SBA Disaster Loan Patterns
(Group 2)

Team Members:

Ritika Shrivastava ritshriv@iu.edu

Yukta Muthreja ymuthrej@iu.edu

Shikha Punjabi shpunjab@iu.edu

Abhirav Dhiren Lande ablande@iu.edu

Introduction

The Small Business Administration (SBA) Disaster Loan Program plays a critical role in supporting individuals and businesses affected by disasters in the United States. Established to provide financial assistance during times of crisis, the program aims to help communities recover by offering low-interest loans to cover uninsured losses. These loans are crucial for rebuilding infrastructure, restoring economic stability, and providing relief to those impacted by events such as hurricanes, floods, and fires. However, questions often arise about the program's effectiveness in addressing verified losses and its ability to equitably distribute aid across different regions and disaster types.

This project aims to evaluate the effectiveness of the SBA Disaster Loan Program by analyzing patterns in approved loan amounts, verified losses, and recovery ratios across the United States. Specifically, it focuses on identifying the key drivers of variation in loan approval rates, examining how different disaster types influence approved loan amounts, and understanding geographical disparities in the program's response. Additionally, the study explores temporal trends to uncover long-term patterns in disaster recovery funding.

Our research delves into the nuanced factors that influence disaster loan distributions, recognizing that these financial resources represent more than statistical data—they are lifelines for communities rebuilding after catastrophic events. What drives the effectiveness of disaster recovery funding? Our goal is to advance our understanding of disaster support systems, identify key factors that impact loan approvals, and provide actionable insights that can help policymakers and administrators optimize resource allocation. By systematically analyzing these foundational elements, we seek to illuminate the broader context of disaster recovery support and its profound significance in helping communities' resilience in the face of adversity.

Statement of goals

- **Investigate the Effectiveness of the SBA Disaster Loan Program**
  Question: How effective is the SBA Disaster Loan Program in addressing the losses incurred during disasters in the US?
  Why do you care: Understanding the program's effectiveness is crucial for evaluating its impact on disaster recovery and identifying areas for improvement.
  Why should we care: The SBA Disaster Loan Program plays a vital role in supporting individuals and businesses affected by disasters, and assessing its effectiveness can inform policy decisions and resource allocation.
- **Analyze Variation in Loan Approval Rates**
  Question: What are the key drivers of variation in loan approval rates across different geographical divisions and time periods?
  Why do you care: Identifying the factors influencing loan approval rates can help policymakers and program administrators optimize the program's design and implementation.
  Why should we care: Understanding the factors driving variation in loan approval rates can ensure that the program is equitable, efficient, and effective in supporting disaster recovery efforts.

- **Examine the Impact of Disaster Types on Loan Amounts**
  Question: Are specific disaster types (e.g., hurricanes, floods, fires) associated with higher or lower average approved loan amounts?
  Why do you care: Analyzing the relationship between disaster types and loan amounts can provide insights into the program's responsiveness to different types of disasters.
  Why should we care: Understanding how disaster types influence loan amounts can help policymakers and program administrators allocate resources effectively and ensure that the program meets the needs of affected individuals and businesses.
- **Investigate the Distribution of Approved Loan Amounts Across Urban and Rural Regions**
  Question: How does the distribution of approved loan amounts compare across urban and rural regions affected by disasters?
  Why do you care: Examining the distribution of approved loan amounts across urban and rural regions can reveal potential disparities in access to disaster recovery resources.
  Why should we care: Understanding the distribution of approved loan amounts across different regions can inform policy decisions aimed at promoting equitable access to disaster recovery resources and supporting the resilience of affected communities.

About the Data

The dataset analyzed in this project provides detailed records from the SBA Disaster Loan Program, specifically focusing on home loans issued to individuals impacted by disasters. This dataset captures a wealth of information, enabling a comprehensive analysis of recovery patterns across regions, disaster types, and time periods. Key features of the dataset include:

- SBA Physical Declaration Number: SBA's unique physical disaster declaration number
- SBA EIDL Declaration Number: SBA's unique economic injury disaster declaration number
- FEMA Disaster Number: FEMA's unique number for declared disaster
- SBA Disaster Number: SBA's unique disaster number
- Damaged Property City Name: City location of the damaged property
- Damaged Property Zip Code: Zip Code location of the damaged property
- Damaged Property State Code: State Code location of the damaged property
- Total Verified Loss (Original): Total $ amount of losses verified by SBA
- Verified Loss Real Estate (Original): $ amount of Real Estate losses verified by SBA
- Verified Loss Content (Original): $ amount of the damaged property Content (personal property, auto, etc.) verified by SBA
- Total Approved Loan Amount (Original): Total $ amount of approved SBA disaster home loan
- Approved Amount Real Estate (Original): $ amount of approved Real Estate loan proceeds
- Approved Amount Content (Original): $ amount of approved Content loan proceeds

Note: The verified loss and approved loan totals in this report are "original" values only and may not reflect the most updated values. Verified losses are damage estimates only and are calculated by SBA loss verifiers (inspectors). Verified losses and approved loan amounts may change when new information is provided to SBA. For example, a disaster suvivor may discover additional damages which increase the overall costs for repair and increase the required SBA disaster loan amount. Similarly, the SBA disaster loan could be decreased if the disaster survivor receives additional recoveries (i.e. insurance or grant assistance) following the original approval decision which duplicate SBA's assistance.

The dataset comprises thousands of records, with each representing an individual loan application. Each record includes both the applicant's verified losses and the approved loan amounts, making it possible to evaluate recovery ratios and identify gaps between needs and funding.

Challenges and Limitations

While the dataset provides extensive details, there are certain challenges and limitations. Missing or incomplete data in some records, particularly in verified losses and approved amounts, may impact the comprehensiveness of the analysis. Additionally, regional biases, such as disparities in the number of applications submitted from urban versus rural areas, could affect the interpretation of results. Despite these challenges, the dataset remains a robust resource for understanding the effectiveness of the SBA Disaster Loan Program.

By leveraging this dataset, the project is positioned to uncover critical insights into geographical disparities, disaster-specific trends, and long-term funding patterns. These findings will inform the analysis and help address the research questions related to recovery effectiveness and equity.

Data Preprocessing

To prepare the dataset for analysis, a series of preprocessing steps were undertaken to address data quality issues, transform variables for analysis, and create features relevant to the research questions. These steps ensured the dataset was consistent, clean, and ready for statistical evaluation and modeling.

Handling Missing Values

Missing values were present in critical columns, including verified losses and approved loan amounts. Records with extensive missing data in key columns were excluded to prevent biases in the analysis.

Log Transformations

Logarithmic transformations were applied to both approved loan amounts and verified losses. These transformations normalized the skewed distributions of financial data, allowing for better visualization and analysis of proportional relationships. Log-transformed data also improved the interpretability of trends in loan amounts and recovery ratios

Feature Engineering

Several new features were created to facilitate analysis and enhance insights:

- **Geographical Grouping:** Geographic regions were categorized into divisions, such as coastal vs. inland and urban vs. rural, for comparative analysis.
- **Disaster Type Encoding:** Disaster types were encoded into categorical variables to simplify analysis and visualization of their impact on loan approvals.
- **Temporal Aggregation:** Data was aggregated by year and region to explore long-term trends and identify patterns over time.

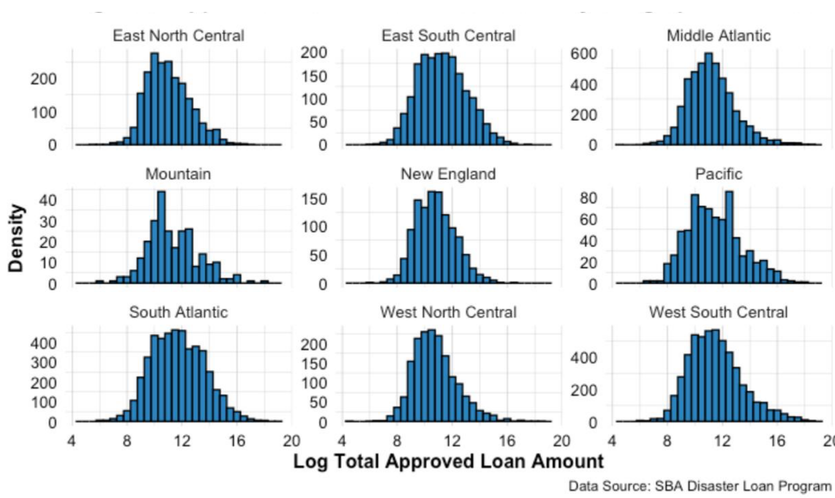
Data Cleaning

Outliers were detected and treated, particularly in columns such as verified losses and approved loan amounts. Extreme outliers were examined to determine whether they reflected valid data or errors, and appropriate adjustments were made.

Data Exploration

Approved Loan Amounts Across Geographical Regions

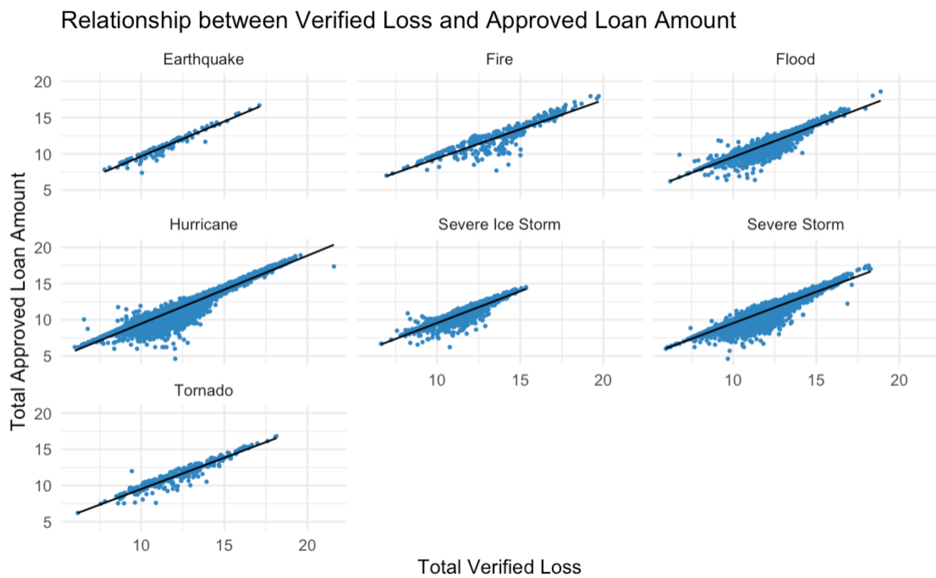
The distribution of approved loan amounts across different regions of the United States provides an insight into the equity and geographic responsiveness of the SBA Disaster Loan Program.



The graphs display the frequency distribution or density of the log total approved loan amounts, providing insights into the patterns and variability of the disaster loan amounts within each region. Some key observations from the graph:

1. The regions exhibit varying distributions, with some showing more concentrated peaks (e.g., New England, Pacific) and others having a more dispersed pattern (e.g., East North Central, South Atlantic).
2. The scale of the loan amounts differs significantly across regions, with the Middle Atlantic region displaying the highest total approved loan amounts compared to the other regions.
3. The shapes of the distributions suggest that the loan amounts follow a skewed or non-normal distribution, with a long right tail indicating the presence of large outlier loan amounts in some regions.
4. Comparing the distributions can reveal potential disparities in the distribution of disaster recovery funding across different geographic areas, which aligns with the project's goal of investigating the equity and effectiveness of the SBA Disaster Loan Program.

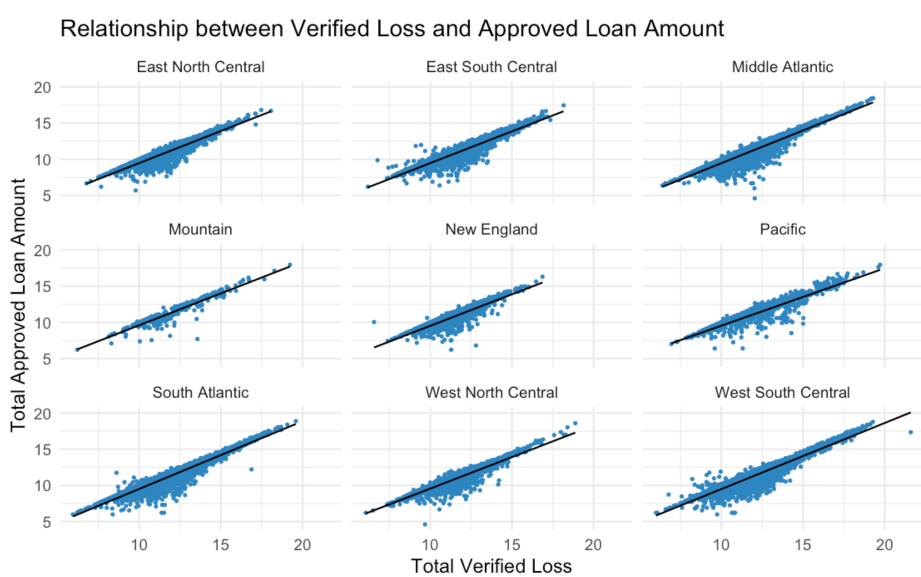
Tri-variate Analysis : Verified Loss vs Approved Loan Amount vs Incident Type



The scatterplots in Image 1 reveal striking differences in the scale and distribution of verified losses and approved loan amounts across different disaster type such as Earthquake, Fire, Flood, Hurricane, Severe Ice Storm, and Tornado. The plots reveal distinct patterns for each disaster type:

1. Earthquake: The data points form a tight, linear relationship, indicating a strong correlation between verified losses and approved loan amounts.
2. Fire: The data points also show a fairly linear relationship, though with more scatter compared to Earthquakes.
3. Flood: The data points exhibit a broader, more dispersed distribution, suggesting a weaker correlation between verified losses and approved loan amounts.
4. Hurricane and Severe Ice Storm: These disaster types show a similar pattern to Flood, with a more scattered distribution and a less pronounced linear relationship.
5. Tornado: The data points are the less scattered, indicating a relatively moderate correlation between verified losses and approved loan amounts.

Tri-variate Analysis : Verified Loss vs Approved Loan Amount vs Geographical Division

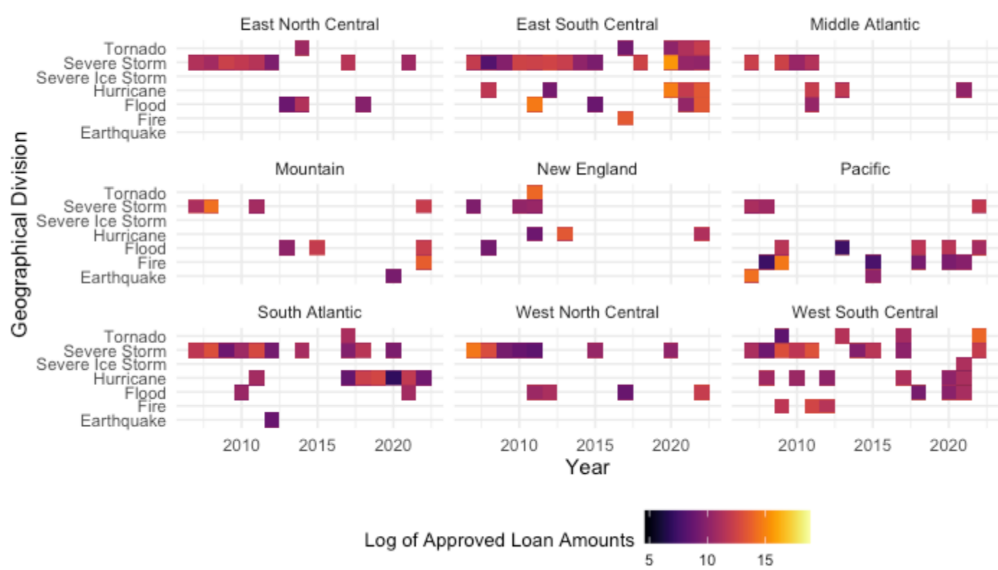


The scatterplots in Image 1 offer a more nuanced view of the relationship between verified losses and approved loan amounts across geographic regions. These visualizations provide valuable insights into the responsiveness and equity of the SBA Disaster Loan Program's allocation of recovery resources.
A few key observations:

1. The data points exhibit a generally linear relationship, indicating that as the log total verified losses increase, the log total approved loan amounts also tend to increase, suggesting the SBA Disaster Loan Program is responsive to the scale of verified losses.
2. The regions that appear to follow a more linear trend are Mountain, Pacific, and West South Central. These regions have a tighter clustering of data points around the linear trend line, suggesting a stronger correlation between the two variables.
3. The correlation between Total Verified Loss and Approved Loan Amount is not perfect, as indicated by the scatter of data points around the linear trend lines. Some regions, such as New England and West North Central, show more variability in the relationship between the two variables.

Overall, the graph demonstrates a generally positive linear relationship between Total Verified Loss and Approved Loan Amount, with some regions exhibiting a stronger correlation than others

Hypervariate Analysis: Approved Loan Amount vs Geographical Division vs Incident Type vs Year



The heatmap provides an analysis of the log of approved loan amounts for different disaster types across various geographic regions and over time. Here are the key insights that can be drawn from the analysis:

Geographical Disparities:

- The heatmap reveals significant disparities in the distribution of approved loan amounts across different geographic regions.
- Certain regions, such as the West South Central and Pacific, tend to have higher approved loan amounts compared to other regions like Mountain and New England.
- This suggests that the SBA Disaster Loan Program may not be equitably allocating resources across all affected regions, potentially leaving some areas underserved.

Disaster Type Influence:

- The heatmap shows that the type of disaster event can influence the approved loan amounts.
- For instance, regions affected by earthquakes and fires generally have higher approved loan amounts compared to those affected by tornadoes and severe storms.
- This indicates that the SBA Disaster Loan Program may respond differently to various disaster types, potentially prioritizing certain types of events over others.

Model Fitting

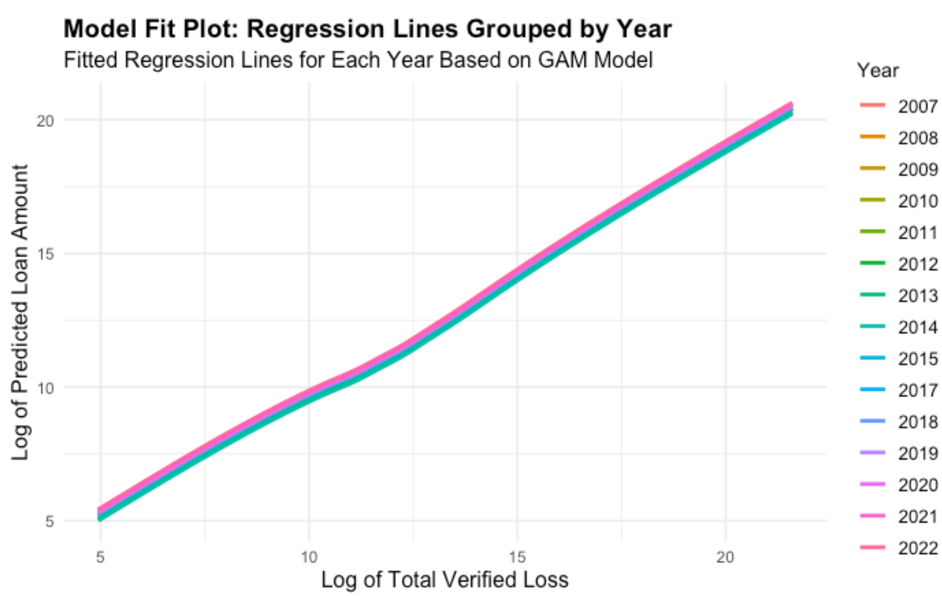
In the graphs above, the data is concentrated in the center, with a tendency toward the lower side, suggesting the potential presence of a non-linear relationship. This observation is supported by the curve visible in the scatterplot of log(Predicted Loan Amount) versus log(Total Verified Loss). Consequently, we select this predictor as the smoothing variable. Additionally, the other variables are included as factors in the model.

The next model we tested was a Mixed Effects model, chosen because our data spans multiple geographical areas and years. To account for unobserved factors such as economic policies and circumstances that may have varied across years, we included year and geographical area as random effects in the model.

The final model we selected was a linear model, with log(Total Verified Loss) as the continuous predictor and Year, Geographical Division, and Incident Type included as categorical factors. We tested all valid interactions across the models, but the results indicated that all models performed best without including interactions.

Model <chr>	AIC <dbl>	BIC <dbl>
GAM	42408.80	42767.19
Linear	43948.84	44247.58
Mixed Effects	44147.18	44454.00

Here GAM is the better model based on BIC and AIC, but when we look at the graph below, we see that there is no significant non-linear effect apart from the group. This is the case for the rest of the factors, too.



For this reason, it would be better not to over-complicate the model and to fit a linear model instead. The linear regression summary with the co-efficient table is attached in the appendix.

The baseline of this models are:
Incident Type: Dam/Levee Break
Year: 2007
Geographical Division: Pacific

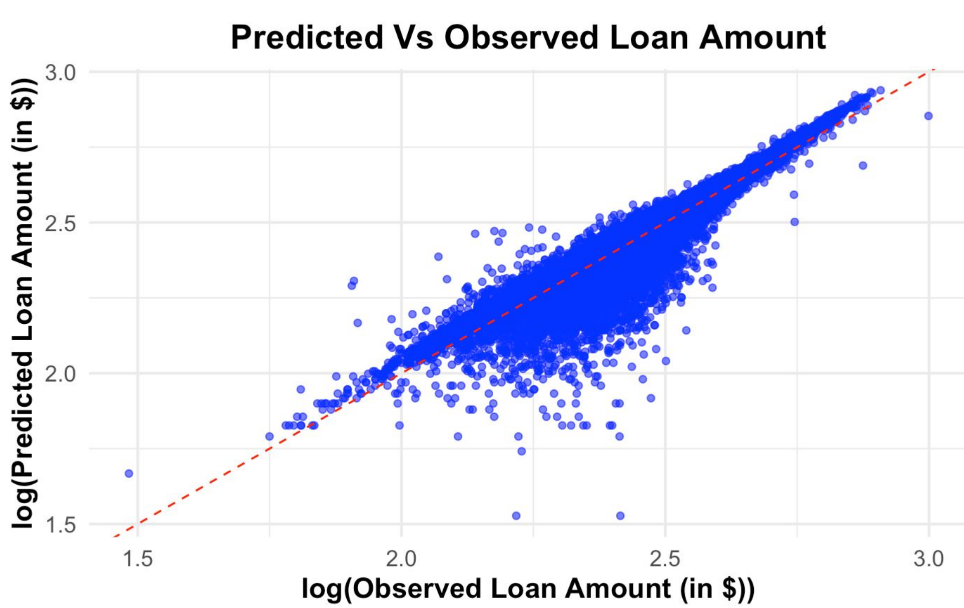
Interpretations:
Intercept: When log(Total Verified Loss) = 0 (i.e. Total Verified Loss = $1), and all the factors are held constant at the baseline, the Total Approved Loan Amount will be exp(0.6118) = $1.844.
Total Verified Loss: Since both the response and continuous are log-transformed, when the Total Verified Loss increases by 1%, the Total Approved Loan Amount will increase by 0.896 percent, while holding all the other variables / factors at baseline level.

Incident Type (Fire): When incident type changes from the baseline (Dam / Levee Break) to Fire, the scale changes by exp(-0.6189) = 0.539, meaning that Fire incidents are approximately granted a 46.1 % lower Total Approved Loan Amount than Dam / Levee Incidents, while holding all other variables at baseline/constant.

Year (2009): When year changes from the baseline (2007) to 2009, the scale changes by exp(-1.493) = 0.225 , meaning that in year 2009, incidents are approximately granted a 77.5 % lower Total Approved Loan Amount than the year 2007, while holding all other variables at baseline/constant.

Geographical Division (Mountain): When geographical division changes from the baseline (Pacific) to Mountain, the scale changes by exp(1.126) = 3.0833 , meaning that incidents in the Mountain geographical division are approximately granted a 208 % higher Total Approved Loan Amount than the Pacific geographical division, while holding all other variables at baseline/constant.

These are the interpretations for one level change from the baseline for the factors.



The model is a good fit for lower values, but as the loan amount increases, the values become more spread out. Still, it is a much better fit than any of our other observed models. There are a few extreme outliers, but they are few in a lot.

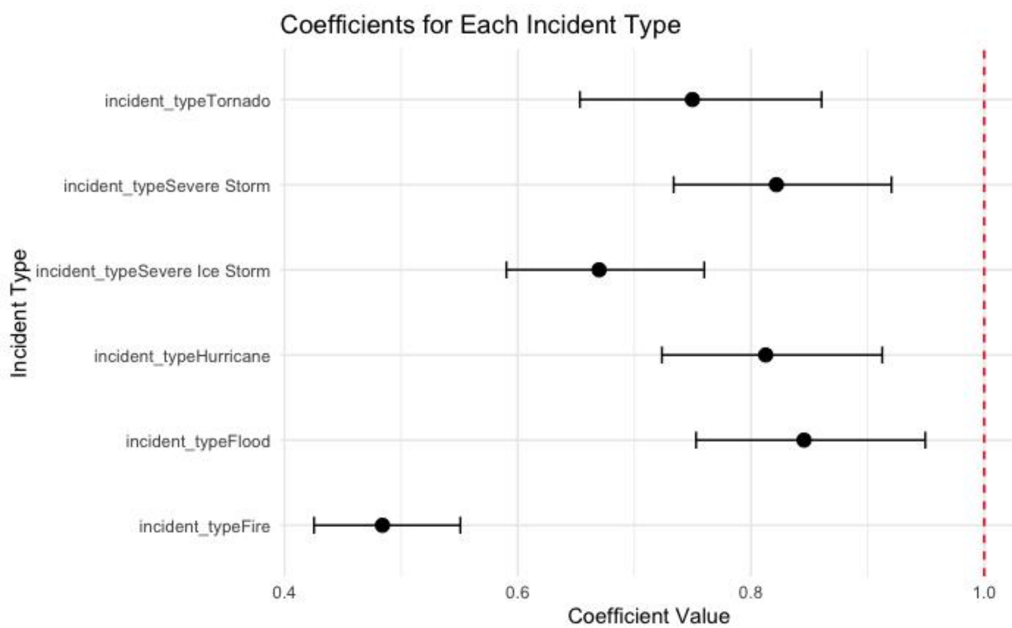
Model Results

By focusing on the correlation coefficient, the analysis is able to directly address the core research questions, uncover meaningful patterns, and identify areas for potential program enhancements - all of which are crucial for improving the effectiveness and equity of the SBA Disaster Loan Program in supporting communities affected by disasters.

The model results provide valuable insights into the factors influencing the log of total approved loan amount in the SBA Disaster Loan Program. The direction and magnitude of the coefficients, along with their associated confidence intervals, offer a nuanced understanding of these dynamics.

Baseline Reference (Red Vertical Line at 1):

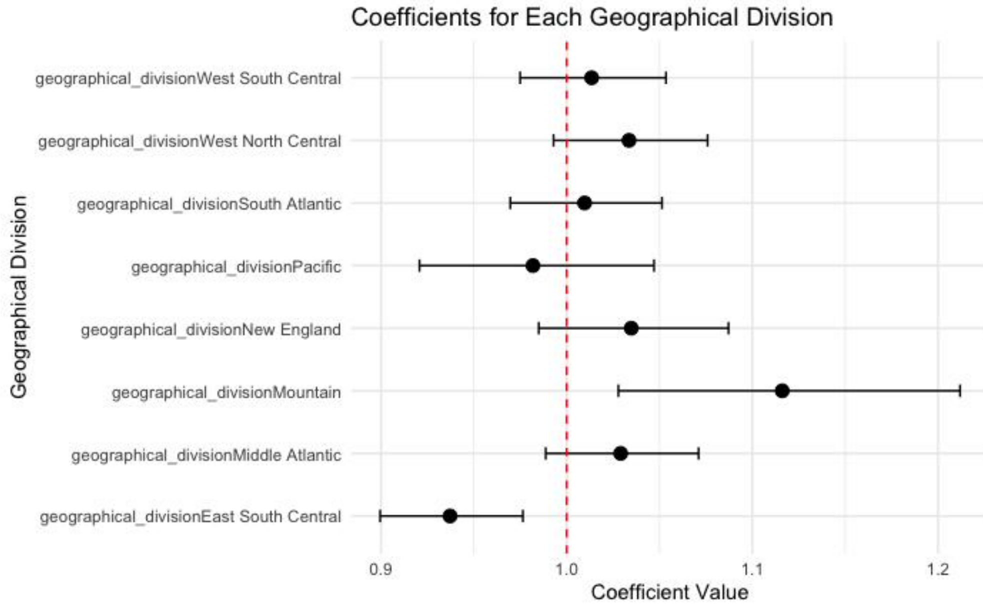
A coefficient value of 1 indicates no effect relative to the baseline.
Coefficients > 1 indicate a positive effect, meaning loans increased compared to the baseline.
Coefficients < 1 indicate a negative effect, meaning loans decreased compared to the baseline.



Baseline is Earthquake

- Tornado has a coefficient around 0.7, indicating total approved loan amounts are 70% of earthquake amounts.
- Severe Ice Storm has a coefficient of approximately 0.6, meaning approved loan amounts are 60% of those for earthquakes.
- Hurricane, Flood and Severe Storm shows a coefficient of roughly 0.8, indicating approved loan amounts at 80% of earthquake levels.
- Fire has the lowest coefficient at about 0.4, meaning approved loan amounts are only 40% of earthquake amounts.

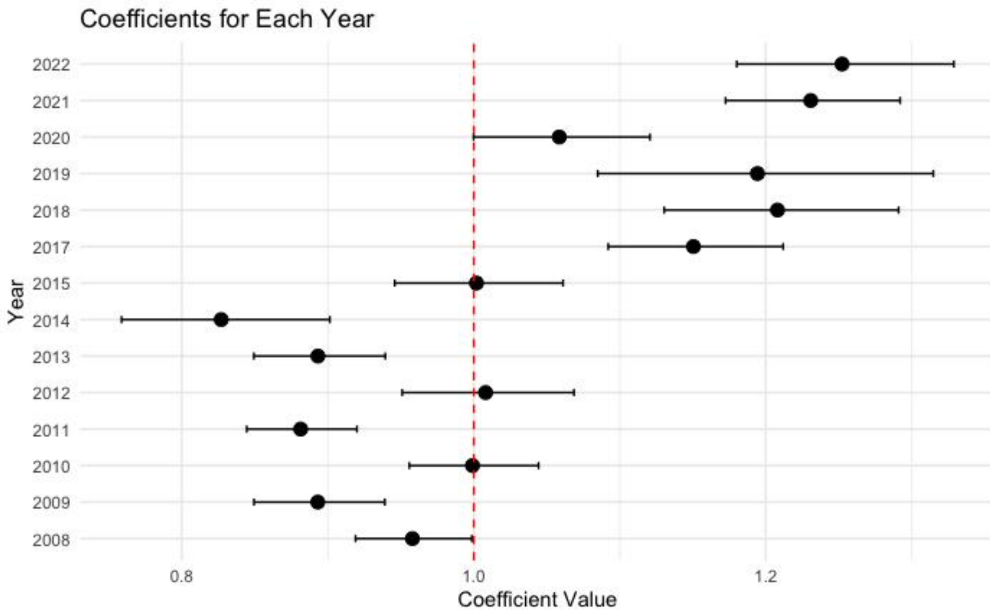
- All incident types show coefficients less than 1.0 (the red dashed line representing earthquakes), confirming that earthquakes consistently result in higher approved loan amounts than all other disaster types.



Baseline is East North Central

- Mountain Region shows the strongest positive effect, with a coefficient approximately 1.15, indicating about 15% higher effect compared to East North Central. The confidence interval for this region is notably wide, suggesting less precision in the estimate.
- New England and West North Central both show positive effects of approximately 1.05, representing a 5% increase compared to the baseline. Their confidence intervals do not cross the reference line, indicating statistical significance.
- Middle Atlantic and South Atlantic display similar modest positive effects around 1.02-1.03, though their confidence intervals slightly overlap with the baseline.
- East South Central shows the most negative effect, with a coefficient of approximately 0.92, representing an 8% decrease compared to the baseline. This is the only region showing a clearly negative and statistically significant difference.
- Pacific region shows a neutral to slightly negative effect (coefficient ≈ 0.98) but has a relatively wide confidence interval that crosses the baseline.

Baseline Overlap: Pacific, Middle Atlantic, and South Atlantic regions have confidence intervals that overlap with the reference line at 1.0



Baseline = Year 2007

The plot visualizes exponentiated coefficients for the variable Year, where the values represent the multiplicative effect of the year on the predicted log total approved loan amount (exponentiated to return to the original scale).

The overall trend shows a general increase in coefficient values from 2008 to 2022, with some fluctuations:

- 2008-2015: Coefficients are mostly below or close to the baseline (1.0), indicating similar or slightly lower effects compared to 2007.
- 2017-2022: A clear shift occurs, with coefficients consistently above the baseline, suggesting stronger effects in recent years.

Specific Observations

- Lowest Point: The year 2014 shows the lowest coefficient, approximately 0.85, indicating about a 15% decrease in effect compared to 2007.
- Highest Point: 2022 has the highest coefficient, around 1.25, suggesting a 25% increase in effect compared to the baseline.
- Transition Year: 2015 appears to be a transition point, with its coefficient very close to 1.0, marking a shift from generally negative to positive effects.

Precision of Estimates

- Most Precise: Years 2013-2015 have the narrowest confidence intervals, indicating more precise estimates.
- Least Precise: 2019 has the widest confidence interval, suggesting less certainty in its estimate.
- Magnitude of Effects

The largest positive effect is observed in 2022, with an approximate 25% increase over the baseline.
The largest negative effect is in 2014, with an approximate 15% decrease from the baseline.

This analysis reveals a clear trend of increasing coefficients over time, with a notable shift occurring around 2015-2017, and the most recent years showing the strongest positive effects relative to 2007.

Conclusion, Limitations and Future Work

To address the central research question on the effectiveness of the SBA Disaster Loan Program, this analysis examines the topic through several key sub-questions. The goal is to provide a comprehensive understanding of the factors that influence loan approval rates, the role of disaster type, and the distribution of funding across different geographic regions.

Answer to the research questions

Effectiveness of the SBA Disaster Loan Program

The effectiveness of the Loan Program is significantly dependent of the drivers in the loan approval process as well as the disaster type and geographical division. To answer this question, we have analyzed each aspect individual in relation to the loan approved amounts.

Drivers of Variation in Loan Approval Rates

The Small Business Administration (SBA) Disaster Loan Program exhibits complex variations in loan approval amount across different geographical and temporal dimensions. Geographical factors emerge as a primary driver of variation, with significant disparities observed across regional boundaries. Regions such as the Middle Atlantic demonstrate consistently higher approved loan amounts compared to regions like Mountain and New England, suggesting an uneven distribution of disaster recovery resources.

Temporal analysis reveals another critical dimension of variation. The program's loan approval response has shown marked differences especially after the year 2020, which may be linked towards the policy changes after Covid, displaying more pronounced patterns of loan allocation. These temporal variations indicate that the SBA Disaster Loan Program's approach to funding has not remained static but has evolved and responded differently to disaster scenarios over time.

Disaster Type Impact on Loan Amounts

The impact of disaster types on approved loan amounts reveals significant nuances in the SBA Disaster Loan Program's response. Earthquakes and fires consistently receive higher approved loan amounts, indicating a potentially more robust support mechanism for these specific disaster types. In contrast, tornadoes and severe storms receive comparatively lower loan approvals, suggesting a more limited recovery support for these event types. Floods and hurricanes present a more complex picture, with a more scattered distribution of loan approvals relative to losses. Tornadoes represent the most challenging category, showing the weakest correlation between verified losses and approved amounts.

Quantitative insights further illuminate these variations. Earthquakes and floods consistently show the highest verified losses, while tornadoes report the lowest. Fires, hurricanes, and severe ice storms occupy an intermediate position, indicating a nuanced approach to disaster type assessment in loan approval processes.

Urban and Rural Loan Distribution

The distribution of approved loan amounts across urban and rural regions reveals significant disparities that raise important questions about the equity of disaster recovery funding. Regions with stronger urban characteristics demonstrate a clear advantage in loan allocation, with urban-centric areas like the East North Central and Middle Atlantic regions receiving higher predicted loan amounts relative to their verified losses.

Limitations

While the analysis provides valuable insights, it is important to note several limitations that should be considered when interpreting the findings. Firstly, the dataset, despite its breadth, may not capture the full complexity of disaster recovery dynamics, as it is primarily focused on home loans issued to individuals. The inclusion of commercial and community-level loans could shed additional light on the program's broader impact.

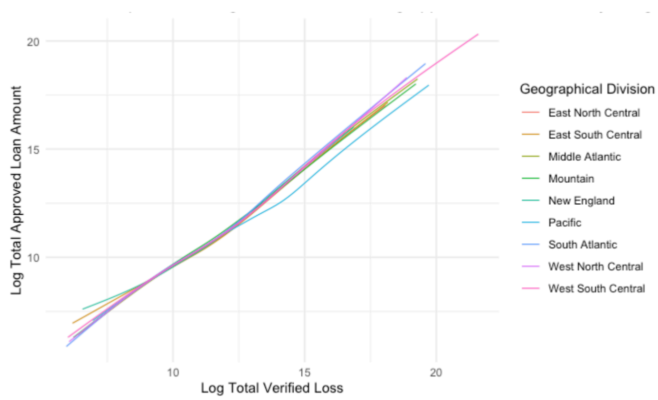
Moreover, the analysis relied on aggregated regional and temporal data, which may obscure nuances at the local or individual level. Incorporating more granular geographic information and individual-level characteristics could yield a richer understanding of the factors shaping loan approval patterns.

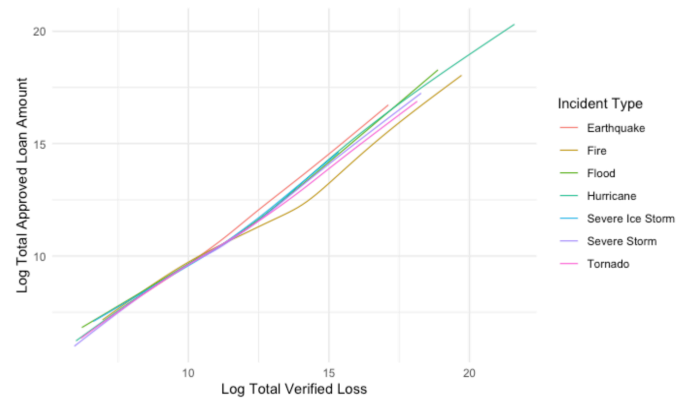
Finally, the study's scope was limited to the SBA Disaster Loan Program, and a comprehensive evaluation of disaster recovery efforts would benefit from considering the role of other federal, state, and local initiatives, as well as private sector and community-based support mechanisms.

Future Work

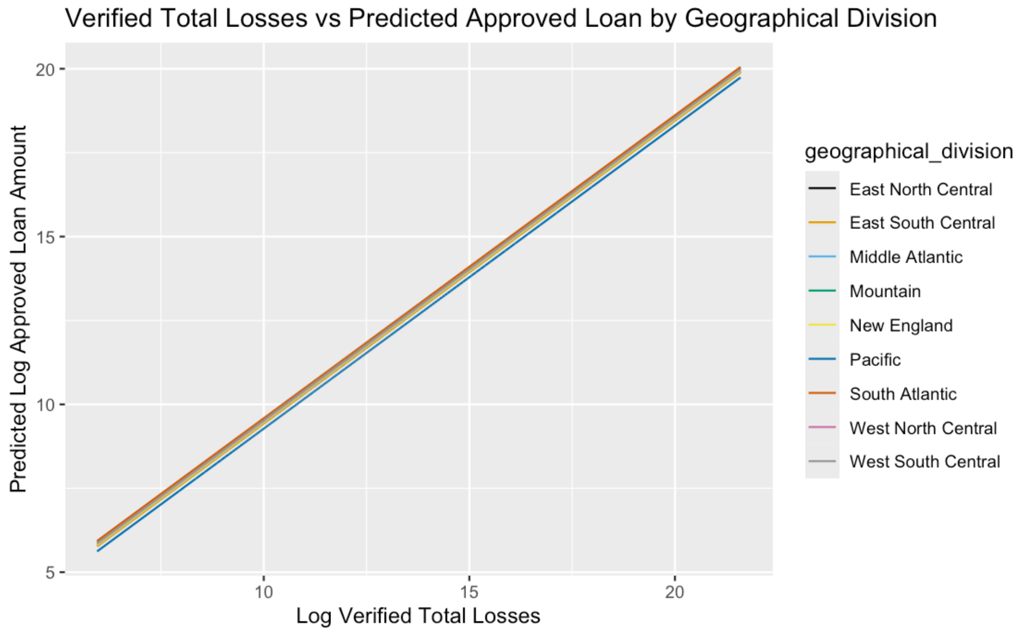
Building upon the insights gained from this research, several avenues for future exploration and program enhancements emerge. Expanding the analytical lens to include a wider range of disaster recovery resources and support programs could provide a more holistic understanding of the ecosystem in which the SBA Disaster Loan Program operates. Additionally, incorporating qualitative data and stakeholder perspectives could help contextualize the quantitative findings and identify opportunities for improving program design and implementation.

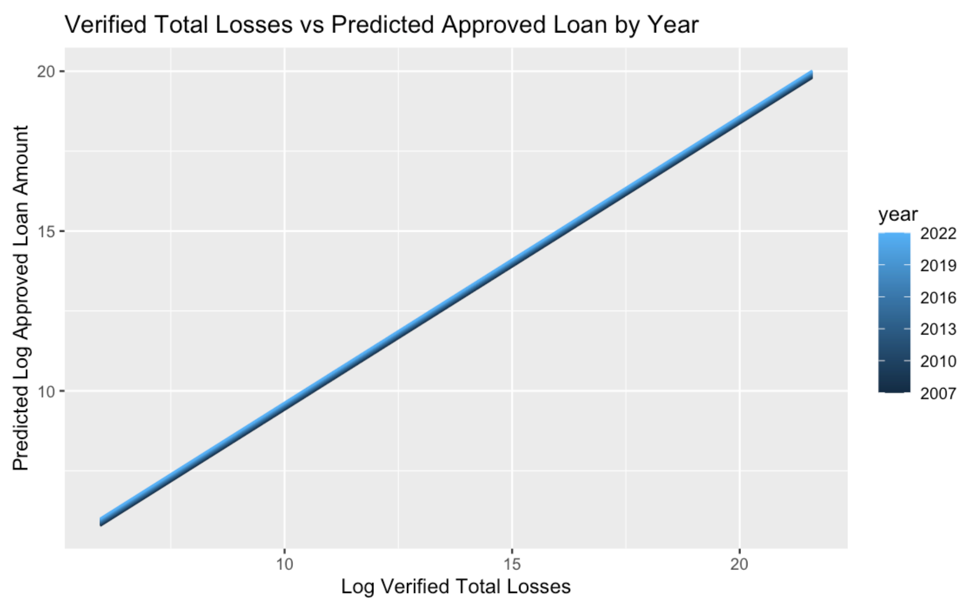
Appendix





While analyzing the performance of the model, we experimented with fitting a prediction grid based on the model, however the results were not very conclusive and since the lines were merging together, we decided to plot a correlation grid instead to identify more nuanced patterns.


Verified Total Losses vs Predicted Approved Loan by Geographical Division
Predicted Log Approved Loan Amount
Log Verified Total Losses
geographical_division
East North Central
East South Central
Middle Atlantic
Mountain
New England
Pacific
South Atlantic
West North Central
West South Central


Verified Total Losses vs Predicted Approved Loan by Year
Predicted Log Approved Loan Amount
Log Verified Total Losses
year

```
Call:
lm(formula = log(total_approved_loan_amount) ~ log(total_verified_loss) +
    incident_type + geographical_division + as.factor(year),
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6295 -0.2131  0.1680  0.3901  3.3681

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                             6.118e-01  1.336e-01   4.581 4.65e-06 ***
log(total_verified_loss)                8.962e-01  2.189e-03 409.370  < 2e-16 ***
incident_typeEarthquake                -9.542e-02  1.323e-01  -0.721 0.470831
incident_typeFire                      -6.189e-01  1.328e-01  -4.662 3.15e-06 ***
incident_typeFlood                     -3.235e-02  1.280e-01  -0.253 0.800439
incident_typeHurricane                 -6.046e-02  1.280e-01  -0.472 0.636679
incident_typeMud/Landslide              1.583e-01  6.258e-01   0.253 0.800313
incident_typeOther                     -7.476e-01  3.314e-01  -2.255 0.024112 *
incident_typeSevere Ice Storm          -2.480e-01  1.307e-01  -1.897 0.057803 .
incident_typeSevere Storm              -4.416e-02  1.276e-01  -0.346 0.729179
incident_typeTornado                   -1.437e-01  1.332e-01  -1.079 0.280623
incident_typeTyphoon                   -1.497e-01  1.774e-01  -0.844 0.398774
incident_typeVolcanic Eruption          1.040e-01  2.826e-01   0.368 0.712909
geographical_divisionEast North Central -3.035e-02  2.750e-02  -1.104 0.269784
geographical_divisionEast South Central -1.016e-01  2.675e-02  -3.798 0.000146 ***
geographical_divisionMiddle Atlantic   -3.628e-03  2.532e-02  -0.143 0.886071
geographical_divisionMountain           1.126e-01  4.499e-02   2.502 0.012342 *
geographical_divisionNew England       -7.215e-06  2.957e-02   0.000 0.999805
geographical_divisionPacific           -1.810e-02  3.648e-02  -0.496 0.619779
geographical_divisionSouth Atlantic    -3.593e-02  2.150e-02  -1.672 0.094622 .
geographical_divisionWest North Central -1.073e-03  2.722e-02  -0.039 0.968568
geographical_divisionWest South Central -2.758e-02  2.404e-02  -1.147 0.251377
as.factor(year)2008                    -4.212e-02  2.109e-02  -1.997 0.045830 *
as.factor(year)2009                    -1.493e-01  2.479e-02  -6.023 1.74e-09 ***
as.factor(year)2010                    -2.185e-03  2.255e-02  -0.097 0.922807
as.factor(year)2011                    -1.272e-01  2.173e-02  -5.853 4.89e-09 ***
as.factor(year)2012                     3.151e-02  2.950e-02   1.068 0.285400
as.factor(year)2013                    -1.103e-01  2.557e-02  -4.315 1.60e-05 ***
as.factor(year)2014                    -1.840e-01  4.384e-02  -4.197 2.71e-05 ***
as.factor(year)2015                     5.570e-03  2.914e-02   0.191 0.848421
as.factor(year)2017                     1.649e-01  2.604e-02   6.332 2.46e-10 ***
as.factor(year)2018                     2.029e-01  3.359e-02   6.041 1.56e-09 ***
as.factor(year)2019                     2.126e-01  4.809e-02   4.420 9.90e-06 ***
as.factor(year)2020                     8.943e-02  2.840e-02   3.149 0.001639 **
as.factor(year)2021                     2.114e-01  2.465e-02   8.578  < 2e-16 ***
as.factor(year)2022                     2.381e-01  2.935e-02   8.115 5.11e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6105 on 23686 degrees of freedom
Multiple R-squared:  0.8903,    Adjusted R-squared:  0.8901
F-statistic:  5490 on 35 and 23686 DF,  p-value: < 2.2e-16
```

Linear Model Summary