Predicting the costs of natural disasters

- By Mike Stegmuller

Problem Statement:

Natural disasters cause various amounts of death, damage, and destruction throughout the country and can happen at any time. Predicting how much a natural disaster will cost next time they hit is crucial to response times effective management of the destruction. The problems that I am looking to explore are how much different natural disasters cost to repair damage, the number of deaths they cause and the effect the cost of natural disasters have on inflation rate. Looking at natural disasters and their costs from 1980 to 2023, I will try to predict how much a disaster will cost so the US can be better prepared.

Data Wrangling:

I used 3 datasets about US billion dollar Weather and climate Disaster from 1980 to present data from the government. The datasets contain U.S. disaster cost assessments of the total, direct losses ($) inflicted by: tropical cyclones, inland floods, drought & heat waves, severe local storms (i.e., tornado, hail, straight-line wind damage), wildfires, crop freeze events and winter storms. The datasets also show the number of deaths associated with each disaster. I also used inflation datasets to see if there was a correlation between inflation and the disaster costs.

Data Cleaning

1. The first part of data wrangling after I read the datasets was to merge the state disaster dataset with the state cost data set. I also looked at the mean, max, and standard deviation of the cost of the disasters.

2. Created data frame with the number of deaths per disaster

3. Grouped the dates by year and created a data frame based on the year and disaster

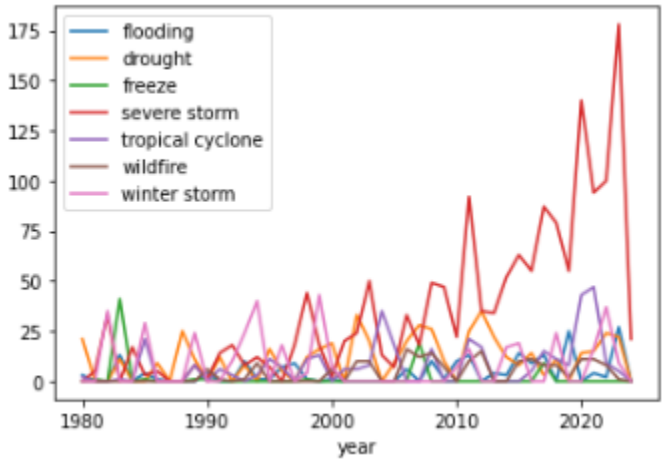
Exploratory Data Analysis

The data consists of 7 different types of disasters: Drought, Flooding, Freeze, Severe Storm, Tropical Cyclone, Wildfire and Winter Storm.

The data I wanted to explore first was to see which type of disaster has cost the most per event. Tropical cyclones have cost the US the most at almost $1.4 billion dollars.

Severe Storms have been on the rise. These are local storms that do alot of damage. Due to severe storms being on the rise they have the highest total cumulative costs.

I also wanted to see which disaster caused the most deaths. Tropical cyclones and droughts cause the most deaths of all the disasters.



Droughts

Droughts cost the second most of all the disasters. Droughts last the most days compared to the other disasters, which is why I hypothesize they are the second most costly disaster. Kansas and Texas have suffered the most droughts since 1980. Texas has endured the most costs due to droughts followed by Kansas and North Dakota.

Flooding

The amount of floods per year vary, but more are trending in recent years. Louisiana has had the most in the US. About half of Louisiana is below sea level, and has many areas prone to flooding. Iowa, a state along the Mississippi river, has had the most costly damage due to floods. Most of the states in the Mississippi river basin have issues with flooding except California. California seems like a dry temperate climate, but when it does rain it causes a lot of damage which is costly, especially in southern California, where it rains less often.

Freezes

The amount of damage caused by freezes cost the least amount of money compared to the other disasters. The states that have the most number of freezes and damage costs due to freezing are states with warmer climates that are not prepared for freezes. States that are prepared for freezes have the least amount of damage costs. Florida and California, specifically

southern California,, known for their sunny days and lower temperature climate are examples of states with infrastructure not prepared to handle freezes. When the rare freezes hit these areas they do a lot of damage.

Severe Storms

Severe storms are local storms that cause various and excessive amounts of damage. They cost the most amount of damage. Texas leads the US with the most in total severe storms and severe storm costs.

Tropical Cyclones

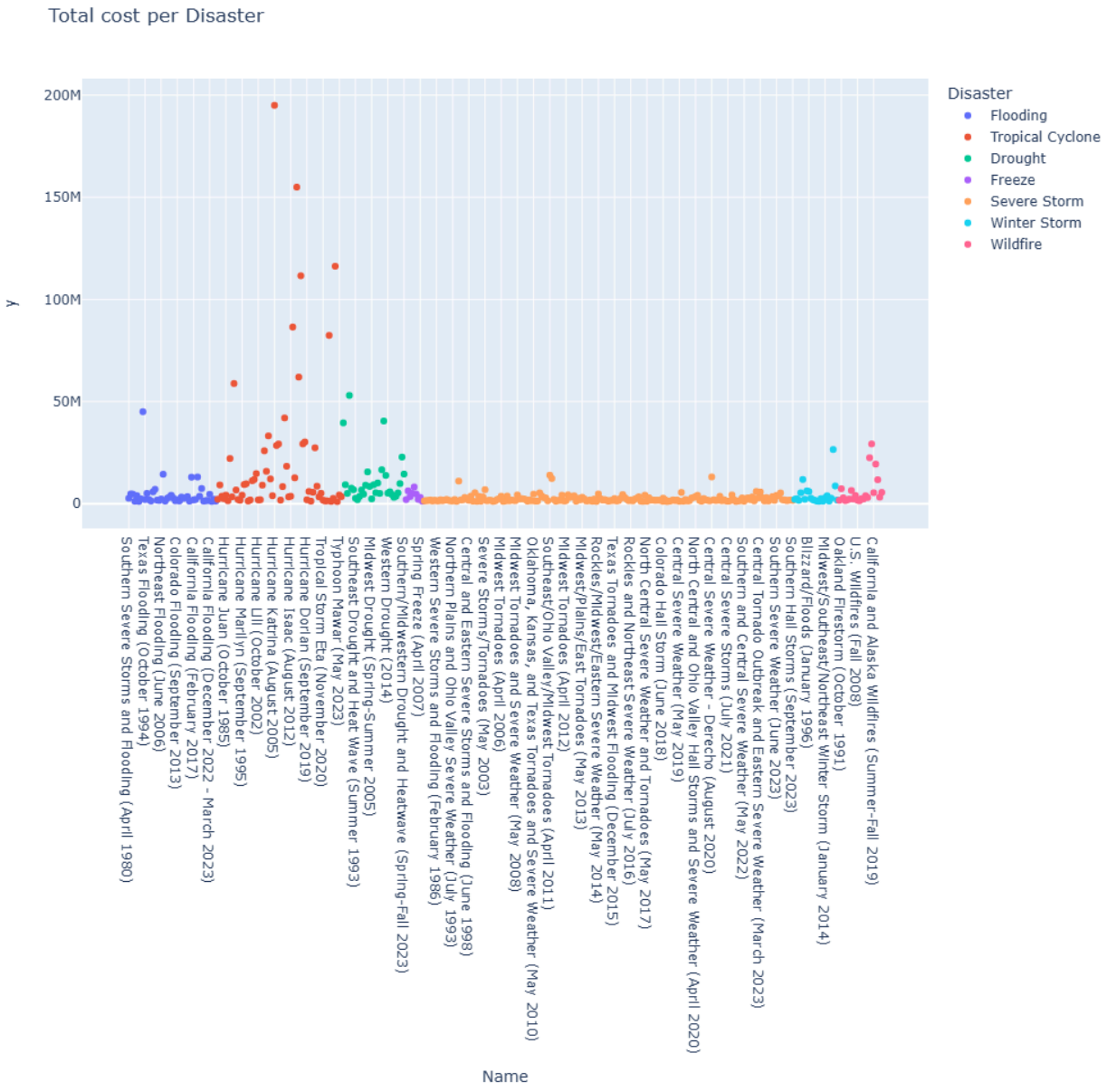
Tropical cyclones cause the most amount of damage per event. The more days a tropical cyclone lasts the more expensive the damage costs will be. Florida leads the US in both total tropical cyclones and most damage costs caused by a tropical cyclone. The states along the Atlantic seaboard and gulf of Mexico have the most damage from tropical cyclones.

Wildfires

Million dollar wildfire disasters happen the least often, but incur a lot of damage costs when they do occur. States on the west coast have the most problems with wildfires due to the dry climates and vast empty forest areas. California has had the most wildfires and has had the most damage costs.

Winter Storms

Winter storms vary per year but are trending downward as the years go on. They tend to affect the eastern half of the US more than the western half. This could be due to the higher populations and more infrastructure available for the storms to damage. Texas has the most damage costs due to winter storms. This may be due to a combination of unpreparedness for when winter storms occur, but also a high population with lots of infrastructure.



Analysis

What I noticed about the data was that there was a pattern between the number of days a disaster lasts and how much a disaster will cost and the amount of deaths a disaster will cause. Tropical cyclones have the highest number of deaths and costs per day. The number of days a disaster last will be a feature in model selection.

Model Selection

Since I was trying to predict a price, I decided to use different types of regression models. I chose linear regression, decision tree regression, and random forest regression as the machine learning models to attempt to predict the price of natural disasters. For my feature selection, I decided on the disaster type, deaths, number of days and the inflation rate value to determine the cost a disaster would cost. I also looked at the mean absolute error.

Linear Regression

The linear regression model didn't have a good model score for the test data. Since the model score didn't score very high, I decided to use the cross value score, best of 20 scores. While there were some good scores, most were low or negative. The highest cross value score was good with .87, so there were some irritations where the data had a good predictive performance. The training data did not indicate a good predictive performance as well.

Decision Tree Regression

The decision tree regressor was slightly better than the linear regression. The model score was better for the training data when compared to the test data. Once again I used the cross value scores. There were higher scores when using the cross value scores, but still had many negative and low scores.

Random Forest Regression

The random forest regressor score had a positive value, though it was low. The cross values scores had mostly positive scores. The training data did have a good model score and had mostly positive scores close to 1%. Based upon the results the random forest regression model is the best model to predict the cost of disasters.

Prediction/Takeaways

While the random forest was the best machine learning model the test data did not score high with a model score of only .29. The training data did score higher with .92 and the cross value scores were mostly positive. What this tells me is that there may not be enough features to consistently make an accurate prediction. I decided to add population to the model. I used csv data with the population per year. When comparing the scores to the data without the

population added, the mean and min cross value scores were still negative but closer to 0. The max cross value score and model accuracy score were slightly lower.

Future Data Research

For future research I would think about adding more information to the data to compare, such as what specific types of infrastructure were damaged. When adding population into the data it did not change much. Knowing the type of infrastructure that was damaged and how much it would cost to fix is needed. For example, maybe one severe storm does a lot of damage to infrastructure with cheap easy fixes as opposed to a severe storm that destroys a vital bridge or major building.

Ideas for Future Research

1. Research the best ways to rebuild infrastructure so it can survive the next disaster. When rebuilding infrastructure, make the new infrastructure more resistant to the disaster that destroyed the infrastructure in the first place. For example, when rebuilding in an area that is prone to flooding, make new buildings out of material that is waterproof, or have ditches, levies, dams, or drainage areas that only fill up when there is a flood so the water has a place to go.

2. I would research what can be done to prevent as many deaths as possible in all situations by using the money to better the urgent response teams. Researching what factors in the disasters that cause deaths should be done. If an area needs flood boats during a flood or vehicles needed to traverse heavy snow, they should be in place.

3. I would try to see if there are any patterns in the weather systems that they notice that lead up to a disaster. If there are any “super storms” or once in a lifetime storms that could be on the way and what leads up to them. Researching the patterns will better help with preparation. For example, we know the oceanic and air temperatures that contribute to tropical cyclones forming and monitoring that. Once the temperatures reach the tropical cyclone forming ‘danger zone’ there should be a preparation system in place already preparing for possible damage. Or maybe knowing ‘X’ amount of dry, hot days could possibly trigger a wildfire and prepare for a possible fire before it starts. Another example would be using the amount of rain per second in a storm to determine if an area will flood and how soon in order to evacuate people if necessary.