CSV PROCESS

state\_ppp\_loans/

**RAW FILES:**

data/public\_150k\_plus\_211121.csv

data/public\_150k\_plus\_220102.csv - latest

**EXCEL:**

data/public\_150k\_plus\_211121\_add.csv

* NAICS > NA
* BorrowerZIP #####-#### > #####

Loan # 4204237701 (VA > VT in BorrowerState)

data/naics\_codes.csv – spreadsheet NAICS code sub-sector industry information

**WORKBOOKS**

**data/state\_data/workbooks/{state}\_workbook**

using {state}.csv –

* Loan by Lender Overview
* Loan by Lender, Industry
* Loan by Lender, Range
* Loan by Lender, Business Type
* Loan by Lender, Zone Filters
* Loan by City Overview
* Loan by City, Industry
* Loan by City, Range
* Loan by City, Business Type
* Loan by City, Zone Filters
* Loan by City, Range, Non-Profit
* Loan Approval Quarter, Lender, Business Type
* Loan by Business Type, Age

**data/state\_data/industry/industry\_workbooks/{state}\_industry**

using {state}\_industry.csv –

* State Industry Overview
  + Industry > Industry Subsector > LongName
    - Loan Count, Amount, JobsReported
* State by Industry, Sector, Borrower
  + Industry > Industry Subsector > LongName > Borrower
    - Loan Count, Amount, JobsReported
* State by Industry, Range, Borrower
  + Industry > Industry Subsector > LongName > Loan\_Range, Borrower drop-down
    - Loan Count, Amount, JobsReported
* State by Quarter, Industry, Sector, Borrower
  + Years > Quarters > Date Approved > Industry > Industry Subsector > LongName > Borrower
    - Loan Count, Amount, JobsReported
* State by Quarter, Lender, Industry
  + Years > Quarters > Date Approved > ServicingLender > Industry > Industry Subsector > LongName > Borrower
    - Loan Count, Amount, JobsReported
* State by Quarter, Business Age, Industry
  + Years > Quarters > Date Approved > BusinessAgeDescription > Industry > Industry Subsector > LongName > Borrower
    - Loan Count, Amount, JobsReported

**JUPYTER NOTEBOOK**

**File:** main\_label\_pipe.ipynb

Read-in: public\_150k\_plus\_211121\_add.csv

+ Estimated\_Lender\_Percent (1%, 3%, 5%)

+ Estimated\_Lender\_Profit (Approved \* %)

+ DateApproved\_ts (timestring conversion of DateApproved)

+ ForgivenessDate\_ts (timestring conversion of ForgivenessDate, Blank Fields > -ts)

+ Industry (Labels for 1 of 21 NA codes)

+ SBAOfficeCode (Labels for 1 of 94 codes)

+ Loan\_Range (Labels for 1 of 5 segmentations)

+ Job\_Range (Labels for 8 Ranges). \* possible use in NAICS comp \*did not run

+ full\_add (combination of address+city+state)

Exported As: (“data/ppp\_pipe.csv”)

\* Division of main DataFrame by State

\* Conversion of Each State DataFrame to CSV

Exported As: (“data/state\_data/ as {statename}.csv”)

**File:** main\_labeling\_geo.ipynb

Read-in: (“data/ppp\_pipe.csv”)

* Geo-coding AK – AZ, VA

+ Lat, Long (coordinates from full\_add)

* Division of main DataFrame by State
* Conversion of AK – AZ, VA DataFrame to CSV

Exported As: (“data/state\_data/geo/{statename}\_geo.csv”)

**File:** state\_industry\_label\_pipe.ipynb

Read-In: "data/state\_data/{state}.csv"

Read-IN: “data/naics\_codes.csv”

* Combines DataFrame with correspondent Industry Sub-Sector information from NAICSCode spreadsheet

+ IndustrySubSector, 4DigitNAICS

+ TRIIndustrySector

+ LongName

Exported As (“data/state\_data/industry/{state}\_industry.csv”)

+ Job\_Range (Labels for 8 Ranges). \* possible use in NAICS comp \*ran on VA

Exported As: ("data/state\_data/industry/va\_industry\_job.csv") for testing

File: poly\_test.ipynb

**Folders/Files:**

Data/notfound/notfound\_{state}.csv

* Jupyter notebook spreadsheet export of addresses not found in geocode process
  + By State

Data/<150K/public\_up\_to\_150k\_{#1-12}\_211121

* 12 sheets, state to state alphabetical order all loans <150K

Data/alt\_data/SQGDB

* 255 files YEAR:QUARTER GDP by State, 05-2021, by Industry

qcew\_python\_3x\_example.py

* Example Labor Statistics Query

RESOURCES:

**TIGER/Shapefiles**

Input FIPS – return cords for State, County, Census Tract, Block, Block Group

**Census Geocoder**

Input Address, Return FIPS (batch processing, file export, 2021 updated, leaky results)

**FCC API**

Input Lat, Long, Return FIPS (batch processing, json export, not updated)

**GOOGLE MAPS GEOCODER API**

Input Address, Return Lat, Long

poly\_test.ipynb

selenium\_census\_script.ipynb

[main\_label\_pipe.ipynb](http://localhost:8888/notebooks/state_ppp_loans/main_label_pipe.ipynb)

* Takes in raw file, NA, SBA Labels

[main\_labeling\_geo.ipynb](http://localhost:8888/notebooks/state_ppp_loans/main_labeling_geo.ipynb)

* Geocoding main file, coords output in state-to-state file (fields differ form main\_label)

[state\_industry\_label\_pipe.ipynb](http://localhost:8888/notebooks/state_ppp_loans/state_industry_label_pipe.ipynb)

* Expanded NAICS codes \*add to main label process, (12,182 NAICS not matched of total)

[selenium\_census\_script\_fipscheck.ipynb](http://localhost:8888/notebooks/state_ppp_loans/selenium_census_script_fipscheck.ipynb)

* Richmond, VA FIPS tests

[selenium\_census\_script\_fipscheck\_VA.ipynb](http://localhost:8888/notebooks/state_ppp_loans/selenium_census_script_fipscheck_VA.ipynb)

* Virginia State FIPS matching on Lat, Long (fields differ from main\_geo, main\_label)

[selenium\_census\_script\_fipscheck\_CA.ipynb](http://localhost:8888/notebooks/state_ppp_loans/selenium_census_script_fipscheck_VA.ipynb)

* California State FIPS matching on Lat, Long (fields differ from main\_geo, main\_label)

[poly\_test.ipynb](http://localhost:8888/notebooks/state_ppp_loans/poly_test.ipynb)

* Shapefile pairing tests with accordant FIPS codes

[fips\_coord\_mapping.ipynb](http://localhost:8888/notebooks/state_ppp_loans/fips_coord_mapping.ipynb" \t "_blank)

* Script for mapping Shapefiles to FIPS Codes, state-by-state
* Maps to census file for population reference using GISJoin from NHGIS files \*polygon, plotly

DOCUMENTS:

State\_Processing\_Workbook.docx

* Index for State\_Processing\_Chart.xlsx – tracks location labeling process state-by-state