**Overview**

Main folder: [\\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 sum of the automation process results in a total of seven excel files, found in Tribal\_Master

* Output:

1\_Original\_Data: Unchanged original retailer data from FDA.

2\_Revised\_Data\_Change\_Log: Revised data, with columns to denote changes

3\_Revised\_Data\_Clean: Revised retailer data with no denoted changes

4\_Existing\_Sources\_Documentation: Links to documentation files for each existing retailer

5\_New\_Retailers: List of new retailers identified by Nielsen data

6\_Nielsen\_Supporting\_Data: Nielsen data for new retailers

7\_New\_Sources\_Documentation: Links to documentation files for each new retailer

Many things happen in the processing code, found in the home Tribal\_Master folder. The [Tribal readme](file:///C:\Users\lpatterson\AnacondaProjects\Tribal_Master\Tribal%20Readme.docx) describes the first 3 steps that occur prior to ownership web scraping (step 4). Step 5 is cleanup and documentation.

There are several aspects of the output we will need to QA/manually inspect, detailed below.

**Manual Ownership Retrieval**

As some states have only a few records/cannot be scraped, we will have to manually retrieve ownership information and documentation for some retailers. Each states’ business search website and number of records is recorded below. Each state with a “No” in the “build scraper” column must be manually retrieved.

<file:///\\afsv03\Research\FDA\2732%20-%20FDA%20Tribal%20Tobacco%20Retailers\Technical\Task%202%20-%20Retailers%20List\State%20business%20websites%20final%20v2.xlsx>

I have started the retrieval process; the states I have completed have a “Y” in the “manual scraping completed?” column. To preserve replicablility, all manual edits are made via Python in " Tribal\_Master\4a\_manual\_owner\_doc.py". As you complete states, fill this column out. The steps of manual retrieval are:

1. Pick a state and navigate to the state SoS website.
2. Open the retailer list and filter so only records from that state are showing (NOTE: use the values found in the columns ending in “\_update” when present):

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

1. Enter the DBA name of the retailer into the search query.
2. Scan results for a matching SoS record. Look for an entity with a similar name, and/or matching address
3. If you do not find a record:
   1. Play around with additional queries with simplified versions of the DBA Name
   2. start a new block of code in file 4a for that retailer and mark all fields as “Not found” and move on to the next record.
4. If you do find a record:
   1. Save a screenshot of the website record in the following folder/filename:
      1. “Tribal\_Master/output/documentation/New Retailer/IMPAQID\_####\_SoS\_Website\_Record.png”, if a new retailer
      2. “Tribal\_Master/output/documentation/Existing Retailer/[Insert REI value] \_SoS\_Website\_Record.png”, if an existing retailer
   2. Save any of the following documents that are freely available in the same folders, with the same filename structure except replacing “Website\_Record” with as a PDF file type:
      1. Articles of incorporation/formation/registration, the latest annual report, the latest agent change notice.
   3. Start a new block of code in file 4a for that retailer. “df” fields are the actual data. Input appropriate values from the website/latest documentation. Doc\_df are excel formula hyperlinks eventually written into the “Sources\_Documentation” excel files. Replace the file path and name with the appropriate value to reference the source document you used to determine the entity (est\_entity) and agent (est\_agent). If other documentation was present, list one of the other files in “est\_other”

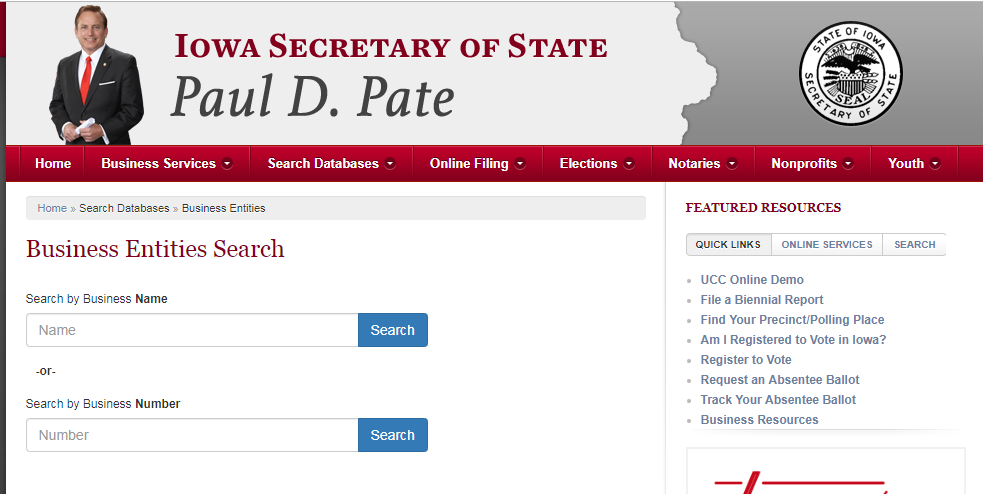
***Walkthrough***

This section walks you through how to do each of the steps for a sample retailer.

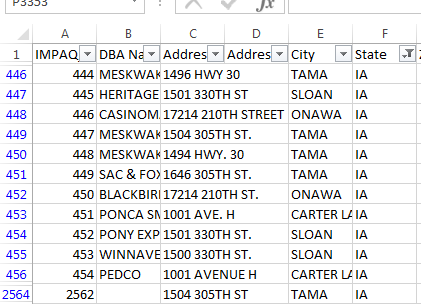
1. *Pick a state and navigate to the state SoS website.*

We’ll pick an Iowa Retailer for this.

<https://sos.iowa.gov/search/business/(S(qmg0ztbr3yolul21ntulm545))/search.aspx>



1. *Open the retailer list and filter so only records from that state are showing:*

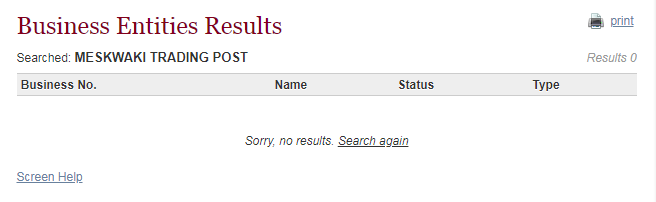


1. *Enter the DBA name of the retailer into the search query.*

We’ll pick IMPAQ ID 444: “MESKWAKI TRADING POST”

1. *Scan results for a matching SoS record. Look for an entity with a similar name, and/or matching address*

No results found.

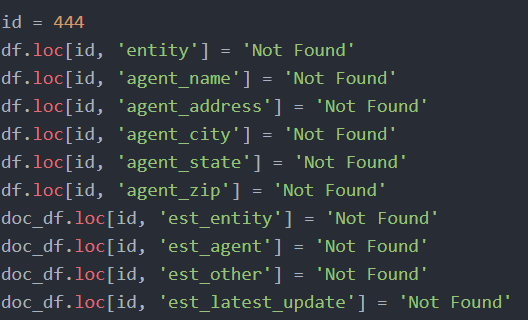


1. *If you do not find a record:*
   1. *Play around with additional queries with simplified versions of the DBA Name*

I tried “Meskawi Trading” and “Meskwaki” but still got no relevant results.

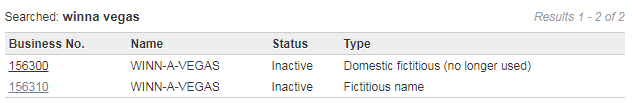
*b. start a new block of code in file 4a for that retailer mark all fields as “Not found” and move on to the next record.*

We did not find a record so we’ll do just that:

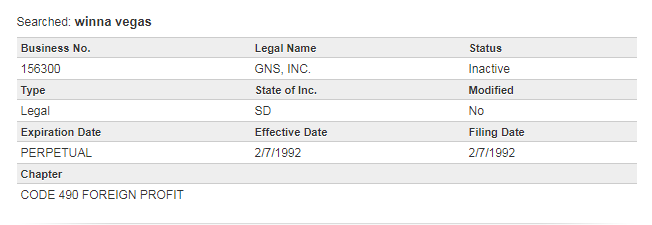


1. *If you do find a record:*

Let’s use another store where we did find a record: IMPAQ ID 445, “Heritage Express-Winna Vegas. The term “Winna Vegas” produces two results:

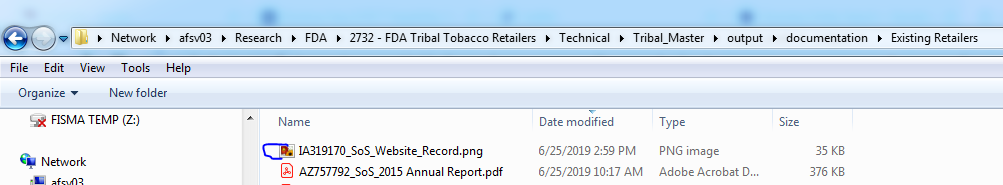


We are not looking for ficticious names, but rather the legal entity owning it. However, the legal entity is named in the ficticious name entry. First, let’s figure out which is the right record. The establishment address is in Sloan, IA. No. 156300’s registered agent is located in Sloan, IA, while 156310 is located in Sioux City, IA. We conclude 156300 is the correct record to use. The owning legal entity is GNS, Inc:



* 1. *Save a screenshot of the website record in the following folder/filename:*
     1. *“Tribal\_Master/output/documentation/New Retailer/IMPAQID\_####\_SoS\_Website\_Record.png”, if a new retailer*
     2. *“Tribal\_Master/output/documentation/Existing Retailer/[Insert REI value] \_SoS\_Website\_Record.png”, if an existing retailer*

This is an existing retailer, so we’ll save it in the appropriate folder with its REI.

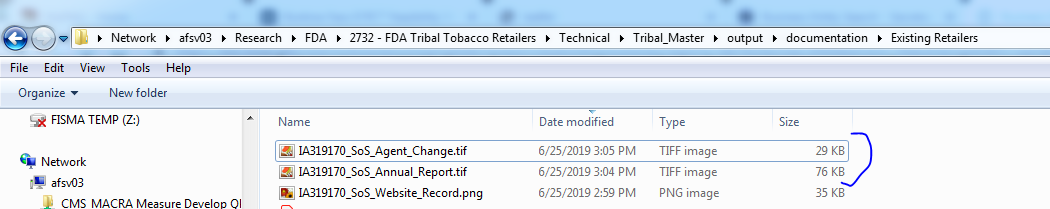


* 1. *Save any of the following documents that are freely available in the same folders, with the same filename structure except replacing “Website\_Record” with as a PDF file type:*
     1. *Articles of incorporation/formation/registration, the latest annual report, the latest agent change notice.*

Clicking on filings reveals several documents associated with the entity, but only a few are both retrievable and relevant for our purposes (latest annual report and agent change)

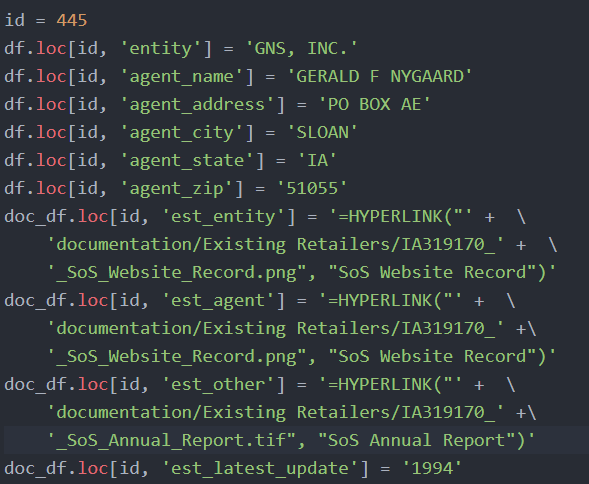


We will download and save in the same folder and appropriate filenames:



* 1. *Start a new block of code in file 4a for that retailer. “df” fields are the actual data. Input appropriate values from the website/latest documentation. Doc\_df are excel formula hyperlinks eventually written into the “Sources\_Documentation” excel files. Replace the file path and name with the appropriate value to reference the source document you used to determine the entity (est\_entity) and agent (est\_agent). If other documentation was present, list one of the other files in “est\_other”*

Start a new block of code for this retailer and fill in the values from the website record. Note that “est\_latest\_update” will be the latest date a document was submitted to So; in this case, 1994.



**Verifying changes to existing entities are sensible**

Make all changes with a separate Python file in a similar fashion to 4a\_manual ownership changes

Changes to verify are correct in “2\_Revised\_Data\_Change\_Log”/”3\_Revised\_Data\_Clean”:

* 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.

**Verifying Documentation Links**

In files 4 and 7, ensure links work properly. Verify file is present/edit with code if not working.