

PANDAS profiling

This function is one of Plugins Operation. You can find the movie in [ARGOS RPA+ video tutorial](#).

Need help?

Technical contact to tech@argos-labs.com

May you search all operations,

- [Actions](#)
- [Verifications](#)
- [System Calls](#)
- [Interactives](#)

 pandas p...	PANDAS profiling
	Author: Jerry Chae
	This plugin is based on the pandas profiling tool. It is a tool to analyze the input data and to give a fixed format HTML as output. Please learn more at
	<ul style="list-style-type: none">• https://pypi.org/project/pandas-profiling/• https://github.com/pandas-profiling/pandas-profiling

Input format
Following 6 files can be used as input.
<ul style="list-style-type: none">• .csv• .tsv• .xls• .xlsx• .xlm• Json
Output format
Only HTML file are supported as output.

How to set parameters (properties and advanced).

Plugin: pandas profiling

Operation name: Operation 1

Plugin version: 2.507.1215

Properties

In file: [] ...

Out file: [] ...

Title of the rep... []

HTML Style: {'style':{'full_width':True}}

Select Range, e... []

Advanced

Sheet Name: 0

Header: 0

Index Col: []

Use Cols: []

Data Type: []

CSV Sep: ,

Encoding: utf-8

Return value

Result type: String

Variable name: {{my.a}}

1: If you want to give Title to the report, put it here.

2: If you are only using specific range of data, put here.

1. Specify file-path of input and output files. If you want give Title to the HTML report, put the Title here. If you wan to analyze a certain range from the data file specify here.
2. When reading Excel file, specify sheet name.
 - Defaults to 0: 1st sheet as a DataFrame.
 - 1: 2nd sheet as a DataFrame.
 - "Sheet1": Load sheet with name "Sheet1".
 - [0, 1, "Sheet5"]: Load first, second and sheet named "Sheet5" as a dict of DataFrame.
 - None: All sheets.
3. Specify what row you have the headers: Row (0-indexed) to use for the column labels of the parsed DataFrame. If a list of integers is passed those row positions will be combined into a MultiIndex. Use None if there is no header.
4. Column (0-indexed) to use as the row labels of the DataFrame. Pass None if there is no such column. If a list is passed, those columns will be combined into a MultiIndex. If a subset of data is selected with usecols, index_col is based on the subset.
5. If None, then parse all columns.
 - If str, then indicates comma separated list of Excel column letters and column ranges (e.g. "A:E" or "A,C,E:F"). Ranges are inclusive of both sides.
 - If list of int, then indicates list of column numbers to be parsed.
 - If list of string, then indicates list of column names to be parsed.
6. Data type for data or columns.

E.g. {'a': np.float64, 'b': np.int32} Use object to preserve data as stored in Excel and not interpret dtype. If converters are specified, they will be applied INSTEAD of dtype conversion.
7. Delimiter to use. If sep is None, the C engine cannot automatically detect the separator, but the Python parsing engine can, meaning the latter will be used and automatically detect the separator by Python's builtin sniffer tool, csv.Sniffer. In addition, separators longer than 1 character and different from '\s+' will be interpreted as regular expressions and will also force the use of the Python parsing engine. Note that regex delimiters are prone to ignoring quoted data. Regex example: '\r\t'.
8. Return Value stores the complete file-path of "Out File".

Output Examples

1)

UN Population Data Overview Variables Interactions Correlations Missing values Sample Duplicate rows

Overview

Overview Reproduction Warnings 4

Dataset statistics		Variable types	
Number of variables	9	CAT	6
Number of observations	64350	NUM	3
Missing cells	39625		
Missing cells (%)	6.8%		
Duplicate rows	30		
Duplicate rows (%)	< 0.1%		
Total size in memory	16.5 MiB		
Average record size in memory	268.1 B		

2)

Variables

Country or Area
Categorical
CARDINALITY HIGH

Distinct count	240
Unique (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	251.4 KiB

Toggle details

Year
Real number (R₃₂)

Distinct count	72	Mean	1991.060202020202
Unique (%)	0.1%	Minimum	1948
Missing	0	Maximum	2019
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	502.8 KiB

Toggle details

Area
Categorical

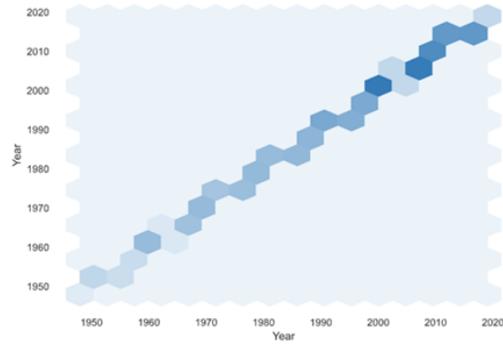
Distinct count	4
Unique (%)	< 0.1%
Missing	0
Missing (%)	0.0%

3)

Interactions

Year Source Year Value

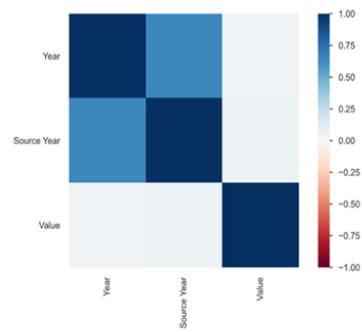
Year Source Year Value



4)

Correlations

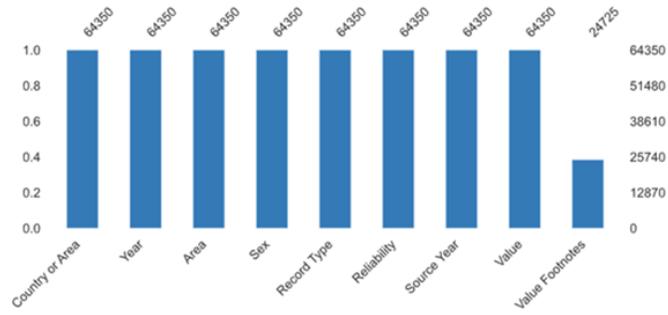
Pearson's r Spearman's ρ Kendall's τ Phik (ϕ_k) Cramér's V (ϕ_c)



5)

Missing values

Count Matrix Dendrogram



6)

Sample

First rows

Country or Area	Year	Area	Sex	Record Type	Reliability	Source Year	Value	Value Footnotes
0 Afghanistan	2019	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	30725560.0	1
1 Afghanistan	2019	Total	Male	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	15643298.0	1
2 Afghanistan	2019	Total	Female	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	15082272.0	1
3 Afghanistan	2019	Urban	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	7761765.0	1
4 Afghanistan	2019	Urban	Male	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	3921208.0	1
5 Afghanistan	2019	Urban	Female	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	3840557.0	1
6 Afghanistan	2019	Rural	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	22963795.0	1
7 Afghanistan	2019	Rural	Male	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	11722080.0	1
8 Afghanistan	2019	Rural	Female	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	11241715.0	1
9 Afghanistan	2018	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	2019	30075018.0	1

Last rows

Country or Area	Year	Area	Sex	Record Type	Reliability	Source Year	Value	Value Footnotes
64340 Zimbabwe	1959	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	1959	3730000.0	NAN
64341 Zimbabwe	1958	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	1959	3610000.0	NAN
64342 Zimbabwe	1957	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	1959	3500000.0	NAN
64343 Zimbabwe	1956	Total	Both Sexes	Estimate - de facto	Final figure, incomplete/questionable reliability	1959	3380000.0	NAN

All Plugins

- Arithmetic Op
- File Conv
- Excel Advanced
- Excel Macro
- File Folder Op
- File Monitor
- JSON Select
- REST API
- SQL
- SSH Command
- String Manipulation
- Web Extract
- Email Read Mon
- Time Stamp
- Work Calendar
- Google Translate
- Google Cloud Vision API
- Google TTS
- Fairy Devices mimi AI
- LINE Notify
- IBM Speech to Text
- Email IMAP ReadMon
- Drag and Drop
- Rossum
- MS Azure Text Analytics
- IBM Visual Recognition
- Convert CharSet
- Detect CharSet
- Bot Collabo
- NAVER OCR
- Excel AdvII
- PANDAS I
- AWS Textra Rekog
- XML Extract
- Simple SFDC
- PANDAS profiling
- MS Word Extract
- AD LDAP
- SSH Copy
- Password Generate
- Newuser-SFDC
- PDF2Doc
- Tesseract
- Google Drive
- Google Sheets
- Google Token