VAST 2019 Mini Challenge 1: Crowdsourcing for Situational Awareness

Faiz Ali Gola
MS Information Technology and Analytics
Rochester Institute of Technology
New York, USA
fg6340@rit.edu

Gulnaaz Shaikh
MS Information Technology and Analytics
Rochester Institute of Technology
New York, USA
gs3058@rit.edu

Vikashini Sabarinathan
MS Information Technology and Analytics
Rochester Institute of Technology
New York, USA
vs3224@rit.edu

***Abstract*- This paper provides a complete analysis for VAST 2019 Mini challenge 1. St. Himark, a fictional city, is affected by earthquakes in the month of April. The government has released an application through which people input their perception of the earthquake intensity in different areas. Given the dataset, we have analyzed areas of the fictional city St. Himark are most affected, along with the localities that are providing reliable reports. The dataset cleaning and visualization are done using Tableau and Python. Various conclusions are drawn based on the results of these visualizations.**

Keywords—data visualization, analysis, Tableau

I. Introduction

This paper outlines Mini Challenge 1: Crowdsourcing for situational awareness. The challenge revolves around the fictional place called St. Himark, where the government has released an application for people to input their responses in case of an emergency like earthquakes, etc. The city is sub-divided into 19 locations from where the data is collected. Each person input's their response on a scale of 0 to 10, where 10 is the highest intensity of the earthquake. Subsequently, these readings are further analyzed to provide relief based on the severity of the situation.

II. Data Description

A. CSV File

The data for MC1 includes one (CSV) file spanning the entire length of the event, containing (categorical) individual reports of shaking/damage by neighborhood over time. Reports are made by citizens at any time; however, they are only recorded in 5-minute batches/increments due to the server configuration. Furthermore, delays in the receipt of reports may occur during power outages.
mc1-reports-data.csv fields:
- time: timestamp of incoming report/record, in the format YYYY-MM-DD hh:mm:ss
- location: id of neighborhood where person reporting is feeling the shaking and/or seeing the damage
- {shake_intensity, sewer_and_water, power, roads_and_bridges, medical, buildings}: reported categorical value of how violent the shaking was/how bad the damage was (0 - lowest, 10 - highest; missing data allowed)

B. Shakemaps

Also included are two shakemap (PNG) files which indicate where the corresponding earthquakes' epicenters originate as well as how much shaking can be felt across the city

C. Map

The following represents the 19 different locations of St. Himark city.

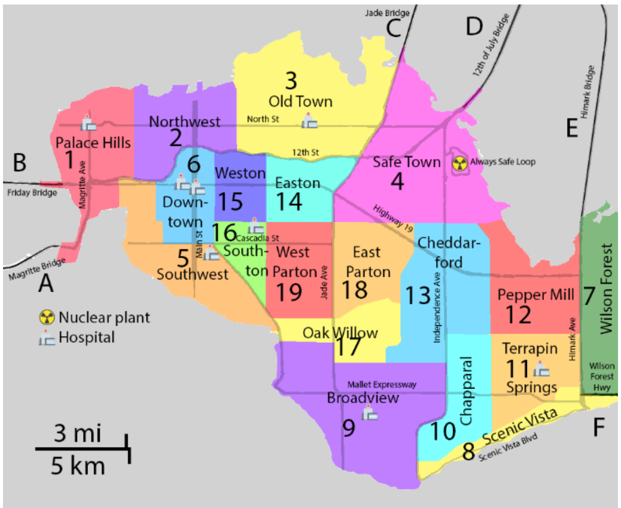

Fig 1: Map of St. Himark

III. Methodology

The main questions we are trying to address here as follows:

- 1. How would you prioritize neighborhoods for response? Which parts of the city are hardest hit?
- 2. Compare the reliability of neighborhood reports. Which neighborhoods are providing reliable reports?

To answer these questions, we have utilized python libraries to clean and sort the data as required, followed by the implementation of Tableau Workbook to visualize solutions for these problems.

Furthermore, answering the questions like how should we prioritize neighbourhoods, we have calculated the averages of the shake intensities at buildings, roads and bridges, and the actual shake intensity. Moreover, these visualizations (Fig 2) can be used to derive conclusions and answers to our main problem. Coming to the second part of the question, we analysed all the parts of the city for averages of buildings, medical, power, sewer and water, roads and bridges, along with the actual shake intensities. The shake intensities give us the overall idea of how bad or worse the intensity of the earthquake was in a particular locality.

Subsequently, for the second question, we visualized the parts of the city based on averages of reports generated in every location. Just by looking at Fig 4 we can analyze further which areas were most affected. Furthermore, the Fig 3 can also help us understand the anomalies present in the dataset which are discussed further.

IV. RESULTS

A. *How would you prioritize neighborhoods for response? Which parts of the city are hardest hit?*

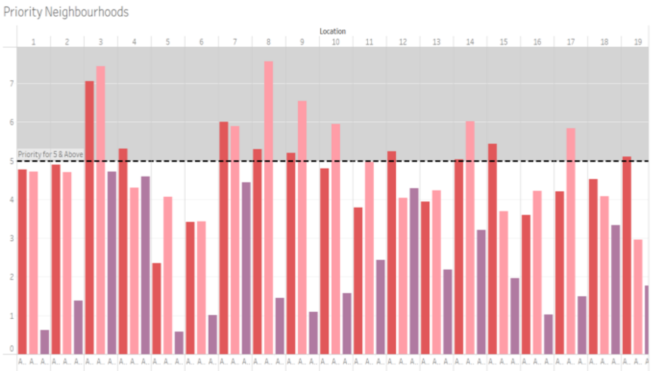

Fig 2: Priority Neighbourhood Visualization

We analysed this visualization for Averages of Roads and Bridges, Sewer and Water and the Shake intensity for each of the 19 locations in St. Himark. To prioritize the locations based on this visualization, we found it interesting to presume that a intensity 5 or greater could be a high intensity earthquake and therefore, emergency responders should prioritize such areas.

Fig 3: Location Wise Reporting of averages

From Fig 3 we can get a clearer picture of which areas are affected the most in every category, along with how some areas have a high shake intensity but reports have not been generated as much, perhaps due to chaos and mishap in such areas.

B. *Compare the reliability of neighborhood reports. Which neighborhoods are providing reliable reports?*

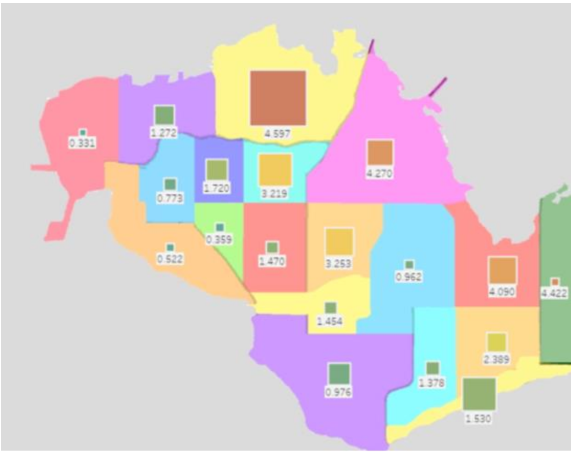

Fig 4: St. Himark map with location wise reporting averages

The reliability of reports can be figured using Fig 4, where we can obtain the number of reports generated in each location, for example, Locations 3, 8 and 12 have generated high number of reports, but that does not necessarily mean that those locations were the most affected. From Fig 3, we can conclude that Location 8, even though had high volume of reports made, the shake intensity was below 5.

V. CONCLUSION

In conclusion, we can say that using Tableau we have visualized the given data to find answers to two questions from VAST 2019- Mini Challenge 1 (Crowdsourcing for situational awareness). Firstly, we found that based on people's response the localities that should be prioritized are Location 3- Old Town and Location 8- Scenic Vista. However, when we visualized the actual shake intensities for these locations, it can be concluded that Location 8 had high number of reports but the intensity of the earthquake was significantly lower. These anomalies can be found in the data due to the perception of individuals. The intensity of earthquakes can vary for individual to individual. Location 3 was the most affected area in St. Himark.

Furthermore, we have also compared the reliability reports of the neighbourhoods. People from Location 8- Scenic Vista, and Location 9- Broadview, were responding that the earthquake was of a high intensity (Figure 3). But if we look at the shake intensity the earthquake was not that bad. Similarly, Location 7- Wilson Forest had little to no input in the sewer and water damages of that locality but as we observe the earthquake occurred was of a high intensity. These neighbourhoods were providing unreliable reports. Subsequently, Location 3- Old Town was providing reliable reports as it had high scale intensities recorded by both people as well as the actual shake intensity.

In future, we would like to explore more data, perhaps with time and analyse the results based on every hour of input received. This would significantly impact the emergency responder's decisions and allow better disaster management techniques.

REFERENCES

[1] VAST Challenge 2019: Mini-Challenge 1 dataset [Data file]. https://visualdata.wustl.edu/varepository/VAST%20Challenge%202019/challenges/Mini-Challenge%201/dataset/MC1.zip

[2] VAST challenge 2019: Disaster at St. Himark! https:// vast-challenge.github.io/2019/. Accessed: 2019-10-19.

[3] Visual analytics benchmark repository. https://www.cs.umd.edu/hcil/varepository/benchmarks.php. Accessed: 2019-10-19.

[4] Adulyanukosol, N. (2019). Earthquake Damage Report Interactive Dashboard Using Bayesian Structural Time Series and Value-Suppressing Uncertainty Palettes. https://doi.org/10.1109/vast47406.2019.8986916

[5] Scott-Brown, J. (2019). *VAST Challenge 2019: Mini-Challenge 1* [Video]. Vimeo. https://vimeo.com/347855541

[6] J. Scott-Brown, "Visualization of Spatial-Temporal Distribution of Damage Reports," 2019 IEEE Conference on Visual Analytics Science and Technology (VAST), Vancouver, BC, Canada, 2019, pp. 146-147, doi: 10.1109/VAST47406.2019.8986929.