kibana

12/18/2021

ASSIGNMENT 2

DATA VISUALIZATION WITH KIBANA

BDAT 1002

Submitted to: Prof. Saber Amini

Abhijeet Singh(200489554)

Introduction

NYC Open Data is a vast trove of City Government datasets that have been made available to the public. One such dataset, 311 Requests collected since 2010, will be the focus of this analysis. This 311 service request data is updated daily and contains information about more than 24 million service requests and has about 41 columns in total. 311 is a phone number used in the U.S. that allows callers to access non-emergency municipal services, report problems to the government agencies, and request information.

Assignment Objectives

- Expose students to ELK stack as an analytic tool to analyze streaming realistic big data.
- To gain experience working with open-ended data analysis questions.
- Know more about Logstash, Elasticsearch, advanced queries, charts, maps and dashboards using Kibana

Problem Background

You have been hired as a data analyst by the city of New York to gain valuable insights from their huge data set for 311 service requests. Your task is to use the ELK stack you successfully installed and configured in your GCP platform. Successful completion of this task includes creating a Logstash configuration file (a sample is given to you) as well as a geo-point template (for maps), creating a GCP instance and firing Logstash to ingest the NYC 311 service request data into Elasticsearch and using Kibana to analyse and visualize the results as per the questions given. The required results are: Code for your Logstash configuration file and geo-point template, results for the analytical questions (tables, charts, tag clouds, maps and dashboard) in MS Word or PDF document. Where applicable, show the syntax/code or capture screenshots for all your analysis.

PART 1

Steps to get started with ELK Tool

- As instructed, we downloaded Elasticsearch, Kibana and Logstash on our GCP VM instance to begin with the analysis.
- All the three files were unzipped, configuration changes such as modifying the network host, local host and the number of nodes to Elasticsearch and Kibana were carried out.
- The two files were made to run on consoles simultaneously.
- The firewall rules were created separately for Elasticsearch and Kibana.
- We used the External IP address and the Kibana port (5601) to open Kibana on the browser.
- The dataset in our instance under Logstash was loaded.
- We downloaded the Logstash configuration file and changed the path of our dataset.
- After running, an index was created in Kibana containing our dataset.
- Since the answers are based on visualizations, the visualizations are created by using the hosting url generated.

Analysis on using ELK analytics tool:

Elastic Search is used for storing and running large volumes of data.

Provides the users with real time data analysis, aggregate data to gain statistics and visualizations.

Open source analytics engine that gives multiple hosting options.

<u>As a team, each of us created the VM instances in our own systems, to create a backup system for one another. If one system fails, we can share and coordinate the work accordingly.</u>

Logstash configuration file:

```
input {
   file {
      path => “/home/cathiemosh19/logstash-7.5.1/nycinfo.csv”
      start_position => “beginning”
     sincedb_path => “/dev/null”
     }
  }

filter {
        csv {
              separator =>”,”
              columns => [“Unique Key”,”Created Date”,”Closed Date”,”Agency”,”Agency
Name”,”Complaint Type”,”Descriptor”,”Location Type”,”Incident Zip”,”Incident
Address”,”Street Name”,”Cross Street 1”,”Cross Street 2”,”Intersection Street 1”,”Intersection
Street 2”,”Address Type”,”City”,”Landmark”,”Facility Type”,”Status”,”Due
Date”,”Resolution Description”,”Resolution Action Updated Date”,”Community
Board”,”BBL”,”Borough”,”X Coordinate (State Plane)”,”Y Coordinate (State Plane)”,”Open
Data Channel Type”,”Park Facility Name”,”Park Borough”,”Vehicle Type”,”Taxi Company
Borough”,”Taxi Pick Up Location”,”Bridge Highway Name”,”Bridge Highway
Direction”,”Road Ramp”,”Bridge Highway Segment”,”Latitude”,”Longitude”,”Location”]
                   }

date{    match => [“Created Date”, “MM/dd/yyyy hh:mm:ss a”]
         target => “Created Date”
}
date{    match => [“Closed Date”, “MM/dd/yyyy hh:mm:ss a”]
         target => “Closed Date”
}
date{    match => [“Due Date”, “MM/dd/yyyy hh:mm:ss a”]
         target => “Due Date”
}
date{    match => [“Resolution Action Updated Date”, “MM/dd/yyyy hh:mm:ss a”]
         target => “Resoultion Action Updated Date”
}
    mutate {convert => [“Incident Zip”,”integer”]}
    mutate {convert => [“BBL”,”integer”]}
    mutate {convert => [“X Coordinate (State Plane)”,”integer”]}
    mutate {convert => [“Y Coordinate (State Plane)”,”integer”]}
    mutate {convert => [“Latitude”,”float”]}
    mutate {convert => [“Longitude”,”float”]}
    mutate {copy =>
                  { “Longitude” => “[location][lon]”
                   “Latitude” => “[location][lat]” }
             }
    mutate {replace => { “Location” => “%{Longitude},%{Latitude}” }}
    }

output {
 elasticsearch {
 hosts => “localhost”
 index => “nycinfo”

  }
stdout {codec => dots}
}
```

Geo-point template code:

```
PUT _template/geotemplate
{
  “index_patterns”:[“nycinfo”],
  “settings”:{},
    “mappings”:{
      “properties”:{
      “Location”:{
          “type”:”geo_point”
        }
      }
    },
    “aliases”: {}
}
```

PART 2

Data Analysis and Visualizations

Visualizations and Analysis:

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

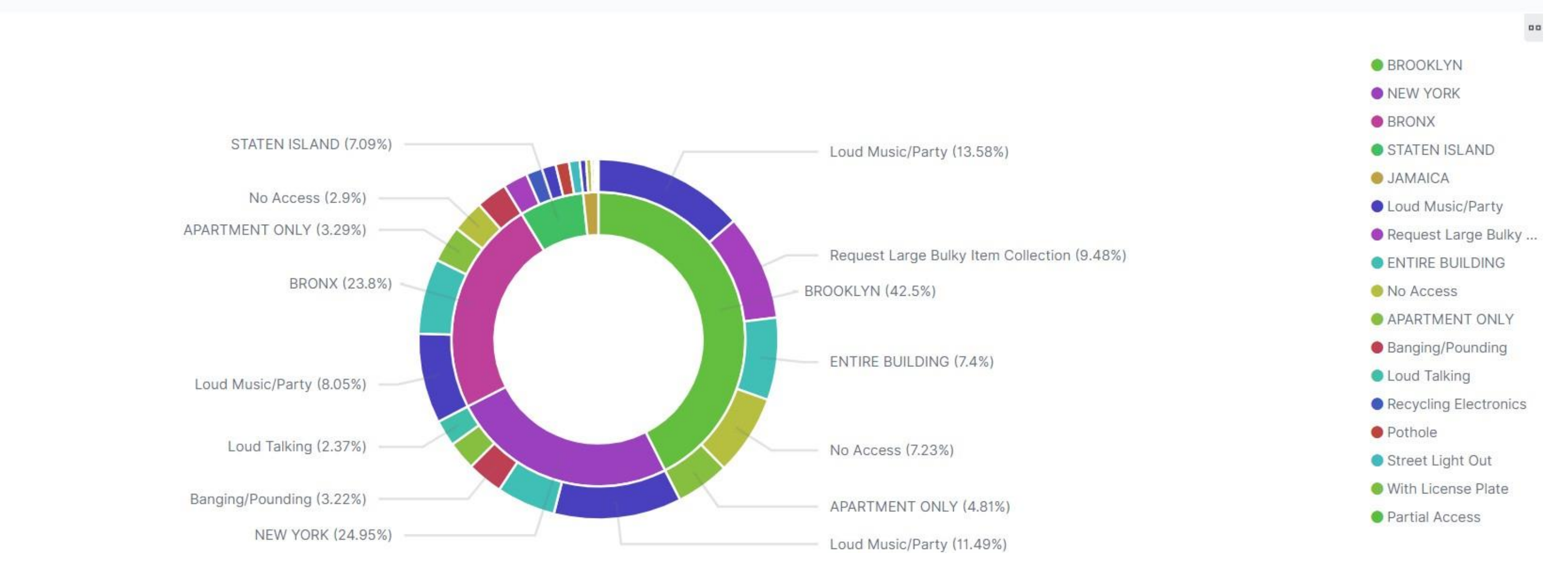


Fig. 1.1

Analysis: The 5 cities with the highest calls are:

1. **Brooklyn** – Loud Music Party(21.63), Entire Building(7.4No Access(2.9), Request Large Bulky Item Collection(9.48%) and Apartment Only(8.1%).
2. **New York**- Loud Music/ Party, Entire Building, Banging/Pounding, Apartment Only, N/A
3. **Bronx**- Entire Building, Loud Music/Party, Apartment Only, No Access, Banging/Pounding
4. **Staten Island** – Recycling Electronics, Request Large Bulky Item Collection, Loud Music Party, Pothole, Street Light Out.
5. **Jamaica** – Loud Music Party, No Access, With License Plate, Banging/Pounding, Partial Access.

What are the top 5 location types with top 5 Descriptors?

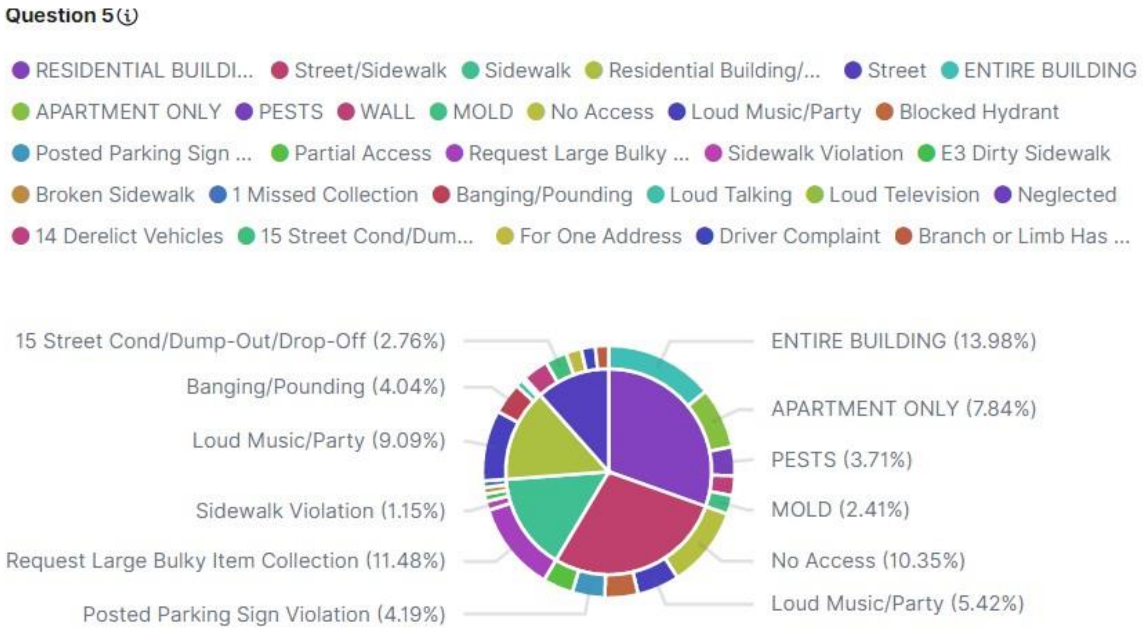


Fig. 1.2

Analysis: The top 5 location types with highest number of Descriptor calls are:

1. Residential Buildings
2. Street/Sidewalk
3. Residential Building/House
4. Sidewalk
5. Street
    - Residential Buildings have maximum number of complaints regarding Heat/hot water provisions in the data loaded.
    - Streets and Sidewalks combined have 2.76 % and 1.15% of complaints regarding Illegal Parking.
    - Residential Buildings and Houses complain about noise from other residences.
    - There have been about 20,000 complaints regarding the street conditions.

What is the count of Complaint status for the top 5 cities?

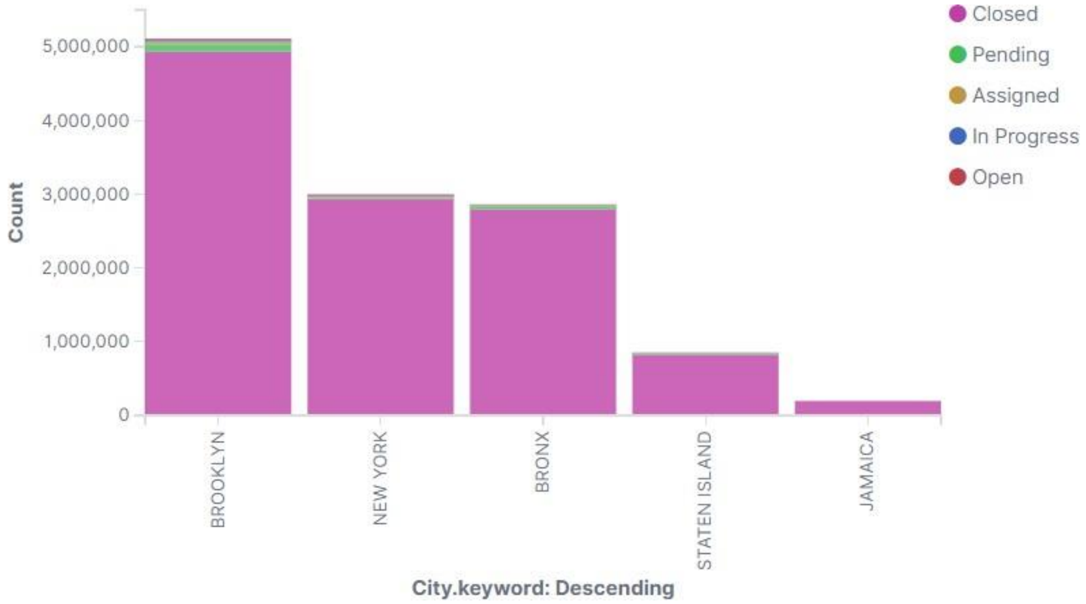


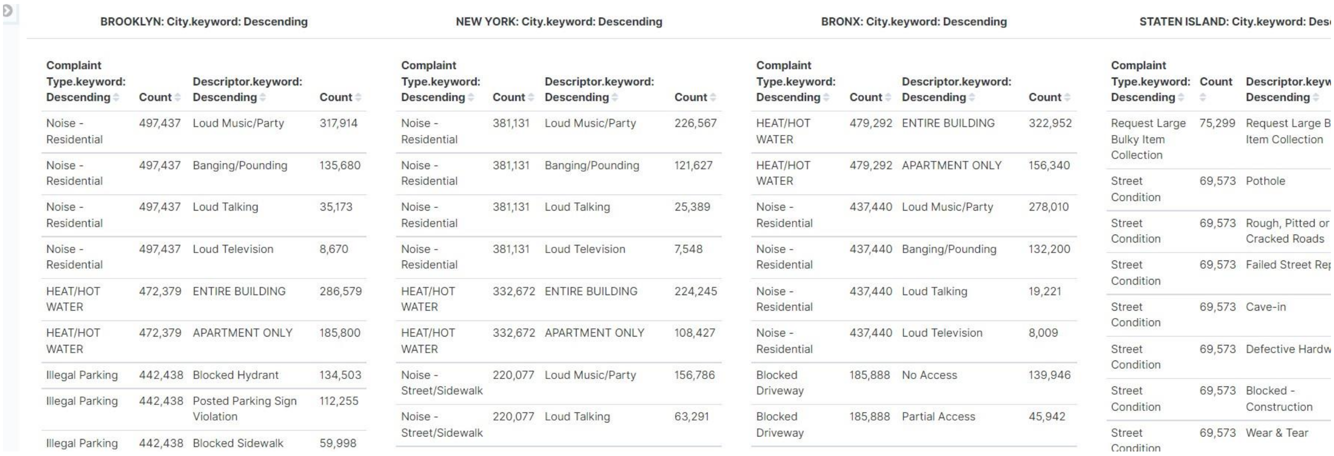
Fig 1.3

Analysis:

- Brooklyn has the highest number of complaints that are assigned as ‘Closed’ status that imply they are either resolved or disregarded depending on their severity.
- New York has about 3,000,000 complaints that are resolved and closed
- Jamaica has the least number of closed complaints.
- Very few complaints are assigned to be looked after and even fewer complaints are open.

Analysis and Visualizations:

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.



BROOKLYN: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Noise - Residential	497,437	Loud Music/Party	317,914
Noise - Residential	497,437	Banging/Pounding	135,680
Noise - Residential	497,437	Loud Talking	35,173
Noise - Residential	497,437	Loud Television	8,670
HEAT/HOT WATER	472,379	ENTIRE BUILDING	286,579
HEAT/HOT WATER	472,379	APARTMENT ONLY	185,800
Illegal Parking	442,438	Blocked Hydrant	134,503
Illegal Parking	442,438	Posted Parking Sign Violation	112,255
Illegal Parking	442,438	Blocked Sidewalk	59,998

NEW YORK: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Noise - Residential	381,131	Loud Music/Party	226,567
Noise - Residential	381,131	Banging/Pounding	121,627
Noise - Residential	381,131	Loud Talking	25,389
Noise - Residential	381,131	Loud Television	7,548
HEAT/HOT WATER	332,672	ENTIRE BUILDING	224,245
HEAT/HOT WATER	332,672	APARTMENT ONLY	108,427
Noise - Street/Sidewalk	220,077	Loud Music/Party	156,786
Noise - Street/Sidewalk	220,077	Loud Talking	63,291

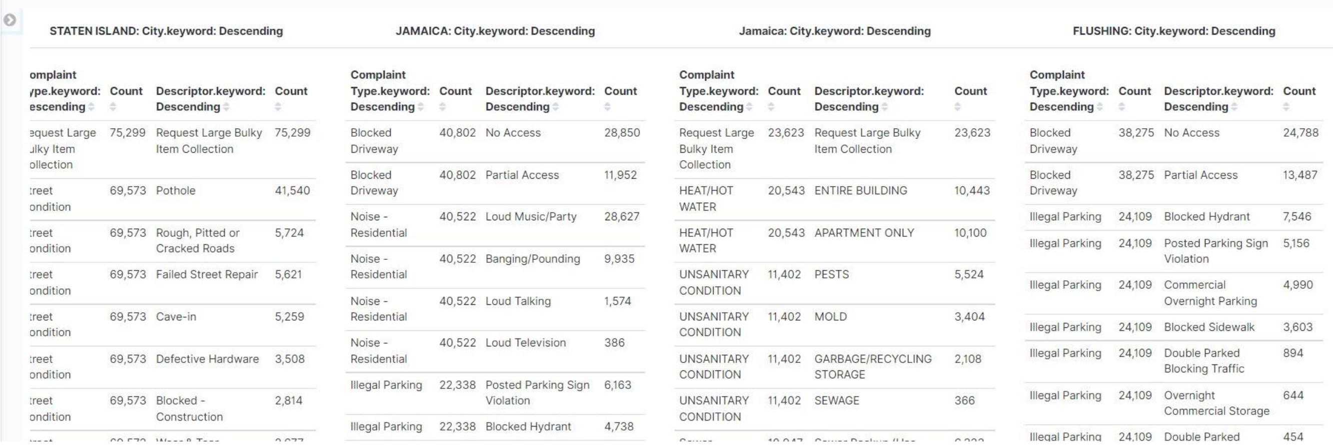
BRONX: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
HEAT/HOT WATER	479,292	ENTIRE BUILDING	322,952
HEAT/HOT WATER	479,292	APARTMENT ONLY	156,340
Noise - Residential	437,440	Loud Music/Party	278,010
Noise - Residential	437,440	Banging/Pounding	132,200
Noise - Residential	437,440	Loud Talking	19,221
Noise - Residential	437,440	Loud Television	8,009
Blocked Driveway	185,888	No Access	139,946
Blocked Driveway	185,888	Partial Access	45,942

STATEN ISLAND: City.keyword: Des

Complaint Type.keyword: Descending	Count	Descriptor.keyw Descending
Request Large Bulky Item Collection	75,299	Request Large B Item Collection
Street Condition	69,573	Pothole
Street Condition	69,573	Rough, Pitted or Cracked Roads
Street Condition	69,573	Failed Street Rep
Street Condition	69,573	Cave-in
Street Condition	69,573	Defective Hardw
Street Condition	69,573	Blocked - Construction
Street Condition	69,573	Wear & Tear

Fig. 1.4



STATEN ISLAND: City.keyword: Descending

omplaint ype.keyword: escending	Count	Descriptor.keyword: Descending	Count
equest Large ulky Item ollection	75,299	Request Large Bulky Item Collection	75,299
treet ondition	69,573	Pothole	41,540
treet ondition	69,573	Rough, Pitted or Cracked Roads	5,724
treet ondition	69,573	Failed Street Repair	5,621
treet ondition	69,573	Cave-in	5,259
treet ondition	69,573	Defective Hardware	3,508
treet ondition	69,573	Blocked - Construction	2,814

JAMAICA: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Blocked Driveway	40,802	No Access	28,850
Blocked Driveway	40,802	Partial Access	11,952
Noise - Residential	40,522	Loud Music/Party	28,627
Noise - Residential	40,522	Banging/Pounding	9,935
Noise - Residential	40,522	Loud Talking	1,574
Noise - Residential	40,522	Loud Television	386
Illegal Parking	22,338	Posted Parking Sign Violation	6,163
Illegal Parking	22,338	Blocked Hydrant	4,738

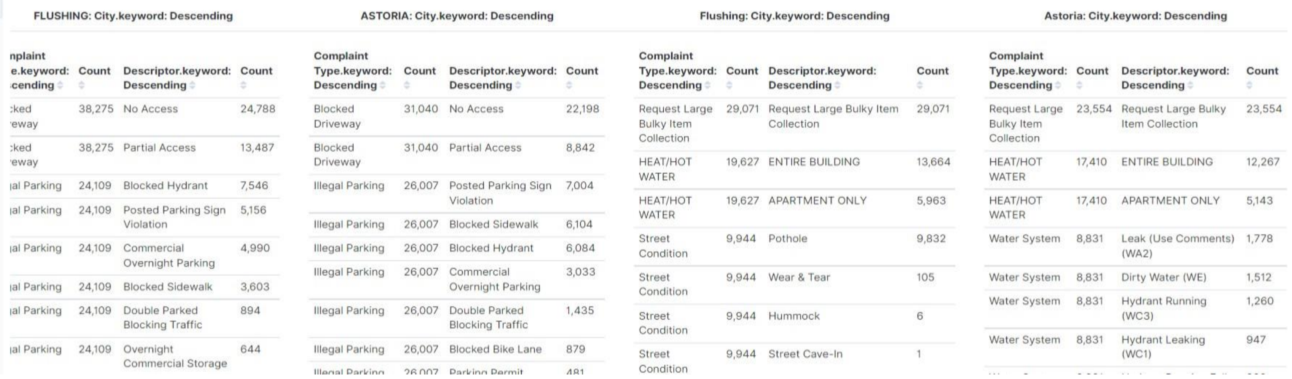
Jamaica: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	23,623	Request Large Bulky Item Collection	23,623
HEAT/HOT WATER	20,543	ENTIRE BUILDING	10,443
HEAT/HOT WATER	20,543	APARTMENT ONLY	10,100
UNSANITARY CONDITION	11,402	PESTS	5,524
UNSANITARY CONDITION	11,402	MOLD	3,404
UNSANITARY CONDITION	11,402	GARBAGE/RECYCLING STORAGE	2,108
UNSANITARY CONDITION	11,402	SEWAGE	366

FLUSHING: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Blocked Driveway	38,275	No Access	24,788
Blocked Driveway	38,275	Partial Access	13,487
Illegal Parking	24,109	Blocked Hydrant	7,546
Illegal Parking	24,109	Posted Parking Sign Violation	5,156
Illegal Parking	24,109	Commercial Overnight Parking	4,990
Illegal Parking	24,109	Blocked Sidewalk	3,603
Illegal Parking	24,109	Double Parked Blocking Traffic	894
Illegal Parking	24,109	Overnight Commercial Storage	644
Illegal Parking	24,109	Double Parked	454

Fig. 1.5



FLUSHING: City.keyword: Descending

nplaint e.keyword: cending	Count	Descriptor.keyword: Descending	Count
cked eway	38,275	No Access	24,788
cked eway	38,275	Partial Access	13,487
al Parking	24,109	Blocked Hydrant	7,546
al Parking	24,109	Posted Parking Sign Violation	5,156
al Parking	24,109	Commercial Overnight Parking	4,990
al Parking	24,109	Blocked Sidewalk	3,603
al Parking	24,109	Double Parked Blocking Traffic	894
al Parking	24,109	Overnight Commercial Storage	644

ASTORIA: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Blocked Driveway	31,040	No Access	22,198
Blocked Driveway	31,040	Partial Access	8,842
Illegal Parking	26,007	Posted Parking Sign Violation	7,004
Illegal Parking	26,007	Blocked Sidewalk	6,104
Illegal Parking	26,007	Blocked Hydrant	6,084
Illegal Parking	26,007	Commercial Overnight Parking	3,033
Illegal Parking	26,007	Double Parked Blocking Traffic	1,435
Illegal Parking	26,007	Blocked Bike Lane	879
Illegal Parking	26,007	Parking Permit	481

Flushing: City.keyword: Descending

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	29,071	Request Large Bulky Item Collection	29,071
HEAT/HOT WATER	19,627	ENTIRE BUILDING	13,664
HEAT/HOT WATER	19,627	APARTMENT ONLY	5,963
Street Condition	9,944	Pothole	9,832
Street Condition	9,944	Wear & Tear	105
Street Condition	9,944	Hummock	6
Street Condition	9,944	Street Cave-In	1

Astoria: City.keyword: Descending

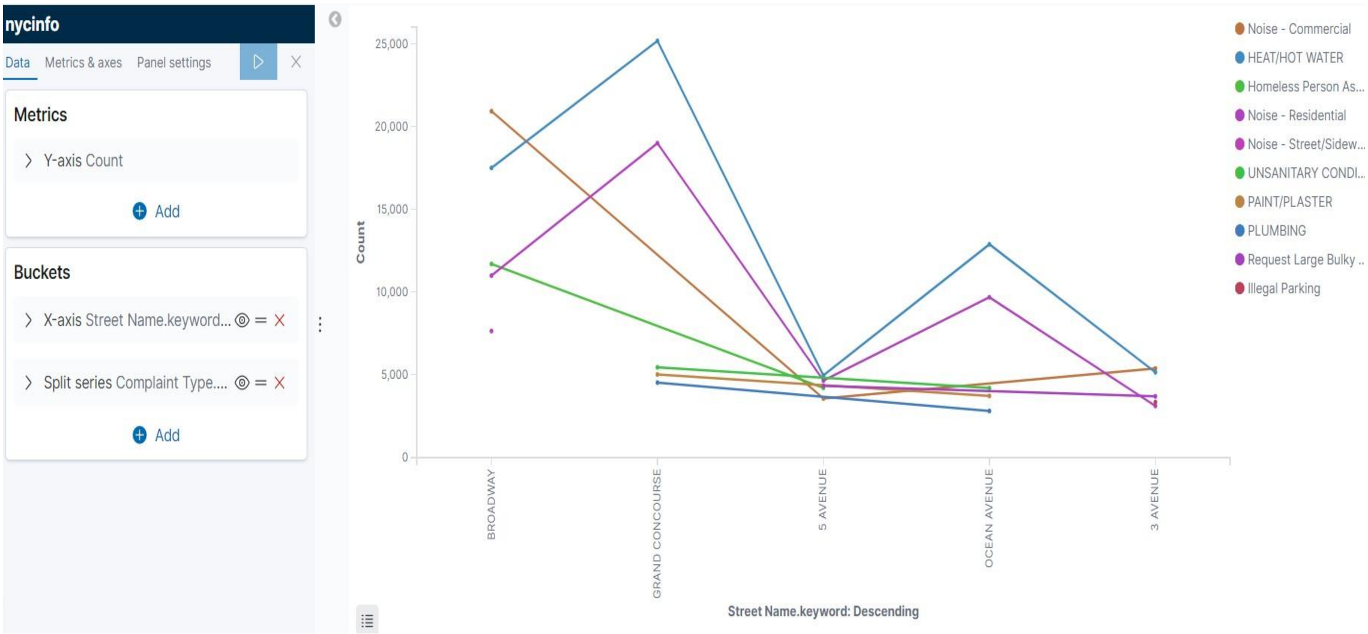
Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count
Request Large Bulky Item Collection	23,554	Request Large Bulky Item Collection	23,554
HEAT/HOT WATER	17,410	ENTIRE BUILDING	12,267
HEAT/HOT WATER	17,410	APARTMENT ONLY	5,143
Water System	8,831	Leak (Use Comments) (WA2)	1,778
Water System	8,831	Dirty Water (WE)	1,512
Water System	8,831	Hydrant Running (WC3)	1,260
Water System	8,831	Hydrant Leaking (WC1)	947

Fig. 1.6

Analysis: The top 10 cities with the highest calls are:

1. Brooklyn
2. New York
3. Bronx
4. Staten Island
5. JAMAICA
6. Jamaica
7. FLUSHING
8. Astoria
9. Flushing
10. Ridgewood

What are the top 5 streets with the highest complaints? What are the top 5 complaints for each of these streets?



nycinfo

Metrics

Y-axis Count

Buckets

X-axis Street Name.keyword...

Split series Complaint Type....

Fig. 1.7

Analysis: The top 5 streets with the highest number of complaints are:

1. Broadway
2. Grand Concourse
3. 5 Avenue
4. Ocean Avenue
5. 3 Avenue
   - Broadway street has over 25000 complaints regarding Heat/Hot water, Noise from Commercial areas. It has around 21000 complaints regarding Noise from Residential areas and lack of assistance for the Homeless.
   - Grand Concourse street has the highest number of complaints regarding Heat/Hot water provision crossing the mark of 25000 complaints.

- Almost all the streets have very few complaints regarding plumbing operations.

What are the top 5 Taxi Pick-Up locations?

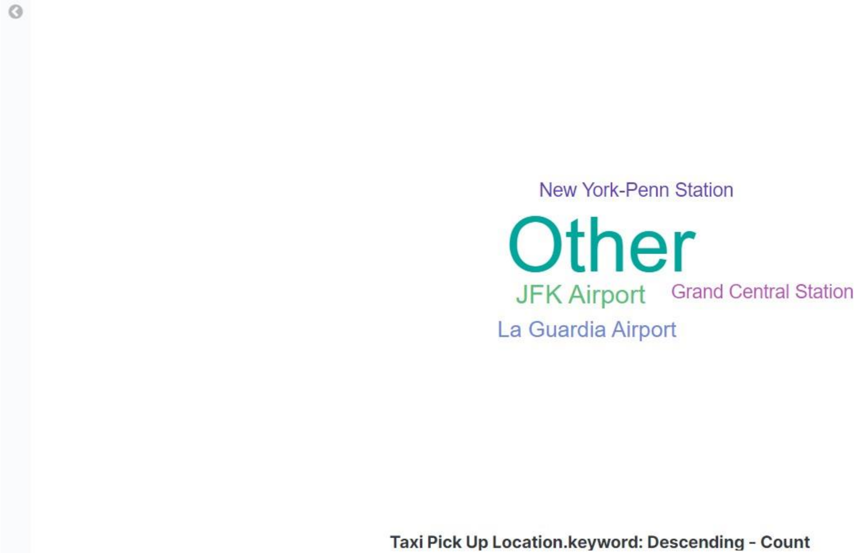


Fig. 1.8

Analysis: The top 5 Taxi Pick-up Locations are

- Other
- JFK Airport
- La Guardia Airport
- Grand Central Station
- New York – Penn Station

Analysis and Visualizations:

Create a tag cloud representing the top 20 call descriptors.

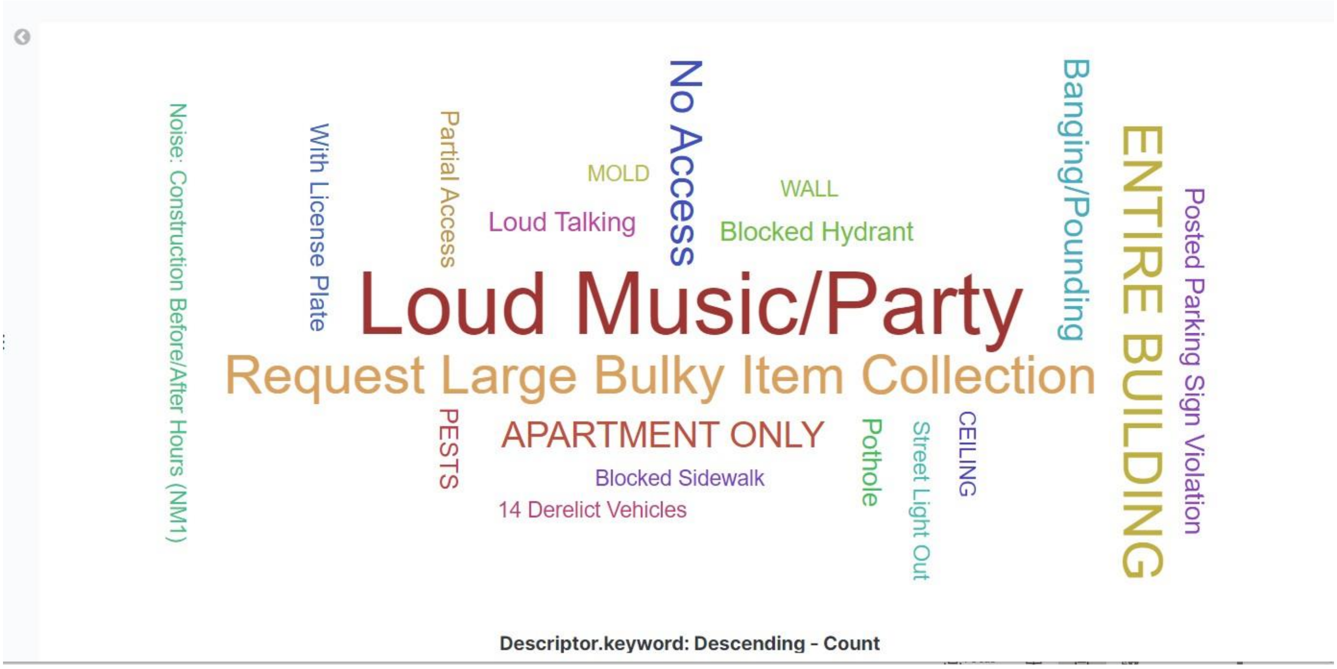


Fig. 1.9

Analysis:

The top 20 call descriptors are as shown in the tag cloud with Loud Music/Party making up for maximum complaint calls.

What are the top 5 agencies in City Of Brooklyn?

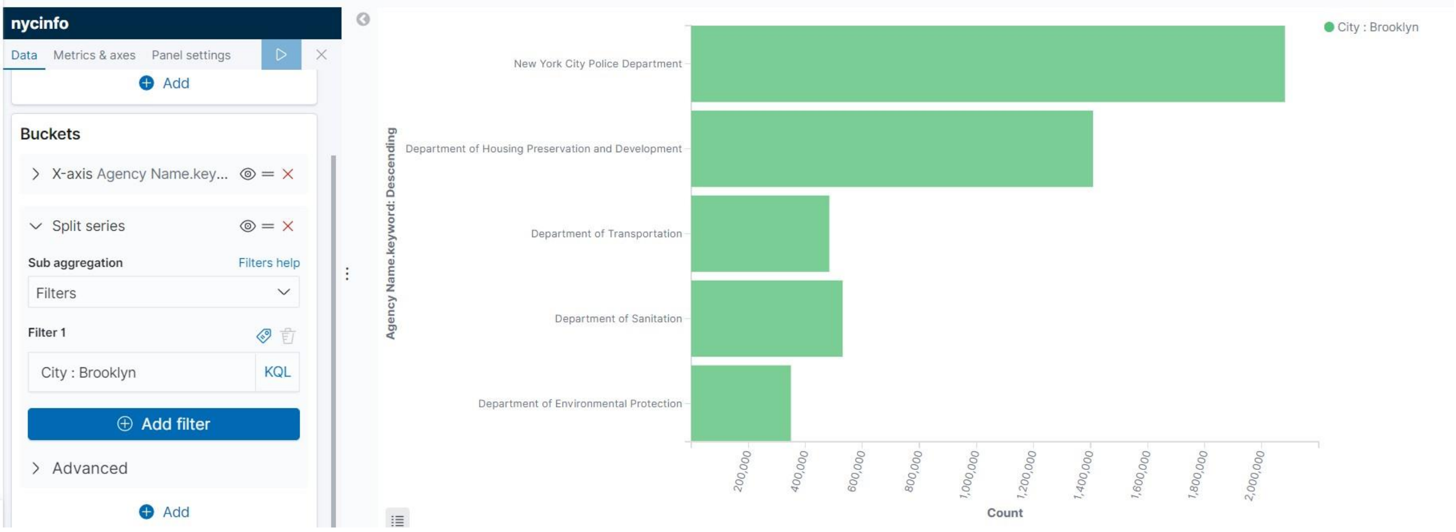
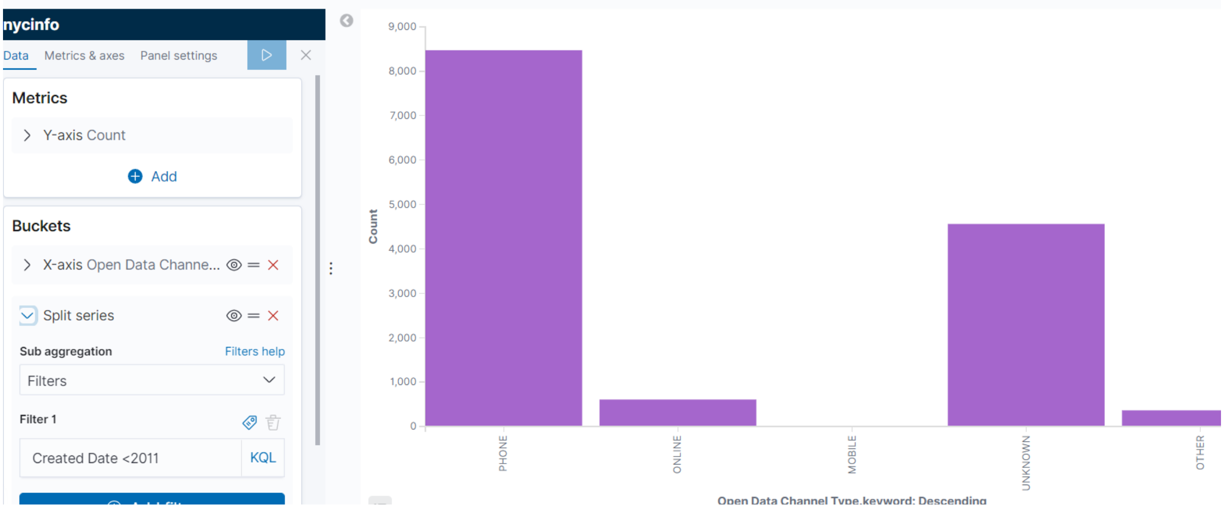


Fig. 1.10

Analysis:

New York City Police Department is the agency doing/receiving maximum calls in the city of Brooklyn, followed by Department of Housing Preservation and Development, Department of transportation, Department of Sanitation and Department of Environmental Protection. Used the filter in KQL for the city of Brooklyn.

9. Show the count of calls made through Data Channel types before the year 2011.



nycinfo

Data | Metrics & axes | Panel settings

Metrics

Y-axis Count

Add

Buckets

X-axis Open Data Channe...

Split series

Sub aggregation — Filters help

Filters

Filter 1

Created Date <2011 — KQL

Fig. 1.11

Analysis:

The maximum number of calls made before 2011 is through Phone Data Channel Type.

The second most number of calls were made through Unknown sources.

Coordinate MAP done collectively.

A coordinated map of all the major call descriptors in each city

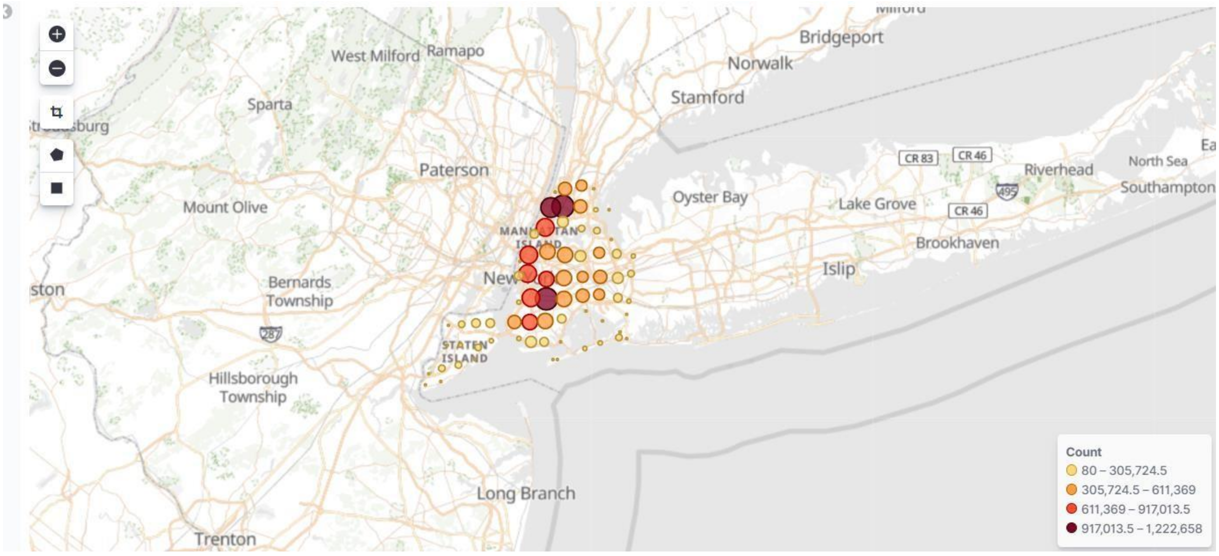


Fig. 1.12

Code:

```
input {
   file {
      path => "/home/abhijeetsingh2506/311_service.csv"
      start_position => "beginning"
      sincedb_path => "/dev/null"
      }
  }

filter {
        csv {
              separator =>","
              columns => ["Unique Key","Created Date","Closed Date","Agency","Agency
Name","Complaint Type","Descriptor","Location Type","Incident Zip","Incident
Address","Street Name","Cross Street 1","Cross Street 2","Intersection Street 1","Intersection
Street 2","Address Type","City","Landmark","Facility Type","Status","Due
Date","Resolution Description","Resolution Action Updated Date","Community
Board","BBL","Borough","X Coordinate (State Plane)","Y Coordinate (State Plane)","Open
Data Channel Type","Park Facility Name","Park Borough","Vehicle Type","Taxi Company
Borough","Taxi Pick Up Location","Bridge Highway Name","Bridge Highway
Direction","Road Ramp","Bridge Highway Segment","Latitude","Longitude","Location"]
                  }

date{   match => ["Created Date", "MM/dd/yyyy hh:mm:ss a"]
        target => "Created Date"
}
date{   match => ["Closed Date", "MM/dd/yyyy hh:mm:ss a"]
        target => "Closed Date"
}
date{   match => ["Due Date", "MM/dd/yyyy hh:mm:ss a"]
        target => "Due Date"
}
date{   match => ["Resolution Action Updated Date", "MM/dd/yyyy hh:mm:ss a"]
        target => "Resoultion Action Updated Date"
}
    mutate {convert => ["Incident Zip","integer"]}
    mutate {convert => ["BBL","integer"]}
    mutate {convert => ["X Coordinate (State Plane)","integer"]}
    mutate {convert => ["Y Coordinate (State Plane)","integer"]}
    mutate {convert => ["Latitude","float"]}
    mutate {convert => ["Longitude","float"]}
    mutate {copy =>
                  { "Longitude" => "[location][lon]"
                    "Latitude" => "[location][lat]" }
              }
    mutate {replace => { "Location" => "%{Latitude},%{Longitude}" }}
    }
output {
 elasticsearch  {
 hosts => "localhost"
 index => "nycinfo"
  }
stdout {codec => dots}
}
```

Dashboard Creation

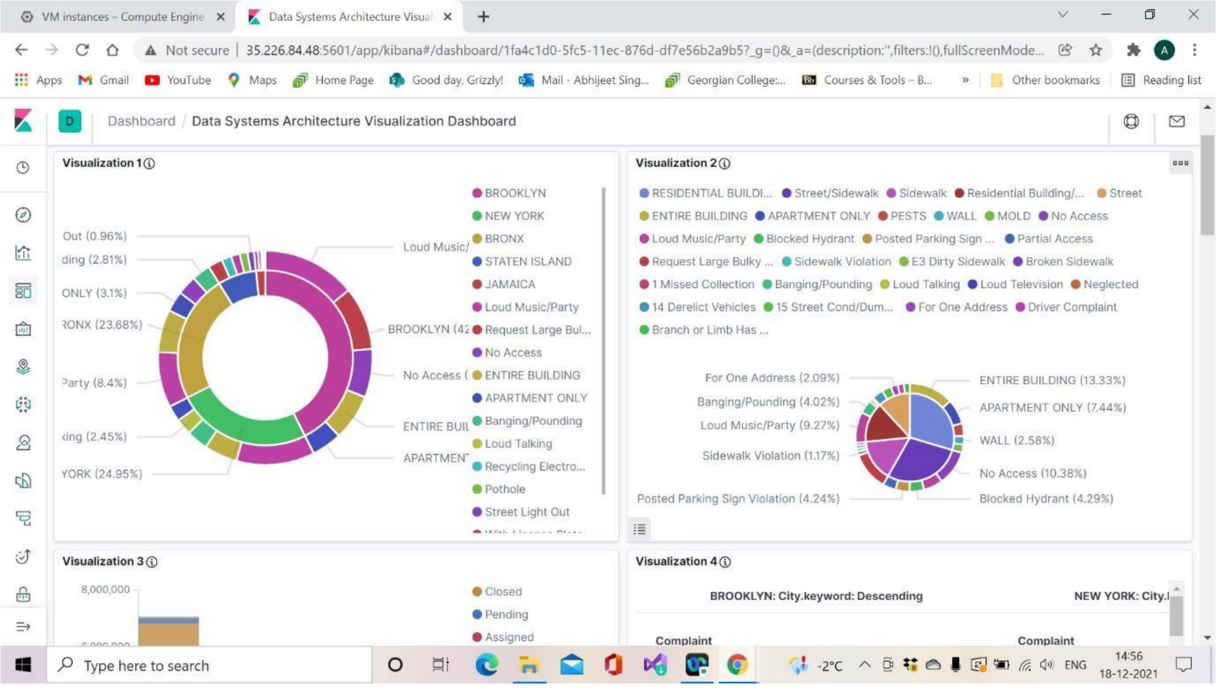
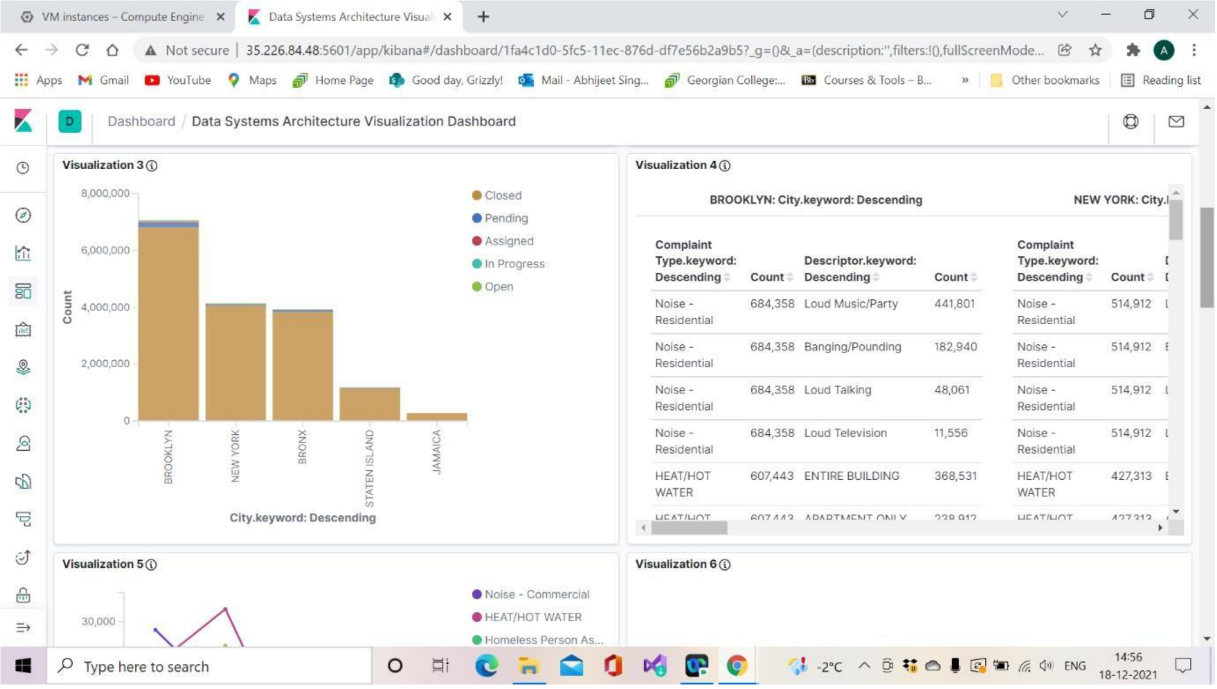


Fig. 1.13



BROOKLYN: City.keyword: Descending | NEW YORK: City.

Complaint Type.keyword: Descending	Count	Descriptor.keyword: Descending	Count	Complaint Type.keyword: Descending	Count
Noise - Residential	684,358	Loud Music/Party	441,801	Noise - Residential	514,912
Noise - Residential	684,358	Banging/Pounding	182,940	Noise - Residential	514,912
Noise - Residential	684,358	Loud Talking	48,061	Noise - Residential	514,912
Noise - Residential	684,358	Loud Television	11,556	Noise - Residential	514,912
HEAT/HOT WATER	607,443	ENTIRE BUILDING	368,531	HEAT/HOT WATER	427,313
HEAT/HOT	607,443	APARTMENT ONLY	238,912	HEAT/HOT	427,313

Fig. 1.14

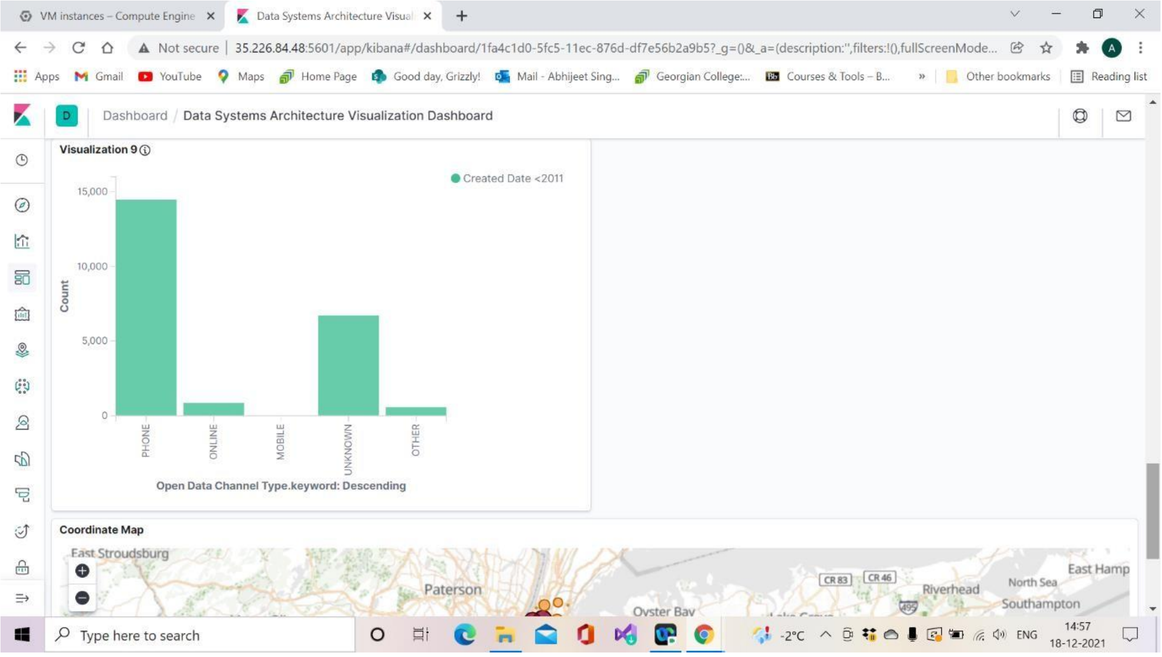


Fig. 1.15

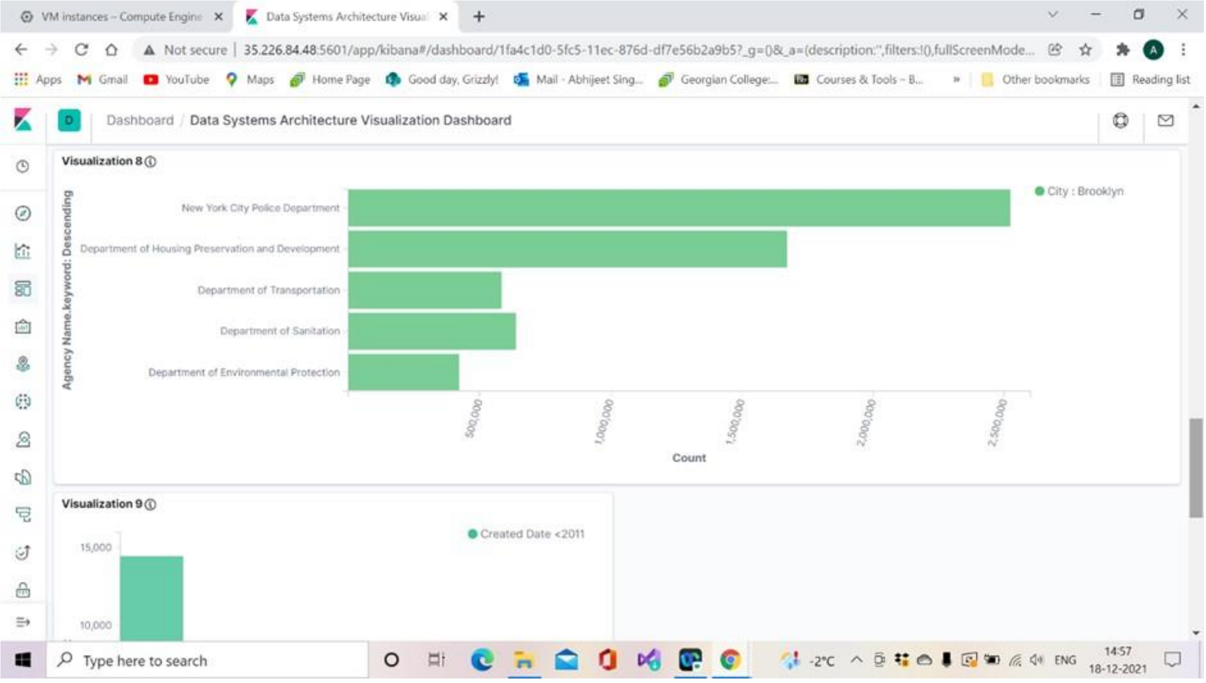

Dashboard / Data Systems Architecture Visualization Dashboard
Visualization 8
City : Brooklyn
New York City Police Department
Department of Housing Preservation and Development
Department of Transportation
Department of Sanitation
Department of Environmental Protection
Agency Name.keyword: Descending
500,000
1,000,000
1,500,000
2,000,000
2,500,000
Count
Visualization 9
Created Date <2011
15,000
10,000

Fig. 1.16

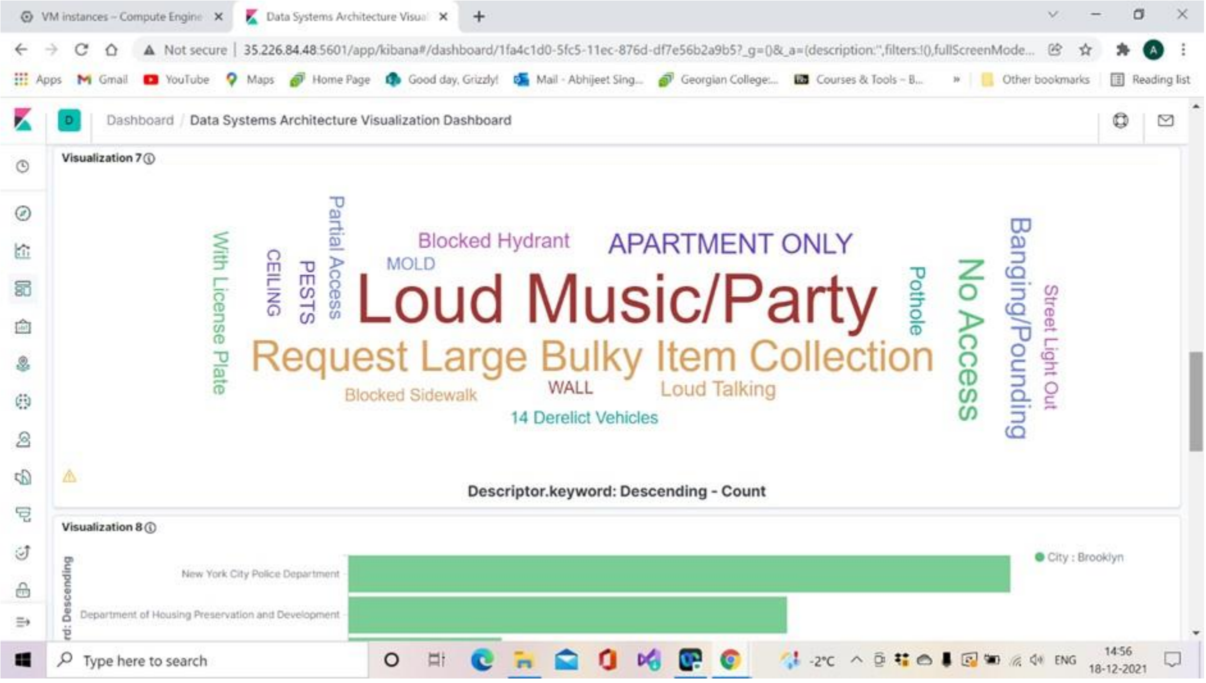


Fig. 1.17

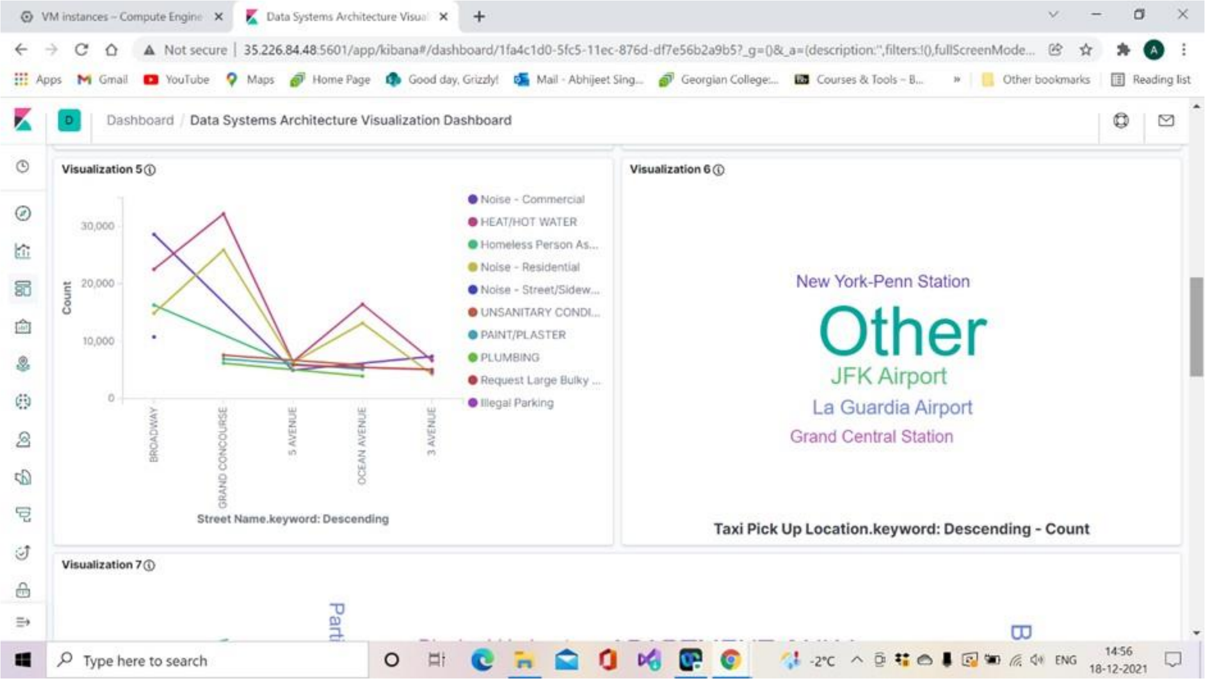


Fig. 1.18

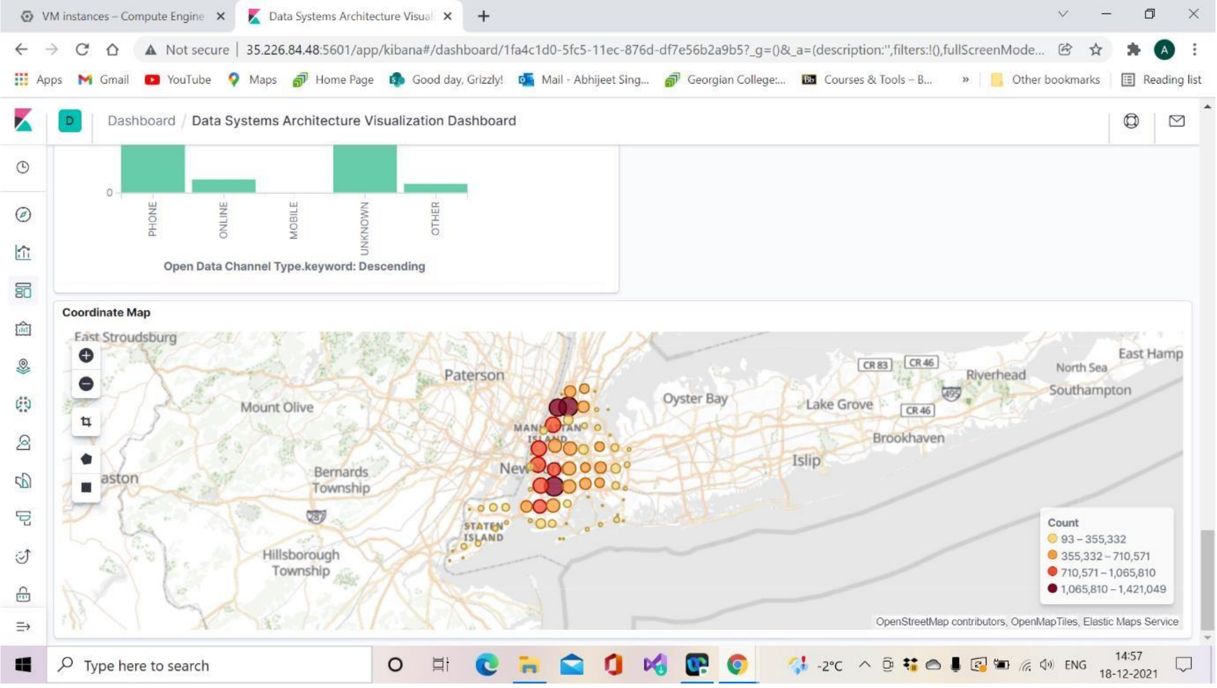
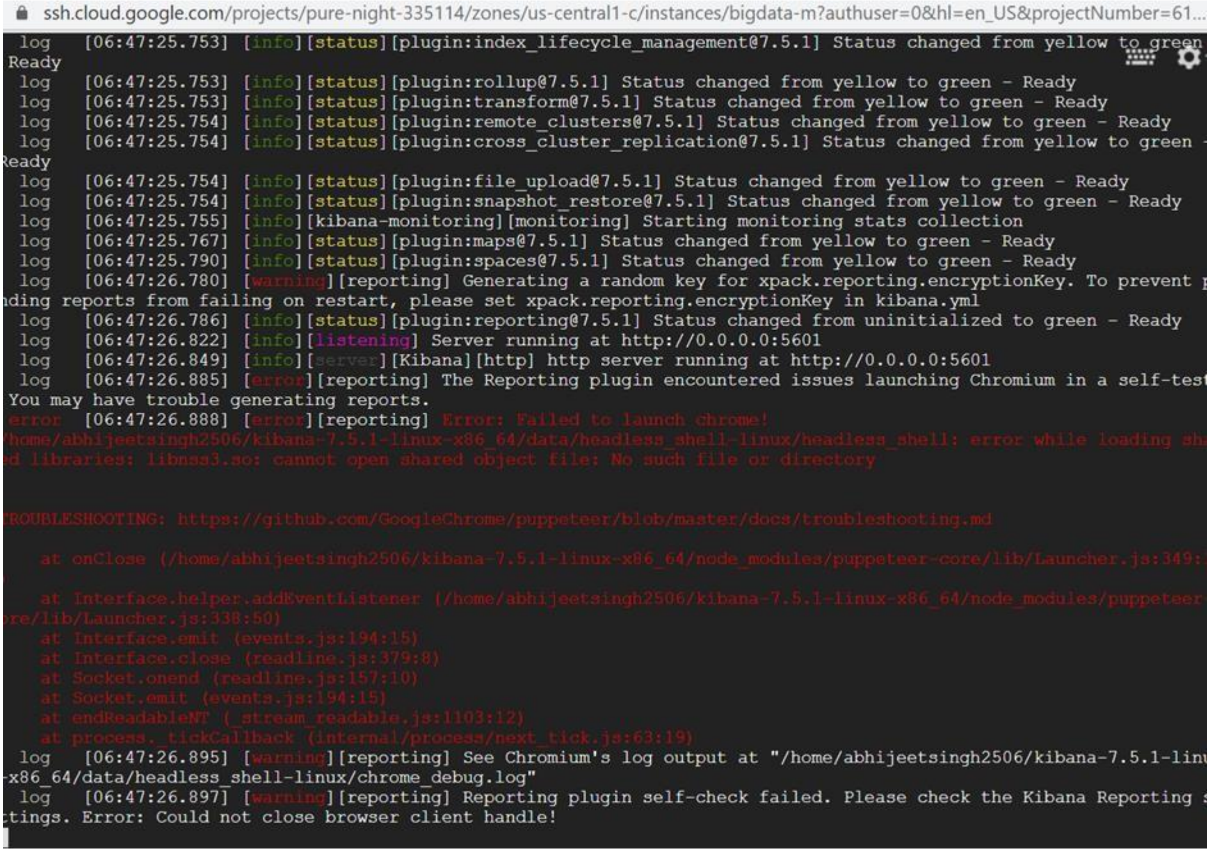


Fig.1.19

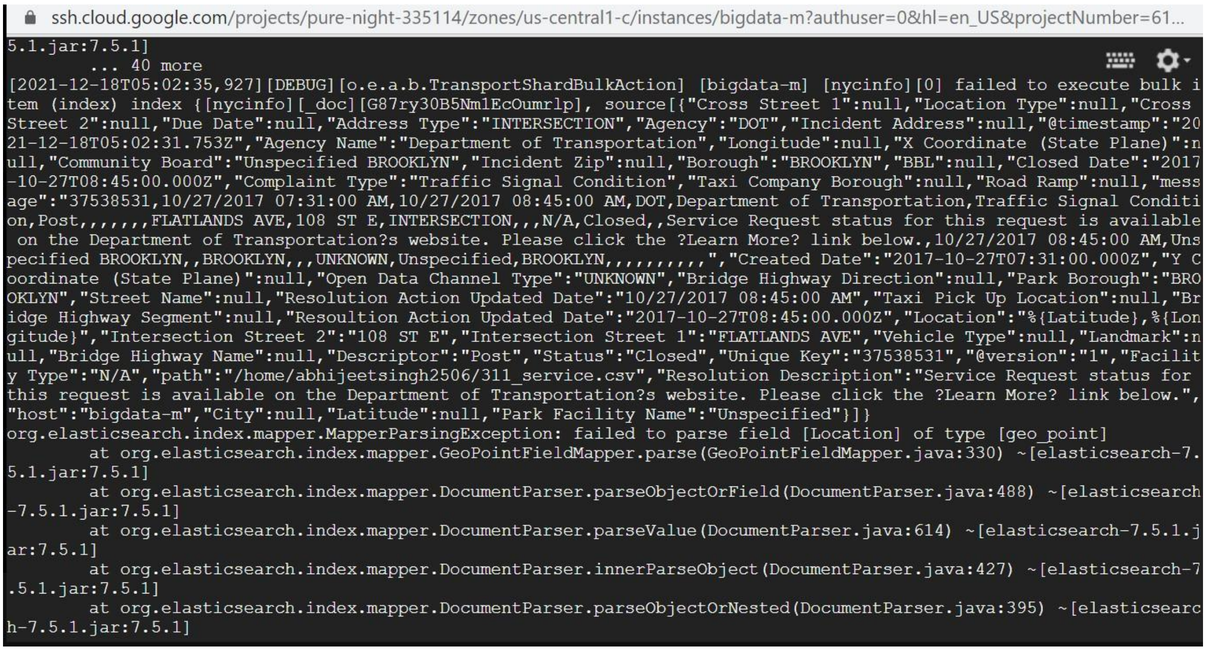
Snippets of Code running behind the Visualizations hosted on Kibana:



ssh.cloud.google.com/projects/pure-night-335114/zones/us-central1-c/instances/bigdata-m?authuser=0&hl=en_US&projectNumber=61...

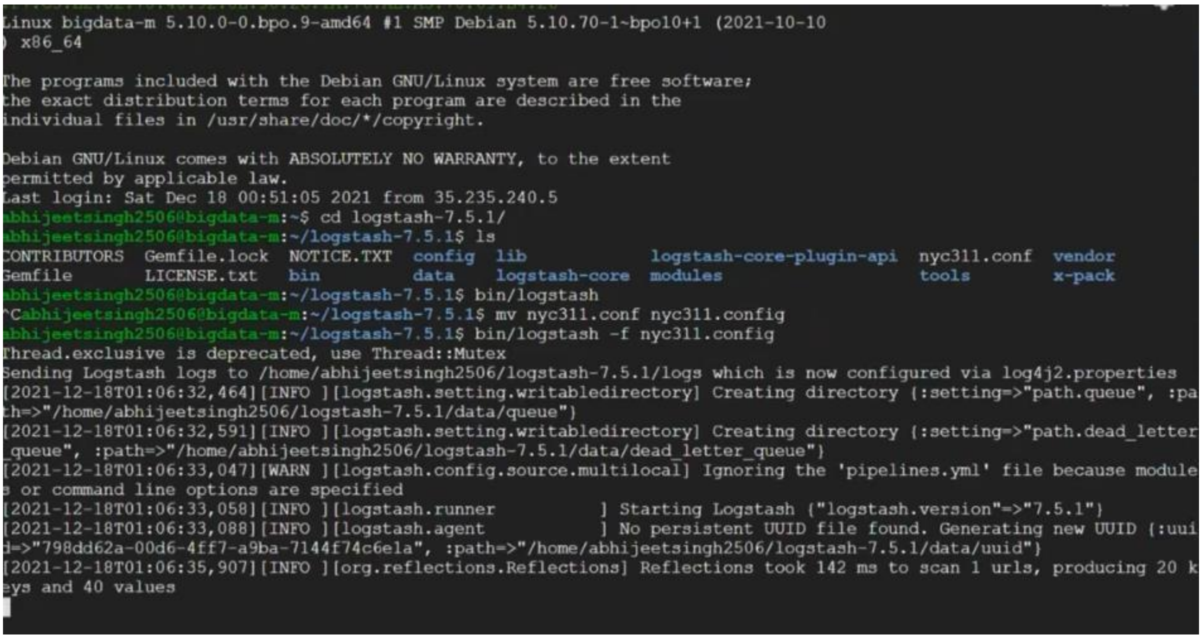
```
 log   [06:47:25.753] [info][status][plugin:index_lifecycle_management@7.5.1] Status changed from yellow to green
 Ready
 log   [06:47:25.753] [info][status][plugin:rollup@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.753] [info][status][plugin:transform@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.754] [info][status][plugin:remote_clusters@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.754] [info][status][plugin:cross_cluster_replication@7.5.1] Status changed from yellow to green
Ready
 log   [06:47:25.754] [info][status][plugin:file_upload@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.754] [info][status][plugin:snapshot_restore@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.755] [info][kibana-monitoring][monitoring] Starting monitoring stats collection
 log   [06:47:25.767] [info][status][plugin:maps@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:25.790] [info][status][plugin:spaces@7.5.1] Status changed from yellow to green - Ready
 log   [06:47:26.780] [warning][reporting] Generating a random key for xpack.reporting.encryptionKey. To prevent
nding reports from failing on restart, please set xpack.reporting.encryptionKey in kibana.yml
 log   [06:47:26.786] [info][status][plugin:reporting@7.5.1] Status changed from uninitialized to green - Ready
 log   [06:47:26.822] [info][listening] Server running at http://0.0.0.0:5601
 log   [06:47:26.849] [info][server][Kibana][http] http server running at http://0.0.0.0:5601
 log   [06:47:26.885] [error][reporting] The Reporting plugin encountered issues launching Chromium in a self-tes
 You may have trouble generating reports.
 error  [06:47:26.888] [error][reporting] Error: Failed to launch chrome!
/home/abhijeetsingh2506/kibana-7.5.1-linux-x86_64/data/headless_shell-linux/headless_shell: error while loading sh
ed libraries: libnss3.so: cannot open shared object file: No such file or directory


TROUBLESHOOTING: https://github.com/GoogleChrome/puppeteer/blob/master/docs/troubleshooting.md

    at onClose (/home/abhijeetsingh2506/kibana-7.5.1-linux-x86_64/node_modules/puppeteer-core/lib/Launcher.js:349:
)
    at Interface.helper.addEventListener (/home/abhijeetsingh2506/kibana-7.5.1-linux-x86_64/node_modules/puppeteer
ore/lib/Launcher.js:338:50)
    at Interface.emit (events.js:194:15)
    at Interface.close (readline.js:379:8)
    at Socket.onend (readline.js:157:10)
    at Socket.emit (events.js:194:15)
    at endReadableNT (_stream_readable.js:1103:12)
    at process._tickCallback (internal/process/next_tick.js:63:19)
 log   [06:47:26.895] [warning][reporting] See Chromium's log output at "/home/abhijeetsingh2506/kibana-7.5.1-linu
-x86_64/data/headless_shell-linux/chrome_debug.log"
 log   [06:47:26.897] [warning][reporting] Reporting plugin self-check failed. Please check the Kibana Reporting s
ttings. Error: Could not close browser client handle!
```

Fig.1.20


ssh.cloud.google.com/projects/pure-night-335114/zones/us-central1-c/instances/bigdata-m?authuser=0&hl=en_US&projectNumber=61...

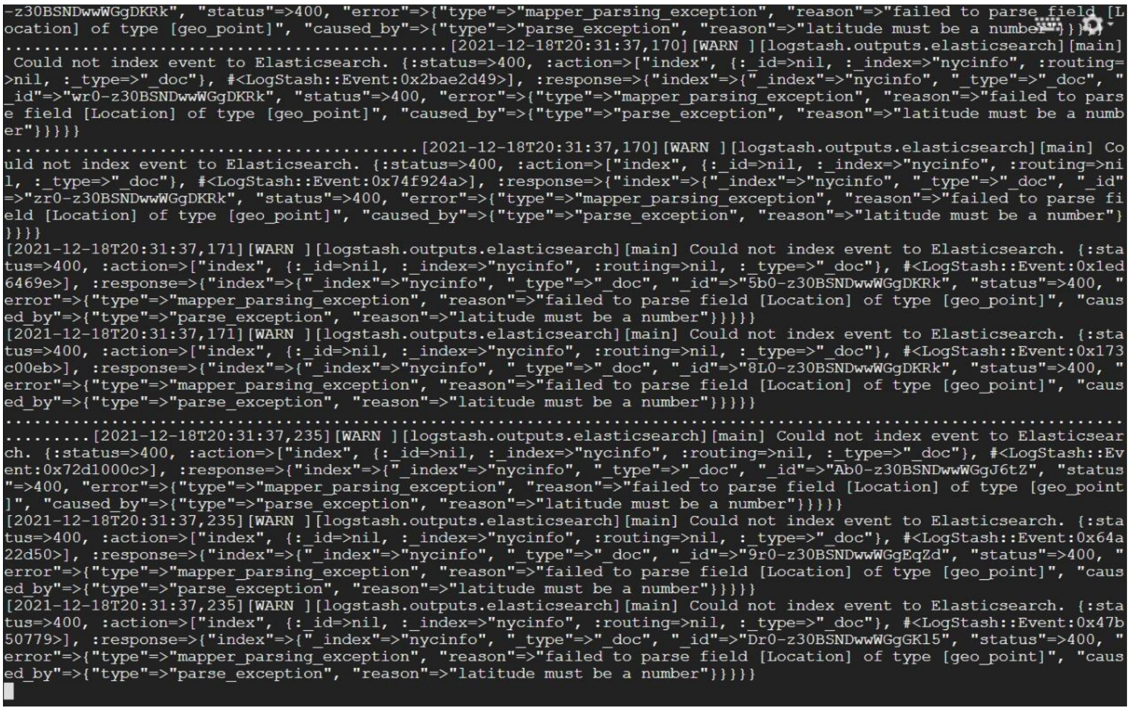
```
5.1.jar:7.5.1]
        ... 40 more
[2021-12-18T05:02:35,927][DEBUG][o.e.a.b.TransportShardBulkAction] [bigdata-m] [nycinfo][0] failed to execute bulk i
tem (index) index {[nycinfo][_doc][G87ry30B5Nm1EcOumrlp], source[{"Cross Street 1":null,"Location Type":null,"Cross
Street 2":null,"Due Date":null,"Address Type":"INTERSECTION","Agency":"DOT","Incident Address":null,"@timestamp":"20
21-12-18T05:02:31.753Z","Agency Name":"Department of Transportation","Longitude":null,"X Coordinate (State Plane)":n
ull,"Community Board":"Unspecified BROOKLYN","Incident Zip":null,"Borough":"BROOKLYN","BBL":null,"Closed Date":"2017
-10-27T08:45:00.000Z","Complaint Type":"Traffic Signal Condition","Taxi Company Borough":null,"Road Ramp":null,"mess
age":"37538531,10/27/2017 07:31:00 AM,10/27/2017 08:45:00 AM,DOT,Department of Transportation,Traffic Signal Conditi
on,Post,,,,,,,FLATLANDS AVE,108 ST E,INTERSECTION,,,N/A,Closed,,Service Request status for this request is available
 on the Department of Transportation?s website. Please click the ?Learn More? link below.,10/27/2017 08:45:00 AM,Uns
pecified BROOKLYN,,BROOKLYN,,,UNKNOWN,Unspecified,BROOKLYN,,,,,,,,,,","Created Date":"2017-10-27T07:31:00.000Z","Y C
oordinate (State Plane)":null,"Open Data Channel Type":"UNKNOWN","Bridge Highway Direction":null,"Park Borough":"BRO
OKLYN","Street Name":null,"Resolution Action Updated Date":"10/27/2017 08:45:00 AM","Taxi Pick Up Location":null,"Br
idge Highway Segment":null,"Resoultion Action Updated Date":"2017-10-27T08:45:00.000Z","Location":"%{Latitude},%{Lon
gitude}","Intersection Street 2":"108 ST E","Intersection Street 1":"FLATLANDS AVE","Vehicle Type":null,"Landmark":n
ull,"Bridge Highway Name":null,"Descriptor":"Post","Status":"Closed","Unique Key":"37538531","@version":"1","Facilit
y Type":"N/A","path":"/home/abhijeetsingh2506/311_service.csv","Resolution Description":"Service Request status for
this request is available on the Department of Transportation?s website. Please click the ?Learn More? link below.",
"host":"bigdata-m","City":null,"Latitude":null,"Park Facility Name":"Unspecified"}]}
org.elasticsearch.index.mapper.MapperParsingException: failed to parse field [Location] of type [geo_point]
        at org.elasticsearch.index.mapper.GeoPointFieldMapper.parse(GeoPointFieldMapper.java:330) ~[elasticsearch-7.
5.1.jar:7.5.1]
        at org.elasticsearch.index.mapper.DocumentParser.parseObjectOrField(DocumentParser.java:488) ~[elasticsearch
-7.5.1.jar:7.5.1]
        at org.elasticsearch.index.mapper.DocumentParser.parseValue(DocumentParser.java:614) ~[elasticsearch-7.5.1.j
ar:7.5.1]
        at org.elasticsearch.index.mapper.DocumentParser.innerParseObject(DocumentParser.java:427) ~[elasticsearch-7
.5.1.jar:7.5.1]
        at org.elasticsearch.index.mapper.DocumentParser.parseObjectOrNested(DocumentParser.java:395) ~[elasticsearc
h-7.5.1.jar:7.5.1]
```

Fig.1.21



```
Linux bigdata-m 5.10.0-0.bpo.9-amd64 #1 SMP Debian 5.10.70-1~bpo10+1 (2021-10-10
) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Dec 18 00:51:05 2021 from 35.235.240.5
abhijeetsingh2506@bigdata-m:~$ cd logstash-7.5.1/
abhijeetsingh2506@bigdata-m:~/logstash-7.5.1$ ls
CONTRIBUTORS  Gemfile.lock  NOTICE.TXT  config  lib             logstash-core-plugin-api  nyc311.conf  vendor
Gemfile       LICENSE.txt   bin         data    logstash-core  modules                   tools        x-pack
abhijeetsingh2506@bigdata-m:~/logstash-7.5.1$ bin/logstash
^Cabhijeetsingh2506@bigdata-m:~/logstash-7.5.1$ mv nyc311.conf nyc311.config
abhijeetsingh2506@bigdata-m:~/logstash-7.5.1$ bin/logstash -f nyc311.config
Thread.exclusive is deprecated, use Thread::Mutex
Sending Logstash logs to /home/abhijeetsingh2506/logstash-7.5.1/logs which is now configured via log4j2.properties
[2021-12-18T01:06:32,464][INFO ][logstash.setting.writabledirectory] Creating directory {:setting=>"path.queue", :pa
th=>"/home/abhijeetsingh2506/logstash-7.5.1/data/queue"}
[2021-12-18T01:06:32,591][INFO ][logstash.setting.writabledirectory] Creating directory {:setting=>"path.dead_letter
_queue", :path=>"/home/abhijeetsingh2506/logstash-7.5.1/data/dead_letter_queue"}
[2021-12-18T01:06:33,047][WARN ][logstash.config.source.multilocal] Ignoring the 'pipelines.yml' file because module
s or command line options are specified
[2021-12-18T01:06:33,058][INFO ][logstash.runner          ] Starting Logstash {"logstash.version"=>"7.5.1"}
[2021-12-18T01:06:33,088][INFO ][logstash.agent           ] No persistent UUID file found. Generating new UUID {:uui
d=>"798dd62a-00d6-4ff7-a9ba-7144f74c6e1a", :path=>"/home/abhijeetsingh2506/logstash-7.5.1/data/uuid"}
[2021-12-18T01:06:35,907][INFO ][org.reflections.Reflections] Reflections took 142 ms to scan 1 urls, producing 20 k
eys and 40 values
```

Fig.1.22



```
-z30BSNDwwWGgDKRk", "status"=>400, "error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [L
ocation] of type [geo_point]", "caused_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
.........................................[2021-12-18T20:31:37,170][WARN ][logstash.outputs.elasticsearch][main]
 Could not index event to Elasticsearch. {:status=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=
>nil, :_type=>"_doc"}, #<LogStash::Event:0x2bae2d49>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "
_id"=>"wr0-z30BSNDwwWGgDKRk", "status"=>400, "error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to pars
e field [Location] of type [geo_point]", "caused_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a numb
er"}}}}}
......................................[2021-12-18T20:31:37,170][WARN ][logstash.outputs.elasticsearch][main] Co
uld not index event to Elasticsearch. {:status=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>ni
l, :_type=>"_doc"}, #<LogStash::Event:0x74f924a>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"
=>"zr0-z30BSNDwwWGgDKRk", "status"=>400, "error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse fi
eld [Location] of type [geo_point]", "caused_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}
}}}}
[2021-12-18T20:31:37,171][WARN ][logstash.outputs.elasticsearch][main] Could not index event to Elasticsearch. {:sta
tus=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>nil, :_type=>"_doc"}, #<LogStash::Event:0x1ed
6469e>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"=>"5b0-z30BSNDwwWGgDKRk", "status"=>400, "
error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [Location] of type [geo_point]", "caus
ed_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
[2021-12-18T20:31:37,171][WARN ][logstash.outputs.elasticsearch][main] Could not index event to Elasticsearch. {:sta
tus=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>nil, :_type=>"_doc"}, #<LogStash::Event:0x173
c00eb>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"=>"8L0-z30BSNDwwWGgDKRk", "status"=>400, "
error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [Location] of type [geo_point]", "caus
ed_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
.......................................................................................................
........[2021-12-18T20:31:37,235][WARN ][logstash.outputs.elasticsearch][main] Could not index event to Elasticsear
ch. {:status=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>nil, :_type=>"_doc"}, #<LogStash::Ev
ent:0x72d1000c>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"=>"Ab0-z30BSNDwwWGgJ6tZ", "status
"=>400, "error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [Location] of type [geo_point
]", "caused_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
[2021-12-18T20:31:37,235][WARN ][logstash.outputs.elasticsearch][main] Could not index event to Elasticsearch. {:sta
tus=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>nil, :_type=>"_doc"}, #<LogStash::Event:0x64a
22d50>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"=>"9r0-z30BSNDwwWGgEqZd", "status"=>400, "
error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [Location] of type [geo_point]", "caus
ed_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
[2021-12-18T20:31:37,235][WARN ][logstash.outputs.elasticsearch][main] Could not index event to Elasticsearch. {:sta
tus=>400, :action=>["index", {:_id=>nil, :_index=>"nycinfo", :routing=>nil, :_type=>"_doc"}, #<LogStash::Event:0x47b
50779>], :response=>{"index"=>{"_index"=>"nycinfo", "_type"=>"_doc", "_id"=>"Dr0-z30BSNDwwWGgGKl5", "status"=>400, "
error"=>{"type"=>"mapper_parsing_exception", "reason"=>"failed to parse field [Location] of type [geo_point]", "caus
ed_by"=>{"type"=>"parse_exception", "reason"=>"latitude must be a number"}}}}}
```

Fig.1.23

PART 3
CONCLUSION

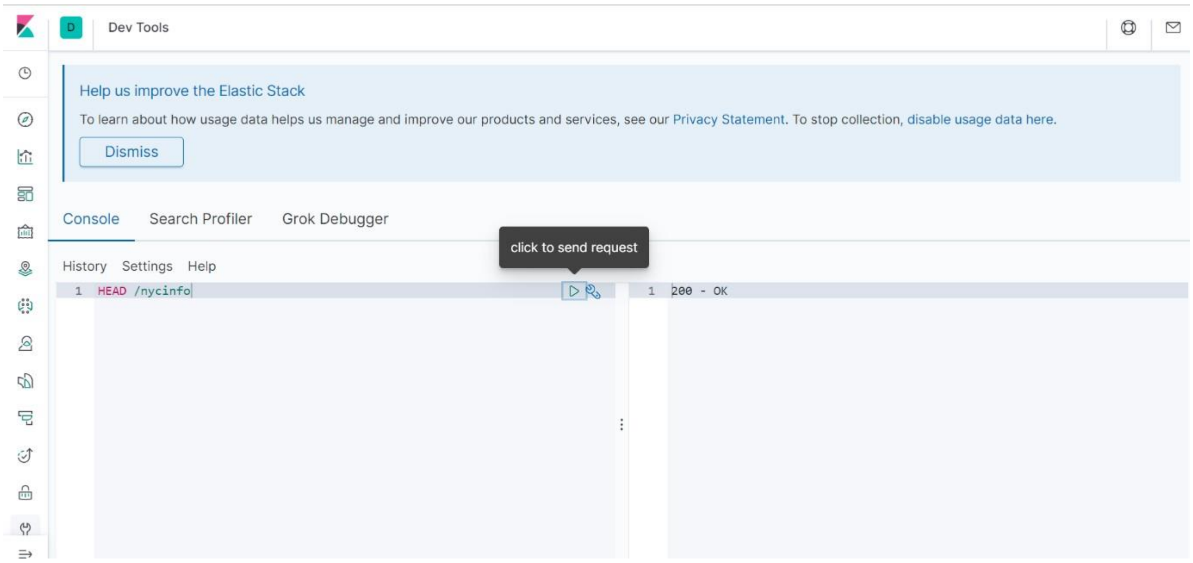
- Brooklyn is the city with the highest number of service calls.
- Loud Music/Party is the reason for the most number of complaints.
- Broadway street has the highest number of complaints/service calls.
- Most complaints are resolved in the early stages thus receiving a 'Closed' status.
- The commonly used open data channel is phone before 2011.
- Familiarity with ELK and Kibana Tools for huge dataset.
- Analysis of data by gaining insight into KQL(Kibana Query Language).

REFERENCES

https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

APPENDICES

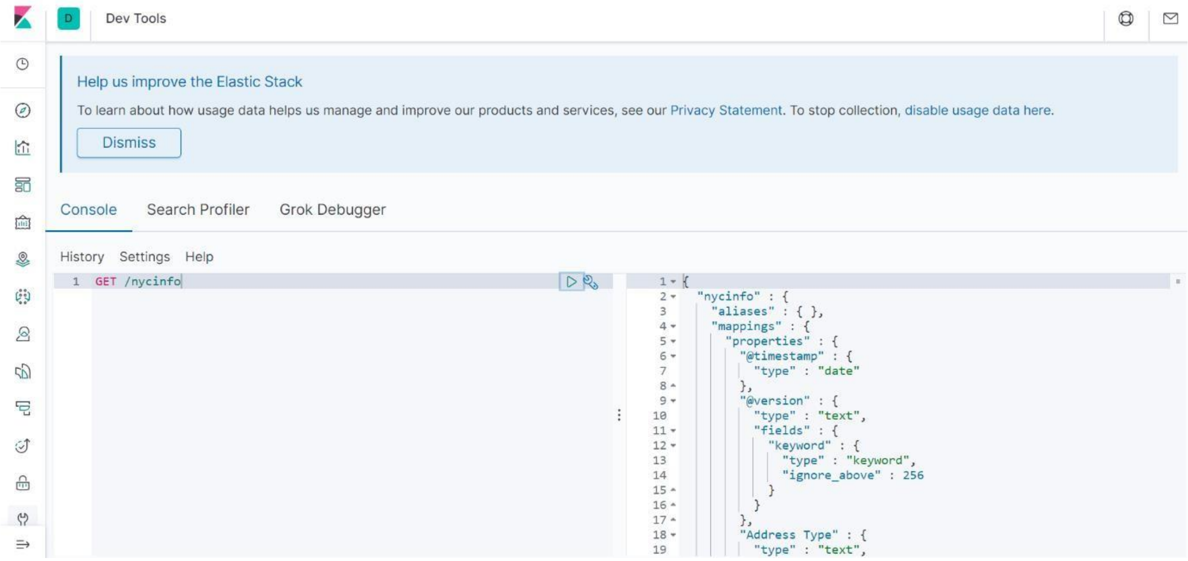
To find out whether the index exists in Kibana


Dev Tools
Help us improve the Elastic Stack
To learn about how usage data helps us manage and improve our products and services, see our Privacy Statement. To stop collection, disable usage data here.
Dismiss
Console
Search Profiler
Grok Debugger
History
Settings
Help
click to send request

```
HEAD /nycinfo
```

```
200 - OK
```

To check the structure of the index



Help us improve the Elastic Stack

To learn about how usage data helps us manage and improve our products and services, see our Privacy Statement. To stop collection, disable usage data here.

```
GET /nycinfo
```

```
{
  "nycinfo" : {
    "aliases" : { },
    "mappings" : {
      "properties" : {
        "@timestamp" : {
          "type" : "date"
        },
        "@version" : {
          "type" : "text",
          "fields" : {
            "keyword" : {
              "type" : "keyword",
              "ignore_above" : 256
            }
          }
        },
        "Address Type" : {
          "type" : "text",
```