**Project Background and Approach:**

The project involves the analysis of a vast dataset containing millions of records related to fire incidents in New York City. The dataset includes various variables such as dispatch response time, incident response time, and incident dates. The approach to this project can be summarized as follows:

1) **EC2** **Provisioning**: Setting up Amazon Elastic Compute Cloud (EC2) instances.

2) **Containerization**: Utilizing container technology to manage and deploy the project.

3) **Terminal** **Navigation**: Navigating and working with the system through the command-line terminal.

4) **Python** **Scripting**: Developing and using Python scripts for data manipulation and analysis.

5) **OpenSearch** **and** **Visualizations**: Employing OpenSearch for data storage and visualization.

**Project Process:**

- The first step was to create an Elasticsearch (ES) domain. The project utilized a master username and password for accessing the ES domain's endpoint and the OpenSearch dashboard URL.

- An EC2 instance was established and accessed via a web browser. This instance played a crucial role in setting up the required folder structure and project files.

- Preliminary tests were conducted to ensure that data retrieval was functioning correctly. Initially, a small number of rows were retrieved using the page\_size parameter for testing purposes.

In summary, this project aimed to analyze a substantial dataset of NYC fire incidents using various technologies and tools, including EC2, containerization, Python scripting, and OpenSearch for effective data storage and visualization.

**DOCKER RUN COMMAND:**

docker run -e **INDEX\_NAME**="fire" -e **DATASET\_ID**="8m42-w767" -e **APP\_TOKEN**="HE1n5LIqNun5UJUdHk0cwXw6F" -e **ES\_HOST**="https://search-nyc-fire-data-f4543eu5y6wx7xz7pfbegnu2dy.us-east-2.es.amazonaws.com" -e **ES\_USERNAME**="anmolmaheshwari" -e **ES\_PASSWORD**="Consistent10." bigdataproject1:1.0 --page\_size=2000 --num\_pages=3200

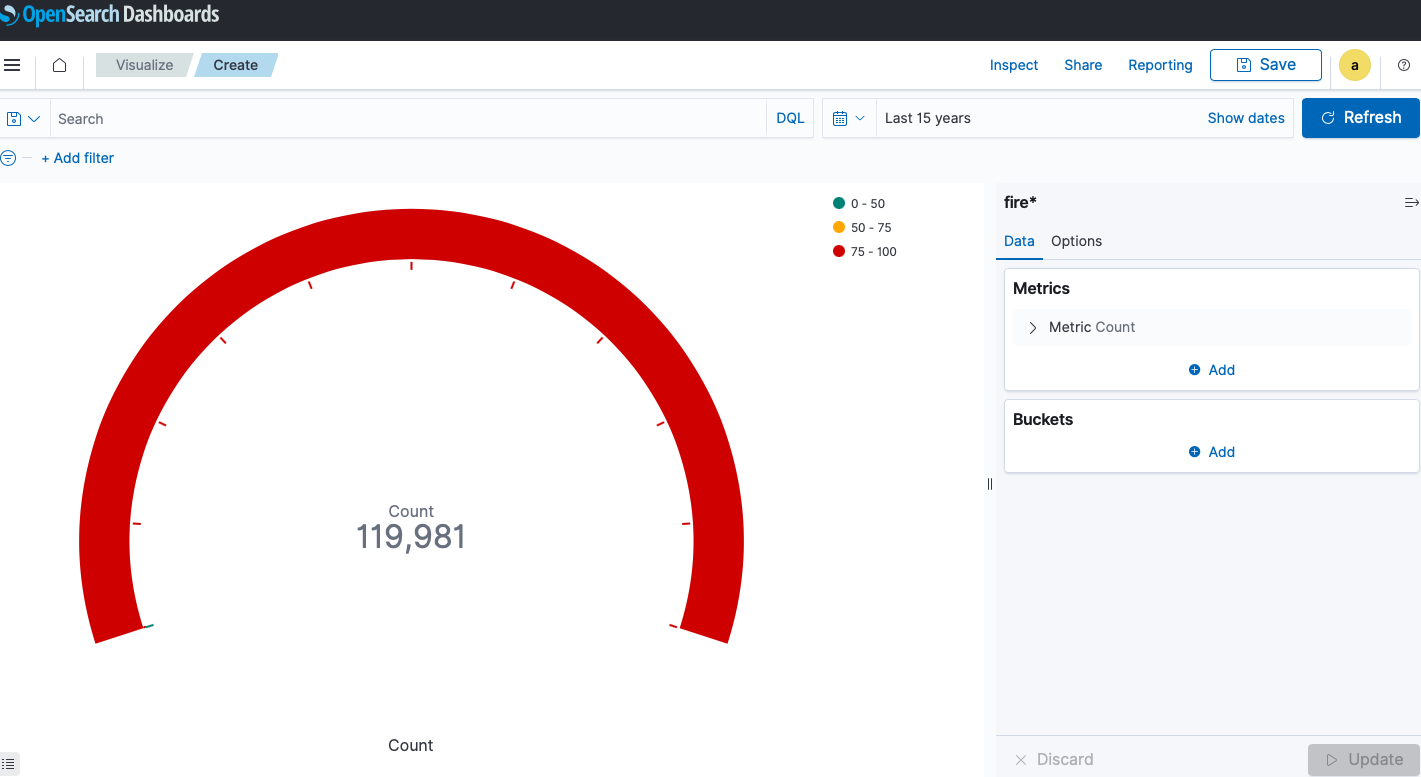
**OpenSearch Dashboards URL:**

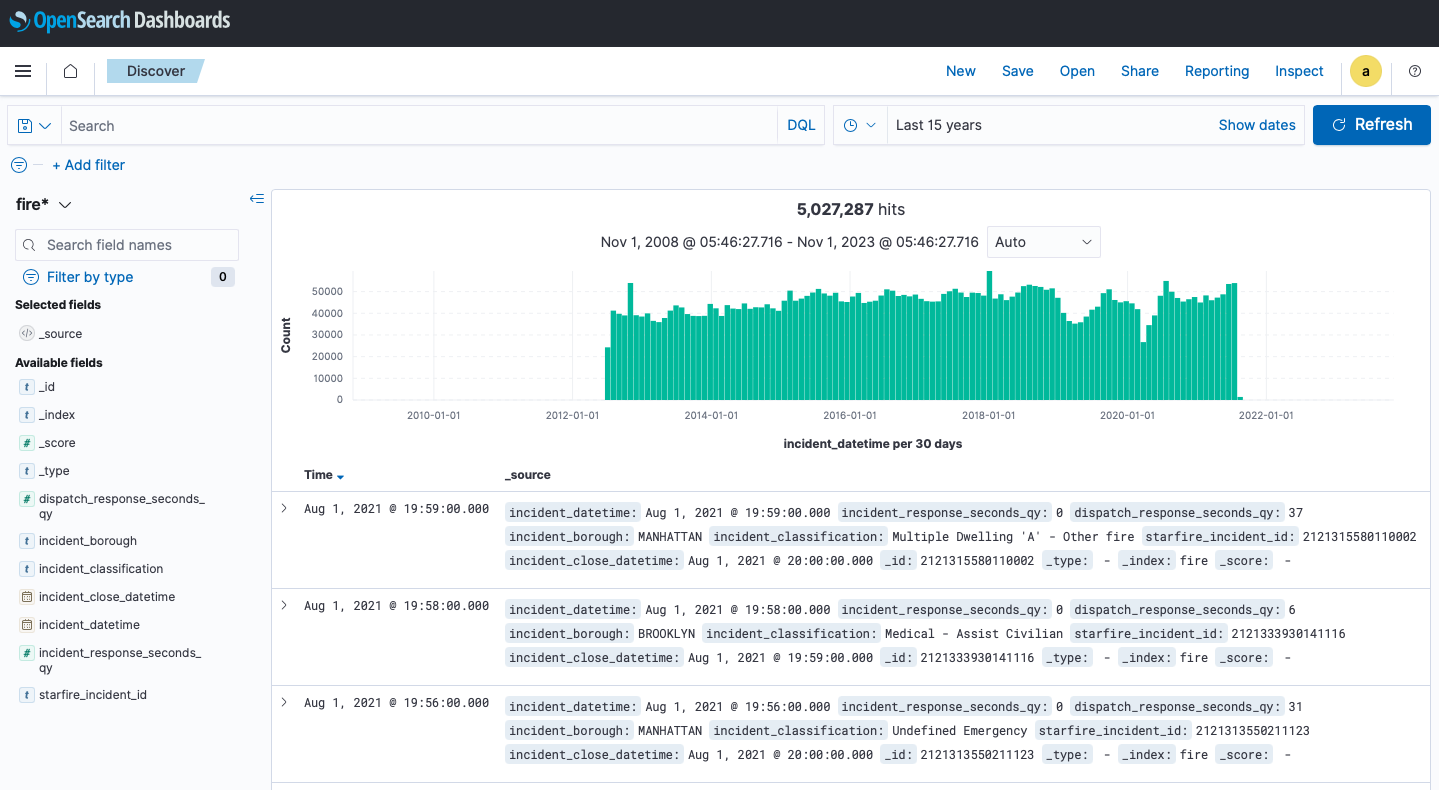
<https://search-nyc-fire-data-f4543eu5y6wx7xz7pfbegnu2dy.us-east-2.es.amazonaws.com/_dashboards>

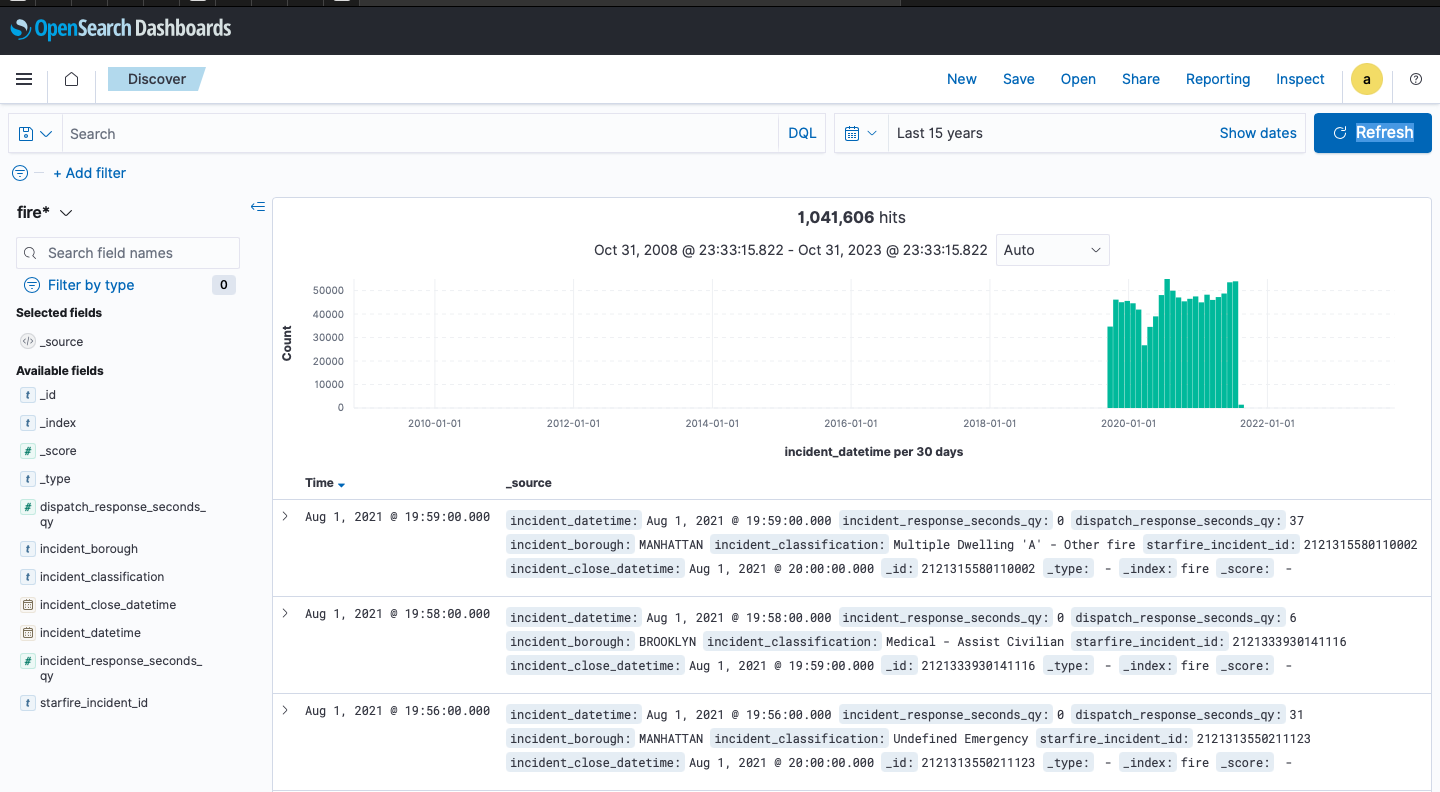
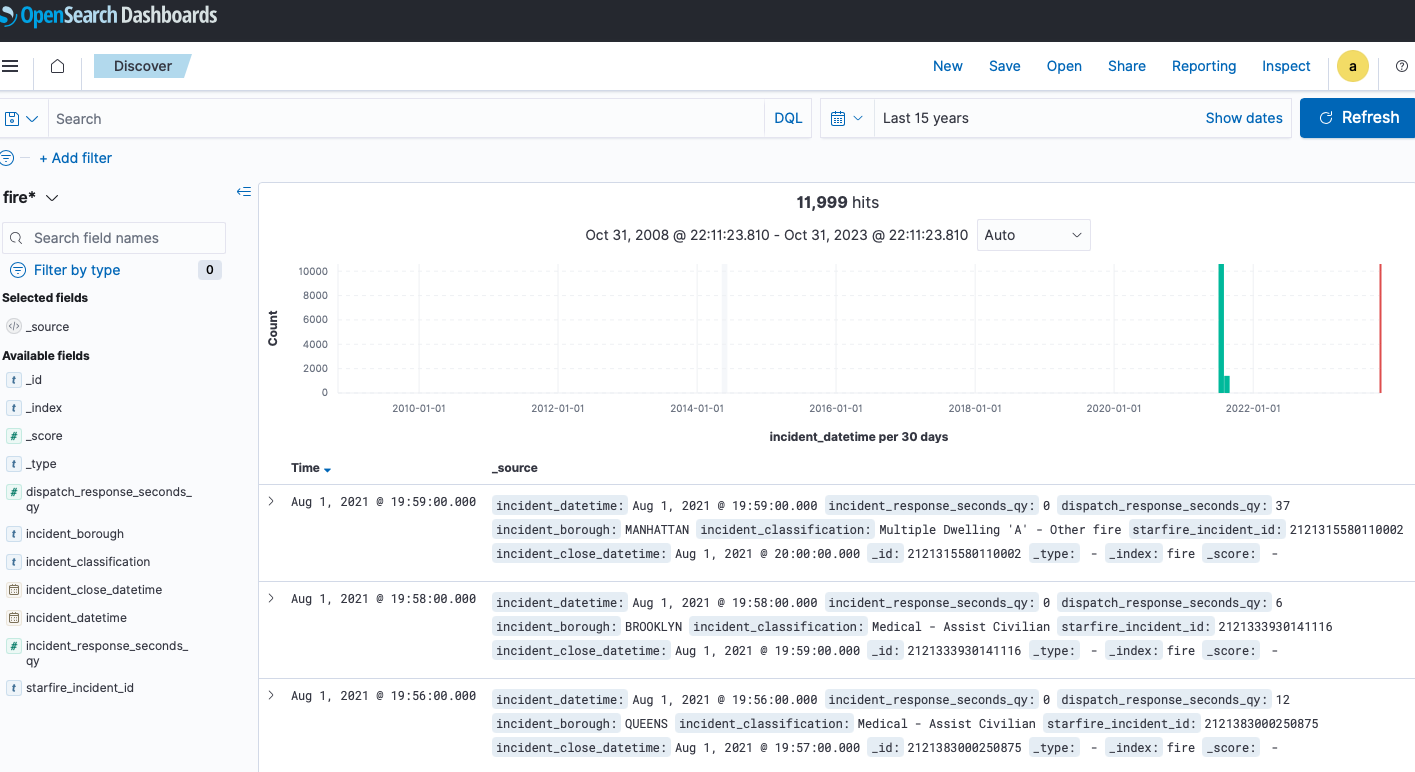
**Domain endpoint:**

<https://search-nyc-fire-data-f4543eu5y6wx7xz7pfbegnu2dy.us-east-2.es.amazonaws.com>

**Gauge chart for 1M+ rows and hits for 5M+:**



****



**VISUALIZATIONS:**

**Visual 1:**

**Q) Which Borough in New York City has the highest count of fire incidents?**

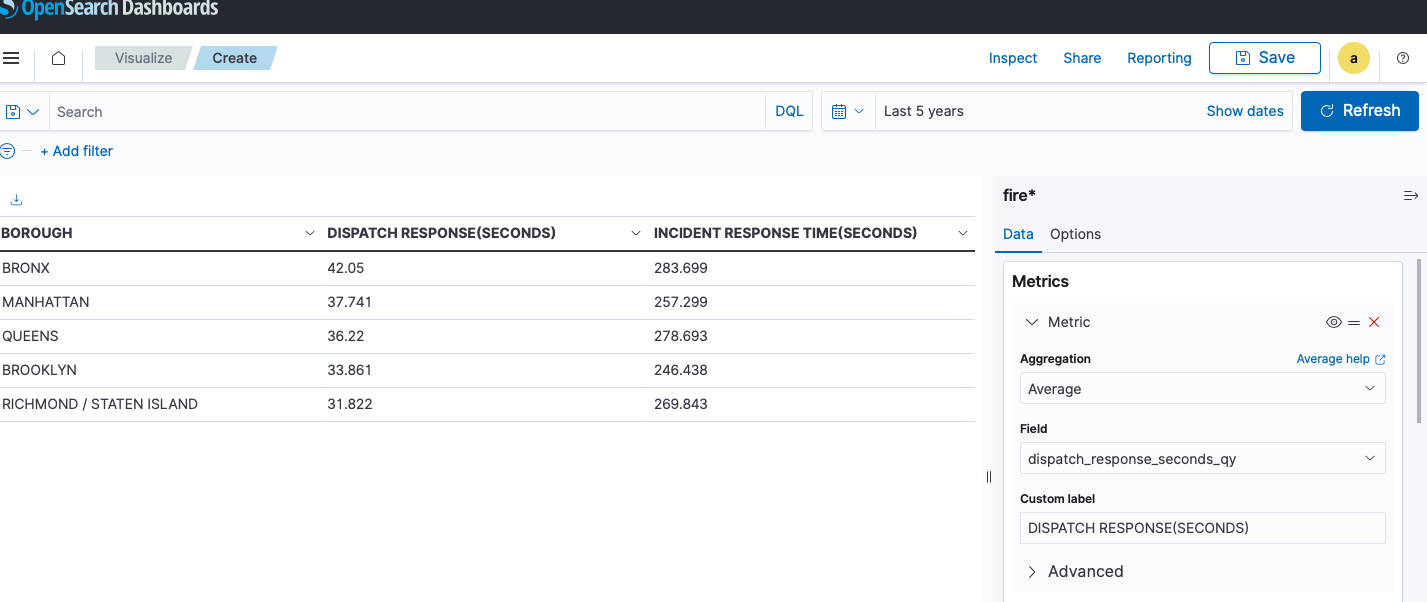
-Brooklyn is highest with count above 5 lakhs.



**Visual 2:**

**Q) Which NYC borough has the fastest average incident response time, and how does it compare to the borough with the slowest response time?**

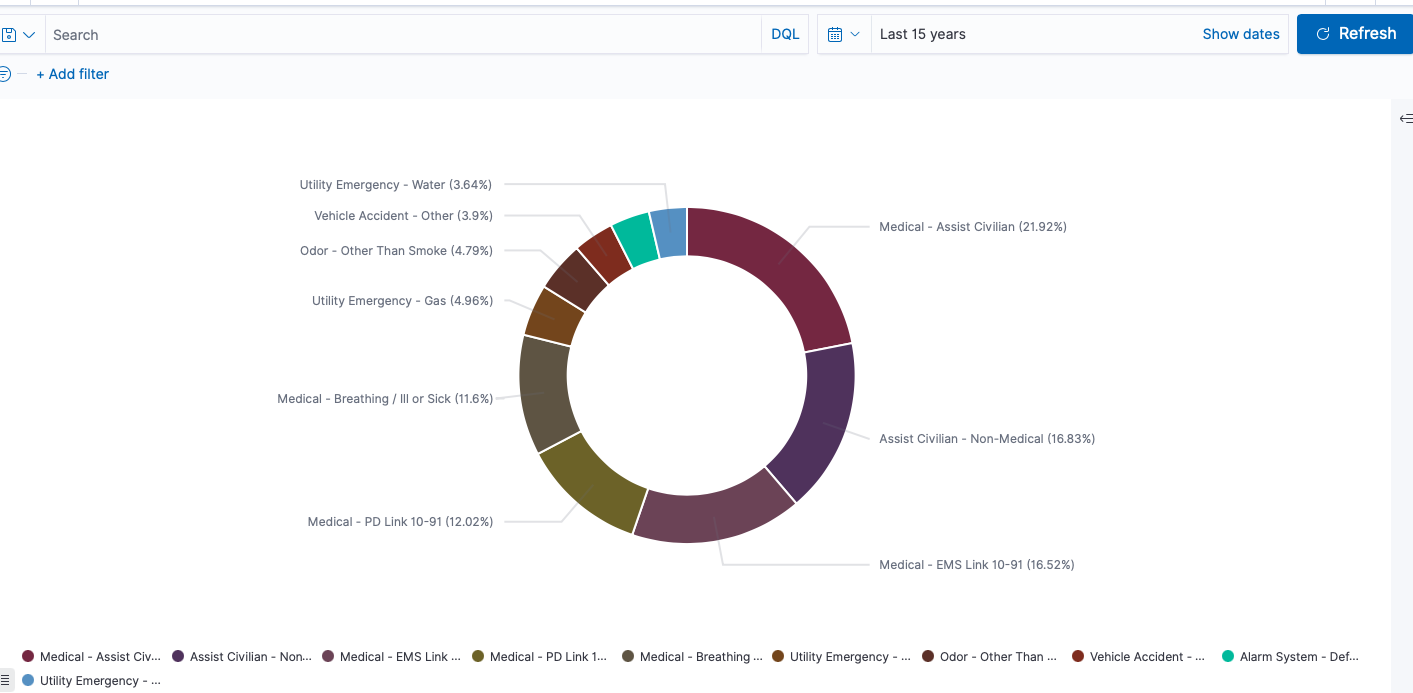
-Manhattan leads with the quickest average incident response time, typically responding to incidents in about 5 minutes. On the other hand, Staten Island lags behind with the slowest average incident response time, averaging around 7 minutes for response.



**Q) Which incident classification/category is most prevalent in the boroughs of NYC?**

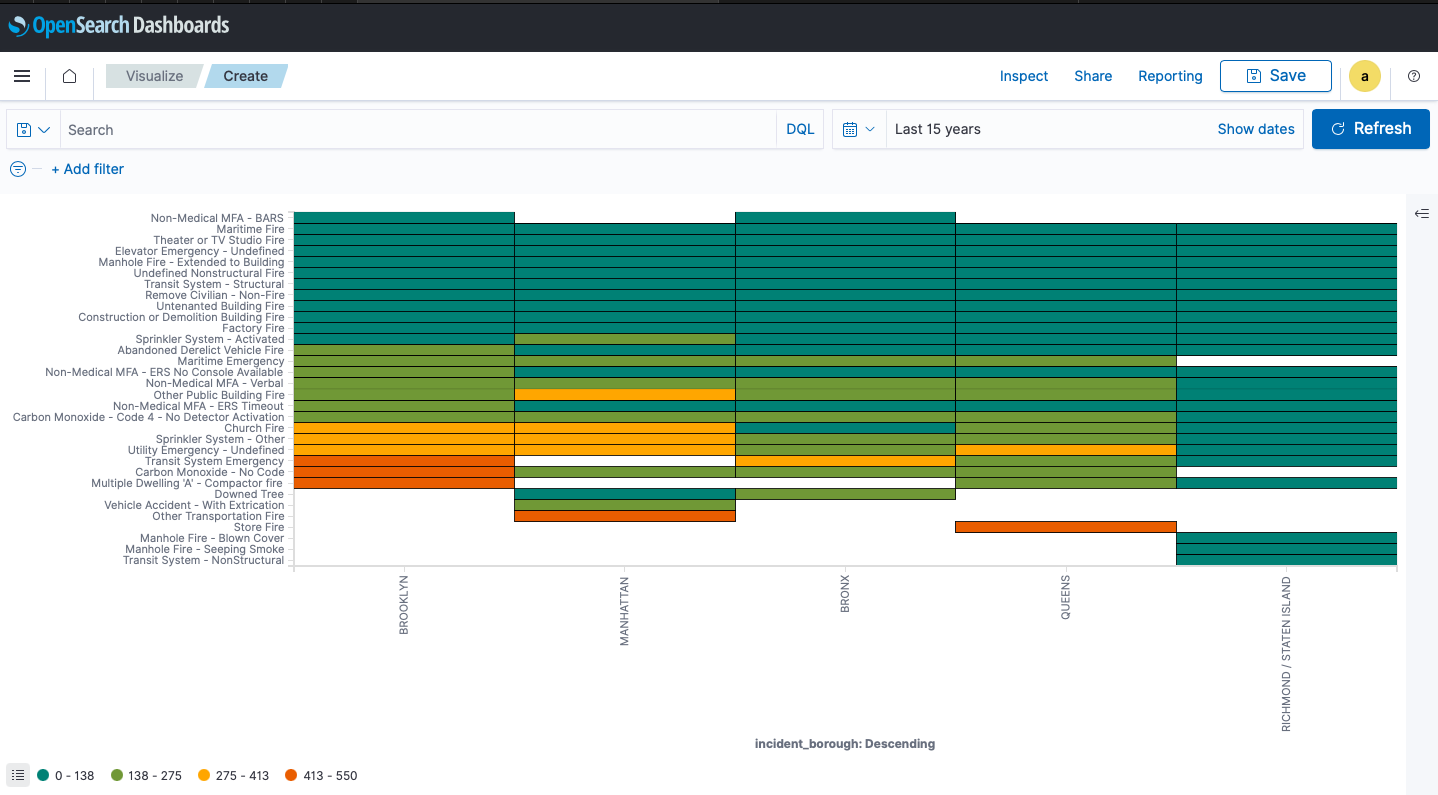
- Medical Incidents is the most prevalent category.

**- Incident Classification and weightage via PIE CHART:**

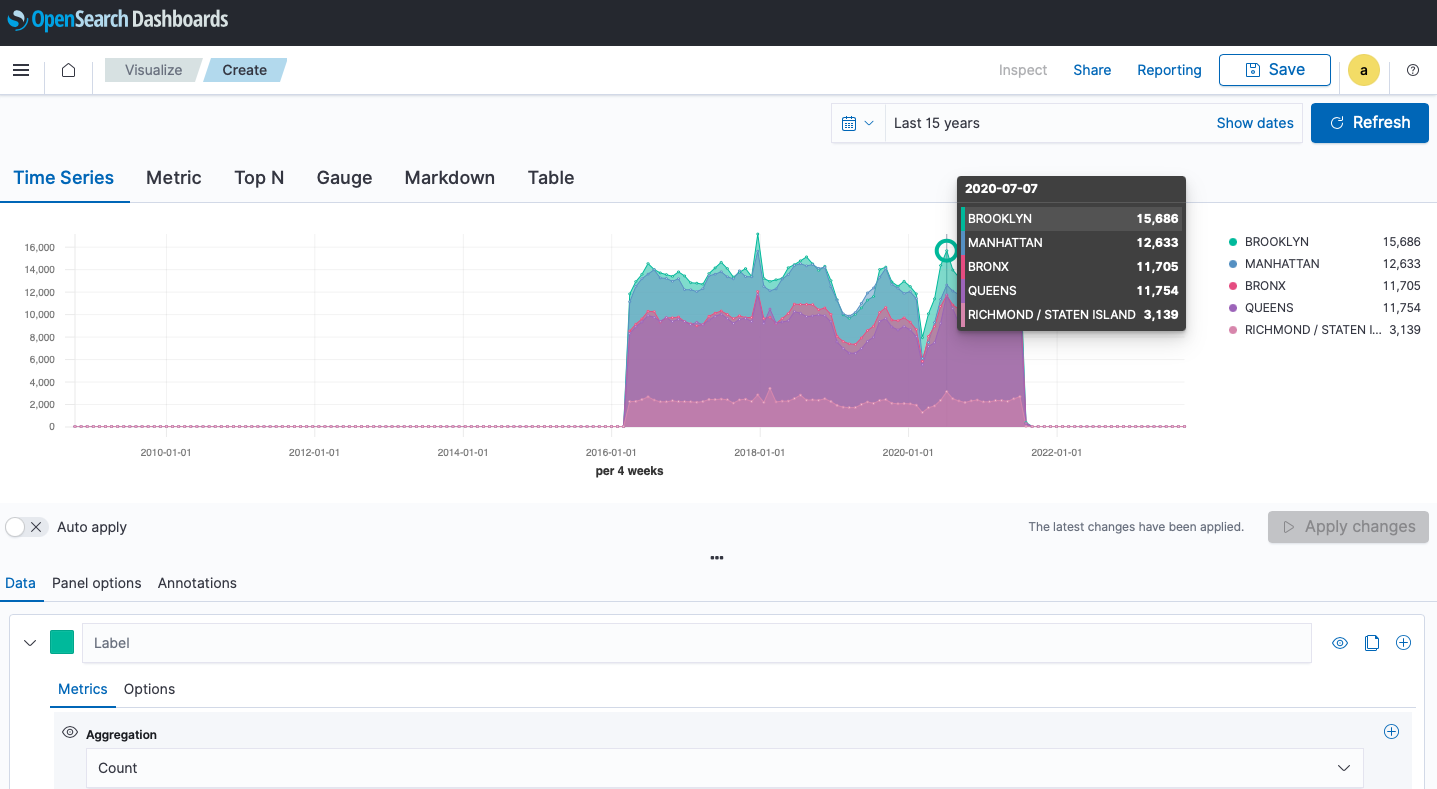


**Q) How is the incident response time distributed amongst the different boroughs of NYC?**

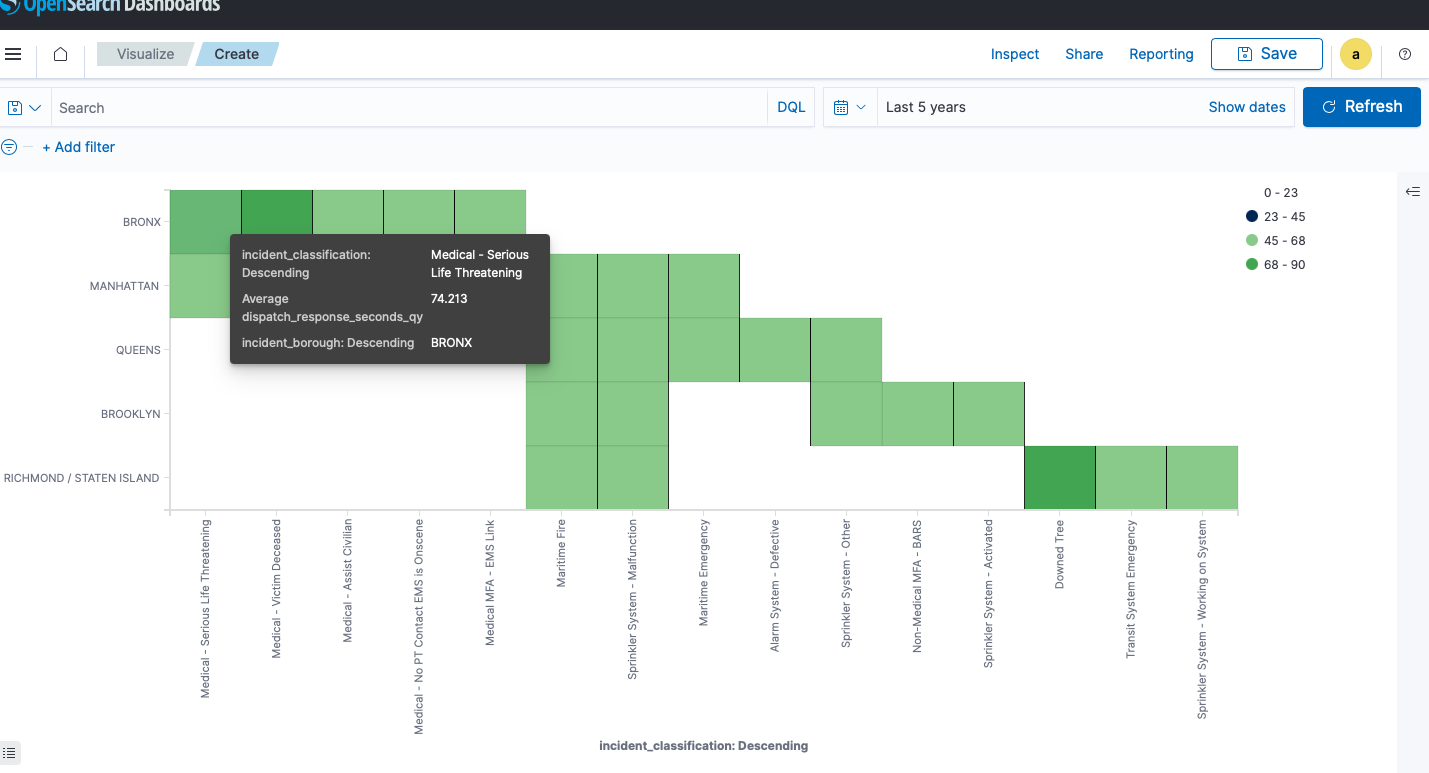
- Bronx has the longest incident response time rate in most of the categories while Staten Island and Brooklyn have lesser incident response times.



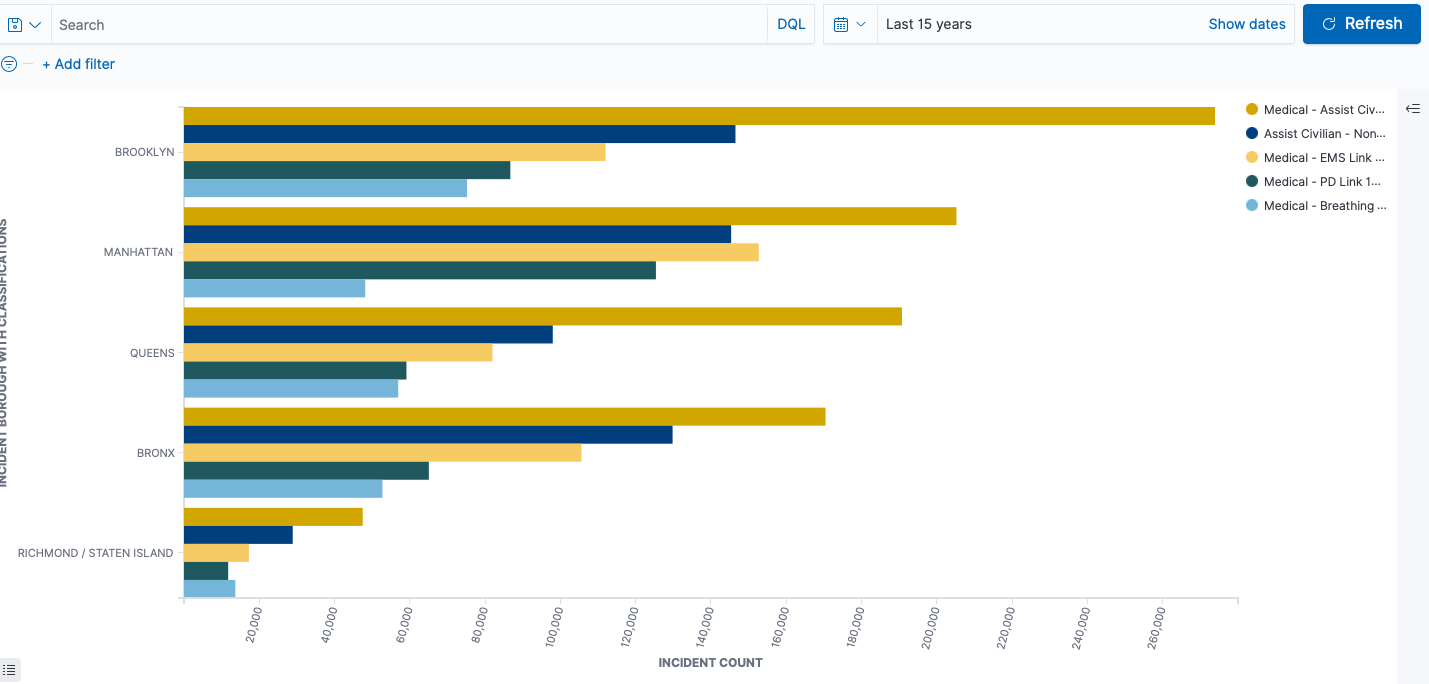
**-Incident counts in different borough over the time period via TIME SERIES PIPELINE:**



**-Classification of incidents with allocation in different broughs with average details:**



**-Classification based on borough with specified categories and their divisions per borough:**



- **Fluctuating trend in the response times over the period:**

