* 8.8mil rows
* Cert col will allow connection w Bank Call Report data
* First\_draw\_ppp\_originator\_lvl: Aggregated by
  + Loan count: count of records from the Masterfile
  + CERT (col L): FDIC (depository banks)
  + Use pivot table to find total counts
* Files that csbs have created are not original PPP data
  + Original file contains missing values present
  + Only subsets
* Added FDIC cert so that other data can be connected
* Request for additional info if needed
  + [data@csbs.or](mailto:data@csbs.or)g
  + [tsiems@csbs.org](mailto:tsiems@csbs.org) (etc)
* Exclude second draw PPP
  + To avoid duplication at company/bank lvl analysis
* Find trends at different levels
  + Creativity!
* How to define community bank:
  + Well-known banks: Complex
    - Derivatives
    - Trading, etc
  + Smaller community banks = “Cookie cutter banks”
    - Deposits > loans > make $ on interest
* Community banks
  + more connected to communities
  + found to be instrumental in PPP rollout
  + PPP Vol: Disproportionately larger than banks’ size
    - How? Why?
    - Complexity is more of the focus
    - Size is more clear-cut
  + FDIC def available
    - 6~7 parameters must be satisfied
* Bank call report: 10,000 banks in the US
  + Types: National, state, others
* Majority = 4,000 = community
  + Mostly state-chartered
* Must be FDIC insured
* Quarterly performance is available in bank call report
  + Next office hr: In-depth
* Significance of missing values for forgiven amt: Will check
  + Business either not applied for forgiveness or did not follow up
* CSBS data’s date stamp: Jun 30, 2021
* NAICS code: North American Industrial Category Code (naics.com)
  + Code to categorize business
  + 6 digits in a code
    - Very specific
    - First two #s: Industry type
  + Interesting to look at: Tells us the type of business
* Pull in data from other sources
  + Where loans are going to
  + Bureau labor statistics: For employment stats
  + Do not limit analysis to CSBS dataset
* Objectives: Where
* Goals: What
* Strategies: How
* Measures: Metrics
  + Make use of financial ratios to assess
  + Accuracy rate: >= 70%
* Know your data
* Formulating a strong and compelling hypothesis is crucial
  + Solid statement
  + Data-driven analysis supporting your hypothesis
* Stories are compelling

Next meeting agenda (tentatively: 25th Feb - Fri) :

* Clean datasets
* Merge datasets
* Set up rules on how to deal with null values
* Trial run on sample data
* Recheck hypothesis
* Set up an appt with Wilck to touch base with our progress

https://www.csbs.org/newsroom/data-corner-ppp-data-highlights-significance-state-chartered-banks-csbs-analysis-shows