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?

**(C) book ch 7“After Reading Excel in Python, how to get the clean data for IS report spec?**

* pandas.merge connects rows in DataFrames based on one or more keys
* pandas.concat glues or stacks together objects along an axis.
* combine\_first instance method enables splicing together overlapping data to fill in missing values in one object with values from another

1. Joint two dataframe using Key - Database-style DataFrame Merges
2. How to spot the invalid data?
3. Remove duplicate data
4. If you know the average value or the maximum value of certain data set. You can do it very easily in Python.

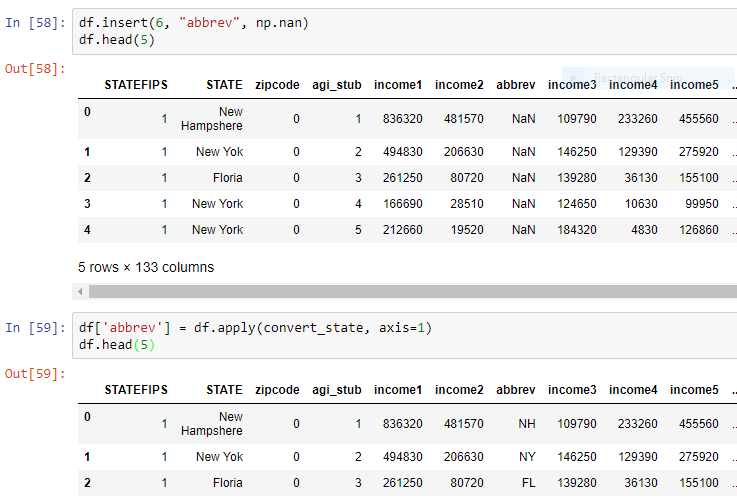
This will help to spot an unusual value to validate the data.

1. Instead of using Vlookup or Sumif in Excel, how to get a clean and only valid data for data analysis?

The data has a column for state, but some with incorrect spelling. In Excel, you can use vlookup to search for invalid data and sumif to add up data.

In Python, there’s an easier way to do it. Google and you can import fuzzywuzzy library and follow the code to covert the incorrect state to correct state.





**(D) How to mockup the report so IS can understand the report spec?**

Once you have the idea of how the data layout, you can prepare a report spec to IS department.

You can avoid lots of pitfall and back and forth with the IS department, because you know the data set might have invalid data.

Project: use the sample (census data) to identify the top income by zip code in the New York Area with family deduction of 3 or more. The project is to identify for local government funding.

1. Us the correct data for insurance and reinsurance pricing
2. Once the data is valid and contain the correct information. Then can prepare IS Spec report.
3. Add a trigger to the loss/policy for potential reinsurance? – add value
4. For example, you have a loss report with policy number, policy effective date, loss date…etc.
5. How to make this data useful?
6. How to mock up the report as a base projection model for FS projection?

