1. The original data set, us\_disaster\_declarations.csv, was loaded into Jupyter Notebook and filtered using Pandas. Columns housing data that was inconsequential to the analysis were removed from the data set.

(pic: cleaningdata\_removecolumns)

1. Outlier incident types that were at risk of skewing the results in this analysis were removed from the dataset.

(pic: cleaningdata\_remove\_outlier\_disasters)

1. The US territories, American Somoa, Guam, Northern Mariana Islands, Puerto Rico and U.S. Virgin Islands, were removed from the dataset due to the propensity for these areas to have significantly more disastrous weather conditions than the rest of the country.

(cleaningdata\_removeterritoriesskew)

1. The date strings were changed into proper date format.

(pic: cleaningdata\_changingdateformats)

1. A new column was created to house the numeric month value each incident began so analysis could be conducted on what time of the year incidents occurred.

(pic: cleaningdata\_creatingmonthcolumn)

1. A new column was created, incident\_duration, by subtracting the incident\_begin\_date column from the incident\_end\_date column This allows analysis to be conducted on the duration of each incident. The data type of the column was then changed from an object to an integer.

(cleaningdata\_creatingincidentdurationcolumn)

1. All zero values in the incident duration column represent incidents that were less than a day in length and were changed to the number 1 in the data set.

(pic: cleaningdata\_changingzerovaluestoone)

1. A new column was created to house the year each incident began.

(cleaningdata\_creatingcolumnforyear)

1. To conduct analysis on the US regions the disasters occurred in, a dictionary was created that contained the state abbreviations as well as their respective regions. The function “get\_region” was created using .apply() to match the region to the states in a new column called “regions”.

(cleaningdata\_createdictforregions) (cleaningdata\_createdictforregions2)

1. A new column was created that combined the values from all four “program\_declared” columns in the dataset.

(cleaningdata\_programsdelcared)

1. Additional columns that were proven to be obsolete to this analysis were dropped from the data set.

(cleaningdata\_droppingadditionalcolumns)

1. A new dataframe was created to couse the disaster number, incident type, incident month, incident begin year and incident duration.

(cleaningdata\_creatingincidentdeclarationdf)

1. A new data frame was created to house the disaster number, incident type, designated area, state, region and programs declared columns.

(cleaningdata\_creatingincidentlocationdf)