**COMPENSATION DATADIARY**

**Combine “OFFICERS” sheet and “Main” sheet**

1. Inserted one column called StandNAME and ran function:

= Proper(D2)

1. There is no FYE data in “Officers” sheet so inserted another column “FILE DATE”

Ran function: =VLOOKUP(C2,'[20160115 GUIDESTAR YEAR-END COMPENSATION DATA FILE -- RAW.xlsx]Main'!$C:$M,11,0)

Inserted one more column “FYE”

Ran function: =year(F2)

3)Inserted one more column “TOTALREVENUE”

Ran function:

=VLOOKUP(C55,'[20160115 GUIDESTAR YEAR-END COMPENSATION DATA FILE -- RAW.xlsx]Main'!$C:$K,9,0)

1. Inserted all year employee

**INSERT TWO COLUMNS “TITLE2” & “DOC”**

1. Title clean

RUN PYTHON FILE “clean.py” in your terminal: run the command line-“python clean.py”

How did I tell who are “CEO, CFO, COO, CMO, CNO AND CIO”s :

Lowercased all titles, then see if the title includes “ceo or chief executive officer”.

The script goes below:

title\_input = row["title"].lower()

doc\_input = row["standname"].lower()

if any(c in title\_input for c in ("ceo","chief executive officer")):

current\_title = "CEO"

elif any(c in title\_input for c in ("cfo","chief financial officer")):

current\_title = "CFO"

elif any(c in title\_input for c in ("coo","chief operating officer")):

current\_title = "COO"

elif any(c in title\_input for c in ("cmo","chief marketing officer")):

current\_title = "CMO"

elif any(c in title\_input for c in ("cno","chief nursing officer")):

current\_title = "CNO"

elif any(c in title\_input for c in ("cio","chief information officer")):

current\_title = "CIO"

else:

current\_title = "OTHER"

1. Doc?

How did I tell if someone is a doctor?

First I lowercased all the names and see if the name starts with (‘dr ‘, ‘dra.’ Or ‘dr.’)

Or it endswith (("md","md phd","do","d.o."))

Or title includes ("physician","surgeon","surgery","dermatology","orthopedic","gastronenter","cardiologist","md","anesthesiologist","radiology"))

Then it fills the column ‘doc’ with “Y”

The script goes below:

if doc\_input.startswith(('dr ','dra.','dr.')) or doc\_input.endswith(("md","md phd","do","d.o.")) or any (c in title\_input for c in ("physician","surgeon","surgery","dermatology","orthopedic","gastronenter","cardiologist","md","anesthesiologist","radiology")):

doc = "Y"

1. Deduplicate

Run python file “clean.py” uncomment output\_compare()

Comment #insert\_title()

Detailed script is in “clean.py”

How did I determine if these records are duplicates? First proper org\_name and name.

If the record in output.csv has the same fiscal year, org\_name and name as any record in input.csv, it won’t be written in final\_output.csv.