



# PPP loans in America



By Tenzing Palden  
and  
Cristian Marcatoma

## Information Before Entering the Project

- The **Paycheck Protection Program (PPP)** is a federal loan program designed to provide financial assistance to businesses affected by the COVID-19 pandemic.
- The PPP is administered by the **Small Business Administration (SBA)** and is available to eligible businesses with 500 or fewer employees. The loan terms include a low interest rate and the potential for loan forgiveness if certain requirements are met.
- The maximum loan amount available under the PPP is **\$10 million**.

# Business Problem


- PPP loans are intended to help businesses maintain their payroll and cover certain expenses, such as rent and utilities, during the economic downturn caused by the pandemic.
- Shortage of customers



## Business Impact

- Conducting a data analysis project on PPP loans can provide valuable insights for both the government and small businesses.
- For the government, a thorough analysis can help identify areas for improvement in the program and ensure that funds are being distributed effectively.
- For small businesses, a data analysis project can help them understand how PPP loans are being used by other companies in their industry, and provide them with information that can help them make more informed decisions about how to use the funds they receive.

# Database → U.S. Small Business Administration

U.S. Small Business Administration

DatasetsOrganizationsGroupsAbout


Search

Home / Organizations / Office of Capital Access / **PPP FOIA**

### PPP FOIA

Followers  
**0**

Organization








U.S. Small Business

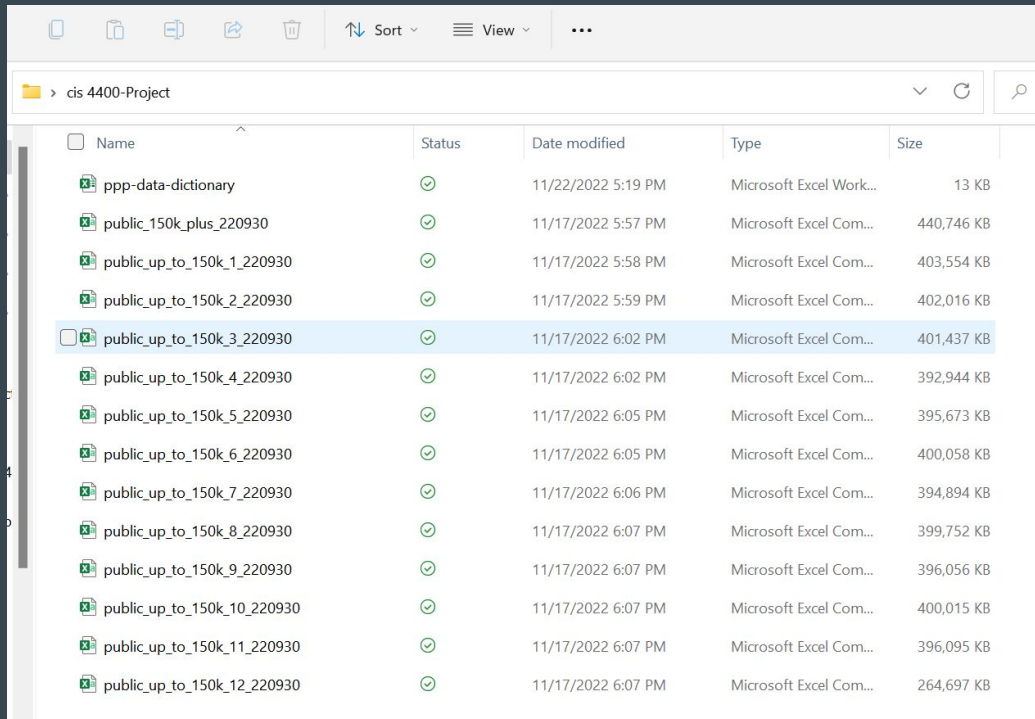
DatasetGroupsActivity Stream

## PPP FOIA

### Data and Resources

	public_150k_plus_220930.csv	<a href="#">Explore</a>
	public_up_to_150k_1_220930.csv	<a href="#">Explore</a>
	public_up_to_150k_2_220930.csv	<a href="#">Explore</a>
	public_up_to_150k_3_220930.csv	<a href="#">Explore</a>
	public_up_to_150k_4_220930.csv	<a href="#">Explore</a>

# Extraction



The screenshot shows a Windows File Explorer window with the address bar set to 'cis 4400-Project'. The main area displays a list of files in a table format. The files are all Microsoft Excel Comma-separated values (CSV) files, each with a green checkmark in the Status column. The file 'public\_up\_to\_150k\_3\_220930' is highlighted in blue.

Name	Status	Date modified	Type	Size
ppp-data-dictionary	✓	11/22/2022 5:19 PM	Microsoft Excel Work...	13 KB
public_150k_plus_220930	✓	11/17/2022 5:57 PM	Microsoft Excel Com...	440,746 KB
public_up_to_150k_1_220930	✓	11/17/2022 5:58 PM	Microsoft Excel Com...	403,554 KB
public_up_to_150k_2_220930	✓	11/17/2022 5:59 PM	Microsoft Excel Com...	402,016 KB
public_up_to_150k_3_220930	✓	11/17/2022 6:02 PM	Microsoft Excel Com...	401,437 KB
public_up_to_150k_4_220930	✓	11/17/2022 6:02 PM	Microsoft Excel Com...	392,944 KB
public_up_to_150k_5_220930	✓	11/17/2022 6:05 PM	Microsoft Excel Com...	395,673 KB
public_up_to_150k_6_220930	✓	11/17/2022 6:05 PM	Microsoft Excel Com...	400,058 KB
public_up_to_150k_7_220930	✓	11/17/2022 6:06 PM	Microsoft Excel Com...	394,894 KB
public_up_to_150k_8_220930	✓	11/17/2022 6:07 PM	Microsoft Excel Com...	399,752 KB
public_up_to_150k_9_220930	✓	11/17/2022 6:07 PM	Microsoft Excel Com...	396,056 KB
public_up_to_150k_10_220930	✓	11/17/2022 6:07 PM	Microsoft Excel Com...	400,015 KB
public_up_to_150k_11_220930	✓	11/17/2022 6:07 PM	Microsoft Excel Com...	396,095 KB
public_up_to_150k_12_220930	✓	11/17/2022 6:07 PM	Microsoft Excel Com...	264,697 KB

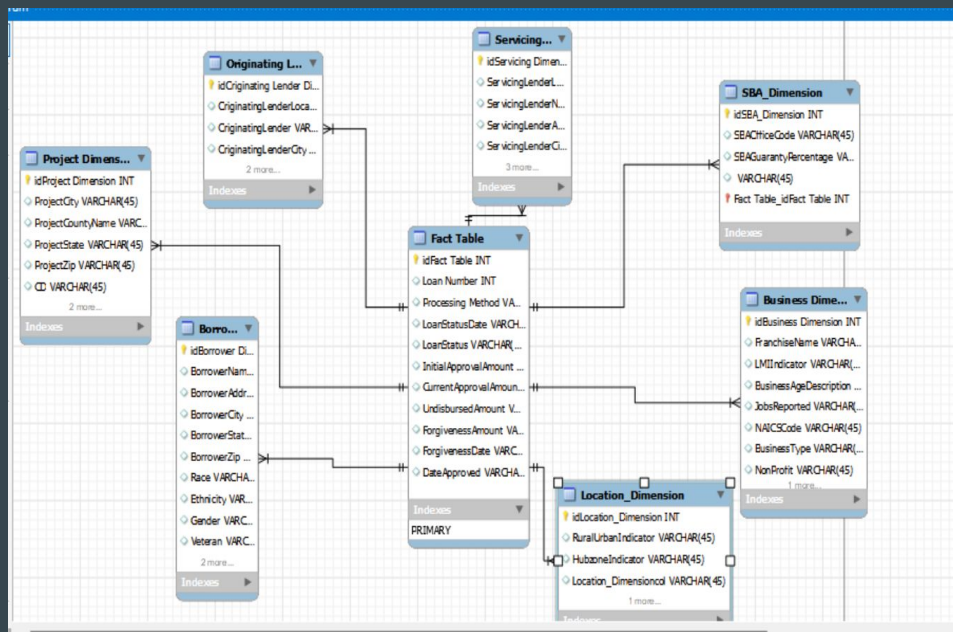
We download each CSV file onto our local machine.

Later these files were used to upload into our Google Cloud Buckets




























Having Data in house made the process easier to organize and simpler to work with.

# Data modeling with SQL Workbench and Data Dictionary

File Home Insert Draw Page Layout Formulas Data Review View Automate Help		
C11		
Fact Table		
A	B	C
1	<b>Field Name</b>	<b>Field Description</b>
2	LoanNumber	Loan Number (unique identifier)
3	DateApproved	Loan Funded Date
4	SBAOfficeCode	SBA Origination Office Code
5	ProcessingMethod	Loan Delivery Method (PPP for first draw; PPS for second draw)
6	BorrowerName	Borrower Name
7	BorrowerAddress	Borrower Street Address
8	BorrowerCity	Borrower City
9	BorrowerState	Borrower State
10	BorrowerZip	Borrower Zip Code
11	LoanStatusDate	Loan Status Date
12	LoanStatus	- Loan Status is replaced by 'Exemption 4' when the loan is disbursed but not Paid In Full or Charged Off
13	Term	Loan Maturity in Months
14	SBAGuarantyPercentage	SBA Guaranty Percentage
15	InitialApprovalAmount	Loan Approval Amount (at origination)
16	CurrentApprovalAmount	Loan Approval Amount (current)
17	UndisbursedAmount	Undisbursed Amount
18	FranchiseName	Franchise Name
19	ServicingLenderLocationID	Lender Location ID (unique identifier)
20	ServicingLenderName	Servicing Lender Name
21	ServicingLenderStreetAddress	Servicing Lender Street Address



# Buckets- Google Cloud Storage- Dataset

▼	 public_150k_plus_220930		
	 Borrower_Dimension		
	 Business_Dimension		
	 Fact_Table		
	 Location_Dimension		
	 Originating_Lender_...		
	 Project_Dimension		
	 SBA_Dimension		
	 Servicing_Dimension		

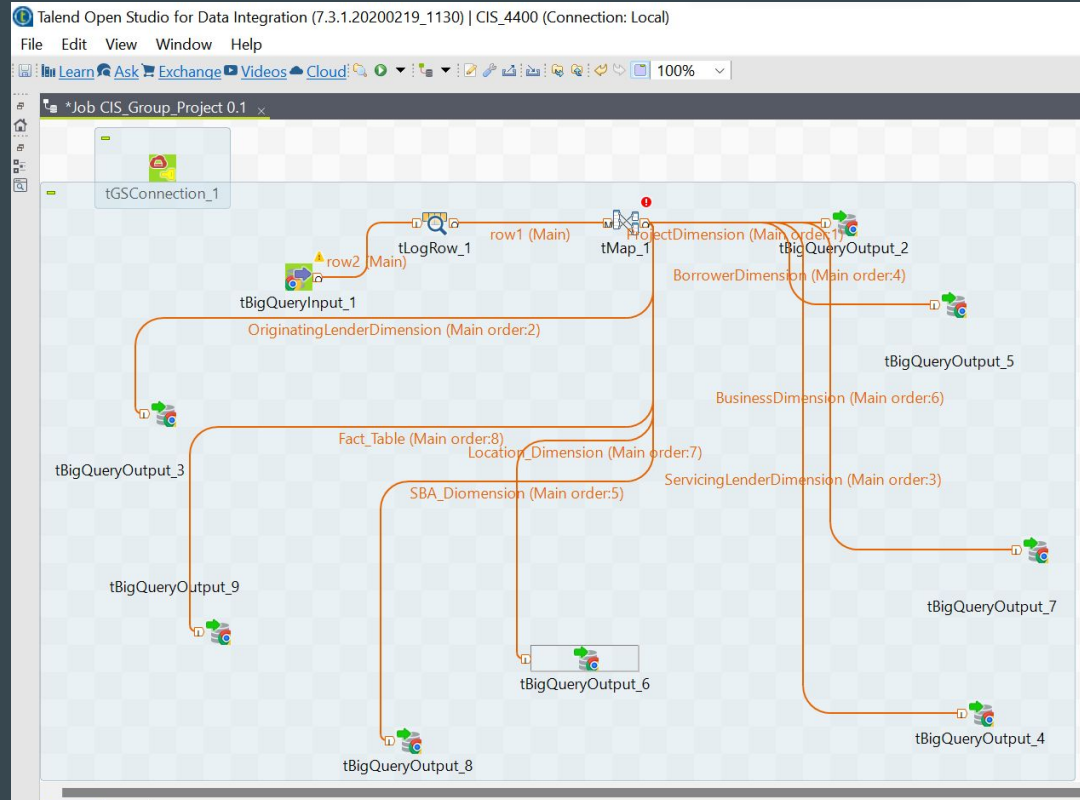
We download each CSV file onto our local machine.

Later these files were used to upload into our Google Cloud Buckets

Having Data in house made the process easier to organize and simpler to work with.



# Data Integration with Talend



# PYTHON CODING TO FIND OUT GOALS

```
top20countryfreq = new_df.value_counts("OriginatingLender")[:20]
```

Python

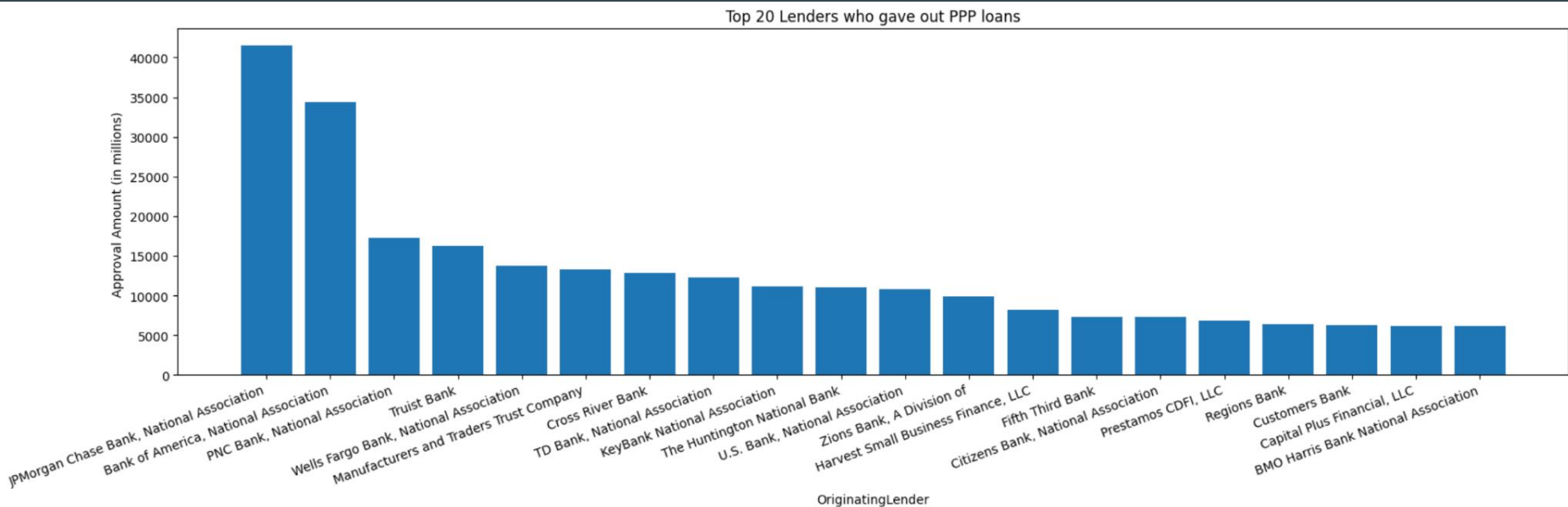
```
countriesum=new_df.groupby(['OriginatingLender']).agg({"CurrentApprovalAmount": "sum"}).sort_values("CurrentApprovalAmount", ascending=False)  
top20countriesum=countriesum[:20]
```

Python

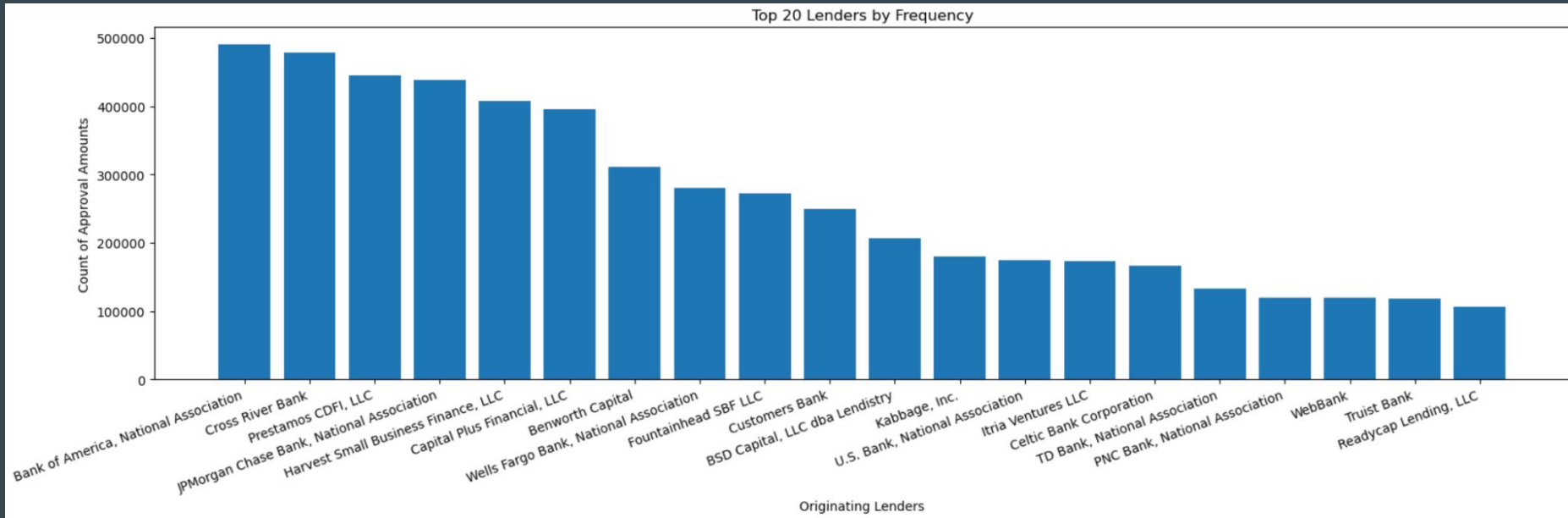
```
state_list = [ "AE", 'AK', 'AL', 'AR', 'AZ', 'CA', 'CO', 'CT', 'DC', 'DE', 'FL', 'GA',  
               'HI', 'IA', 'ID', 'IL', 'IN', 'KS', 'KY', 'LA', 'MA', 'MD', 'ME',  
               'MI', 'MN', 'MO', 'MS', 'MT', 'NC', 'ND', 'NE', 'NH', 'NJ', 'NM',  
               'NV', 'NY', 'OH', 'OK', 'OR', 'PA', 'RI', 'SC', 'SD', 'TN', 'TX',  
               'UT', 'VA', 'VT', 'WA', 'WI', 'WV', 'WY']
```

```
statesum= new_df.groupby(["BorrowerState", "OriginatingLender"], ).CurrentApprovalAmount.sum()  
statesum2= new_df.groupby(["BorrowerState", "OriginatingLender"], as_index=False).CurrentApprovalAmount.sum()  
statesum2.to_csv("statesum_update.csv")  
statefreq= new_df.groupby(["BorrowerState", "OriginatingLender"], as_index=False).OriginatingLender.count()  
statefreq.to_csv("freq.csv")
```

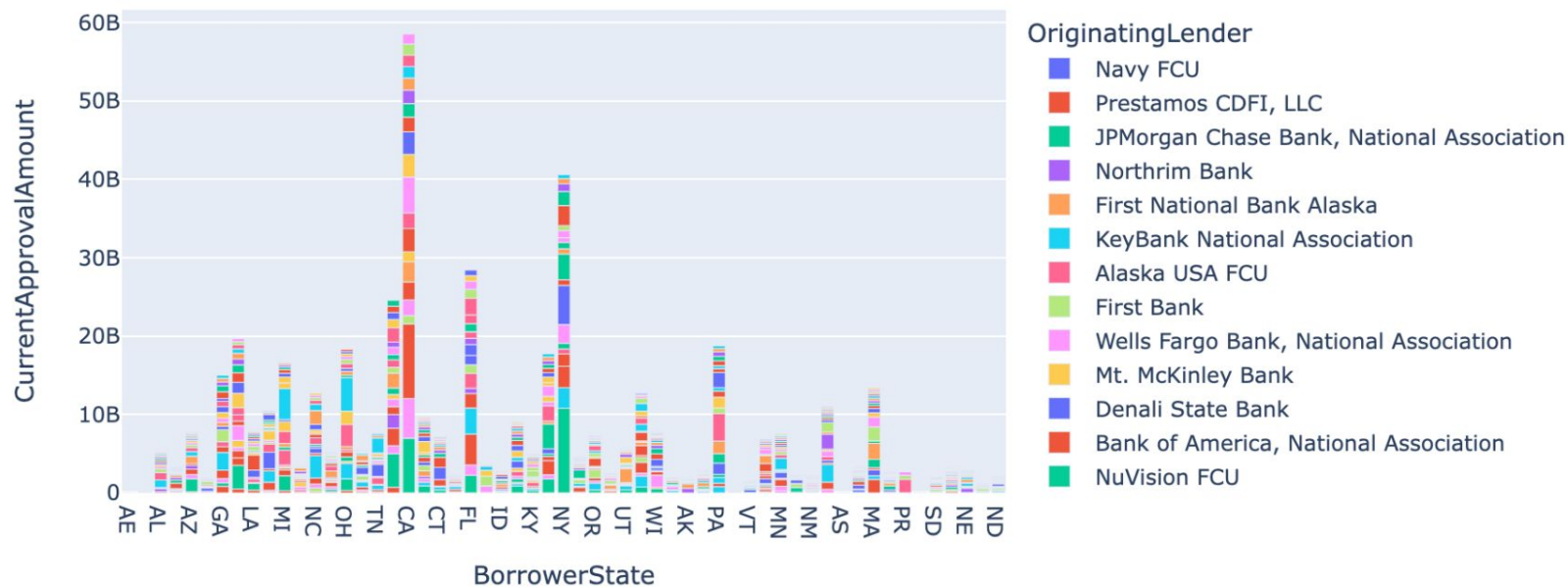
# TOP 20 LENDERS (BANKS) IN AMERICA OF PPP LOANS



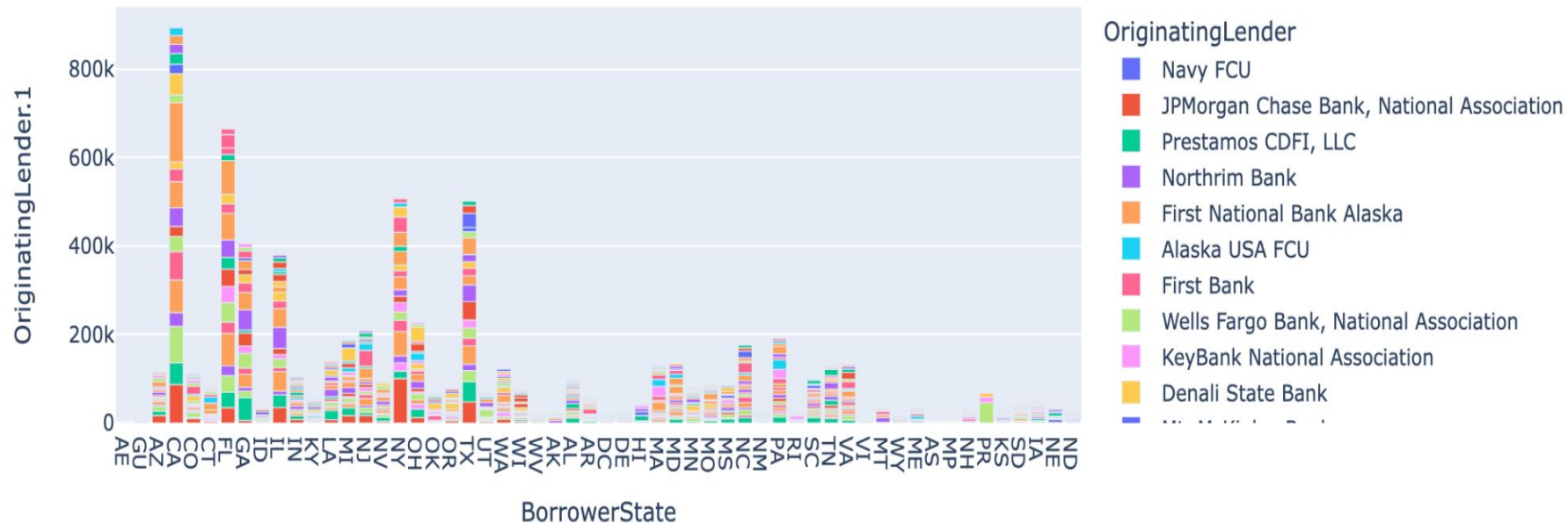
# TOP 20 FREQUENCY (BANKS) IN AMERICA OF PPP LOANS



## Top 20 Lenders for PPP loans per state



## Top 20 Lenders for PPP loans per stat by Amount of loans given



# Insights

- JP Morgan, Bank of America and PNC Bank were the top 3 leading lenders during the pandemic in nation.
- California, New York and Florida were the states that received the most aid.
- An interesting fact is that some army bases around the world received PPP loans
- The lowest amount of loans in the top 20 were given by BMO Harris Bank National Association, Capital Plus Financial, LLC, Customers Bank.



Thank  
you!