Methods of Advance Data Engineering Data Report

Kartik Kanodia

Summer Semester 2024

1. Introduction

The United States of America is frequently affected by various natural disasters, such as hurricanes or wildfires. These events can have devastating impacts on communities, infrastructure, and the economy. Hence, analysing historical data on natural disasters and temperature trends can provide valuable insights into patterns and correlations, which may be essential for developing effective strategies to mitigate risks and enhance resilience. This report presents a descriptive analysis of a dataset containing average temperatures and incidents of various types of natural disasters in the United States of America from year 1990 to 2013. The analysis includes the distribution of incident types, yearly and monthly trends, and a correlation analysis between temperature and natural disaster incidents.

2. Used Data

The report contains two data sources.

Datasource 1: Kaggle

Type: CSV

This dataset originates from the Berkeley Earth Surface Temperature Study and covers global land and ocean temperatures from as early as 1750, with temperature averages and uncertainties provided. The dataset provides geographical granularity at the country level.

Licence: CC BY-NC-SA 4.0

CC BY-NC-SA 4.0 is a type of Creative Commons license. This license allows others to Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) as long as they credit the original creator, use the material for non-commercial purposes and license any new creations under identical terms.

Datasource 2: Kaggle

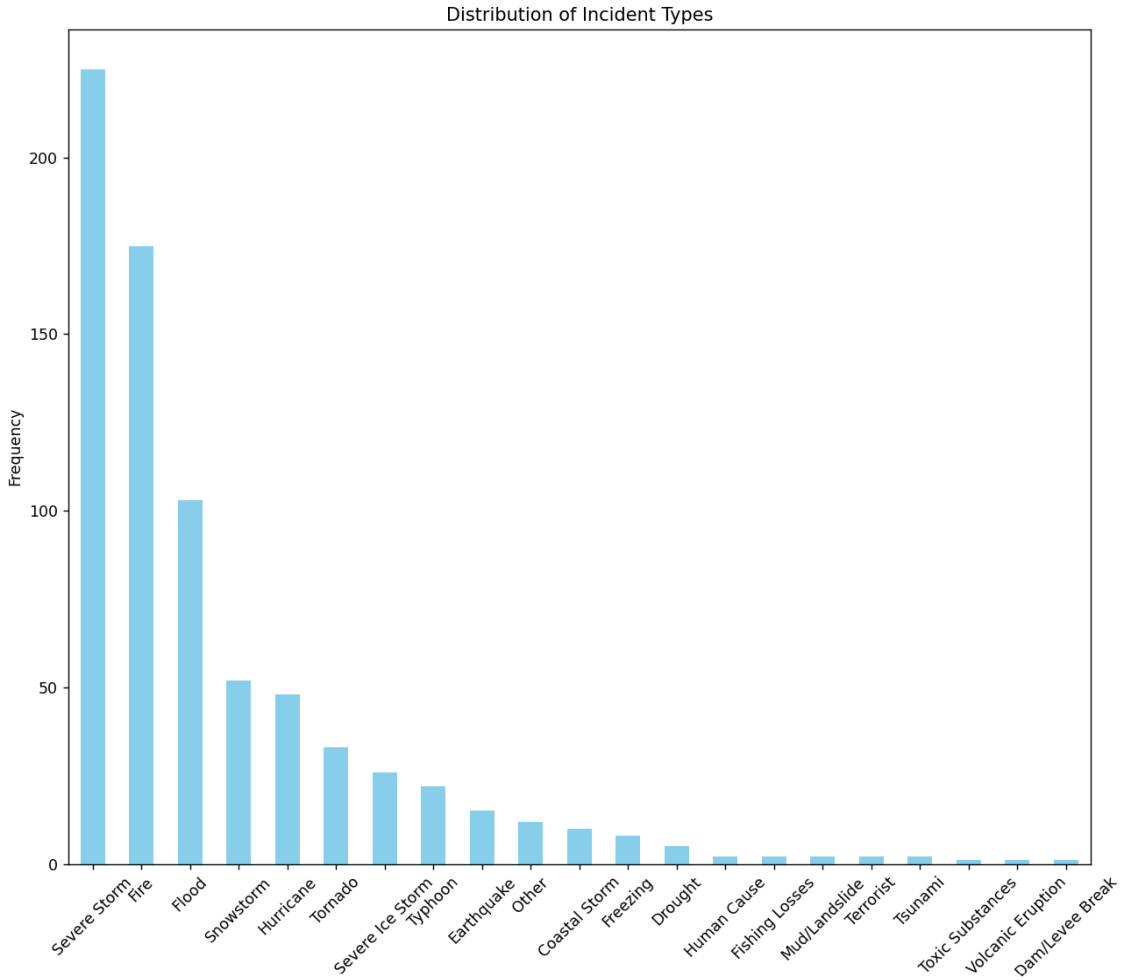
Data Type: CSV

This dataset is a high-level summary of all federally declared disasters that occurred in the United States 1 of America since 1953. The geographical granularity of the raw dataset is at a city level. The dataset provides the type of disaster and also binary flags that indicate whether specific aid programs were triggered in response. Licence: U.S. Government Works license refers to works created by employees of the United States federal government as part of their official duties. Since these works are in the public domain, they are free to use by anyone without needing permission or paying royalties. They can be freely copied, modified, distributed, and performed by anyone.

3. Analysis

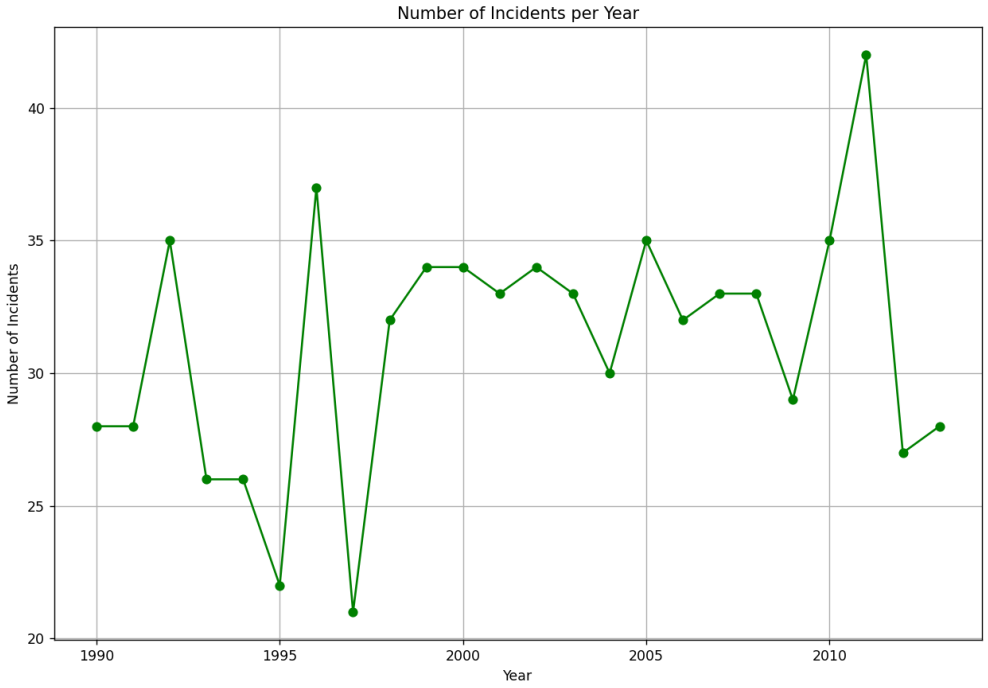
Thise section of the report presents the data analysis using various types of visualisations and their descriptions.

1. Distribution of Incident Types



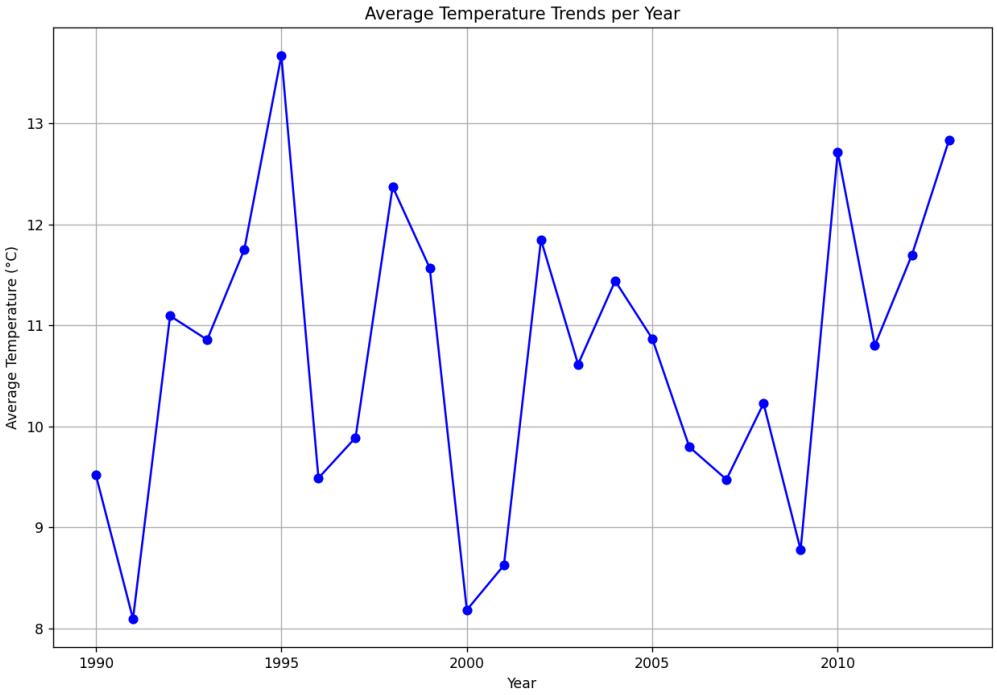
The most frequent incident type are ‘Severe Storm’ followed by ‘Fire’ and ‘Flood’. While, the less common incident types include ‘Tsunami’, ‘Volcanic Eruption’, and ‘Dam/Levee Break’.

2. Number of Incidents per Year



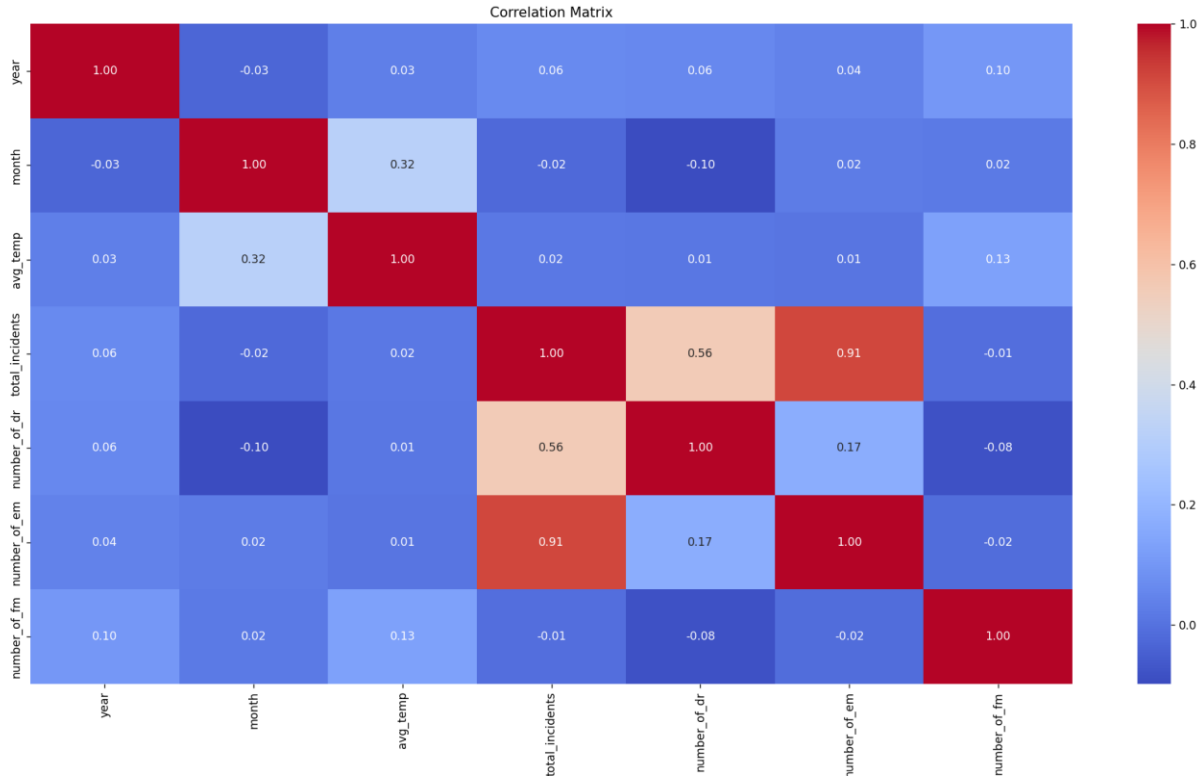
The number of incidents per year shows some variability but generally ranges from 21 to 42 incidents per year, with the highest number of incidents occurring in 1996 and 2011.

3. Average Temperature Trend



At a yearly level a general increase in average temperature, with some fluctuations can be observed.

4. Correlation Analysis



The visualisation shows a weak positive correlation between average temperature and total incidents with a correlation coefficient of 0.016.

5. Conclusions

This analysis presents several important aspects related to temperature trends and natural disaster incidents in the United States of America from years 1990 to 2013. The data showed a wide range of disaster types, with severe storms being the most frequently occurring natural disaster. The analysis further revealed seasonal patterns in natural disaster incident occurrences, with more incidents in the summer months. Though there was a general increase in average temperature over the years in the United Staes of America, the correlation between temperature increases and the number of natural disaster incidents is relatively weak. Hence a correlation could not be established. However, the insights gained form this analysis could be relevant to future studies and aid in disaster preparedness and response planning. Apart from these insights, there are limitations and uncertainties to consider as well. Since, the dataset is limited to the United States it may not capture the global trends and influences. Furthermore, the correlation between temperature rise and natural disaster occurrences is weak, suggesting that other factors, not included in the dataset, may play a significant role in the occurrence of natural disasters. Additionally, the data spans only until 2013, and more recent data would be helpful in identifying current trends. To conclude, the analysis addresses the research question to an extent, however, it cannot provide a comprehensive answer due to these limitations. Future studies should include additional variables and more recent data to gain a deeper understanding of the problem.