# Checking Legal Entity Names Against d/b/a Names

**Summary:** We need a human to validate the results from the automated processes we used to update the list of tribal tobacco retailers.

**Files you need:**

full\_retailer\_list\_org (Excel) - [\\afsv03\Research\FDA\2732 - FDA Tribal Tobacco Retailers\Technical\Tribal\_Master\step\_3\_work\output\full\_retailer\_list\_orig.xlsx](file:///\\afsv03\Research\FDA\2732%20-%20FDA%20Tribal%20Tobacco%20Retailers\Technical\Tribal_Master\step_3_work\output\full_retailer_list_orig.xlsx)

3i\_manual\_list\_changes (Python) - [\\afsv03\Research\FDA\2732 - FDA Tribal Tobacco Retailers\Technical\Tribal\_Master](file:///\\afsv03\Research\FDA\2732%20-%20FDA%20Tribal%20Tobacco%20Retailers\Technical\Tribal_Master)



The “%load” command followed by the full path to the file will load the .py file into the first cell of your Jupyter Notebook

**Instructions:**

* Open full\_retailer\_list\_org
* Scroll through and visually compare the original tribal retailer information in columns C-J with the updates we’ve found in columns P-S

**Things to check for:**

* Changes to business Name: should not excessively deviate from original name
* Changes to business address: should not change great distances; comparison of longitude/latitude easiest way to check this probably.
* Changes to business status: punch business name/address into Google Maps and verify it can/cannot be found.

**If you find a mismatch:**

Don’t make any changes in the excel sheet for reproducibility reasons. We need to keep records of every change made to each retailer. Instead, hard code the changes via python. Once you load the 3i\_manual\_list\_changes.py into Jupyter notebook, write lines of code that change the values in the “\_updated” column names to “Can’t validate”.

**Example:**

IMPAQ\_ID 59, Cameron Conoco, looks like a mismatch. It updated to “Cameron Trading Post RV Park”, which has a latitude that’s ~0.5pts different than the original record. Luke said that the automated process performs poorly with chains like Conoco (a chain of gas stations).

Since it’s a mismatch, we need to write python code to fix the mistake. Each line below changes one cell in row 59.

df.loc[59, “name\_update”] = “Can’t validate”  
df.loc[59, “addr\_update”] = “Can’t validate”

df.loc[59, “lat\_update”] = “Can’t validate”  
df.loc[59, “lon\_update”] = “Can’t validate”  
df.loc[59, “biz\_status”] = “Can’t validate”