12/20/2021

Assignment 2 – Data Visualization with Kibana

BDAT 1002 - Data System Architecture

Group 10 Members	Student ID
Amith Heiden	**200497933**
Karthikeyan Suresh Kumar	**200489370**
Srilekha Sampath Kumar	**200499290**

Submitted to : Dr. Saber Amini

TABLE OF CONTENTS

INTRODUCTION

ELK Stack

The ELK Stack is the world's most popular log management platform. The ELK Stack was a collection of three open-source products — Elasticsearch, Logstash, and Kibana — all developed, managed and maintained by Elastic.

- Elasticsearch is an open source, full-text search, and analysis engine, based on the Apache Lucene search engine.
- Logstash is a log aggregator that collects data from various input sources, executes different transformations and enhancements and then ships the data to various supported output destinations.
- Kibana is a visualization layer that works on top of Elasticsearch, providing users with the ability to analyze and visualize the data.

Background of the Dataset

This dataset is taken from NYC OpenData website and it is associated with remote call taking necessitated by the unprecedented volume 311, handled during the Covid-19 crisis. The main focus of the dataset is all about Service Requests (SR) – When, where, Complaint type, Description, Status of the SR.

Objective

We have been hired as data analysts by the city of New York to provide valuable insights of their huge data set for 311 service requests. Our task is to work in the ELK stack by installing and configured it in GCP platform.

Deliverables

- Code for Logstash configuration file
- Geo-point template (for maps) – Screen shot
- Results for the analytical questions (tables, charts, tag clouds, maps, and dashboard) - Screen shots

ASSIGNMENT APPROACH AND METHODOLOGY

My Team

Students of Big Data Analytics, Georgian College, Fall 2021 intake

Amith Heiden - Data Analyst

Tasks Performed – Creation of Logstash configuration file, geo-point template, Analytical Questions : 3, 4,10

Karthikeyan suresh kumar – Data Analyst

Tasks Performed – Creation of Logstash configuration file, geo-point template, Analytical Questions : 1, 2, 6

Srilekha Sampath kumar - Data Analyst

Tasks Performed - Creation of Logstash configuration file, geo-point template,

Analytical Question : 5, 7, 8, 9

Techniques Used

Prepared cloud environment and the below steps are done after setting up ELK on a GCP cluster:

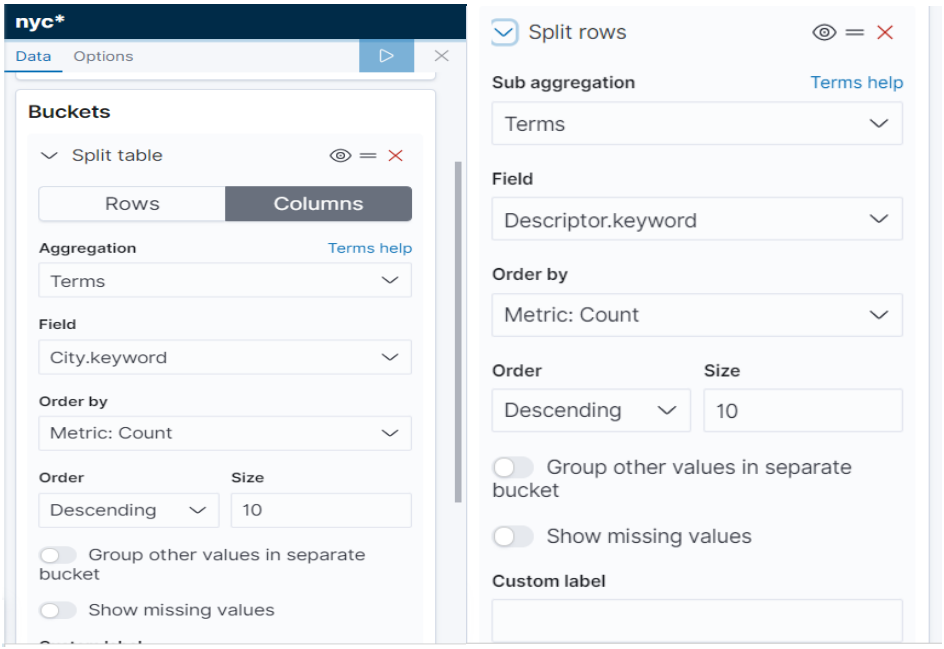
1. Created an index called ‘nycinfo’ with the sample Logstash configuration file (logstash_nyc311.config)
2. Created an index pattern called ‘nycinfo’, which matches the ‘nycinfo’ index created at step 1. Specify the Time Filter option as “I don’t want to use the time filter.” This index is used for the analysis of questions 1&2.
3. Created a template called ‘geotemplate’ to perform analysis of question 4.

ANSWERS TO ANALYTICAL QUESTIONS

Data Visualizations are performed using Kibana

1. Create a table showing the top 10 cities with the highest calls alongside the count of top 10 complaint calls (by Descriptor) in each city.

Created the visualizations -data table- in Kibana.



BROOKLYN: City.keyword: Descending

Descriptor.keyword: Descending	Count
Loud Music/Party	39,961
ENTIRE BUILDING	28,892
Request Large Bulky Item Collection	24,155
No Access	23,978
APARTMENT ONLY	19,329
Banging/Pounding	12,882
Blocked Hydrant	11,030
Posted Parking Sign Violation	8,910
Partial Access	7,900
Loud Talking	6,979

Export: Raw Formatted

NEW YORK: City.keyword: Descending

Descriptor.keyword: Descending	Count
Loud Music/Party	38,279
ENTIRE BUILDING	22,222
Banging/Pounding	12,012
APARTMENT ONLY	11,263
N/A	9,467
Loud Talking	8,001
Driver Complaint	7,281
Noise: Construction Before/After Hours (NM1)	7,038
Request Large Bulky Item Collection	5,289
Posted Parking Sign Violation	3,920

Export: Raw Formatted

BRONX: City.keyword: Descending

Descriptor.keyword: Descending	Count
ENTIRE BUILDING	32,613
Loud Music/Party	28,758
APARTMENT ONLY	15,907
No Access	11,700
Banging/Pounding	11,565
PESTS	6,083
WALL	4,848
CEILING	4,496
MOLD	4,128
FLOOR	4,002

Export: Raw Formatted

STATEN ISLAND: City.keyword: Descending

Descriptor.keyword: Descending	Count
Recycling Electronics	5,062
Request Large Bulky Item Collection	4,460
Pothole	3,251
Loud Music/Party	3,115
Street Light Out	2,468
With License Plate	2,274
No Access	1,705
Posted Parking Sign Violation	1,462
Blocked Hydrant	1,128
Banging/Pounding	1,105

Export: Raw Formatted

Jamaica: City.keyword: Descending

Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	1,319
ENTIRE BUILDING	1,089
APARTMENT ONLY	1,067
Pothole	792
14 Derelict Vehicles	709
Street Light Out	702
Illegal Conversion Of Residential Building/Space	540
Sewer Backup (Use Comments) (SA)	528
PESTS	445
15 Street Cond/Dump-Out/Drop-Off	438

Export: Raw Formatted

JAMAICA: City.keyword: Descending

Descriptor.keyword: Descending	Count
Loud Music/Party	2,398
No Access	2,321
With License Plate	1,095
Partial Access	943
Banging/Pounding	795
Posted Parking Sign Violation	433
Driver Complaint	374
Blocked Hydrant	365
Branch or Limb Has Fallen Down	338
Commercial Overnight Parking	298

Export: Raw Formatted

FLUSHING: City.keyword: Descending

Descriptor.keyword: Descending	Count
No Access	2,068
Loud Music/Party	1,111
Partial Access	1,084
Banging/Pounding	830
With License Plate	605
Blocked Hydrant	593
Sidewalk Violation	437
Posted Parking Sign Violation	377
Branch or Limb Has Fallen Down	364
Loud Talking	353

Export: Raw Formatted

ASTORIA: City.keyword: Descending

Descriptor.keyword: Descending	Count
Loud Music/Party	2,643
No Access	1,922
Partial Access	740
Banging/Pounding	707
Loud Talking	585
Posted Parking Sign Violation	574
Blocked Hydrant	471
Blocked Sidewalk	454
With License Plate	395
For One Address	301

Export: Raw Formatted

Flushing: City.keyword: Descending

Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	1,744
ENTIRE BUILDING	1,463
Pothole	826
APARTMENT ONLY	627
Street Light Out	447
Illegal Conversion Of Residential Building/Space	437
PESTS	245
Controller	191
GARBAGE/RECYCLING STORAGE	163
1 Missed Collection	161

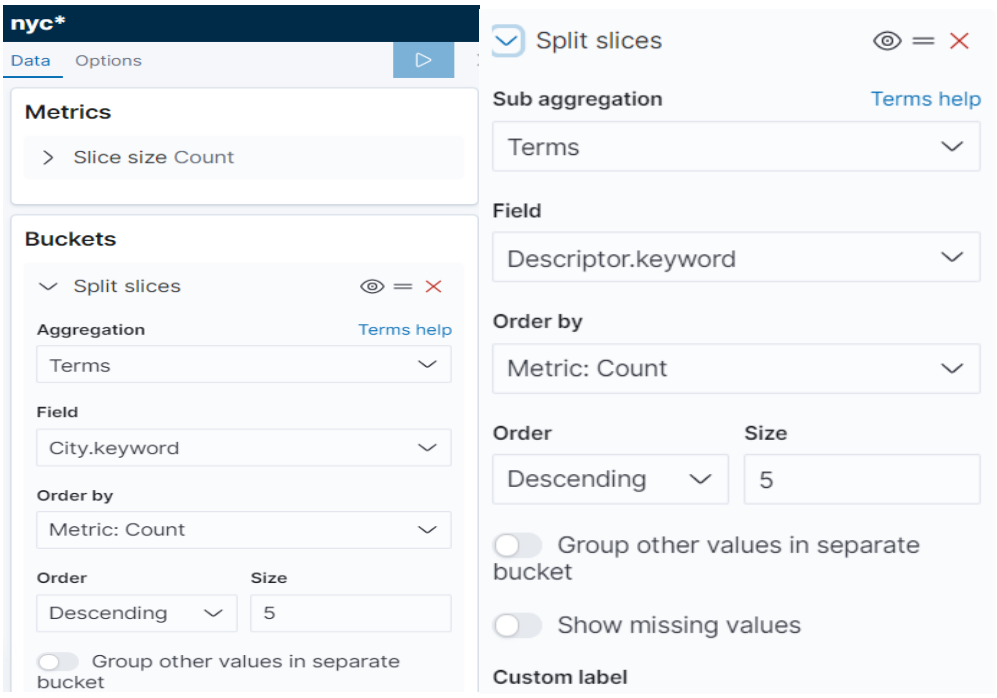
Export: Raw Formatted

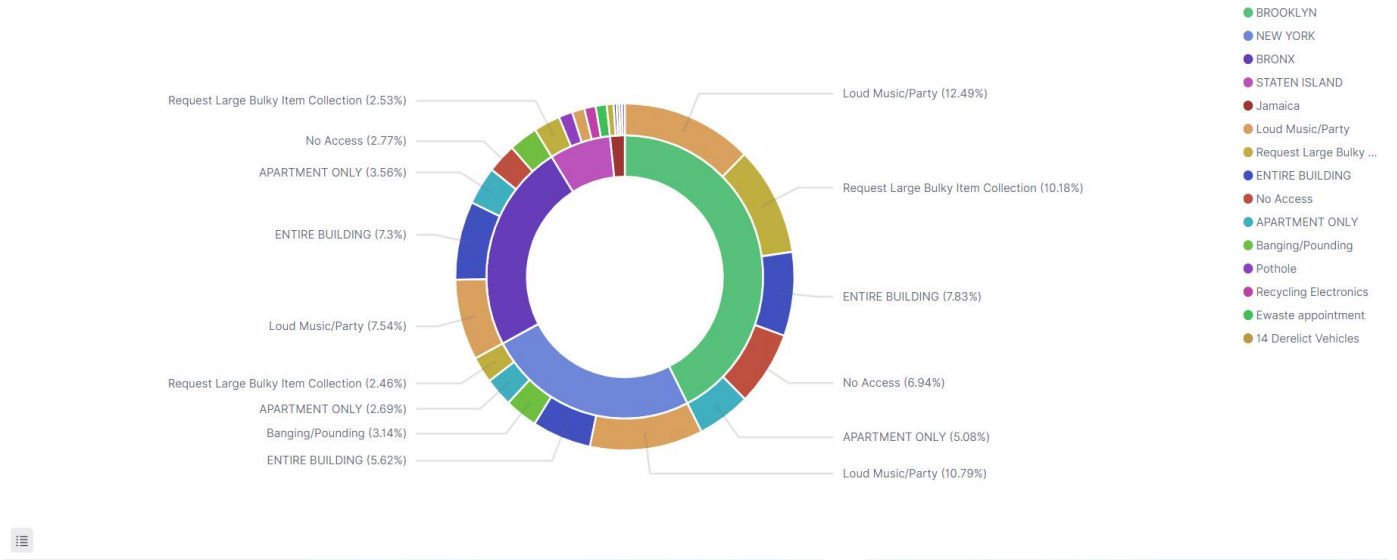
Ridgewood: City.keyword: Descending

Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	4,055
ENTIRE BUILDING	707
Pothole	363
APARTMENT ONLY	315
Street Light Out	293
E15 Illegal Postering	260
PESTS	189
15 Street Cond/Dump-Out/Drop-Off	167
2 Bulk-Missed Collection	166
Leak (Use Comments) (WA2)	162

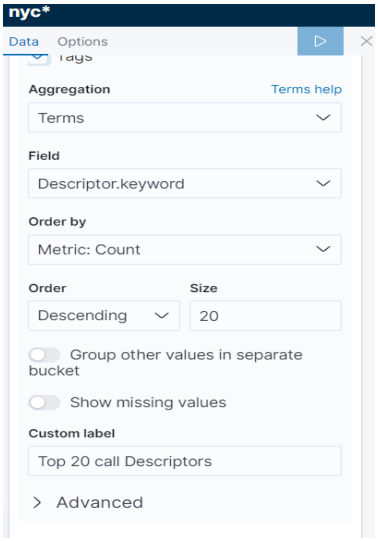
Export: Raw Formatted

2. Create a pie chart showing the top 5 cities with the highest calls alongside the top five calls (Descriptor) in each city





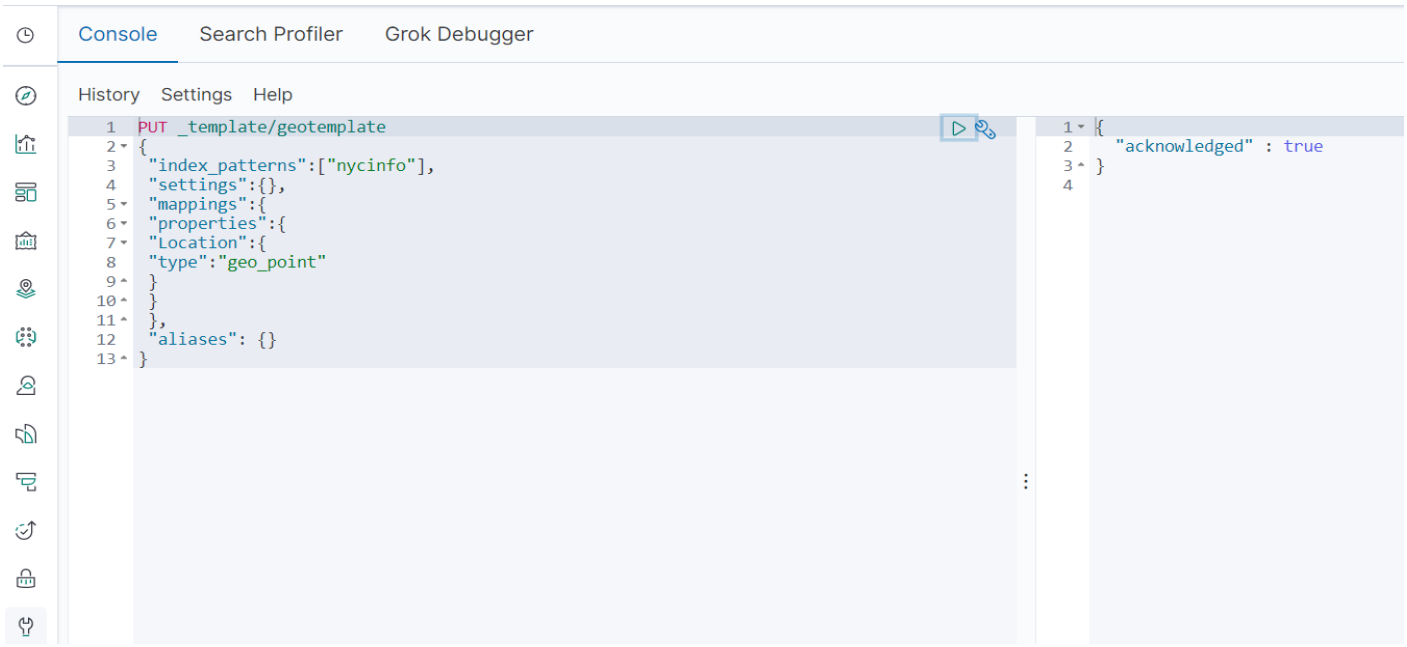
3. Create a tag cloud representing the top 20 call descriptors

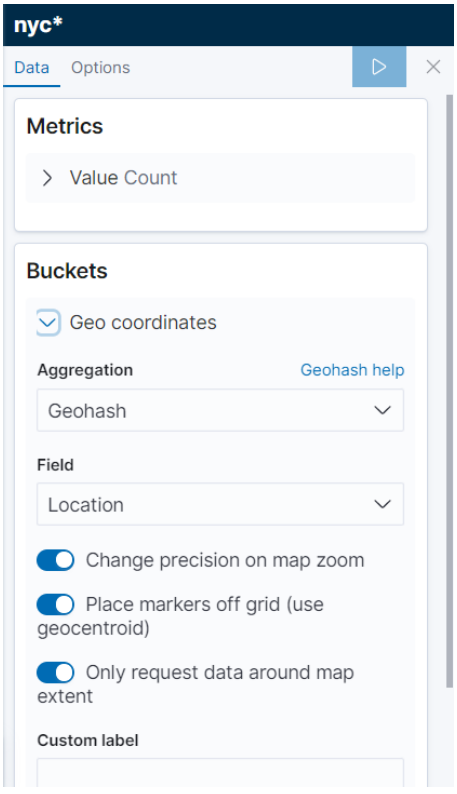


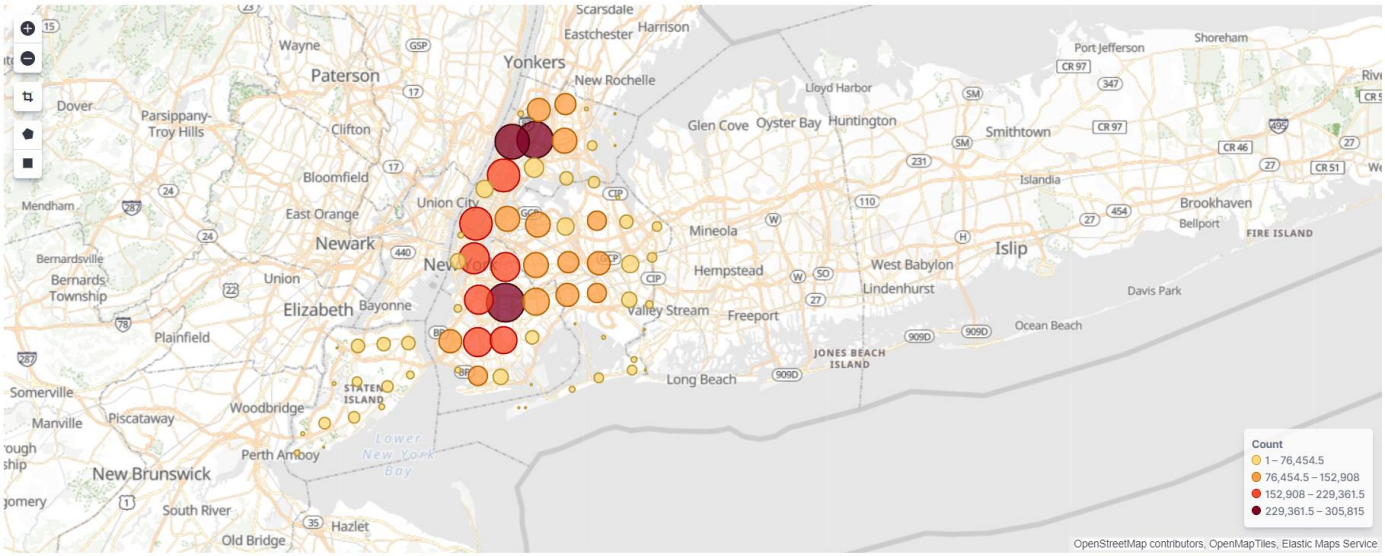
Top 20 call Descriptors - Count

4. Create a coordinated map of all the major call descriptors in each city

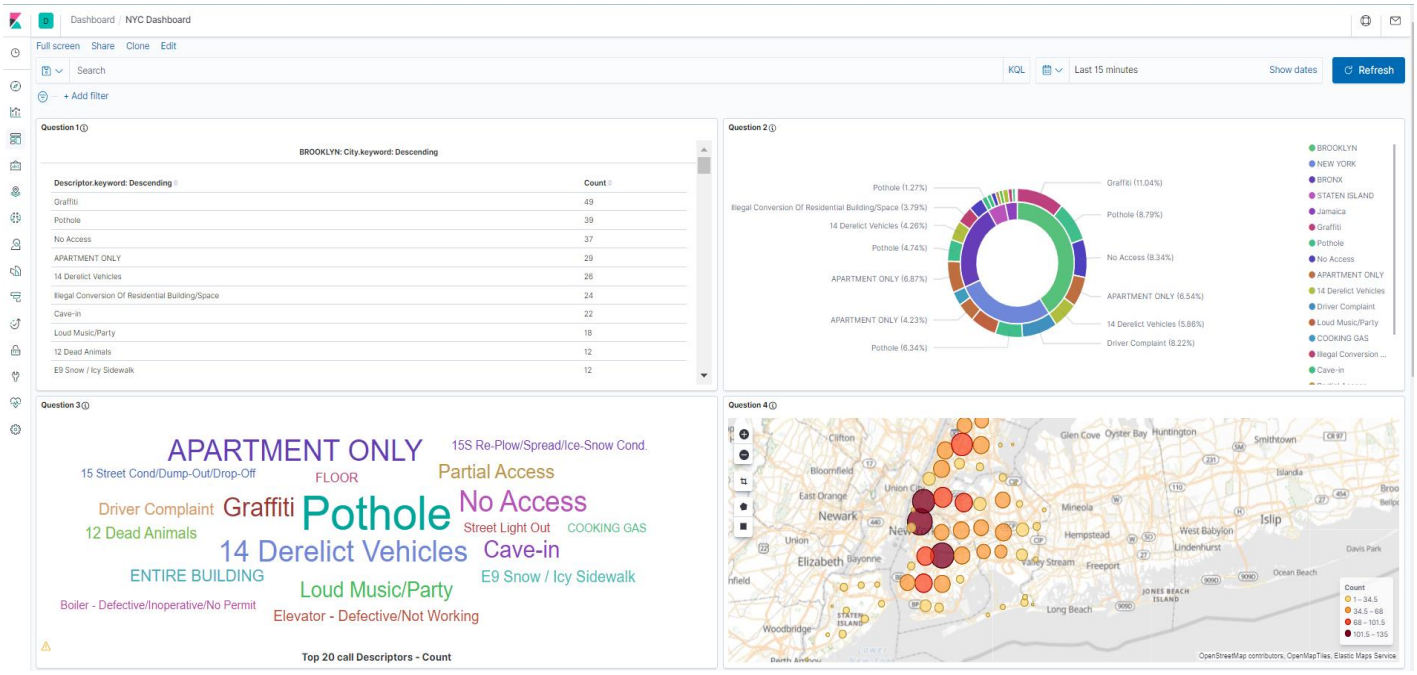
Created a template called ‘geotemplate’



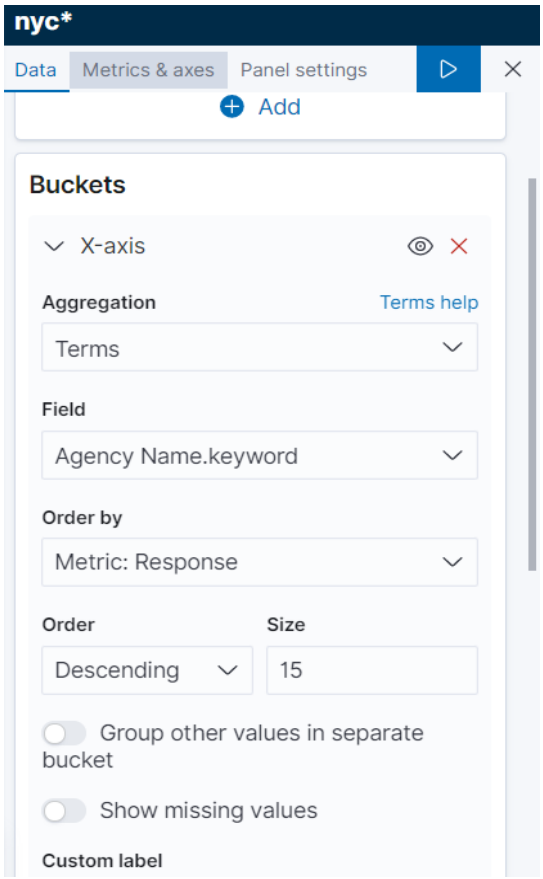


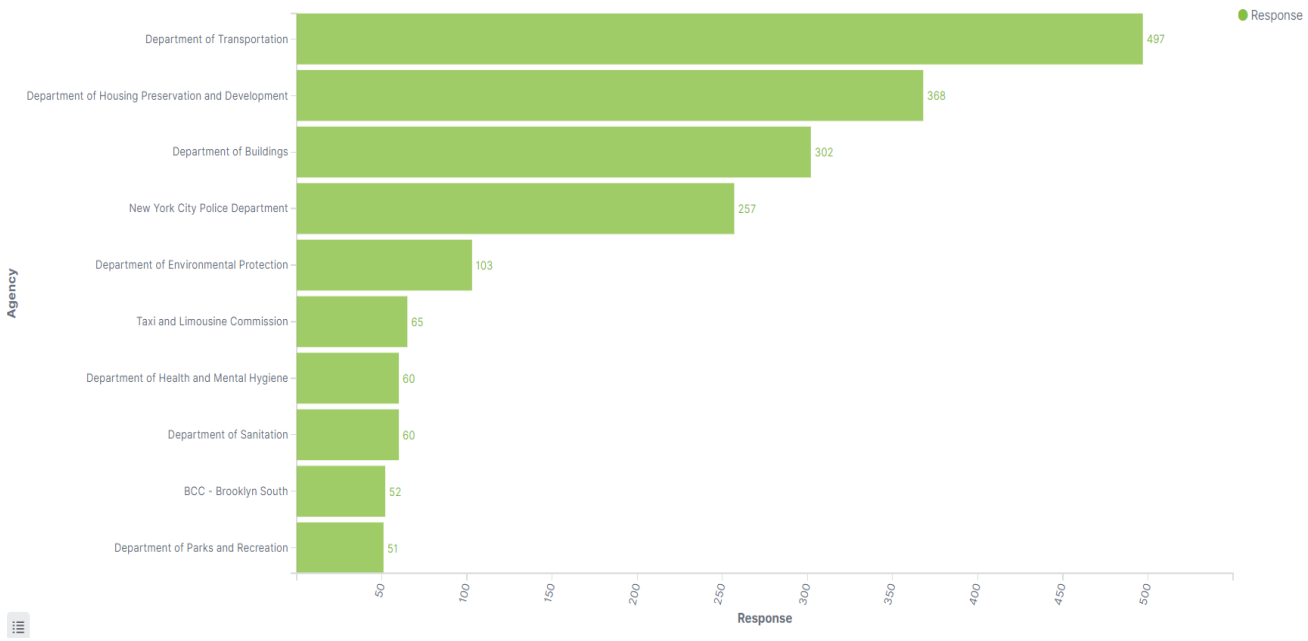


5. Create a dashboard for all visualizations of 1to 4 above

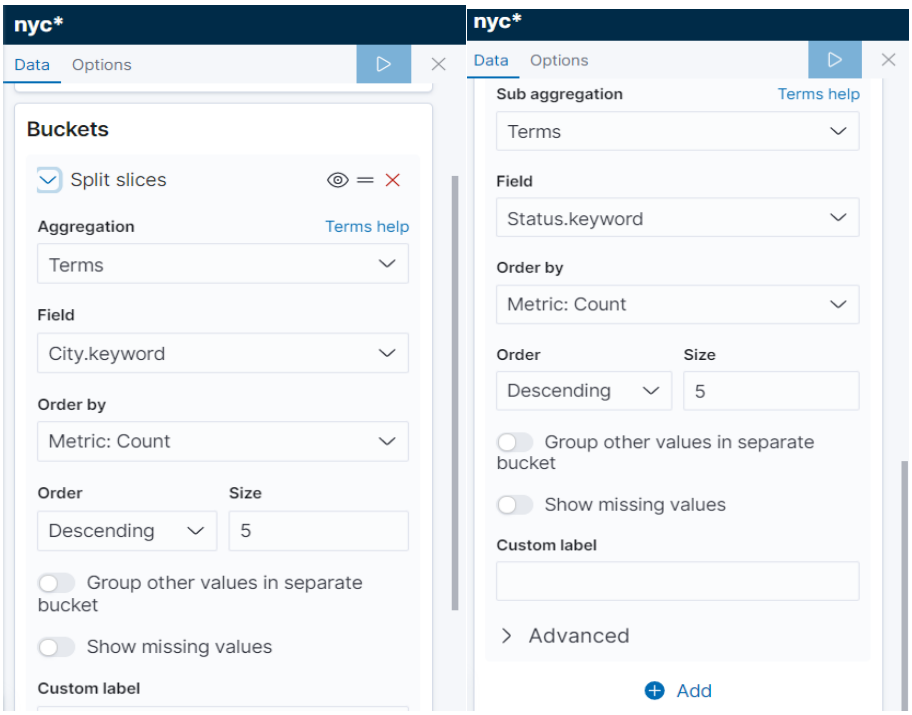


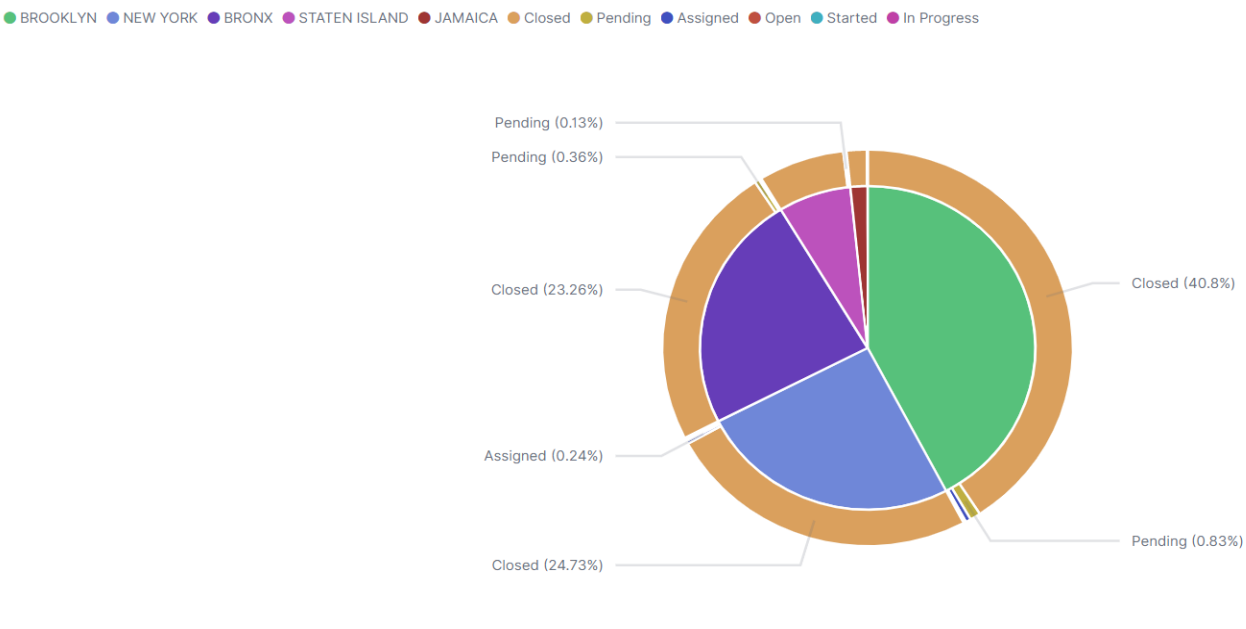
6. What are the Top 10 responding city government agency?



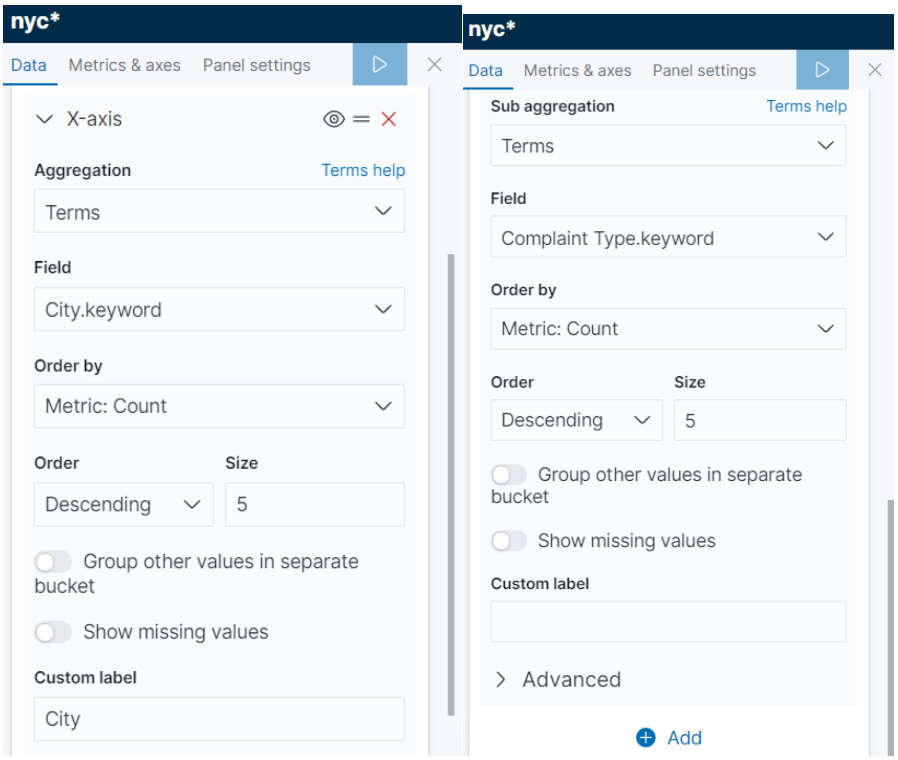


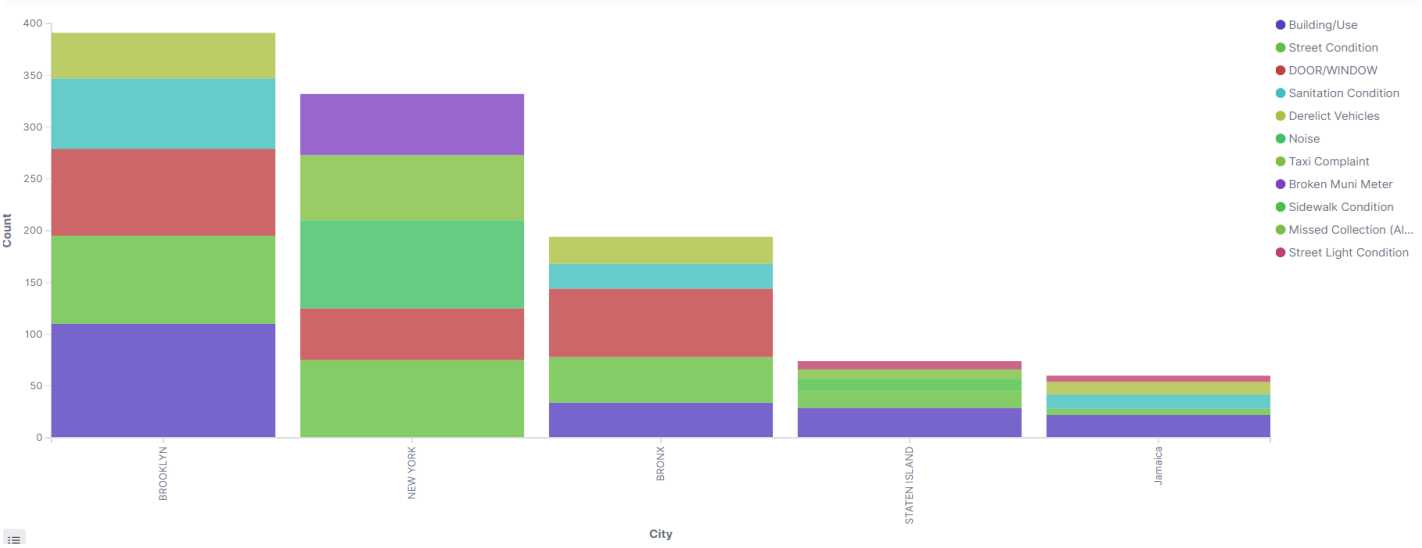
7. What is the Total count of each complaint status of Top 5 cities?



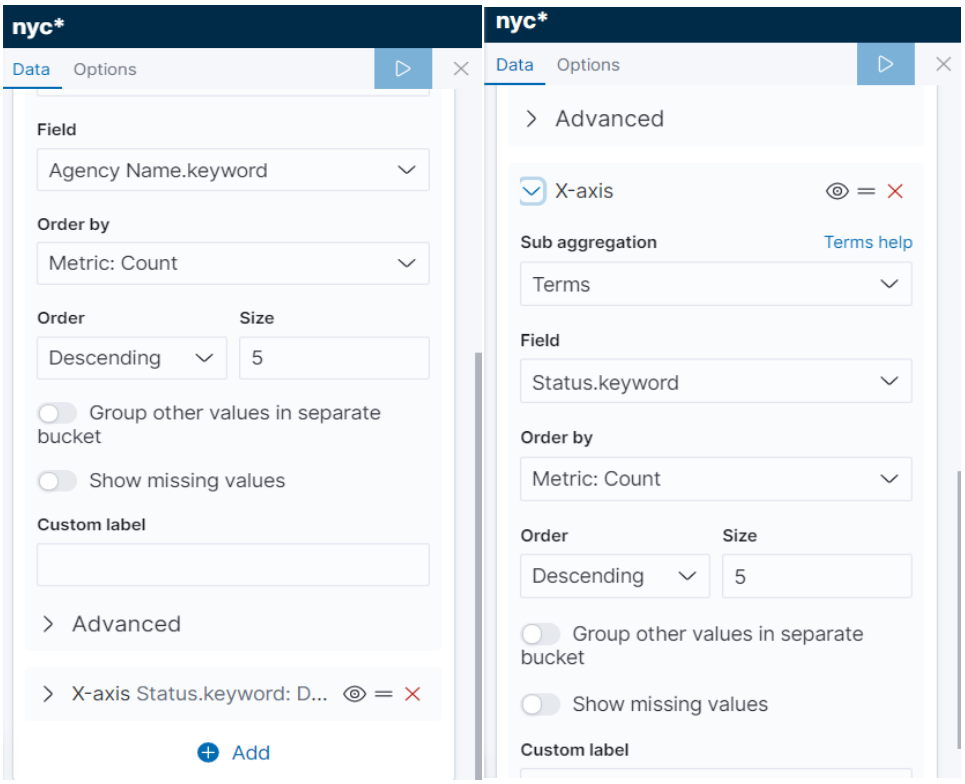


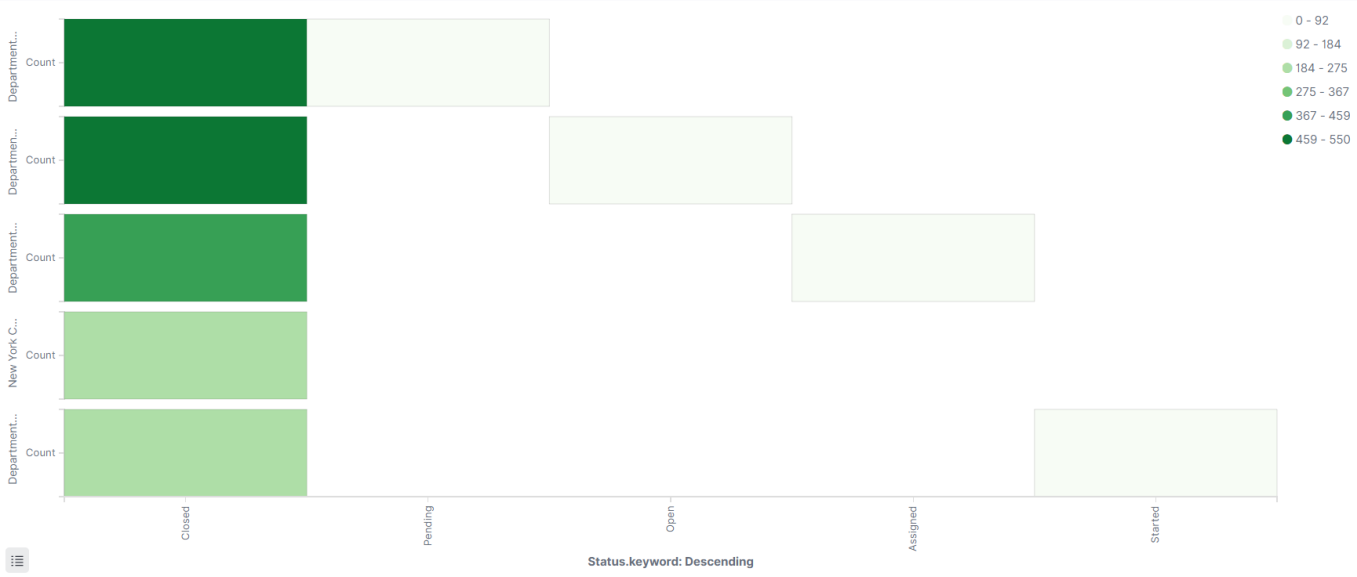
8. What is the total count of Top 5 Complaint types in Top 5 Cities?



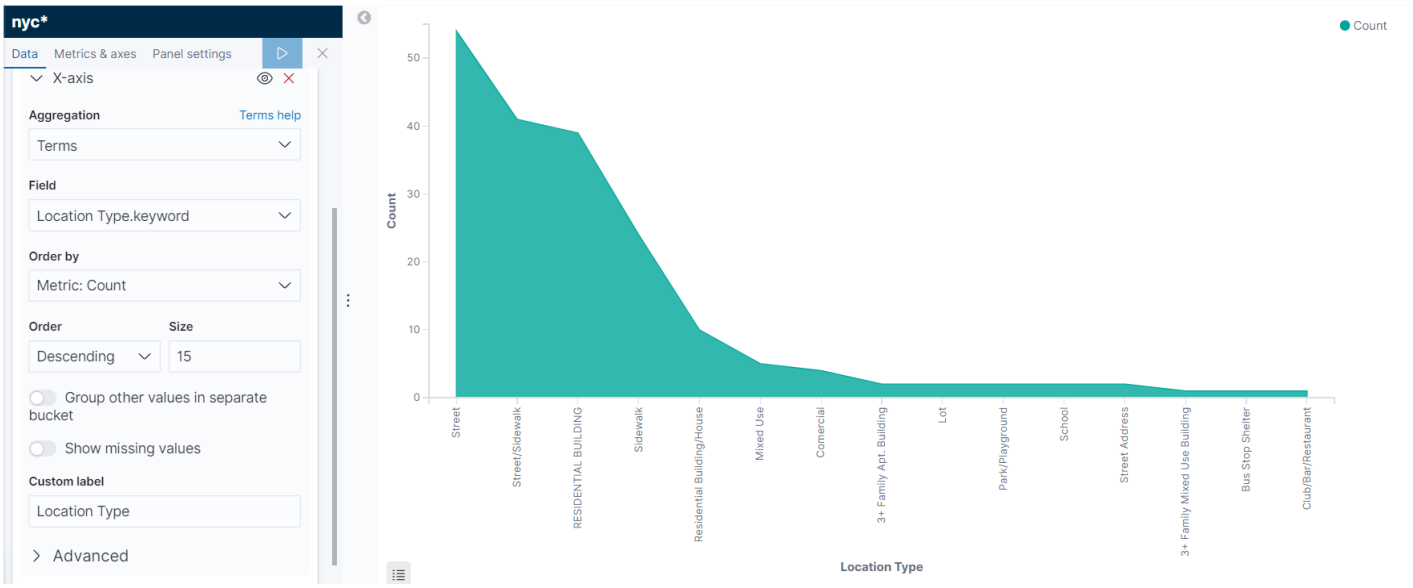


9. What are the Top 5 Service Request Status of Top 5 Government Agencies?

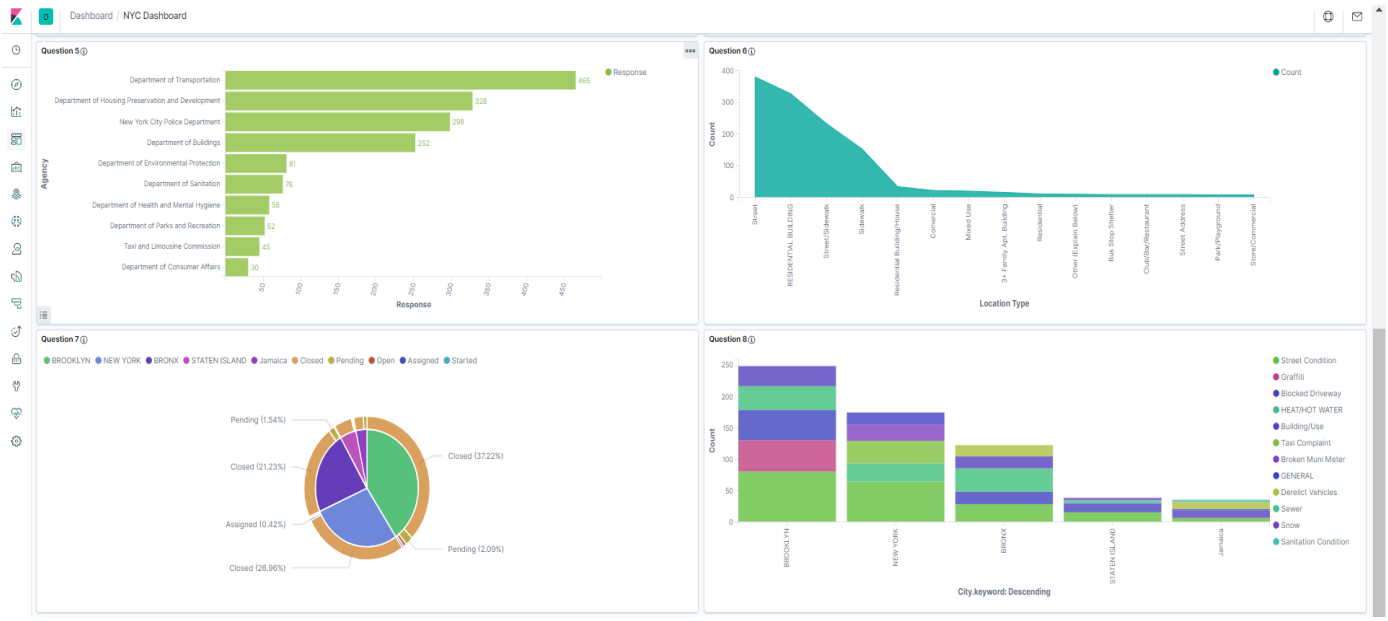


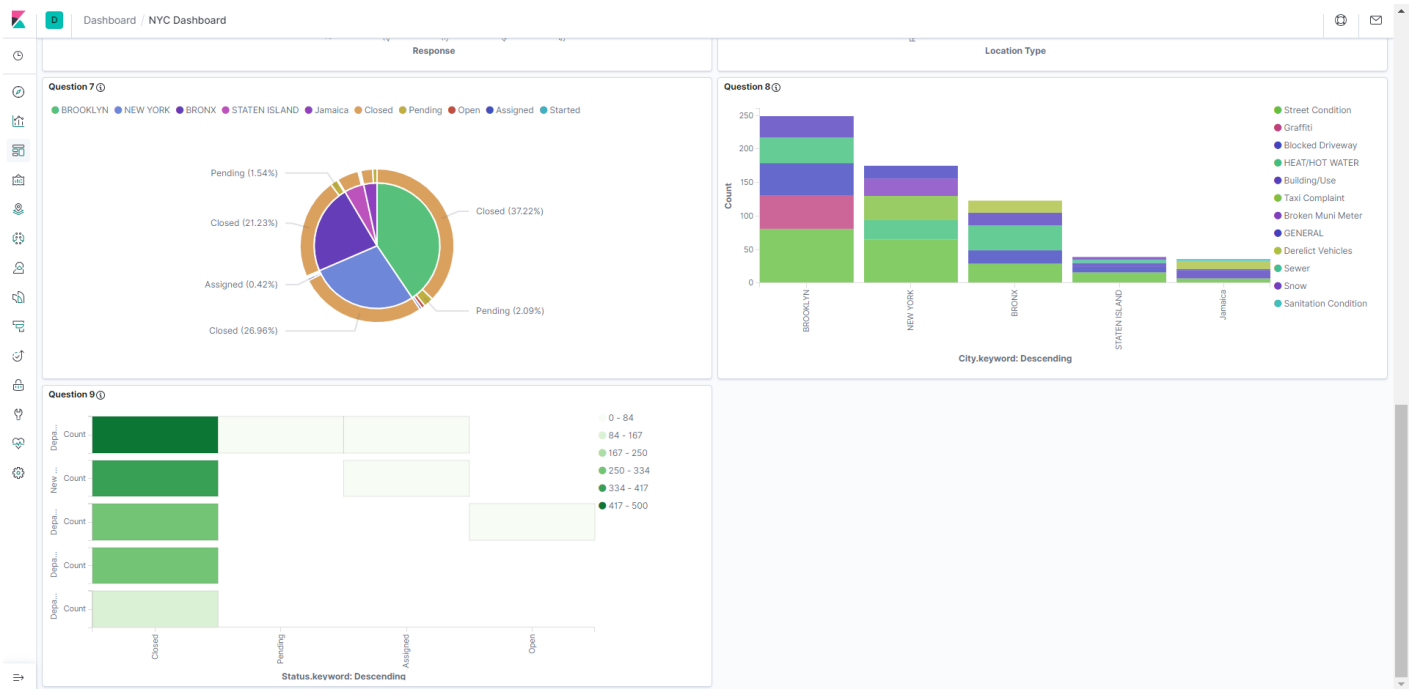


10. What is the top 15 location type based on frequency of SR?



Dashboard for Questions 6 – 10





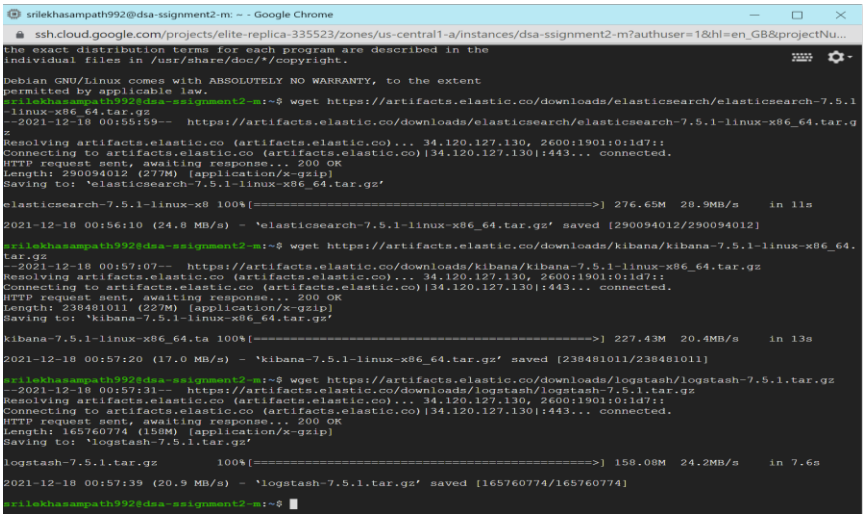
References

College, C. (2021, 12 13). blackboard . Retrieved from blackboard:

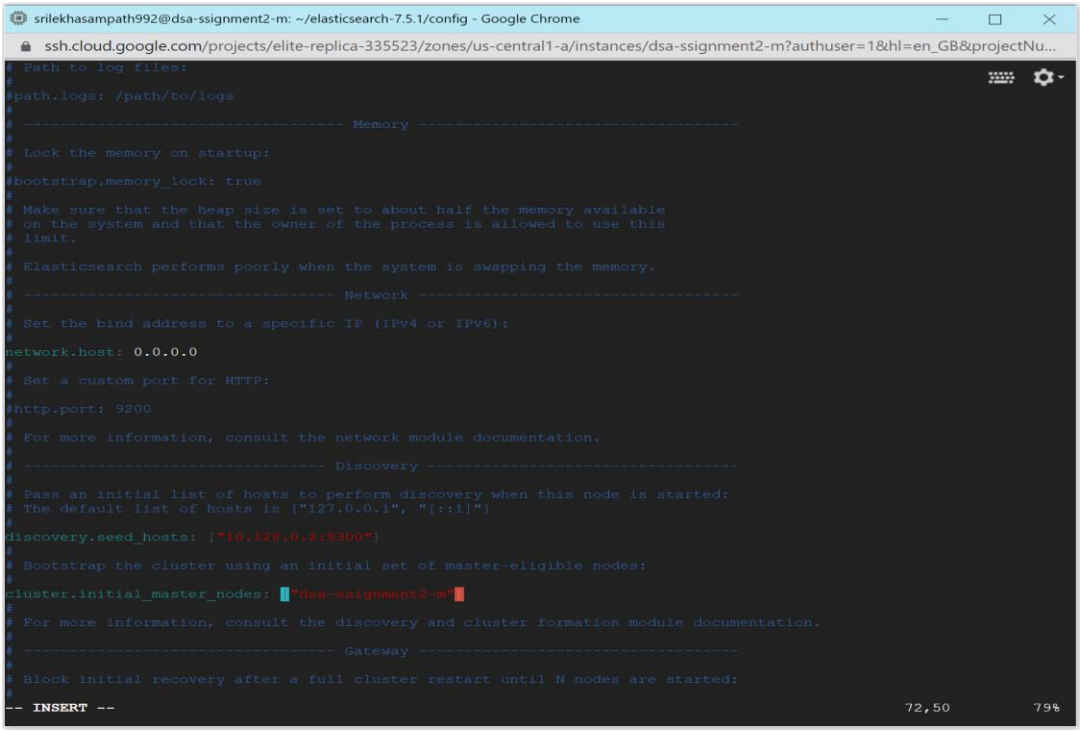
https://gc.blackboard.com/webapps/blackboard/content/listContent.jsp?course_id=_317687_1&content_id=_6464709_1

Appendix

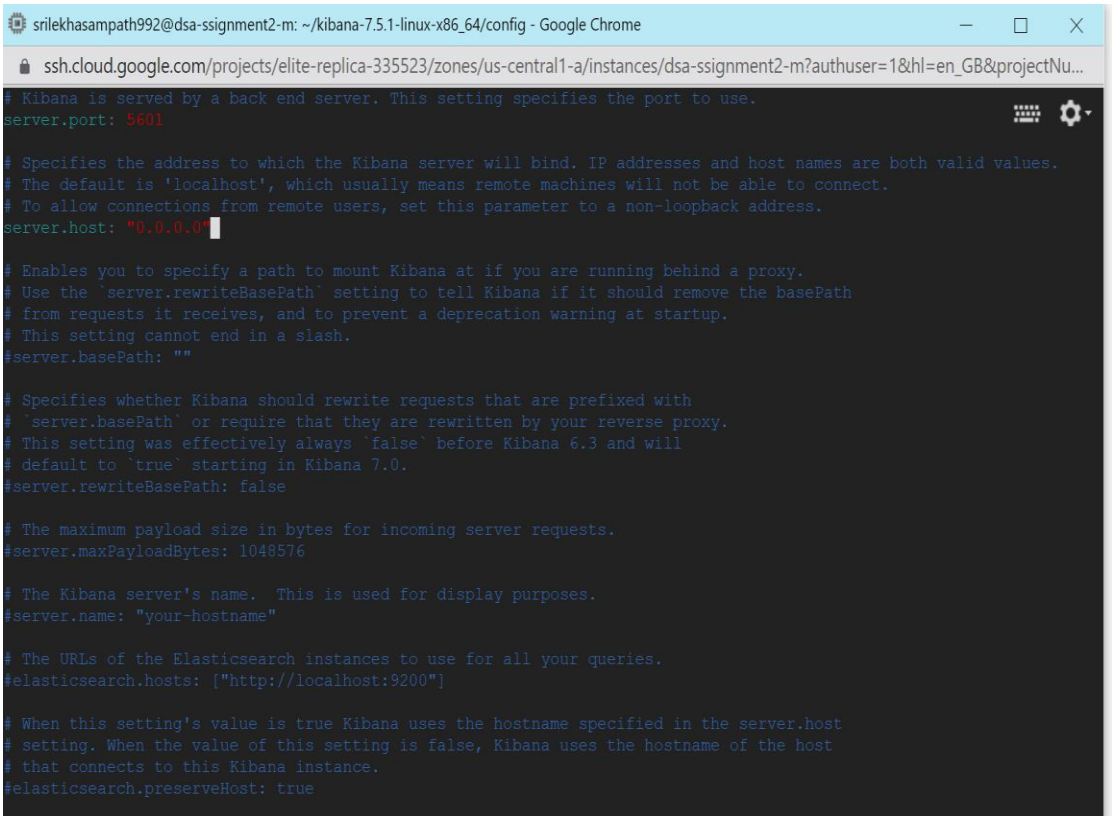
Setting up ELK on a GCP cluster



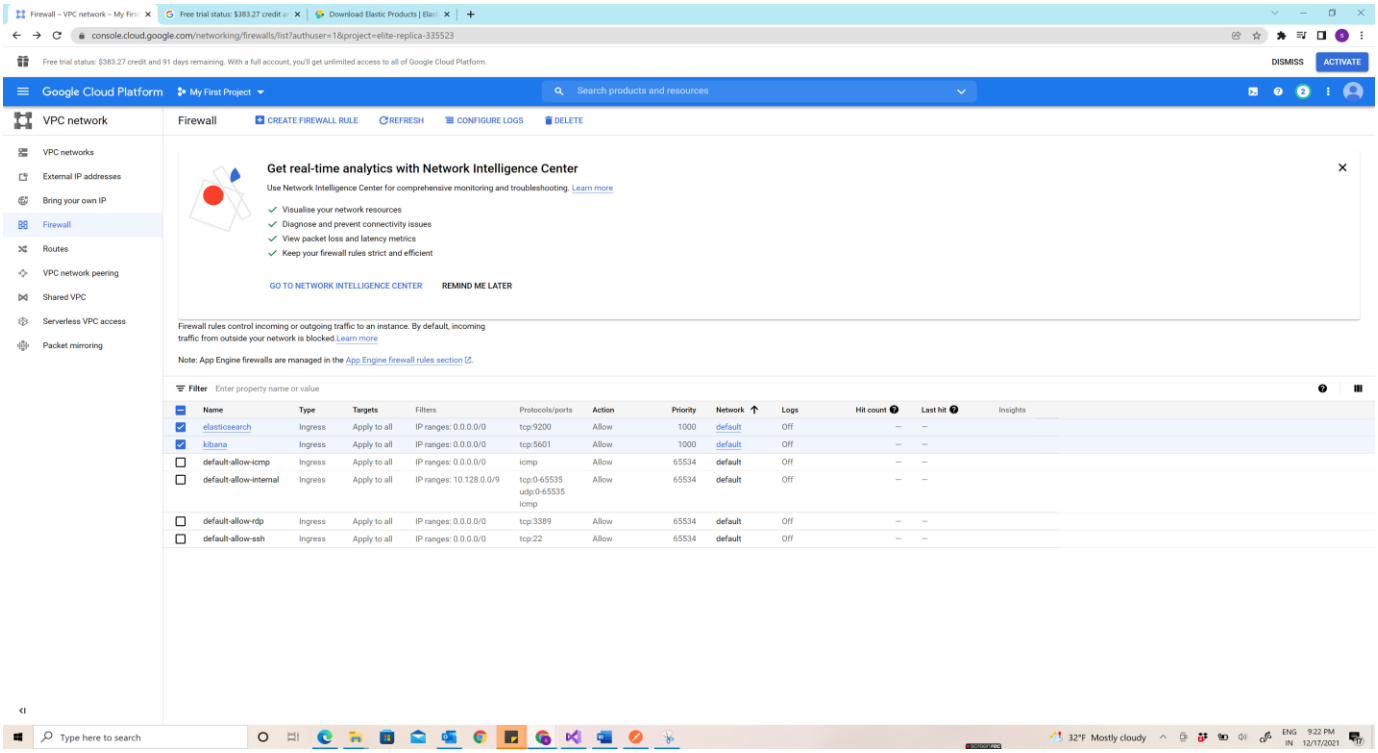
Modifying relevant information in elasticsearch configuration file using vi text editor,



Modifying relevant information in Kibana configuration file using vi text editor,



Created new Firewall rules for elasticsearch and Kibana



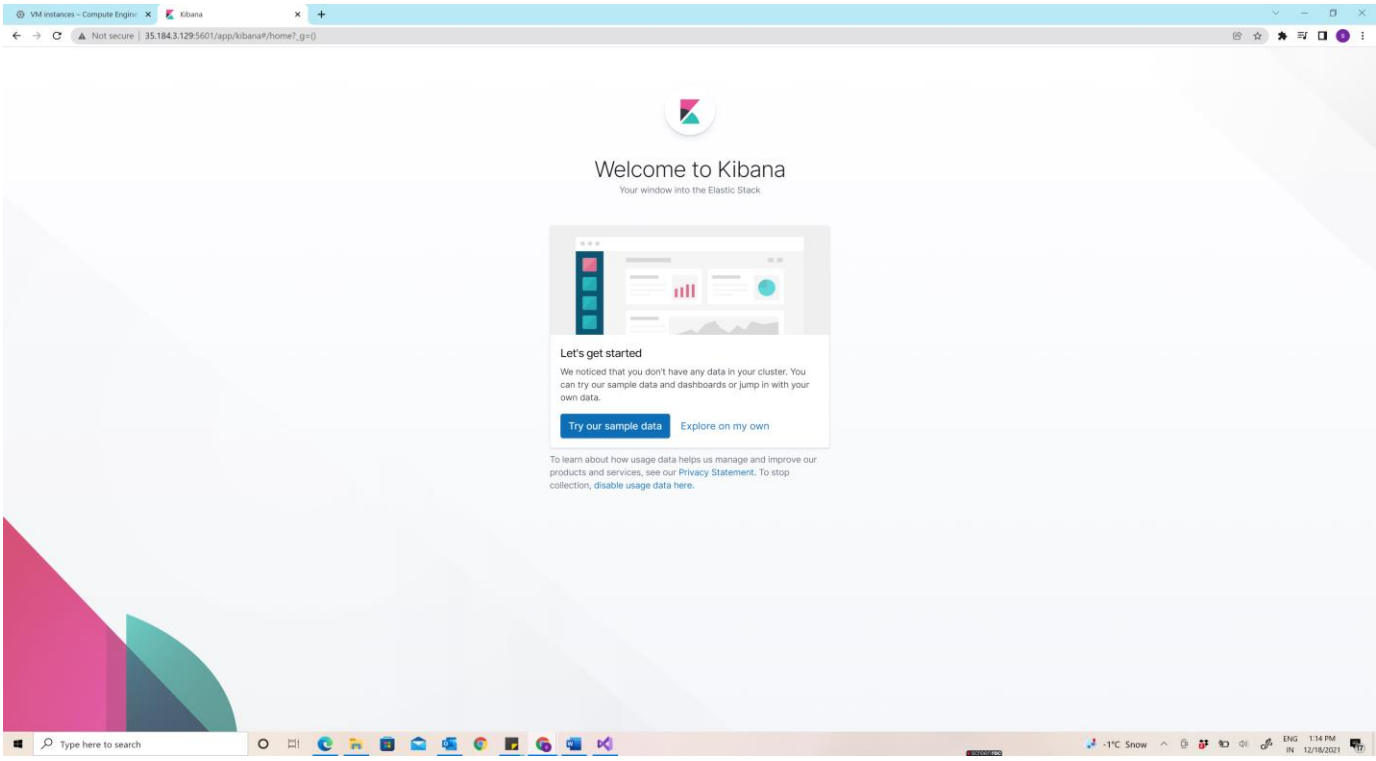
Loaded 311 service request csv file

```
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ wget https://www.dropbox.com/sh/smx7s2f32y4izkk/AADhiDbPkwjLMYfrVDu76PvXa/311_service.csv?dl=0
--2021-12-18 20:19:08--  https://www.dropbox.com/sh/smx7s2f32y4izkk/AADhiDbPkwjLMYfrVDu76PvXa/311_service.csv?dl=0
Resolving www.dropbox.com (www.dropbox.com)... 162.125.3.18, 2620:100:601b:18::a27d:812
Connecting to www.dropbox.com (www.dropbox.com)|162.125.3.18|:443... connected.
HTTP request sent, awaiting response... 301 Moved Permanently
Location: /sh/raw/smx7s2f32y4izkk/AADhiDbPkwjLMYfrVDu76PvXa/311_service.csv [following]
--2021-12-18 20:19:08--  https://www.dropbox.com/sh/raw/smx7s2f32y4izkk/AADhiDbPkwjLMYfrVDu76PvXa/311_service.csv
Reusing existing connection to www.dropbox.com:443.
HTTP request sent, awaiting response... 302 Found
Location: https://uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com/cd/0/inline/BcHLcnR1L3qXZBTfQ_2OQdEJu7TSz6GV-iOFKnJ6aCjJ4btEoJvVFalDgLW2c1Q6jg5fGCva7Iaf87QmifPAk3mI15H98f-DOevA_wWU90Hy7Mtl0VGCqet99StP6-sfK5nyd3pVGds2NCVArM-Hi6du/file# [following]
--2021-12-18 20:19:09--  https://uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com/cd/0/inline/BcHLcnR1L3qXZBTfQ_2OQdEJu7TSz6GV-iOFKnJ6aCjJ4btEoJvVFalDgLW2c1Q6jg5fGCva7Iaf87QmifPAk3mI15H98f-DOevA_wWU90Hy7Mtl0VGCqet99StP6-sfK5nyd3pVGds2NCVArM-Hi6du/file
Resolving uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com (uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com)... 162.125.3.15, 2620:100:601b:15::a27d:80f
Connecting to uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com (uce654e85ef7cefac42c9249924f.dl.dropboxusercontent.com)|162.125.3.15|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 13247207051 (12G) [text/plain]
Saving to: '311_service.csv?dl=0'

311_service.csv?dl=0         100%[=============================================>]  12.34G   107MB/s    in 2m 28s

2021-12-18 20:21:37 (85.4 MB/s) - '311_service.csv?dl=0' saved [13247207051/13247207051]
```

Modifying relevant information in Logstash configuration file using vi text editor,

```
srilekhasampath992@dsa-ssignment2-w-0:~$ ls
elasticsearch-7.5.1                 kibana-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1.tar.gz.1
elasticsearch-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1                 logstash-7.5.1.tar.gz.2
kibana-7.5.1-linux-x86_64           logstash-7.5.1.tar.gz
srilekhasampath992@dsa-ssignment2-w-0:~$ cd logstash-7.5.1/
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ ls
'311_service.csv?dl=0'   Gemfile.lock   bin      lib            logstash-core-plugin-api   vendor
 CONTRIBUTORS            LICENSE.txt    config   logs           modules                    x-pack
 Gemfile                 NOTICE.TXT     data     logstash-core  tools
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ mv '311_service.csv?dl=0' newyork.csv
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ ls
CONTRIBUTORS  Gemfile.lock  NOTICE.TXT  config  lib   logstash-core            modules      tools   x-pack
Gemfile       LICENSE.txt   bin         data    logs  logstash-core-plugin-api  newyork.csv  vendor
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ vi newyork.config
srilekhasampath992@dsa-ssignment2-w-0:~/logstash-7.5.1$ vi newyork.config
```

Starting Kibana using External IP of the cluster and Port number



Creating Index pattern