Understanding Rising Home Insurance Premiums

Spring 2025

Samantha Miller

Figure 1: Source: Unsplash, Jakub Żerdzicki

Introduction

According to *The National Oceanic and Atmospheric Administration* (NOAA), natural disasters in the United States have increased in both frequency and severity over the past few decades. At the same time, homeowners' insurance premiums have steadily risen, leading to concerns by many about the factors that seems to be driving these premium increases.

While inflation has often been blamed as the key factor to the rise in insurance premiums, it is critical to consider the role that natural disasters could be having on these trends *(Cavallo 2023)* . Therefore, the purpose of this research is to examine how natural disaster occurrences, inflation rates, and the change in homeowners' insurance premiums (specifically from 2000 to 2018) correlate with one another. This project aims to analyze the impact of natural disasters and inflation, as key risk factors, on the rising costs of home insurance premiums in the United States.

Data Description and Preprocessing

The datasets used in this project include information on natural disasters, homeowners' insurance premiums, commercial insurance premiums, and inflation rates in the United States. The Natural Disasters Dataset contains the details about the occurrences of disasters worldwide, including disaster type, location, and geolocation. To focus the analysis on just the United States, the dataset was filtered to include only U.S. entries.

The Home Insurance Premiums Dataset tracks the Producer Price Index (PPI) for homeowners' insurance, while the Commercial Insurance Premiums Dataset tracks the PPI for commercial multiple peril insurance. In both cases, the original PPI columns were renamed for clarity, and the observation date was standardized to a datetime format. Similarly, the Inflation Data Dataset, which contains U.S. inflation rates, had its main value column renamed and its dates standardized. Rows with missing or invalid dates were removed to maintain data integrity. After cleaning, the datasets were merged using outer joins on the observation date, and a year column was extracted to support annual analysis. The final merged dataset includes variables for insurance premiums, inflation rates, and disaster occurrences by year.

```
import pandas as pd

from scipy.stats import pearsonr
import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests

from sklearn.linear_model import LinearRegression

import matplotlib.pyplot as plt
```

```
import matplotlib.ticker as mticker
import plotly.express as px


# read data
natural_disasters = pd.read_csv('naturaldisasters.csv', on_bad_lines='skip')
home_data = pd.read_csv('home_insurance_premiums.csv', on_bad_lines='skip')
commercial_data = pd.read_csv('commercial _insurance_premiums.csv', on_bad_lines='skip')
inflation_data = pd.read_csv('US_Inflation.csv', on_bad_lines='skip')

# rename and drop variables
inflation_data.rename(columns={"FPCPITOTLZGUSA": "inflation_rate"}, inplace=True)

home_data.rename(columns={"PCU9241269241262": "Home_Insurance_PPI"}, inplace=True)

commercial_data.rename(columns={"PCU9241269241265": "Commercial_Insurance_PPI"}, inplace=True

natural_disasters = natural_disasters.rename(columns={"id": "disaster_id"})
natural_disasters = natural_disasters[natural_disasters["country"] == "United States"]
natural_disasters = natural_disasters[[
    "disaster_id", "country", "iso3", "year", "geo_id", "geolocation",
    "location", "disastertype", "latitude", "longitude", "adm1"
]]

# observation_date to standardized datetime format
home_data["observation_date"] = pd.to_datetime(home_data["observation_date"], errors='coerce
commercial_data["observation_date"] = pd.to_datetime(commercial_data["observation_date"], er
inflation_data["observation_date"] = pd.to_datetime(inflation_data["observation_date"], erro

# drop rows with missing observation dates
home_data.dropna(subset=["observation_date"], inplace=True)
commercial_data.dropna(subset=["observation_date"], inplace=True)
inflation_data.dropna(subset=["observation_date"], inplace=True)

home_data.dropna(subset=["observation_date"], inplace=True)
commercial_data.dropna(subset=["observation_date"], inplace=True)
inflation_data.dropna(subset=["observation_date"], inplace=True)

# merge datasets
merged_data = pd.merge(home_data, commercial_data, on='observation_date', how='outer')
merged_data = pd.merge(merged_data, inflation_data, on='observation_date', how='outer')
```

```
merged_data['year'] = merged_data['observation_date'].dt.year
merged_data = pd.merge(merged_data, natural_disasters, on='year', how='outer')
```

Analysis

The analysis of how insurance rates could be affected by inflation and natural disaster occurances is divided into several distinct parts. Below are the key questions and areas this study aims to explore in each section:

- **Trends Over Time** (Section)**:** How have homeowners insurance premiums, inflation rates, and natural disaster occurrences changed over time?
- **Linearity Analysis** (Section): Is there a linear relationship between inflation and insurance premiums? Similarly, is the frequency of natural disasters linear to the change in premium rates?
- **Multivariate Linearity Analysis** (Section): When considering both inflation and natural disaster occurrences together, how well can they explain the variation in insurance premium rates?
- **Granger Causality Analysis** (Section): Do inflation or natural disasters Granger-cause changes in insurance premiums? In other words, can past values of one variable statistically predict future changes in insurance rates?
- **Analyzing the Trends and Hotspots of Natural Disasters** (Section): What types of natural disasters are most prevalent across the U.S. from 2000–2023? Which states are most affected, and where are the geographic hotspots?

Trends Over Time

This section investigates how insurance premiums, inflation rates, and natural disaster occurrences have changed over time individually. Simple linear regression models are applied to evaluate the presence of trends. These foundational analyses and visualizations provide a basis for exploring potential causal relationships and interactions among the variables in later sections.

Insurance Rates

```
home_insurance_by_year = merged_data.groupby('year')['Home_Insurance_PPI'].mean()
commercial_insurance_by_year = merged_data.groupby('year')['Commercial_Insurance_PPI'].mean(

# prep data for regression
home_insurance_by_year = home_insurance_by_year.dropna()
commercial_insurance_by_year = commercial_insurance_by_year.dropna()

X_home = home_insurance_by_year.index.values.reshape(-1, 1)  # Independent variable (years)
y_home = home_insurance_by_year.values  # Dependent variable (Home Insurance Premiums)

X_commercial = commercial_insurance_by_year.index.values.reshape(-1, 1)  # Independent varial
y_commercial = commercial_insurance_by_year.values  # Dependent variable (Commercial Insuranc

# fit linear regression models
home_model = LinearRegression()
home_model.fit(X_home, y_home)
home_predicted = home_model.predict(X_home)

commercial_model = LinearRegression()
commercial_model.fit(X_commercial, y_commercial)
commercial_predicted = commercial_model.predict(X_commercial)

# plot with regression lines
plt.figure(figsize=(10, 6))
plt.plot(home_insurance_by_year.index, home_insurance_by_year.values, marker='o', label='Home
plt.plot(commercial_insurance_by_year.index, commercial_insurance_by_year.values, marker='s'
plt.plot(home_insurance_by_year.index, home_predicted, linestyle='--', color='blue', label='I
plt.plot(commercial_insurance_by_year.index, commercial_predicted, linestyle='--', color='rec

plt.title('Homeowners vs Commercial Insurance Premiums with Regression')
plt.xlabel('Year')
plt.ylabel('Insurance Premiums (Producer Price Index June 1 1998=100)')
plt.legend()
plt.grid(True)
plt.show()

print(f"Home Insurance Linear Regression Slope: {home_model.coef_[0]}")
print(f"Commercial Insurance Linear Regression Slope: {commercial_model.coef_[0]}")
```

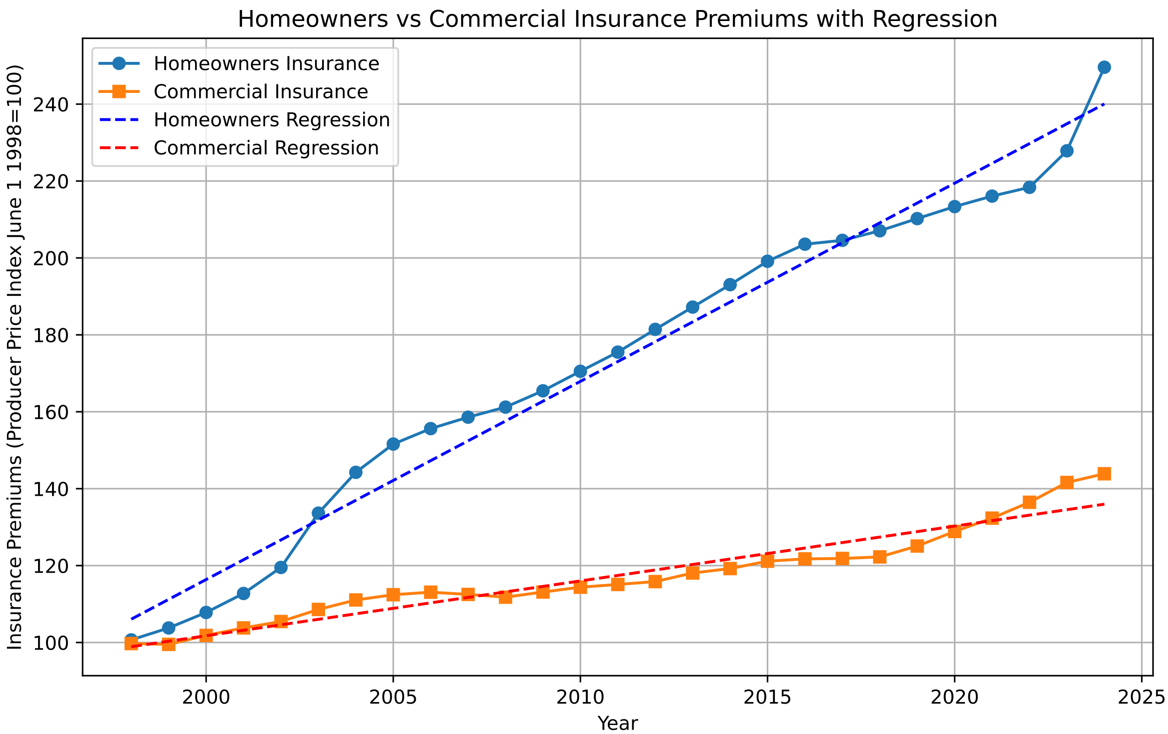


Figure 2: Homeowners vs Commercial Insurance Premiums with Regression

```
Home Insurance Linear Regression Slope: 5.1529773896738185
Commercial Insurance Linear Regression Slope: 1.4252543679865113
```

Figure 2 and the linear regression analysis confirms that homeowners' insurance premiums have been increasing at a significant rate, with a steady annual increase of about 5.15. The rise in premiums confirms the need to investigate the potential factors driving these increases, which is why this project focuses on understanding the relationship between natural disasters, inflation, and insurance premiums. Similarly, commercial insurance premiums are also increasing, but at a slower rate with a slope of 1.43.

Inflation

```
inflation_data_2000 = merged_data[(merged_data['year'] >= 2000) & (merged_data['inflation_ra

X = inflation_data_2000[['year']]  # independent variable (year)
y = inflation_data_2000['inflation_rate']  # dependent variable (inflation_rate)
```

```
inflation_model = LinearRegression()
inflation_model.fit(X, y)


merged_data.loc[merged_data['year'] >= 2000, 'predicted_inflation_rate'] = inflation_model.p

# plot
plt.figure(figsize=(10, 6))
plt.plot(inflation_data_2000['year'],
         inflation_data_2000['inflation_rate'],
         marker='o', linestyle='-', color='b', label='Inflation Rate')
plt.plot(merged_data[merged_data['year'] >= 2000]['year'],
         merged_data[merged_data['year'] >= 2000]['predicted_inflation_rate'],
         linestyle='--', color='r', label='Linear Regression')
plt.title('Inflation Rate Over Time with Linear Regression (2000 and Later)')
plt.xlabel('Year')
plt.ylabel('Inflation Rate')
plt.grid(True)
plt.legend()
plt.show()

print(f"Inflation Linear Regression Slope (2000 and Later): {inflation_model.coef_[0]}")
```

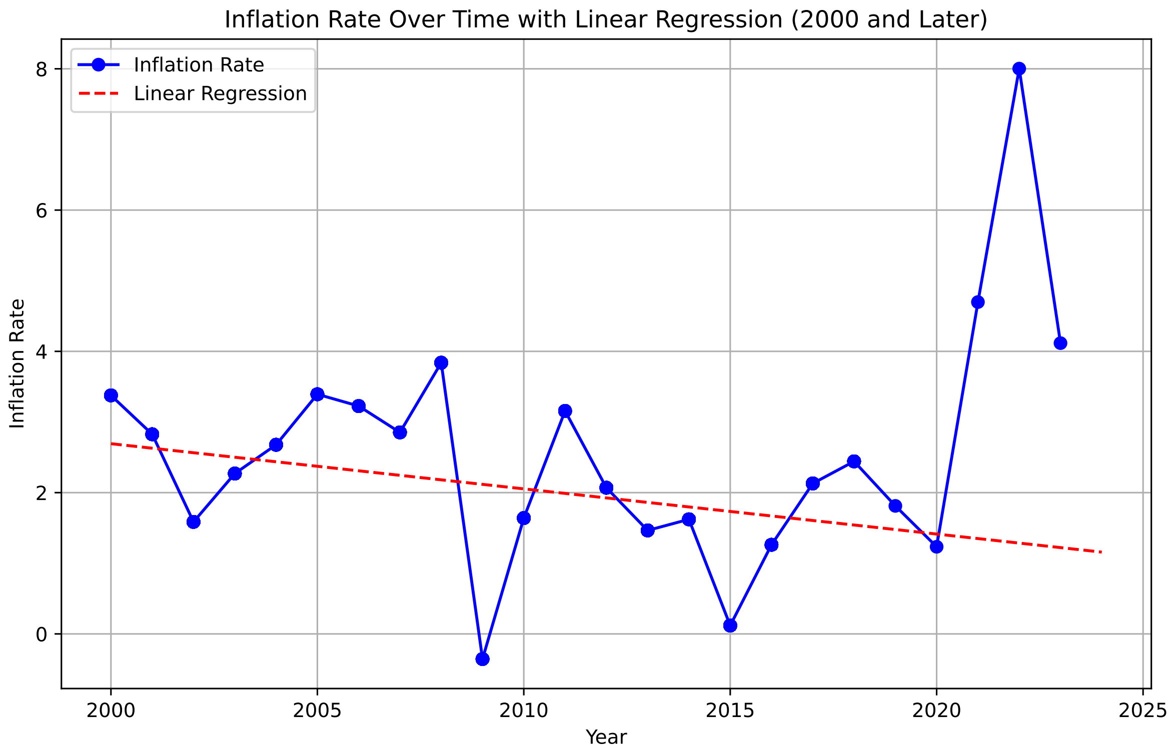


Figure 3: Inflation Rate Over Time with Linear Regression (2000 and Later)

```
Inflation Linear Regression Slope (2000 and Later): -0.06395429870449372
```

The linear regression model in Figure 3 was used to analyze the inflation rates from 2000 . The model predicts inflation based on the year, and the results are plotted, showing both the actual and predicted inflation rates. The regression slope of -0.064 indicates a slight annual decrease in inflation rates from 2000 onward, suggesting a downward trend in inflation over this period. However, there are multiple dips and increases of the actual rates by year. There is a significant increase seen after 2020, which could have an impact on insurance premiums, but is not analyzed in this project due to the natural disaster data having only been updated until 2018.

Natural Disasters

```
natural_disasters_by_year = merged_data.groupby('year')['disaster_id'].count()
# years from 2000 to 2018
natural_disasters_since2000 = natural_disasters_by_year[(natural_disasters_by_year.index >= 
```

```
                                                          (natural_disasters_by_year.index <=

# prep data for regression
natural_disasters_since2000 = natural_disasters_since2000.dropna()

X_disasters = natural_disasters_since2000.index.values.reshape(-1, 1)  # independent variabl
y_disasters = natural_disasters_since2000.values  # dependent variable (Number of Disasters)

# linear regression model
disaster_model = LinearRegression()
disaster_model.fit(X_disasters, y_disasters)
disaster_predicted = disaster_model.predict(X_disasters)

# with regression line
plt.figure(figsize=(10, 6))
plt.plot(natural_disasters_since2000.index, natural_disasters_since2000.values, marker='o', l
plt.plot(natural_disasters_since2000.index, disaster_predicted, linestyle='--', color='red',
plt.title('Natural Disasters Over Time with Regression (2000-2018)')
plt.xlabel('Year')
plt.ylabel('Number of Disasters')
plt.xticks(range(2000, 2019))
plt.legend()
plt.grid(True)
plt.show()

print(f"Natural Disasters Linear Regression Slope: {disaster_model.coef_[0]}")
```

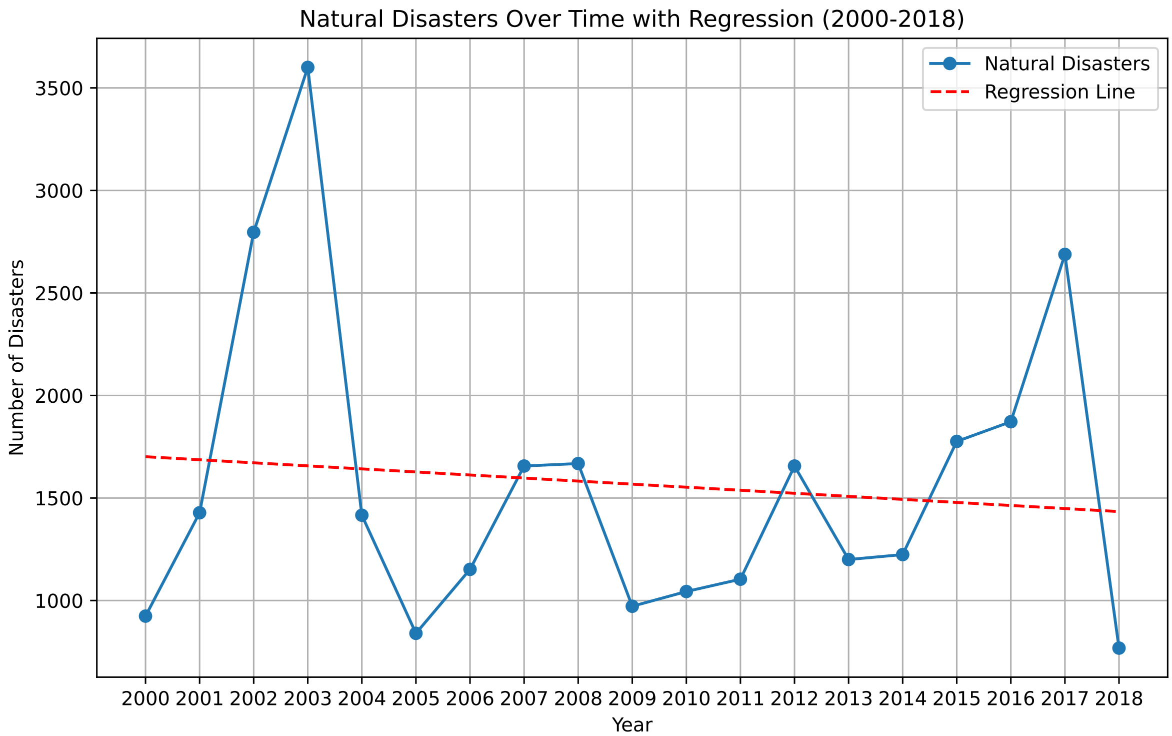


Figure 4: Natural Disasters Over Time with Regression (2000-2018)

```
Natural Disasters Linear Regression Slope: -14.863157894736837
```

This analysis uses a linear regression model, as seen in Figure 4, to explore the number of natural disasters in the United States between 2000 and 2018. The regression line reveals a slight downward trend, with a slope of -14.86, suggesting that the number of disasters decreased by an average of about 15 events per year during this period. However, there are notable fluctuations in the actual frequencies versus the regression line, with a significant peak in 2003 followed by a sharp decline in 2005.

Linearity Analysis

This section analyzes whether changes in home insurance premiums have a linear relationship with inflation rates and the frequency of natural disasters. Pearson correlation tests are used to evaluate these relationships from 2000 to 2018.

Inflation vs Home Insurance Premiums

-**Null Hypothesis:** There is no significant linear relationship between the yearly percentage change in home insurance premiums and inflation rates. (Correlation coefficient = 0)

-**Alternative Hypothesis:** There is a significant linear relationship between the yearly percentage change in home insurance premiums and inflation rates. (Correlation coefficient != 0)

```
Pearson correlation coefficient: 0.06279684885608025
P-value: 0.7984189734928215
Fail to reject the null hypothesis. There is no significant linear relationship between the c
```

The Pearson correlation test was used to evaluate the linear relationship between the yearly percentage change in homeowners' insurance premiums and inflation rates from 2000 to 2018.

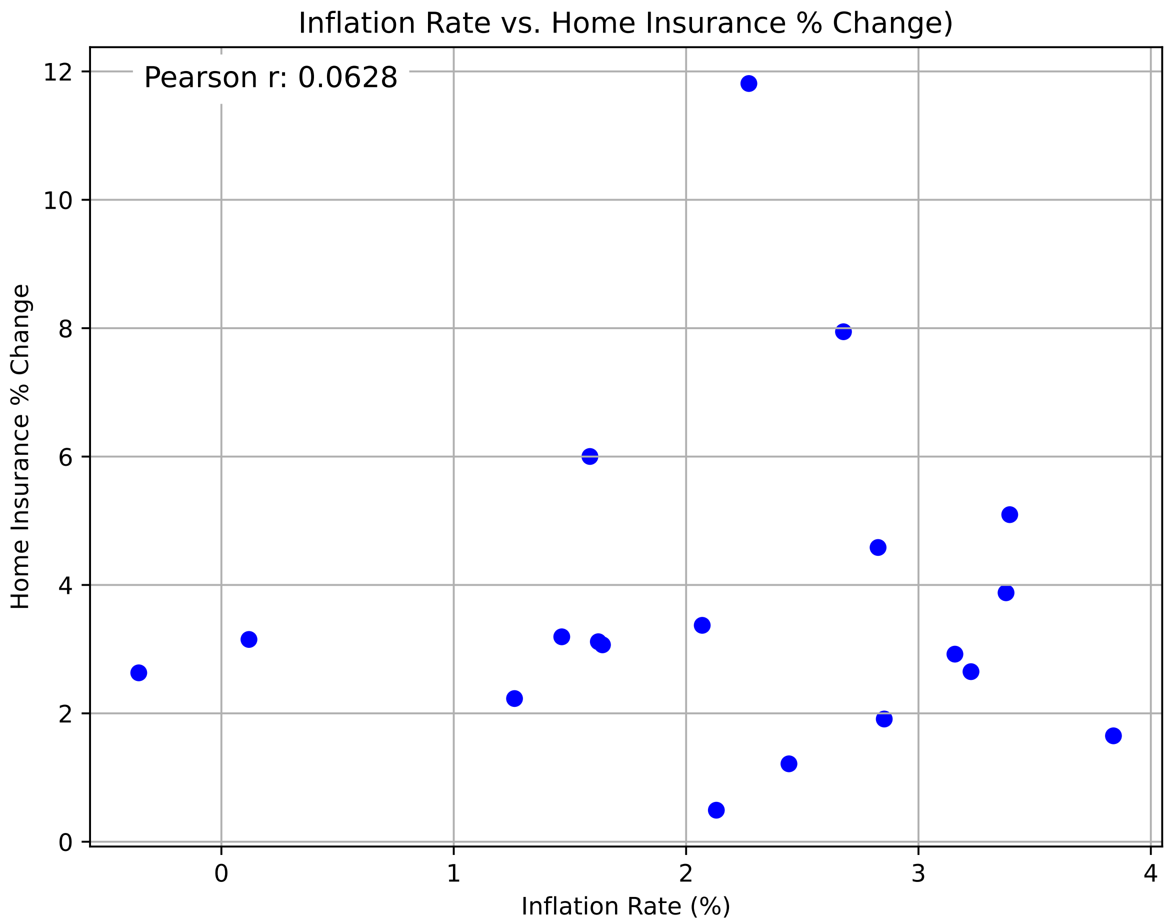
The Pearson correlation coefficient was 0.0628, and the p-value was 0.7984. Since the p-value exceeds the significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is no significant linear relationship between the changes in homeowners' insurance premiums and inflation rates over the period from 2000 to 2018. Figure 5 above visually confirms the weak correlation between the inflation rate and the percentage change in home insurance premiums.

Next, the same test was conducted, but with a shift in the data to compare the inflation rate in a given year to the percentage change in home insurance premiums for the following year. This allows for analysis to determine if home insurance premiums are linearly related to the inflation rate from the prior year, and to assess whether there is a delay in the impact of inflation on insurance rates.

```
# shift home insurance data by one year
home_yearly['home_insurance_pct_change_next_year'] = home_yearly['home_insurance_pct_change']

# merge shifted with inflation data
paired_data_shifted = pd.merge(home_yearly[['year', 'home_insurance_pct_change_next_year']],
                               inflation_yearly[['year', 'inflation_rate']],
                               on='year', how='inner')

paired_data_shifted.dropna(inplace=True)
paired_data_shifted = paired_data_shifted[(paired_data_shifted['year'] >= 2000) & (paired_da


# Pearson correlation test
corr_shifted, p_value_shifted = pearsonr(paired_data_shifted['home_insurance_pct_change_next_
```

```
Text(0.05, 0.95, 'Pearson r: 0.0628')
```

(a) Inflation Rate vs. Home Insurance % Change



(b)

Figure 5

```python
print(f"Pearson correlation coefficient (shifted): {corr_shifted}")
print(f"P-value (shifted): {p_value_shifted}")

# Results
alpha = 0.05
if p_value_shifted < alpha:
    print("Reject the null hypothesis. There is a significant linear relationship between the
else:
    print("Fail to reject the null hypothesis. There is no significant linear relationship be

# scatterplot for inflation rate and next year's insurance rate change
plt.figure(figsize=(8, 6))
plt.scatter(paired_data_shifted['inflation_rate'], paired_data_shifted['home_insurance_pct_c
plt.title('Inflation Rate vs. Next Year Home Insurance % Change')
plt.xlabel('Inflation Rate (%)')
plt.ylabel('Home Insurance % Change (Next Year)')
plt.grid(True)

plt.annotate(f'Pearson r: {corr:.4f}',
             xy=(0.05, 0.95), xycoords='axes fraction',
             fontsize=12, backgroundcolor='white')

plt.show()
```

```
Pearson correlation coefficient (shifted): 0.010169976478017293
P-value (shifted): 0.9670398509254077
Fail to reject the null hypothesis. There is no significant linear relationship between the
```

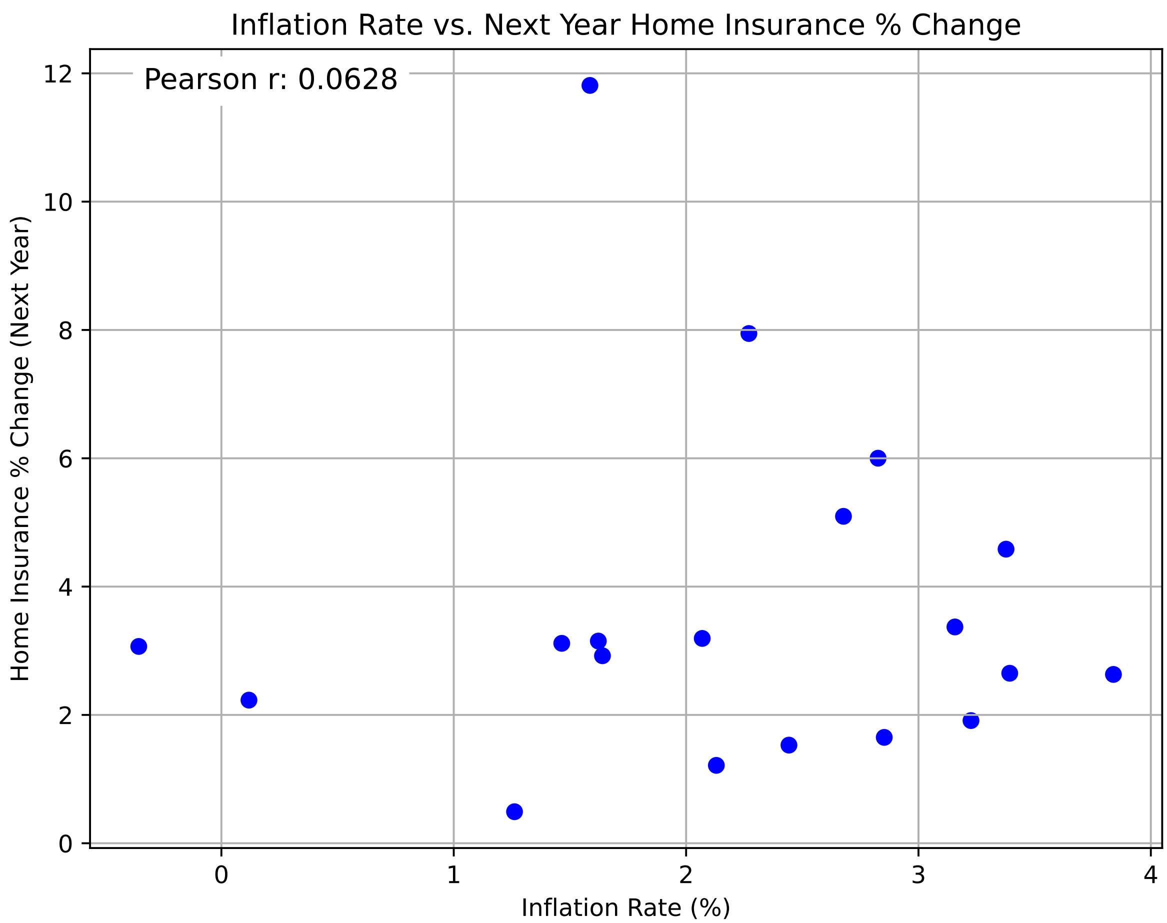


Figure 6: Inflation Rate vs. Next Year Home Insurance % Change

The Pearson correlation test for whether the inflation rate in one year affects the change in home insurance premiums for the next year showed that there is still no significant relationship between them. The correlation was very close to zero, and the p-value was very high (0.967), meaning the result is not statistically significant, suggesting that changes in home insurance premiums for the next year are not strongly influenced by the inflation rate from the previous year. Figure 6 above also visually confirms the weak correlation between the inflation rate and the percentage change in home insurance premiums for the following year.

Natural Disasters vs Home Insurance Premiums

-**Null Hypothesis:** There is no significant linear relationship between disaster frequency and home insurance percentage change.

-**Alternative Hypothesis:** There is a significant linear relationship between disaster frequency and home insurance percentage change.

```
Pearson correlation coefficient: 0.48777700070038277
P-value: 0.034126180998995476
Reject the null hypothesis. There is a significant linear relationship between the disaster c
```

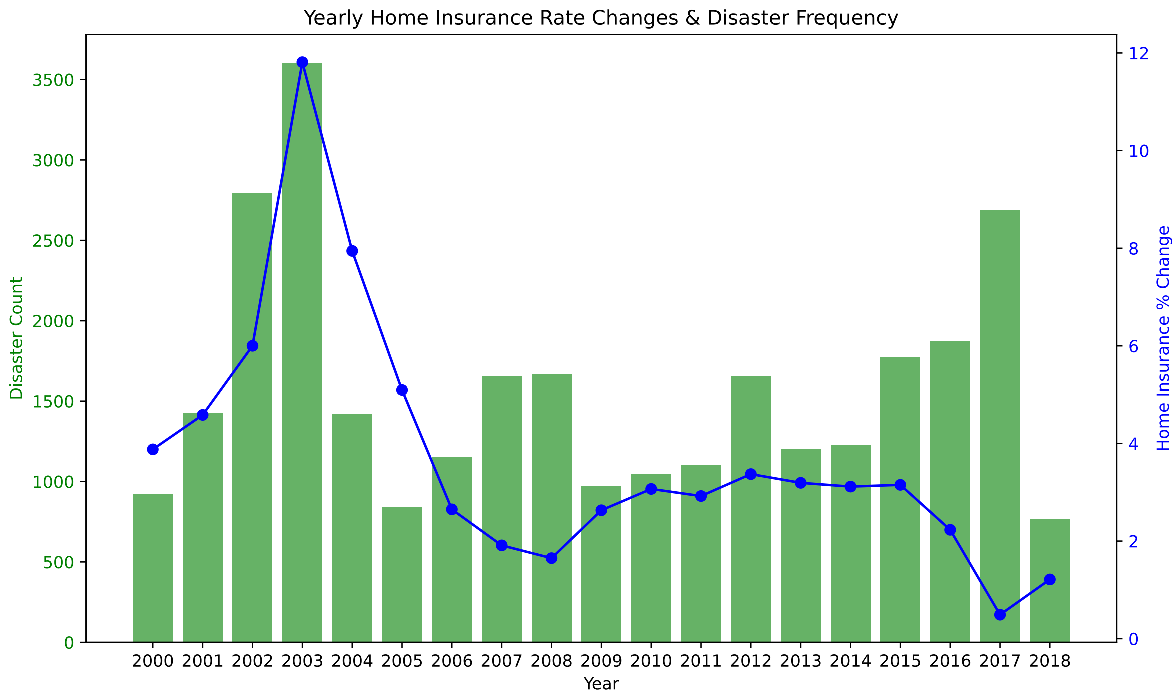


Figure 7: Yearly Home Insurance Rate Changes & Disaster Frequency

Based on the hypothesis test, there is a significant linear relationship between the frequency of natural disasters and the changes in home insurance premiums. The results showed that the correlation coefficient between disaster count and the percentage change in home insurance premiums is 0.49, indicating a positive relationship. The p-value of 0.034 is less than the significance level of 0.05, leading to the rejection of the null hypothesis. This means there is a statistically significant linear relationship between disaster frequency and home insurance premium changes.

Figure 7 provides a clear comparison between the number of natural disasters each year and the changes in home insurance premiums, helping to highlight how these two factors are linearly related be related over time.

Multivariate Linearity Analysis

This section uses multiple linear regression to evaluate the combined influence of inflation and natural disaster frequency on changes in home insurance premiums from 2000 to 2018. By quantifying and visualizing the relative contributions of each factor, the analysis aims to determine which variable more strongly correlates with rising insurance costs.

```
#adjust year 2000-2018
merged_data_filtered = merged_data[(merged_data['year'] >= 2000) & (merged_data['year'] <= 20
disaster_yearly = merged_data.groupby('year')['disaster_id'].count().reset_index()
disaster_yearly.rename(columns={'disaster_id': 'disaster_count'}, inplace=True)

full_data = pd.merge(paired_data, disaster_yearly, on='year', how='inner')
full_data['inflation_pct_change'] = full_data['inflation_rate'].pct_change() * 100
full_data_cleaned = full_data.dropna(subset=['inflation_pct_change', 'disaster_count', 'home_

# independent variables (inflation percentage change and disaster count)
X = full_data_cleaned[['inflation_pct_change', 'disaster_count']]
X = sm.add_constant(X)

# dependent variable (home insurance percentage change)
y = full_data_cleaned['home_insurance_pct_change']

# multiple linear regression model
model = sm.OLS(y, X).fit()

print(f"R-squared: {model.rsquared:.4f}")
print(f"Adjusted R-squared: {model.rsquared_adj:.4f}")
print(f"Coefficients:\n{model.params}")
print(f"P-values:\n{model.pvalues}")

# coef from regression model
beta_1, beta_2 = model.params['inflation_pct_change'], model.params['disaster_count']

# calculate contributions
full_data['Inflation_Impact'] = beta_1 * full_data['inflation_pct_change']
full_data['Disaster_Impact'] = beta_2 * full_data['disaster_count']

# normalize contributions to sum up to 100% of the insurance rate change
full_data['Total_Impact'] = full_data['Inflation_Impact'] + full_data['Disaster_Impact']
full_data['Inflation_Impact'] = (full_data['Inflation_Impact'] / full_data['Total_Impact']) *
full_data['Disaster_Impact'] = (full_data['Disaster_Impact'] / full_data['Total_Impact']) * 1
```

```python
full_data = full_data[full_data['year'] <= 2018]

# stacked area chart with custom colors
plt.figure(figsize=(10,6))
plt.stackplot(full_data['year'], full_data['Inflation_Impact'], full_data['Disaster_Impact']
              labels=['Inflation Impact', 'Disaster Impact'], alpha=0.7,
              colors=['red', 'lightblue'])
plt.xticks(full_data['year'], rotation=45)
plt.gca().xaxis.set_major_locator(mticker.MultipleLocator(1))  #show every year
plt.title('Impact of Inflation vs. Disasters on Home Insurance Rate Increases Over Time')
plt.xlabel('Year')
plt.ylabel('Contribution to Insurance Rate Increase (%)')
plt.legend(loc='upper left')
plt.grid(True, linestyle='--', alpha=0.5)

plt.show()
```

```
R-squared: 0.2812
Adjusted R-squared: 0.1854
Coefficients:
const                   0.685448
inflation_pct_change   -0.001707
disaster_count          0.001917
dtype: float64
P-values:
const                   0.630276
inflation_pct_change    0.445578
disaster_count          0.029843
dtype: float64
```

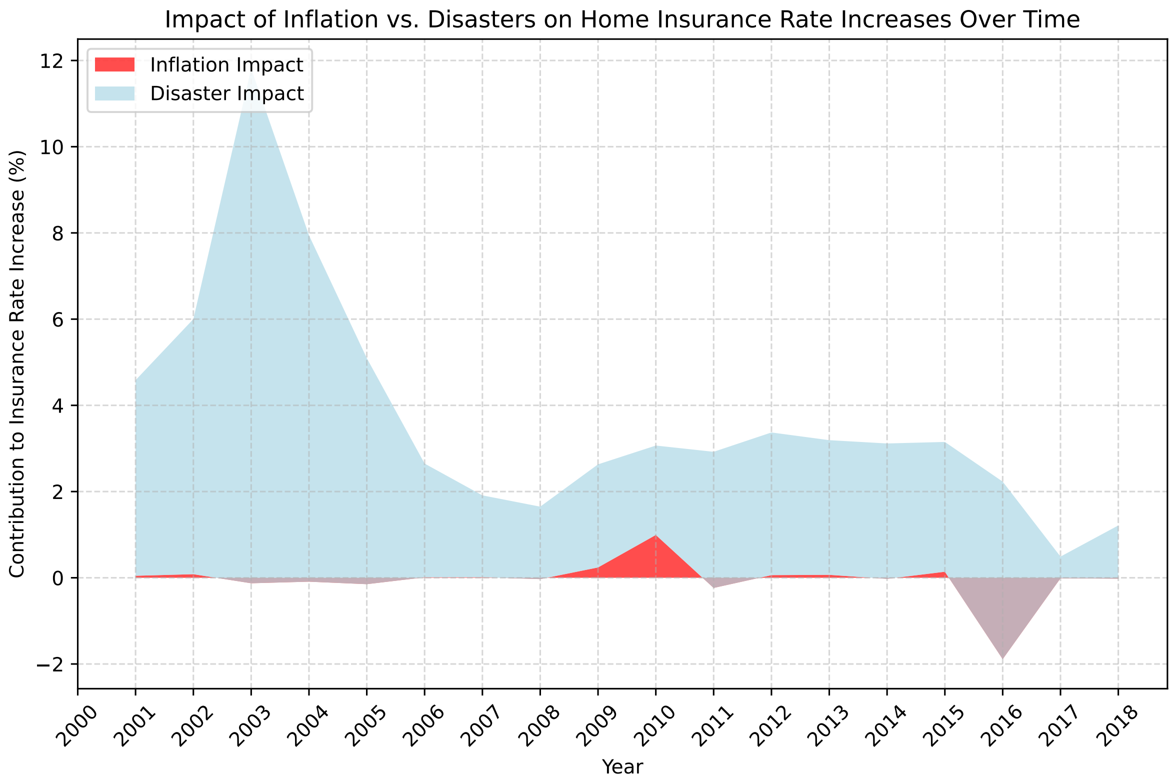


Figure 8: Impact of Inflation vs. Disasters on Home Insurance Rate Increases Over Time

The multiple linear regression analysis examined the actual impact of inflation and natural disaster frequency on changes in home insurance premiums. The results revealed that inflation has a minimal and statistically insignificant effect on home insurance premiums (with a p-value of 0.45). On the contrary, the frequency of natural disasters showed a significant relationship with premium changes, with a p-value of 0.03. This suggests that natural disasters are a more influential factor in driving up home insurance premiums compared to inflation. However, The R-squared value of 0.2812 indicates that approximately 28.12% of the variability in home insurance premium changes is explained by inflation and natural disaster occurances. This suggests that while the model accounts for some of the variation in insurance premiums, there is still a significant portion of the variation that remains unexplained by these factors. This means other potential factors, not included in this analysis, are likely also influencing the changes in home insurance premiums in real life.

The stacked area chart , Figure 8, visualizes these findings, showing how the contributions of inflation and disasters over time on insurance rates. Inflation's impact remains relatively weak, with only small changes in premiums. On the other hand, the influence of disaster frequency is a lot higher, emphasizing the role of disasters in driving premium increases.

Granger Causality Analysis

The Granger Causality Test checks if past natural disasters can predict changes in home insurance premiums in the following years.

```
Granger Causality Test: Disaster Count -> Home Insurance % Change

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=0.9413  , p=0.3473  , df_denom=15, df_num=1
ssr based chi2 test:   chi2=1.1295  , p=0.2879  , df=1
likelihood ratio test: chi2=1.0955  , p=0.2953  , df=1
parameter F test:         F=0.9413  , p=0.3473  , df_denom=15, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=1.0555  , p=0.3782  , df_denom=12, df_num=2
ssr based chi2 test:   chi2=2.9907  , p=0.2242  , df=2
likelihood ratio test: chi2=2.7549  , p=0.2522  , df=2
parameter F test:         F=1.0555  , p=0.3782  , df_denom=12, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=1.2012  , p=0.3637  , df_denom=9, df_num=3
ssr based chi2 test:   chi2=6.4066  , p=0.0934  , df=3
likelihood ratio test: chi2=5.3883  , p=0.1455  , df=3
parameter F test:         F=1.2012  , p=0.3637  , df_denom=9, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=0.3203  , p=0.8549  , df_denom=6, df_num=4
ssr based chi2 test:   chi2=3.2034  , p=0.5244  , df=4
likelihood ratio test: chi2=2.9033  , p=0.5741  , df=4
parameter F test:         F=0.3203  , p=0.8549  , df_denom=6, df_num=4

Granger Causality
number of lags (no zero) 5
ssr based F test:         F=0.2342  , p=0.9240  , df_denom=3, df_num=5
ssr based chi2 test:   chi2=5.4636  , p=0.3620  , df=5
likelihood ratio test: chi2=4.6128  , p=0.4649  , df=5
parameter F test:         F=0.2342  , p=0.9240  , df_denom=3, df_num=5
Lag 1 results:
```

```
Lag 2 results:
```

The Granger Causality Test results indicate that despite the earlier Pearson Correlation tests on linearity and multiple linear regression suggesting a relationship between the frequency of natural disasters and changes in home insurance premiums, there is no significant causal link between the two variables. The p-values for all lags (from 1 to 5) are well above the 0.05 threshold, suggesting that the number of natural disasters in any given year does not predict the changes in home insurance premiums in the following years. This means that while there may be a linear relationship between these variables, the frequency of disasters does not directly cause the increases in insurance rates for the next year, since the causality test fails to reject the null hypothesis.

Analyzing the Trends and Hotspots of Natural Disasters

This section explores the trends and geographic distribution of natural disasters in the U.S. from 2000 to 2023. The analysis begins with an overview of the types of natural disasters and their distribution across the years. It then moves to a state-level analysis, identifying regions most affected by these events, followed by a geographical visualization highlighting disaster hotspots across the country. These visualizations provide insight into the frequency and impact of natural disasters in different regions, shedding light on areas most vulnerable to such events.

```python
# 2000-2023
filtered_data = merged_data[(merged_data['year'] >= 2000) & (merged_data['year'] <= 2023)]

disasters_by_type = filtered_data.groupby('disastertype').size()

plt.figure(figsize=(10, 7))
colors = plt.cm.Paired.colors

plt.pie(
    disasters_by_type,
    startangle=90,
    colors=colors,
    pctdistance=0.85
)



# key
plt.legend(disasters_by_type.index, title='Disaster Types', loc='upper left', bbox_to_anchor=
```

```
plt.show()
```

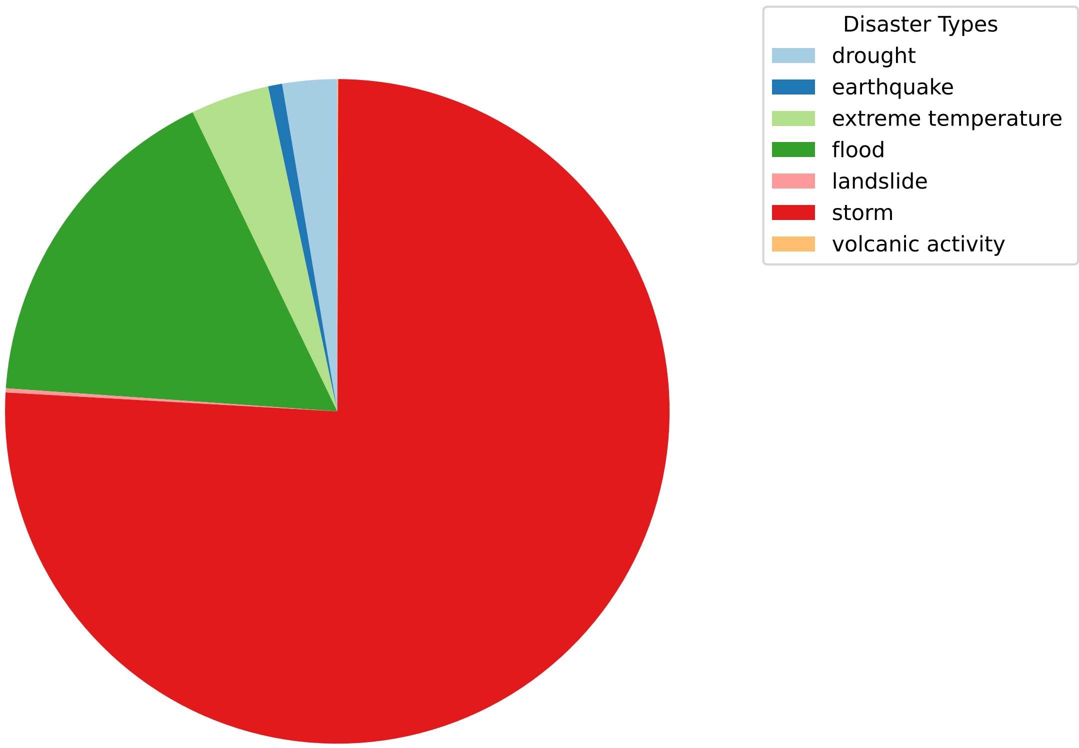


Figure 9: Distribution of Natural Disasters by Type (2000 - 2023)

Figure 9 illustrates the distribution of different types of natural disasters in the U.S. from 2000 to 2023. It shows that storms are the most common natural disaster in the country.

```
# natural disasters by state
disasters_by_state = natural_disasters['adm1'].value_counts().reset_index()
disasters_by_state.columns = ['state', 'num_disasters']

# names to state codes
us_states = pd.read_csv("https://raw.githubusercontent.com/jasonong/List-of-US-States/master/
disasters_by_state = disasters_by_state.merge(us_states, left_on='state', right_on='State', h

# map
```

```python
fig = px.choropleth(disasters_by_state,
                    locations="Abbreviation",
                    locationmode="USA-states",
                    color="num_disasters",
                    hover_name="state",
                    hover_data=["num_disasters"],
                    color_continuous_scale="dense",
                    labels={'num_disasters': 'Number of Natural Disasters'},
                    scope="usa"
                   )

fig.update_layout(
    width=750, height=600,
    coloraxis_colorbar=dict(len=0.6),
    margin=dict(l=0, r=0, t=50, b=0)
)

fig.update_traces(marker_line_color='white', marker_line_width=1.0)

fig.show()
```

```
Unable to display output for mime type(s): text/html
```

(a) Natural Disasters by State in the USA

```
Unable to display output for mime type(s): text/html
```

(b)

Figure 10

Figure 10 displays the number of natural disasters across different U.S. states from 2000 to 2023, where Texas is shown to be the state most vulnerable to occurances.

```python
disasters_geo = natural_disasters.dropna(subset=['longitude', 'latitude'])

# Scatter Geo Plot
fig = px.scatter_geo(disasters_geo,
                     lon='longitude',
                     lat='latitude',
                     scope='usa',
                     color='disastertype',
```

```
                     hover_name='disastertype',
                     hover_data={'longitude': True, 'latitude': True, 'year': True},
                     opacity=0.6,
                     color_discrete_sequence=px.colors.qualitative.Bold)

fig.update_layout(
    geo=dict(
        scope='usa',
        landcolor='lightgray',
        showcountries=False,
        showland=True,
        showlakes=True,
        lakecolor='lightblue'
    ),
    margin=dict(l=0, r=0, t=50, b=0),
    width=800,
    height=600
)

fig.show()
```

```
Unable to display output for mime type(s): text/html
```

Figure 11: Natural Disaster Hotspots in the USA

This scatter geo plot, Figure 11, highlights the geographic distribution of natural disasters across the United States. Each point represents a disaster, color-coded by type, with the hover feature to provide details about the disaster's type, location, and year it occured.

Conclusion

In conclusion, the analysis of natural disasters, inflation rates, and home insurance premiums provides valuable insights into the factors influencing insurance cost changes over time. The statistical tests and regression models revealed that while inflation has a limited impact, natural disasters significantly affect insurance premium increases. However, the Granger causality test suggested no direct causal relationship between natural disasters and premium adjustments, despite having a linear relationshp, suggesting that other underlying factors may also play a role.

The visualizations of natural disaster occurances in the United States highlighted key trends, including the distribution of disasters by type, the state-by-state impact, and geographic hotspots of disaster occurrences. These findings offer a deeper understanding of how natural disasters vary across regions and how they could have a potential influence on insurance rates based on this study.

References

Cavallo, Ben. 2023. "Insurance Inflation: How & Why Rates Are Affected." C&S Insurance; https://www.candsins.com/blog/insurance-inflation/. https://www.candsins.com/blog/insurance-inflation/.

Federal Reserve Bank of St. Louis. 2025. "10-Year Breakeven Inflation Rate [T10YIE]." https://fred.stlouisfed.org/series/T10YIE.

Guha-Sapir, Debarati, Regina Below, and Philippe Hoyois. 2014. "EM-DAT: International Disaster Database." Centre for Research on the Epidemiology of Disasters (CRED).

National Oceanic and Atmospheric Administration. 2023. "Annual Report on u.s. Natural Disasters." https://www.ncei.noaa.gov/access/billions/.

Rosvold, E. L., and H. Buhaug. 2021. "GDIS, a Global Dataset of Geocoded Disaster Locations." *Sci Data* 8: 61. https://doi.org/https://doi.org/10.1038/s41597-021-00846-6.

Rosvold, E., and H. Buhaug. 2021. "Geocoded Disaster (GDIS) Dataset." Palisades, NY: NASA Socioeconomic Data; Applications Center (SEDAC). https://doi.org/10.7927/zz3b-8y61.

U.S. Bureau of Labor Statistics. 2025a. "Producer Price Index by Industry: Premiums for Property and Casualty Insurance: Premiums for Commercial Multiple Peril Insurance [PCU9241269241265]." https://fred.stlouisfed.org/series/PCU9241269241265.

———. 2025b. "Producer Price Index by Industry: Premiums for Property and Casualty Insurance: Premiums for Homeowner's Insurance [PCU9241269241262]." https://fred.stlouisfed.org/series/PCU9241269241262.