Project 1 – Analyzing Millions of NYC Fire Incident Dispatch Data

Project Background

Data published by NYC agencies and other partners is made freely accessible by the NYC Open Data project. Depending on the department and time period, these datasets have anywhere from a few thousand to millions of records. Using EC2, the terminal, Docker, Elasticsearch, and other tools available in AWS, we can ingest, analyze and visualize a dataset from the NYC Open Data Project. We will use a service that provides an Application Programming Interface, or API, to make it simple to query and load the data via HTTP queries in terminal or code. In order to evaluate the data, this project will use a python client of the Socrata API to connect to the Fire Incident Dispatch Data API, load all the data into an Elasticsearch (OpenSearch) instance and proceed to analyze with OpenSearch Dashboards. After the data has been loaded and made available on OpenSearch, we will move on to creating a few visualizations to help us better understand the data.

Make sure you have the following:

- An app token for the NYC Open Data API
- An AWS account (for running EC2, etc)

Here are the steps to create and run a docker image in the command line.

Step 1.) Create Docker Image

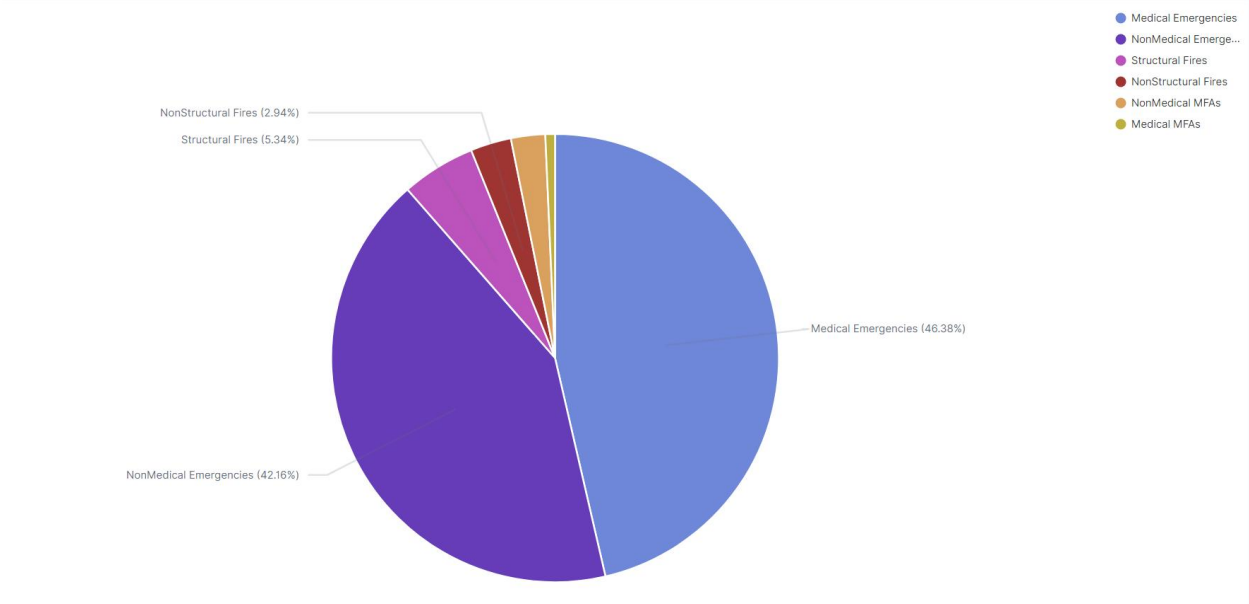
```
docker build -t bigdataproject1:1.0 .
```

Step 2.) Run Docker Image

```
docker run \
-e INDEX_NAME="fire" \
-e DATASET_ID="8m42-w767" \
-e APP_TOKEN= "your_token" \
-e ES_HOST= "your_domain" \
-e ES_USERNAME="your_username" \
-e ES_PASSWORD="your_password" \
bigdataproject1:1.0 --page_size=1000 –num_pages=4
```

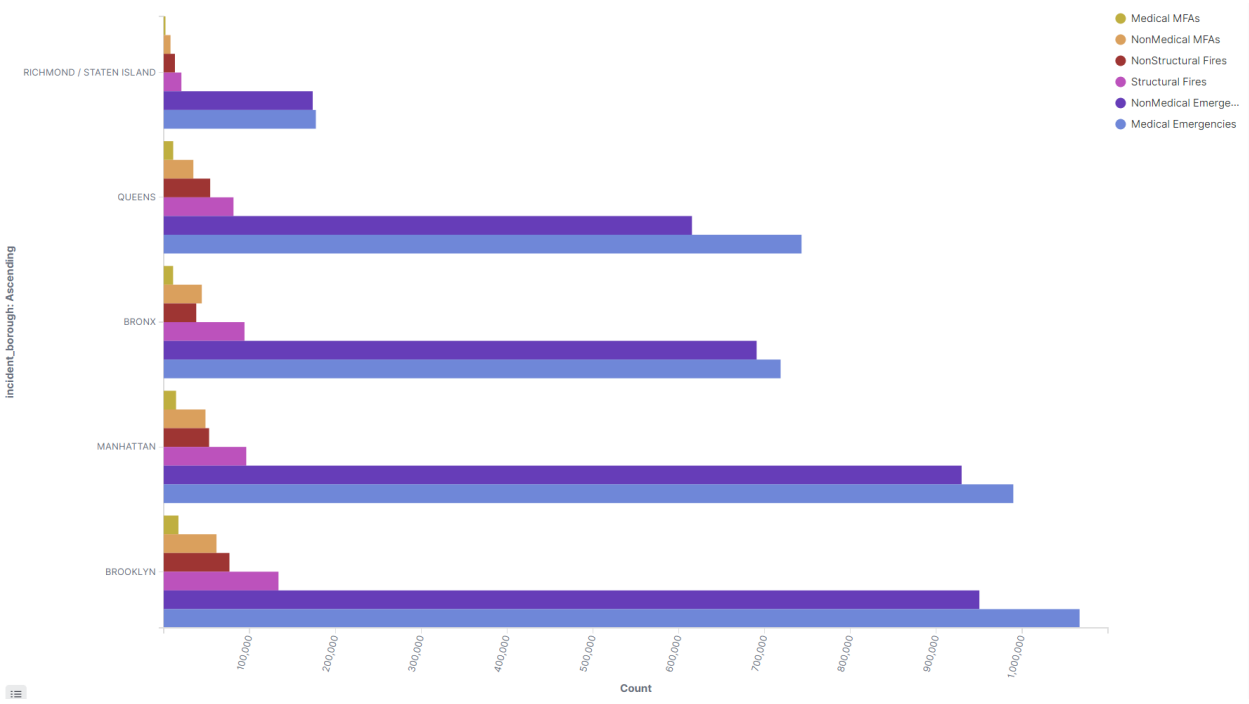
Visualizations

What is the numerical proportion of Incidents for all borough from 2004 to 2022?



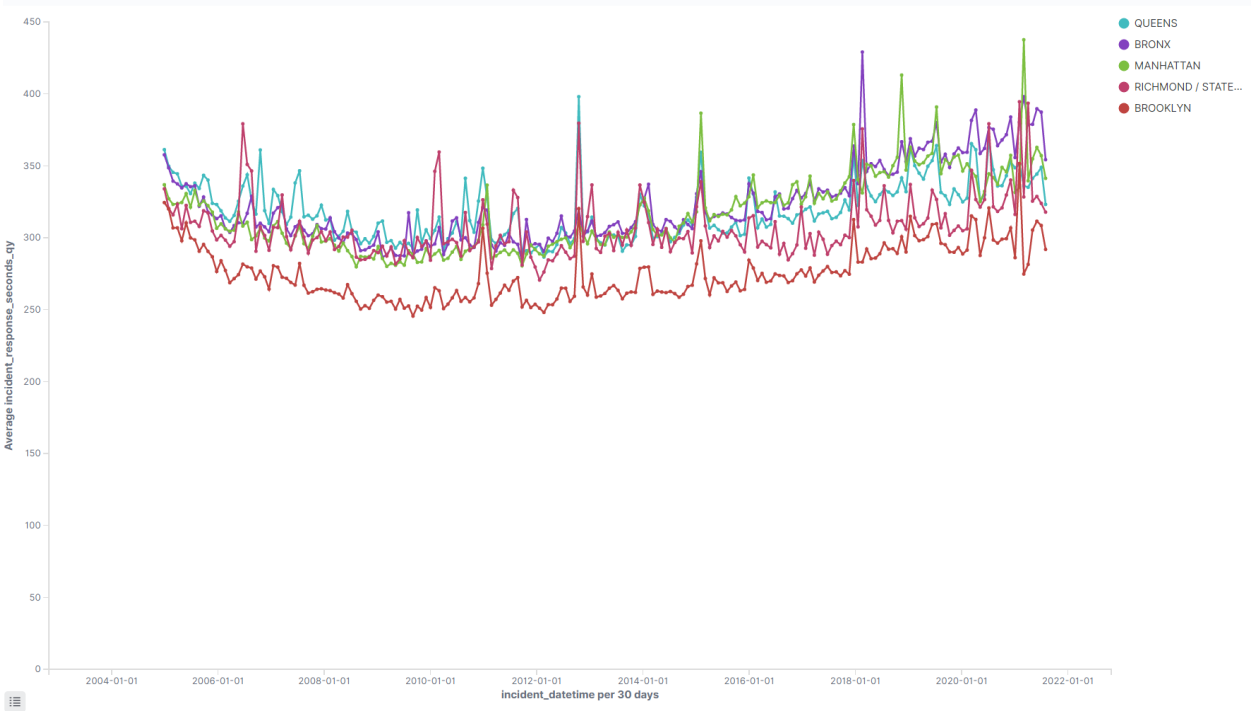
Incident Classification is divided into 6 categories, namely Medical Emergencies, Non-Medical Emergencies, Structural Fires (Building Fire), Non-Structural Fires (Outside and unclassified fires), Non-Medical MFA (Malicious False Alarm) and Medical MFA. Based from the data, the top 2 incident that occur 88% of the time is Medical Emergencies (46%) and Non-Medical Emergencies (42%).

What are the top Incident Classification Group for each Borough?



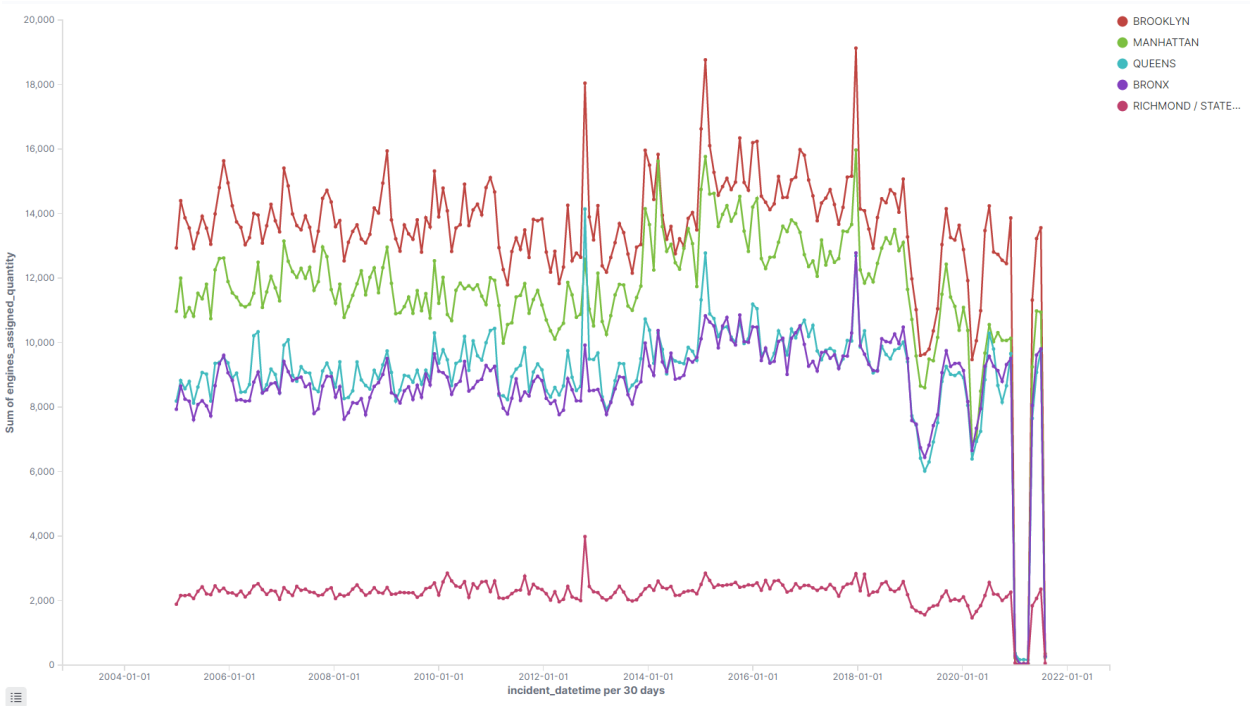
Based from the data, Medical & Non-Medical Emergencies are the top Incident classification for every borough. But it is interesting to know that Staten Island and Brooklyn has the least and the greatest number of total incidents of the 5 boroughs, respectively. Is it also interesting to know that Medical Emergencies often occur much more often compared to Non-Medical Emergencies in Brooklyn and Queens than the other 3 boroughs. Lastly, looking at Bronx, Non-medical MFAs occurs more often than Non-structural fires.

What is Average Incident Response Time per borough from 2004-2022?



Incident response time is defined as the time that passed between when an event was reported to the control center and when the incident management team arrived in person at the incident site. Based from the data, Brooklyn has an overall lower average out of all the 5-borough year after year of around 250-350 seconds.

How many engines are allotted to each borough year over year from 2004 to 2022?



Engines assigned quantity is defined as the number of vehicles sent to the incident site by the control center. Based from the data above, we can see that Staten Island has the least number of engines allotted year over year with a range from 2000-2500. On the other side, Brooklyn typically receives the most engines each year, with a range of 12500–19000.