CIS 9760 Big Data Technologies – Project-1
Hrithik Shukla

Background & Procedure:

The analysis in this report is based on a vast dataset of millions of NYC fire incident records. The key variables analyzed as part of this project have been highlighted in Figure 0 below:

This dataset offers insights into the distribution of fire incidents across NYC's boroughs, temporal trends in response times, efficiency of emergency responses, and the prevalence of different incident types. It is a valuable resource for enhancing fire safety and urban planning in one of the world's largest and most densely populated cities.

Key Variable Name	Description
incident_datetime	Date and time of the incident occurrence
incident_response_seconds_qy	Response time for handling the incident in seconds
dispatch_response_seconds_qy	Response time for dispatching resources in seconds
incident_borough	Location of the incident, e.g., Manhattan
incident_classification	Type or Classification of the incident – e.g. Medical
starfire_incident_id	Identification number associated with the incident
incident_close_datetime	Date and time when the incident was officially closed

Figure 0

Approach Followed:

Amazon Elastic Compute Cloud (EC2) instances
Provisioning
Containerization
Python Scripting and File Creation included the development of 'main.py' for our application logic, a 'Dockerfile' for container configuration, and 'requirements.txt' for defining project dependencies.
OpenSearch Dashboard Visualizations:

Key Steps:

1. ES Domain Setup: We initiated the Elasticsearch (ES) domain, utilizing the master username and password to access the domain endpoint and OpenSearch dashboard URL.

2. EC2 Connection: We connected to an EC2 instance through a web browser, enabling us to create the necessary folder structure and project files.

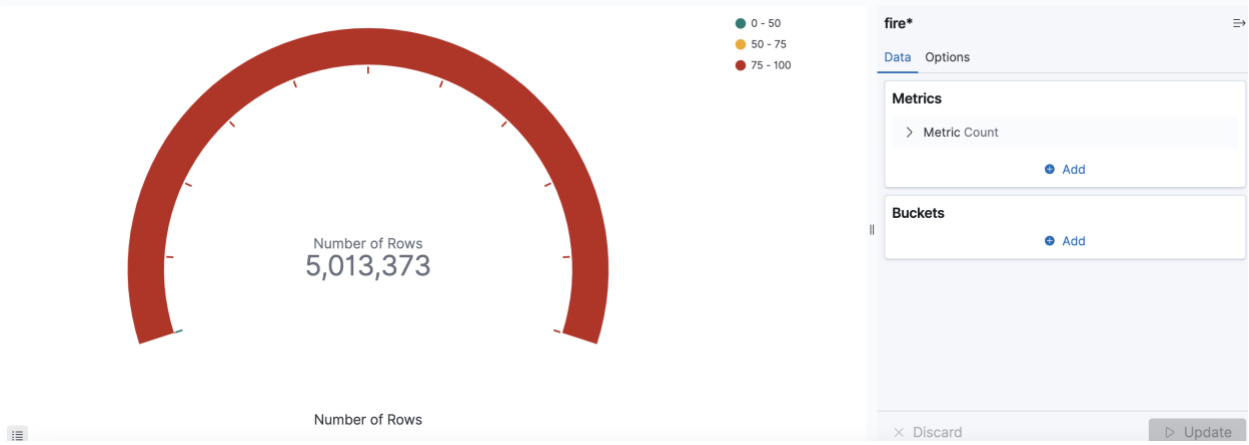
3. Data Testing: To ensure data retrieval, we conducted initial tests, initially fetching a limited number of rows while using the page_size argument.

Commands used for running the Docker file:
docker run -e INDEX_NAME="fire" -e DATASET_ID="8m42-w767" -e APP_TOKEN="1uexBLYJFL6r8FGlOVGQvY8p8" -e ES_HOST="https://search-hrithik-domain-final-fxzmnpzcoin6pq5vvcfjh74mfu.us-east-2.es.amazonaws.com" -e ES_USERNAME="hrithikshukla" -e ES_PASSWORD="Alpha@13711" bigdataproject1:1.0 --page_size=2000 --num_pages=2500
docker run

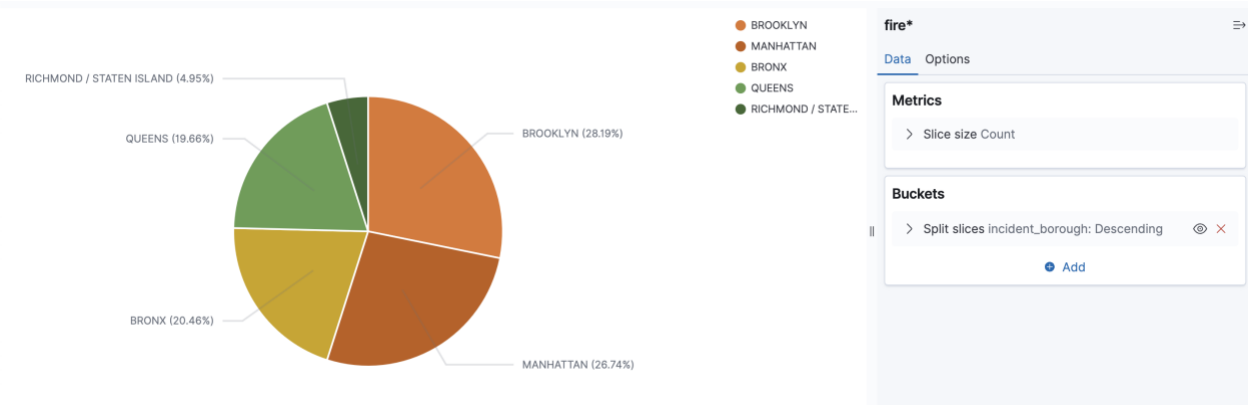
INDEX_NAME="fire"
DATASET_ID="8m42-w767"
APP_TOKEN= "1uexBLYJFL6r8FGlOVGQvY8p8"
ES_HOST= " https://search-hrithik-domain-final-fxzmnpzcoin6pq5vvcfjh74mfu.us-east-2.es.amazonaws.com "
ES_USERNAME="hrithikshukla"
ES_PASSWORD=" Alpha@13711 "

Analysis and Visualization:

Final gauge chart for >5 Million rows:



Question 1: What's the distribution of fire incidents among the five boroughs of New York City?
Figure 1: Pie Chart of Borough Fire Incidents:



The data in Figure 1 shows the percentage distribution of fire incidents across the five boroughs of New York City. Brooklyn leads with 28.2% of incidents, followed by Manhattan at 26.8%. Queens and the Bronx have substantial percentages as well, with 19.6% and 20.5% respectively, while Staten Island has the lowest percentage at 4.9%. This distribution likely reflects the population density and urban development in these boroughs. Brooklyn and Manhattan are the most densely populated areas, which could explain their higher incident rates. Staten Island, being less densely populated, has fewer incidents.

Question 2: How have the average incident response times and dispatch response times evolved over the years?

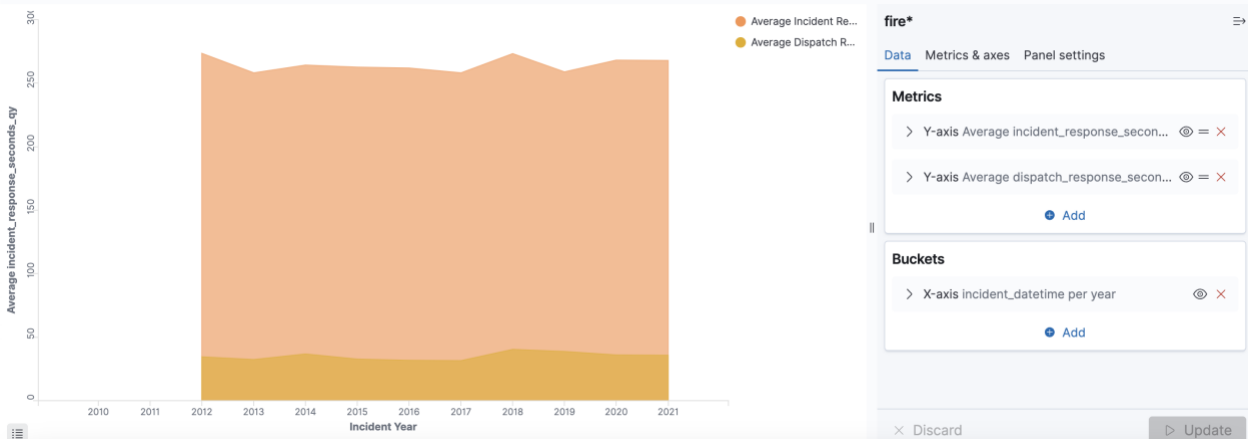


Figure 2: Stacked Area Chart

Figure 2 reveals that both the average incident response time and average dispatch response time peaked in 2018 and have remained relatively steady since then. This stability could be attributed to improved emergency response systems, enhanced training, and allocation of resources. The spike in 2018 might indicate a temporary increase in incidents or an adjustment period for the fire department to adapt to changing conditions.

Question 3: Which boroughs have the shortest and longest average incident and dispatch response times?

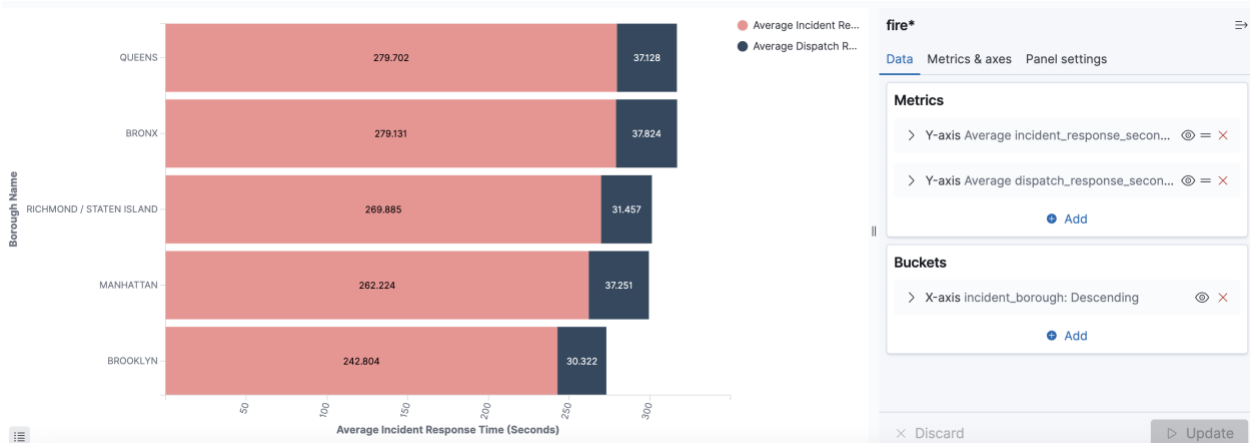


Figure 3: Stacked Horizontal Bar Chart.

The data in Figure 3 shows that Brooklyn has the shortest average incident response time at 242.4 seconds, while Queens has the longest at 278.7 seconds. In terms of dispatch response time, Manhattan boasts the shortest time at 26.8 seconds, whereas the Bronx has the longest at 37.9 seconds.

Brooklyn's shorter incident response time may result from efficient dispatching and proximity to fire stations. Manhattan's quick dispatch response time could be attributed to its well-connected road networks and proximity to emergency services. On the other hand, Queens and the Bronx, with longer times, may experience more complex traffic conditions or challenges related to geography and infrastructure.

Question 4: What are the top 10 major incident types, and which borough has the highest frequency of Medical Assists?

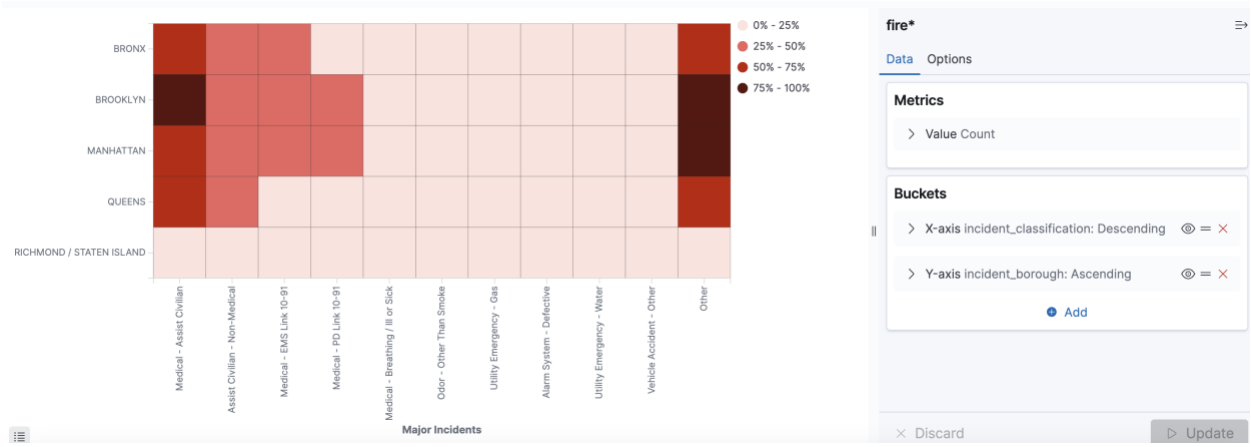


Figure 4: Incident Frequency Heatmap.

Figure 4's heatmap shows that Brooklyn has the highest relative frequency of "Medical Assists for Civilians" and tops the list for most other incident categories. This could be due to Brooklyn's

diverse population and a higher number of healthcare facilities. Moreover, it could reflect the need for more medical assistance in certain neighborhoods within Brooklyn.
The variation among boroughs likely stems from differences in demographics, socioeconomic factors, and urban planning, impacting the types and frequencies of incidents each borough experiences. Brooklyn's leadership in several categories may be due to its size and complexity as a borough within New York City.