

New York state leads nation in fire deaths in 2023. According to the [Firefighters](#) Association of the State of New York, New York holds the leading position in the country due to fire deaths. In this project, I will perform data exploration of Fire Incidents in NYC throughout different NYC boroughs to provide the government with useful information to improve their policy-making decisions.

In the first section of the project, we developed a python script that runs on Docker to build and connect our EC2 instance to the Elastic search domain. This allows us to load over 5M records to the Elastic search index. The data can be publicly accessed from [NYC Open Data](#). Our script supports the python command line interface to send API requests.

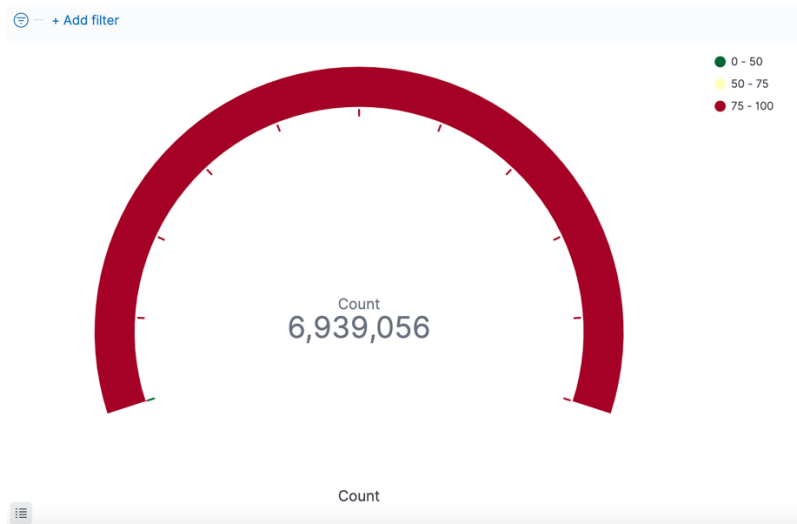
In addition, a docker image was created, containing the Dockerfile, python script, and all the requirements.

```
root@ip-172-31-11-99:/var/snap/amazon-ssm-agent/6312# docker build -t bigdataproject:1.0 .
```

```
root@ip-172-31-11-99:/var/snap/amazon-ssm-agent/6312# docker run -e INDEX_NAME="fire" -e DATASET_ID="8m42-w767" -e APP_TOKEN="Jwia91fbXpcTHOJ35d9nRfg8L" -e ES_HOST="https://search-project1-mariia-v3rbklzow42rdb3ggcskze5gk4.us-east-2.es.amazonaws.com" -e ES_USERNAME="mariiamohyla" -e ES_PASSWORD="cis-9760-Mariia" bigdataproject:1.0 --page_size=10 --num_pages=2
```

```
FROM python:3.9
WORKDIR /app
COPY . /app
RUN pip install -r requirements.txt
ENTRYPOINT ["python", "src/main.py"]
```

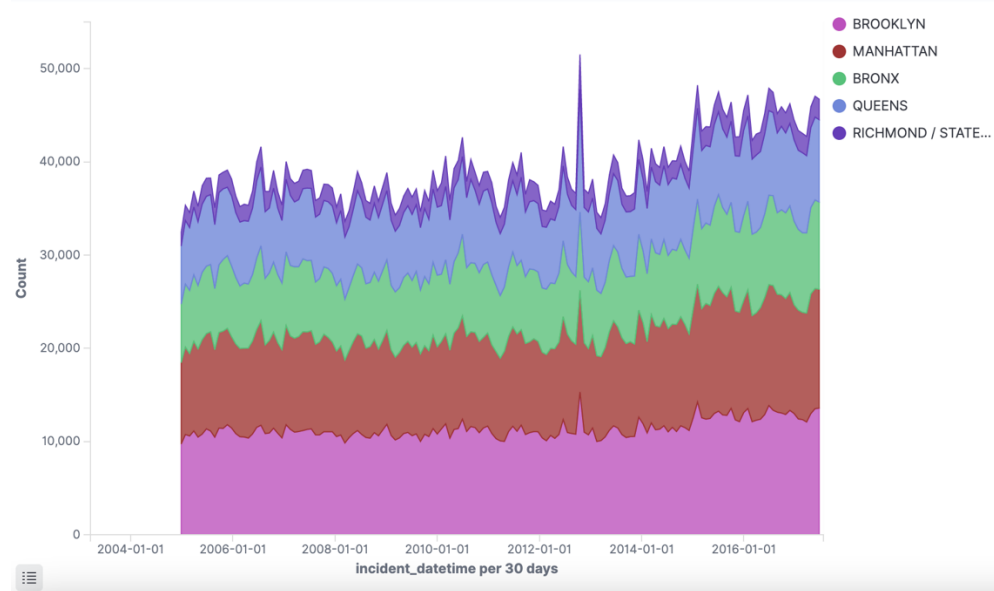
I managed to successfully upload over 6M rows to the Elasticsearch index. This can be seen in the picture below.



In the second section of the project, I used OpenSearch Dashboard (Kibana) to analyze and visualize data from January 2005 until July 2017 to answer the following questions:

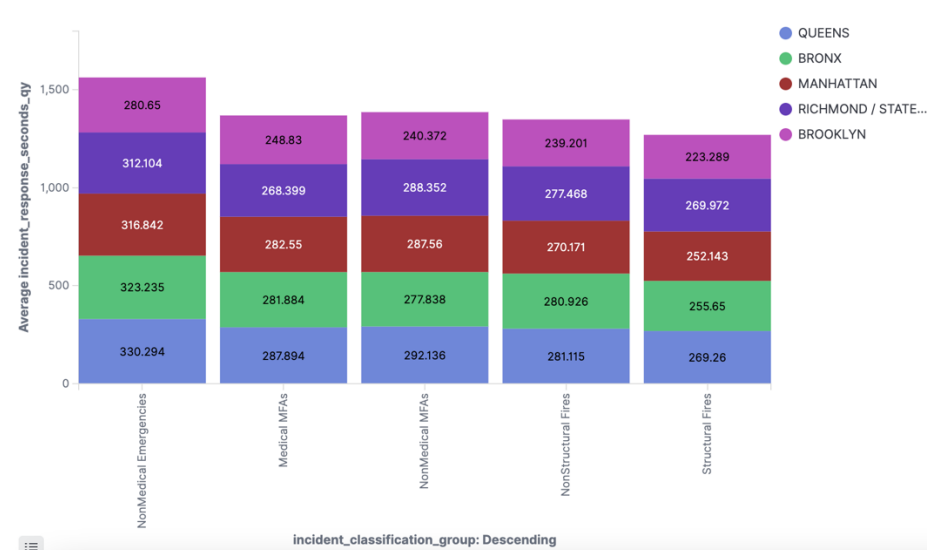
1. Which borough has the highest number of incidents?
2. What is the average incident response time for different types of incidents with respect to the borough?
3. What is the long-term trend of fire incidents for the analyzed years?
4. What is the average incident response time for each borough over time?
5. What are the most common types of incidents?
6. What are the top 10 zip codes that tend to have serious fire incidents?

1. Which borough has the highest number of incidents and how it changes over time?



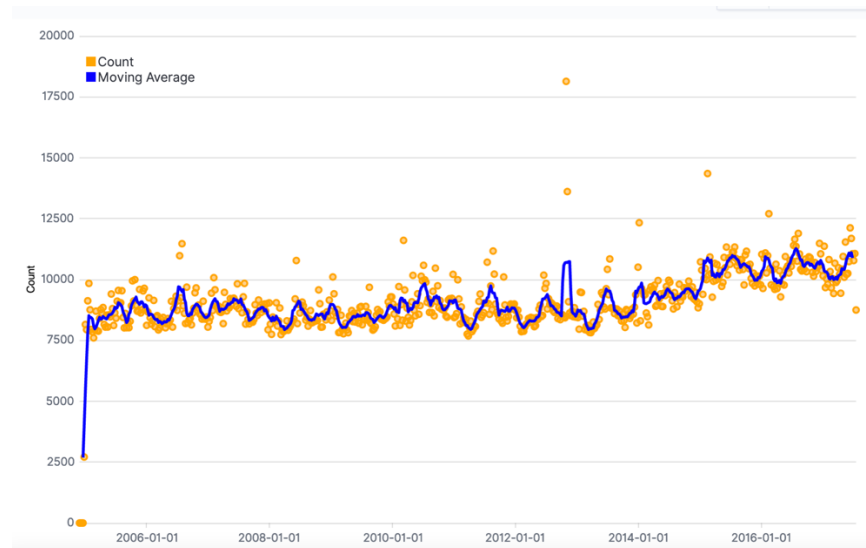
Brooklyn has the highest number of incidents while Staten Island has the lowest.

2. What is the average incident response time for different group types of incidents with respect to the borough?



As we may notice highest average response time is observed in non-medical emergencies in all boroughs. On the other hand, structural fires have a significantly lower response time in all boroughs, except Staten Island. Brooklyn has the lowest response time in all categories, while Queens needs improvements.

3. What is the long-term trend of fire incidents for the analyzed years?



To identify the trend direction of the number of incidents over time, I have performed a moving average analysis. As you can see, on average number of incidents remained stable until the end of 2013. Since then, the moving average is increasing, showing an upward trend in the number of incidents. This may be caused by the urbanization trend.

4. What is the average incident response time for each borough over time?



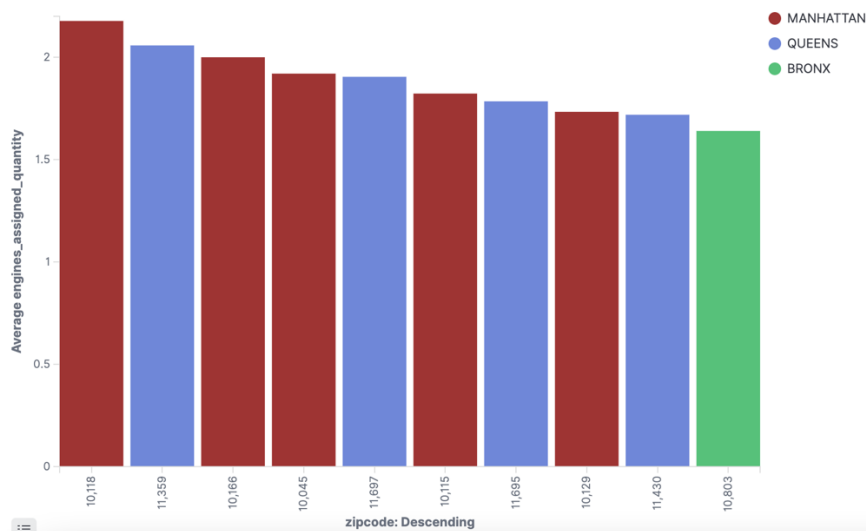
While looking at the histogram, we may observe that since the beginning of 2005, the average response time has slightly decreased in all boroughs. Brooklyn holds first place in having the lowest response time compared to other boroughs, despite the fact that it has the highest number of incidents. Manhattan, Queens, Bronx, and Staten Island fire departments need to decrease their response time.

5. What are the most common types of incidents?

Utility Emergency - Water
Medical - EMS Link 10-91
Assist Civilian - Non-Medical
Medical - PD Link 10-91
Odor - Other Than Smoke
Utility Emergency - Gas

Firefighters are obligated to respond to various complaints and aid people, as we can see the most popular complaints are medical and non-medical assistance. The government can utilize this information to equip locations with EMS providers, to balance out a load of emergency calls.

6. What are the top 10 zip codes that tend to have serious fire incidents?



I selected the top ten zip codes with the highest average number of engines assigned to the incident. I consider that these areas on average required more units compared to other zip codes. In most cases, the zip codes are located in Manhattan or Queens. To be more specific, this information can be used to increase fire-preventative measures in schools, hospitals, and local businesses in that areas.

A few takeaways from the analysis:

- Brooklyn has the highest number of incidents while keeping the lowest incident response time.
- There is a positive trend in the number of incidents that may be caused by urbanization and may require additional fire departments.

- The most popular types of complaints are medical and non-medical assistance, which means it can be more efficient to provide additional emergency room locations than building strategic fire departments.
- The top ten zip codes are identified due to the higher demand for fire engine units compared to other zip codes. Fire-preventative measures in schools, hospitals, and local businesses have to be in place and community outreach programs may be conducted. Also, the government can help increase building inspections.