1.Go to <https://data.sba.gov/dataset/ppp-foia> and download all dataset files.

Diagram

Description automatically generated

2. merge the **first five-digit of BorrowerZip to county fips using the excel ZIP\_COUNTY\_122021**

**If one zip corresponds to multiple county, you only need to match with the county with highest res\_ratio[[1]](#footnote-1)**

Then, aggregate the data based on the **year-month of DateApproved AND Borrower county fip**. What we need to calculate is

* Total number of loans
* Ratio of rural area loan (RuralUrbanIndicator)
* Ratio of Hub zone loan (HubzoneIndicator)
* Ratio of LMI loan (LMIIndicator)
* Avg employees (JobsReported)
* Total employees (JobsReported)
* Avg loan approval (InitialApprovalAmount)
* Total loan approval (InitialApprovalAmount)

Diagram

Description automatically generated with medium confidence

3. aggregate the data based on the **year-month of DateApproved AND** BorrowerState. What we need to calculate is

* Total number of loans
* Ratio of rural area loan (RuralUrbanIndicator)
* Ratio of Hub zone loan (HubzoneIndicator)
* Ratio of LMI loan (LMIIndicator)
* Avg employees (JobsReported)
* Total employees (JobsReported)
* Avg loan approval (InitialApprovalAmount)
* Total loan approval (InitialApprovalAmount)

Graphical user interface, text, application, table, Excel

Description automatically generated

4. aggregate the data based on the **year-month of DateApproved, Borrower county fip**, and first two-digit of NAICS. What we need to calculate is

Table

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

5. aggregate the data based on the **year-month of DateApproved, BorrowerState, and first two-digit of NAICS.** What we need to calculate is

1. the residential proportion; see <https://www.huduser.gov/portal/periodicals/cityscpe/vol22num1/ch12.pdf> [↑](#footnote-ref-1)