**Deferred Revenue Data Checking**

The deferred revenue billings data comes from two sources, the base billings file and the type A billings that have no config. These two tableau reports are copied to the ‘base\_billings’ workbook and imported into python.

We need a dataframe to be created that will track the progress of how we adjust the data as we create our deferred revenue forecast such that we will know the magnitude of any data we are excluding and check to make sure that some process is not dropping data.This dataframe will be exported to excel and reviewed to make sure there were no problems with the processing of this data.

The ‘data\_transformation\_checks.xlsx’ spreadsheet contains what I expect the output from this dataframe to look like. (In the docs folder on github).

For our first pass at this, the dataframe will include the USD equivalent billings for each document currency. The totals will be shown every time we make an assumption about the data that will change the totals and broken out by every category we use to split the data. The function will be called to add totals after each of these steps.

* Dropping currencies that have less than 20 billing periods.
* Split out of Sales Type (Non-Revenue gets dropped here and the Recognized and Professional Services revenue do not make it into the deferred category).
* Split out of deferred by type
* Type D billings by type (checking the totals match)
* Type A billings – dropping odd product configs
* Type A billings – being added by the second data set
  + Checking that the two amount (what was dropped and what was added) match

We need one function that will create the columns to the spreadsheet. This function will have the following inputs

* df\_data\_check: This is the final dataframe that gets sent to excel in the ‘data\_transformation\_checks’ spreadsheet.
* df\_other: The df\_other will be a different dataframe every time the function is called based on how we want to split out the data.
* transform\_type: This will be a string that will be passed to determine what adjustements get made in this function

The function will return the df\_data\_check dataframe with additional columns attached.

The df\_data\_check dataframe will be initialized before it gets called the first time, so we will be able to pass this variable into the function. (At this point it will include only the Document Currency (as an index) and a column showing the initial USD equivalent from the base billing file.

Next you will need to work on sending this thing to Excel

(harder than it sounds if we want to put it into an existing file versus starting a new file.)

Both of us need to understand this.

Options for sending data to excel

* Pandas ExcelWriter functionality

I am not sure if this will allow us to keep column headers that exist in an existing document and spreadsheet, but I need you to determine how that works and if it is possible.

Example syntax:

With pd.ExcelWriter(“output.xlsx”) as writer:

df\_data\_check.to\_excel(writer, sheet\_name = “USD\_check”)

* The module xlwt

<https://pypi.org/project/xlwt/>

* The module xlutils

<https://xlutils.readthedocs.io/en/latest/>

If we can do this with the pandas version (ExcelWriter) then I think we should use that one. If it is not possible to keep header rows and formatting in an existing spreadsheet and simply copy a dataframe into the sheet, then we should use one of the other modules (xlutils looks like it may contain the ability to modify an excel function versus create one or simply add data to a new sheet.)