**Analyzing San Francisco Fire Department Service Responses**

A Project Report

Presented to

DATA-228 Sec 21

Spring 2023

By

Aishwarya Milind Joshi [016700658]

Dharani Aningi [016671200]

Sai Keerthana Kattige [016676348]

Supriya Kamble [016002584]

Vrushali Menthe [016669770]

07 May 2023

Copyright © 2021

Aishwarya Joshi

Dharani Aningi

Sai Keerthana Kattige

Supriya Kamble

Vrushali Menthe

ALL RIGHTS RESERVED

**ABSTRACT**

**Analyzing San Francisco Fire Department Service Responses**

By Aishwarya Joshi, Dharani Aningi, Sai Keerthana Kattige, Supriya Kamble, Vrushali Menthe

Uncertain events such as fires, natural disasters, accidents, and even medical emergencies are becoming common in our daily lives. To help us in these situations, the first response is often given by the fire department after an emergency call has been placed. A survey conducted by NFPA’s (National Fire Protection Association) in 2017 shows that a US fire department responds every 24 seconds to service calls. Not only wildfires, the fire department helps in situations such as infrastructure fire, investigation of gas leaks, EMS services etc. The data related to all these events is stored in the department database and is already humongous. For our analysis, we have taken into account the data of the San Francisco Fire Department which was founded in 1866 to provide fire prevention and suppression along with emergency services to an estimated 1.5M people of San Francisco. Every year the department responds to 73,000 EMS calls as stated in their reports. We are planning to use this data to understand and analyze trends and patterns between emergency service demand and the department's response times. The dataset used for this project is enriched with valuable information about the department calls and responses with over 6M records (6,199,147) from 2000 till date. Additionally, the data can also be used to pinpoint areas for improvement and assess how well the fire department's current practices are working. A cloud-based system will be utilized to achieve this because it offers appropriate data utilization along with additional advantages like dependability, security, and scalability. The entire data warehousing process will be managed by the AWS Cloud platform, with Tableau serving as the reporting and visualization tool.

**Acknowledgements**

We would like to thank Professor Andrew H. Bond and TA [Revathi Boopathi](https://sjsu.instructure.com/courses/1559246/users/4528496) for their constant guidance and assistance throughout the course work and for the successful development of our project.

**Table of Contents**

**Chapter 1 Introduction**

* 1. Project goals and objectives
  2. Problem and motivation
  3. Project application and impact
  4. Project results and deliverables
  5. Project report structure

**Chapter 2 Project Background and Related Work**

* 1. Background and used technologies

2.2 State-of-the-art technologies

* 1. Literature survey

**Chapter 3 System Requirements and Analysis**

3.1 Domain and business requirements

3.2 Customer-oriented requirements

3.3 System function requirements

3.4 System behavior requirements

3.5 System performance and non-functional requirements

3.6 System context and interface requirements

3.7 Technology and resource requirements

**Chapter 4 System Design**

* 1. System architecture design

4.2 System interface and connectivity design

4.3 System user interface design (for software project only)

4.4 System design problems, solutions, and patterns

**Chapter 5 System Implementation**

5.1 System implementation summary

5.2 System implementation issues and resolutions

5.3 Used technologies and tools.

5.4 Data Visualizations using Tableau

**Chapter 6 System Testing and Experiment**

6.1 Testing and experiment scope

6.2 Testing report

**Chapter 7 Conclusion and Future Work**

7.1 Project summary

7.2 Future work

**References**

**List of Figures**

Figure 1. Fire Department Response Process…………………………………………………..14

Figure 2. System Architecture design…………………………………………………………..24

Figure 3. DataSF website……………………………………………………………………….28

Figure 4. Data Upload to AWS S3……………………………………………………………...28

Figure 5. Workflow execution………………………………………………………………….29

Figure 6. Glue ETL job………………………………………………………………………....30

Figure 7. Glue Crawler………………………………………………………………………….30

Figure 8. Sample query in Athena……………………………………………...……………….31

Figure 9. Connection of AWS Athena to Tableau…………………………..………………….31

Figure 10. Tableau Dashboard……………………………………..…………………………...32

Figure 11. Hosted Website……………………………………..……………………………….33

Figure 12. Hosted Website……………………………………..……………………………….33

Figure 13. Data extract………………………..………………………………………………...34

Figure 14. Number of Responses per call type………………………………………………….36

Figure 15. Satellite View of Incident Areas….…………………………………………………37

Figure 16. Number of Responses over the years…………………………….………………….38

Figure 17. Number of Responses to different Call Type groups.…………….…………………39

Figure 18. Number of Responses given by each Battalion..…………………………………….40

Figure 19. City-wise Call Responses……...…………………………………………………….41

**List of Tables**

Table 1. Project phases, deliverables and dates scheduled to complete the deliverables………11

Table 2. Software and Hardware Requirements………………………………………………..23

Table 3. Used AWS services…………………………………………………………………...36

**Chapter 1 Introduction**

**1.1 Project Goals and Objectives**

Every day we come across many emergency situations such as fires, natural disasters, accidents, medical emergencies etc. In these situations, usually the responses are given by the Fire department of that particular area. These departments need to be prompt with their responses as situations can be very fatal. All this data related to the calls, their dispatched units, call date, response date and time is stored in their database. For this project, we are considering the dataset of the San Francisco Fire Department consisting of the calls and responses records. The project aims to understand the needs of the department and support them in crisis management situations. The goal of the project is to improve response time of the department by analyzing the data and identifying the areas of improvement. As a result, they will be able to make decisions and respond to emergencies quickly thus resulting in saving more lives.

The project’s objective is to use a cloud-based platform to store and analyze the data of the department. The development process will be managed by the AWS cloud platform using services like AWS S3, AWS Athena, AWS Glue Studio etc. which offers various advantages like dependability, security, and scalability. A Tableau dashboard will be created to visualize the data and gain insights from it. Also, this analysis would help them to collaborate with other emergency responders and stakeholders to promote effective and coordinated responses to emergency situations, thus enhancing public safety.

**1.2 Problem and Motivation**

It is getting overwhelming day by day to manage resources efficiently and deliver emergency services on time for the department which is the major problem. They come across many difficult and complex situations which require quick thinking and efficient decision making for allocating proper resources at proper time. Longer response times would bring on ineffective resource allocation which would result in casualties and damage to the property.

The motivation behind this project is to use cloud-based services like AWS Cloud Platform to process and analyze the data to address these challenges ultimately improving the response service. The AWS services would be helpful in collecting, storing, and analyzing this vast amount of data which would then be used for getting insights about various improvement areas for the department. They would also be able to allocate resources according to the need, making decisions based on the data, and optimizing their operations.

**1.3 Project Application and Impact**

This project could have a positive impact from a very scholarly standpoint. Firstly, it gives researchers a chance to utilize the data analytics and cloud-based technologies behind the emergency response services sector, a field that has traditionally and wholly relied on manual procedures and sparse technology use. This project can contribute to the creation of new knowledge and insights about emergency services by making use of contemporary data analysis techniques and cloud computing methodologies. Secondly, this project has a great potential to advance academic knowledge in emergency services. Researchers can then utilize this information to better match and understand the major factors that influence the calls for emergency services as well as the most efficient processes and ways to allocate resources and handle emergencies.

The healthcare sector is one of the industries that would benefit from this project. Extremely large databases are used in hospitals and other medical facilities to store the patient data. Medical professionals can make use of the insights generated by the analysts by analyzing patient data to optimize resource allocation in emergencies by using data analytics and cloud-based technology. For instance, hospitals can use our processes to more effectively understand patient needs and allocate resources accordingly. By doing so, the burden on the healthcare system can be minimized and lives can be saved.

The insurance sector is yet another industry that this technology can help. With the aid of this technology, insurance companies can more precisely assess risk and modify their policies as required. Insurance companies can improve their understanding of the likelihood of future emergencies and modify their policies to offer better coverage by analyzing data on previous incidents. For instance, insurance companies can modify their policies to offer better coverage in a certain specific area if data analysis reveals that a given area is more vulnerable.

In conclusion, the project will significantly impact society in terms of enhancing global emergency services, enhancing safety, and saving lives. The project can benefit local communities, emergency services, and insurance companies by optimizing resource allocation and emergency response processes using data analytics and cloud-based technology. Organizations can improve their emergency response times and resource allocation by utilizing data analytics and cloud-based technology. This can enhance the general safety and wellbeing of people and communities while also helping to save lives and lessen property damage. In summary, by advancing knowledge and enhancing emergency services, the application of the project results can have a significant impact on academia, business, and society.

**1.4 Project Results and Deliverables**

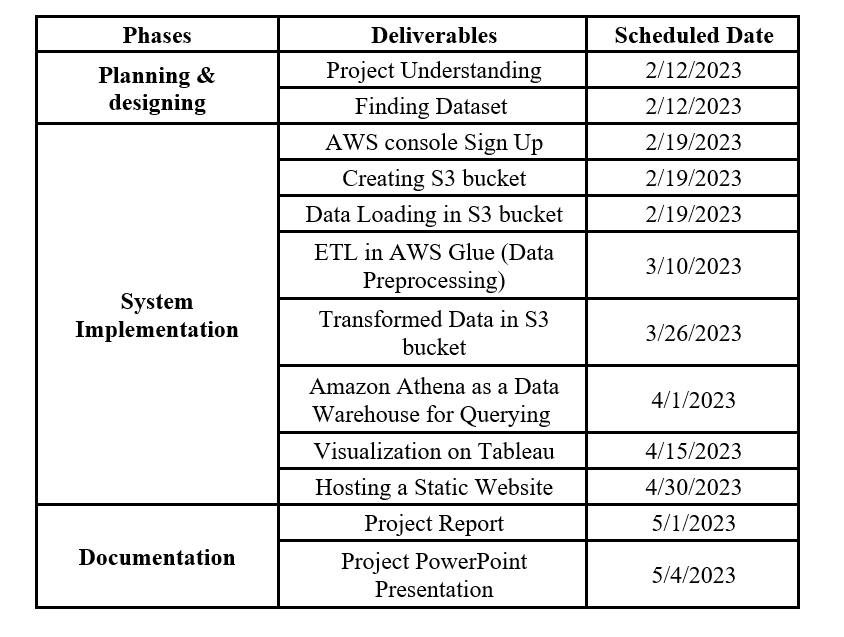
The project aimed to transform and analyze SFFD data by leveraging AWS services such as AWS S3, AWS Glue Studio, and AWS Athena and bring out insights useful to the fire department. AWS S3 was used for storing the data, Glue for ETL processing, and Athena for querying the transformed data. Tableau is used to create a dashboard to present various visualizations and insights such as the number of responses given over the years, different types of calls, and responses per call type or the most responsive battalion of all etc.

In addition, a static website was hosted using Amazon S3 and Route 53. Amazon S3 provides a scalable storage service, while Route 53 is a DNS service which provides reliable and scalable routing of web traffic to resources hosted on AWS. The website hosted the dashboard visualizations, providing a user-friendly interface to view and interact with the data.

Along with the website, a detailed report was created which documented the project methodology, architecture, implementation, and results. The report also includes insights and recommendations based on the analyzed data, which could be used to make informed decisions. For example, the trends identified in call types and response times could be used to optimize resource allocation and improve the emergency response process. Overall, by leveraging AWS services and tools, the team was able to efficiently process and analyze the data and present the insights in a user-friendly manner.

**Table 1**

*Project phases, deliverables and dates scheduled to complete the deliverables.*

**

*Note.*A table showing planned activities for this project.

**1.5 Project Report Structure**

The project report is structured in a way that showcases all the important factors utilized for the development of this project. The first chapter focuses on introducing the general idea of the project. It discusses the project goals and objectives that we have finalized. The problem and motivation behind this project as well as its applications and impacts are discussed in this chapter. The results and deliverables from this project are also provided. The second chapter provides the background about this project. The technologies used and literature survey papers are discussed in this chapter. Chapter 3 showcases system requirements including software and hardware requirements, business, and customer-oriented requirements. Chapter 4 provides information on system design involving system architecture, interface design, system connectivity design etc. while Chapter 5 explains the system implementation along with technologies used and problems encountered. Chapter 6 showcases the test and experiments performed for this project. Chapter 7 provides the conclusion and future scope of this project. Along with all these chapters, related tables and figures are added at respective sections to provide more information.

**Chapter 2 Project Background and Related Work**

**2.1 Background and Used Technologies**

The San Francisco Fire Department (SFFD) is in charge of handling a variety of situations, such as fires, medical crises, and events involving hazardous chemicals. The department covers an area of roughly 47 square miles and provides services to nearly 880,000 persons. Over 1,500 firemen and paramedics work for the department, which has 44 fire stations throughout the city. In addition to fire prevention and suppression, the SFFD provides emergency assistance to an estimated 1.5 million people.

According to San Francisco Fire Department’s data, annually they provide response to more than 73,000 EMS calls which itself is a big achievement. Due to its prominent response to the emergencies the fire department holds a strong reputation for providing high-quality emergency response services. Typically, the fire department claims that they arrive at the scene of an emergency within six minutes of response time after receiving the call of service. After studying the fire department’s working pattern, we can summarize their typical response process in a few steps which include Emergency Response, Incident Assessment, Response Implementation and Response Completion. Figure 1 shows SFFD’s response process in detail. SFFD’s Response Process include following steps:

**Emergency Response.** The Universal Emergency Number is used to contact the San Francisco Fire Department in case of an emergency. Respective units are dispatched to the scene when dispatchers in the communication center ascertain the emergency's type and location.

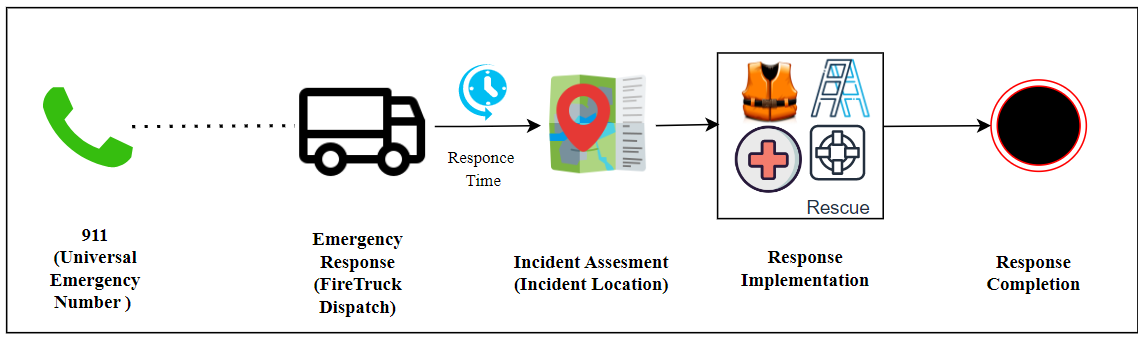
**Incident Assessment**. To decide on the best course of action, the SFFD units evaluate the scene as soon as they arrive. Potential risks are identified, the emergency's scope is assessed, and the resourcesrequired to control the crisis are determined.

**Response Implementation**. The SFFD units carry out the necessary response after finishing the initial assessment. Depending on the emergency, this could entail providing medical care, rescue efforts, or other sorts of aid.

**Response Completion.** After the incident is over, the SFFD units execute any necessary cleanup or other chores and return to their stations. A debriefing may also be held by the incident commander who is in charge of the respective incident, examines the reaction and identifies areas for improvement.

**Figure 1**

*Fire Department Response Process*

*Note*. The above figure shows SFFD’s response process.

The San Francisco’s Fire department collects a significant amount of data on its operations, including response times, incident types, and locations out of which improving response time is the key measure for any fire department. The San Francisco Fire Department (SFFD) uses a number of cutting-edge technologies to handle crises as quickly and effectively as feasible and to fulfill the standards for response times. One illustration of such a technology is the computer-aided dispatch (CAD) system. A piece of software called the CAD system helps emergency dispatchers receive, prioritize, and route emergency calls to the appropriate responders. The technology immediately dispatches the closest available responders to the emergency scene using real-time data to establish their location.

The SFFD also uses GPS tracking technologies. All SFFD vehicles have GPS systems installed, enabling dispatchers to track the whereabouts of every vehicle in real time. Due to their proximity to the incident, this aids dispatchers in making decisions regarding which units to send to an emergency. Moreover, GPS tracking enables more effective emergency vehicle routing to the emergency scenes. All these systems generate humongous data, and this data is used to track performance, identify trends in SFFD’s performance. For getting sufficient insights about the current running system there is a need to store and analyze this data. SFFD is still utilizing the traditional ways to preserve the data which sometimes impacts SFFD’s Response Process and thus, affects the response time.

**2.2 State-of-the-art Technologies**

Moving away from traditional technologies, being aware of the latest trends and making use of the leading data management and analytical tools can aid in improvement of response time. The real-time data produced by SFFD is quite large and would need a lot of storage space. With the use of cloud computing, cloud service providers can set up computing facilities in their data centers for the purpose of processing and storing data through the Internet. Cloud computing is a useful tool for lowering the cost of storing and processing a lot of data for both individuals and organizations. Amazon Web Services (AWS), the largest public cloud provider, provides a variety of services, including data management, analytics, compute, and storage.

The objective of our project is to improve the fire department's response times through the usage of a variety of AWS services. We are utilizing AWS Simple Storage Services (S3), a cloud storage solution, for the data storage because it offers virtually endless storage space, exceptional data durability, 99.99% uptime, and low data access latency. Pay-as-you-go billing, which essentially means paying for the data you store and the time you store it, is another option supported by AWS S3. AWS can be used to introduce the newest developing technology, such as Big Data. Amazon MapReduce is commonly used for big data processing tasks, such as log analysis, data warehousing, and machine learning. It is designed to be highly scalable and can handle petabytes of data. By using these emerging technologies to store, process and analyze data, we can help identify areas of improvement and thereby enable SFFD to improve their processes and the response time even further.

**2.3 Literature Survey**

To uncover trends and patterns to enhance response times and resource allocation, the Springfield Fire Department examined its incident data from 2014 to 2016. The study looked at occurrences including fires, medical emergencies, and responses to dangerous items. To illustrate the distribution of incidents throughout the city and identify hotspots with greater incidence density, the analysis used GIS mapping. By sending more resources to places with higher event rates, this improved resource distribution. The study also identified specific incident types that were more likely to happen in locations, such as medical emergencies in downtown areas and incidents involving hazardous chemicals on highways. The study also looked at reaction times and discovered that while overall response times met national criteria, there were several areas where they might be improved, such as reducing response times for medical crises. A data management system should be put in place to better track incident data and support decision-making, according to the report. In all, the study used data analysis and GIS mapping techniques to uncover patterns and trends in event data, enabling better resource allocation and faster response times. The study's conclusions can be applied to enhance the Springfield Fire Department's efficiency and general level of safety (Springfield Fire Department Analysis Report, 2022)

For the community's safety, it is crucial to ascertain the connection between the fire department's reaction time and the fire's outcome. The study displays a statistical analysis of the fire departments' response times for US residents between 2002 and 2017. According to NFIRS data, the log normal distribution is a good approximation for the empirical distribution of reaction time. Additionally, the article examines the impact of longer reaction times on various metrics, including reported fire spread categories, estimated monetary property, and reported flame damage. An analysis was carried out to check the impact of NFPA 1710, The proportion of fires that caused "extreme" damage to at least one storey and the proportion of damaged property value were shown to be inversely related to department response times. To determine how many burned regions there were as a result of fire, a technique known as power law distribution was developed. A Monte-Carlo methodology was developed to verify whether the data is consistent with the claim. The various approaches for assessing the consistency and quality of data from NFIRS reports are described at the end. One of the study's main findings is a measurement of how response time affects several markers of fire loss. It was discovered that all average fire metrics had increased between 3 and 13 minutes of response time, increasing the likelihood of highly catastrophic fire mishaps at longer response times (Buffington et al.,2019).

The research paper proposes a real-time predictive model for fire department response times in the city of San Francisco, California. By offering precise and timely estimates of response times, the authors hope to increase the effectiveness and efficiency of emergency services. These predictions can aid fire departments in more efficient resource allocation and quicker emergency response. The San Francisco Fire Department provided statistics on fire incidences and response times to the author during the years of 2003 and 2013. The data includes specifics regarding each incident's location, type, time of day, and reaction time. The author used this information to create a prediction model that, using the incident's location and other pertinent information, could forecast the response time for a specific incident. The predictive model was based on machine learning algorithms, specifically random forests, and gradient boosting machines. The technologies which aided in model development are Amazon Web Services, Apache Spark, and python. The model was trained on historical data using machine learning algorithms, and the model's precision was assessed by comparing predictions to actual response times for a portion of the data. The scientists discovered that the predictive model could forecast response times with accuracy for a variety of incident kinds and locations. Additionally, they discovered that the model was capable of making predictions in real-time, indicating that it may be utilized to give emergency responders accurate and timely information. To make the model more scalable and efficient, the authors implemented a distributed computing framework using Apache Spark. This allowed them to process the large amounts of data quickly and efficiently, and to update the model in real-time as new data became available (Lian et al.,2019).

**Chapter 3 System Requirements and Analysis**

**3.1 Domain and Business Requirements**

The project's primary objective is to provide insights into data collected from diverse sources, allowing fire department officials to make well-informed choices and take required steps. The Fire Department Calls for Service system will collect, and store data related to fire incidents and other emergency calls for service. The data will be evaluated to discover trends and patterns, and reports and visualizations will be created to assist fire department officials in making educated decisions. The system will also offer fire department officials and the public with real-time updates on the status of events. The system will allow fire department officials to respond to situations more effectively by giving timely and accurate information.

**3.2 Customer-oriented Requirements**

The system must have a user-friendly interface that allows fire department personnel to quickly and effortlessly access the information they require. This necessitates a straightforward and simple interface that presents only the information users need. Officials should be able to easily find the data they need by searching, filtering, and sorting it in the system. The system should also allow for specialized inquiries and the export of data in several formats for additional research.

The technology should also allow the general public to report emergencies and receive information about service requests. The system must allow the general public to readily report emergencies and provide crucial details like their location, the nature of the problem, and contact information. The system should also provide real-time updates on the status of any response teams that have been dispatched or are on way. Additionally, the system must respect all applicable data protection rules and regulations while handling the personal information of persons taking part in emergency calls.

**3.3 System Function Requirements**

The Fire Department Calls for Service project’s system function requirements should ensure that the public and fire department authorities may easily examine information about emergency calls and fire events.

**Data Collection and Storage***.* The system should be able to collect and store data from a variety of sources, such as emergency calls, dispatch records, and fire incident reports, among others. In addition, the system must ensure the safety and confidentiality of the data it obtains by keeping it in a secure manner that protects its privacy and security.

**Real-time updates.** The system should provide instant access to up-to-date information about ongoing events for both the public and the fire service. Additionally, it should have the ability to alert the relevant authorities and guide the appropriate response teams to the incident's location, depending on its severity.

**Data Analysis.** To identify patterns and trends in the data, the system must be able to analyze it. The system should provide analysis of the data, which should include the quantity of service requests, the fire department's response time, the location and severity of incidents, and other relevant information.

**Reporting and Visualization.** The system should be able to provide reports and visualizations to help the fire department staff make educated decisions. The reports should contain a summary of the incidents, the reaction time, and other relevant data. The system should also be able to create customized reports and visualizations based on the requirements of the user.

**User management and security.** For authorized users to use the system, secure login credentials must be enabled. The system should limit access to the data to those who have been given permission to see and update it. The system should also be able to define and manage user roles and permissions so that users can only access the data that is relevant to their job function.

**3.4 System Performance and Non-functional Requirements**

The Fire Department Calls for Service system's performance and non-functional requirements are essential to ensure that it is effective, dependable, and secure. The following conditions must be satisfied -

**Scalability.** The system needs to be able to manage a lot of data because the number of reported occurrences may rise rapidly. The system must be built to analyze data fast and handle spikes in data traffic without sacrificing performance.

**Reliability and availability.** To ensure that emergency calls are never dismissed, the system must be reliable and available around-the-clock. Because any downtime or system failure can cause significant harm to the public safety system, the system should be designed to minimize downtime or interruptions.

**3.5 System Behavior Requirements**

The Fire Department Calls for Service project's system behavior requirements outline the expected actions of the system in different scenarios. The primary requirement is for the system to reliably and safely collect and store data. This necessitates stringent methods for data gathering and storage to safeguard against data loss and safeguards against unwanted data access.

Additionally, it's critical that the system update fire department personnel and the general public in real time on the development of situations. As a result, the system needs a way to update the incident status in real time and share that data with the appropriate people.

The third requirement is precise data analysis and trend/pattern reporting. This necessitates sophisticated analytic capabilities on the side of the system in order to identify trends and patterns in the data and make conclusions that can guide the decisions of fire department officials.

**3.6 System Context and Interface Requirements**

The main requirements for the system to access the Fire Department Calls for Service dataset are the System Context Requirements. These specifications outline the functionality that the system must offer in order to meet user needs. Users should be able to access and search the dataset using a number of different criteria, including call number, incident number, call type, call date, and other pertinent data. This means that the system needs to have a simple search interface that enables users to find and get pertinent data fast. Users should be able to filter data using the system's date range, location, and other options. This means that the system should enable users to apply filters to the data to narrow their search results and obtain the information they require.

The system must first offer a simple user interface that enables users to navigate and interact with the data rapidly. This means that the system must offer an easy-to-use and understandable user interface, especially for people who are new with the dataset. Users should also be able to sort data in the system using other criteria, such as call type, call date, and response time. Additionally, the system must be adaptable to a range of devices, including PCs, laptops, and mobile ones. Any device should be able to access it, and it should provide a consistent user experience.

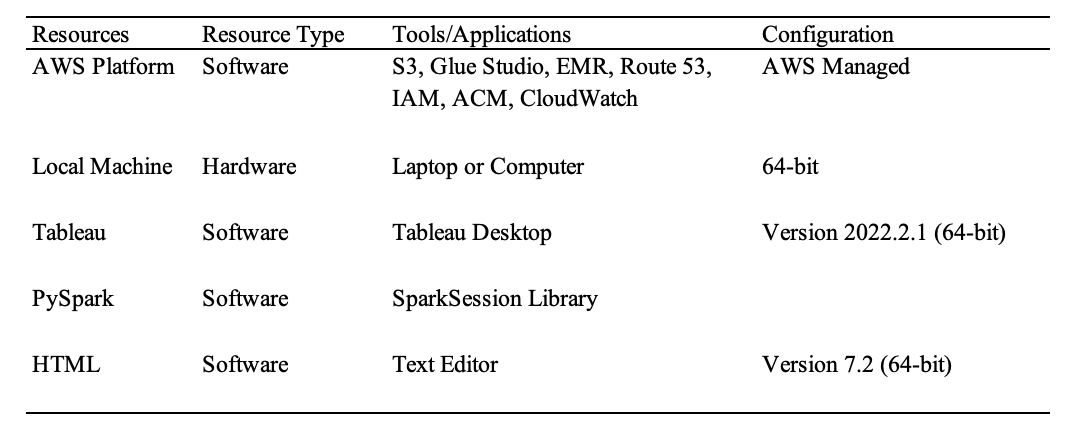
**3.7 Technology and Resource Requirements**

The hardware and software configurations for this project are listed in the table below. The Hardware configuration includes a 1.2 GHz quad-core Intel Core i7 processor, which offers ample processing capacity for a variety of applications. An Intel Iris Plus Graphics chipset with 1536 MB of dedicated RAM provides graphics capabilities, which should be suitable for most graphics-related applications. Furthermore, the system includes 16GB of high-speed RAM which should be enough to handle a variety of applications and operations.

The project's software configuration includes MacOS Ventura 13.3.1 as the operating system, and an Amazon Web Services (AWS) platform comprising S3 for cloud-based storage, Glue Studio for data processing and transformation, Athena for data analysis, CloudWatch for log analysis, and Route 53 for domain name services. Additionally, the Tableau is leveraged for visualization and business intelligence purposes.

**Table 2**

*Software and Hardware Requirements*

**

*Note.* Software and hardware requirements

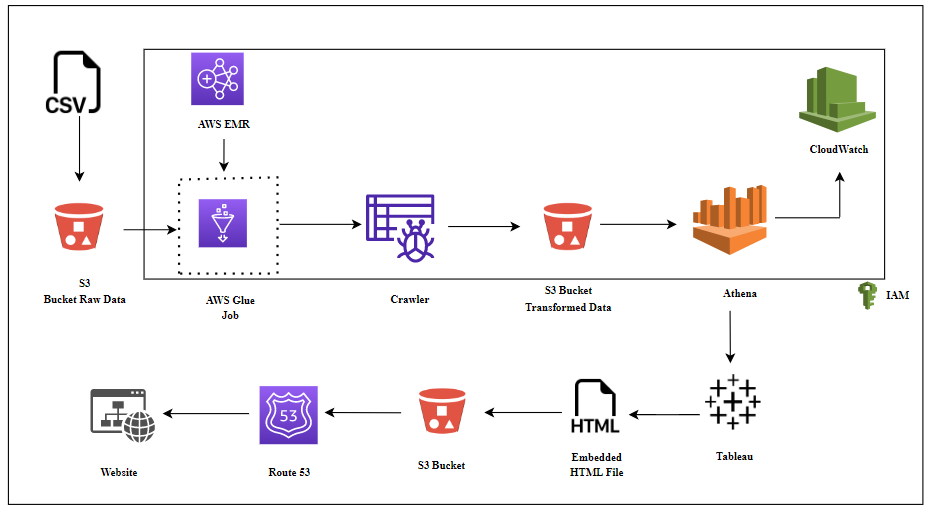
**Chapter 4 System Design**

**4.1 System Architecture Design**

The system architecture process begins with the collection of raw data from the San Francisco government website. This data is then stored in an Amazon S3 bucket, which serves as a central repository for data storage and management. A virtual environment is set up using Amazon Elastic MapReduce (EMR), to perform data transformation tasks. The transformed data is processed using an AWS Glue Job and crawled using a crawler. The transformed data is then stored in the S3 bucket's output folder in Parquet format.

**Figure 2**

*System Architecture design*

****

*Note.* The project follows above given system architecture.

Once the data is reflected in the data catalog after execution of crawler, it can be queried using Amazon Athena, an interactive query service that allows users to analyze data in Amazon S3 using SQL. All log files generated during the process are stored in CloudWatch, an AWS monitoring service that provides operational insights and analysis for applications running on the AWS infrastructure. The transformed data is connected to Tableau for visualization purposes. After visualizing the data in Tableau Public, the embedded code is used in an HTML script to be uploaded into S3 bucket. Finally, the S3 bucket is linked to Route53, a highly scalable and reliable DNS service provided by AWS, to host a static website for data analysis and visualization.

**4.2 System interface and connectivity design**

Tableau includes a connection that allows users to connect to Amazon Athena and access data in Amazon S3. This integration is achievable via Athena's ODBC driver, which allows the connector to connect to and obtain data from S3. With the connection established, users may use Athena data to build and personalize visualizations and dashboards in Tableau. The connection facilitates the integration of Athena with Tableau, allowing users to take advantage of the full capabilities of both platforms while also providing seamless data analysis and visualization. Also, the embedded code generated from Tableau Public is integrated in an HTML script that is hosted by using AWS Route 53 and AWS S3 services, thus benefiting from Route 53's sophisticated traffic routing capabilities.

**4.3 System user interface design**

In order to better understand the San Francisco Fire Department responses, we created a website that incorporates Tableau visualizations created using AWS Athena. By offering in-depth research and insights into the department's data, the website attempts to improve comprehension and knowledge of its activities. Users can browse a range of interactive Tableau visualizations on the website that provide important data on fire incidents, emergency calls, response times, and more. These visualizations help a wide range of consumers access complex data that is simple to grasp. Users can learn important information about response patterns and incident trends by combining various visualizations and data. The Fire Department Service Calls and Response dataset covers data ranging from the year 2000 till date. This dataset provides useful details on the different call types, priority levels, geographic locations, response times, and resource needs, allowing for more informed decision-making and enhanced emergency response. In order to manage and analyze data, AWS Athena is integrated with the Tableau Desktop. Large datasets kept in the cloud are efficiently queried using Athena, giving instantaneous insights. The website gives dynamic, current visuals that reflect the most recent data by integrating Tableau to Athena.

In conclusion, the website provides a thorough platform for learning and understanding the data from the San Francisco Fire Department. It promotes better decision-making, transparency, and public participation via interactive visualizations, a central dashboard, and cloud-based analytics.

**4.4 System design problems, solutions, and patterns**

The problems faced when creating a system or architecture, such as scalability, performance, and cost-effectiveness, are referred to as system design problems. Finding the finest data warehousing solution for storing and managing enormous amounts of data is one of these issues. Popular data warehouse solution AWS Redshift provides quick and scalable performance for analytical workloads. Redshift, however, can be expensive and challenging to set up and operate. To overcome this issue, we chose Amazon Athena as our data warehousing option. Athena is a serverless, interactive query service that allows users to run normal SQL queries to evaluate data in Amazon S3. It is affordable, requires little maintenance of the infrastructure, and is easy to set up and operate. Athena is well suited for handling massive datasets since it automatically scales and has good query performance. AWS Athena offers a complete end-to-end solution for data processing, storage, and analysis when combined with other AWS services like AWS Glue and Amazon Quick Sight. Additionally, a variety of data formats are supported by Athena, such as CSV, JSON, and Parquet, enabling users to query data saved in different formats without the need for data transformation.

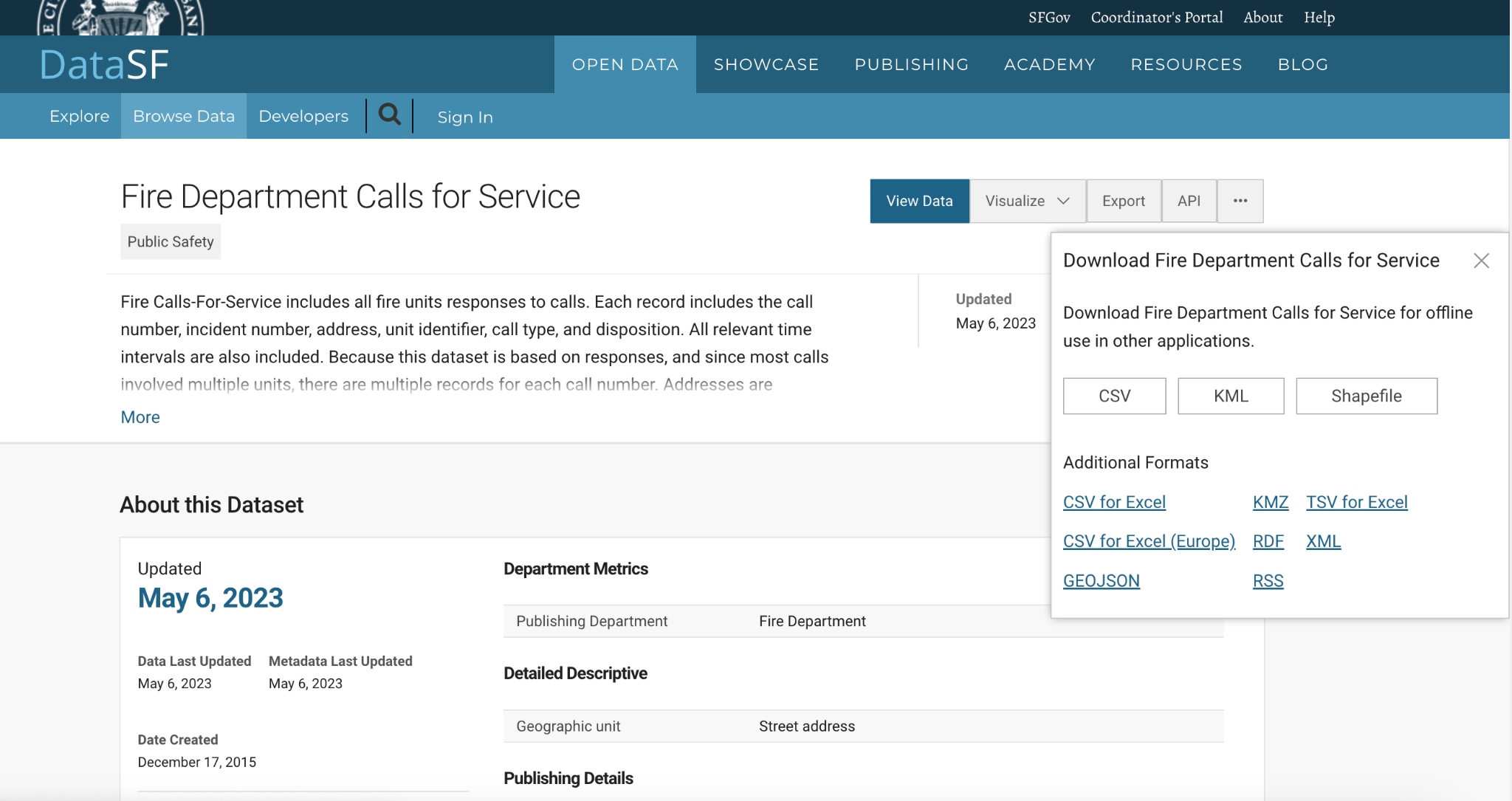
**Chapter 5 System Implementation**

**5.1 System Implementation Summary**

Starting with data collection, we have downloaded 2.2 GB of data in CSV format from the given website of the SF Fire Department in a local machine. This data dump is then uploaded into the AWS S3 bucket under a folder named input.

**Figure 3**

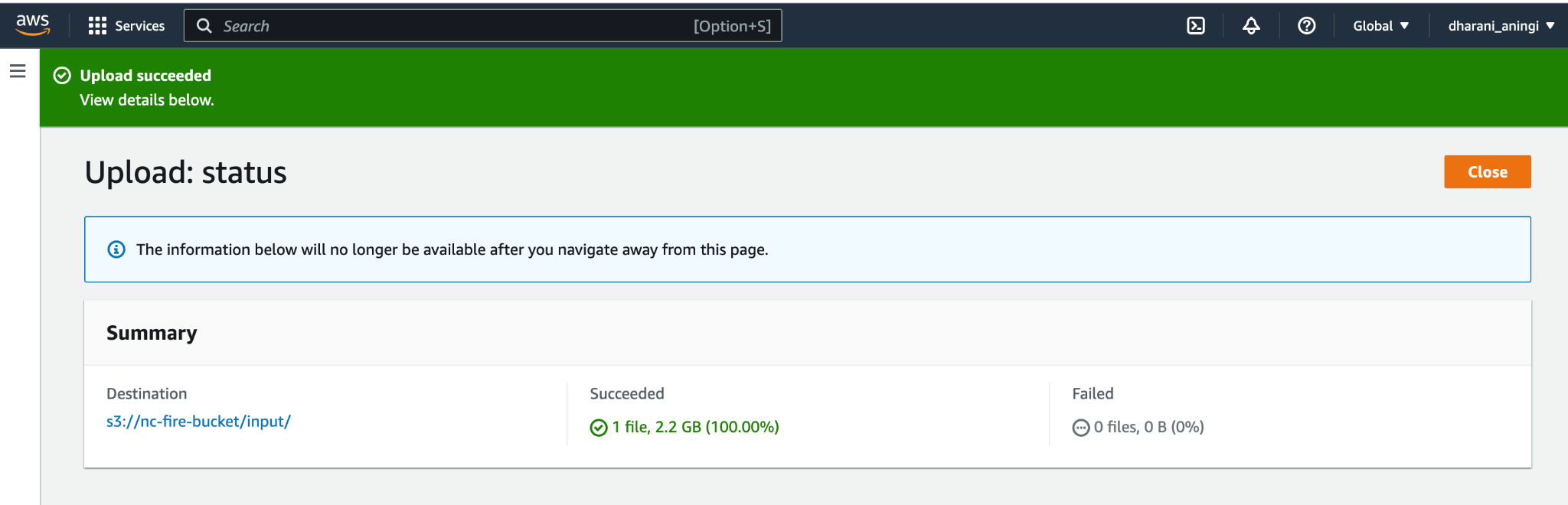
*DataSF website*



*Note.* The data is downloaded from the above website.

**Figure 4**

*Data Upload to AWS S3*

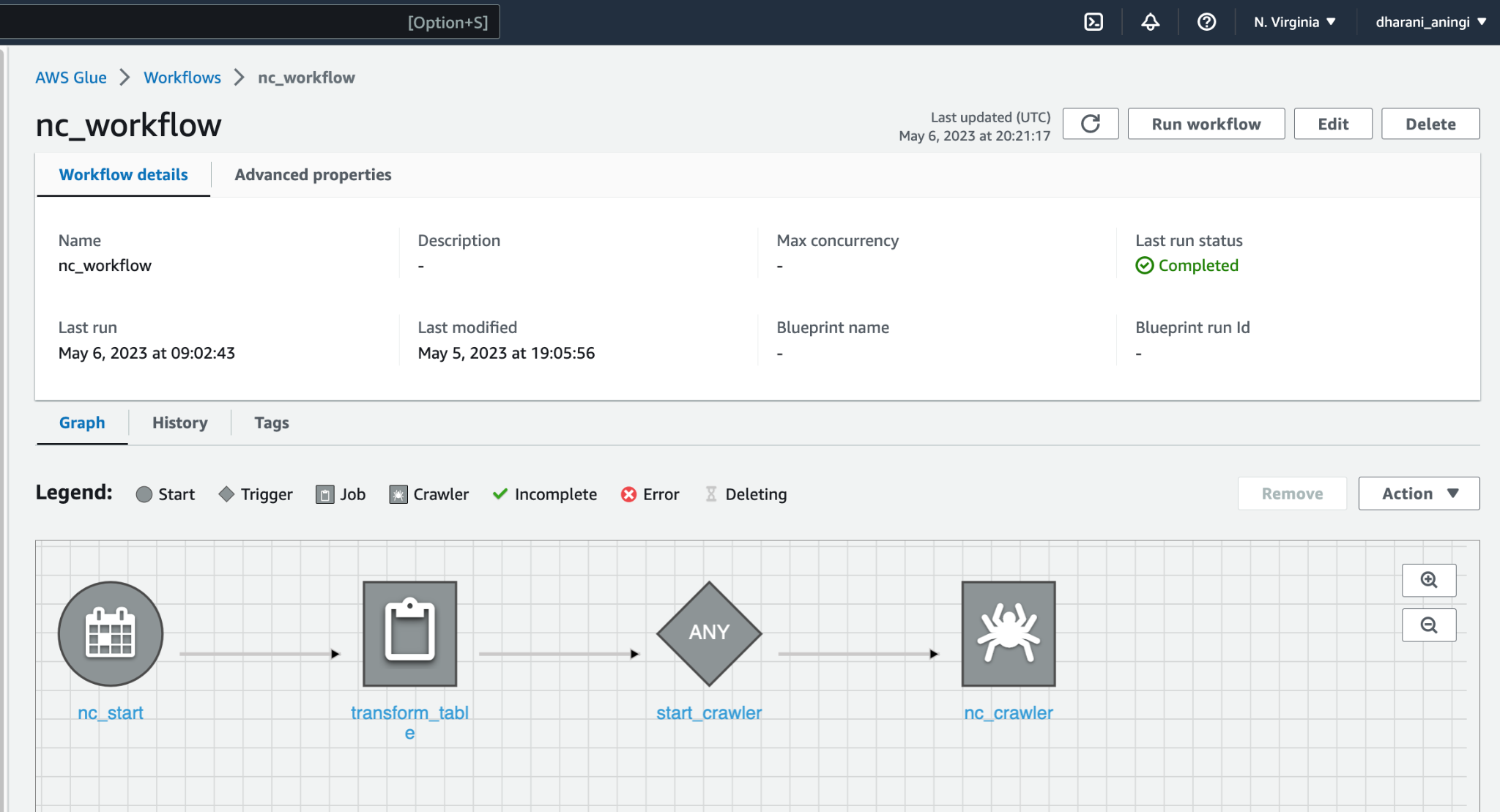


*Note.* Data upload done in AWS S3 bucket.

A workflow named nc\_workflow is executed to run the Glue job that performs ETL operations on the raw data in S3 bucket. The ETL activities performed are removing null values, removing spaces between column names. Some columns such as City, Original Priority, Fire Prevention District etc. needed transformation as the values were not consistent. After all these operations, the transformed data is saved in parquet format in an output folder in S3 bucket.

**Figure 5**

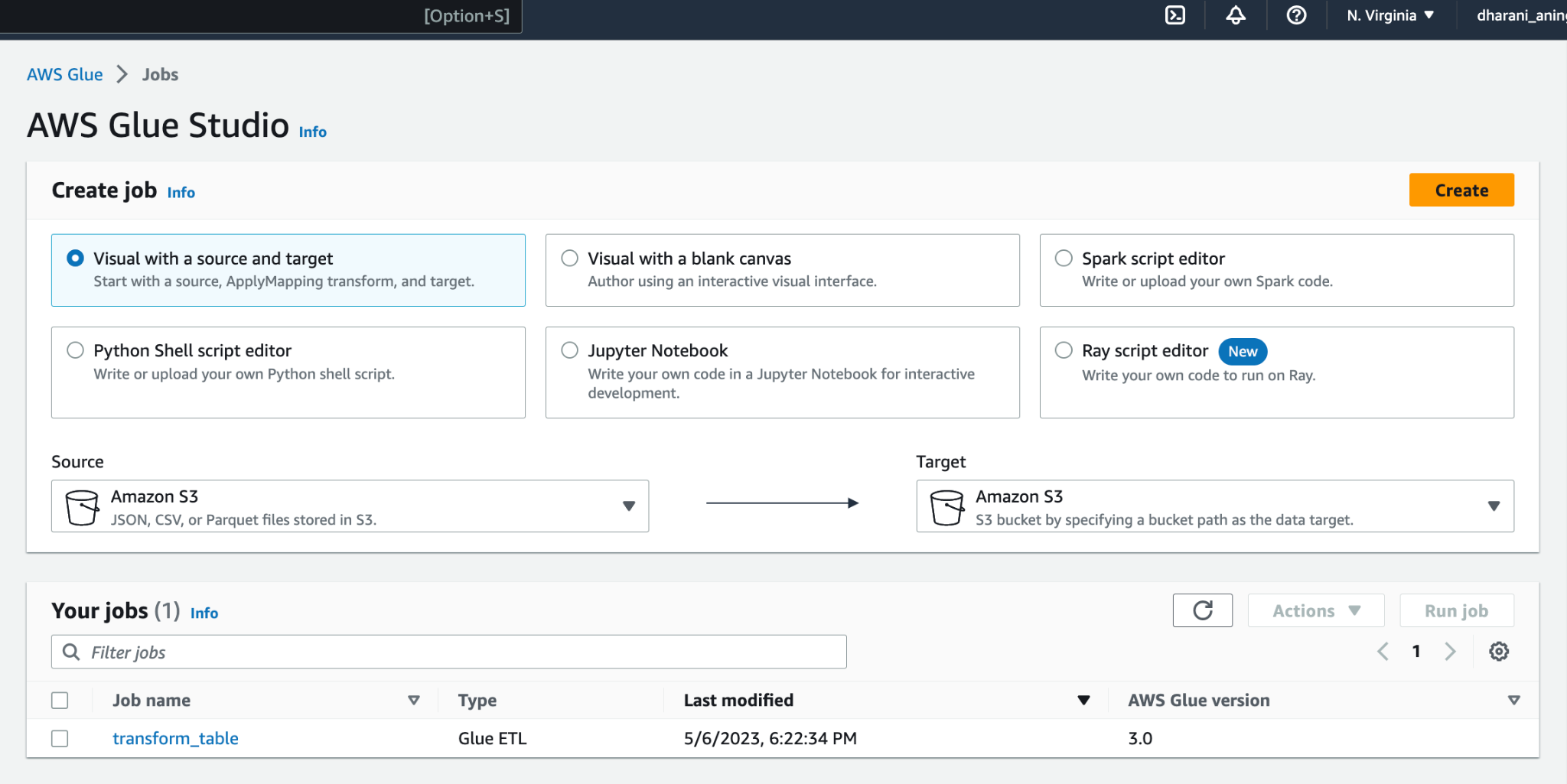
*Workflow execution*



*Note.* The workflow shows ETL job and crawler automation.

**Figure 6**

*Glue ETL job*

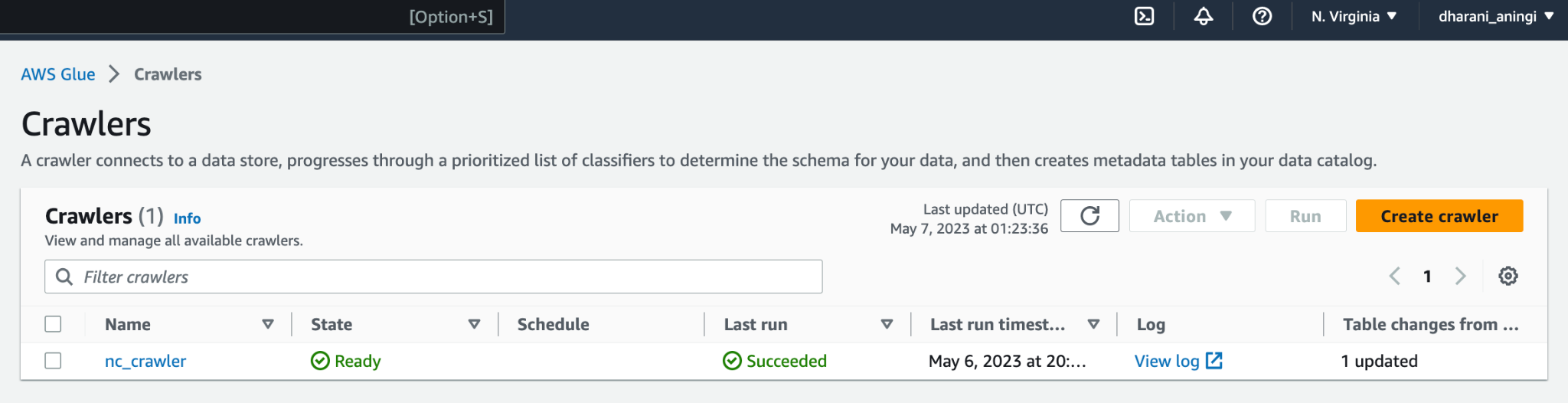


*Note.* A glue ETL job created for data processing and transformation.

This workflow also includes a crawler to crawl the data from the output folder and create a table in AWS Data Catalog under a nc\_db database. The workflow is scheduled to run every day at 9.00 AM. After running the workflow, data gets loaded into a table named output which is then queried using AWS Athena.

**Figure 7**

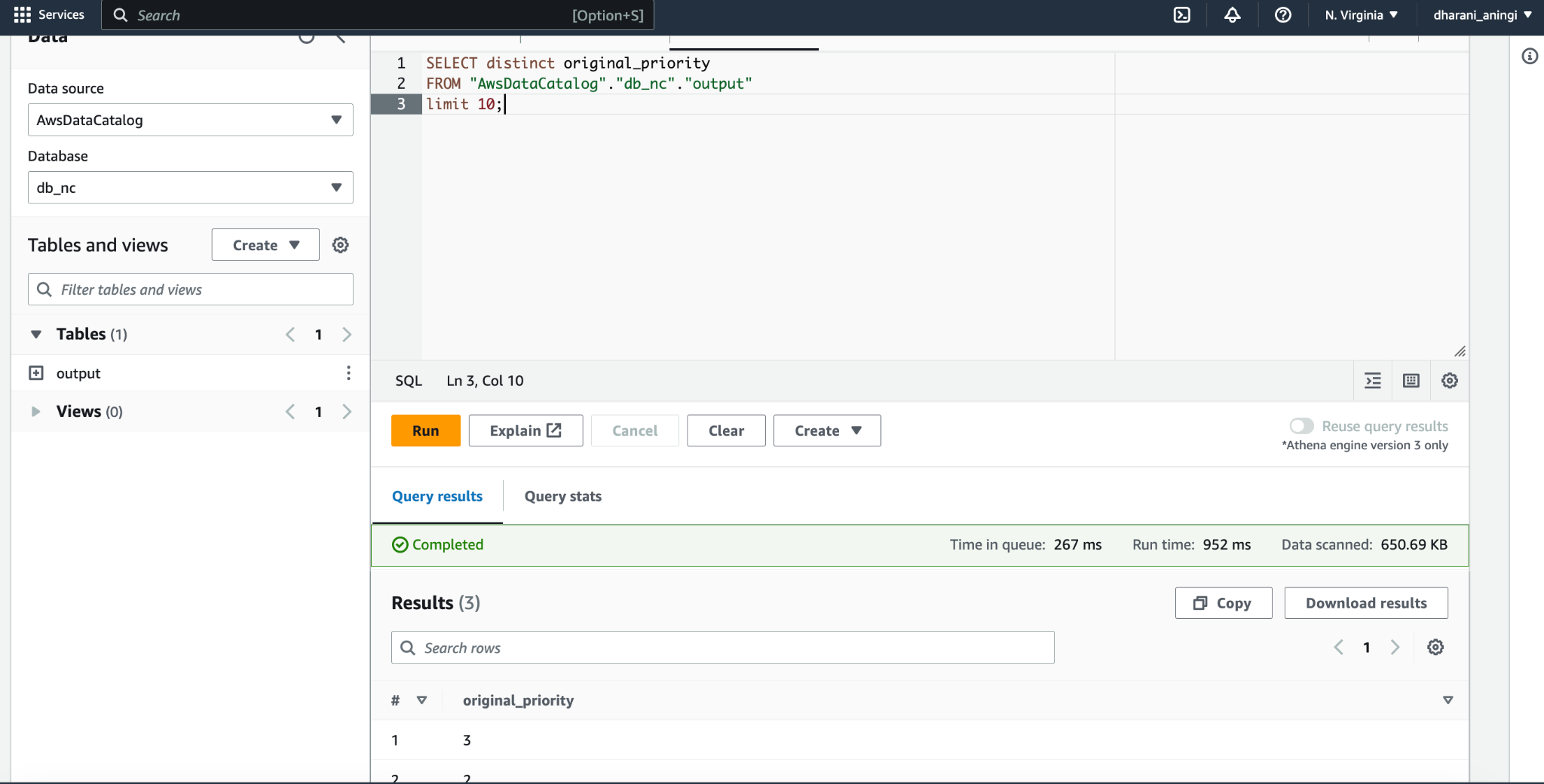
*Glue Crawler*



*Note.* A glue crawler for creating table schema.

**Figure 8**

*Sample query in Athena*

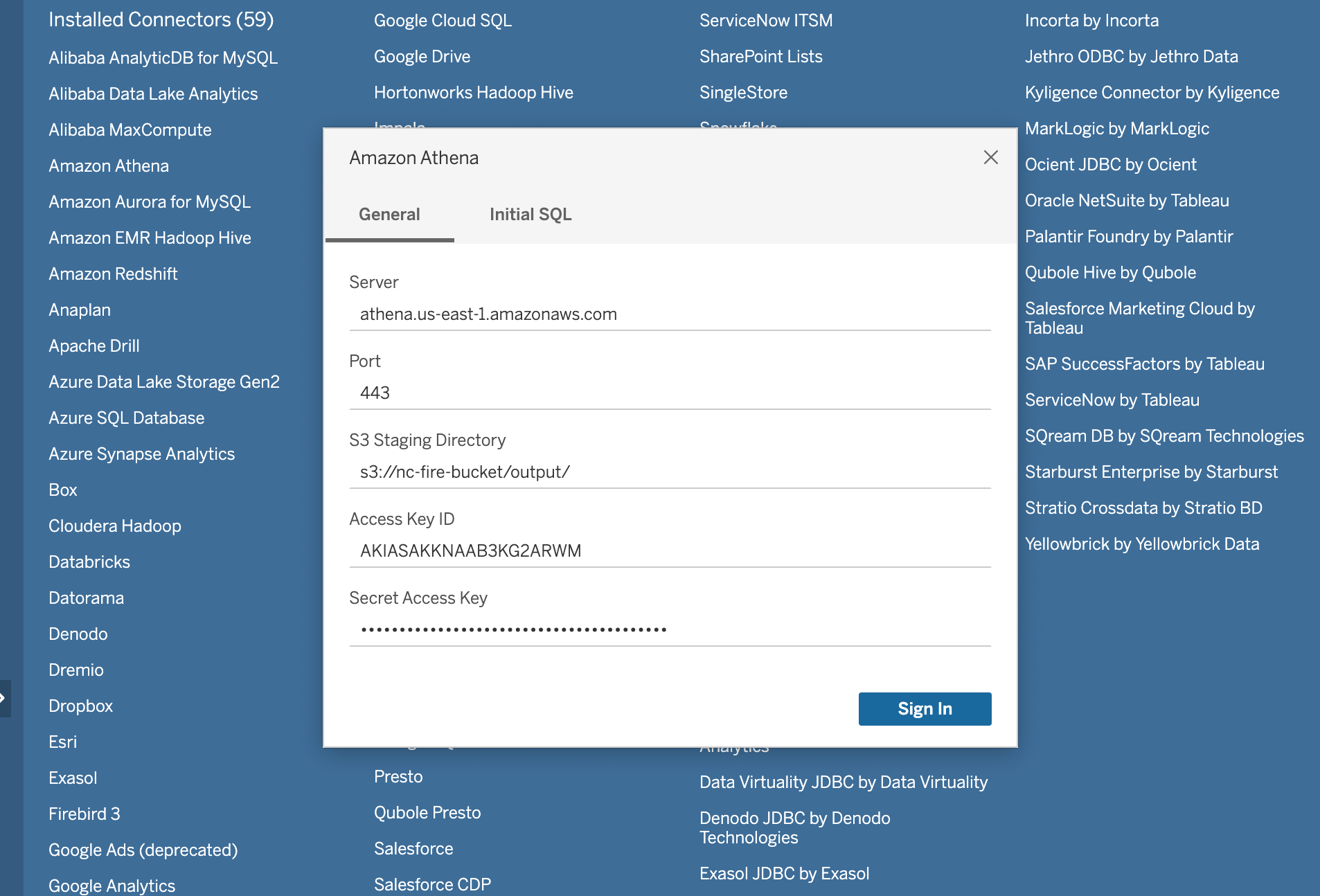


*Note.* A sample query executed on transformed data.

We have made use of Tableau for performing data analytics. A connection is made in Tableau for AWS Athena by providing the Server name, Port and the S3 staging directory. Access Key ID and Secret Access Key are obtained from an IAM user.

**Figure 9**

*Connection of AWS Athena to Tableau*

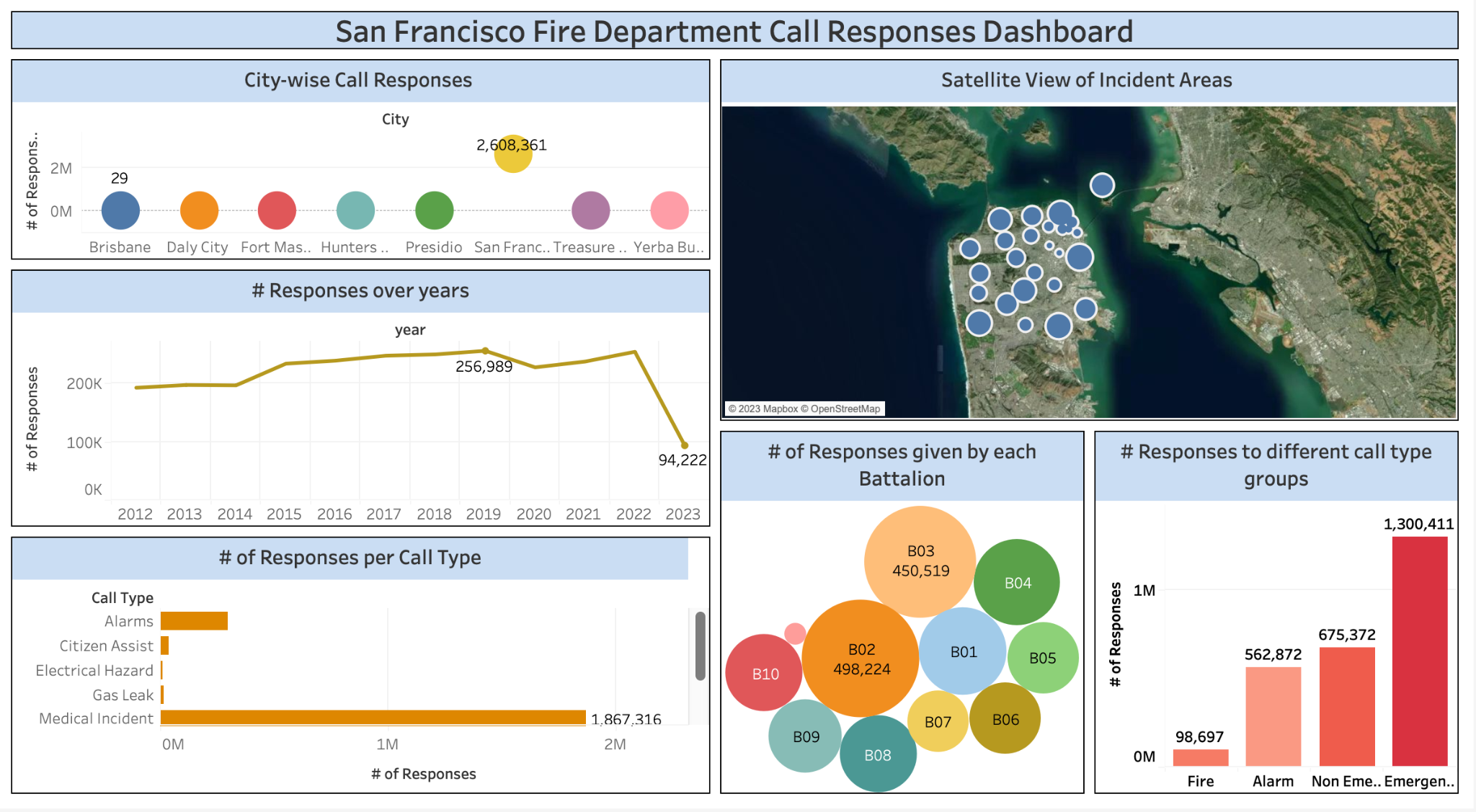


*Note.* Connection parameters between AWS Athena and Tableau.

A dashboard is created for all the visualizations in Tableau and published on Tableau Public by using the in-built server connection from Tableau. After the dashboard was published, the embed code of the dashboard is used in a HTML script. This script is uploaded into a S3 bucket and hosted using a simple routing policy in a DNS hosted zone offered by AWS Route 53 service.

**Figure 10**

*Tableau Dashboard*



*Note.* A dashboard is created in Tableau using the data from AWS Athena.

**Figure 11**

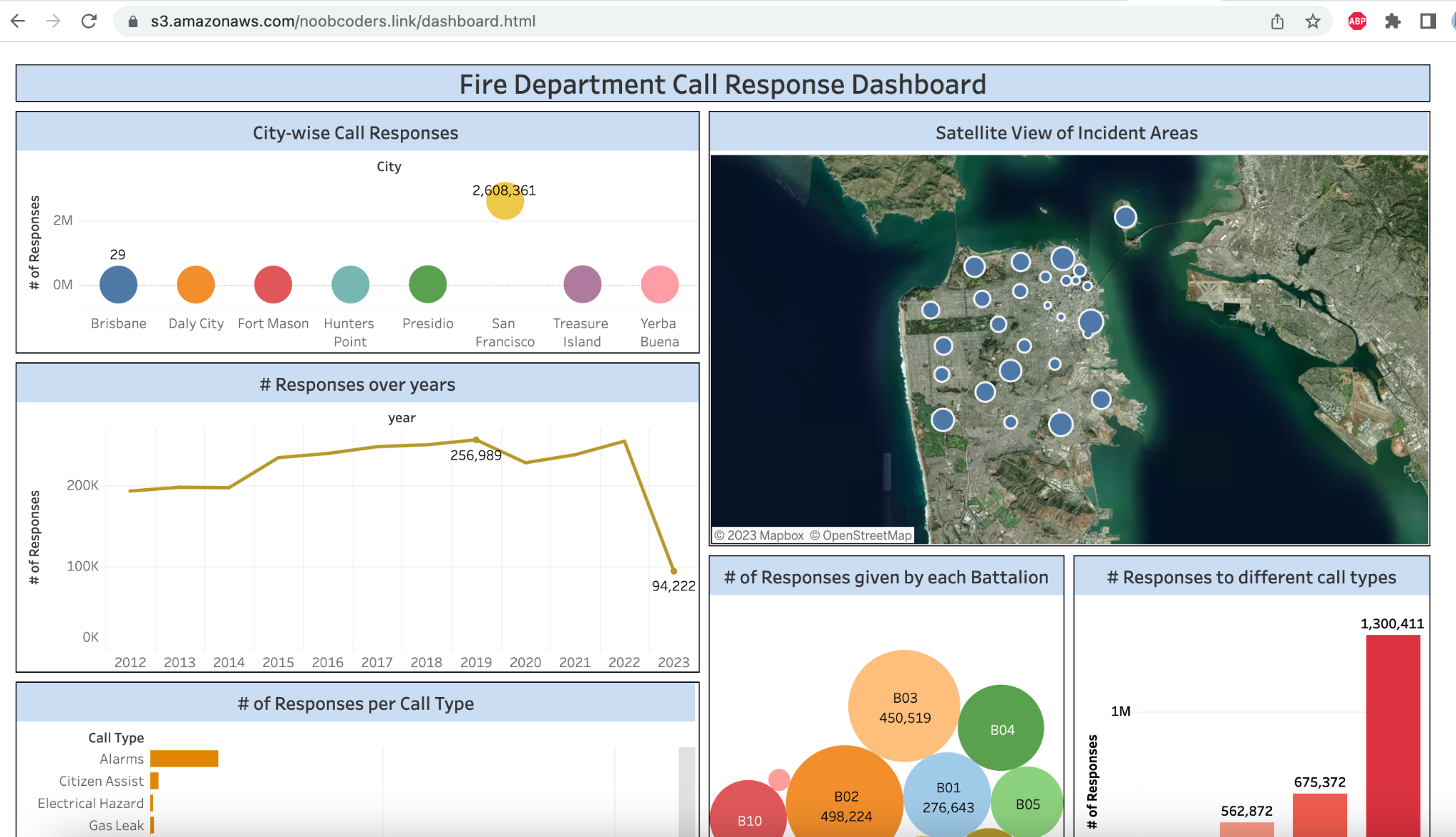
*Hosted Website*

****

*Note.* A website showing Tableau Visualizations.

**Figure 12**

*Hosted Website*

**

*Note.* A website showing Tableau dashboard.

**5.2 System Implementation Issues and Resolutions**

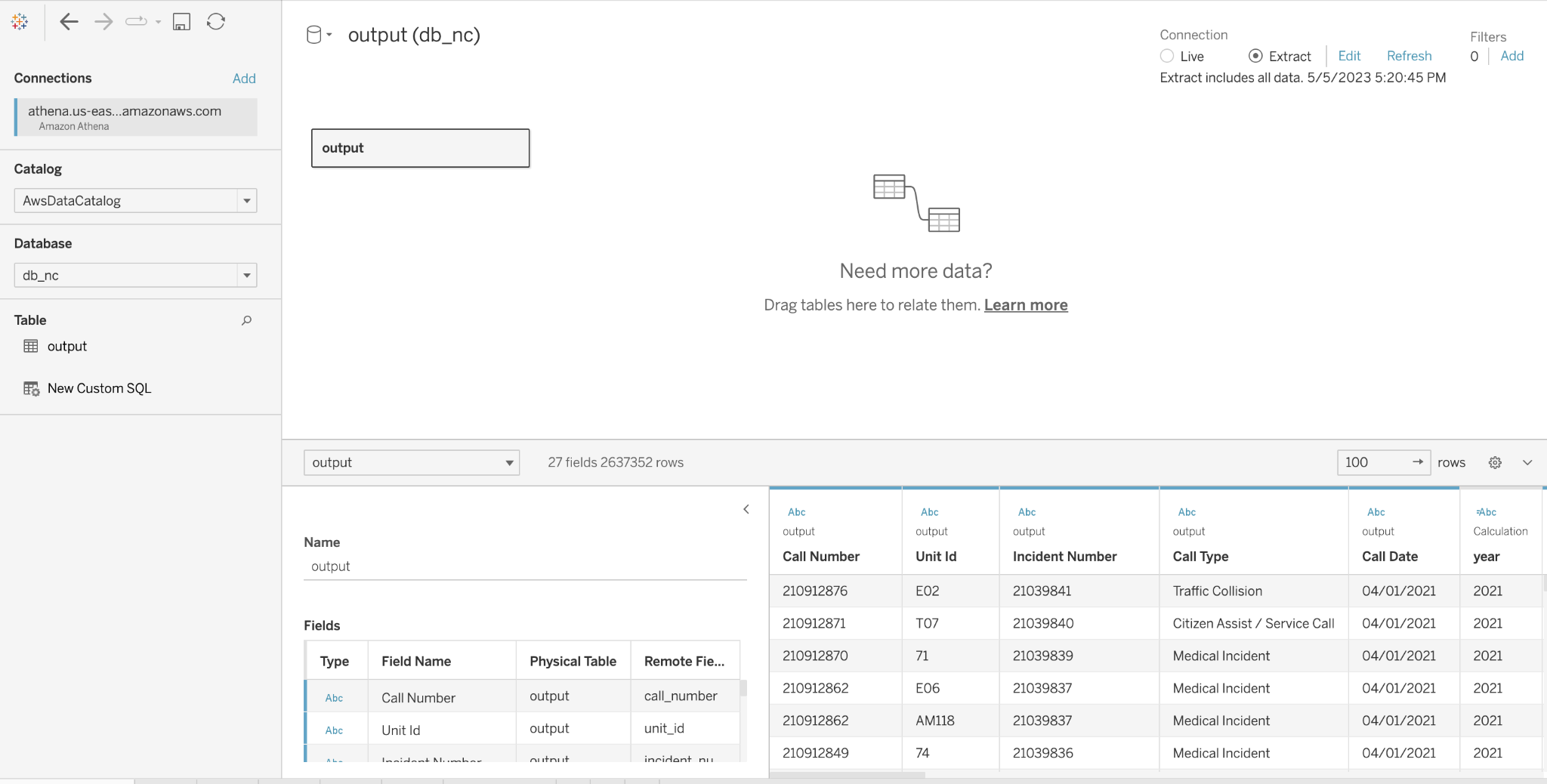
We faced many obstacles that needed attention during the implementation of this project. One of the issues was the execution of the ETL job which was failing because it was not able to access the S3 bucket to read the input file. After inspecting, we got to know that we need to update permissions for the bucket by unclicking the Block public access setting. This way the job was able to read the input file from the bucket and perform ETL operations.

The other issue was the connection of Athena to Tableau. Athena was not able to connect to Tableau because of a missing JDBC driver. After placing the driver in the Tableau/drivers folder, Athena got successfully connected to Tableau. It also gave Invalid Username/Password errors, however those got resolved after providing new Access Key ID and Secret Access Key.

One more issue we encountered was when publishing the dashboard to Tableau Public. The workbook was not able to upload/save on the site because of the live connection to Athena server as it is an external database connection. So, we had to create an extract of the data from Athena which was stored locally and then use it to publish the dashboard to Tableau Public.

**Figure 13**

*Data extract*

**

*Note.* Extract connection used to take an extract of the data from AWS Athena.

**5.3 Used Technologies and Tools**

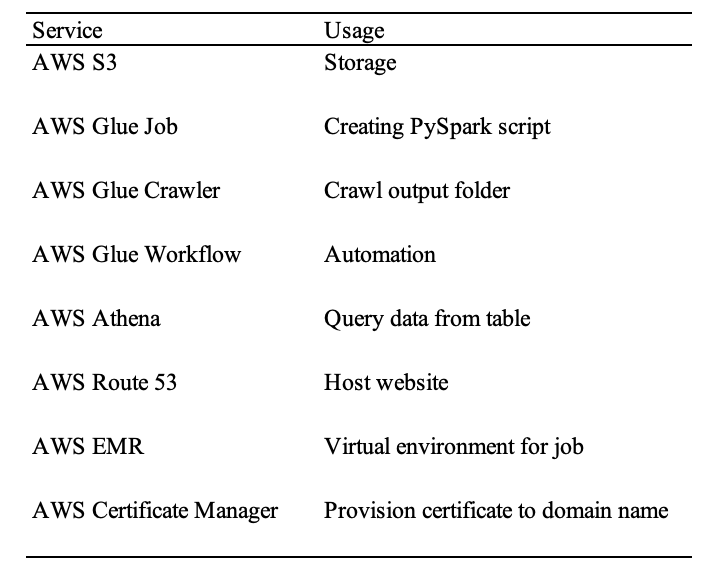
The "Analyzing Fire Department Service Calls and Responses" project combines a variety of Amazon Web Services (AWS) technologies and tools to collect, process and analyze data on fire department service responses. The services like AWS S3, AWS Glue, AWS EMR, AWS Athena, AWS Glue Crawler, Route 53, AWS Certificate Manager etc. are utilized in implementing this project. Also, Tableau is used for visualization purposes and creating a website using the embed code received after publishing the dashboard to Tableau Public.

The AWS S3 is a cloud-based storage service used to store our data in the cloud. In this project, the initial load of 2.2 GB of data is stored in S3 bucket. After the execution of the ETL job, the processed and transformed data is also stored in the S3 output folder. The static website hosting is also carried out with the help of AWS S3 bucket. AWS Glue Studio is used to create a Glue ETL job for performing ETL operations on the raw data. A Glue Crawler is also created to crawl the output folder and create a schema in AWS Data Catalog. AWS Athena is used to query the table and view the data.

AWS EMR is used to provide a virtual execution environment to the Glue job by providing distributed computing power for processing this large amount of data. Route 53 is used to provide a unique domain name to host the website while AWS Certificate Manager is used to provision and manage SSL/TLS certificates for our domain names. We have made use of Tableau to create interesting visualizations from the data and understand the relations between them. Tableau Public is also used to publish the dashboard online and generate an embed code for the dashboard.

**Table 3**

*Used AWS services*

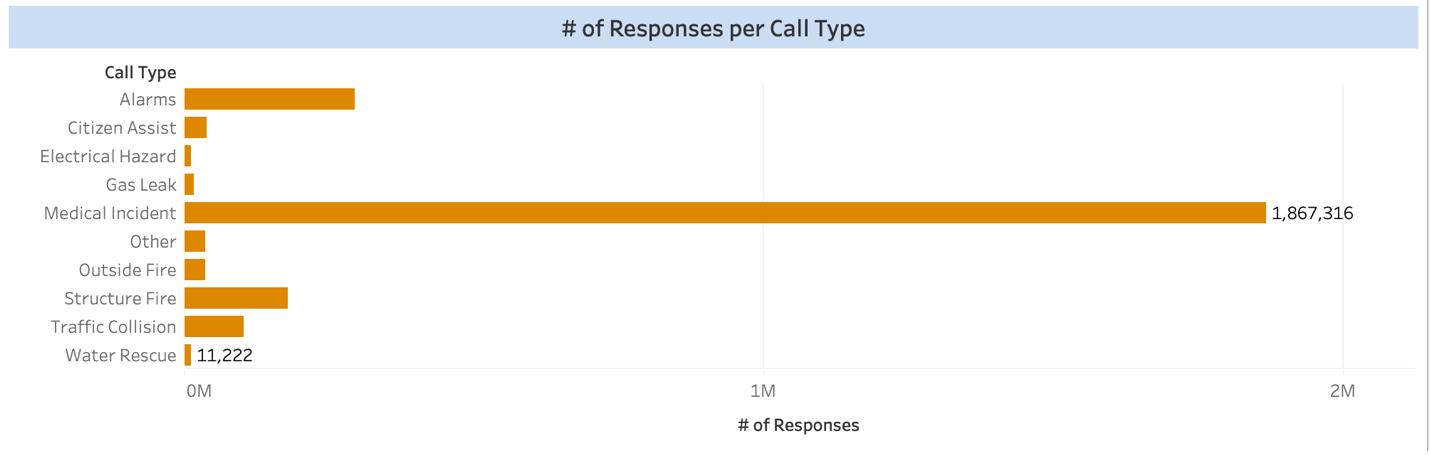
**

*Note*. Above table shows the used AWS services for implementing this project.

**5.4 Data Visualization Using Tableau**

**Figure 14**

*Number of Responses per Call Type*

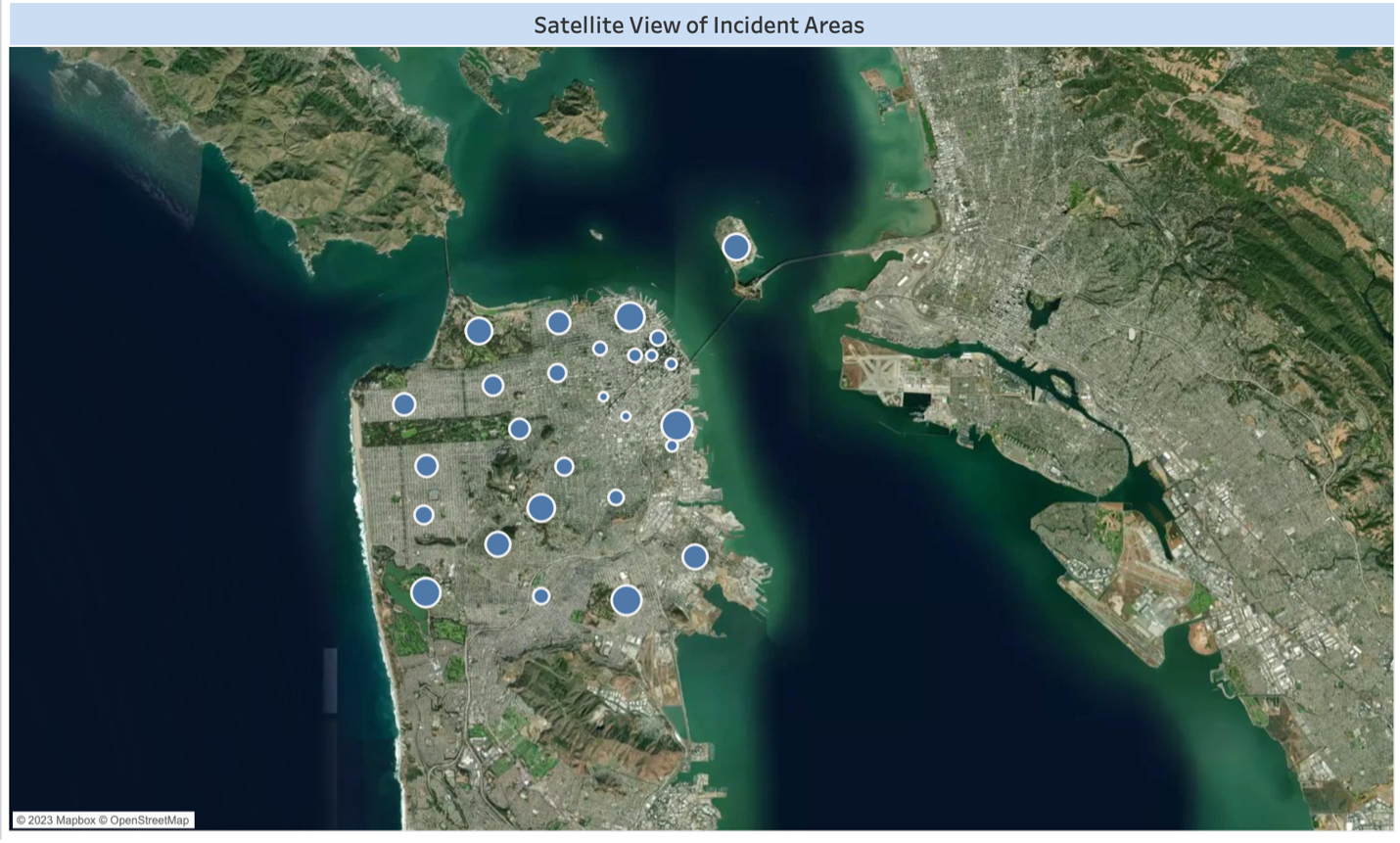


*Note.* A bar plot showing number of responses per Call Type.

The above graph depicts the number of responses per call type. We can see that 1,867,316 medical incidents accounted for the majority of responses. Water Rescue had the fewest responses, with only 11,222. Alarms receive the second-highest number of responses, followed by structure fire and traffic collision. Compared to the prior Call Types, the Other Call Type did not receive as many responses.

**Figure 15**

*Satellite View of Incident Areas*

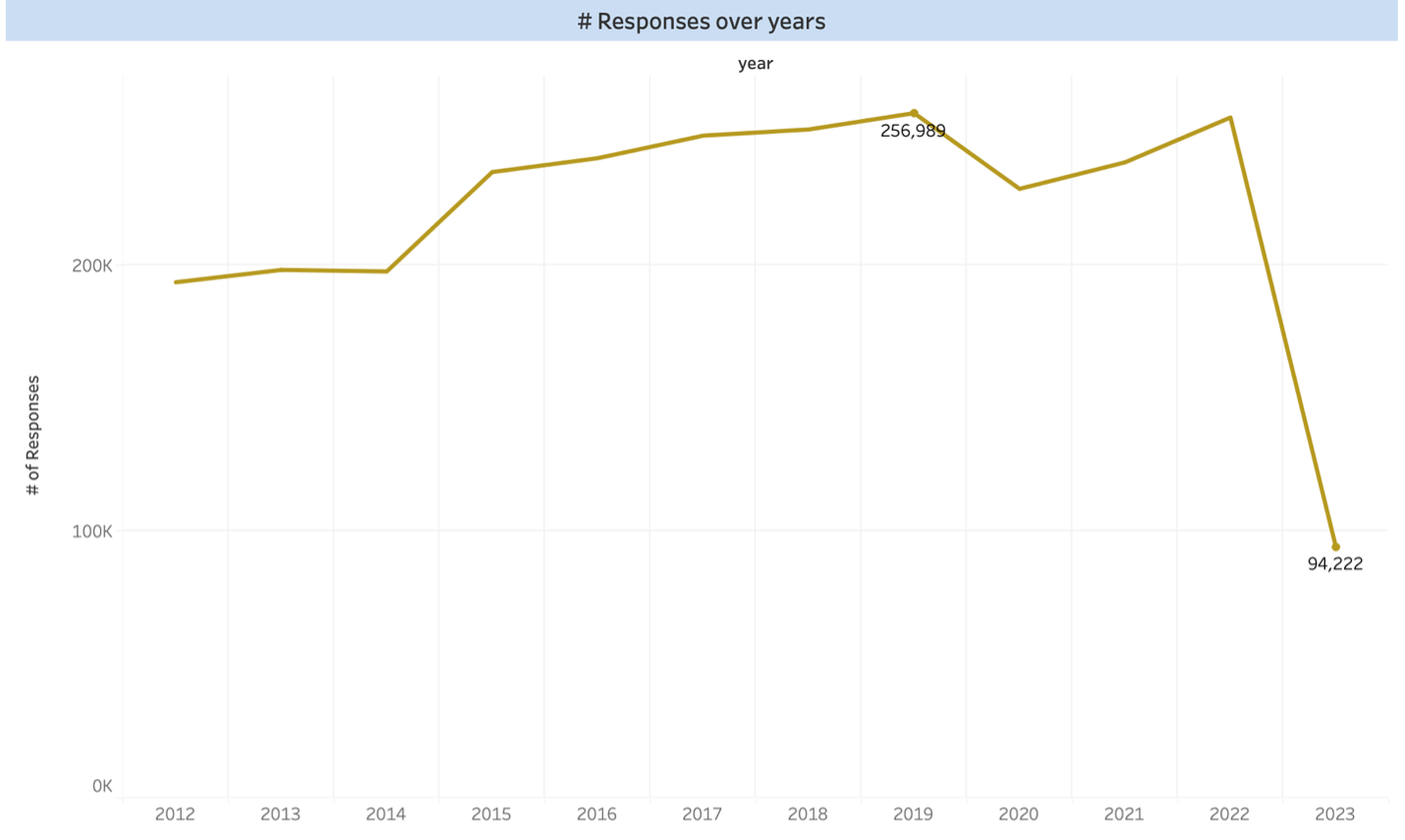


*Note.* A map showing satellite view of incident areas.

The above map is a satellite view of the areas where the SF fire department has responded to service calls. Some of the areas have bigger circles indicating areas having most number of responses from the department. Bigger the circles indicates more number of responses. Smaller the circle indicates very few number of Responses.

**Figure 16**

*Number of responses over the years*

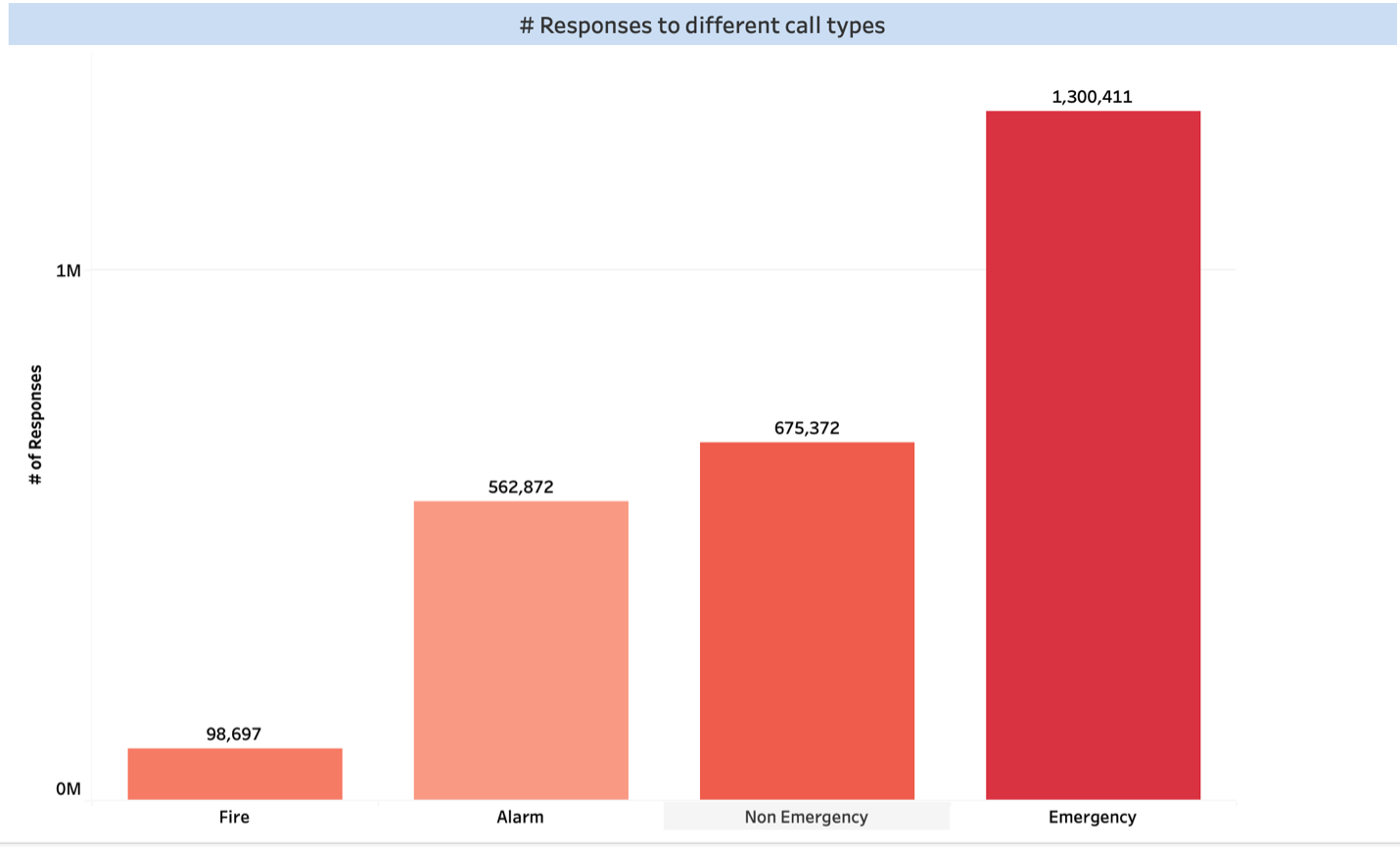


*Note.* A line plot showing number of responses over years.

Above is a line chart showing the number of Responses given by the fire department during the years. The data starts from the year 2012 till the year 2023. We can see that most number of Call Responses are seen in the year 2019 with 256,989 Responses. Followed by the year 2022 with 255,350 Responses. Data from the year 2023 is yet to get updated for the call responses.

**Figure 17**

*Number of Responses to different Call Types*

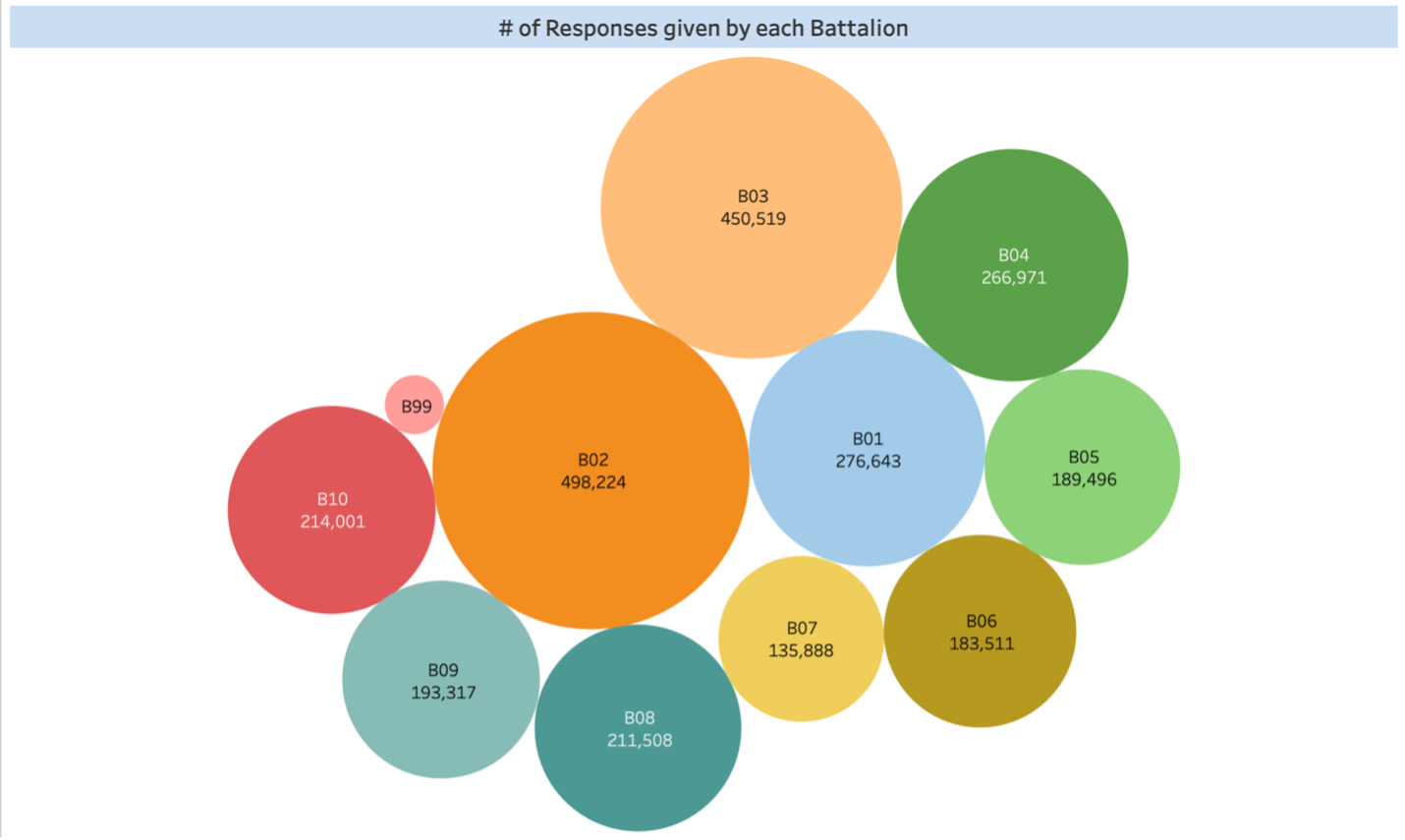


*Note.* A bar plot showing number of responses per Call Type group.

The above graph is a bar graph which illustrates Responses to different call types. The Call Types are categorized as Emergency, Non-Emergency, Alarm and Fire. The most Responses received were for Emergency Call Type which is about 1,300,411. Non-Emergency comes in second with 675,372. Followed by the Alarm Call Type which are 562,872 and the least Responses are for Fire Call Types which are 98,697.

**Figure 18**

*Number of Responses given by each Battalion*

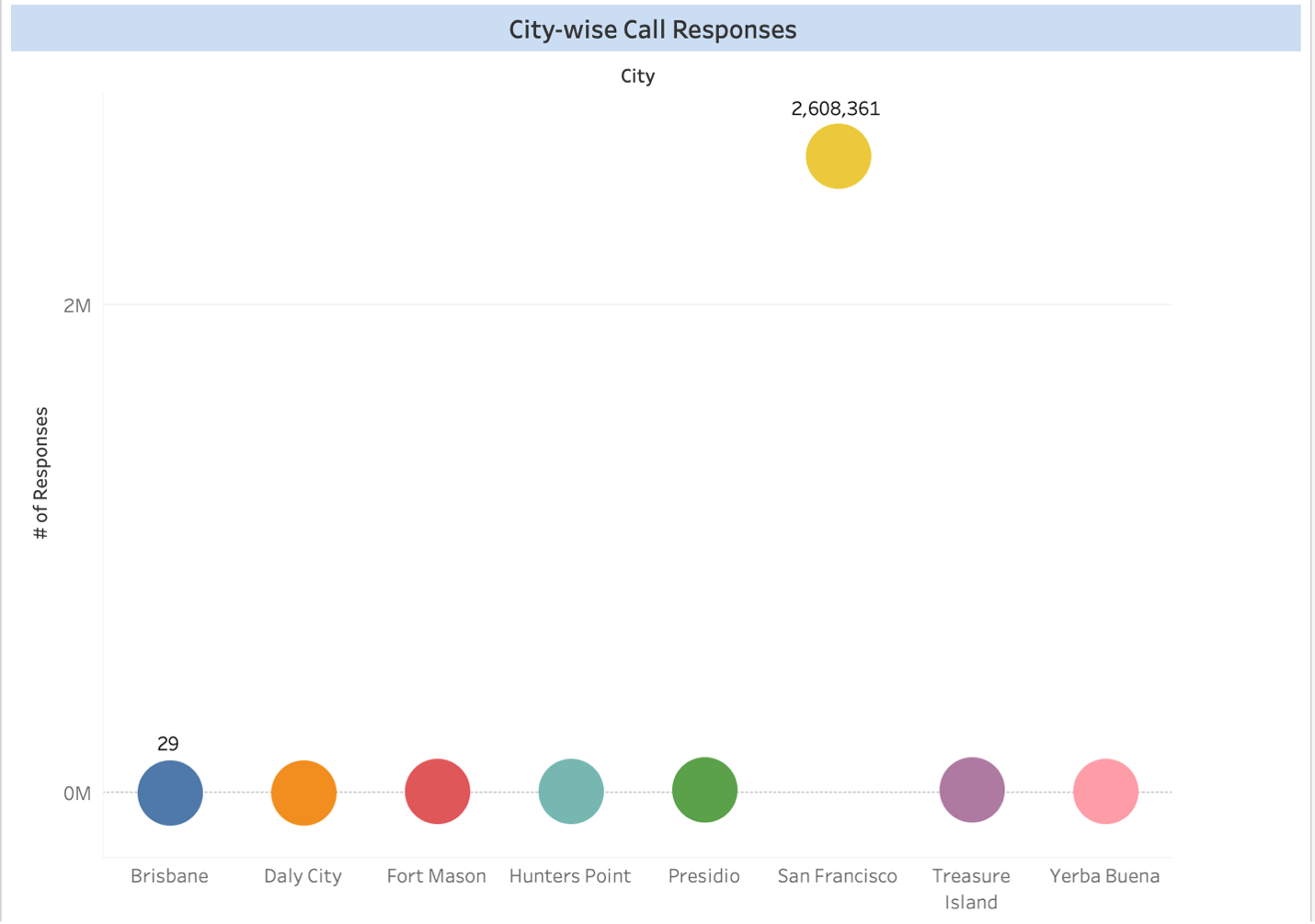


*Note.* A bar plot showing number of responses per Battalion.

The above graph illustrates the Responses given by each Battalion. The size of Bubbles indicates the number of Responses. The bigger the Bubble, the more number of responses it has given. According to the graph the most Responses were given by the B02 with 498,224. Followed by B03 with 450,519 Responses. The least number of Responses are given by B99 which are 17,272 Responses.

**Figure 19**

*City-wise Call Responses*



*Note.* A bar plot showing number of responses per City.

The above depicts City-wise Call Response. The City which had most Responses is San Francisco having almost 2,608,361 Responses. The next City having major Responses are Presidio with 12,295 Responses. The city with least Responses is Brisbane with only 29 Responses. Overall, we can see cities like San Francisco, presidio, Treasure Island are having majority Responses and cities like Daly City, Yerba Buena and Brisbane have least number of Responses.

**Chapter 6 System Testing and Experiment**

**6.1 Testing and Experiment Scope**

The test process and experiment scope involve a comprehensive set of test procedures and experiments aimed at evaluating the performance, functionality, and reliability of the system components and system. The test focuses on verifying the correctness, completeness, and consistency of the system components and system. The test process focuses not only on AWS services but also Tableau. For this project, the testing was carried out for individual components like connection between AWS S3 and Glue ETL job, AWS Athena connection with Tableau, execution of Glue workflow, connection between Tableau Desktop and Tableau Public etc. The testing was done after the prerequisites were completed for the components.

**6.2. Testing report**

Below tables shows the test report involving the tests performed on the system to check for its effectiveness, connectivity with external applications.

Test Case 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Test No** | 1 |  |  |
| **Test Title** | Connection between S3 and Glue Job | | |
| **Test Purpose** | Read input file from S3 via Glue Job | | |
| **Test Setup** | Script is created and executed | | |
| **Prerequisites** | Input file availability in S3 | | |
| **Procedure** | Reading file from S3 using a PySpark script and generate output file in S3 bucket | | |
| **Checks** | Data table availability in Athena for query | | |
| **Expected Results** | Output files is S3 | | |
| **Result** | Output file is generated in S3 bucket | | |
| **Reason for Failure** | - | | |
| **Remarks** | Successfully completed | | |

Test Case 2

|  |  |  |  |
| --- | --- | --- | --- |
| **Test No** | 2 |  |  |
| **Test Title** | Execute Glue Workflow | | |
| **Test Purpose** | To execute Glue job and Crawler under one Workflow | | |
| **Test Setup** | Glue workflow is created | | |
| **Prerequisites** | Glue job and Crawler are created | | |
| **Procedure** | Execute Glue Workflow | | |
| **Checks** | NA | | |
| **Expected Results** | Successful execution of Glue workflow | | |
| **Result** | Glue Workflow executed successfully | | |
| **Reason for Failure** | - | | |
| **Remarks** | N/A | | |

Test Case 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Test No** | 3 |  |  |
| **Test Title** | Connection between Tableau Desktop and AWS Athena | | |
| **Test Purpose** | To connect AWS Athena to Tableau Desktop to create visualizations | | |
| **Test Setup** | Access Key ID and Secret Access Key | | |
| **Prerequisites** | Data availability tables in Athena | | |
| **Procedure** | Configuring settings for Athena and Tableau | | |
| **Checks** | NA | | |
| **Expected Results** | Connection successful between Athena and Tableau | | |
| **Result** | Athena is connected to Tableau successfully and data is available for visualizations | | |
| **Reason for Failure** | - | | |
| **Remarks** | N/A | | |

Test Case 4

|  |  |  |  |
| --- | --- | --- | --- |
| **Test No** | 4 |  |  |
| **Test Title** | Connect Tableau Desktop to Tableau Public | | |
| **Test Purpose** | To get embed code for dashboard in Tableau Public and host it using S3 | | |
| **Test Setup** | Publish dashboard to Tableau Public | | |
| **Prerequisites** | Dashboard is created | | |
| **Procedure** | Use Publish Workbook feature from Tableau Desktop | | |
| **Checks** | NA | | |
| **Expected Results** | Dashboard visible in Tableau Public | | |
| **Result** | Dashboard published in Tableau Public | | |
| **Reason for Failure** | - | | |
| **Remarks** | We used extracted data from Athena to connect to Tableau Public | | |

**Chapter 7 Conclusion and Future Work**

**7.1 Project Summary**

The project started by understanding the problem and background of the San Francisco Fire Department. After which, the dataset search and collection activities were carried out. AWS services such as S3, Glue, and Athena were leveraged for storing the data and performing transformations. The transformed data was then visualized using Tableau, and a dashboard was built to showcase the trends and insights. The project was successful in providing valuable insights of the SFFD data, such as the number of responses over the years, the number of different types of calls, the number of responses per call type, and more.

In addition, a static website is hosted using Amazon S3 and Route 53, providing easy access to the dashboard visualizations and a report consisting of project methodology, architecture, implementation, and results is created. Overall, this project helped in demonstrating the effective use of AWS services such as AWS S3, AWS Glue Studio, AWS Route 53, AWS Athena, CloudWatch etc. for data transformation and analysis, as well as the data visualization in communicating insights and recommendations. The team gained valuable experience in working with AWS tools and services, and extensive use of Pyspark and SQL for data processing, and Tableau for visualization. Throughout the project, we worked collaboratively, which gave us valuable experience in working in a team environment and developing effective communication and collaboration skills. We also learnt how to manage project timelines and deliverables, as well as how to present project results and recommendations in a clear and concise manner. Overall, this project provided a unique and challenging opportunity to apply the skills and knowledge in a real-world setting, and it has greatly enhanced our abilities as a data analyst.

**7.2 Future work**

There are several future directions and potential areas for future work that can be explored based on the results and findings of this project. One possible direction is to conduct an in-depth analysis of the response times of SFFD to emergency calls. SFFD collects data that includes the time a call was received and the time the first response unit arrived on the scene. This information could be used to optimize the response times and improve emergency services for the citizens of San Francisco. Another research area is to analyze the correlations between different variables in the SFFD dataset. For example, we could investigate the relationship between the type of incident and the response time, or between the location of the incident and the number of firefighters required. By identifying these correlations, we could gain a better understanding of the underlying factors that contribute to successful emergency response and inform future decision-making processes.

In addition, we could also consider integrating external datasets with the SFFD data to provide more context and insights. For instance, we could include weather data to determine if weather conditions are responsible for a number of emergency calls. This way, we could provide actionable recommendations that can inform policies and strategies.

Finally, the possibility of creating predictive models that can forecast emergency incidents and their potential impact can also be studied. Leveraging machine learning algorithms and historical data, we could develop models that can predict the likelihood of an incident occurring in a specific area, its severity, and the resources that may be required to respond to it. These predictive models could assist emergency services in planning and allocating resources in a more effective and efficient manner.

**References**

Buffington, T., & Ezekoye, O. A. (2019). *Statistical Analysis of Fire Department Response*

*Times and Effects on Fire Outcomes in the United States*. Fire Technology, 55(6),

2369–2393. <https://doi.org/10.1007/s10694-019-00870-4>

Creation of Static Website. May 04, 2023.

<http://noobcoders.link.s3-website-us-east-1.amazonaws.com>

*Fire Department Analysis Report*. [www.springfield.il.us](http://www.springfield.il.us). Retrieved May 07, 2023 from

<https://www.springfield.il.us/Docs/SpringfieldILFireAnalysis012722FinalReport.pdf>

*Fire Department Calls for Service | DataSF | City and County of San Francisco*. (2023, May 7). <https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3>

Xu Lian, Sarah Melancon, Jon-Ross Presta, Adam Reevesman, Brian J. Spiering, Diane

Myung-Kyung Woodbridge (2019, August 19). *Scalable Real-time Prediction and*

*Analysis of San Francisco Fire Department Response Times*. IEEE Xplore. Retrieved May 06, 2023 from <https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00154>