

Check CSV Equivalence Using Polars

2025-11-20

Table of contents

Today I Learned added on 2025-11-20, learned on 2025-11-20; edited on 2026-06-14.

I found myself reviewing two PRs made by the same individual and each PR had a `csv` file containing forecasts. The `csvs` seemed identical, but I wasn't sure, so I wanted a way to check for equivalence using Python.

Given that, at present, I prefer `polars` over `pandas`, I searched `polars` for a method and found the method `equals` (see <https://docs.pola.rs/api/python/stable/reference/dataframe/api/polars.DataFrame.equals.html>) in the `DataFrame` class .

Example `csv` file as `example_01.csv`:

```
column_1,column_2
45,98
32,31
17,100
```

Example `csv` file as `example_02.csv`:

```
column_1,column_2
87,98
32,31
17,100
```

First, the non-Python, Unix command `diff` (the standard method I've employed for equivalence testing) can also be used.

```
diff example_01.csv example_02.csv

# output:
# 2c2
# < 45,98
# ---
# > 87,98
```

In Python, using the `equals` method of `polars`, which outputs `True` or `False`, instead of the actual differences if there are any:

```
import polars as pl

df_01 = pl.read_csv("example_01.csv")
df_02 = pl.read_csv("example_02.csv")
```

```
# output: False  
df_01.equals(df_02)
```