**SQL Queries**

1. vsql\_ppa\_competition\_report.sql: The main SQL script that drives the report. I run this script once per day using the python script discussed below. Please see the script itself for commentary. The general gist is to pull data at an ad call, impression, and click level from the log data – then transpose this data into having 1 row where each advertiser and price is displayed. This row is duplicated with each of the 5 potential players (including the publisher) occupying the first slot. I want to eventually turn this SQL script into a map reduce job
2. vsql\_ppa\_competiton\_report\_full\_weeks.sql: Pulls for each date what week it belongs to, and whether that week has a full 7 days worth of data. If it doesn’t my report filters non full 7 day weeks out .

**Python script**

1. Fill\_Competition6.py: Because vsql\_ppa\_competition\_report.sql takes so long to run. I wrote a python script to run it once for each day. The python script is broken out into 5 functions for each of the 5 parts of the sql query. The final function creates a new table with the data for that day. The reason it creates a new table and not an insert is because Inserts do not seem to be committing with python. This is something we need to investigate in the future. Once the script is run for each day I Insert each new daily table created into the final table using the excel file attached.

There’s a bunch of code in the Python script to define string variables for month day and year. These variables are simply used to keep track of where in the loop the python script is. I also use an error handler which tells the script to restart at the most recent point in the loop in case there is an error. I did this because occasionally the connection drops or times out, and I want the script to keep running without needing manual intervention.

**Excel File**

1. Load Competition Data.xls: A simple Excel file I built to generate the insert scripts for each of the daily tables generated by the python script

**Tableau Report**

Most of the report is pretty self explanatory and there are not any super complicated calculations if you understand the structure of the data behind it. The main drivers of the report of the calculated metrics starting with AGGR. I labeled them AGGR as they are all aggregate functions.