**Excel Maker – To create two outputs:**

* Excel Creative Data
* Excel Placement Data

**There are two datasets used for that:**

‘FY19Q2\_CEAnalytics\_DCM\_AzureDirectCreative\_10.8-QTD.csv’

‘FY19Q2\_CEAnalytics\_DCM\_AzureDirect\_10.8-QTD.csv’

**Note:** The dictionaries for mapping and data cleaning are at the Excels at the different folders for each step

The second dataset is exactly the same one as the first one, but it contains two extra columns: creative ID and Creative name.

Each placement id (variable at both datasets) can contain multiple creative ID, that is why the first dataset (creative) is larger than the first one.

**Excel Maker:**

1. **Folder ‘1. First Step - DCM Pulls’** in this folder there is the Excel CSV. It is just the data we are going to use for the project.
   1. At this step we will use that dataset: ‘FY19Q2\_CEAnalytics\_DCM\_AzureDirectCreative\_10.8-QTD.csv’
2. **Folder ‘2. Second Step - Cleaning DCM Pulls for Cost Correction’.** 
   1. We are going to use only the following excel: ‘FY19Q1\_DCM\_CREATIVE\_Cleaning\_Template-AzureDirect.xlsb’
      1. What is the process – that needs to be done in Python:
         1. (Step 1) Copy data from Folder 1 into the Excel at Folder 2 into the tab Data Dump – 1
            1. Make sure that the column names are the same, if any formulas, these are recreated in Python
         2. (Step 2) Pivot tab – pivot the data.
         3. (Step 3) The data is ready to be copied pasted into the next excel cleaning step – that of course in python it will be all at the same script.
         4. Now we have the master data ready
3. **Folder ‘2.1. Third Step - Mapping Landing Page Urls’.** In this folder there is an excel ‘LandingPageURL\_Azure\_Q2\_10.8-QTD.xlsx’. How it works, the excel tab ‘LandingPageURL\_Azure\_Q2\_10.8-QT’ is the data we are going to use (pulled from a datasource and pasted there).

\*\* to ignore tab ‘Mapping for placement’

The tab ‘Mapping for Creative’ contains a pivot using ‘LandingPageURL\_Azure\_Q2\_10.8-QT’ data (excel tab at same excel).

After pivoting the data using the variables Creative ID and Landing Page URL, there is a conditional formatting applied. That conditional formatting changes the column b into red color when the Creative ID appears more than once. Afterwards, the color red is hidden so that that the table does not contain any Creative ID that appears more than once.

\*\* here I want it differently, instead of hiding all creative ID that appears more than once – so if the creative ID appears more than once, now it does not appear at all; what needs to happen instead is that the code creates a table that does not contain any duplicated creative id. So if a creative id appears twice and has two different landing page url, it only returns the first page landing for the duplicated creative id – and one creative id.

Then the values of the pivot table are copied into the tab Creative 1:1 – we do not need that step in python. That data will be mapped into the Master data using Creative ID to return (or populate) landing page URL.

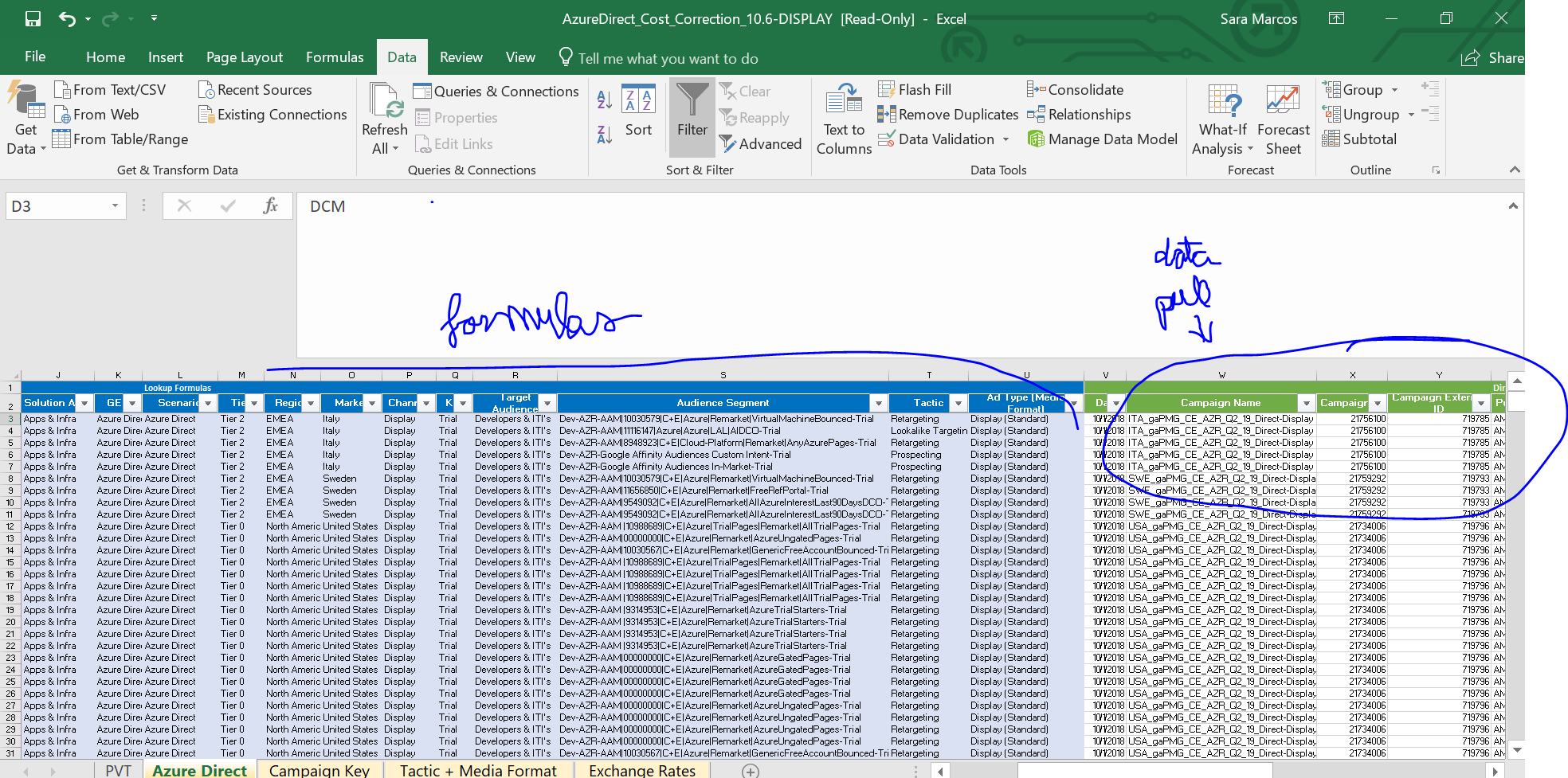
1. **Folder ‘3. Doing Cost Correction + QA’.** In this folder there are two Excels:
   1. **Creative**. Inside this folder there is 1 Excel.

\*\* here we create master\_Creative

* + 1. How it works: the output from before will go into the right (column starting at AA). And the cleaning happens at columns A – Z columns.
    2. All the dictionaries used to clean are at the Excel
  1. **Placement.** Inside this folder there are 2 Excels. There are two Excels to make the data manipulation easier – since the cleaning is done in Excel, they divide it in two Excels. But they are the same. The only difference is that for xx the ‘channel’ is display and for xx excel the ‘channel’ is social.

\*\* here we create master\_Placement

* + 1. How it works: the output from before will go into the right (column starting at Y). And the cleaning happens at A-U columns.
    2. Things to note: All the data from the master would go into the right but **without** the variables ‘creative name’ and creative ID’.



How we want now in Python: Display + Social in one file. The data that will populate the columns V-AO will be a subset of the master (without creative ID and creative name). Please double check that the names of the master match as the columns V-AO.

Columns from AP to AU are calculated functions.

1. **Folder ‘Master File’.**
   1. ‘Q2\_AzureDirect\_Creative\_MASTER\_Data.xlsb’
      1. Here the data from the previous step goes into that excel to clean further. Ignore columns from AS.
      2. They copy paste the data here and scroll down the formulas. All dictionaries are either created internally (like weeks in market) – there should be a way to calculate that column easier in python; and other dictionaries are all available. Other columns contain simple arithmetic functions.

\*\* data is ready for master\_creative\*\*

* 1. ‘Q2\_AzureDirect\_MASTER\_Data.xlsb’
     1. same as above. Here we do not need to ignore anything.

\*\* data is ready for master\_placement\*\*