pandas II

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 II

Author: Jerry Chae

This is the second release in our pandas plugin series. It will give you more flexibility and more agility as to the pre-process sequence of your data. Data-Scientists can try their cleansing and filtering in tools like Jupyter Notebook and apply it the flow immediately to this plugin for super quick and easy distribution and deployment in the field.

Primary Features

This plugin runs python statement(s) on pandas with one (1) dataframe.

- https://en.wikipedia.org/wiki/Pandas_(software)
- https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

Prerequisite

This plugin requires Python and Regular Expression skills.

⚠ **Please note** that the pandas solution is a large software using numerous Python machine learning sub-modules. The bot will take more than just a few minutes to download them to be ready. But this is just for the "first run". As to the second run on, the local VENV will be used to avoid downloading unless new pandas II version has been selected to replace what was in the bot originally.

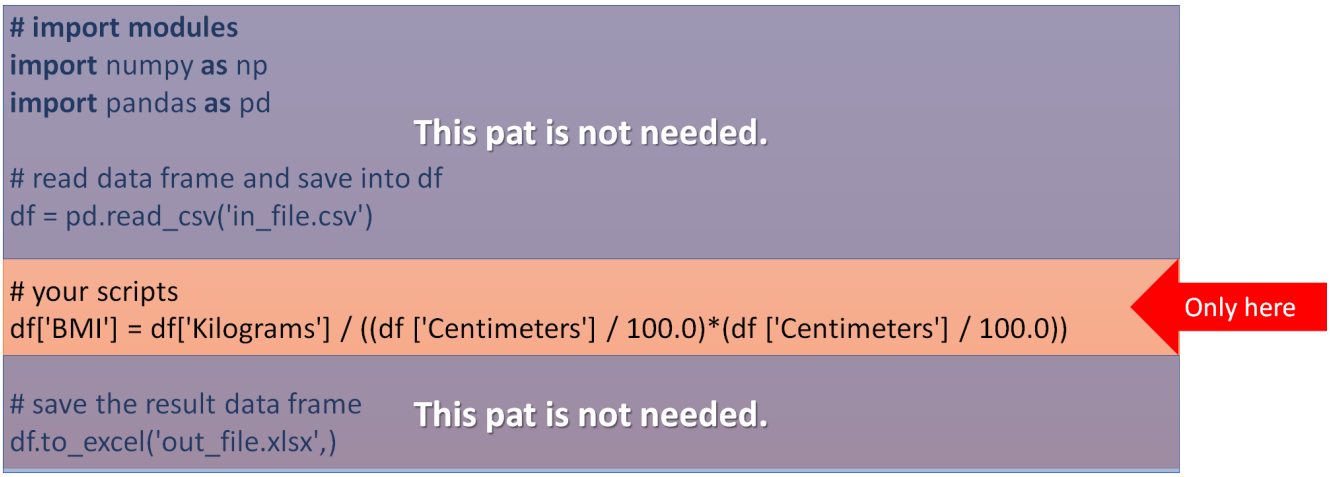
And, **YES** --- if no pre-processing is performed, this plugin can act as **a file format converter**. Again, you will face a long first run download time because of the reason described above.

Update 2021.03.11

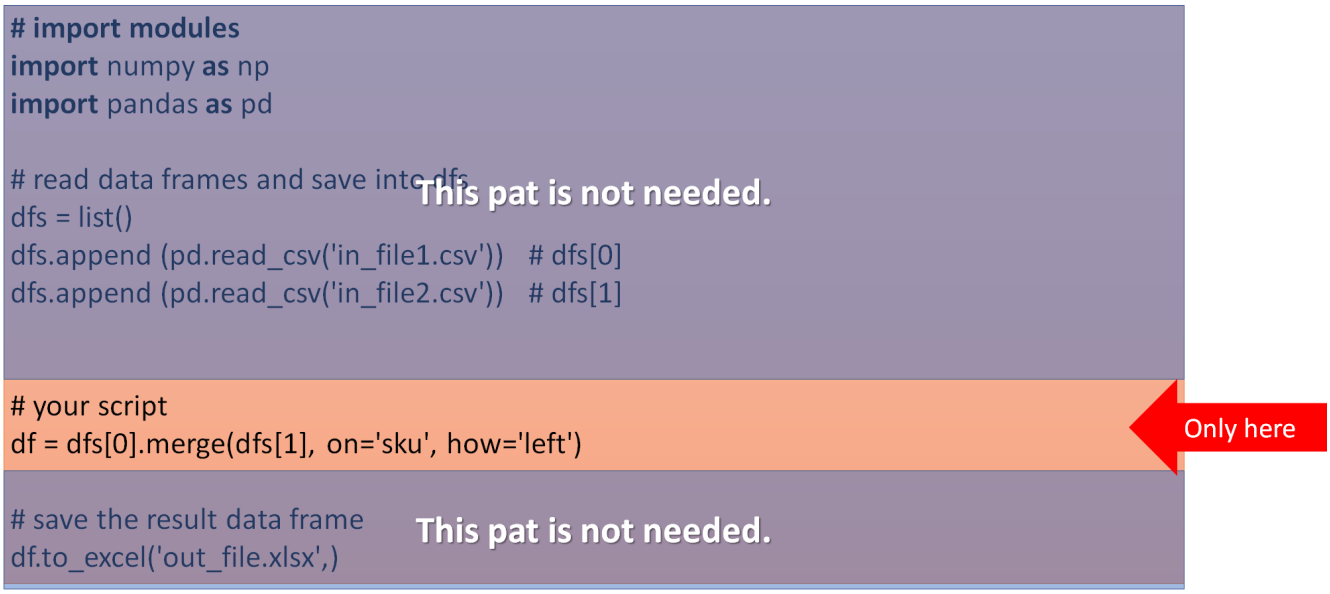
You only need the BODY part of your pandas statements to drive the pandas-II and -III plugins.

The pandas-II and -III plugins have integrated the **importing, reading, and the saving parts**, you only need the body part of your statements. For example, when your pandas statements look like below you only need one line in the pandas-II and -III plugins.

For pandas-II



For pandas-III



Update 2021.02.22

Sample Statements and Use of “df” and “dfs” variables

1. **You must use “df” and “dfs” as variables for data-frames**

   a. As variable for the dataframes with the Python statements in pandas II and III plugins, it is required to use "df" and "dfs" to represent dataframes (all in small cases).

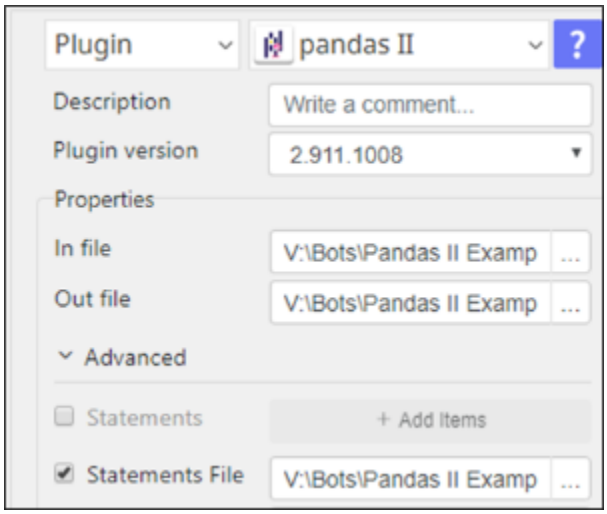
   b. As for pandas III, the multiple dataframes ("dfs“) will take [n] as index (it is zero based as the first set of dataframe becomes dfs[0]) as shown in examples below.

2. **For pandas II Statements**

a. The “In file” will be the data frame stored at "df" Python variable

b. All pandas functionality is working with "df" data frame including Reshaping at statements File

c. Processed results of statement’s execution will continue to be stored in the same "df“ variable and eventually be the “Out file”



3. pandas II Statements Example

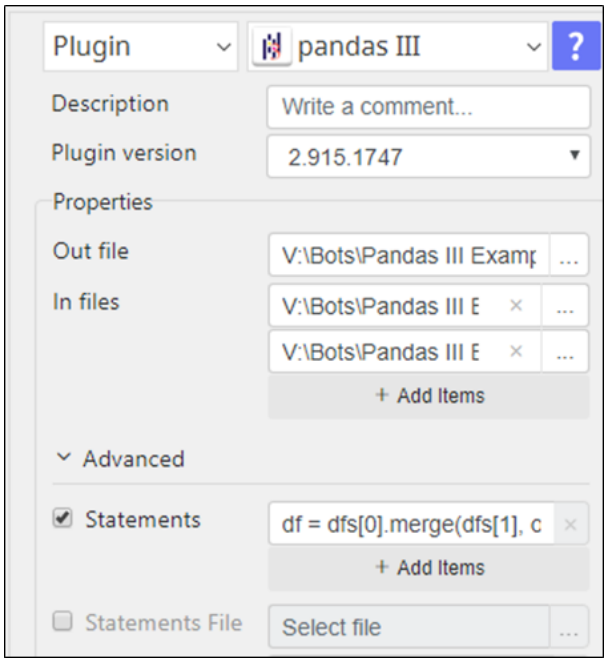
- df['BMI'] = df['Kilograms'] / ((df ['Centimeters'] / 100.0)*(df ['Centimeters'] / 100.0))
- df = df.sort_values('BMI', ascending=False)
- df = df.sort_values('BMI', ascending=False).groupby('Gender').head(5)

4. pandas III Statements

a. “In files” will be a data frame stored at "dfs[0]", "dfs[1]",... Python variable (zero base index)

b. All pandas functionality is working with "dfs[n]" data frames including merge

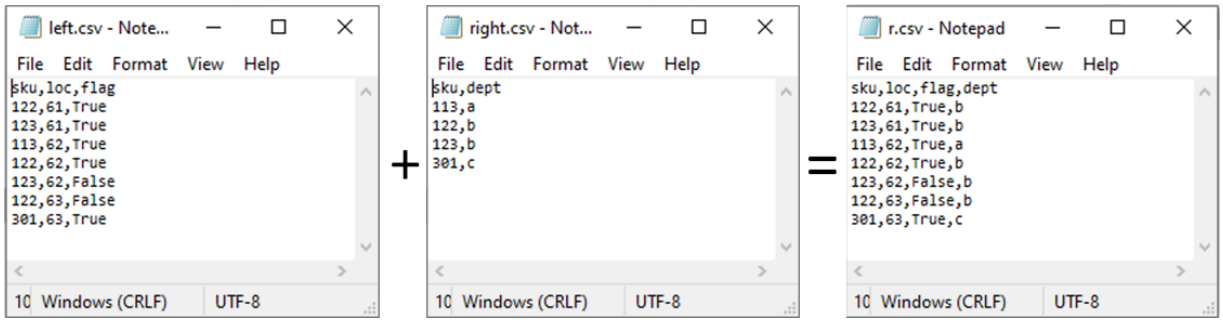
c. Processed results of statement’s execution will continue to be stored in the same "df“ variable and eventually be the “Out file”



5. pandas III Statements Example

- df = dfs[0].merge(dfs[1], on='sku', how='left')

Above Python represents the process illustrated below (just like vlookup feature in Excel)



Input, Output, Features, and Parameters.

Required Input

1. Input File: One data file (dataframe).

   Supported input formats are **.xlsm, .xls, xlsm, .csv, .tsv, and .json**

2. Output File: One data file.
   Supported input formats are .**xlsm, .xls, xlsm, .csv, .tsv, and .json**

Optional Input

3. Enter a **Python statement**, or multiple statements. Also a text file that contains a list of statements can be used as input.
4. When input file multiple **sheets,** you can select which sheet to be processed.
5. You can designate which row you can use as **header (variable)** for your processing.
6. You can specify a column to be used as the **index** of the dataframe.
7. You can specify which column(s) to be or not to be processed.
8. You can determine specific pandas **datatypes** for your column.
9. You can determine what character to use to **separate your data** (default is comma).
10. You can specify **encoding** technology of the input file (default is UTF-8).
11. You can select to either **show or hide the index** column in your output file.

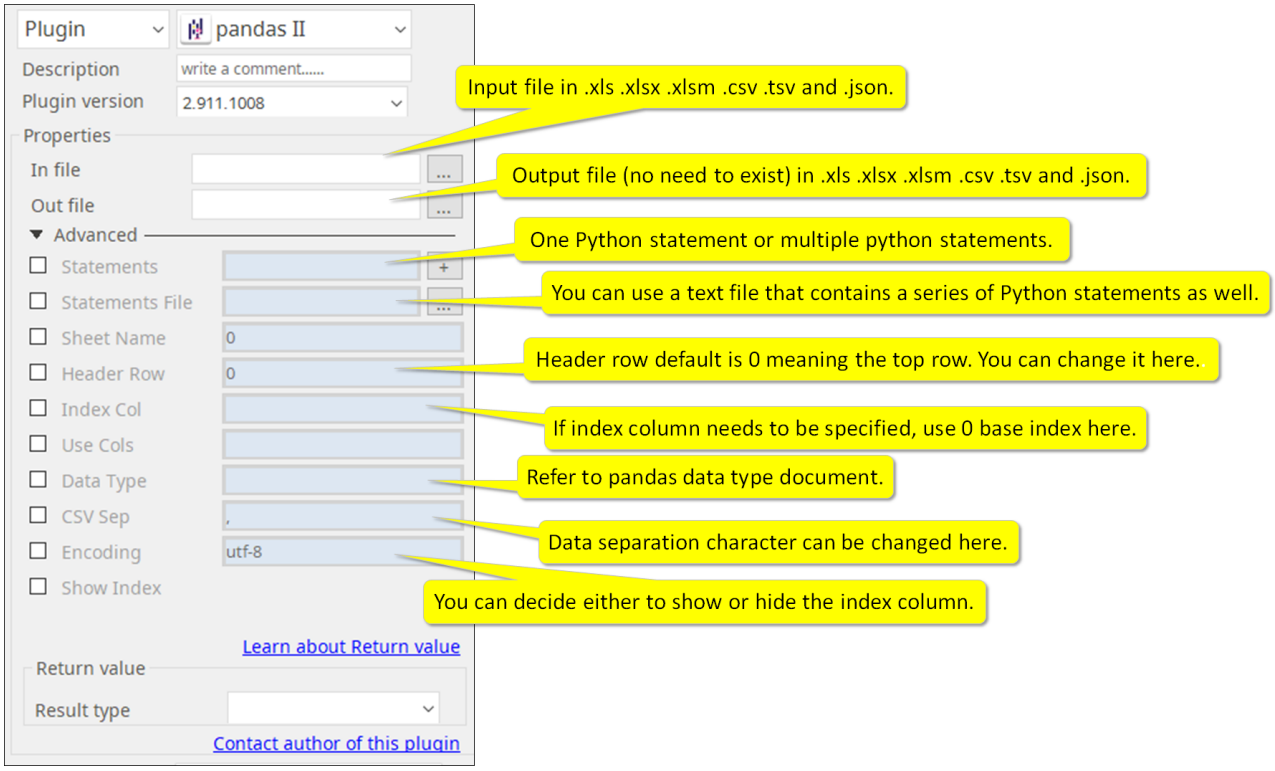
How to set parameters

pandas-II plugin parameters are 100% compatible to pandas read_excel specifications

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_excel.html

Please refer the parameters on the right in the pandas document above.

- Sheet Name sheet_name
- Header Row header
- Index Col index_cols
- Use Col usecols
- Data Type dtypes



Text from Image

All Plugins

- AD LDAP
- Adv Send Email
- Arithmetic Op
- Attach Image
- AWS S3
- AWS Textra Rekog
- Bot Collabo
- Chatwork GetMessage
- Chatwork Notification
- Clipboard
- Codat API
- Convert CharSet
- Convert Image
- Convert Image II
- Create Newfile
- CSV2XLSX
- Data Plot I
- Detect CharSet
- Dialog Calendar
- Dialog Error

- Dialog File Selection
- Dialog Forms
- Dialog Info
- Dialog Password
- Dialog Question
- Dialog Text Entry
- Dialog Text Info
- Dialog Warning
- Drag and Drop
- Dynamic Python
- Email IMAP ReadMon
- Email Read Mon
- Env Check
- Env Var
- Excel Advanced
- Excel AdvII
- Excel AdvIII
- Excel Copy Paste
- Excel Formula
- Excel Macro
- Excel Newfile
- Excel Simple Read
- Excel Simple Write
- Excel Update
- Fairy Devices mimi AI
- File Conv
- File Folder Exists
- File Folder Op
- File Status
- Folder Monitor
- Folder Structure
- Google Calendar
- Google Cloud Vision API
- Google Drive
- Google Search API
- Google Sheets
- Google Token
- Google Translate
- Google TTS
- Html Extract
- HTML Table
- IBM Speech to Text
- IBM Visual Recognition
- JSON Select
- JSON to from CSV
- LINE Notify
- MongoDB
- MS Azure Text Analytics
- MS Word Extract
- NAVER OCR
- Newuser-SFDC
- PANDAS I
- pandas II
- pandas III
- PANDAS profiling
- Parsehub
- Password Generate
- PDF2Doc
- PDF Miner
- PDF SplitMerge
- PowerShell
- Print 2 Image
- Python Selenium
- QR Generate
- QR Read
- Regression
- Rename File
- REST API
- Rossum
- Scrapy Basic
- Screen Recording START
- Screen Recording STOP

- Screen Snipping
- Simple SFDC
- Slack
- Speed Test
- SQL
- SSH Command
- SSH Copy
- String Manipulation
- Sys Info
- Telegram
- Tesseract
- Text Read
- Text Write
- Time Diff
- Time Stamp
- Web Extract
- Windows Op
- Work Calendar
- XML Extract
- Xtracta Get Doc
- Xtracta Tracking
- Xtracta Upload
- ZipUnzip