Federal Emergencies and Disasters

Contents

Data description

Natural disasters are so sudden, devastating, and unavoidable. We cannot change the nature of disasters, but we can be prepared as we understand their scope. Analysis of previous data would help us in this endeavour in terms of cost incurred for each type of disaster and flow the efforts towards the areas that are affected the most.

For this report I am using the Federal Emergencies and Disasters dataset from Kaggle. This dataset holds the data for each state and their counties, the type of declarations made by the President in the face of these disasters and the duration of such declarations. The data presented here has declarations from 1953 to 2017

Data Overview

The data contains 14 columns as described below

1. Declaration number – Serves as an identifier for the declaration by President
2. Declaration type – This is a categorical variable defining the type of declaration (Disaster, Emergency, Fire)
3. Declaration Date – The date of declaration
4. State – The 2-letter code of the state for which declaration was issued (OH, TX etc.)
5. County – The county for which the declaration was issued
6. Disaster type – Type of disaster (Hurricane, Flood, Storm etc.)
7. Disaster title – This column is similar to disaster type, however this single columns holds multiple comma separated values which add more details on the type of disaster
8. Start Date – Start date of the disaster
9. End Date – End date of the disaster
10. Close Date – Closing date of the declaration
11. Individual Assistance Program – A categorical column with only two values Yes/No, which mentions if assistance is provided to the individuals
12. Individual and Households program – A categorical column with only Yes/No to mention if assistance is provided to both individuals and the entire households
13. Public Assistance Program – A categorical column with only Yes/No to mention if assistance is provided for public units

14. Hazard Mitigation program – Another categorical column with yes/No to mention if mitigation measures are taken during this period.

Normalized Data

The data in this data set is not normalized. It can possibly be normalized into the following tables

- DeclarationDetails
- DisasterDetails
- LocationDetails
- AssistanceDetails

Data Consistency

- The declaration numbers do not start from DR-1 or EM-1, which indicate that there might have been some missing entries about which we need to be cautious about.
- County, EndDate, CloseDate have missing values.
- Though there are 46,185 rows in the data set, there are only 3427 distinct declaration numbers. This indicates that there are multiple entries for each declaration filling in different details.

Conclusion

- Accompanied with this report is an SQL file and a Tableau workbook.
- The dataset did not have any numerical data to form aggregations for.
- We can create heatmaps, treemaps and barcharts from the data.
- From the SQL commands run as part of this project, we can see that the year 2005 has the highest number of disasters.
- Storm is the most common disaster type