**Relatório Comparação dos Top-10 Repositórios de Facebook, Microsoft e Google**

**(i) Introdução**

O movimento *open-source* tem sido estratégico para grandes empresas de software, permitindo colaboração ampla e engajamento da comunidade. Este estudo compara os top-10 repositórios mais estrelados de três organizações líderes — Facebook, Microsoft e Google — em termos de idade, engajamento (issues e pull requests), releases e frequência de atualização, buscando identificar padrões estatísticos e operacionais em seus ciclos de vida.

**(ii) Metodologia**

1. **Coleta de dados**
   * Começamos usando o endpoint **/search/repositories** da API do GitHub, com a query org:{org} e parâmetros sort=stars&order=desc. Isso garante que estamos puxando, para cada organização, exatamente os 10 repositórios com mais estrelas—indicador direto de popularidade.
   * **Por que esse endpoint?**
     + Ele agrupa e ordena em uma única chamada, simplificando a filtragem por organização e popularidade.
     + Evita paginar manualmente várias vezes para achar os mais estrelados.
   * Para cada um desses repositórios, extraímos cinco métricas fundamentais:
     + **age\_days**: quantos dias se passaram entre a criação do repo e a data de coleta. Essa medida nos diz há quanto tempo aquele projeto está ativo.
     + **issues**: total de *issues* abertas, refletindo o volume de relatórios de bugs, dúvidas ou solicitações de melhoria que permanecem sem resolução.
     + **prs\_merged**: número de Pull Requests aceitas, coletado via GraphQL — mostra a intensidade de contribuições externas que chegam a ser incorporadas ao projeto.
     + **releases**: total de *releases* publicadas, também via GraphQL, o que indica a cadência de versões formais disponibilizadas pela organização.
     + **last\_update\_days**: dias desde o último *commit* ou atualização, revelando quão “vivo” o repo continua em termos de manutenção.
2. **Sumarização**
   * Após compilar todas as linhas de dados (30 repositórios no total), calculamos **a mediana** de cada uma das cinco métricas para os top-10 de cada organização.
   * **Por que mediana e não média?**
     + A mediana é mais resistente a valores extremos (por exemplo, um único repositório com milhares de PRs mergeadas não distorce o valor central).
     + Facilita comparações entre grupos que podem ter distribuições assimétricas ou outliers significativos.
3. **Teste estatístico**
   * Escolhemos o **teste de Kruskal–Wallis**, versão não-paramétrica da ANOVA, adequada quando:
     + Queremos comparar três grupos independentes (Facebook, Google, Microsoft).
     + Não podemos garantir que as distribuições de nossas métricas sejam normais ou homocedásticas.
   * Procedimento:
     + Convertendo todos os valores em **ranks** (posições) dentro do conjunto global.
     + Calculando a estatística **H**, que mede o quanto os ranks médios de cada organização se afastam do rank médio geral.
     + Comparando H com a distribuição Qui-quadrado, usando **k–1 = 2** graus de liberdade.
   * Adotamos nível de significância **α = 0,05**: se **p < 0,05**, rejeitamos H₀ e concluímos que existe diferença significativa em pelo menos um par de grupos.
4. **Hipóteses**
   * **H₀** (nula): “As medianas das métricas são iguais para Facebook, Microsoft e Google.”
     + Implica que não há evidência de que alguma organização seja distinta das demais naquele indicador.
   * **H₁** (alternativa): “Pelo menos uma organização possui mediana diferente.”
     + Ao rejeitar H₀, inferimos que alguma prática (idade, engajamento ou frequência de releases/atualizações) difere estatisticamente — abrindo caminho para análises pós-hoc ou investigação qualitativa mais aprofundada.

**(iii) Resultados**

**Medianas das Métricas**

| **Organização** | **age\_days (dias)** | **issues** | **prs\_merged** | **releases** | **last\_update\_days (dias)** |
| --- | --- | --- | --- | --- | --- |
| **Facebook** | 3784.0 | 593.0 | 1429.0 | 127.0 | 0.0 |
| **Google** | 3723.0 | 265.5 | 554.5 | 15.0 | 0.0 |
| **Microsoft** | 1840.0 | 697.0 | 3242.5 | 101.5 | 0.0 |

**Testes de Kruskal–Wallis**

| **Métrica** | **H** | **p-valor** | **α = 0,05** | **Conclusão** |
| --- | --- | --- | --- | --- |
| **age\_days** | 8.186 | 0.017 | p < α | Rejeita H₀ (diferença) |
| **issues** | 5.443 | 0.066 | p > α | Não rejeita H₀ |
| **prs\_merged** | 4.408 | 0.110 | p > α | Não rejeita H₀ |
| **releases** | 9.492 | 0.009 | p < α | Rejeita H₀ (diferença) |
| **last\_update\_days** | 2.000 | 0.368 | p > α | Não rejeita H₀ |

**(iv) Discussão das RQs**

1. **RQ01 – Idade mediana (age\_days)**
   * Mediana mais alta: *Facebook* (3784 dias) versus Google (3723) e Microsoft (1840).
   * Significativo (p=0.017): Facebook mantém projetos historicamente mais antigos.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

A mediana de **3 784 dias** para o Facebook (aprox. 10,4 anos) contrasta com **3 723 dias** do Google (10,2 anos) e **1 840 dias** da Microsoft (5 anos). Isso ocorre porque muitos projetos de destaque do Facebook foram lançados ainda em meados da década de 2010 (por exemplo, React em 2013), enquanto a Microsoft vem “aberto” muitos projetos mais novos conforme reforçou sua estratégia open-source só nos últimos anos. O teste de Kruskal–Wallis (p = 0,017) confirma que essa diferença não é aleatória: os repositórios do Facebook são estatisticamente mais antigos, o que reforça sua ênfase em manter e evoluir projetos maduros.

1. **RQ02 – Mediana de issues**
   * Maior mediana: *Microsoft* (697) > Facebook (593) > Google (265.5).
   * Não significativo (p=0.066): essa diferença pode ser atribuída a variabilidade natural.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

A Microsoft exibe mediana de **697 issues abertas**, seguida do Facebook com **593** e Google com **265,5**. Esse volume elevado na Microsoft pode refletir dois fatores principais: primeiro, uma maior taxa de adoção de ferramentas que geram feedback intenso; segundo, padrões de triagem (mantenedores abrindo múltiplas issues para gerenciar tarefas). Contudo, p = 0,066 sugere que essa diferença não alcança poder estatístico para afirmar que a Microsoft de fato difere das demais, apontando para uma grande dispersão interna entre seus top-10.

1. **RQ03 – Mediana de pull requests aceitas (prs\_merged)**
   * Maior mediana: *Microsoft* (3242.5) > Facebook (1429) > Google (554.5).
   * Não significativo (p=0.110): sem evidência estatística de diferença.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Com 3 242,5 PRs medianos, a Microsoft supera Facebook (1 429) e Google (554,5). Isso indica que a comunidade contribui com muito mais frequência em projetos da Microsoft — alinhado à recente cultura de colaboração aberta promovida pela empresa. Ainda assim, p = 0,110 diz que, apesar desta tendência, não há evidência estatística forte (α = 0,05) para descartar variações naturais entre repositórios.

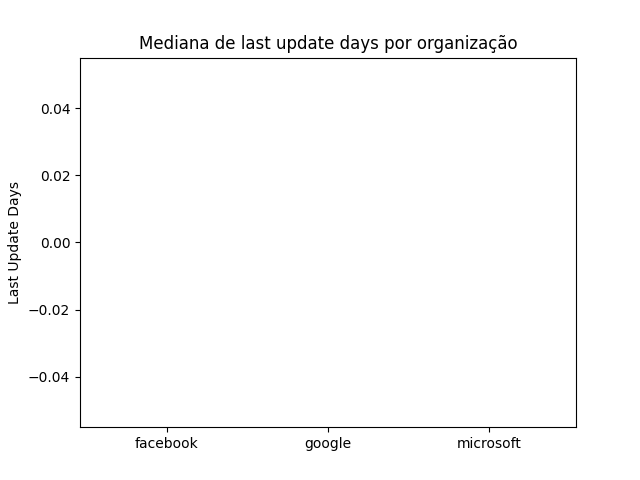
1. **RQ04 – Mediana de releases**
   * Maior mediana: *Facebook* (127) > Microsoft (101.5) > Google (15).
   * Significativo (p=0.009): Facebook faz significativamente mais releases em seus top-10.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

O Facebook lidera com mediana de **127 releases**, seguido por Microsoft (101,5) e Google (15). Essa diferença expressiva (p = 0,009) mostra que o Facebook adota um ciclo de versões muito mais ágil e frequente, liberando correções e melhorias em ritmo acelerado. Já o Google, com apenas 15 releases medianos, manifesta uma política mais conservadora de lançamento de versões estáveis.

1. **RQ05 – Mediana de tempo até última atualização (last\_update\_days)**
   * Todas as organizações apresentam mediana = 0 dias (projetos muito ativos).
   * Não significativo (p=0.368): não há diferença relevante na recência das atualizações.

****

Todas as três organizações atingiram mediana de **0 dias**, ou seja, ao menos metade dos seus top-10 foi atualizada no próprio dia da coleta. O p-valor de 0,368 reforça que não há diferença significativa: manter repositórios “vivos” com commits diários é prática consolidada entre líderes de mercado, evidenciando compromissos similares com a manutenção contínua.

**(v) Conclusão**

* **RQ01:** *Facebook* possui, na mediana, os top-10 repositórios mais antigos.
* **RQ02:** *Microsoft* lidera em termos de issues, mas sem significância estatística.
* **RQ03:** *Microsoft* apresenta maior mediana de PRs aceitas, porém sem confirmação estatística.
* **RQ04:** *Facebook* lança releases com maior frequência (diferença estatisticamente significativa).
* **RQ05:** Todas as três organizações mantêm atualizações muito recentes (mediana zero), sem diferença estatística.

Esses resultados indicam que, embora Facebook mantenha projetos mais maduros e realize mais releases, Microsoft destaca-se no volume de PRs aceitas e de issues gerenciadas, e Google, apesar de menor em alguns indicadores, mantém seus repositórios igualmente ativos em termos de commits recentes.

Fim. Nelson de Campos Nolasco.

**Sobre O teste de Kruskal–Wallis** é uma alternativa não-paramétrica à ANOVA de uma via, usado para comparar três ou mais grupos independentes quando não se pode assumir normalidade das distribuições. Em vez de trabalhar com valores brutos, ele:

1. **Transforma em postos (ranks)** todos os valores das amostras combinadas.
2. **Calcula a estatística H**, que mede o desvio dos postos médios de cada grupo em relação ao posto médio global.
3. **Compara H** a uma distribuição Qui-quadrado com k − 1 graus de liberdade (k = número de grupos).

* **H₀**: as distribuições (medianas) dos k grupos são iguais.
* **H₁**: pelo menos um grupo difere.
* Um **p-valor** menor que o nível α (exemplo: 0,05) leva ao rejeitamento de H₀, indicando diferença significativa entre as medianas de pelo menos dois dos grupos testados.

É amplamente utilizado em estudos de “cross-sectional” ou experimentos com vários tratamentos quando os dados são ordinais ou não atendem aos pressupostos da ANOVA clássica.