Enhancing Emergency Response Efficiency in San Francisco

Anirudh Garg
Courant Institute of Mathematical Sciences
New York University
New York, USA

Haardik Dharma
Courant Institute of Mathematical Sciences
New York University
New York, USA

Rhea Chandok
Courant Institute of Mathematical Sciences
New York University
New York, USA

***Abstract*—**

In this paper, we analyze and gain actionable insights into the efficiency of emergency response times and patterns related to fire incidents and emergency medical services (EMS) in San Francisco. Utilizing publicly available datasets on dispatched calls, fire incidents, and EMS response times, we conduct a comprehensive analysis to assess response efficiency, identify common emergency types, and uncover temporal patterns in incident frequency. The findings from this study provide valuable insights to assist public safety departments and policymakers in enhancing preparedness, optimizing resource allocation for emergency services in San Francisco, and helping residents avoid high-incident areas.

Keywords*—*analytics, fire, calls, ems, response times, San Francisco

I. Introduction

Emergency response services play a critical role in ensuring public safety and minimizing the impact of incidents in urban environments. In San Francisco, a city known for its diverse geography and dense population, the efficiency of fire and emergency medical services (EMS) is paramount. This project focuses on analyzing the performance and patterns of these essential services to provide data-driven insights for improvement.

In our analysis, we delve into various aspects of emergency response and incident patterns in San Francisco, leveraging the rich datasets available. Our comprehensive approach examines both fire and medical emergency services, providing a multifaceted view of the city's emergency response landscape.

The study encompasses several key areas of investigation, including response time efficiency across different regions for both medical and fire emergencies, the geographical distribution of fire incidents, and the prevalence of various emergency types. We also explore temporal trends, analyzing yearly and monthly patterns in emergency occurrences to identify potential seasonal influences or long-term shifts. Additionally, our research examines the distribution of call types and their regional variations, offering insights into the diverse nature of emergencies across San Francisco's neighborhoods. To assess the human impact of fire incidents, we investigate fatalities and injuries concerning specific zip codes. Furthermore, we conduct a comparative analysis of turnaround times for medical and fire responses, including exploring factors influencing fire response durations. This comprehensive approach aims to provide a holistic understanding of emergency services in San Francisco, uncovering areas for improvement and informing strategic decision-making in public safety management.

Our approach leverages big data technologies to handle the volume and complexity of the data. We employ Hadoop Distributed File System (HDFS) for efficient storage and accessibility and utilize MapReduce for data profiling, cleaning, and analysis. This allows us to process large-scale datasets and extract meaningful patterns and insights. We use Trino to run queries on our data and Tableau to visualize the results of our analysis.

The findings from this analysis have the potential to significantly impact public safety strategies and community well-being in San Francisco. By identifying areas for improvement in response times, understanding the distribution of incident types, and recognizing temporal patterns, emergency services can better prepare for and respond to incidents. This project aims to contribute to the enhancement of public safety and the optimization of emergency response resources in San Francisco, with several potential benefits:

1. **Improved resource allocation:** Emergency services can strategically distribute personnel and equipment based on identified high-risk areas and peak incident times.
2. **Enhanced community awareness:** Residents and newcomers can make informed decisions about housing choices by accessing data on emergency incident patterns in different neighborhoods.
3. **Data-driven policy making:** Local governments can use insights on incident distributions and response efficiencies to inform urban planning and public safety policies.
4. **Public health insights:** Patterns in medical emergencies can provide valuable data for public health officials to address recurring health issues in different communities.
5. **Business planning:** Companies considering relocation or expansion in San Francisco can factor in emergency response data when choosing locations.

Ultimately, this project seeks to create a safer, more resilient San Francisco by leveraging data analytics to improve emergency preparedness and response across the city.

II. Motivation

Emergency response services are the lifeline of urban safety, yet their effectiveness can be significantly enhanced through data-driven insights. Emergency incidents rarely occur randomly. The vast archives of historical emergency data contain valuable patterns and trends that can revolutionize how we approach public safety. By leveraging these extensive datasets, we can uncover hidden factors influencing emergency response times, incident frequencies, and outcomes that may not be immediately apparent to even the most experienced first responders.

III. Related Work

The analysis of emergency response efficiency has been a growing area of interest due to its potential to improve public safety outcomes and optimize resource allocation. Several studies have explored various aspects of emergency response using data analytics techniques. This section highlights some of the key research relevant to our project.

Predictive modeling for Emergency Medical Services (EMS) has gained significant attention in recent years. Researchers have employed various machine learning and data mining techniques to forecast EMS demand and optimize resource allocation. The Predictive Model for Emergency Medical Services (EMS) using Machine Learning and Data Mining paper [1] developed a predictive model for EMS using machine learning algorithms such as Random Forest and decision trees. Their approach demonstrated the potential of these techniques in predicting EMS demand patterns and improving response times.

In the realm of ambulance allocation optimization, Acuna et al. [2] proposed a model to address the overcrowding problem in US emergency departments. Their research focused on a case study in Florida and utilized a bi-objective mixed-integer programming model to optimize ambulance allocation while considering both patient waiting times and hospital overcrowding. This study highlighted the importance of efficient resource allocation in improving emergency response services.

The impact of response times on fire outcomes has been the subject of significant research. The paper “Statistical Analysis of Fire Department Response Times and Effects on Fire Outcomes in the United States” [3] conducted a statistical analysis of fire department response times and their effects on fire outcomes in the United States. The study revealed a strong correlation between shorter response times and improved fire outcomes, emphasizing the critical nature of rapid emergency response.

Recent research has also focused on analyzing national EMS delays. A comprehensive study by the Journal of Emergency Medical Services [4] examined the frequency and types of EMS delays from 2017 to 2022. This research provided valuable insights into the various factors contributing to response time delays, including dispatch issues, traffic conditions, and resource availability. Understanding these delay patterns is crucial for developing strategies to improve overall EMS performance.

Our project builds upon these works by integrating multiple datasets to provide a comprehensive analysis of emergency response efficiency and incident patterns in San Francisco.

IV. Design and Implementation

A. Design Details

Figure 1 shows the design details of our project. Our project follows a systematic approach to process and analyze emergency response data from San Francisco. The design consists of multiple stages, from data ingestion to visualization, leveraging big data technologies for efficient processing. The third step was to generate Impala tables, so we could access and query the data seamlessly. The implementation pipeline consists of several key stages:

1. **Data Ingestion and Storage:** Three primary datasets (described in part B of this section) from SF.GOV are ingested into HDFS.
2. **Data Processing and Cleaning:** MapReduce jobs process each dataset to remove invalid or incomplete records, standardize location data and timestamps, transform data into consistent formats, and generate part files for each dataset.

Details:

Fire and EMS Dispatched Calls Dataset

The dataset underwent profiling and cleaning to ensure structured and consistent data for analysis, leveraging MapReduce jobs. Initially, unnecessary columns were dropped, retaining only key fields like *Call Number*, *Incident Number*, *Call Type*, *Call Date*, and others, narrowing the dataset to 20 essential columns. Data transformation included standardizing date formats to *YYYYMMDD* and unifying city names, consolidating variations like *SFO* and *San Francisco* under a single name. Records were filtered to include incidents between January 1, 2014, and October 31, 2024, reducing the dataset from 900,000 to 520,000 rows.

A value count analysis was conducted on critical columns, such as *Call Type* and *Unit Type*, using custom MapReduce programs. For instance, *Call Type* revealed 15 unique categories, with "Medical Incident" being the most frequent (65,000 occurrences). Similarly, *Unit Type* analysis identified "Engine" as the predominant type, accounting for 40% of the records. These operations provided a comprehensive understanding of the data structure and content, ensuring a robust foundation for further analysis.

Fire Incidents Dataset

The dataset, originally containing 66 columns, was cleaned to retain 30 relevant attributes by removing columns unrelated to emergency response (e.g., *ignition_cause*, *area_of_fire_origin*). Data was filtered to include incidents from January 1, 2014, to October 31, 2024, reducing records from 684,000 to 349,000.

Key metrics, including fatalities, injuries, response times, and turnaround times, were profiled using MapReduce. Analysis revealed a right-skewed distribution, with mean response times (9.39 minutes) significantly higher than the median (4.65 minutes) and similar trends for turnaround times (mean: 56.41 minutes, median: 17.50 minutes), indicating outliers from rare, delayed incidents.

The cleaned dataset enabled focused analysis on response efficiency, removing noise from irrelevant or incomplete data.

EMSA Response Times Dataset

The dataset underwent cleaning and profiling to ensure validity and readiness for analysis. Unnecessary columns like *rowid*, *data as of*, and *data loaded at* were removed, reducing the dataset to essential fields. Records were filtered to include only those from January 1, 2014, to October 31, 2024. Rows with missing or invalid values in the *response time min* column were excluded, dropping 12,345 rows due to invalid dates and 8,910 rows due to missing or invalid response times. Using MapReduce, the data cleaning step involved a Mapper to parse and filter records while tracking dropped rows and a Reducer to aggregate the cleaned data, retaining 678,432 valid rows for analysis.

The profiling step used another Mapper-Reducer pair to compute key statistics for *response time min*. The statistics revealed a minimum response time of 2 minutes, a maximum response time of 256 minutes, an average response time of 34.8 minutes, and a standard deviation of 15.3 minutes. These outputs provided insights into response efficiency, with cleaned data and profiling statistics highlighting the distribution and quality of the dataset.

3. **Data Integration:** Cleaned part files are then used to create Hive tables. A derived table is generated by combining relevant information from all three datasets. Key fields are joined to enable comprehensive analysis across datasets.
4. **Analysis and Visualization:** Processed data is exported to Tableau for visualization and analysis. Various metrics are calculated including - Response time distributions, Incident type frequencies, Geographical patterns, Temporal trends, etc.

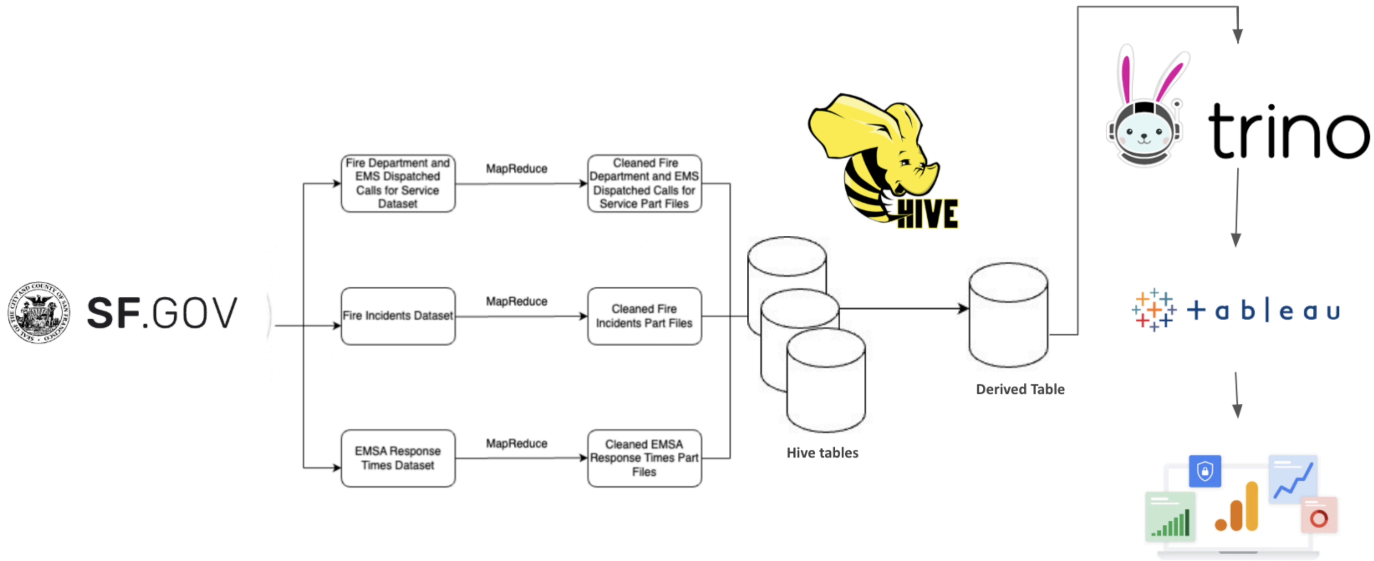

Figure 1: Flow diagram of the project

B. Description of Datasets

As the datasets by SF.GOV are updated daily, we cleaned all of our datasets to only include records from the last decade, specifically from January 1, 2014, to October 31, 2024.

1. Fire and EMS Dispatched Calls Dataset

The dataset "Fire Department and Emergency Medical Services Dispatched Calls for Service" provided by the City of San Francisco is a valuable resource for analyzing emergency response patterns in the city. It has a total size of 2.6 GB and is publicly available here. It contains detailed information on each call, including timestamps and dates, response times, neighborhood, and call type (Fire, Medical, etc.)

Key fields in the dataset include call number, incident number, call type, response times, address, and priority level.

2. Fire Incidents Dataset

The "Fire Incidents" dataset records specific fire-related incidents and outcomes that the San Francisco Fire Department responded to. This dataset provides a comprehensive overview of non-medical incidents and includes crucial details that help in understanding fire-related emergencies in the city.

Key fields in the dataset include incident number, Alarm date time, Arrival date time, Close date time, and zip code, among others. The dataset is publicly available here.

3. EMSA Response Times Dataset

The "EMSA Response Times Dataset" publishes information on the time it takes for emergency response vehicles to arrive at the scene of a medical incident after it is dispatched. This dataset is derived from the Fire Department Calls for Service dataset and includes responses to 911 calls for service from the city's Computer-Aided Dispatch (CAD) system.

Key fields in the dataset include incident number, call date, and response time in minutes. The dataset is publicly available here.

V. Results

We first looked at the top emergency types. The result is shown in Figure 2. We can see that medical and fire alarms are the most common type of incidents.

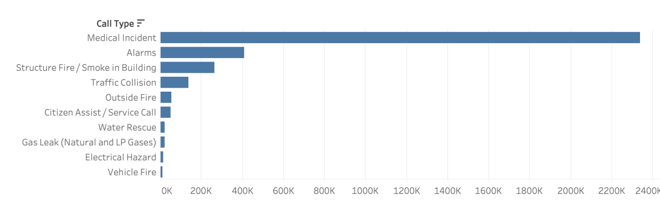

Figure 2: Top 10 Call Types

Further, we looked into the temporal trends in Fire and Medical Emergencies as shown in Figures 3, 4, and 5. In yearly trends, an interesting thing to note here is a slight dip in the year 2020-21. This was because most of the emergencies during this time were due to the COVID-19 pandemic, and that data is not included in this dataset.

Subsequently, in monthly trends, it is apparent that there are more fire emergencies in winter months as compared to summer months as shown in Figure 4. Heating, holiday decorations, winter storms, and candles all contribute to an increased risk of fire during the winter months.

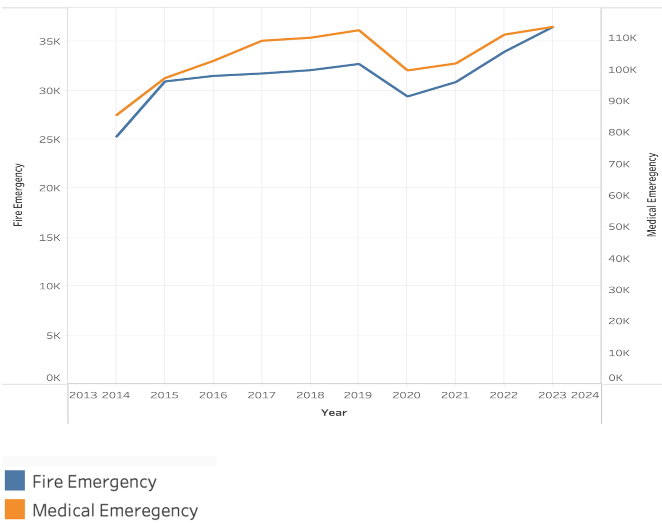

Figure 3: Yearly Trend in Fire and Medical Emergencies

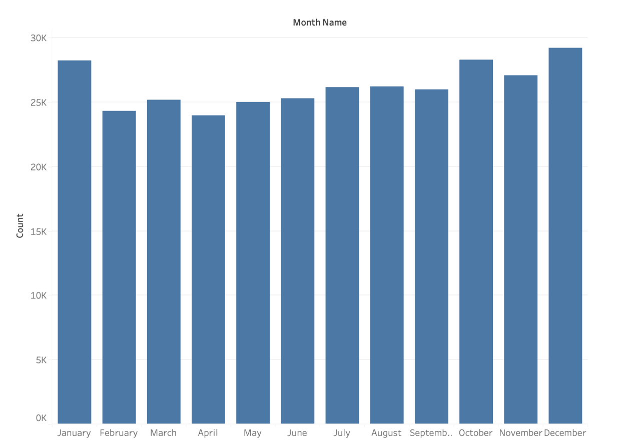

Figure 4: Monthly Trend in Fire Emergencies

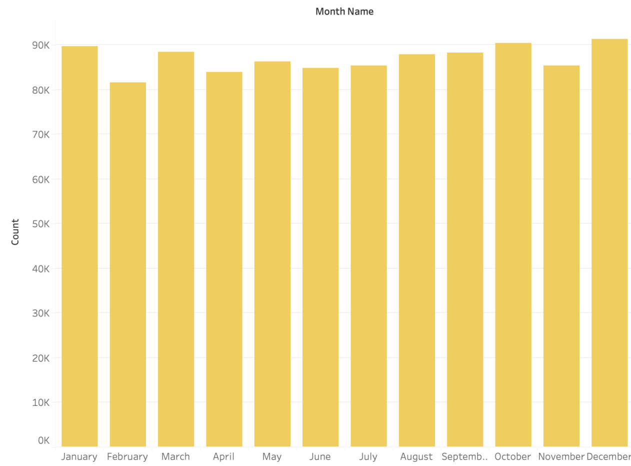

Figure 5: Monthly Trend in Medical Emergencies

Furthermore, we analyzed the life-threatening incidents by region in San Francisco as depicted in Figure 6. We can see a higher proportion of life-threatening incidents in the northeast regions of the city. Some of the prominent neighborhoods in those parts are shown in Figure 7.
These areas are characterized by a combination of:

- High population density, transient populations, and socio-economic challenges.
- Limited accessibility for emergency response vehicles in densely packed urban environments.

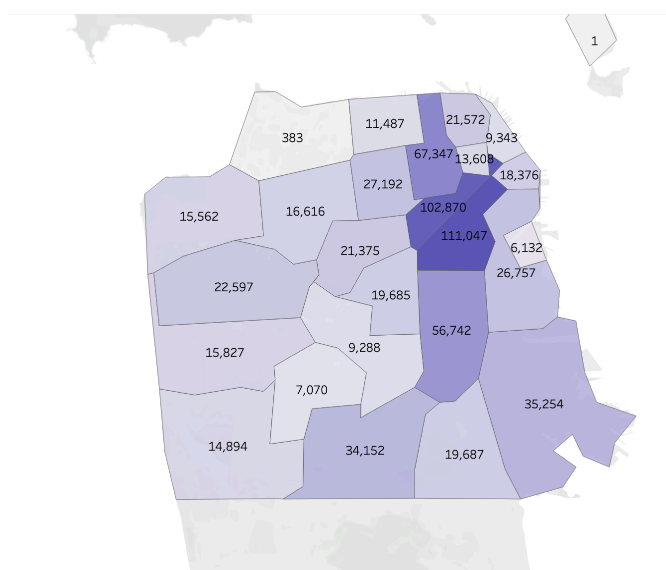

Figure 6: Potentially Life-Threatening Incidents by Area

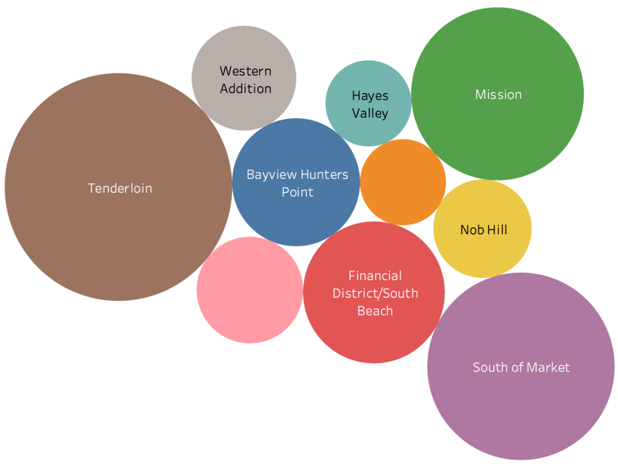

Figure 7: Potentially Life-Threatening Incidents by Neighborhood

Next, we analyzed the average response time in minutes for each neighborhood as shown in Figure 8 (Lighter colors correspond to a faster response, thus lower response times).

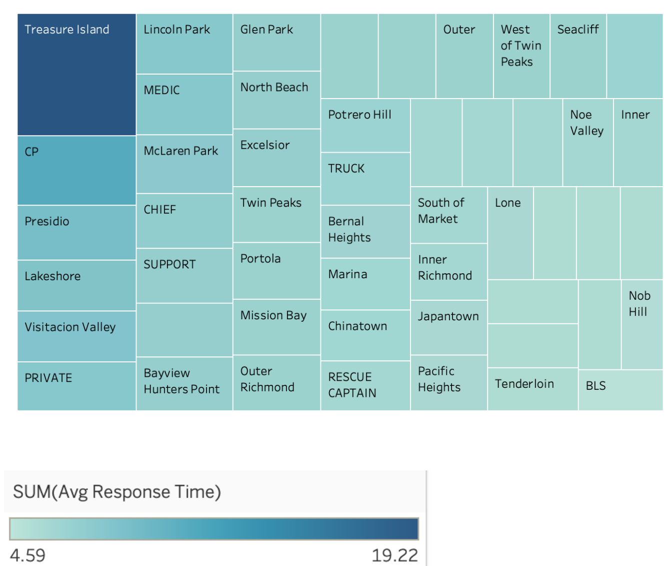

Figure 8: Average Response Time (Minutes) by Neighborhood

It can be logically derived that the worst neighborhoods would have high incident counts and higher response times. Thus, as shown in Figure 9, we plotted them to highlight the importance of strategic resource allocation and localized response units.

Neighborhoods with both high incident counts and elevated average response times include:

- Bayview Hunters Point, Financial District/South Beach, and Sunset/Parkside: Urban zones with high traffic congestion and complex layouts.
- Suburban and Peripheral Areas such as Benicia, Muir Beach, Alameda, San Bruno, San Mateo: Limited proximity to emergency response hubs and extended travel times contribute to delays.

These findings highlight the importance of strategic resource allocation and localized response units.

On the other hand, it would be useful to analyze which would be the best and safest neighborhoods to reside in. Naturally, these would be the ones with low incident counts and low response times. Such neighborhoods are shown in Figure 10 and Figure 11.

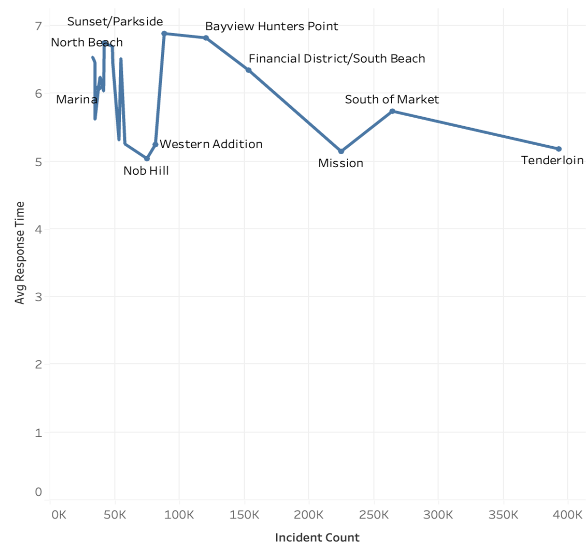

Figure 9: Response Times vs Count of Incidents by Neighborhood

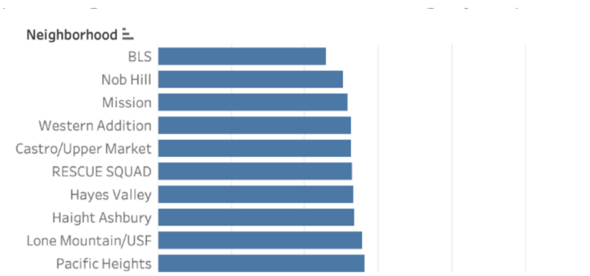

Figure 10: Neighborhoods with the fastest emergency response time

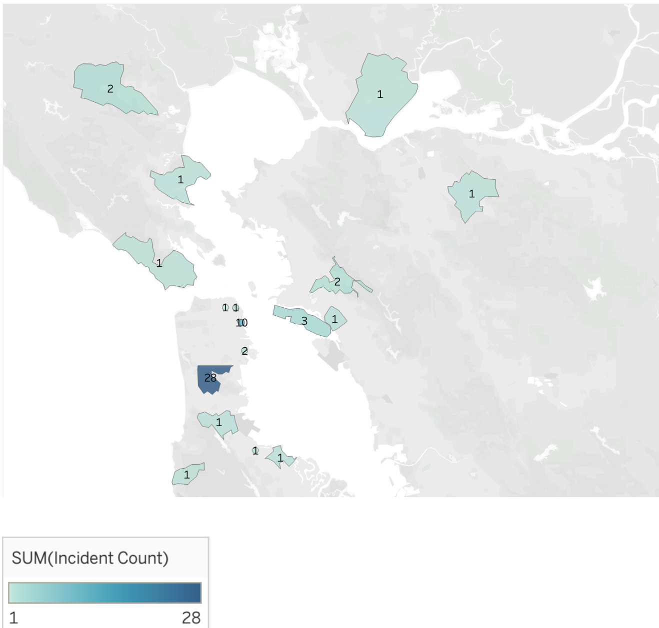

Figure 11: Neighborhoods with Least Number of Emergencies

VI. Future Work

While this study provides a solid foundation for understanding emergency response patterns in San Francisco, there are several avenues for future research to further enhance public safety and optimize resource allocation.

By leveraging advanced machine learning techniques, we can develop predictive models to forecast emergency demand patterns. This would enable proactive resource allocation, allowing emergency services to be strategically positioned in anticipation of peak demand periods. Exploring spatial statistics can help identify clusters of high-incident areas and potential underlying causes, such as building codes, infrastructure deficiencies, or socio-economic factors. This granular analysis can inform targeted interventions and resource allocation strategies.

A deeper analysis of medical emergency patterns can uncover correlations between specific neighborhoods and health conditions. These insights can be valuable for public health officials to implement targeted interventions and improve overall community health. Also, incorporating real-time data sources, such as traffic information and weather conditions, can significantly enhance emergency response planning and execution. By considering dynamic factors, emergency services can optimize dispatch decisions and response routes.

Further, developing a user-friendly platform that provides access to emergency response data and safety resources can empower residents to make informed decisions about their safety and well-being. This platform could include neighborhood-specific incident statistics, safety tips, and emergency contact information.

By pursuing these research directions, we can further optimize emergency response operations in San Francisco, leading to a safer and more resilient city.

VII. Conclusion

This study has successfully analyzed emergency response data in San Francisco, providing valuable insights into response efficiency, incident patterns, and potential areas for improvement. Medical and Fire Alarms dominate call types, suggesting a focus on preventative measures and public education related to these incidents. Additionally, the seasonal variation in fire emergencies suggests a need for adjusted resource allocation during peak winter months.

The higher frequency of life-threatening incidents in the northeast region necessitates targeted strategies and potentially localized response units. Conversely, identifying safe neighborhoods (such as Benicia, Muir Beach, Alameda, San Bruno, etc.) with low incidents and fast response times can empower residents in choosing their place of residence.

Acknowledgments

We would like to thank Prof. Yang Tang for his guidance and support throughout the Realtime and Big Data Analytics course, which provided invaluable insights and direction for this project. We would also like to acknowledge NYU High-Performance Computing (HPC) for granting access to computational resources such as the Dataproc cluster that facilitated efficient data processing and analysis. Further, we are grateful to Tableau for providing a free trial for their software, enabling visualization and analytics for this project.

References

1. Dudhale, A., Mishra, A. and Pise, P., 2023. Predictive Model for Emergency Medical Services (EMS) using Machine Learning and Data Mining.
2. Acuna, J.A., Zayas-Castro, J.L. and Charkhgard, H., 2020. Ambulance allocation optimization model for the overcrowding problem in US emergency departments: A case study in Florida. Socio-Economic Planning Sciences, 71, p.100747.

3. Buffington, T. and Ezekoye, O.A., 2019. Statistical analysis of fire department response times and effects on fire outcomes in the United States. Fire technology, 55, pp.2369-2393.
4. https://www.jems.com/ems-operations/frequency-and-type-of-national-emergency-medical-system-delays-from-2017-2022/
5. https://data.sfgov.org/Public-Safety/Fire-Department-and-Emergency-Medical-Services-Dis/nuek-vuh3/about_data
6. https://data.sfgov.org/Public-Safety/Fire-Incidents/wr8u-xric/about_data
7. https://data.sfgov.org/dataset/EMSA-Emergency-Medical-Services-Response-Times-Dat/faug-73ss/about_data