To EPG Tec Blog Website – last updated 4/11/18

For Accountants who deal with big data set dump from the IS department

To analyze huge data set with endless rows and columns from the data dump, how to use Python to extract the data and perform the meaningful analysis? How to visualize the financial analysis?

The example illustrated is in csv or xlsx format since most likely that IS department would only provide the output data in either format.

Topic:

* Use Python to Import dataset in CSV or Excel data (1st blog)
* To add and perform math average in Python for data validation (2nd blog)
* After Reading Excel in Python, how to get the clean data for IS report spec?
* Put data in SQL Lite format for data analysis
* use Visio for mockup presentation purpose for reporting purpose.
* How to mockup the report so IS can understand the report spec?

**(B****) To add and perform math average in Python for data validation**

Python codes to analysis Excel data, those commends are useful for smaller or larger data sets.

This can be very effective when Hugh dataset need to be verify between IT and financial analysis

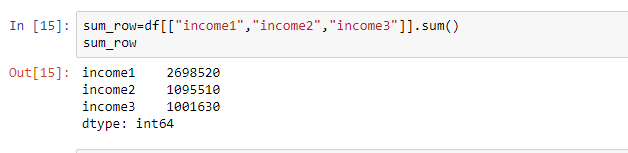
It’s a quick way to exam the accuracy of data

* Sum up columns or rows
* transpose columns
* get the maximum rows in Excel
* get the maximum, mean, average value in the data set
* filter data to isolate invalid data

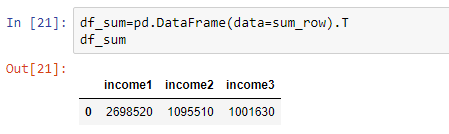
Dow Jone index and DJ Index 30 from 1/29/1985 to 10/25/2017

Use same stock file for adding total, so the blog flow better, easy to follow

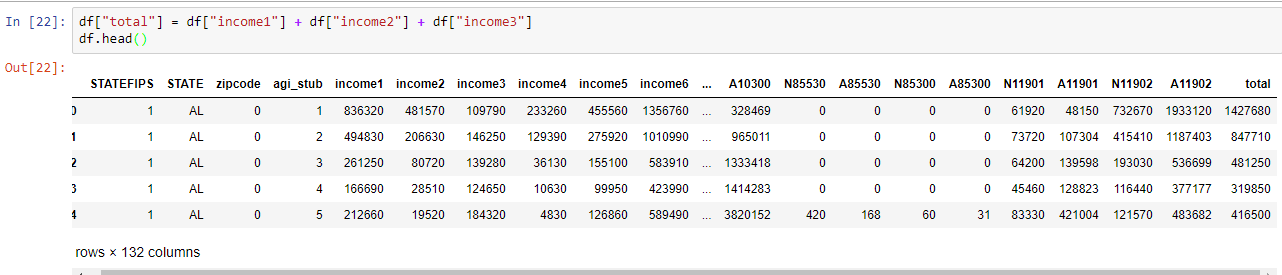
1. How to sum up columns income1, income2 and income3? Once we decide to sum up the total to income1, income2 and income3, we need to save it in Python. Its easy to create a total in Excel, for data integrity, we need to do one more step in Python to add the column “total”.



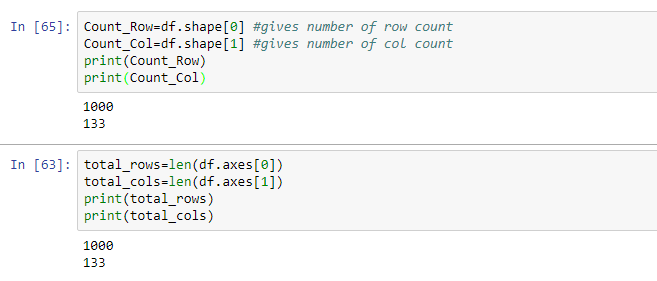
1. How to transpose columns? This will be very helpful to put in the report spec to identify which column of data should be use for report

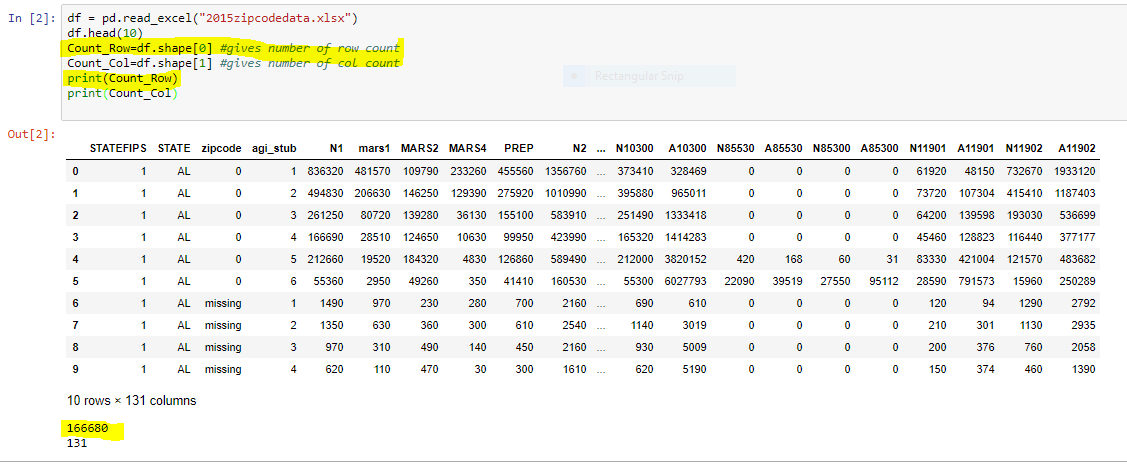


1. How to show the total in Python Panda dataframe?



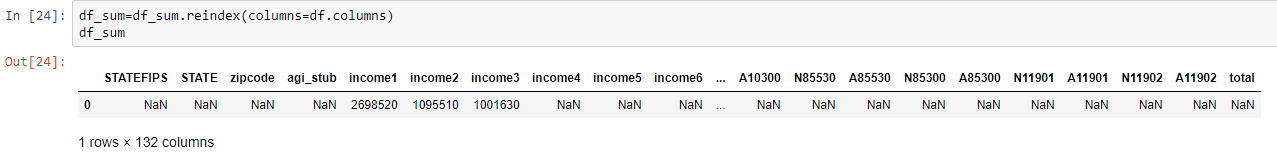
1. How to get the maximum rows in Excel? There are few different ways to get the total rows and columns. This is especially useful for hugh data set.



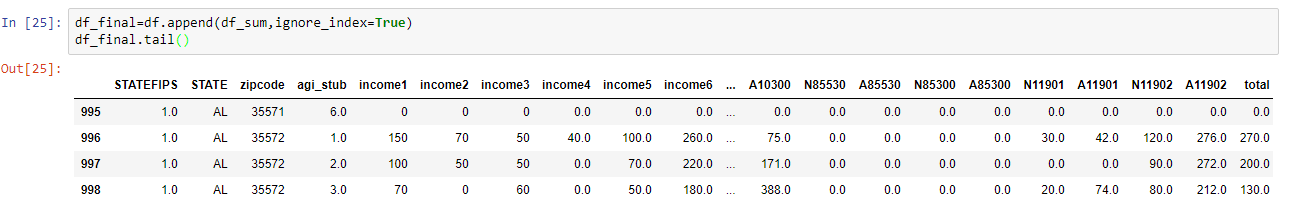


1. How to save the file in Python to Excel? The final thing we need to do before adding the totals back is to add the missing columns.

(1)We use reindex to do this for us. The trick is to add all of our columns and then allow pandas to fill in the values that are missing.



(2) The format in DataFrame is okay now, so we can add it to our existing Excel by using append



(3) save the file to excel



Then open the sample.xlsx in Excel. The “total” created in Python is now showing up in Excel file.

