Speed Data-ing: Hurricane Harvey

FEMA director, William Long, says Harvey is probably the worst disaster in Texas history. The top rainfall total was over 50 inches, and this rainfall total would be the heaviest from any tropical cyclone in the continental U.S. in records dating to 1950. (Washington Post). Thousands of water rescues occurred in the Houston metro area, and Houston was not the only area that was affected.

We created a word cloud of the tweets provided and realized that a lot of what was being said wasn't relevant to what things we thought we might need for this project. Our team then looked at the other CSV files of needs and shelters, and realized with that information, we could make a customized box for each shelter. When we looked at competing emergency boxes on the market, we found that they didn't provide some basic needs required by each shelter, and so we felt that a box made for each shelter would be most useful.

Current solutions for similar products are very expensive ($100 - $300) and are created for personal use by 1 - 4 people, not the shelter. To find the item needs we decided to find what shelters said they needed in the data set we were provided at the competition.

By making word clouds and using natural language processing, we made programs for frequency analysis and word filtration with Python, Power BI, and R. We also used an API called DataMuse to create buckets of similar unbiased keywords so that our results truly reflected our data.

The results we got from our frequent words analysis helped us create 5 distinct buckets that represent various types of boxes: baby, first aid, sanitary, food, other. We used those keywords on the data set and found a consistent ratio of those buckets in unofficial and official shelter needs.

Some problems we encountered were many words that didn't belong in the reports, and null fields in our lists of needs. One way that we could solve this problem in the future is to make checklists for future disaster response needs for shelters, so that people, when panicked, have less opportunity to make spelling mistakes or name specific products. By making a checklist, we

can also have these uploaded directly into a database like Apache Cassandra, and have aggregations performed automatically, and then have a custom box built faster for each shelter. This would also make boxes created even more cost and time efficient, and potentially save more lives.

Citations:

https://www.washingtonpost.com/national/fema-director-says-harvey-is-probably-the-worst-disaster-in-texas-history/2017/08/27/ef01600a-8b3f-11e7-8df5-c2e5cf46c1e2_story.html

https://www.nytimes.com/2017/08/27/us/harvey-pictures-hurricane-storm.html?mcubz=1

https://data.world/aegorov/carolinadatachallenge