**Analise de Métricas de Analise Utilizando Git**

**Discente: Bruno Rodrigues Conceição**

**Docente: José Grimaldo**

Departamento de Ciências Exatas e da Terra (DCET)  
Universidade Estadual da Bahia (UNEB) – Salvador, BA – Brazil

Link do repositório escolhido

“ <https://github.com/pandas-dev/pandas> ” e “ <https://github.com/pola-rs/polars> ”

**Repositórios Escolhido**

Os repositórios escolhidos para analise foram os Pandas e Polars, duas bibliotecas open source de análise de dados em Pyhton, vastamente utilizada na área de ciências de dados.

Pandas é umas das bibliotecas mais populares do Python, oferecendo uma API robusta e rica em funcionalidades, ideal para manipulação de dados de tamanho moderado. Já o Polars se destaca por sua alta performance e eficiência em processamento de grandes volumes de dados.

Pandas: Possuindo 43.3 estrelas e 17.8 Forks e com mais de 3000 contribuidores

Pandas foi criado no ano de 2009

Polars: Possuindo 29.3 estrelas e 1.9 Forks e com mais de 450 contribuidores

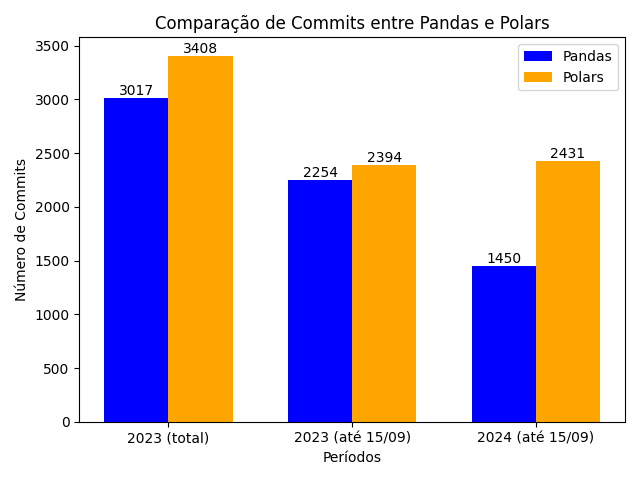
Polars foi criado no ano de 2020

**Tecnologia Escolhida**

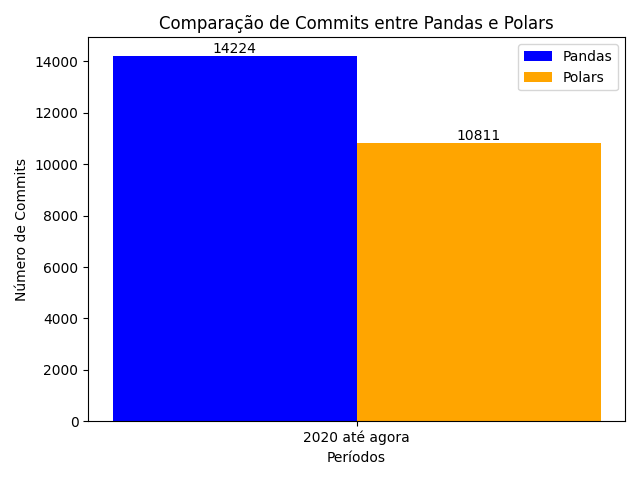
A tecnologia escolhida foi a linguagem de programação Python, junto dela a biblioteca GitPython para desenvolver o código.

**Métrica 1**

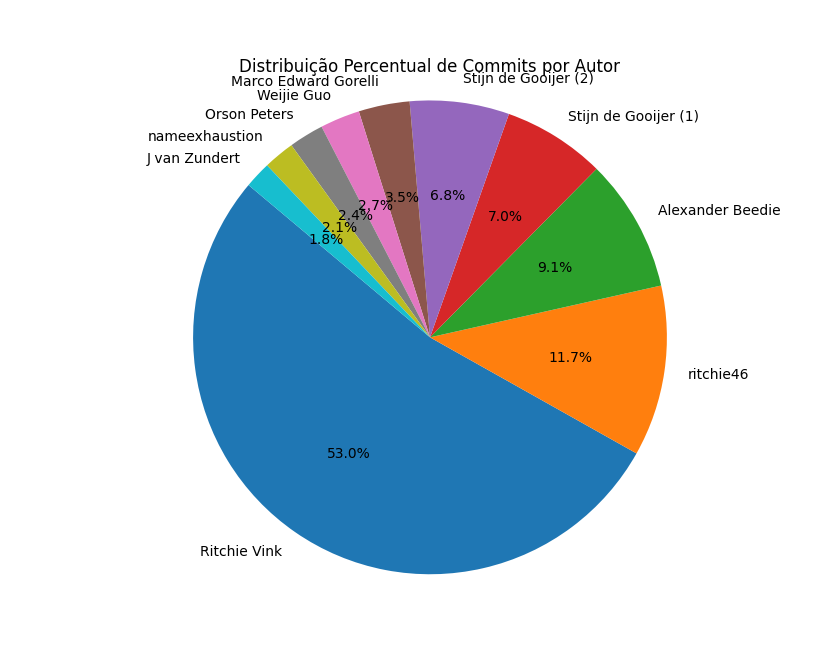
A métrica 1 foi destinada a medir o engajamento da equipe, avaliando aspectos como a quantidade de commits feitos nos anos de 2023 e 2024, as 10 pessoas que mais contribuíram e suas respectivas métricas, a quantidade total de contribuições no repositório e o número de Pull Requests desse repositório.



Comparação de três períodos diferentes, sendo o primeiro referente a todo o ano de 2023, para comparar dois repositórios que tiveram números de commits semelhantes, ambos acima de 3.000 commits. O repositório Polars teve um pouco mais de commits do que o Pandas. Ao comparar o número de commits no recorte de 01/01/2023 até 15/09 do mesmo ano com o período de 01/01/2024 até 15/09, notamos que o Polars mantém uma média semelhante de commits, indicando que o repositório está em constante crescimento. Por outro lado, ao analisarmos o Pandas, observamos uma queda considerável nessa comparação entre 2023 e 2024.

****

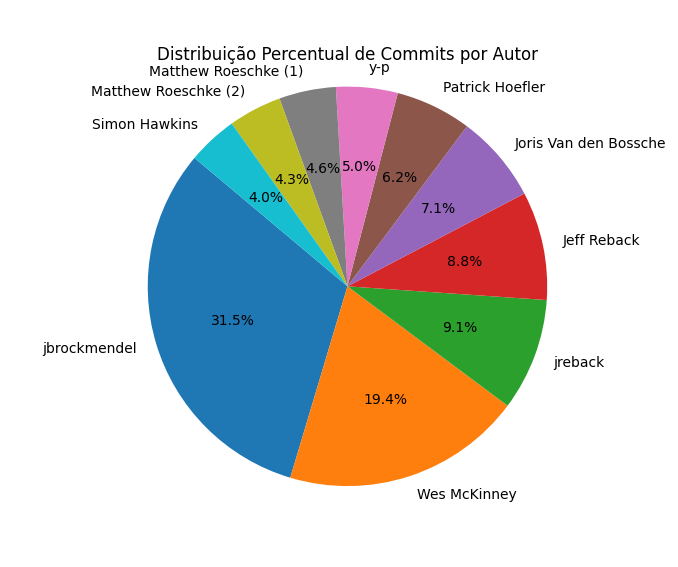
Agora, se pegarmos desde 2020, escolhemos esse ano porque foi quando o repositório Polars foi criado. Ao analisarmos o período desde essa data até a data atual (18 de setembro), podemos ver que, mesmo com a queda em 2024, o Pandas ainda tem o maior número de commits. Isso sugere que 2024 pode ser um ano mais focado em estruturação e novas ideias para a biblioteca, em vez de apenas implementar correções de bugs.

****

**Polars**

* 4515 Ritchie Vink
* 994 ritchie46
* 774 Alexander Beedie
* 597 Stijn de Gooijer
* 578 Stijn de Gooijer
* 297 Marco Edward Gorelli
* 231 Weijie Guo
* 203 Orson Peters
* 181 nameexhaustion
* 155 J van Zundert

Podemos observar uma alta taxa de concentração de commits no repositório Polars, considerando que apenas a primeira conta é dele, representando cerca de mais de 50% dos commits. Esse é um número elevado em comparação com os outros desenvolvedores do projeto, especialmente se analisarmos que o repositório possui mais de 400 contribuidores.

****

**Pandas**

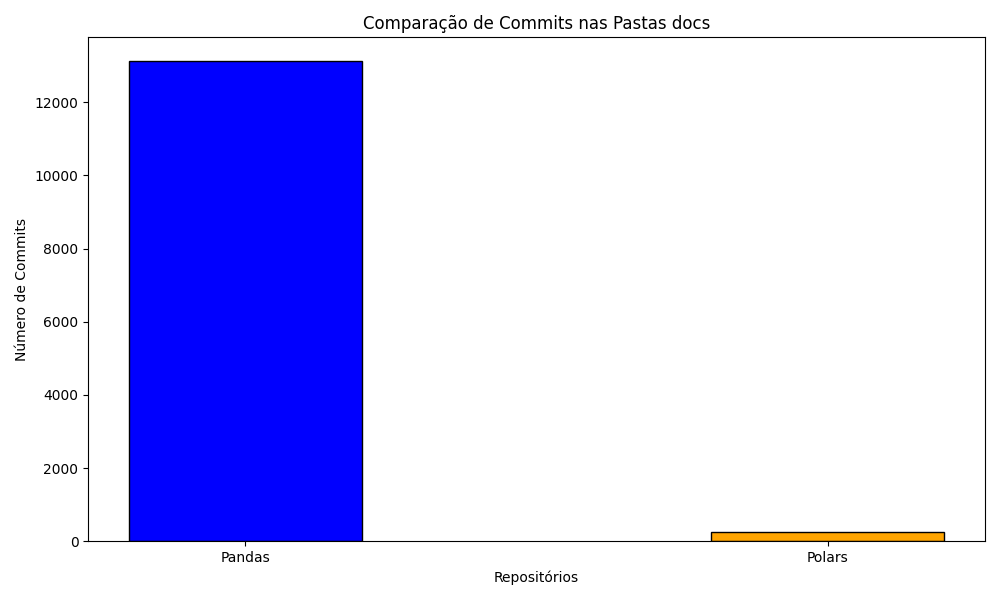
* 4836 jbrockmendel
* 2984 Wes McKinney
* 1398 jreback
* 1349 Jeff Reback
* 1092 Joris Van den Bossche
* 945 Patrick Hoefler
* 765 y-p
* 707 Matthew Roeschke
* 662 Matthew Roeschke
* 622 Simon Hawkins

Por outro lado, no Pandas, vemos uma distribuição mais igualitária entre os desenvolvedores, sendo que o maior contribuinte possui pouco mais de 30% dos commits no repositório, considerando que ele utiliza apenas essa conta.

Podemos concluir que o Polars apresenta um monopólio maior de uma única pessoa, que conhece bem o sistema, ou que essa pessoa foi o grande impulsionador do projeto, o que explica sua alta participação nos commits. Por outro lado, o Pandas mostra uma distribuição mais equilibrada entre os contribuidores.

**Métrica 2**

A métrica 2 foi destinada a avaliar a saúde do projeto, verificando se a documentação é atualizada com frequência e se o readme é ajustado a cada atualização do projeto.

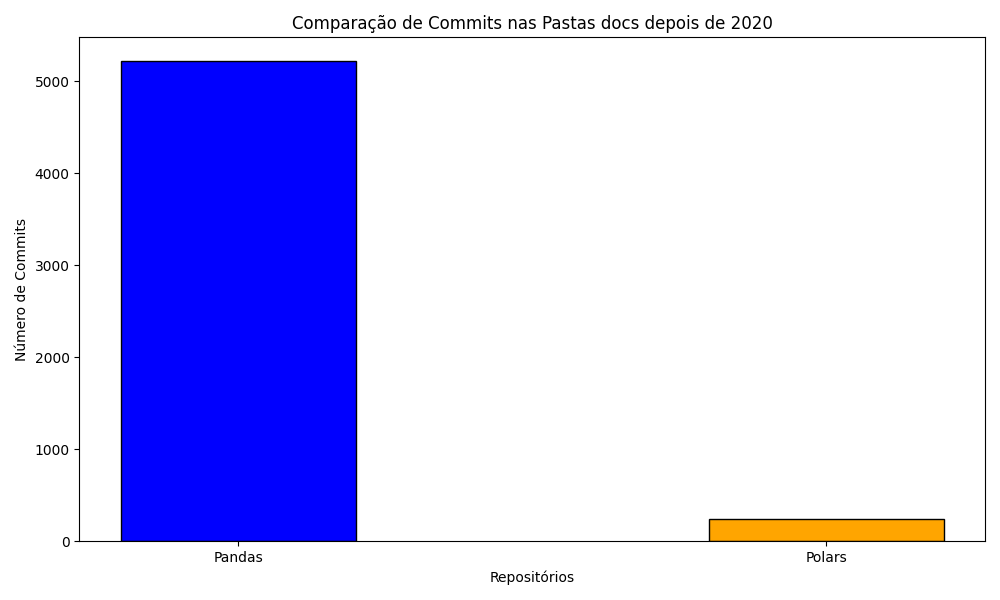


**Polars**

239

**Pandas**

13119

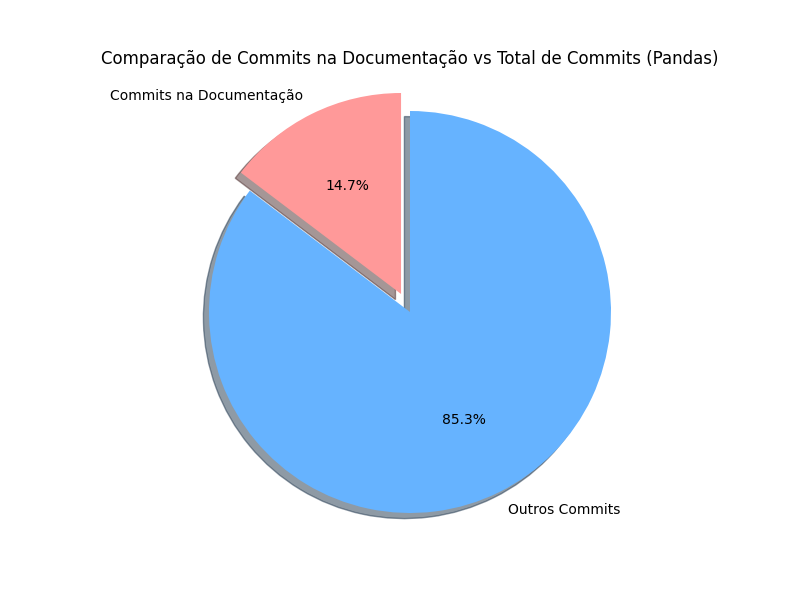
****

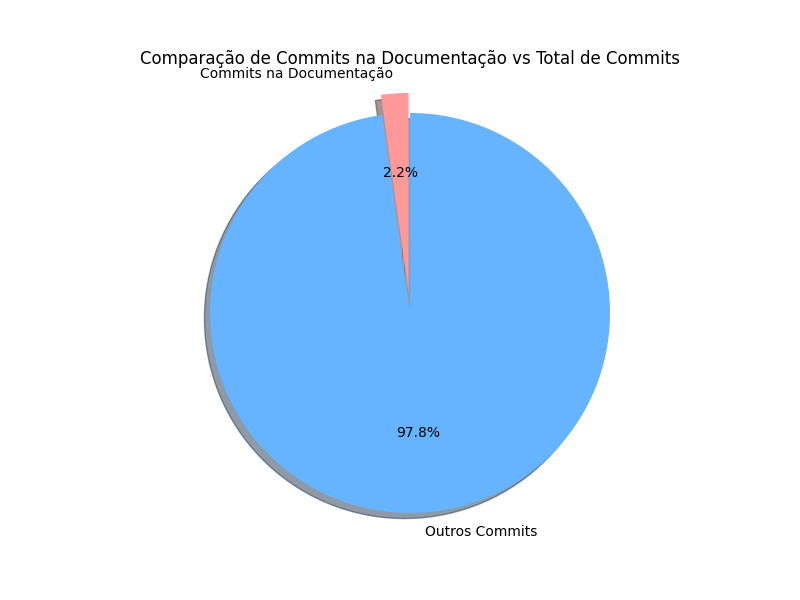
**Polars**

239

**Pandas**

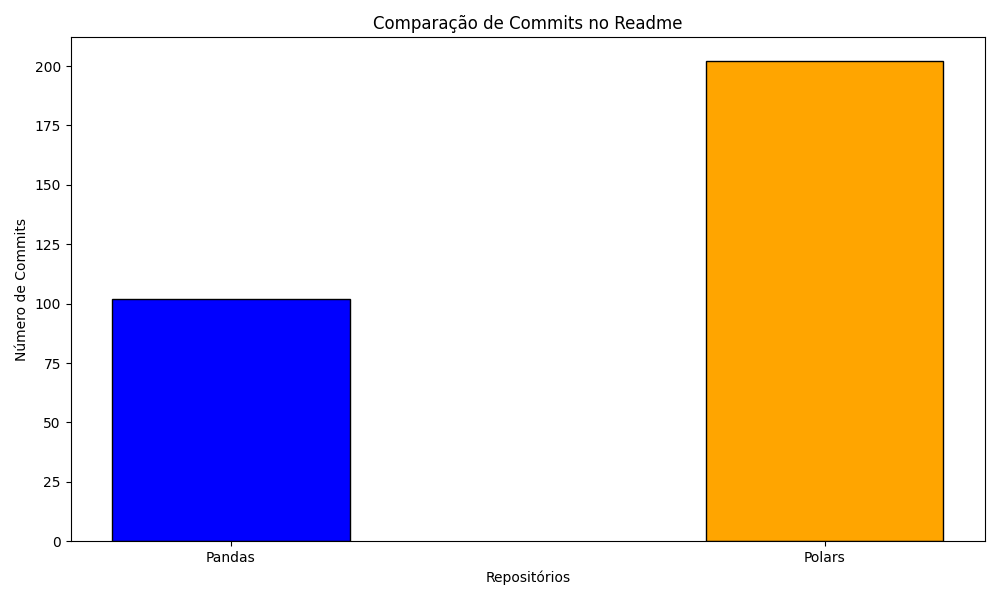
5215





**Analise**

Quando analisamos a métrica de documentação, esperamos uma quantidade relativamente alta de commits, especialmente em repositórios recém-criados. Ao examinarmos os dois repositórios desde sua criação até o período atual (18 de setembro de 2023), notamos uma grande diferença entre eles, com o Pandas ultrapassando 13 mil commits e o Polars registrando cerca de 200. No entanto, ao fazermos um recorte desde a criação do Polars, em 2020, até o período atual, ainda observamos que o número de commits é muito menor em comparação com o Pandas. Isso indica que, mesmo após mais de 10 anos desde sua criação, a documentação do Pandas continua recebendo um número significativo de commits, apresentando uma atenção a mais em sua documentação.

****

**Polars**

202

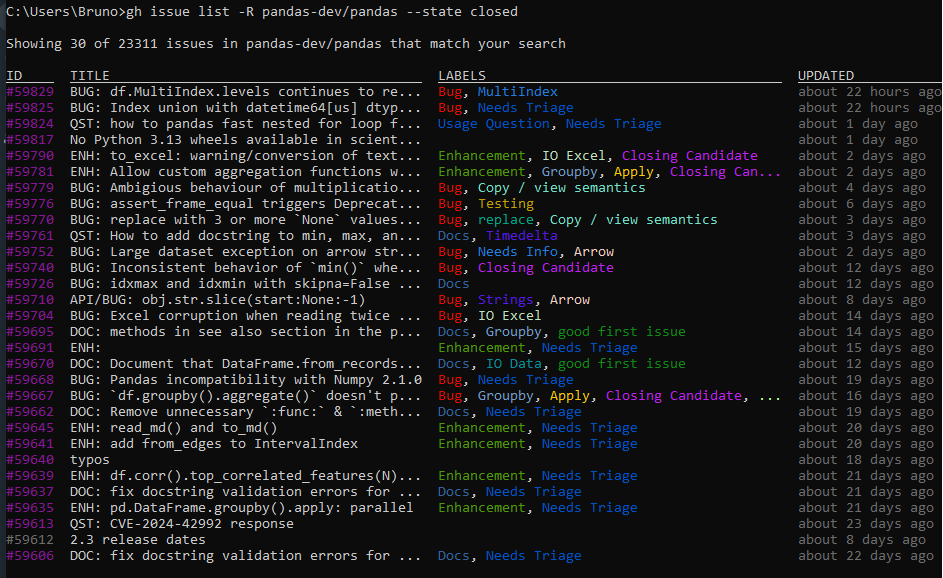
**Pandas**

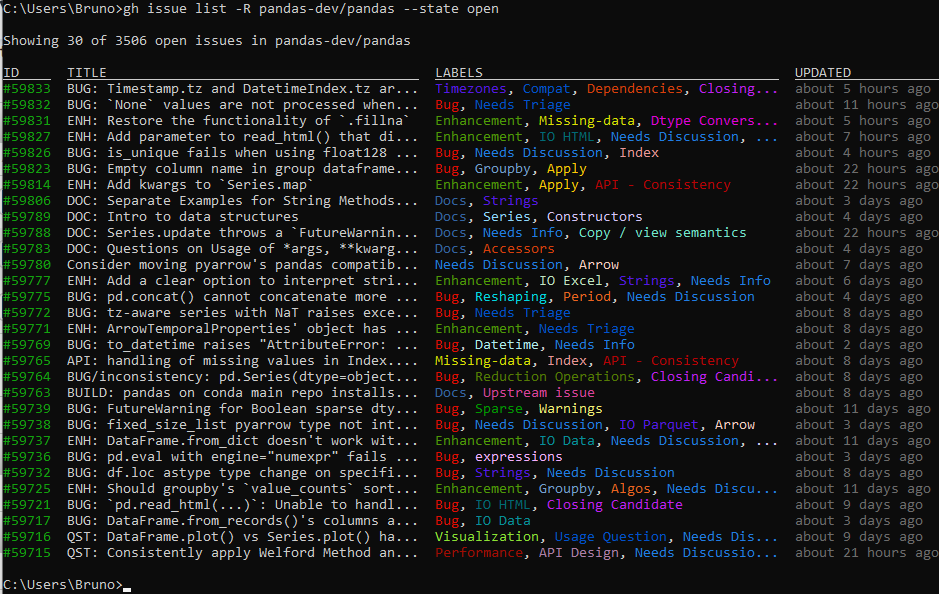
102

Comparamos não apenas a pasta de documentação, mas também o README, que mostra que o Pandas tem praticamente o dobro de commits em comparação com o Polars. O README também serve como documentação para o repositório.

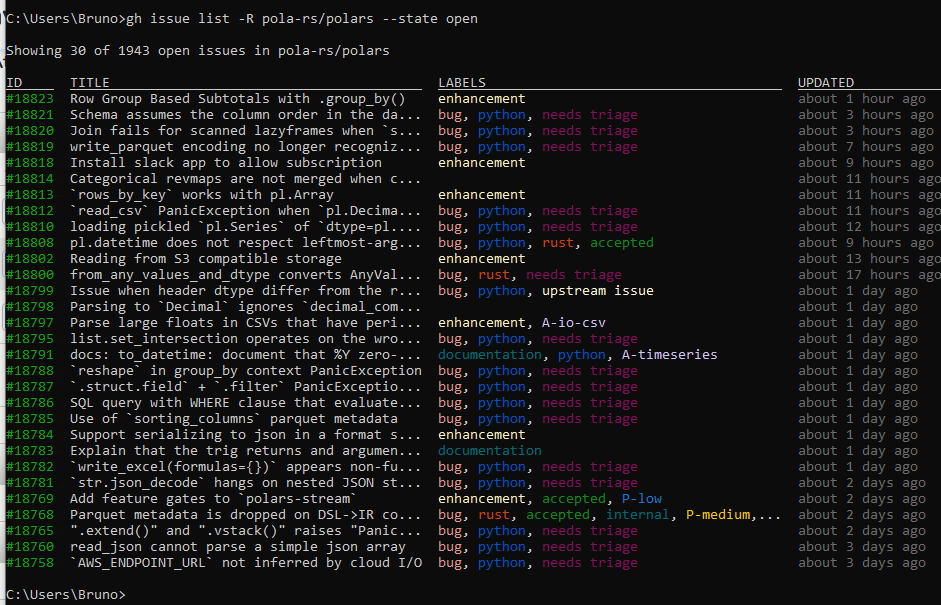
**Métrica 3**

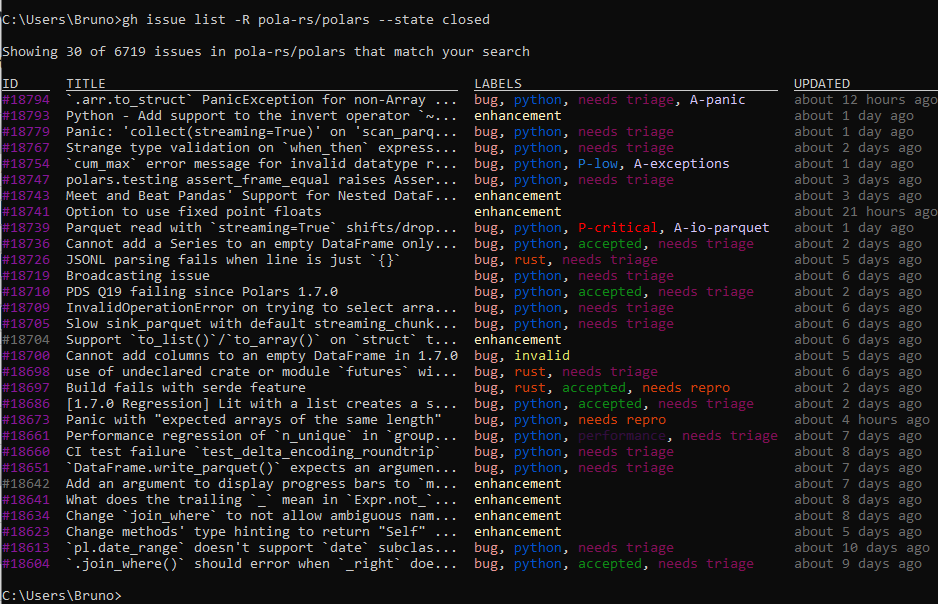
A métrica 3 foi destinada a medir o engajamento da comunidade, verificando o número de issues, que funcionam como um fórum da comunidade do repositório, e o número de Pull Requests também relacionadas a esse repositório.





Pandas conta com 23311 Issues fechados e 3506 issues abertos





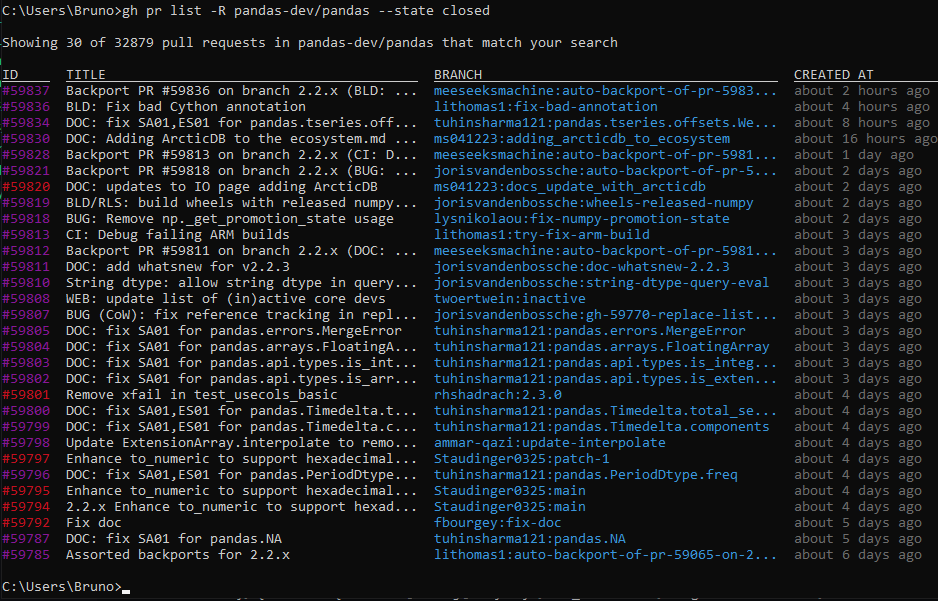
Polars conta com 6719 Issues fechados e 1943 issues abertos

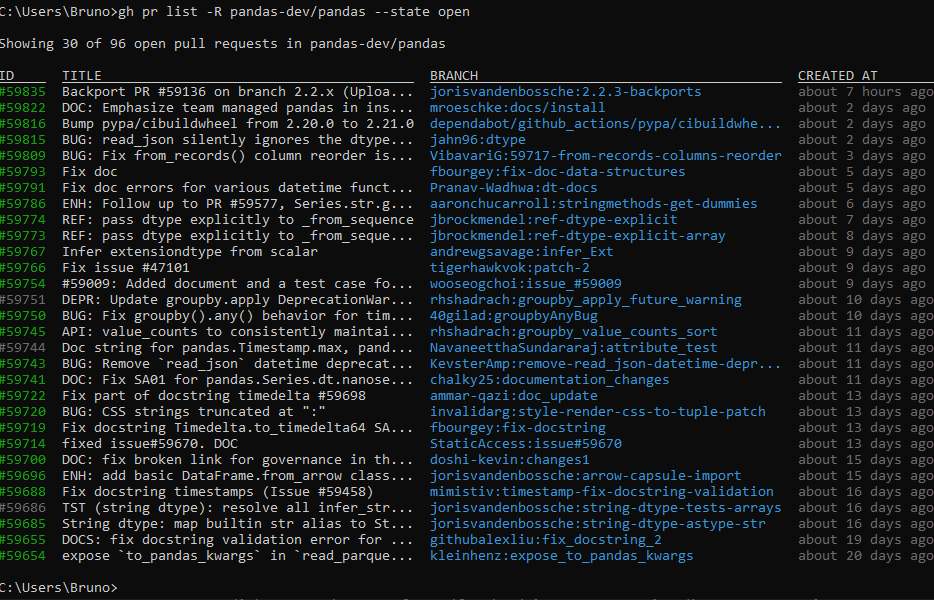
**Analise Issues**

Ao avaliar as issues abertas e fechadas de ambos os repositórios, podemos notar informações sobre a saúde e a atividade da comunidade de desenvolvedores dentro deles.

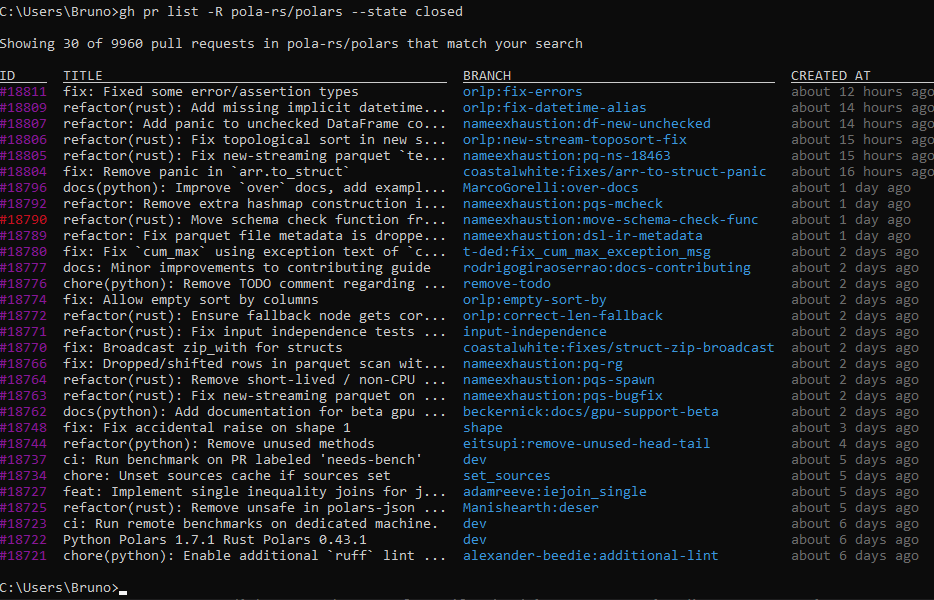
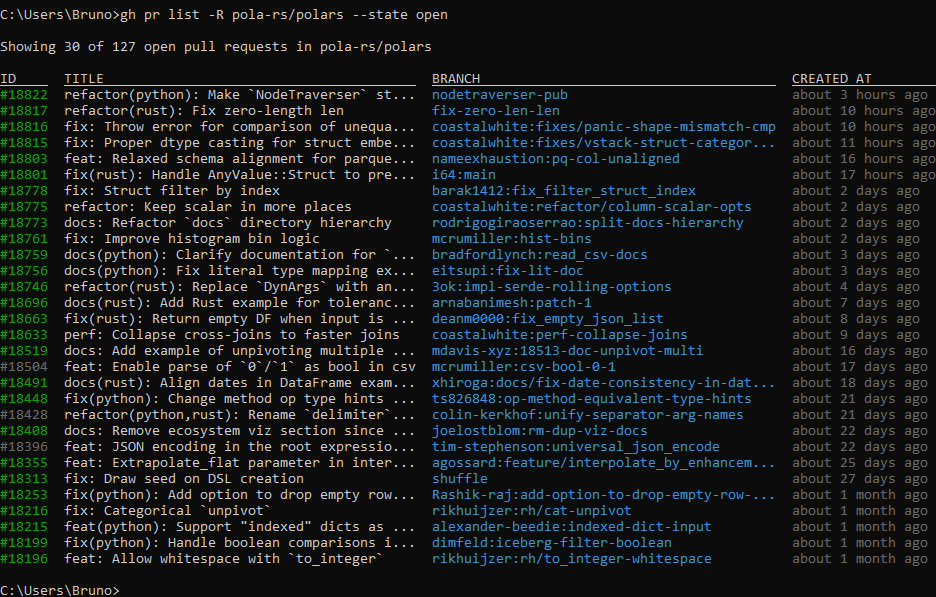
No Pandas, temos um total de 23.311 issues fechadas e 3.506 abertas. Analisamos que a comunidade é bem ativa e engajada em reportar problemas e na resolução, possuindo um número elevado de issues fechadas. Isso indica que muitos problemas e solicitações relatados foram resolvidos ao longo do tempo, refletindo a maturidade do projeto, com aproximadamente 1.550 issues resolvidas por ano. Por outro lado, com 3.506 issues abertas, a quantidade é relevante, mas, em comparação com as fechadas, percebemos que a maioria das questões é resolvida.

Já ao avaliar o Polars, por ser um repositório relativamente recente em comparação com o Pandas, possui 6.719 issues fechadas e 1.943 abertas. Mesmo sendo um repositório novo, apresenta um número considerável de issues fechadas, com cerca de 1.600 issues resolvidas por ano. Em comparação com o Pandas, o número de issues fechadas por ano é um pouco menor. Contudo, por ser um repositório em fase de crescimento e maturação, a quantidade de issues abertas é normal para novos projetos. Podemos notar que o engajamento da comunidade está crescendo, pois há um número significativo de problemas e correções sendo relatados.





Pandas conta com 32879 Pull request fechados e 96 Pull request abertos



Polars conta com 9960 Pull request fechados e 127 Pull request abertos

**Analise PRs**

O Pandas apresenta um total de 32.879 pull requests fechados e apenas 96 abertos. Esse grande número de pull requests fechados indica um histórico robusto de contribuições e melhorias implementadas ao longo do tempo. O número massivo de PRs fechadas demonstra a eficácia da equipe na revisão e integração das contribuições, tanto internas quanto externas, com o apoio da comunidade, refletindo assim em uma biblioteca sólida. O baixo número de PRs abertas sugere que a equipe do projeto se dedica a revisar e aceitar ou não as contribuições de maneira ágil.

Por outro lado, o Polars conta com 9.960 pull requests fechados e 127 abertos. Embora o número de pull requests fechados seja considerável, especialmente para um repositório recente, podemos notar que o número de PRs abertas não é tão elevado. Isso indica que o Polars também possui uma equipe dedicada à revisão do código e, dado que é um repositório relativamente novo, é mais aceitável ter um aumento no número de PRs abertas.

**Conclusão:**

Chegamos à conclusão de que o Polars é uma biblioteca bastante promissora e em ascensão, mas ainda é muito nova, o que pode resultar em um maior número de bugs e problemas na implementação dentro de um projeto. Isso significa que, ao optar pelo Polars, a equipe de implementação pode se deparar com desafios adicionais e a necessidade de adaptar suas soluções à medida que a biblioteca evolui.

Por outro lado, temos o Pandas, uma biblioteca mais consolidada no mercado, com um tempo de vida maior e uma consolidação também maior. Isso a torna mais apropriada para aplicações práticas, especialmente em projetos que exigem estabilidade. A experiência acumulada da comunidade e o vasto conjunto de recursos disponíveis, incluindo documentação, tutoriais e suporte, tornam o Pandas uma escolha confiável para desenvolvedores que buscam eficiência e resultados rápidos.

**Referência**

Pandas. Disponível em “ <https://github.com/pandas-dev/pandas> ” acessado em 16/09/2024.

Polars. Disponível em “ <https://github.com/pola-rs/polars> ” acessado em 16/09/2024.

GitPython. Disponível em “https://gitpython.readthedocs.io/en/stable/index.html ” acessado em 16/09/2024.

Github CLI. Disponível em “https://cli.github.com” acessado em 18/09/2024.