Validation with goodtables

Katerina Drakoulaki

katerina.drakoulaki@gmail.com - @KDrakoulaki

Frictionlessdata.io . github.com/frictionlessdata

What is Frictionless Data?

Data cleaning

What's the license for this data?

What does this data value mean?

Checking data quality

Who created this data?

Can I use this Excel data in Python?

Structure errors

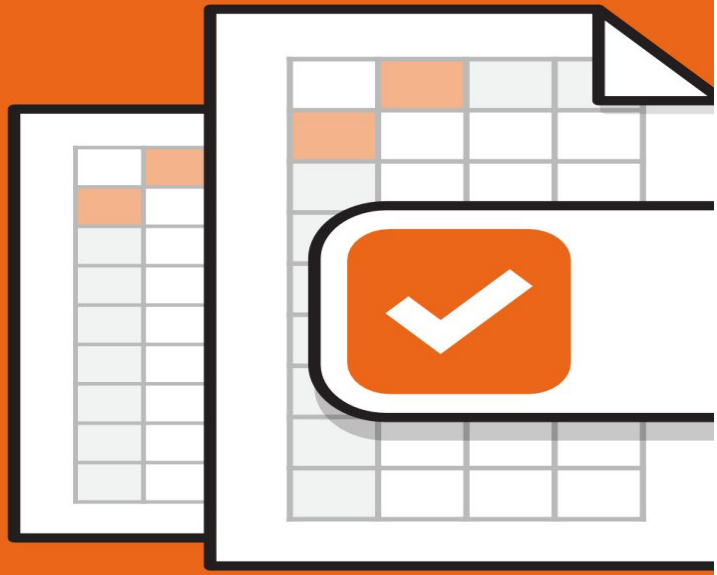
These checks validate that the structure of the file are valid.

check	description
blank-header	There is a blank header name. All cells in the header row must have a value.
duplicate-header	There are multiple columns with the same name. All column names must be unique.
blank-row	Rows must have at least one non-blank cell.
duplicate-row	Rows can't be duplicated.
extra-value	A row has more columns than the header.
missing-value	A row has less columns than the header.

FRICTIONLESS DATA

Field guide

Validating data with try.goodtables

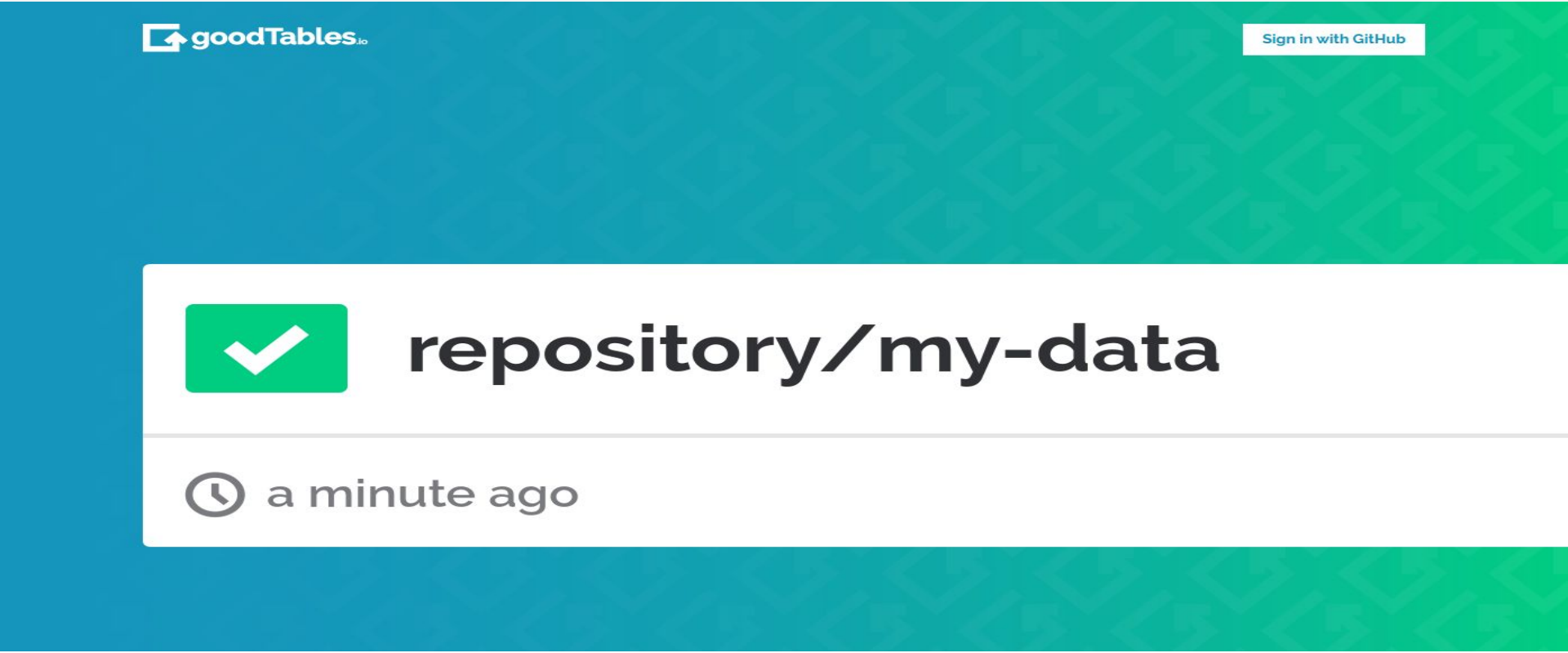


Try.goodtables.io

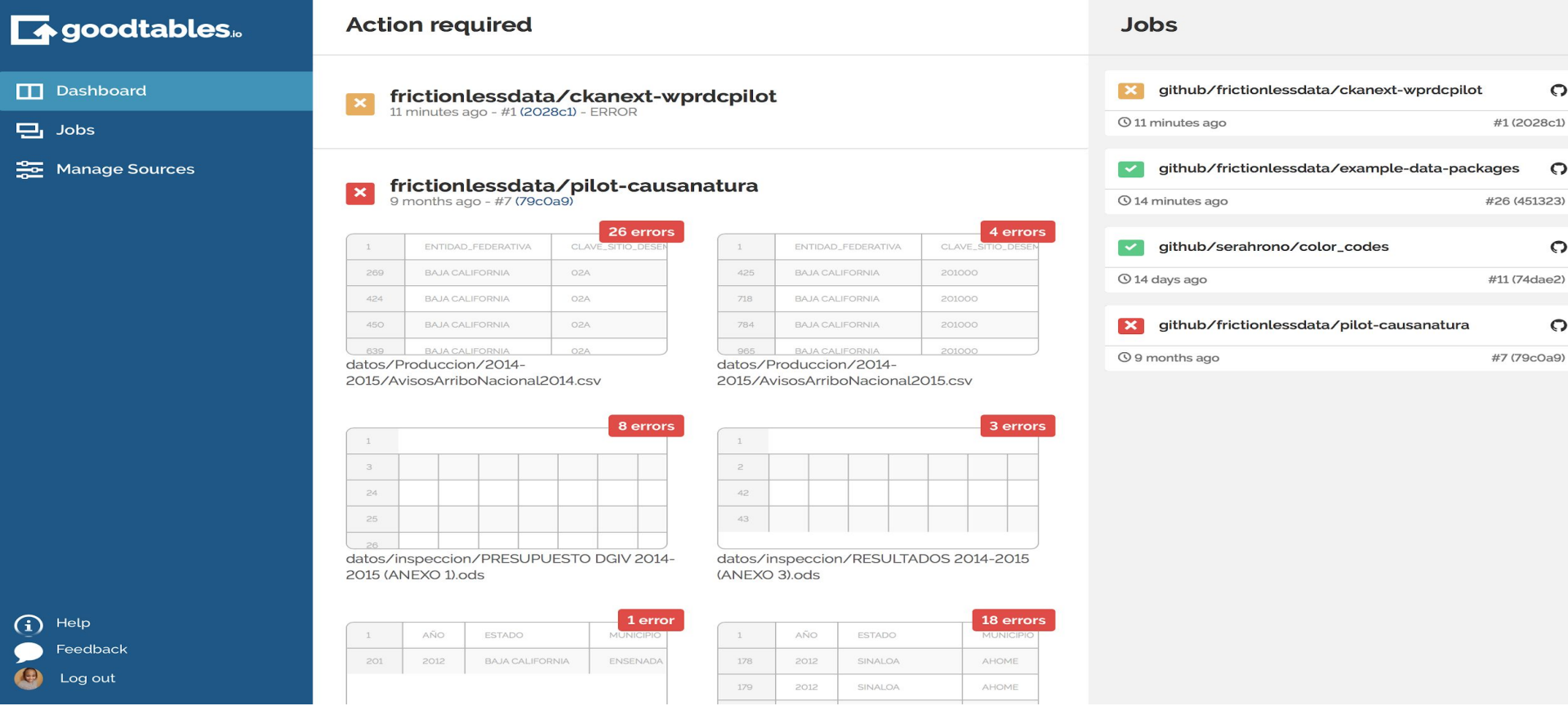
https://bit.ly/2P6oogW Or lots of example data:

https://github.com/frictionlessdata/goodtables-py/tree/master/data

Continuous data validation: http://goodtables.io



Frictionless Data
Fellows



http://goodtables.io/

Frictionless Data Fellows

Validate using the command line

1. more details on installation here
2. change directories to where your file is
3. check the validation >goodtables file_I_want_to_validate.csv

Thank you for your time!