ETL process

Wine

1. Created database and tables, but relaxed the NULL on the load files
   1. Made changes to eliminate the id not allowing dups for testing
2. Had to fix some text lengths
3. Load into Postgres – getting & failure due to the registry error
   1. Set up loader…

COPY xxxx

FROM 'C:\Users\edwar\Desktop….”

DELIMITER ',' CSV HEADER;

* 1. Had to fix permissions with this:

<https://stackoverflow.com/questions/14083311/permission-denied-when-trying-to-import-a-csv-file-from-pgadmin>

address\_load (same as above on theloader etc)

Merging:

Uppcase the winery names into a special variable

Match the winery “like” the operating names

* Should match on exact name
* The Like will match without the trailers like:
  + Vintners, Cellars, Vineyards, Winery,Wines, Wine

Log of activity:

* Downloaded all the addresses from the government site
* Delete and recreated the database with clean sql
* Load all the California addresses into the address\_load table
* Check the counts, and the final count
* Tested the merge
* Started first major merge at 19:11 pm 4/25
  + Finished quickly – 24,245 matches
* Need to
  + Summarize by winery / operator name
    - Verify matches
  + Summarize by wine
    - Eliminate dups
  + Look over nulls/missing data and get down to 10K

Adding the geocoordinates

* Loaded all addresses for California into one file
* Deleted all with no operating name since these are needed for a match
* Deleted all with missing street address, since will not be able to geocode
* Saved as Cal\_Add\_All\_Clean.csv

Name clean up:

* Remove
  + Wines