Report for Assignment 4

Marcus Jakobsson
Toto Roomi
Rémi Grasset
Kerem Robin Yurt
Anton Bölenius

March 2024

1 Project

Name: Pandas
URL: https://github.com/pandas-dev/pandas
Pandas is a large, widely used graphing and calculation tool in python. Often used in Jupyter notebooks.

2 Onboarding Experience

We chose to move on from Jabref to Pandas. The documentation for Pandas was very detailed and instructive. Some tests were failing, however the main branch was also failing tests on the 4 latests commits. Jabrefs documentation was slightly easier to navigate than Pandas though, this is due to the sheer scale of Pandas.

3 Effort Spent

For each team member, how much time was spent in:

1. plenary discussions/meetings;
2. discussions within parts of the group;
3. reading documentation;
4. configuration and setup;
5. analyzing code/output;
6. writing documentation;

7. writing code;

8. running code?

For setting up tools and libraries (step 4), enumerate all dependencies you took care of and where you spent your time, if that time exceeds 30 minutes.

4 Overview of Issue(s) and Work Done

Title: BUG: to_json does not preserve nanosecond precision of DatetimeIndex with time zone information
URL: bug report , our fix
Summary in one or two sentences. Scope (functionality and code affected).

It was difficult finding a bug as the vast majority we found of reported bugs were either not bugs at all, solved or pending review from repo admins. In the end we found the linked bug, it was a problem where nanosecond precision was lost somewhere in the execution. After exhaustive investigation we finally traced the problem down to a the file "pandas/_libs/src/datetime/pd_datetime.c" where the function convert_pydatetime_to_datetimestruct() did not handle nanoseconds even though the rest of the code had made space for it. In our solution we added a handler for nanoseconds.

5 Requirements for the New Feature or Requirements Affected by Functionality Being Refactored

Optional (point 3): trace tests to requirements.

6 Code Changes

6.1 Patch

we added the following line to the afformationed file on line 74:
"out − > ps = 1000 * PyLong_AsLong(PyObject_GetAttrString(obj, "nanosecond"));"
We also created a test that makes sure nanoseconds work.

```
   git diff ...
git diff ddc3144 682ba46
```

Optional (point 4): the patch is clean. Optional (point 5): considered for acceptance (passes all automated checks).

7 Test Results

Overall results with link to a copy or excerpt of the logs (before/after refactoring).

8 UML Class Diagram and Its Description

8.1 Key Changes/Classes Affected

Optional (point 1): Architectural overview.
Optional (point 2): relation to design patterns.

9 Overall Experience

What are your main take-aways from this project?
Our main take-away is that contributing to a large project doesn't have to be a scary and overwhelming task but rather, much like university assignments, just a small problem you have to solve using a general accepted convention. The community is open to discuss and give feedback on the pull requests, which is a great opportunity to contribute and also evolve.

What did you learn?
We've learned how to create tests, document and execute them in a large python project using "pytest". Additionally we've learned the convention used in solving issues on github. For instance by writing "take" on issues you'd like to get assigned to.

How did you grow as a team, using the Essence standard to evaluate yourself?

Optional (point 6): How would you put your work in context with best software engineering practice? Optional (point 7): Is there something special you want to mention here?

10 P+

Criteria 1 - Architecture ???

Criteria 3 - Tests ???

Criteria 4 - Clean Patch

- **a)** We have no obsolete code.
- **b)** We didn't add any debug code.
- **c)** We didn't add any unnecessary white spaces.

Criteria 5 - Patch Accepted

For now we are waiting on the PR to get accepted.

Criteria 7 - Extraordinary

In assignment 3 we had the opportunity to work on a open-source project where we implemented a coverage tool. This time we got the opportunity to contribute in the source code and help the community resolve some issues. Contributing to an open-source project is new for most of us and having an assignment that introduces to that is a good first step. We chose to work on pandas which is a Python library. The library is towards data analysis and closely related areas. Most of the group members has at least once used the library and was glad to contribute to a project they have used before.