Project Report

GitHub URL

https://github.com/thiago-preira/UCDPA_ThiagoPereira

Abstract

This project aims to show the impact of Severe Storm in the USA using data provided by the National Center of Environment Information. The analysis was done over events from the past 10 years, revealing the most common events, the distribution of these events and damage and deaths caused.

Introduction

The National Centers for Environment Information maintains one of the most significant archives on Earth, managing more than 30 petabytes of data and information that spans the entire spectrum of Earth's environmental systems and cycles with comprehensive oceanic, atmospheric, and geophysical data.

The Storm Events Database is an integrated database of severe weather events across the United States from 1950 to this year, with information about the occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, disruption to commerce.

This dataset allows identifying the most severe storms and damage caused by them, as well as define correlations between the features that compose the dataset.

Dataset

The datasets are provided via bulk download. There are 3 files linked by the event ID number. Details, locations and fatalities

- Events Details file: The storm event description and data captured
- Events Location file: The storm location data
- Event fatalities file: the fatalities related to the storm

The full description of all features as described here. The datasets relevant to this project are Storm details and fatalities file

Event Details file

Named as StormEvents_details-ftp_v1.0_dYYYY_cYYYYMMdd.csv where dYYYY = data year and cYYYYMMdd = file creation date

- episode_id Ex: 61280, 62777, 63250 ID assigned by NWS to denote the storm episode;
- event_id Ex: 383097, 374427, 364175 ID assigned by NWS for each individual storm event contained within a storm episode; links the record with the same event in the storm_event_details, storm_event_locations and storm_event_fatalities tables (Primary database key field).
- state Ex: GEORGIA, WYOMING, COLORADO The state name where the event occurred (no State ID's are included here; State Name is spelled out in ALL CAPS).
- year Ex: 2000, 2006, 2012 The four digit year for the event in this record.
- event_type Ex: Hail, Thunderstorm Wind, Snow, Ice (spelled out; not abbreviated)
- injuries_direct Ex: 1, 0, 56 The number of injuries directly caused by the weather event.
- injuries_indirect Ex: 0, 15, 87 The number of injuries indirectly caused by the weather event.
- deaths_direct Ex: 0, 45, 23 The number of deaths directly caused by the weather event.
- deaths_indirect Ex: 0, 4, 6 The number of deaths indirectly caused by the weather event.
- damage_property Ex: 10.00K, 0.00K, 10.00M The estimated amount of damage to property incurred by the weather event
- damage_crops Ex: 0.00K, 500.00K, 15.00M The estimated amount of damage to crops incurred by the weather event
- magnitude Ex: 0.75, 60, 0.88, 2.75 The measured extent of the magnitude type ~ only used for wind speeds (in knots) and hail size (in inches to the hundredth).
- tor_f_scale Ex: EF0, EF1, EF2, EF3, EF4, EF5 Enhanced Fujita Scale describes the strength of the tornado based on the amount and type of damage caused by the tornado. The F-scale of damage will vary in the destruction area; tor_length Ex: 0.66, 1.05, 0.48 Length of the tornado or tornado segment while on the ground (in miles to the tenth).
- tor_width Ex: 25, 50, 2640, 10

Storm Data Fatality File

Named StormEvents_fatalities-ftp_v1.0_dYYYY_cYYYYMMdd.csv.gz where dYYYY = data year and cYYYYMMdd = file creation date

- fatality_id Ex: 17582, 17590, 17597, 18222 ID assigned by NWS to denote the individual fatality that occurred)
- event_id Ex: 383097, 374427, 364175 ID assigned by NWS for each individual storm event contained within a storm episode; links the record with the same event in the storm_event_details, storm_event_locations and storm_event_fatalities tables (Primary database key field)
- fatality_type Ex: D , I (D = Direct Fatality; I = Indirect Fatality; assignment of this is determined by NWS software; details below are from NWS Directve 10-1605 at http://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf, Section 2.6)
- fatality_date Ex: 4/3/2012 00:00 MM/DD/YYYY hh:mm (time is usually 00.00)
- fatality_age Ex: 38, 25, 69, 54 The age in years of the fatality (sometimes 'null' if unknown)
- fatality_sex Ex: M, F The gender of the fatality (sometimes 'null' if unknown)
- fatality_location Ex: UT, OU, MH, PS

Implementation Process

First all files need to be downloaded from a bulk download website. The website provides files from 1950 until the current year, but for this experiment will be using just the 10 years of data, from 2011 to 2021. The library BeatifulSoup allows scraping the website and downloading all csv datasets. The desired files have the extension as .csv.gz and a name pattern well defined. The files will be saved to the datasets folder.

After downloading the datasets, they need to be loaded into Pandas Dataframe. As mentioned above, only a subset of features will be used. The strategy is to join the Storm details and fatalities dataset for each year, and append to a parent dataset that will hold all data.

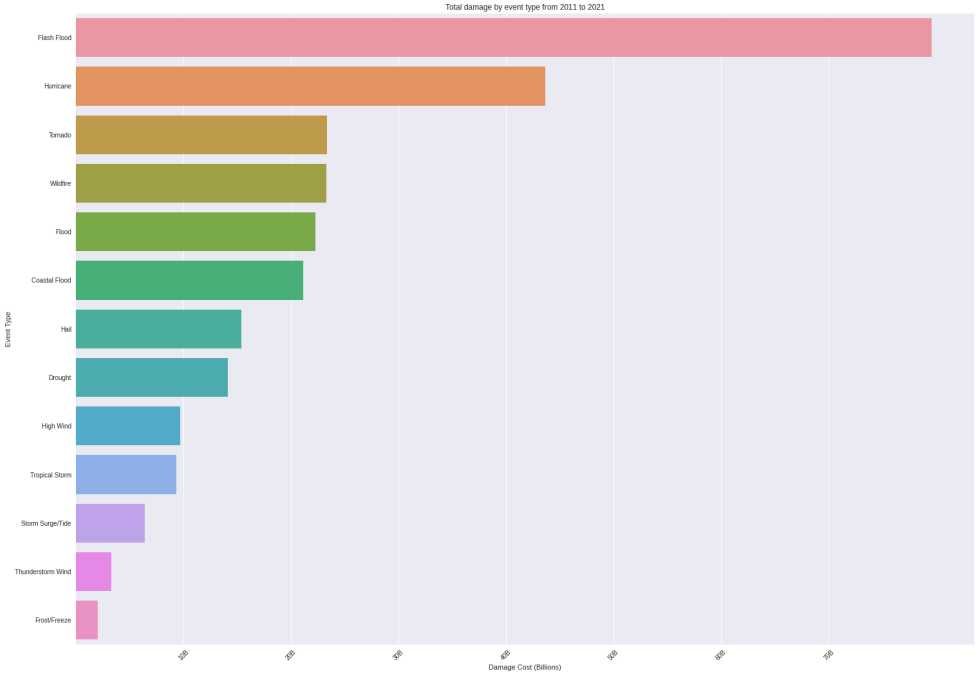
After some cleaning, the analysis starts by inferring the distribution of events and the damage that they have caused. To apply the calculation of total damage, a sum of damage_crops and damage_property will be assigned as a new column called damage_total. The results will be filtered by states that had more than 10k events and damage values greater than a billion. Further, present the number of event types and how life threatening they are. Again, a summary of injuries_direct and injuries_indirect and deaths_direct and deaths_indirect is needed for the calculation.

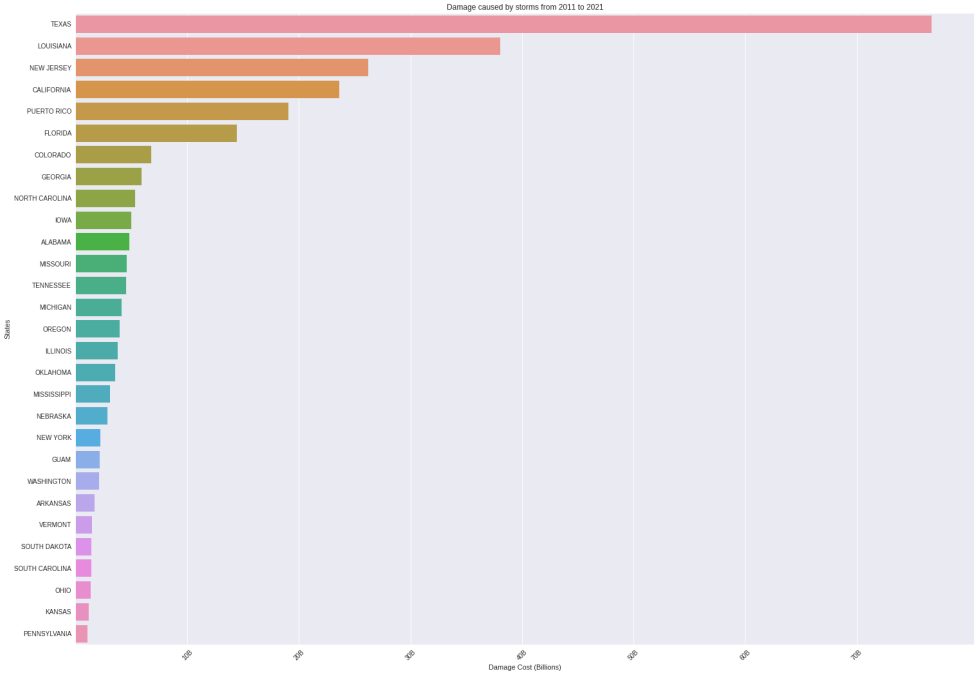
In order to present the distribution of deaths related event types, fatalities dataset needs to be merged with storm events dataset. Only the event_id and event_type columns are needed.

The result will be filtered by the top 5 deadliest events.

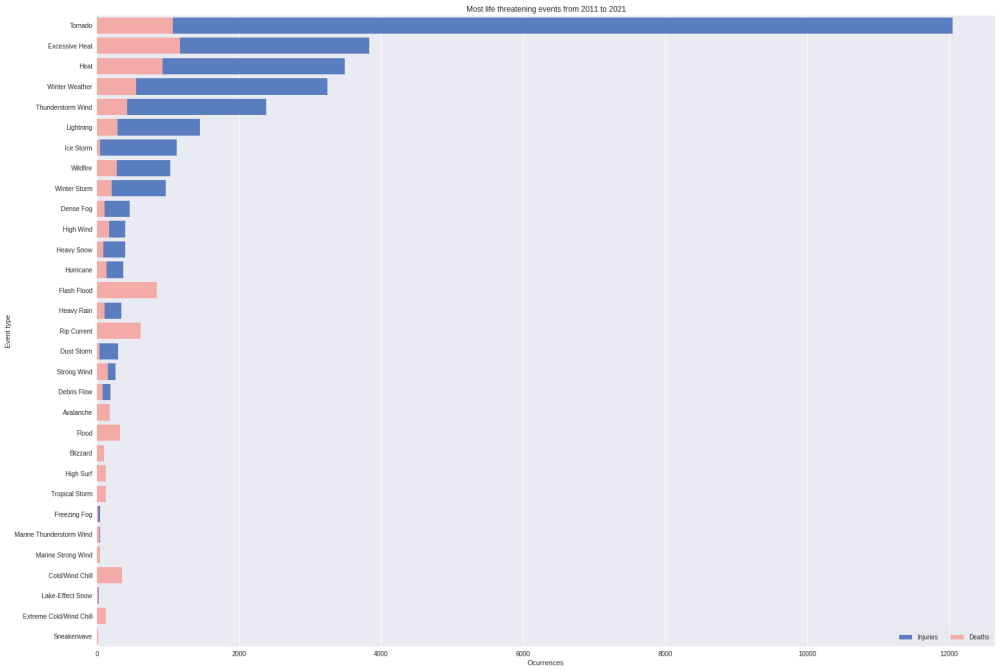
Finalizing used the RandomForestRegressor model to predict how much damage a Tornado event would cause, starting with the feature selections

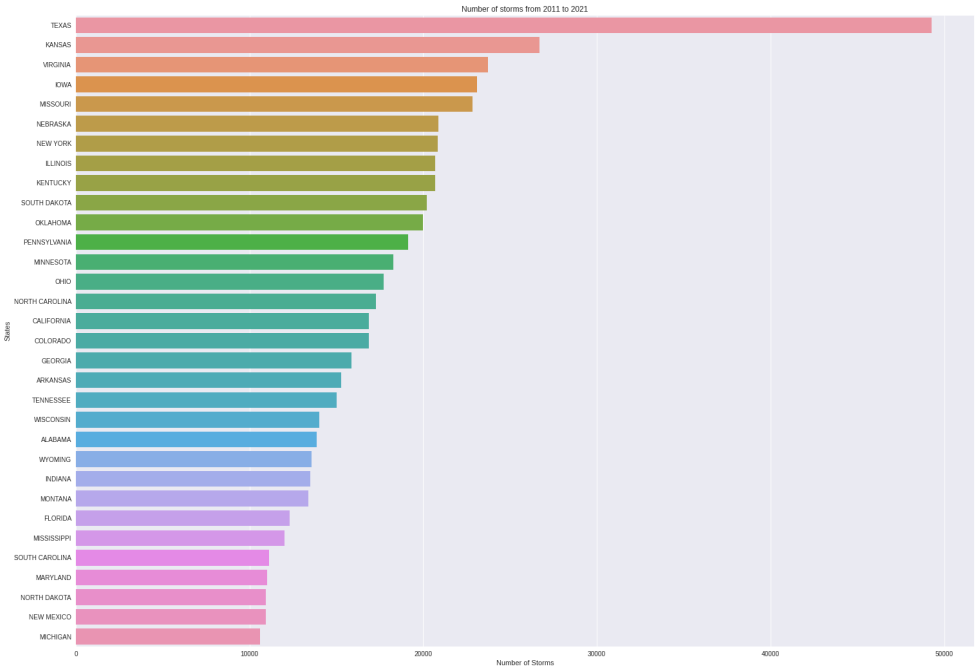
Results


Total damage by event type from 2011 to 2021
Flash Flood
Hurricane
Tornado
Wildfire
Flood
Coastal Flood
Hail
Drought
High Wind
Tropical Storm
Storm Surge/Tide
Thunderstorm Wind
Frost/Freeze
Event Type
100
200
300
400
500
600
700
Damage Cost (Billions)


Damage caused by storms from 2011 to 2021
TEXAS
LOUISIANA
NEW JERSEY
CALIFORNIA
PUERTO RICO
FLORIDA
COLORADO
GEORGIA
NORTH CAROLINA
IOWA
ALABAMA
MISSOURI
TENNESSEE
MICHIGAN
OREGON
ILLINOIS
OKLAHOMA
MISSISSIPPI
NEBRASKA
NEW YORK
GUAM
WASHINGTON
ARKANSAS
VERMONT
SOUTH DAKOTA
SOUTH CAROLINA
OHIO
KANSAS
PENNSYLVANIA
States
100
200
300
400
500
600
700
Damage Cost (Billions)


Sex:M | Event: Tornado
Sex:M | Event: Winter Weather
Sex:M | Event: Thunderstorm Wind
Sex:M | Event: Heat
Sex:M | Event: Excessive Heat
Sex:F | Event: Tornado
Sex:F | Event: Winter Weather
Sex:F | Event: Thunderstorm Wind
Sex:F | Event: Heat
Sex:F | Event: Excessive Heat
Ocurrences
0
20
40
60
80
100
Death Age


Most life threatening events from 2011 to 2021
Tornado
Excessive Heat
Heat
Winter Weather
Thunderstorm Wind
Lightning
Ice Storm
Wildfire
Winter Storm
Dense Fog
High Wind
Heavy Snow
Hurricane
Flash Flood
Heavy Rain
Rip Current
Dust Storm
Strong Wind
Debris Flow
Avalanche
Flood
Blizzard
High Surf
Tropical Storm
Freezing Fog
Marine Thunderstorm Wind
Marine Strong Wind
Cold/Wind Chill
Lake-Effect Snow
Extreme Cold/Wind Chill
Sneakerwave
Event type
Injuries
Deaths
0
2000
4000
6000
8000
10000
12000
Ocurrences



Insights

- **Texas** is by far the state with more events and damage caused by storms,
- **Texas** has more than 70 billion in damage in the past 10 years.
- The event that causes more damage is **Flash Flood** with more than 75 billions in damage.
- **Tornados, Winter Weather, Thunderstorm Winds, Heat and Excessive Heat** are the events that caused more deaths.
- **Heat** kills more infant and elderly people

References

- DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce
- The Severe Weather Data Inventory (SWDI): a Geospatial Database of Severe Weather Data at the NOAA National Centers for Environmental Information (NCEI)