

RQ 02. Sistemas populares recebem muita contribuição externa?

Métrica: total de pull requests aceitas

- Média de pull requests aceitos: 3228.26
- Mediana pull requests aceitos: 613.5

RQ 03. Sistemas populares lançam releases com frequência?

Métrica: total de releases

- Média de releases: 101.76
- Mediana de releases: 32.5

RQ 04. Sistemas populares são atualizados com frequência?

Métrica: tempo até a última atualização (calculado a partir da data de última atualização)

- Média de dias desde última atualização:

- Mediana de dias desde última atualização:

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

Métrica: linguagem primária de cada um desses repositórios

(Verificar referência e enunciado)

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

Métrica: razão entre número de issues fechadas pelo total de issues Relatório Final:

- Média do percentual de issues fechadas: 4469.19 (falta colocar em %)
- Mediana do percentual de issues fechadas: 1558.5 (falta colocar em %)

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

(Verificar referência)

DISCUSSÃO (HIPÓTESES x VALORES OBTIDOS)