INFO-656 Machine Learning

Can proximity to Natural Disasters predict the home insurance cost of a residential area?

Nishant Jain

Overview

The purpose of this project is to predict the Home Insurance costs of different residential areas that lie within different disaster zones.

Purpose of the research.

Impact of Natural Disasters

Natural disasters – such as hurricanes, cyclones, earthquakes, mudslides, floods, wildfires, volcanic eruptions – are likely increasing in frequency due to climate change. These events bring with them a host of issues, including humanitarian, public health, environmental and infrastructural problems.

As a result the areas that are more prone to natural disasters incur a higher cost of insurance. Is that Home Insurance Price justified or can the home insurance prices be predicted based on the counts of these disasters?
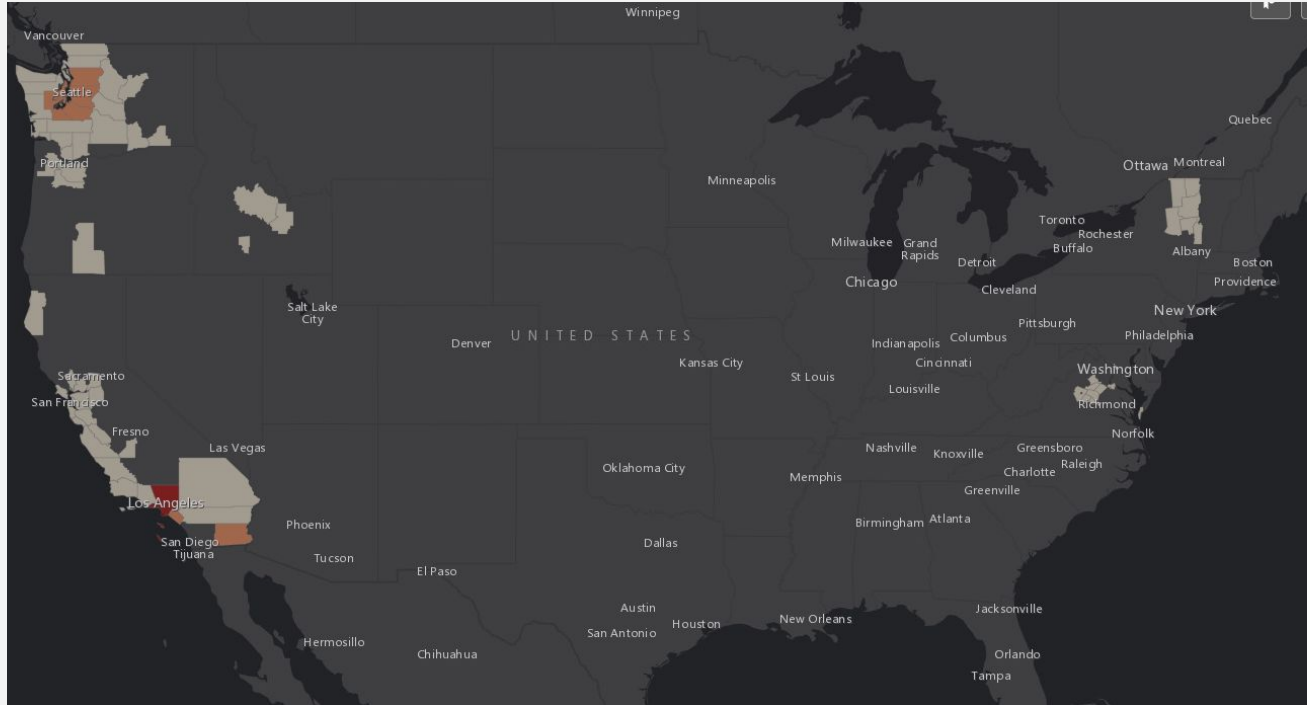
Understanding the different layers
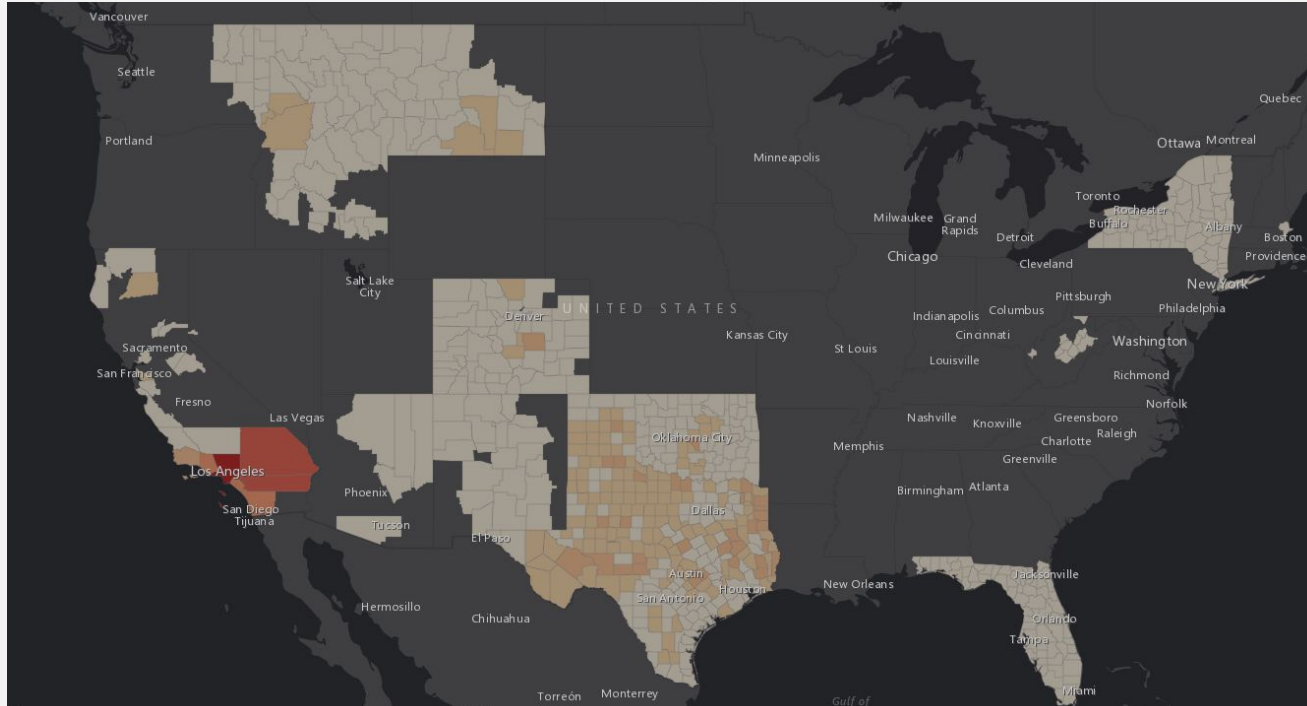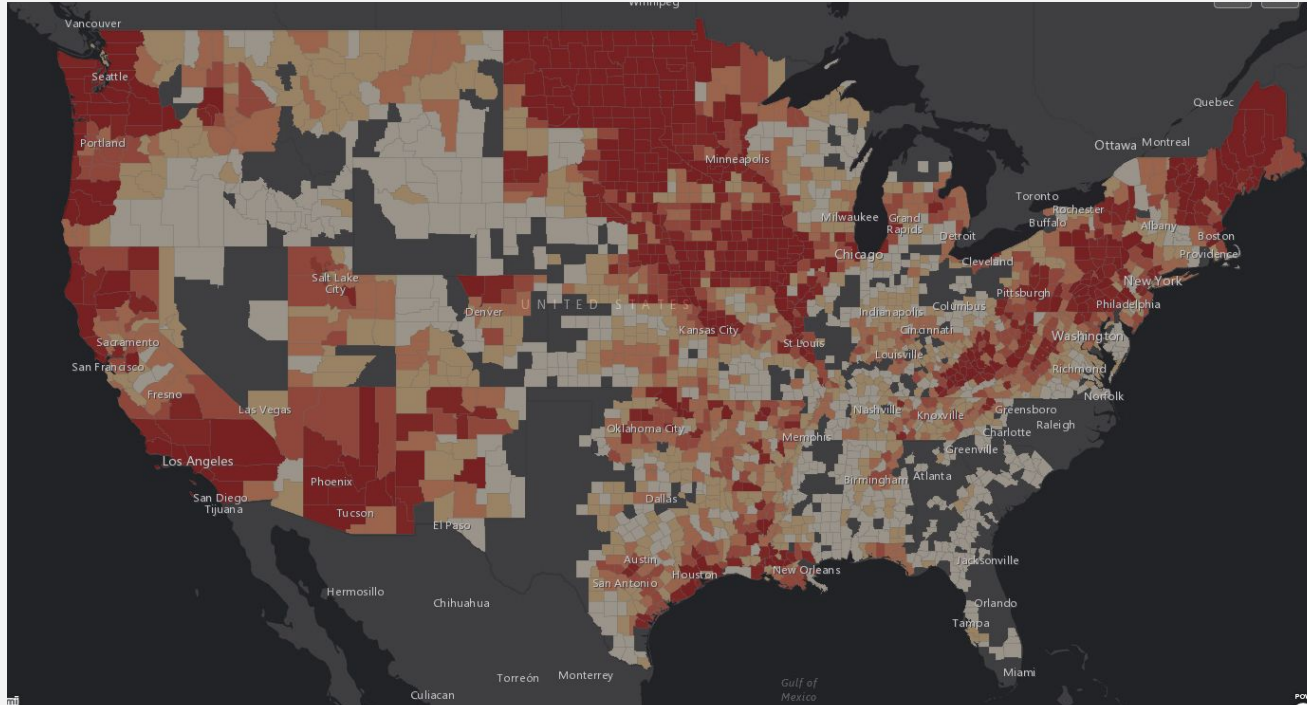
Disaster Layers
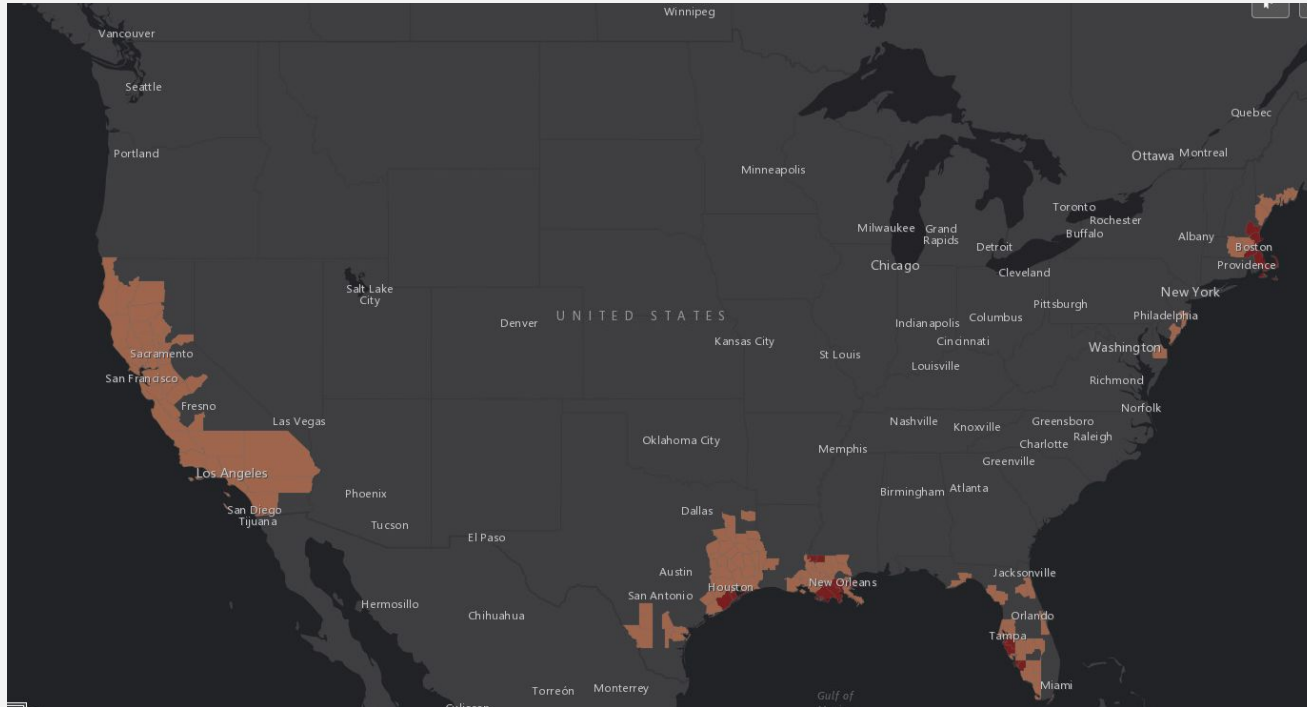
Coastal Storms.

Disaster Layers

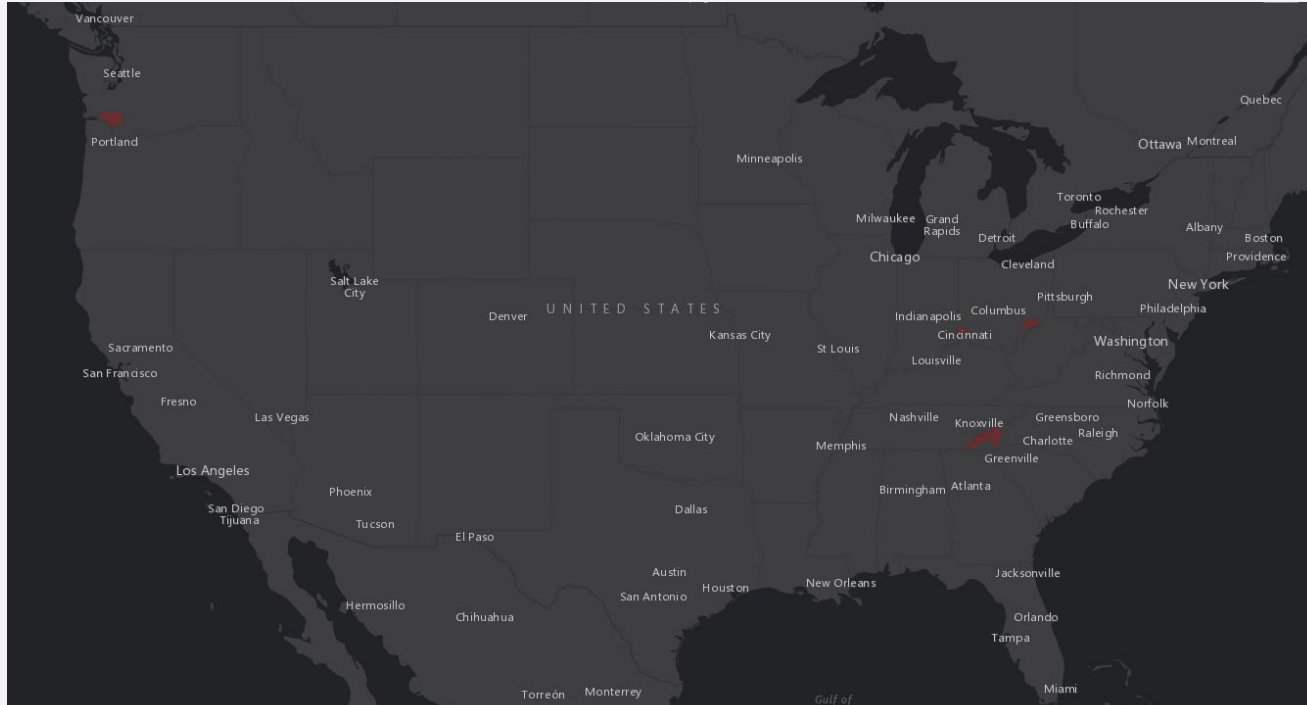
EarthQuakes

Disaster Layers

Fires

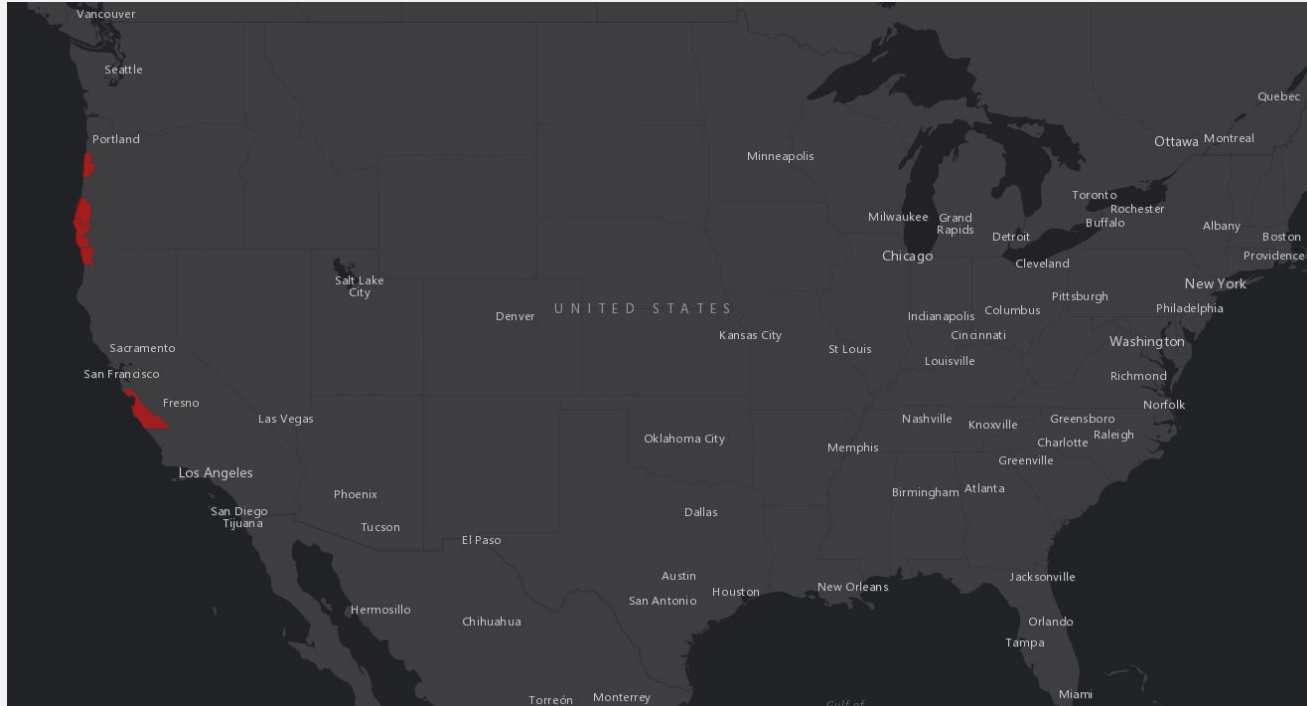# Disaster Layers

Floods

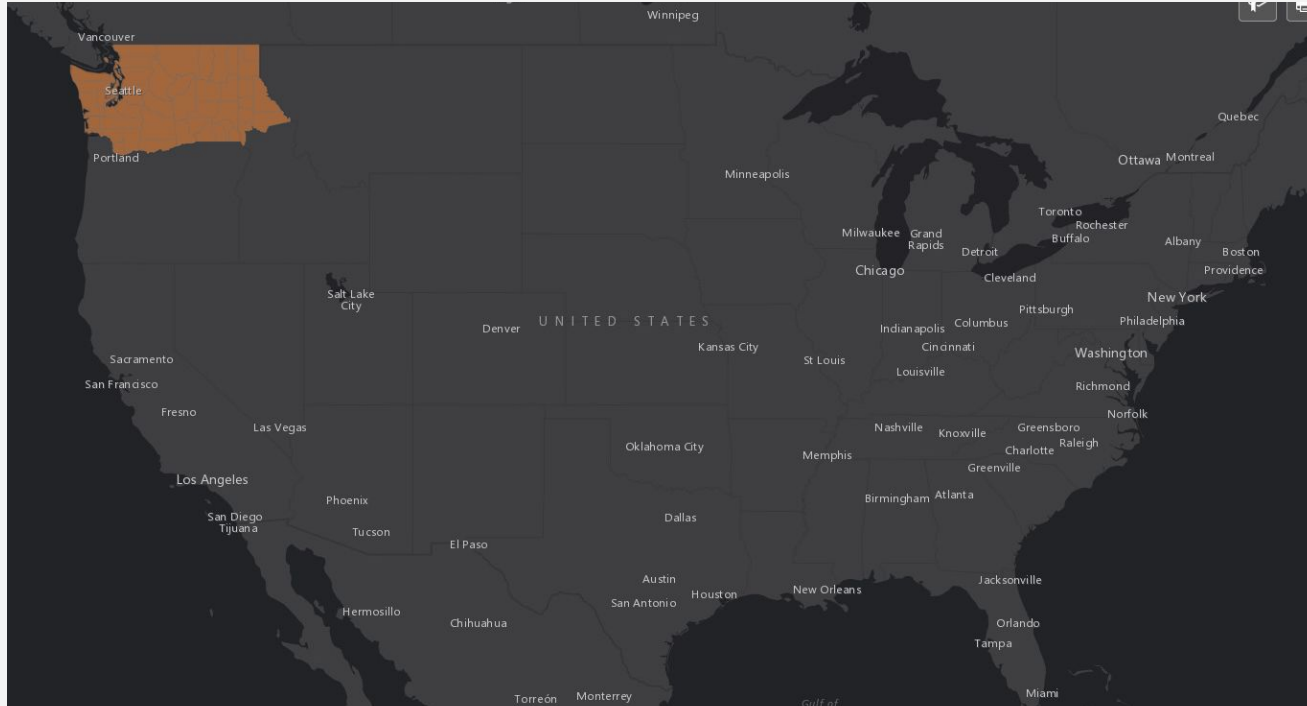# Disaster Layers

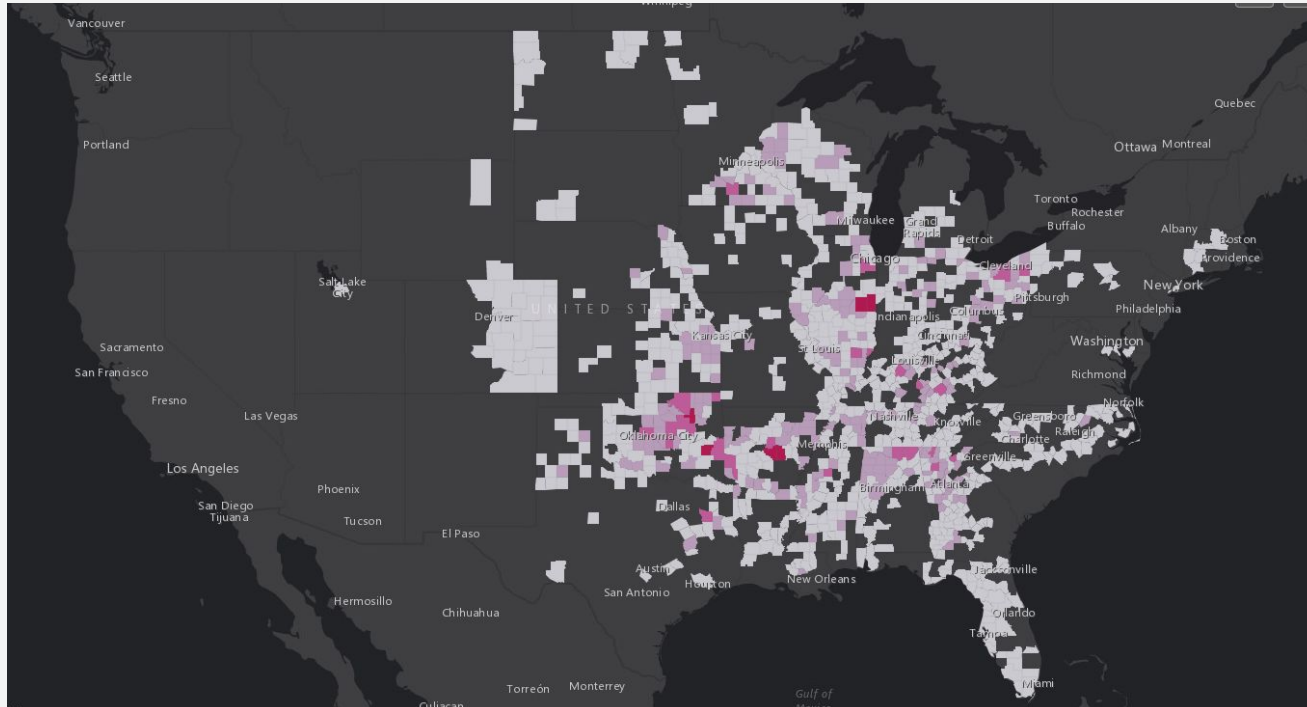Hurricanes

Disaster Layers

Landslides

Disaster Layers

Tsunamis

Disaster Layers

Volcanos

Disaster Layers

Tornadoes

Data for Home Insurance Costs

Average statewide premium

$600 $800 $1000 $1200 $1400 $1600 $1800 $2000

Source: Insurance Information Institute
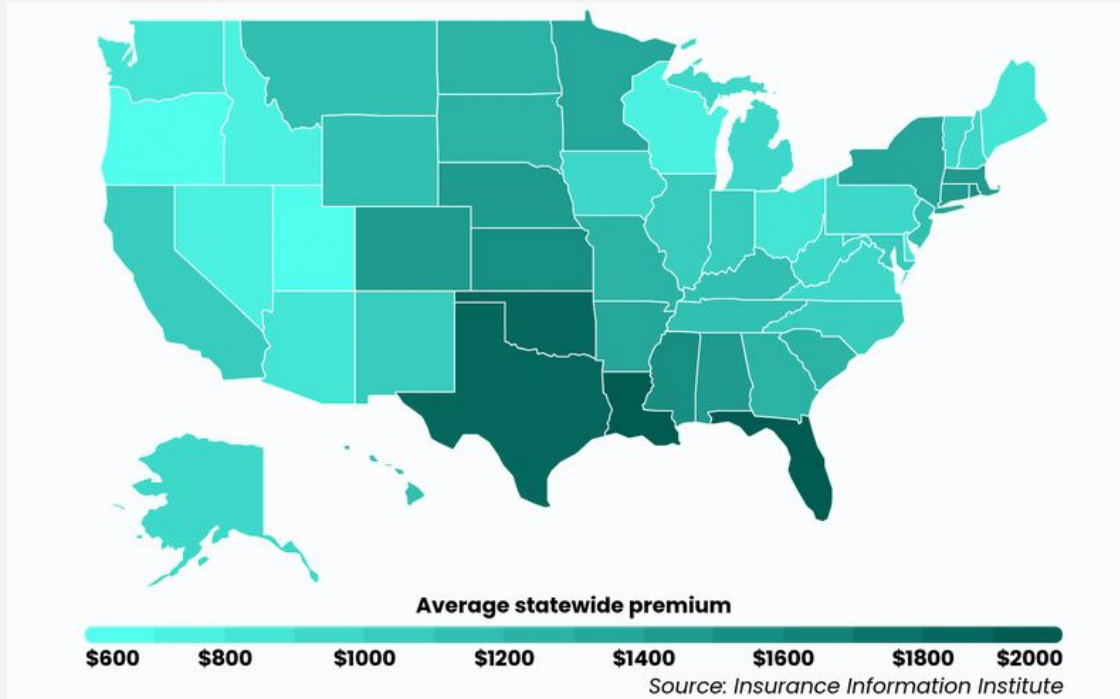
A sample dataset was downloaded for the average home insurance costs across all the states in the US. The map shows the distribution of all the values.

Preparing the Data for Predictions

Data Arrangement

The different disaster layers that were mapped and a count of all the different features was calculated for each of the states and exported as a .csv file.

Also, for the home insurance data a .csv file was generated that shows the average home insurance cost across all the states in the United States.

Finally, a file was generated with the state Name, Average insurance Cost and the number of incidents.

Building the Model & Training

Model

To build the model for this project linear regression modelling was used. The goal of regression is to learn a pattern and then situate a point within the pattern. As Regression works best with numerical variables, that was the best matching modelling technique for this project.

Once the model was built the model was trained on the sample home insurance data collected from different states.

Predictions and Results

Predictions

Once the model was built, trained and saved the predictions were made based on the disaster layers data for all the states.

```
lm_insuranceCosts <- lm(InsuranceCost ~ TotalDisasters, data=allincidents)
summary(lm_insuranceCosts)

unseenCost1 <- data.frame(TotalDisasters = 2500)
predict(lm_insuranceCosts, unseenCost1)
```
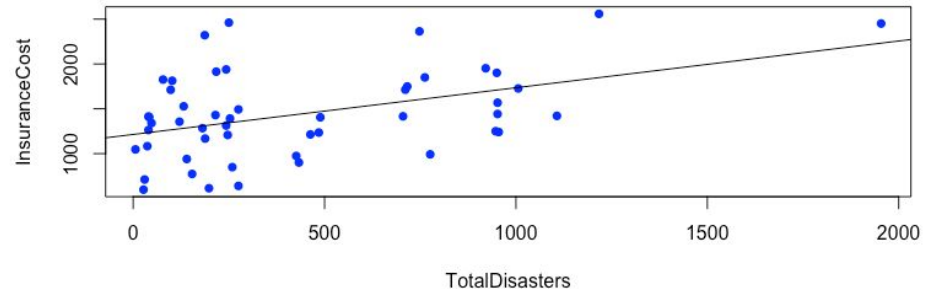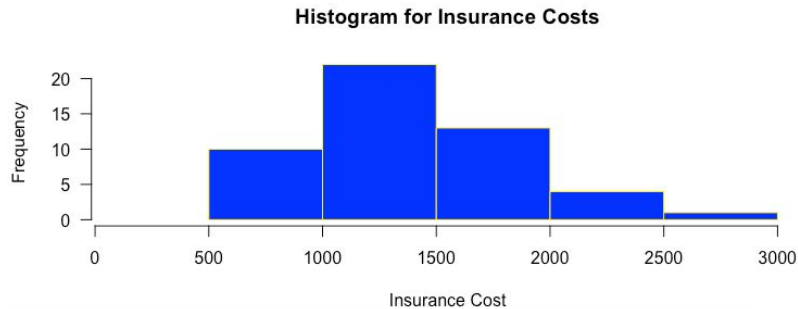
Conclusion

Final Analysis

After all the predictions were made based on the number of disasters in each state, it can be inferred that to accurately predict the insurance cost for a state a lot of other variables are needed but with increasing disasters the cost seems to be increasing. The following plots shows the average insurance cost based on disaster counts vs the average insurance cost by state.

Thank you!