## Step Two: Project Proposal

**Problem Statement:**

In using the *NOAA: Severe Storm Even Details* dataset (see below), this Capstone Project seeks to analyze and visualize the evolution of severe weather events in the United States (on a year-over-year basis), across various event types.

In conjunction with secondary datasets, it may then be possible to assess the impact of severe weather events on homeownership and/or home insurance premiums to help determine the following:

- How is the occurrence of severe weather events evolving? Is it possible to observe an evolution, in which certain areas / states have recently become more prone to such events?
- Of the various event types, are certain weather events occurring with greater frequency?
- Are individuals and/or families migrating to states where the occurrence of severe weather events is greatest?
- Are individuals and/or families migrating to states where home insurance premiums are rising the fastest due to the greater risk of severe weather events occurring?

**Data Source(s):**

Primary: NOAA: Severe Storm Event Details
    Update Frequency: **Monthly**

Secondary:
- Housing Vacancies and Homeownership by State (1995 – 2022)
    Update Frequency: **Annually**

- Average Premiums For Homeowners And Renters Insurance By State (2007 – 2020)
    Update Frequency (*with a multi-year lag)*: **Annually**

- US Residential Real Estate Data
    Update Frequency: **Monthly**

**Data Cleaning & Joining Summary:**

First and foremost, to isolate subsets of data from the datasets listed above to ensure that the information spans the same period, and to aggregate the information at the same level (based on the data sources listed above, some data is provided at a state level, whereas others provide more granular information at a zip-code level). As such, an additional dataset may have to be referenced (*Zip Code Tabulation Area (ZCTA)* dataset) for the purposes of data aggregation.

**Tools / Technologies to be Used:** *TBD*

- Apache Airflow and Spark
- Azure
- Tableau

**Project Goals:**

1. Establish a data pipeline that is capable of aggregating data from each of the sources listed above, which can be scheduled to retrieve information in accordance to when each dataset is updated.

2. Analyze the year-over-year trends in severe weather events and identify patterns of evolution in frequency, intensity, and types of severe weather events.

3. Determine if areas that are considered at higher risk of severe weather events are also areas where people are moving to and/or are experiencing the most acute rise in home insurance premium.

4. Create visualizations to illustrate these trends.

5. *Post Data Engineering:* Develop predictive models to forecast future trends in severe weather event occurrence and their impact on home insurance premiums.