

Validation with goodtables



Katerina Drakoulaki

katerina.drakoulaki@gmail.com - @KDrakoulaki

[Frictionlessdata.io](https://frictionlessdata.io) . github.com/frictionlessdata



Open Knowledge
Foundation

@okfn

<https://okfn.org>

What is Frictionless Data?

Data
cleaning

What's the license
for this data?

What does this
data value mean?

FRICTION



Checking
data
quality

Who created
this data?

Can I use this
Excel data in
Python?

Structure errors

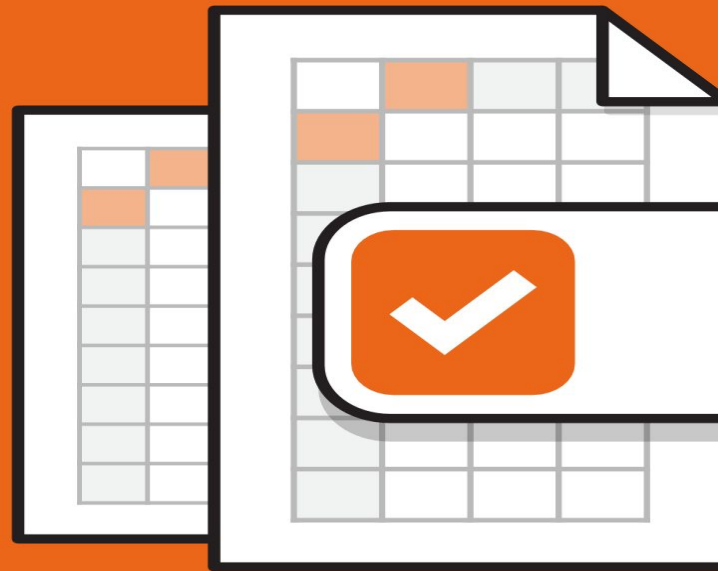
These checks validate that the structure of the file are valid.

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://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>



Sign in with GitHub



repository/my-data



a minute ago

Action required

frictionlessdata/ckanext-wprdcipilot
11 minutes ago - #1 (2028c1) - ERROR

frictionlessdata/pilot-causanatura
9 months ago - #7 (79c0a9)

26 errors

1	ENTIDAD_FEDERATIVA	CLAVE_SITIO_DESEN
269	BAJA CALIFORNIA	02A
424	BAJA CALIFORNIA	02A
450	BAJA CALIFORNIA	02A
639	BAJA CALIFORNIA	02A

datos/Produccion/2014-2015/AvisosArriboNacional2014.csv

4 errors

1	ENTIDAD_FEDERATIVA	CLAVE_SITIO_DESEN
425	BAJA CALIFORNIA	201000
718	BAJA CALIFORNIA	201000
784	BAJA CALIFORNIA	201000
965	BAJA CALIFORNIA	201000

datos/Produccion/2014-2015/AvisosArriboNacional2015.csv

8 errors

1							
3							
24							
25							
26							

datos/inspeccion/PRESUPUESTO DGIV 2014-2015 (ANEXO 1).ods

3 errors

1							
2							
42							
43							

datos/inspeccion/RESULTADOS 2014-2015 (ANEXO 3).ods

1 error

1	AÑO	ESTADO	MUNICIPIO
201	2012	BAJA CALIFORNIA	ENSENADA

18 errors

1	AÑO	ESTADO	MUNICIPIO
178	2012	SINALOA	AHOME
179	2012	SINALOA	AHOME

Jobs

github/frictionlessdata/ckanext-wprdcipilot
11 minutes ago #1 (2028c1)

github/frictionlessdata/example-data-packages
14 minutes ago #26 (45132c3)

github/serahrono/color_codes
14 days ago #11 (74dae2)

github/frictionlessdata/pilot-causanatura
9 months ago #7 (79c0a9)

http://goodtables.io/

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!