Repo: **mapdata\_finalization\_notebooks**

**Processes Raw PPP Loan Data obtained from SBA Public Data Directory for 2021 and 2023.**

**Scripts:**

**main\_label\_pipe.ipynb -** processes raw data through pipeline of cleaning, formatting, and integrating associated data sources to expand on given database.

**main\_hashing.ipynb -** creates HASHcode column for dataframe with variables for Industry, Industry Subsector, Loan Range, Job Range, Business Type, Business Age, and PPP/PPS program.

**main\_concat.ipynb –** combines final outputs of main\_label\_pipe, main\_hashing, and Geo processed FIPS files for finalized output to be used in state aggregation.

**review/**

**process\_checks.ipynb –** Comparison Check of File Size, Columns (Names/Count), Row Counts, dtypes for all main script process inputs/outputs for 2021 and 2023 files. Includes overview of: Raw input files, ppp\_pipe and ppp\_pipe23, ppp\_hash and ppp\_hashSort, ppp\_hpgf, ppp\_geo, ppp\_geo\_fips.

**raw\_compare.ipynb –** Comparison of Raw 2021 and Raw 2023 file updates, changes, additions, deletions.

**update\_compare.ipynb -** Addresses in Raw 2023 file compared to addresses in final 2023 output file.

**null\_updates.ipynb –** Outputting of files for select columns containing Null values.

**dataviews/**

**dataviews.ipynb –** Exploratory Data Post Pipe process:

figs/

DataFrame: (Grouped) SBA Office Label, Code, (Aggregated) Loan Number(count).

Plot: Box, SBA Office Label vs Current Approval Amount(sum)

DataFrame: Grouped (Borrower State), (Aggregated) Loan Count, Current Approval Amount, Average Loan Amount

Plot: Box, Borrower State vs Current Approval Amount (values)

Plot: Bar, Borrower State vs Current Approval Amount (sum)

DataFrame: (Grouped) Servicing Lender, Location ID, (Aggregated) Loan Number(count), Current Approval Amount(sum), Average Loan Amount

DataFrame: (Grouped) Originating Lender, Location ID, (Aggregated) Loan Number(count), Current Approval Amount(sum), Average Loan Amount

DataFrame: Lender Match, filtered where Servicing Lender != Originating Lender

DataFrame: From filtered Lender Match DataFrame (Grouped) Servicing Lender, Location ID, Originating Lender, (Aggregated) Loan Number(count), Current Approval Amount(sum), Estimated Lender Profit (sum)

DataFrame: (Grouped) Date Approved, (Aggregated) Loan Number (count)

Plot: Scatter, Date Approved (x), Loan Number (count) (y) for Selected Range

DataFrame: (Grouped) Date Approved, (Aggregated) Count Each State Column on Date

Plot: Scatter, Date Approved (x), State Columns (count) (y) for Selected Range based on Data Available.

doc/

pipe\_descriptions.docx - detailed process pipeline descriptions

pipeline.png - pipeline process flowchart

tsfilelist.txt - repository file listing with size, last modification time