

Análise de Métricas de Análise 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 análise foram os Pandas e Polars, duas bibliotecas open source de análise de dados em Python, vastamente utilizada na área de ciências de dados. Pandas é uma 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

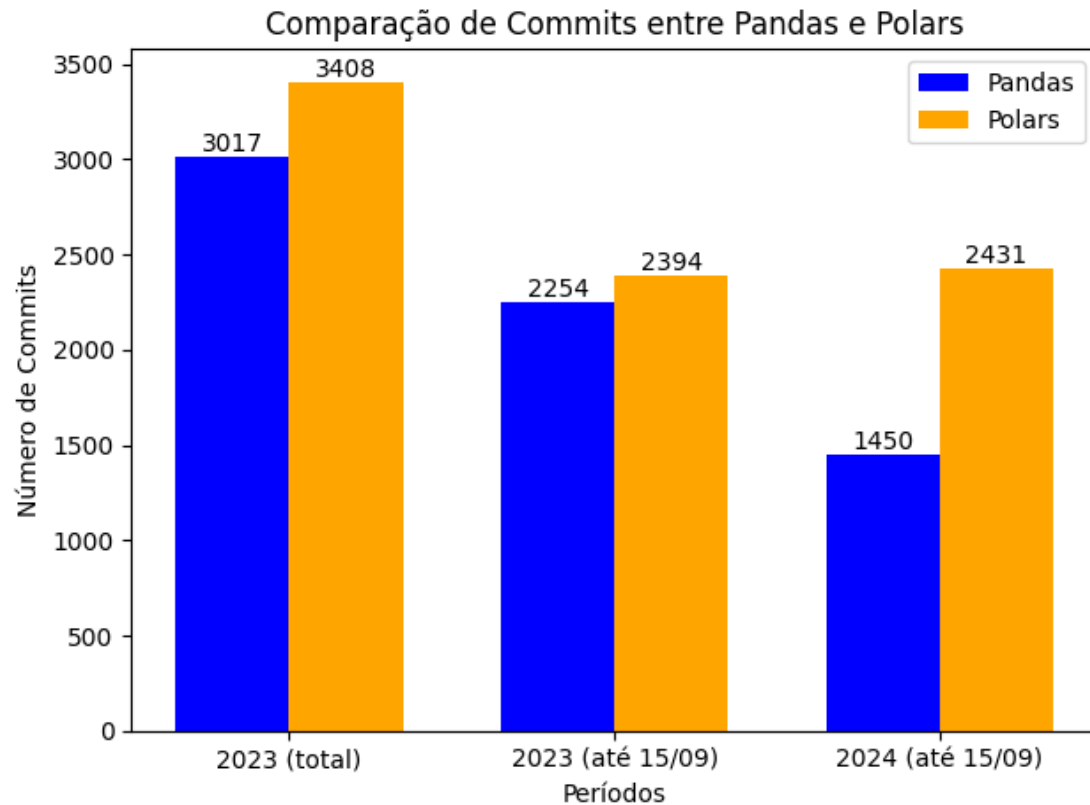
Polars foi criado no ano de 2020

Tecnologia Escolhida

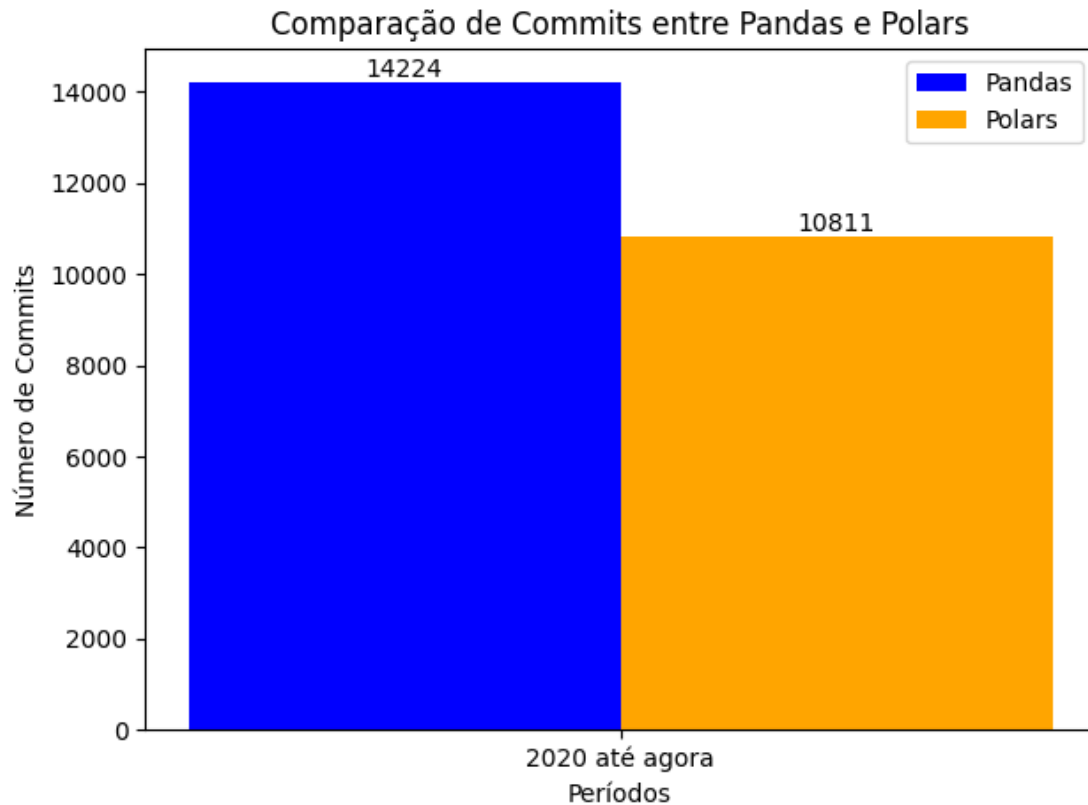
A tecnologia escolhida foi a linguagem de programação Python, junto dela a biblioteca GitPython para desenvolver o código.

Métrica 1

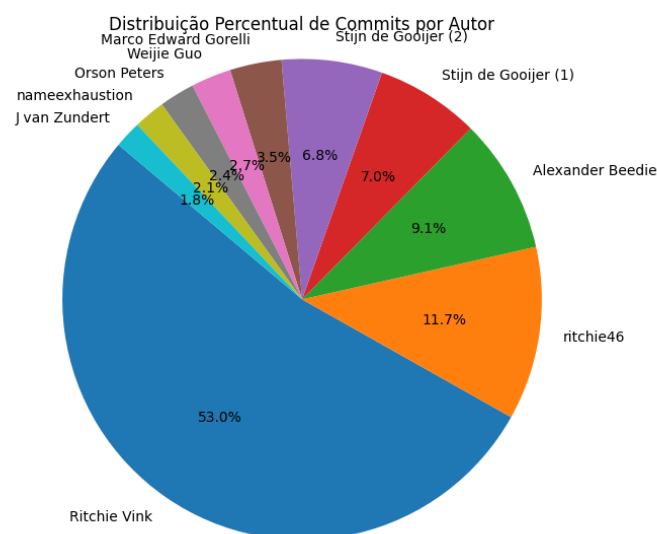
A métrica 1 foi destinada a medir o engajamento da equipe, avaliando aspectos como a quantidade de commits feitos nos anos de 2023 e 2024, as 10 pessoas que mais contribuíram e suas respectivas métricas, a quantidade total de contribuições no repositório e o número de Pull Requests desse repositório.



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.



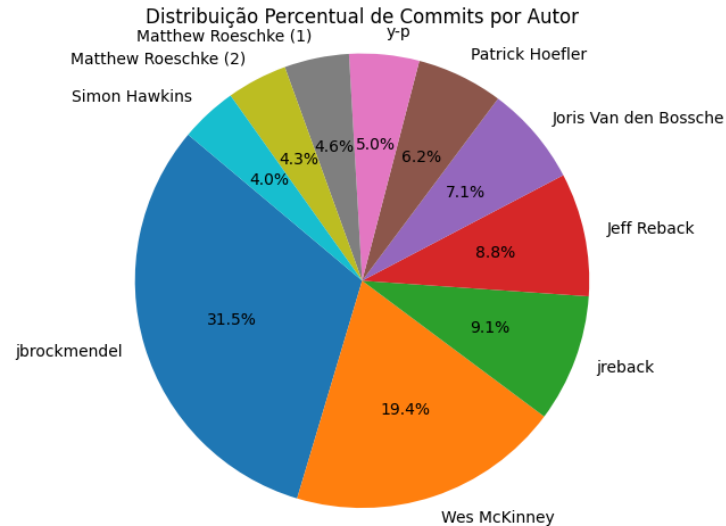
Agora, se pegarmos desde 2020, escolhemos esse ano porque foi quando o repositório Polars foi criado. Ao analisarmos o período desde essa data até a data atual (18 de setembro), podemos ver que, mesmo com a queda em 2024, o Pandas ainda tem o maior número de commits. Isso sugere que 2024 pode ser um 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 |

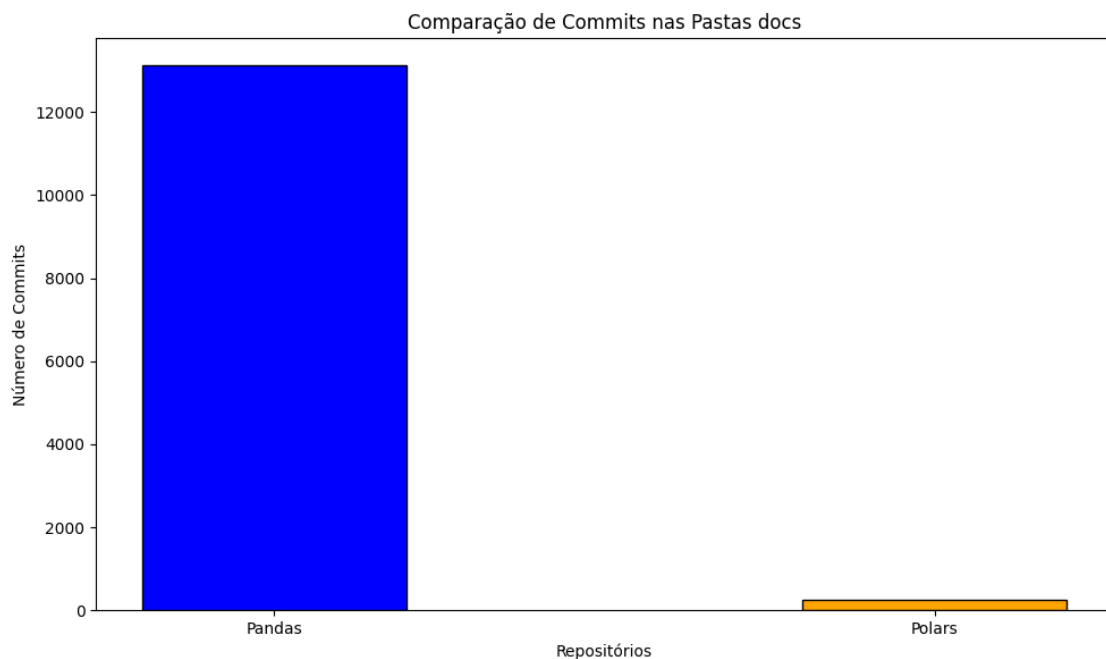
- 707 Matthew Roeschke
- 662 Matthew Roeschke
- 622 Simon Hawkins

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.

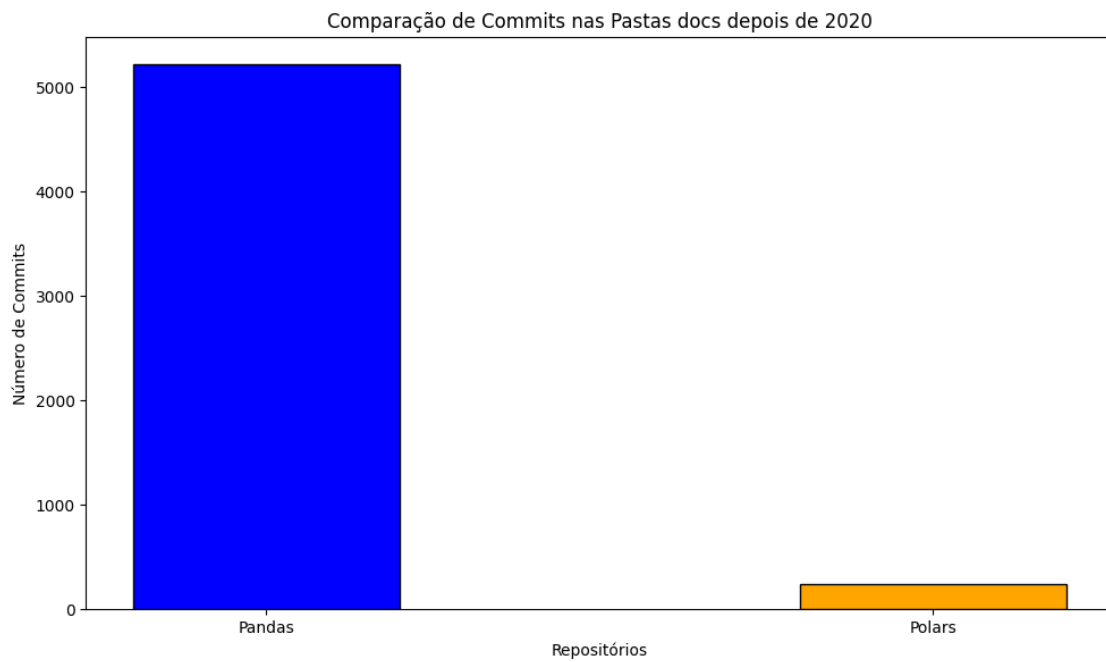


Polars

239

Pandas

13119



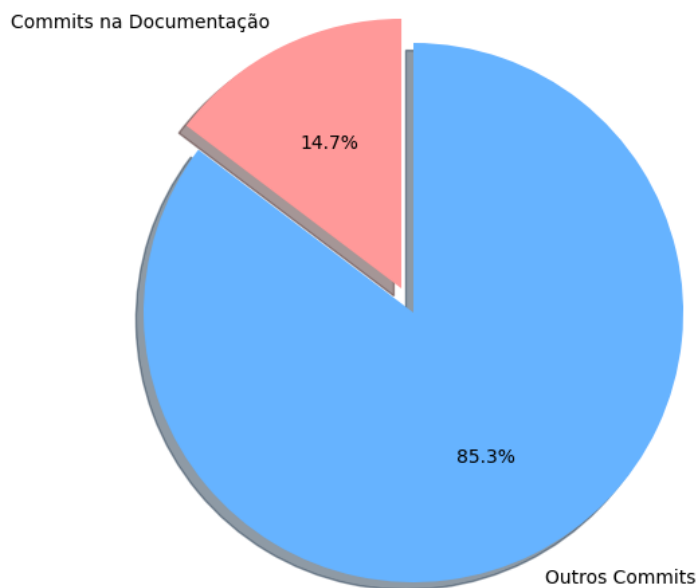
Polars

239

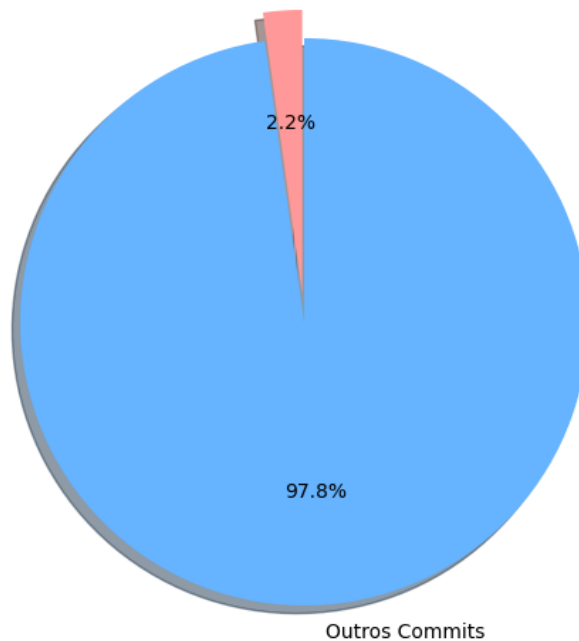
Pandas

5215

Comparação de Commits na Documentação vs Total de Commits (Pandas)

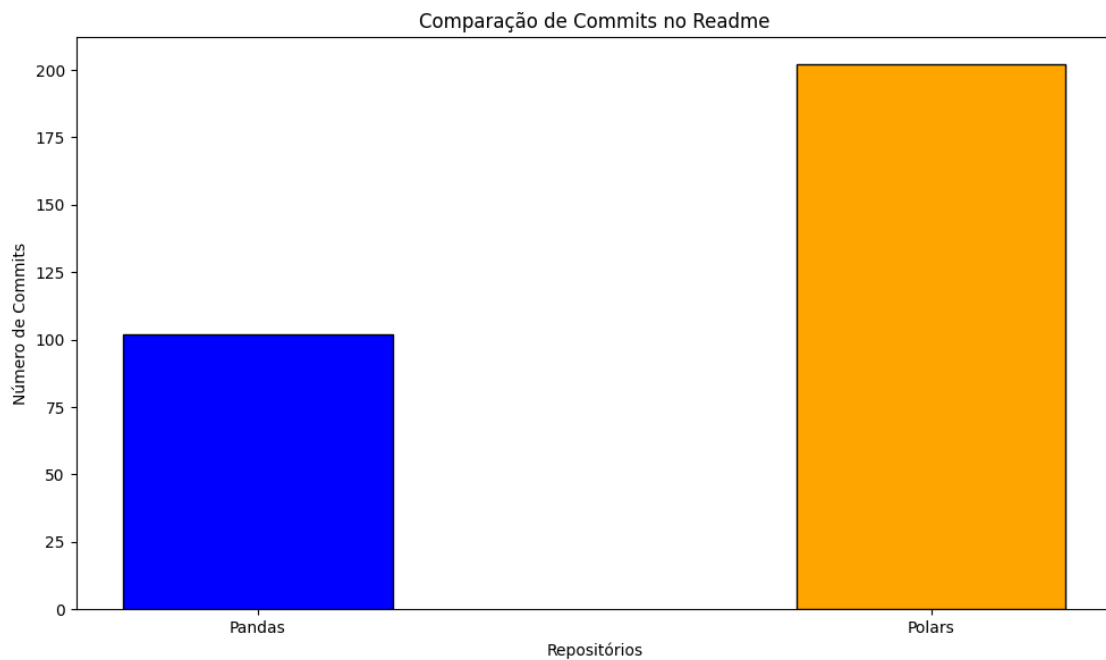


Comparação de Commits na Documentação vs Total de Commits
Commits na Documentação



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.


```
C:\Users\Bruno>gh issue list -R pandas-dev/pandas --state closed

Showing 30 of 23311 issues in pandas-dev/pandas that match your search



| ID     | TITLE                                         | LABELS                                      | UPDATED            |
|--------|-----------------------------------------------|---------------------------------------------|--------------------|
| #59829 | BUG: df.MultiIndex.levels continues to re...  | Bug, MultiIndex                             | about 22 hours ago |
| #59825 | BUG: Index union with datetime64[us] dtype... | Bug, Needs Triage                           | about 22 hours ago |
| #59824 | QST: how to pandas fast nested for loop f...  | Usage Question, Needs Triage                | about 1 day ago    |
| #59817 | No Python 3.13 wheels available in scient...  |                                             | about 1 day ago    |
| #59798 | ENH: to_excel: warning/conversion of text...  | Enhancement, IO Excel, Closing Candidate    | about 2 days ago   |
| #59781 | ENH: Allow custom aggregation functions w...  | Enhancement, Groupby, Apply, Closing Can... | about 2 days ago   |
| #59779 | BUG: Ambiguous behaviour of multiplicatio...  | Bug, Copy / view semantics                  | about 4 days ago   |
| #59776 | BUG: assert_frame_equal triggers Deprecat...  | Bug, Testing                                | about 6 days ago   |
| #59770 | BUG: replace with 3 or more `None` values...  | Bug, replace, Copy / view semantics         | about 3 days ago   |
| #59761 | QST: How to add docstring to min, max, an...  | Docs, Timedelta                             | about 3 days ago   |
| #59752 | BUG: Large dataset exception on arrow str...  | Bug, Needs Info, Arrow                      | about 2 days ago   |
| #59740 | BUG: Inconsistent behavior of `min()` whe...  | Bug, Closing Candidate                      | about 12 days ago  |
| #59726 | BUG: idxmax and idxmin with skipna=False ...  | Docs                                        | about 12 days ago  |
| #59710 | API/BUG: obj.str.slice(start=None:-1)         | Bug, Strings, Arrow                         | about 8 days ago   |
| #59704 | BUG: Excel corruption when reading twice ...  | Bug, IO Excel                               | about 14 days ago  |
| #59695 | DOC: methods in see also section in the p...  | Docs, Groupby, good first issue             | about 14 days ago  |
| #59691 | ENH:                                          | Enhancement, Needs Triage                   | about 15 days ago  |
| #59670 | DOC: Document that DataFrame.from_records...  | Docs, IO Data, good first issue             | about 12 days ago  |
| #59668 | BUG: Pandas incompatibility with Numpy 2.1.0  | Bug, Needs Triage                           | about 19 days ago  |
| #59667 | BUG: `df.groupby().aggregate()` doesn't p...  | Bug, Groupby, Apply, Closing Candidate, ... | about 16 days ago  |
| #59662 | DOC: Remove unnecessary `:func:` & `:meth...  | Docs, Needs Triage                          | about 19 days ago  |
| #59645 | ENH: read_md() and to_md()                    | Enhancement, Needs Triage                   | about 20 days ago  |
| #59641 | ENH: add from_edges to IntervalIndex          | Enhancement, Needs Triage                   | about 20 days ago  |
| #59640 | typos                                         |                                             | about 18 days ago  |
| #59639 | ENH: df.corr().top_correlated_features(N)...  | Enhancement, Needs Triage                   | about 21 days ago  |
| #59637 | DOC: fix docstring validation errors for ...  | Docs, Needs Triage                          | about 21 days ago  |
| #59635 | ENH: pd.DataFrame.groupby().apply: parallel   | Enhancement, Needs Triage                   | about 21 days ago  |
| #59613 | QST: CVE-2024-42992 response                  |                                             | about 23 days ago  |
| #59612 | 2.3 release dates                             |                                             | about 8 days ago   |
| #59606 | DOC: fix docstring validation errors for ...  | Docs, Needs Triage                          | about 22 days ago  |



C:\Users\Bruno>gh issue list -R pandas-dev/pandas --state open

Showing 30 of 3506 open issues in pandas-dev/pandas



| ID     | TITLE                                         | LABELS                                      | UPDATED            |
|--------|-----------------------------------------------|---------------------------------------------|--------------------|
| #59833 | BUG: Timestamp.tz and DatetimeIndex.tz ar...  | Timezones, Compat, Dependencies, Closing... | about 5 hours ago  |
| #59832 | BUG: `None` values are not processed when...  | Bug, Needs Triage                           | about 11 hours ago |
| #59831 | ENH: Restore the functionality of `.fillna`   | Enhancement, Missing-data, Dtype Convers... | about 5 hours ago  |
| #59827 | ENH: Add parameter to read_html() that di...  | Enhancement, IO HTML, Needs Discussion, ... | about 7 hours ago  |
| #59826 | BUG: is_unique fails when using float128 ...  | Bug, Needs Discussion, Index                | about 4 hours ago  |
| #59823 | BUG: Empty column name in group dataframe...  | Bug, Groupby, Apply                         | about 22 hours ago |
| #59814 | ENH: Add kwargs to `Series.map`               | Enhancement, Apply, API - Consistency       | about 22 hours ago |
| #59806 | DOC: Separate Examples for String Methods...  | Docs, Strings                               | about 3 days ago   |
| #59789 | DOC: Intro to data structures                 | Docs, Series, Constructors                  | about 4 days ago   |
| #59788 | DOC: Series.update throws a `FutureWarnin...  | Docs, Needs Info, Copy / view semantics     | about 22 hours ago |
| #59783 | DOC: Questions on Usage of `*args, **kwarg... | Docs, Accessors                             | about 4 days ago   |
| #59780 | Consider moving pyarrow's pandas compatib...  | Needs Discussion, Arrow                     | about 7 days ago   |
| #59777 | ENH: Add a clear option to interpret stri...  | Enhancement, IO Excel, Strings, Needs Info  | about 6 days ago   |
| #59775 | BUG: pd.concat() cannot concatenate more ...  | Bug, Reshaping, Period, Needs Discussion    | about 4 days ago   |
| #59772 | BUG: tz-aware series with NaT raises exce...  | Bug, Needs Triage                           | about 8 days ago   |
| #59771 | ENH: ArrowTemporalProperties' object has ...  | Enhancement, Needs Triage                   | about 8 days ago   |
| #59769 | BUG: to_datetime raises "AttributeError: ...  | Bug, Datetime, Needs Info                   | about 2 days ago   |
| #59765 | API: handling of missing values in Index...   | Missing-data, Index, API - Consistency      | about 8 days ago   |
| #59764 | BUG/inconsistency: pd.Series(dtype=object...  | Bug, Reduction Operations, Closing Candi... | about 8 days ago   |
| #59763 | BUILD: pandas on conda main repo installs...  | Docs, Upstream issue                        | about 8 days ago   |
| #59739 | BUG: FutureWarning for Boolean sparse dty...  | Bug, Sparse, Warnings                       | about 11 days ago  |
| #59738 | BUG: fixed_size_list pyarrow type not int...  | Bug, Needs Discussion, IO Parquet, Arrow    | about 3 days ago   |
| #59737 | ENH: DataFrame.from_dict doesn't work wit...  | Enhancement, IO Data, Needs Discussion, ... | about 11 days ago  |
| #59736 | BUG: pd.eval with engine="numexpr" fails ...  | Bug, expressions                            | about 3 days ago   |
| #59732 | BUG: df.loc astype type change on specifi...  | Bug, Strings, Needs Discussion              | about 8 days ago   |
| #59725 | ENH: Should groupby's `value_counts` sort...  | Enhancement, Groupby, Algos, Needs Discu... | about 11 days ago  |
| #59721 | BUG: `pd.read_html(...): Unable to handl...   | Bug, IO HTML, Closing Candidate             | about 9 days ago   |
| #59717 | BUG: DataFrame.from_records()'s columns a...  | Bug, IO Data                                | about 3 days ago   |
| #59716 | QST: DataFrame.plot() vs Series.plot() ha...  | Visualization, Usage Question, Needs Dis... | about 9 days ago   |
| #59715 | QST: Consistently apply Welford Method an...  | Performance, API Design, Needs Discussio... | about 21 hours ago |



C:\Users\Bruno>
```

Pandas conta com 23311 Issues fechados e 3506 issues abertos

```
C:\Users\Bruno>gh issue list -R pola-rs/polars --state open

Showing 30 of 1943 open issues in pola-rs/polars



| ID     | TITLE                                         | LABELS                                      | UPDATED            |
|--------|-----------------------------------------------|---------------------------------------------|--------------------|
| #18823 | Row Group Based Subtotals with .group_by()    | enhancement                                 | about 1 hour ago   |
| #18821 | Schema assumes the column order in the da...  | bug, python, needs triage                   | about 3 hours ago  |
| #18820 | Join fails for scanned lazyframes when `s...  | bug, python, needs triage                   | about 3 hours ago  |
| #18819 | write_parquet encoding no longer recogniz...  | bug, python, needs triage                   | about 7 hours ago  |
| #18818 | Install slack app to allow subscription       | enhancement                                 | about 9 hours ago  |
| #18814 | Categorical revmaps are not merged when c...  |                                             | about 11 hours ago |
| #18813 | `rows_by_key` works with pl.Array             | enhancement                                 | about 11 hours ago |
| #18812 | `read_csv` PanicException when `pl.Decima...  | bug, python, needs triage                   | about 11 hours ago |
| #18810 | loading pickled `pl.Series` of `dtype=pl...   | bug, python, needs triage                   | about 12 hours ago |
| #18808 | pl.datetime does not respect leftmost-arg...  | bug, python, rust, accepted                 | about 9 hours ago  |
| #18802 | Reading from S3 compatible storage            | enhancement                                 | about 13 hours ago |
| #18800 | from_any_values_and_dtype converts AnyVal...  | bug, rust, needs triage                     | about 17 hours ago |
| #18799 | Issue when header dtype differ from the r...  | bug, python, upstream issue                 | about 1 day ago    |
| #18798 | Parsing to `Decimal` ignores `decimal.com...  |                                             | about 1 day ago    |
| #18797 | Parse large floats in CSVs that have peri...  | enhancement, A-io-csv                       | about 1 day ago    |
| #18795 | list.set_intersection operates on the wro...  | bug, python, needs triage                   | about 1 day ago    |
| #18791 | docs: to_datetime: document that %Y zero-...  | documentation, python, A-timeseries         | about 1 day ago    |
| #18788 | `reshape` in group by context PanicExcepti... | bug, python, needs triage                   | about 1 day ago    |
| #18787 | `struct.field` + `filter` PanicExceptio...    | bug, python, needs triage                   | about 1 day ago    |
| #18786 | SQL query with WHERE clause that evaluate...  | bug, python, needs triage                   | about 1 day ago    |
| #18785 | Use of `sorting_columns` parquet metadata     | bug, python, needs triage                   | about 1 day ago    |
| #18784 | Support serializing to json in a format s...  | enhancement                                 | about 1 day ago    |
| #18783 | Explain that the trig returns and argumen...  | documentation                               | about 1 day ago    |
| #18782 | `write_excel(formulas={})` appears non-fu...  | bug, python, needs triage                   | about 1 day ago    |
| #18781 | `str.json_decode` hangs on nested JSON st...  | bug, python, needs triage                   | about 2 days ago   |
| #18769 | Add feature gates to `polars-stream           | enhancement, accepted, P-low                | about 2 days ago   |
| #18768 | Parquet metadata is dropped on DSL->IR co...  | bug, rust, accepted, internal, P-medium,... | about 2 days ago   |
| #18765 | ".extend()" and ".vstack()" raises "Panic...  | bug, python, needs triage                   | about 2 days ago   |
| #18760 | read_json cannot parse a simple json array    | bug, python, needs triage                   | about 3 days ago   |
| #18758 | `AWS_ENDPOINT_URL` not inferred by cloud I/O  | bug, python, needs triage                   | about 3 days ago   |



C:\Users\Bruno>
```

```
C:\Users\Bruno>gh issue list -R pola-rs/polars --state closed

Showing 30 of 6719 issues in pola-rs/polars that match your search



| ID     | TITLE                                             | LABELS                                 | UPDATED            |
|--------|---------------------------------------------------|----------------------------------------|--------------------|
| #18794 | `arr.to_struct` PanicException for non-Array ...  | bug, python, needs triage, A-panic     | about 12 hours ago |
| #18793 | Python - Add support to the invert operator `~... | enhancement                            | about 1 day ago    |
| #18779 | Panic: 'collect(streaming=True)' on 'scan_parq... | bug, python, needs triage              | about 1 day ago    |
| #18767 | Strange type validation on 'when_then' express... | bug, python, needs triage              | about 2 days ago   |
| #18754 | 'cum_max' error message for invalid datatype r... | bug, python, P-low, A-exceptions       | about 1 day ago    |
| #18747 | polars.testing.assert_frame_equal raises Asser... | bug, python, needs triage              | about 3 days ago   |
| #18743 | Meet and Beat Pandas' Support for Nested DataF... | enhancement                            | about 3 days ago   |
| #18741 | Option to use fixed point floats                  | enhancement                            | about 21 hours ago |
| #18739 | Parquet read with `streaming=True` shifts/drop... | bug, python, P-critical, A-io-parquet  | about 1 day ago    |
| #18736 | Cannot add a Series to an empty DataFrame only... | bug, python, accepted, needs triage    | about 2 days ago   |
| #18726 | JSONL parsing fails when line is just `{}`        | bug, rust, needs triage                | about 5 days ago   |
| #18719 | Broadcasting issue                                | bug, python, needs triage              | about 6 days ago   |
| #18710 | PDS Q19 failing since Polars 1.7.0                | bug, python, accepted, needs triage    | about 2 days ago   |
| #18709 | InvalidOperationError on trying to select arra... | bug, python, needs triage              | about 6 days ago   |
| #18705 | Slow sink parquet with default streaming chunk... | bug, python, needs triage              | about 6 days ago   |
| #18704 | Support `to_list()`/`to_array()` on `struct` t... | enhancement                            | about 6 days ago   |
| #18700 | Cannot add columns to an empty DataFrame in 1.7.0 | bug, invalid                           | about 5 days ago   |
| #18698 | use of undeclared crate or module `futures` wi... | bug, rust, needs triage                | about 6 days ago   |
| #18697 | Build fails with serde feature                    | bug, rust, accepted, needs repro       | about 2 days ago   |
| #18686 | [1.7.0 Regression] Lit with a list creates a s... | bug, python, accepted, needs triage    | about 2 days ago   |
| #18673 | Panic with "expected arrays of the same length"   | bug, python, needs repro               | about 4 hours ago  |
| #18661 | Performance regression of `n_unique` in `group... | bug, python, performance, needs triage | about 7 days ago   |
| #18660 | CI test failure `test_delta_encoding_roundtrip`   | bug, python, needs triage              | about 8 days ago   |
| #18651 | `DataFrame.write_parquet()` expects an argumen... | bug, python, needs triage              | about 7 days ago   |
| #18642 | Add an argument to display progress bars to `m... | enhancement                            | about 7 days ago   |
| #18641 | What does the trailing `_` mean in `Expr.not_...  | enhancement                            | about 8 days ago   |
| #18634 | Change `join_where` to not allow ambiguous nam... | enhancement                            | about 8 days ago   |
| #18623 | Change methods' type hinting to return "Self" ... | enhancement                            | about 5 days ago   |
| #18613 | `pl.date_range` doesn't support `date` subclas... | bug, python, needs triage              | about 10 days ago  |
| #18604 | `join_where()` should error when `_right` doe...  | bug, python, accepted, needs triage    | about 9 days ago   |



C:\Users\Bruno>
```

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.

```
C:\Users\Bruno>gh pr list -R pandas-dev/pandas --state closed
Showing 30 of 32879 pull requests in pandas-dev/pandas that match your search
```

ID	TITLE	BRANCH	CREATED AT
#59837	Backport PR #59836 on branch 2.2.x (BLD: ...	meeseeksmachine:auto-backport-of-pr-5983...	about 2 hours ago
#59836	BLD: Fix bad Cython annotation	lithomas1:fix-bad-annotation	about 4 hours ago
#59834	DOC: fix SA01,ES01 for pandas.tseries.off...	tuhinsharma121:pandas.tseries.offsets.We...	about 8 hours ago
#59830	DOC: Adding ArcticDB to the ecosystem.md ...	ms041223:adding_arcticdb_to_ecosystem	about 16 hours ago
#59828	Backport PR #59813 on branch 2.2.x (CI: D...	meeseeksmachine:auto-backport-of-pr-5981...	about 1 day ago
#59821	Backport PR #59818 on branch 2.2.x (BUG: ...	jorivandenbossche:auto-backport-of-pr-5...	about 2 days ago
#59820	DOC: updates to IO page adding ArcticDB	ms041223:docs_update_with_arcticdb	about 2 days ago
#59819	BLD/RLS: build wheels with released numpy...	jorivandenbossche:wheels-released-numpy	about 2 days ago
#59818	BUG: Remove np._get_promotion_state usage	lysnykolaou:fix-numpy-promotion-state	about 2 days ago
#59813	CI: Debug failing ARM builds	lithomas1:try-fix-arm-build	about 3 days ago
#59812	Backport PR #59811 on branch 2.2.x (DOC: ...	meeseeksmachine:auto-backport-of-pr-5981...	about 3 days ago
#59811	DOC: add whatsnew for v2.2.3	jorivandenbossche:doc-whatsnew-2.2.3	about 3 days ago
#59810	String dtype: allow string dtype in query...	jorivandenbossche:string-dtype-query-eval	about 3 days ago
#59808	WEB: update list of (in)active core devs	twoertwein:inactive	about 3 days ago
#59807	BUG (CoW): fix reference tracking in repl...	jorivandenbossche:gh-59770-replace-list...	about 3 days ago
#59805	DOC: fix SA01 for pandas.errors.MergeError	tuhinsharma121:pandas.errors.MergeError	about 3 days ago
#59804	DOC: fix SA01 for pandas.arrays.FloatingA...	tuhinsharma121:pandas.arrays.FloatingArray	about 3 days ago
#59803	DOC: fix SA01 for pandas.api.types.is_int...	tuhinsharma121:pandas.api.types.is_integ...	about 3 days ago
#59802	DOC: fix SA01 for pandas.api.types.is_arr...	tuhinsharma121:pandas.api.types.is_exten...	about 3 days ago
#59801	Remove xfail in test_uscols_basic	rhshadrach:2.3.0	about 4 days ago
#59800	DOC: fix SA01,ES01 for pandas.Timedelta.t...	tuhinsharma121:pandas.Timedelta.total_se...	about 4 days ago
#59799	DOC: fix SA01,ES01 for pandas.Timedelta.c...	tuhinsharma121:pandas.Timedelta.components	about 4 days ago
#59798	Update ExtensionArray.interpolate to remo...	ammar-qazi:update-interpolate	about 4 days ago
#59797	Enhance to numeric to support hexadecimal...	Staudinger0325:patch-1	about 4 days ago
#59796	DOC: fix SA01,ES01 for pandas.PeriodDtype...	tuhinsharma121:pandas.PeriodDtype.freq	about 4 days ago
#59795	Enhance to numeric to support hexadecimal...	Staudinger0325:main	about 4 days ago
#59794	2.2.x Enhance to numeric to support hexad...	Staudinger0325:main	about 4 days ago
#59792	Fix doc	fbourgey:fix-doc	about 5 days ago
#59787	DOC: fix SA01 for pandas.NA	tuhinsharma121:pandas.NA	about 5 days ago
#59785	Assorted backports for 2.2.x	lithomas1:auto-backport-of-pr-59065-on-2...	about 6 days ago

```
C:\Users\Bruno>
```

```
C:\Users\Bruno>gh pr list -R pandas-dev/pandas --state open

Showing 30 of 96 open pull requests in pandas-dev/pandas
```

ID	TITLE	BRANCH	CREATED AT
#59835	Backport PR #59136 on branch 2.2.x (Uploa...	jorismvandenbossche:2.2.3-backports	about 7 hours ago
#59822	DOC: Emphasize team managed pandas in ins...	mroeschke:docs/install	about 2 days ago
#59816	Bump pypa/cibuildwheel from 2.20.0 to 2.21.0	dependabot/github_actions/pypa/cibuildwhe...	about 2 days ago
#59815	BUG: read_json silently ignores the dtype...	jahn96:dtype	about 2 days ago
#59809	BUG: Fix from_records() column reorder is...	VibavariG:59717-from-records-columns-reorder	about 3 days ago
#59793	Fix doc	fbourgey:fix-doc-data-structures	about 5 days ago
#59791	Fix doc errors for various datetime funct...	Pranav-Wadhwa:dt-docs	about 5 days ago
#59786	ENH: Follow up to PR #59577, Series.str.g...	aaronchucarroll:stringmethods-get-dummies	about 6 days ago
#59774	REF: pass dtype explicitly to _from_sequence	jbrockmendel:ref-dtype-explicit	about 7 days ago
#59773	REF: pass dtype explicitly to _from_seque...	jbrockmendel:ref-dtype-explicit-array	about 8 days ago
#59767	Infer extensiondtype from scalar	andrewgsavage:infer_Ext	about 9 days ago
#59766	Fix issue #47101	tigerhawkvok:patch-2	about 9 days ago
#59754	#59009: Added document and a test case fo...	wooseogchoi:issue_#59009	about 9 days ago
#59751	DEPR: Update groupby.apply DeprecationWar...	rhshadrach:groupby_apply_future_warning	about 10 days ago
#59750	BUG: Fix groupby().any() behavior for tim...	40gilad:groupbyAnyBug	about 10 days ago
#59745	API: value_counts to consistently maintai...	rhshadrach:groupby_value_counts_sort	about 11 days ago
#59744	Doc string for pandas.Timestamp.max, pand...	NavaneethaSundararaj:attribute_test	about 11 days ago
#59743	BUG: Remove 'read_json' datetime deprecat...	KevsterAmp:remove-read_json-datetime-depr...	about 11 days ago
#59741	DOC: Fix SA01 for pandas.Series.dt.nanose...	chalky25:documentation_changes	about 11 days ago
#59722	Fix part of docstring timedelta #59698	ammar-qazi:doc_update	about 13 days ago
#59720	BUG: CSS strings truncated at ":"	invalidarg:style-render-css-to-tuple-patch	about 13 days ago
#59719	Fix docstring Timedelta.to_timedelta64 SA...	fbourgey:fix-docstring	about 13 days ago
#59714	fixed issue#59670. DOC	StaticAccess:issue#59670	about 13 days ago
#59700	DOC: fix broken link for governance in th...	doshi-kevin:changes1	about 15 days ago
#59696	ENH: add basic DataFrame.from_arrow class...	jorismvandenbossche:arrow-capsule-import	about 15 days ago
#59688	Fix docstring timestamps (Issue #59458)	mimistiv:timestamp-fix-docstring-validation	about 16 days ago
#59686	TST (string dtype): resolve all infer_str...	jorismvandenbossche:string-dtype-tests-arrays	about 16 days ago
#59685	String dtype: map builtin str alias to St...	jorismvandenbossche:string-dtype-astype-str	about 16 days ago
#59655	DOCS: fix docstring validation error for ...	githubalexliu:fix_docstring_2	about 19 days ago
#59654	expose 'to_pandas_kwargs' in 'read_parque...	kleinhenz:expose_to_pandas_kwargs	about 20 days ago

```
C:\Users\Bruno>
```

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


```
C:\Users\Bruno>gh pr list -R pola-rs/polars --state open

Showing 30 of 127 open pull requests in pola-rs/polars
```

ID	TITLE	BRANCH	CREATED AT
#18822	refactor(python): Make `NodeTraverser` st...	nodetraverser-pub	about 3 hours ago
#18817	refactor(rust): Fix zero-length len	fix-zero-len-len	about 10 hours ago
#18816	fix: Throw error for comparison of unequa...	coastalwhite:fixes/panic-shape-mismatch-cmp	about 10 hours ago
#18815	fix: Proper dtype casting for struct embe...	coastalwhite:fixes/vstack-struct-categor...	about 11 hours ago
#18803	feat: Relaxed schema alignment for parque...	nameexhaustion:pq-col-unaligned	about 16 hours ago
#18801	fix(rust): Handle AnyValue::Struct to pre...	i64:main	about 17 hours ago
#18778	fix: Struct filter by index	baraki412:fix_filter_struct_index	about 2 days ago
#18775	refactor: Keep scalar in more places	coastalwhite:refactor/column-scalar-opts	about 2 days ago
#18773	docs: Refactor `docs` directory hierarchy	rodrigogiraoserrao:split-docs-hierarchy	about 2 days ago
#18761	fix: Improve histogram bin logic	mcrumiller:hist-bins	about 2 days ago
#18759	docs(python): Clarify documentation for `...	bradfordlynych:read_csv-docs	about 3 days ago
#18756	docs(python): Fix literal type mapping ex...	eitsupi:fix-lit-doc	about 3 days ago
#18746	refactor(rust): Replace `DynArgs` with an...	3ok:impl-serde-rolling-options	about 4 days ago
#18696	docs(rust): Add Rust example for toleranc...	arnabanimesh:patch-1	about 7 days ago
#18663	fix(rust): Return empty DF when input is ...	deanm0000:fix_empty_json_list	about 8 days ago
#18633	perf: Collapse cross-joins to faster joins	coastalwhite:perf-collapse-joins	about 9 days ago
#18519	docs: Add example of unpivoting multiple ...	mdavis-xyz:18513-doc-unpivot-multi	about 16 days ago
#18504	feat: Enable parse of `0`/`1` as bool in csv	mcrumiller:csv-bool-0-1	about 17 days ago
#18491	docs(rust): Align dates in DataFrame exam...	xhilogga:docs/fix-date-consistency-in-dat...	about 18 days ago
#18448	fix(python): Change method op type hints	ts826848:op-method-equivalent-type-hints	about 21 days ago
#18428	refactor(python,rust): Rename `delimiter`...	colin-kerkhof:unify-separator-arg-names	about 21 days ago
#18408	docs: Remove ecosystem viz section since ...	joelostblom:rm-dup-viz-docs	about 22 days ago
#18396	feat: JSON encoding in the root expressio...	tim-stephenson:universal_json_encode	about 22 days ago
#18355	feat: Extrapolate flat parameter in inter...	agossard:feature/interpolate_by_enhancem...	about 25 days ago
#18313	fix: Draw seed on DSL creation	shuffle	about 27 days ago
#18253	fix(python): Add option to drop empty row...	Rashik-raj:add-option-to-drop-empty-row-...	about 1 month ago
#18216	fix: Categorical `unpivot`	rikhuijzer:rh/cat-unpivot	about 1 month ago
#18215	feat(python): Support "indexed" dicts as ...	alexander-beedie:indexed-dict-input	about 1 month ago
#18199	fix(python): Handle boolean comparisons i...	dimfeld:iceberg-filter-boolean	about 1 month ago
#18196	feat: Allow whitespace with `to_integer`	rikhuijzer:rh/to_integer-whitespace	about 1 month ago

```
C:\Users\Bruno>
```

```
C:\Users\Bruno>gh pr list -R pola-rs/polars --state closed

Showing 30 of 9960 pull requests in pola-rs/polars that match your search
```

ID	TITLE	BRANCH	CREATED AT
#18811	fix: Fixed some error/assertion types	orlp:fix-errors	about 12 hours ago
#18809	refactor(rust): Add missing implicit datetime...	orlp:fix-datetime-alias	about 14 hours ago
#18807	refactor: Add panic to unchecked DataFrame co...	nameexhaustion:df-new-unchecked	about 14 hours ago
#18806	refactor(rust): Fix topological sort in new s...	orlp:new-stream-toposort-fix	about 15 hours ago
#18805	refactor(rust): Fix new-streaming parquet `te...	nameexhaustion:pq-ns-18463	about 15 hours ago
#18804	fix: Remove panic in `arr.to_struct`	coastalwhite:fixes/arr-to-struct-panic	about 16 hours ago
#18796	docs(python): Improve `over` docs, add exampl...	MarcoGorelli:over-docs	about 1 day ago
#18792	refactor: Remove extra hashmap construction i...	nameexhaustion:pqs-mcheck	about 1 day ago
#18790	refactor(rust): Move schema check function fr...	nameexhaustion:move-schema-check-func	about 1 day ago
#18789	refactor: Fix parquet file metadata is droppe...	nameexhaustion:dsl-ir-metadata	about 1 day ago
#18780	fix: Fix `cum_max` using exception text of `c...	t-ded:fix_cum_max_exception_msg	about 2 days ago
#18777	docs: Minor improvements to contributing guide	rodrigogiraoserrao:docs-contributing	about 2 days ago
#18776	chore(python): Remove TODO comment regarding ...	remove-todo	about 2 days ago
#18774	fix: Allow empty sort by columns	orlp:empty-sort-by	about 2 days ago
#18772	refactor(rust): Ensure fallback node gets cor...	orlp:correct-len-fallback	about 2 days ago
#18771	refactor(rust): Fix input independence tests ...	input-independence	about 2 days ago
#18770	fix: Broadcast zip_with for structs	coastalwhite:fixes/struct-zip-broadcast	about 2 days ago
#18766	fix: Dropped/shifted rows in parquet scan wit...	nameexhaustion:pq-rg	about 2 days ago
#18764	refactor(rust): Remove short-lived / non-CPU ...	nameexhaustion:pqs-spawn	about 2 days ago
#18763	refactor(rust): Fix new-streaming parquet on ...	nameexhaustion:pqs-bugfix	about 2 days ago
#18762	docs(python): Add documentation for beta gpu ...	beckernick:docs/gpu-support-beta	about 2 days ago
#18748	fix: Fix accidental raise on shape 1	shape	about 3 days ago
#18744	refactor(python): Remove unused methods	eitsupi:remove-unused-head-tail	about 4 days ago
#18737	ci: Run benchmark on PR labeled 'needs-bench'	dev	about 5 days ago
#18734	chore: Unset sources cache if sources set	set_sources	about 5 days ago
#18727	feat: Implement single inequality joins for j...	adamreeve:iejoin_single	about 5 days ago
#18725	refactor(rust): Remove unsafe in polars-json ...	Manishearth:deser	about 5 days ago
#18723	ci: Run remote benchmarks on dedicated machine.	dev	about 6 days ago
#18722	Python Polars 1.7.1 Rust Polars 0.43.1	dev	about 6 days ago
#18721	chore(python): Enable additional `ruff` lint ...	alexander-beedie:additional-lint	about 6 days ago

```
C:\Users\Bruno>
```

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.