An Exploration of the Impact of Disaster Aid on Unemployment

Christopher Ng, Jake Thompson, Mikhaela Anderson, Nick Kartschoke, Ryan Lippe

Dev10 Capstone Project: Data Cohort 20

May 24, 2022

Introduction

After some initial exploratory searches into financial datasets, we came across the Small Business Administration (SBA) disaster loans. It initially sparked our interest and after some thought, we found that we were curious about the data's relationship to unemployment. Our original hypothesis was that we would find that as the total damages done by disasters increased, we would find that unemployment would increase. Realizing that unemployment is a complicated topic we knew we needed more data to support our analysis and ultimately decided to include the S&P 500 index, and unemployment and local government finance data from the U.S. Census. These additional datasets were chosen as we considered them to have strong relationships with unemployment. With these datasets, our final hypothesis is that unemployment can be predicted by a combination of disaster loan data, local government finances, and market data. We will analyze data between 2008 and 2015 on a county level.

Originally, we knew we needed a regression model and considered a few options including LASSO with cross validation and multi-linear regression models but ultimately decided on Elastic Net. We chose Elastic Net because we wanted to avoid over-fitting our model, which LASSO can be prone to. Elastic Net takes penalties from both LASSO and Ridge regression methods and performs variable selection and regularization simultaneously. We will see which factors were the best predictors of unemployment by looking at coefficient weights of the finished model. In addition to predicting unemployment, our analysis will be aimed at answering these questions:

1. Are disaster loans a significant predictor of unemployment?
2. Does the amount of the loan affect mitigating unemployment?
3. Which disasters saw the most loan money awarded?
4. Were local government revenues correlated with increased unemployment rate?
5. What states received the most disaster loans?
6. Is the S&P 500 a significant predictor for unemployment?

Exploratory Data Analysis

Once our datasets were combined as described in the ETL report, we ran descriptive statistics and a correlation matrix to begin our data exploration as seen below in **Figure 1** and **Figure 2**.

	Approved Amount Content	Approved Amount EIDL	Approved Amount Real Estate	Damaged Property Zip Code	Fines and Forfeits	Interest Revenue	Market Value	Property Tax	Total Salaries & Wages	Unemployement	Verified Loss Content	Verified Loss Real Estate	Year
count	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908
unique	461	451	700	1171	184	247	8	277	271	96	1216	1644	8
top	0.0	0.0	0.0	8008.0	0	977	1646.49	99949	202936	7.5	0.0	0.0	2013
freq	1167	1255	1009	33	231	152	752	152	152	157	477	258	752

Figure 1.

	Approved Amount Content	Approved Amount EIDL	Approved Amount Real Estate	Damaged Property Zip Code	Fines and Forfeits	Interest Revenue	Market Value	Property Tax	Total Salaries & Wages	Verified Loss Content	Verified Loss Real Estate	Year	Unemployment
Approved Amount Content	1.000000	0.416740	0.659065	-0.017202	-0.016806	-0.015025	0.039631	-0.009469	-0.014500	0.583050	0.649920	-0.041560	-0.048577
Approved Amount EIDL	0.416740	1.000000	0.359718	0.039691	-0.012043	-0.003591	-0.057778	-0.016077	-0.023734	0.223797	0.287088	-0.077939	-0.116359
Approved Amount Real Estate	0.659065	0.359718	1.000000	0.027232	-0.041942	-0.033125	0.040617	-0.033432	-0.040234	0.417810	0.633326	-0.066324	-0.066842
Damaged Property Zip Code	-0.017202	0.039691	0.027232	1.000000	0.136964	0.051875	-0.060827	0.047237	0.017746	0.007310	0.032540	-0.285273	-0.127382
Fines and Forfeits	-0.016806	-0.012043	-0.041942	0.136964	1.000000	0.659474	0.090892	0.723044	0.696462	-0.012186	-0.025699	-0.065824	-0.160826
Interest Revenue	-0.015025	-0.003591	-0.033125	0.051875	0.659474	1.000000	0.003518	0.589251	0.543271	-0.015774	-0.016960	-0.151293	-0.142499
Market Value	0.039631	-0.057778	0.040617	-0.060827	0.090892	0.003518	1.000000	0.092355	0.128464	0.025304	0.038948	0.628099	-0.113223
Property Tax	-0.009469	-0.016077	-0.033432	0.047237	0.723044	0.589251	0.092355	1.000000	0.954393	-0.019972	-0.022222	-0.004724	-0.045300
Total Salaries & Wages	-0.014500	-0.023734	-0.040234	0.017746	0.696462	0.543271	0.128464	0.954393	1.000000	-0.020373	-0.021845	0.036101	-0.034038
Verified Loss Content	0.583050	0.223797	0.417810	0.007310	-0.012186	-0.015774	0.025304	-0.019972	-0.020373	1.000000	0.923536	-0.061022	-0.068573
Verified Loss Real Estate	0.649920	0.287088	0.633326	0.032540	-0.025699	-0.016960	0.038948	-0.022222	-0.021845	0.923536	1.000000	-0.055770	-0.061441
Year	-0.041560	-0.077939	-0.066324	-0.285273	-0.065824	-0.151293	0.628099	-0.004724	0.036101	-0.061022	-0.055770	1.000000	0.364116
Unemployment	-0.048577	-0.116359	-0.066842	-0.127382	-0.160826	-0.142499	-0.113223	-0.045300	-0.034038	-0.068573	-0.061441	0.364116	1.000000

Figure 2.

One of the datasets we considered was census data on local government finances from 2008-2015. This included data at a county level and had over 500 factors, and one of our early findings was that this data had many nulls because of inconsistent reporting. We were collecting data with the goal of feeding it into a machine learning model to predict unemployment, so we ultimately would need to filter or impute all null values for our numeric columns. With this in mind, we selected the columns “Total Salaries & Wages”, “Property Tax”, “Interest Revenue”, and “Fines and Forfeits” to maximize the amount of data we would have to work with after filtering nulls. There were many other columns that had a notable correlation with unemployment, but when we filtered nulls all of the data was removed. “Fines and Forfeits” and “Interest Revenue” had correlations with unemployment of -0.161 and -0.143, respectively. These values were high relative to other features we were considering. “Property Tax” and “Total Salaries & Wages” both had low correlations, but we thought they might have predictive value as more general indicators of how an economy is doing.

We also looked at data on SBA disaster loans at the county level. When we combined this data with the census data and our other sources, there were 3556 rows removed because of counties that had census data, but never applied for an SBA loan. We were interested in seeing how strong of a predictor of unemployment SBA confirmed damages and loans were, so we only considered counties that had both census data and an SBA loan at some point. Looking at that intersection left us with 1908 rows and 20 columns. We also saw that there were a couple disasters which had many more loans associated with them than the rest, seen in **Figure 3**, and we looked into these outlier disasters for our analysis.

```
NJ-00033    507
NJ-00023    184
NY-00130    138
IL-00025     84
IL-00041     77
           ...
NC-00013      1
PA-00026      1
IN-00052      1
SC-00006      1
NC-00061      1
Name: SBA Disaster Number, Length: 135, dtype: int64
```

Figure 3.

When exploring other datasets we could use to predict unemployment, we saw that S&P 500 market value had a correlation with unemployment of -0.113 and decided to include it as a feature. Despite not having the highest correlation, it ended up being the strongest predictor.

Machine Learning Model

For our machine learning model we used an Elastic Net and applied it to our filtered dataset from our SQL server, which eliminated all zero entries for the disaster number column. For our predictors we used only the numeric columns, except for “Unemployment”, which was our target. We then created our training and testing sets, with 75% being for training and 25% being for testing. We created our first model with the default parameters for Elastic Net and applied it to our dataset. We then found the R^2 score and MSE to obtain a baseline that will be used for comparisons with our tuned models. We then tried to adjust the hyper-parameters “alphas” and “l1_ratio” to increase the performance of the model. For the second model we tried to expand the ranges of the alpha and l1_ratio values, then did a gridsearchCV to see which combination of those two parameters gave the best results. For our third model we tried expanding the alpha value options by using a logspace to obtain more values between 0.1 and 2. While the third model performed better than the previous models, we then applied a logspace to the l1_ratio as well and created our fourth model. The fourth model performed the best so far with a R^2 score of 0.4028 and a MSE of 1.989.

Another thing we checked was if removing a column would lead to a better performing model. We found the descriptive statistics for the numeric predictors for general information for each of them. Then, we ran a correlation matrix to see the relationships between each numerical predictor and the “Unemployment” value. From the correlation matrix we concluded that the “Total Salaries & Wages'' variable had the lowest correlation. So we then created a new Elastic Net model on the dataset, dropping the “Total Salaries & Wages'' column, with log spaces for both the alpha and l1_ratio parameters. This model had a MSE of 1.987 and a R^2 of 0.3711. When comparing this model to the fourth model, we can conclude that the fourth model had a higher R^2 and slightly higher MSE. Even though the fourth model is more complex, we believe that it is the more optimal model for predicting employment.

Results

1. Are disaster loans a significant predictor of unemployment?

We saw that disaster loans were overall very weak predictors of unemployment, both in general and relative to other factors we investigated. Our model minimized the weights of these factors, but did not improve when they were removed. We also saw that the approved amount of economic injury disaster loans (EIDL) was the disaster loan factor with the strongest correlation to unemployment, with a correlation of -0.116. This was interesting, because the negative correlation implies that places which received greater amounts in loans saw slightly lower unemployment.

2. Does the amount of the loan affect mitigating unemployment?

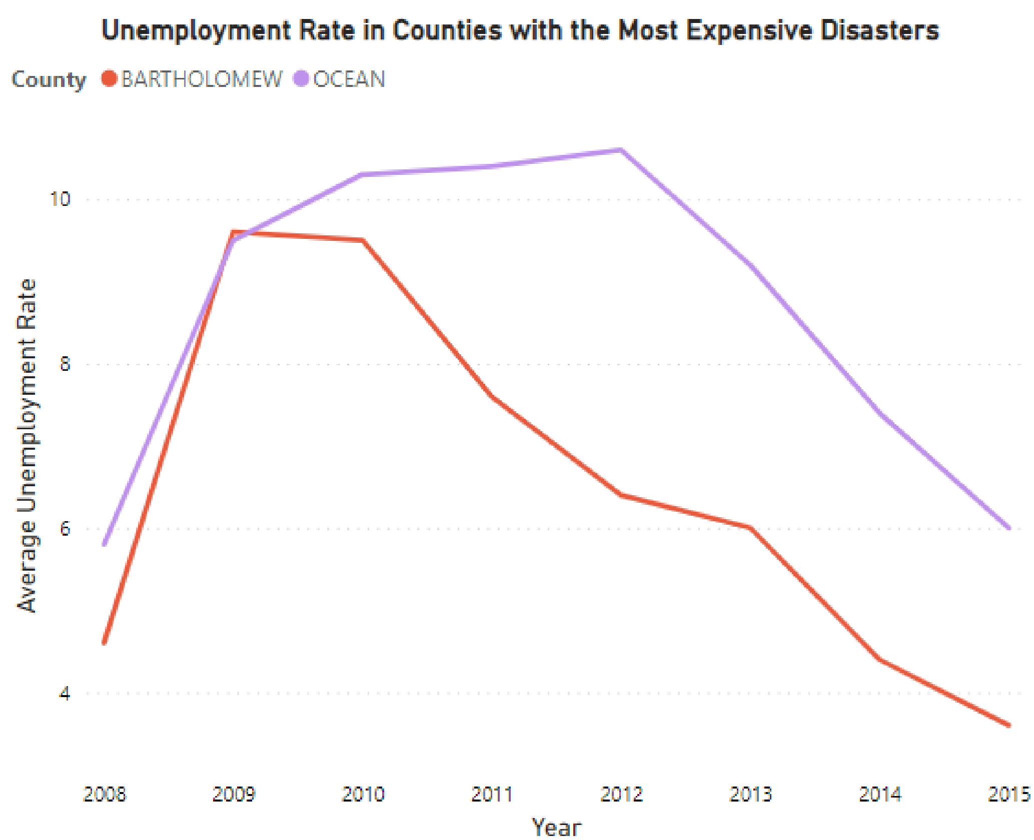

Figure 4.

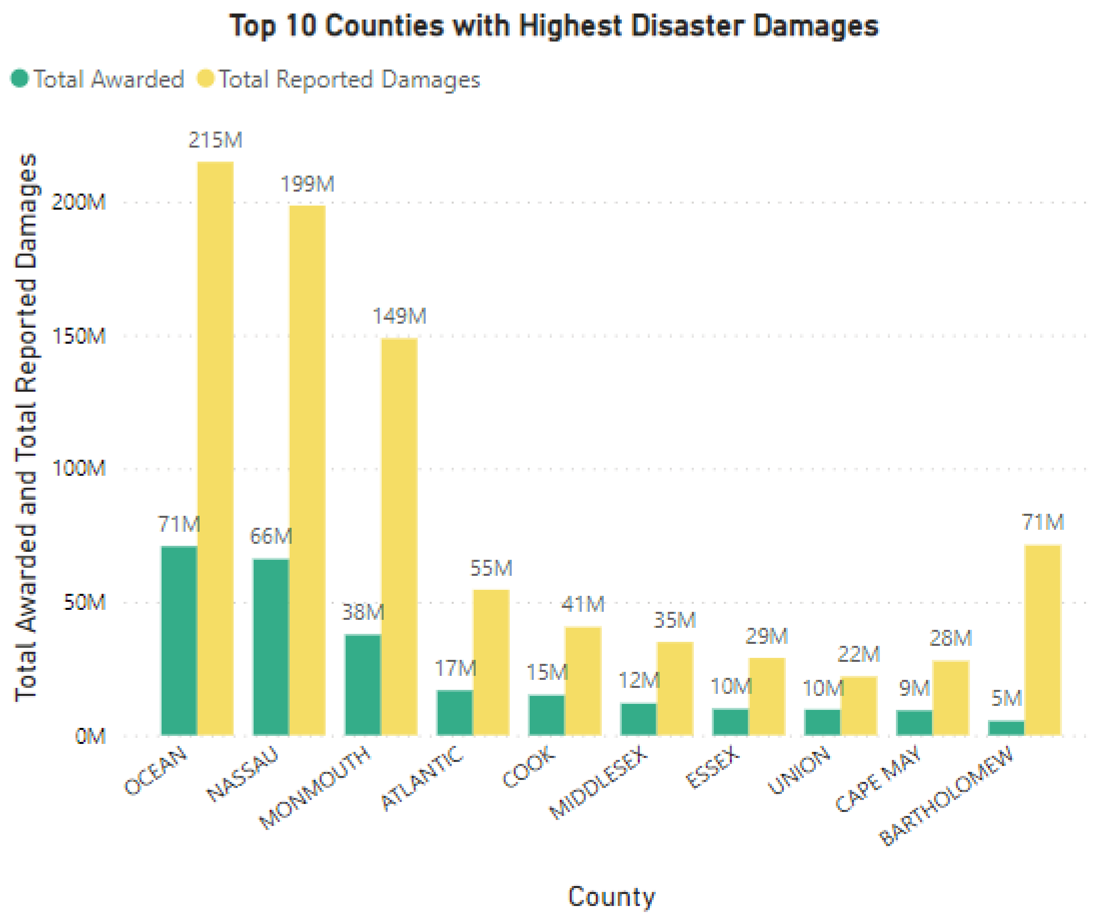

Figure 5.

It appears that the amount of disaster loans doesn't significantly mitigate unemployment. In the cases of Ocean and Bartholomew County, we see that despite reporting much higher awarded loans and a higher proportion of damages awarded as loans (**Figure 5**), Ocean County saw higher unemployment (**Figure 4**). Looking at the counties with the highest disaster loan awards, unemployment rates vary widely, but they all appear to move with a similar trend despite very different award amounts. This doesn't necessarily rule out a causal relationship between loan amounts and unemployment rates, but our analysis did not find any evidence for one. The disaster loan factor which our model weighed the most heavily was the approved amount of EIDL, but this was still a very weak predictor relative to other factors.

3. Which disasters saw the most loan money awarded?

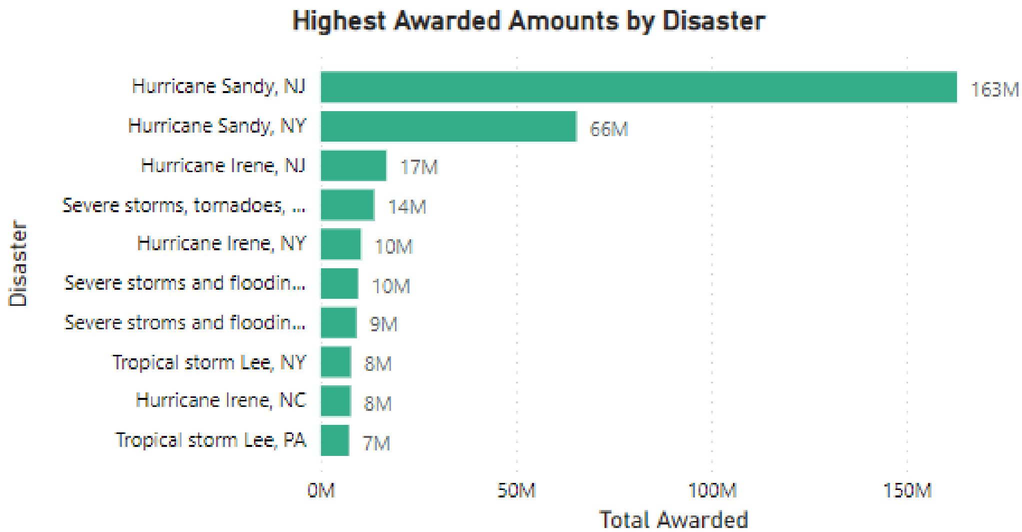

Figure 6.

By far the most destructive single disaster in terms of loans awarded was Hurricane Sandy, which caused $229 million in disaster loans just in New York and New Jersey (**Figure 6**) Hurricane Irene was the next largest cause of disaster loans. In general, severe storms and flooding was the disaster category which incurred the most loans.

4. Were local government revenues correlated with increased unemployment rate?

We saw that local government revenues, specifically from fines and forfeits as well as revenue from interest on debts held by local governments, were both weighed more heavily by our model than any SBA disaster loan factors. Revenue from fines and forfeits was our second best predictor of unemployment overall, though the relationship was still weak. Fines and forfeits had a correlation with unemployment of -0.161, and interest revenue had a correlation of -0.143. This implies that governments which saw greater revenue from these sources saw slightly lower unemployment.

5. What states received the most disaster loans?

The top 3 recipients of disaster loans between 2008-2015 were New Jersey, with $180,667,100, New York, with $88,670,300, and Illinois, with $17,681,200. New Jersey and New York received significantly more loans than any other state, and we saw that they were both affected by Hurricane Sandy (**Figure 6**) which was the most destructive single disaster we found.

6. Is the S&P 500 a significant predictor for unemployment?

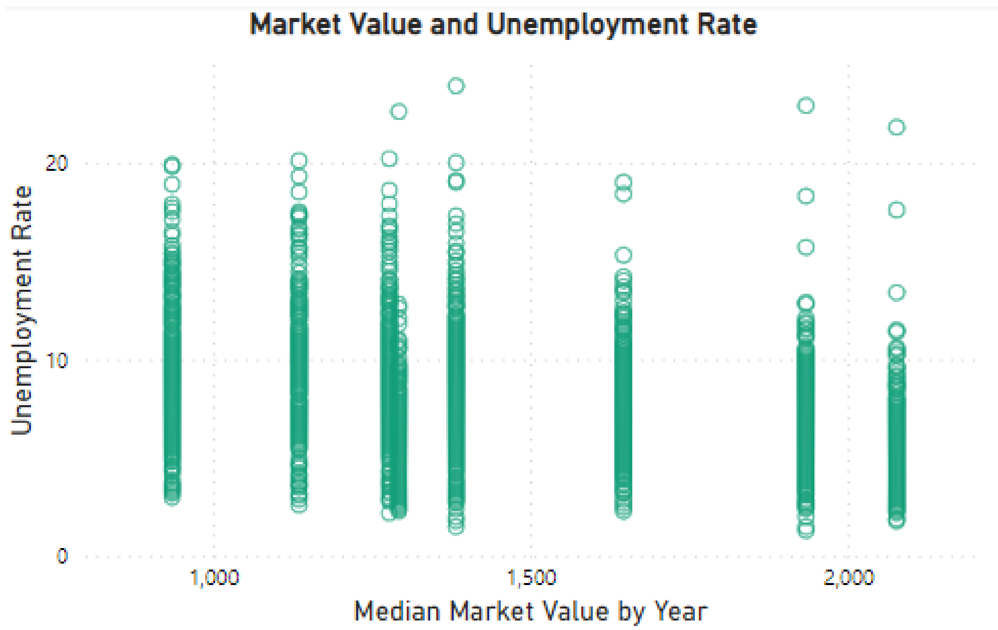

Figure 7.

Market value from the S&P 500 (**Figure 7**) was the strongest predictor of unemployment from all the features we included in our model by a wide margin, with a coefficient weight over 10x greater than government revenue from fines and forfeits, which was our next strongest predictor. Interestingly, it was our strongest predictor despite having a fairly weak correlation with unemployment of -0.1132. Fines and forfeits, interest revenue, and approved amount of EIDL all had stronger correlations despite being weaker predictors.

Conclusion

We successfully made an Elastic Net model to predict unemployment, and our final R^2 score was 0.4028. We saw that the factors we chose were overall weak predictors of unemployment, with S&P 500 market value being the strongest predictor, and with local government revenue from fines and forfeits being the next strongest predictor. Unemployment is complex and is influenced by many factors, so for additional progress we recommend trying to improve the model by considering more features. The scope of our research included correlation and using features as predictors, but for further analysis to investigate causation, we recommend the features be compared with unemployment in later years. For example, we could consider whether disaster loans received impacted the next year's unemployment rate, or unemployment over the next 5 years.

We were interested in seeing whether disaster loans were a significant predictor of current unemployment, and we saw that they were not. Further research is needed to see whether disaster loans are predictors of future unemployment as mentioned above. We also researched whether the amount of loans affected unemployment, and we saw that unemployment appeared insensitive to the amount of disaster loans.

Next, we looked into which disasters saw the most loan money awarded, and we found that the most destructive disasters from 2008-2015 in terms of disaster loans were Hurricane Sandy and Hurricane Irene. In general, severe storms and flooding were the largest category of disasters.

We then considered whether local government revenues correlated with unemployment. We saw that government revenue from fines and forfeits was the second strongest predictor of unemployment overall, and had a weak negative correlation with unemployment of -0.161.

The next question we looked into was which states received the most in disaster loans, and we saw that the top recipients were New Jersey, New York, and Illinois. Each of these states saw significant loans due to storms, with the top 2 being caused by Hurricane Sandy.

The final question we researched was whether the S&P 500 was a significant predictor of unemployment. We saw that it was our strongest predictor by a wide margin despite being less correlated with unemployment than several other features we considered. For future research, we recommend considering more market factors as predictors of unemployment due to this being much better than other features we included.

Bibliography

Ravaliya, Jay. *US Unemployment Rate by County*, 1990-2016. https://www.kaggle.com/datasets/jayrav13/unemployment-by-county-us

U.S. Small Business Administration. *Disaster Loan Data.* 2008-2015. https://data.sba.gov/dataset/disaster-loan-data

United States Census Bureau. *State & Local Government Finance Historical Datasets and Tables.* https://www.census.gov/programs-surveys/gov-finances/data/datasets.html.

Willden, Chase. *Sp500.* Accessed 2022. https://data.world/chasewillden/stock-market-from-a-high-level

World Population Review. *List of State Abbreviations.* Accessed 2022. https://worldpopulationreview.com/states/state-abbreviations