# Analysis of emergency response services data to determine efficient positions of the response teams

Harsh Yadav(hy1217@nyu.edu), Datta Sainath D(dsd298@nyu.edu), Madhu Kiran Gudivada(mg5309@nyu.edu

***Abstract*— The fire service exists today in an environment constantly inundated with data, but data is seen of little use in the real world in which first responders live and work. The fire service is changing, it is not fighting fires as much as it is doing EMS, HAZMAT, inspections, investigations, prevention, and other nontraditional but important tasks which are vital to the community. Balancing limited resources and justifying daily operations and finances in the face of tough economic times is essential to every department.**

***Keywords—Hadoop, fire service, R, Pyspark, Tableau, Google Maps API, data analysis, data visualization, python.***

## I. INTRODUCTION

The primary objective of the project is to derive statistical techniques for analyzing data typically collected in fire departments. As improved resource allocation and statistical analysis of emergency medical calls can determine the impact of providing another paramedic unit in the field. The other goal of the project is to visualize incidents in an area and find the most efficient position for future fire stations.

## II. MOTIVATION

Motivation for the project stems from the belief that fire departments collect an immense amount of data, but do very little with it. The reports can provide a beneficial service to fire departments by yielding insights into the nature of fires and injuries in their jurisdictions. Most of the training on fire-fighting is based on a curriculum that has been in place for many years. It makes sense to see how training matches characteristics of fires in a particular jurisdiction. This is not to say that other training is not as important, since an exception can always occur. However, knowing more about the fires in a jurisdiction can improve services.

## III. DATA

### III.I. DATA DESCRIPTION

We have gathered data from the NYC Open Data website. The Fire Incident Dispatch Data contains data that is generated by the Starfire Computer Aided Dispatch System. The data spans from the time the incident is created in the system to the time the incident is closed in the system. It covers information about the incident as it relates to the assignment of resources and the Fire Department's response to the emergency. The data contains key details like DATE/TIME, ZIP CODE, INCIDENT_RESPONSE_SECONDS etc. which plays a significant role in the analysis of the data.

### III.II. DATA EXTRACTION

For extracting the coordinates of different incident locations for our project, we have used the Google Maps API. We used the incident address as input and extracted the longitude and latitude values of that address using the API and then used it for our analysis for visualizing as well as for finding the best locations for Fire Stations within a zip code using the K means clustering algorithm.

## IV. TECHNOLOGIES USED

For analyzing the Fire Incident Data, we have used PySpark to clean the data from null values and then analyzed the cleaned data using the Apache Zeppelin Notebook. The results are then interactively visualized using R and heat maps are created using Tableau software. We also used Spark ML to identify efficient positions for response teams and simulated them on a map.

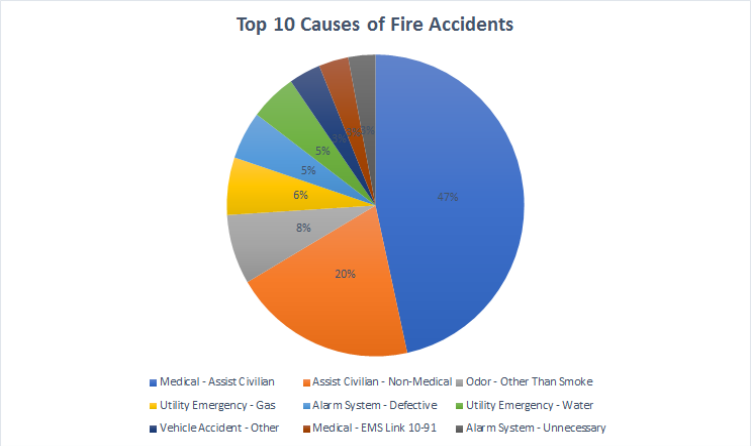
## V. PROJECT ARCHITECTURE

In our project we have performed analysis on the EMS and data and for that we have taken the data from the Open Data NYC website. Then we converted the incident locations to coordinate values. From this data, we created various visualizations using R, Tableau and Google Maps API. Further, we used the Spark ML libraries to find the optimal locations for fire stations within a zip code to reduce the response time to any incident location.

**The architecture diagram is attached at the end of the report.**

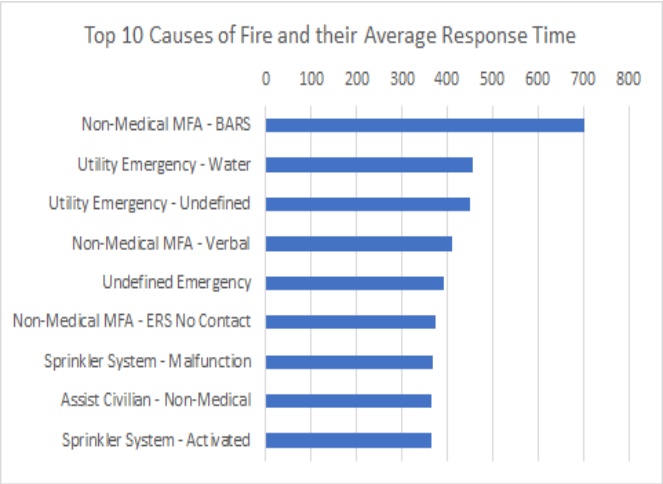
## VI. EXPERIMENTAL RESULTS

### VI.I. Top 10 Causes of Fire Incidents



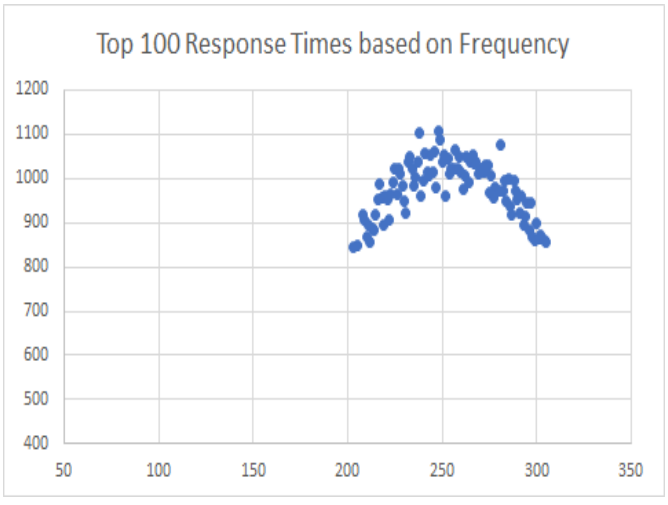
2016 were identified and visualized using a pie chart that showed a huge majority, as high as 47% of the incidents were related to assisting the EMS response team.

### VI.II. Top 10 Causes having the highest Response Time



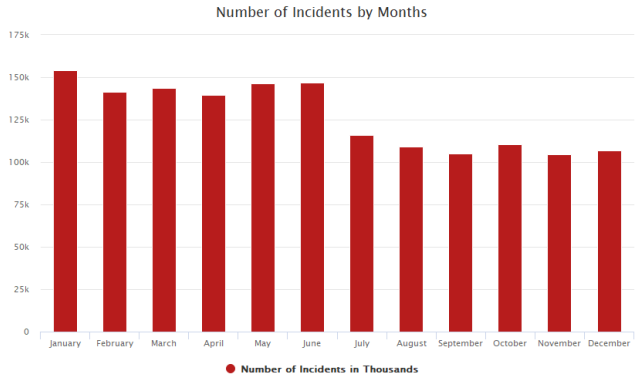

Here we tried identifying the top 10 causes of fire that take the highest average response time. This helps us identify what type of emergency is taking up huge amount of time and will help the fire department to come up with better strategies to combat these problems. We also observed that a vast number of incidents were water related issues which took up a lot of time.

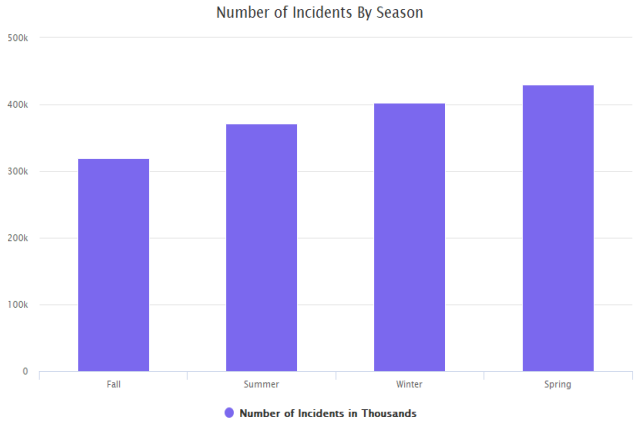
### VI.III. Top 100 Response Timings Based on Frequency




Now we tried to see a scatter plot of the top 100 response times based on their frequency. The above graph shows that most of the responses take about 200 secs to 300 secs. This is called as critical time range that defines the overall performance of the fire response teams.

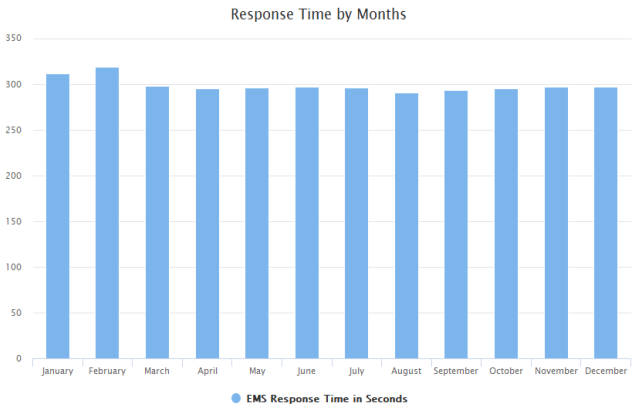
### VI.IV Number of Incidents per Month and Season

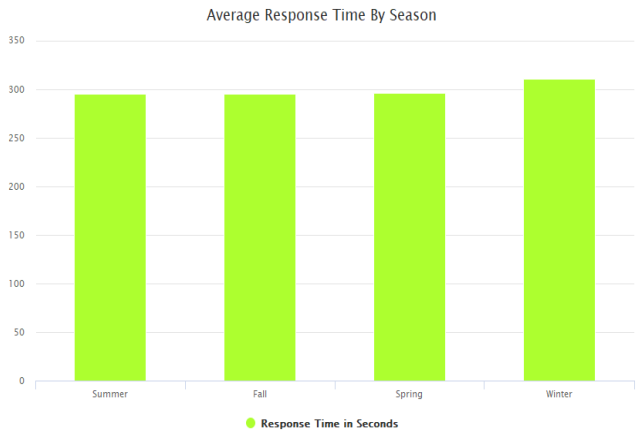



Now we tried to establish a relation between the months and the number of incidents. The highest number of incidents happened during January followed by May and June. To get a better understanding, we grouped them into seasons which showed that most incidents occur during Spring season. This data helps us to identify when the fire department need most volunteers which is a common program New York City.

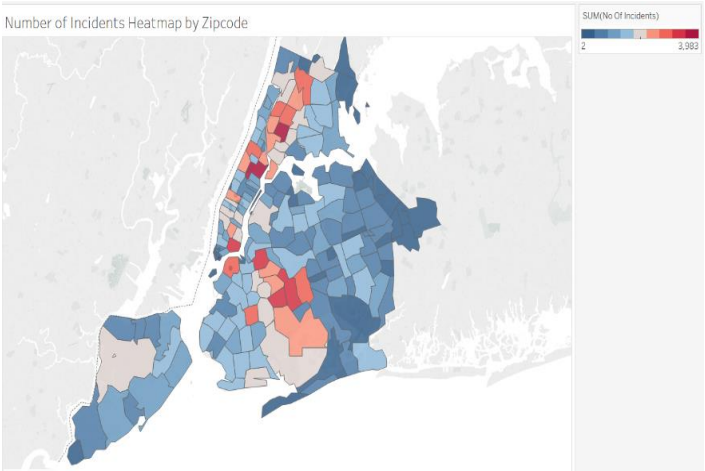
### VI.V. Average Response time per Month






Next, we have tried to establish a relation between the average response time (in seconds) per month. This showed us that February and January months had higher average response time than others. When we grouped the data to show the average response time per season we can clearly see that Winter has the highest average response time as it snows in NYC and it usually takes more time to reach the same destination in Winter compared to other seasons.

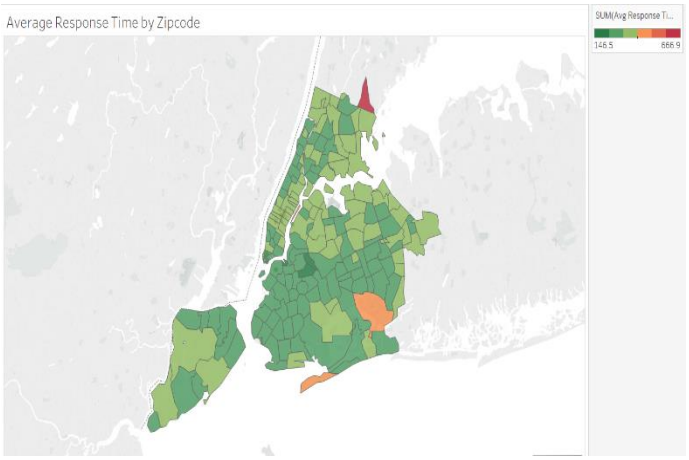
### VI.VI Number of Incidents per Zip code




Now, to identify other trends we created a heatmap based on the number of incidents per zip code. The heatmap shown above has zip codes with as less as 2 incidents (which are shown in dark blue) and as high as 3200 incidents (which are show in dark red). This shows that some areas require informative and training programs that would help reduce the number of incidents. Also, proper inspection of smoke alarm in areas with substantially elevated risk will help avoid huge accidents.

We have published an interactive heat map for better visualization:
https://public.tableau.com/profile/harsh.yadav#!/vizhome/NumberofFireIncidentsinNYC/No_ofIncidentsbyZipcode

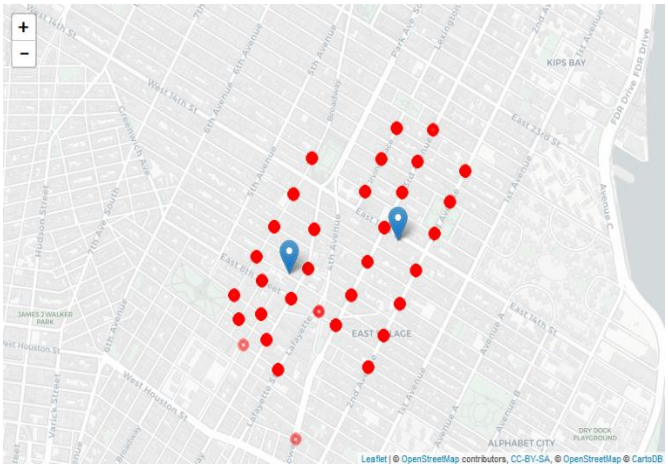
### VI.VII Average Response Time by Zip Code




Here, we can see that the average response time per zip code seems to be almost the same in most of the zip codes except a few outliers. This shows that all the fire stations are equally equipped, and the personnel have been trained similarly.

https://public.tableau.com/profile/harsh.yadav#!/vizhome/AverageResponseTimeatindidentbyFireDepartmentbyZipcodeinNYC/Sheet1?publish=yes

### VI.VIII Identifying Fire Incidents in Zip Code 10003




For the secondary objective we tried to identify the most efficient position of response team by analyzing the locations of previous incident locations. We used K-Means algorithms with K=2 to predict the top 2 positions for a new fire station. The above graph shows the previous incidents in red and the efficient future positions of the response teams in blue markers.

Such analysis can help to identify the most efficient positions but may not a reliable position as a fire station requires various specific norms. But, if the city is able to provide such locations for the response teams then the response time can be reduced significantly.

## VII. CONCLUSION

From the given data, we can see that a lot of statistical analysis can be done. This analysis can help us to come up with solutions that could not only improve the efficiency of the fire department but also save human life and prevent property damage. The analysis can help change volunteer and fire personnel training to match the characteristics of incidents in their areas.

**ARCHITECTURAL FLOW:**