Paycheck Protection Program Loan Data

SUMMARY

The Paycheck Protection Program (PPP) is a nearly $1 trillion business loan program started in 2020 under the Trump administration to provide relief to businesses struggling due to the Coronavirus epidemic. This dataset represents only businesses who received loans of more than $150,000, and can be further explored to gain insights in the industry.

DATA SOURCE

The data used for this project is open-source data, publicly available on Kaggle - Paycheck Protection Program Loan Data.

LIMITATIONS AND ETHICS

➢ No personally identifiable information is included in this data
➢ 13 observations were removed due to missing the borrower's State
➢ 51,044 businesses have not disclosed the length of time they have been operating for (BusinessAgeDescription column, left Unanswered)
➢ 719 BusinessType missing
➢ 29,646 ForgivenessAmount missing
➢ ProcessingMethod PPS represents a second draw. This means that 291,668 loans from the data set are second loans received by the businesses, a year later. These businesses are listed twice for this reason

DATA PROFILE

The original data set is made up of 53 columns and 968,531 observations. The cleaned data set is made up of 20 columns and 968,518 observations.

The following columns were removed due to insufficient/ irrelevant data: SBAOfficeCode, BorrowerName, BorrowerAddress, SBAGuarantyPercentage, UndisbursedAmount, FranchiseName, ServicingLenderLocationID, ServicingLenderAddress, ServicingLenderCity, ServicingLenderState, ServicingLenderZip, HubzoneIndicator, LMIIndicator, ProjectCity, ProjectCountyName, ProjectState, ProjectZip, CD, NAICSCode, Race, Ethnicity, UTILITIES_PROCEED, PAYROLL_PROCEED,

MORTGAGE_INTEREST_PROCEED, RENT_PROCEED, REFINANCE_EIDL_PROCEED, HEALTH_CARE_PROCEED, DEBT_INTEREST_PROCEED, OriginatingLenderLocationID, OriginatingLenderCity, OriginatingLenderState, Gender, Veteran

#	Columns	Description	Time Variant / Invariant	Data type
1	LoanNumber	Unique identifier for each loan	Invariant	Quantitative
2	DateApproved	Date that the loan was funded	Variant	Quantitative
3	ProcessingMethod	Loan Delivery Method (PPP for first draw; PPS for second draw)	Invariant	Qualitative
4	BorrowerCity	Borrower City	Invariant	Qualitative
5	BorrowerState	Borrower State	Invariant	Qualitative
6	BorrowerZip	Borrower Zip Code	Invariant	Qualitative
7	LoanStatusDate	Loan Status Date is blank when the loan is disbursed but not Paid In Full or Charged Off	Variant	Quantitative
8	LoanStatus	Loan Status Description - Loan Status is replaced by 'Exemption 4' when the loan is disbursed but not Paid in Full or Charged Off	Invariant	Qualitative
9	Term	Loan Maturity in Months	Invariant	Quantitative
10	InitialApprovalAmount	Loan Approval Amount(at origination)	Invariant	Quantitative
11	CurrentApprovalAmount	Loan Approval Amount (current)	Invariant	Quantitative
12	ServicingLenderName	Servicing Lender Name	Invariant	Qualitative
13	RuralUrbanIndicator	Rural or Urban Indicator (R/U)	Invariant	Qualitative
14	BusinessAgeDescription	Business Age Description	Invariant	Qualitative
15	JobsReported	Number of Employees	Invariant	Quantitative
16	BusinessType	Business Type Description	Invariant	Qualitative
17	OriginatingLender	Original Lender Name	Invariant	Qualitative
18	NonProfit	'Yes' if Business Type = Non-Profit Organization or Non-Profit Childcare Center or 501(c) Non Profit	Invariant	Qualitative
19	ForgivenessAmount	Forgiveness Amount	Invariant	Quantitative
20	ForgivenessDate	Forgiveness Paid Date	Variant	Quantitative

QUESTIONS TO EXPLORE

1) What do the businesses who have not yet paid the loan in full have in common?
2) Where are most businesses that received a loan greater than $150,000 located?
3) What is the relationship between the number of employees and the loan amount granted?
4) Which lender approved the highest loan amounts?
5) How fast are loans paid off?