BDAT1002 Data Systems Architecture

Final Project Report

Group 10

Members:

Aanandita Chavan

Amjad El Baba

Prateek Singh

Contents

Introduction

Cities have benefited greatly from the merger of technology with a three-number hotline. The result: better service delivery for citizens and more data about how to run local government.

311 is a non-emergency phone number that people can call in many cities to find information about services, make complaints, or report problems like graffiti or road damage. Even in cities where a different phone number is used, 311 is the generally recognized moniker for non-emergency phone systems.

311 was first launched in 2003 and since its inception, 311 has evolved with technological advances into a multi-channel service that connects citizen with government, while also providing a wealth of data that improves how cities are run.

People can access 311 over the phone, using Facebook or Twitter, by sending a text message to 311-NYC or through a 311 smartphone app for iPhone and Android.

Basically, 311 is type of a citizens' hotline number. It is a simple way of asking questions and reporting issues without going through complex municipal guidelines or typing emergency issues.

When 311 is called, the person calling is connected to a trained representative who will help out to solve the problem. The representatives can forward calls or complaints to the local authorities or local utilities to get the problem fixed or answer questions.

While 311 isn't available nationally, many metropolitan cities have the 311 service in place. Cities like New York, San Francisco and Los Angeles have mobile apps for 311 services to make reporting issues or asking questions easier.

311 is designed to take calls which are made to 9-1-1 in case of non emergencies. However, even non-emergencies can escalate to emergencies quickly if required actions are not taken immediately.

Here's a list of a few of the most common reasons to call 311:

- Sidewalk and road repairs
- Noise complaints
- Abandoned vehicles
- Large debris blocking roadways
- Garbage bin replacements
- Dangerous animal complaints
- Damaged traffic signals or road signs
- Graffiti removal
- Leaking hydrant or sewer problems

911 is the number to call in case of emergency. 311 is the number to call to obtain information and access to non-emergency City government services. 911 calls are for life threatening emergencies. Reducing the call traffic towards 911 and directing it toward 311 will help free up operators in the 911 line to tend towards serious emergencies.

About the Dataset

“311 service requests from 2010 to present” is a dataset provided by NYC Open Data. NYC Open Data is a website that provides open data to everyone. Its mission is to engage its audience in information that is produced and used by the city government. This data is used by various organizations and users to learn and perform data analytics.

NYC 311’s mission is to provide the public with quick and easy access to the government services and information and provide the best customer service. Everyday, NYC 311 services receives thousands of requests related to non emergency services like noise complaints, plumbing issues or abandoned or illegally parked cars. These requests are received by NYC311 and forwarded to responsible or relevant authorities. The authorities the responds to the request, addresses it and then closes it.

“311 service requests from 2010 to present” is acquired from NYC Open Data Website . It is basically a dataset which consists of various parameters that describe the type of 311 calls made to various cities in New York. This data is updated daily and it consists data from the year 2010 till the present date. This is a public dataset.

The dataset consists of 30.8 million rows and 41 columns. Each row is a 311-service request. Some of the columns of the dataset are the complaint type, the city, the created date of the service request, the closed date of the service request etc.

Below is the description of all the columns in the dataset:

Column Name	Description
Unique Key	Unique identifier of a Service Request (SR) in the open data set
Created Date	Date SR was created
Closed Date	Date SR was closed by responding agency
Agency	Acronym of responding City Government Agency
Agency Name	Full Agency name of responding City Government Agency
Complaint Type	This is the fist level of a hierarchy identifying the topic of the incident or condition. Complaint Type may have a corresponding Descriptor (below) or may stand alone.
Descriptor	This is associated to the Complaint Type, and provides further detail on the incident or condition. Descriptor values are dependent on the Complaint Type, and are not always required in SR.
Status	Status of SR submitted
Due Date	Date when responding agency is expected to update the SR. This is based on the Complaint Type and internal Service Level Agreements (SLAs).
Resolution Action Updated Date	Date when responding agency last updated the SR.
Resolution Description	Describes the last action taken on the SR by the responding agency. May describe next or future steps.
Location Type	Describes the type of location used in the address information
Incident Zip	Incident location zip code, provided by geo validation.
Incident Address	House number of incident address provided by submitter.
Street Name	Street name of incident address provided by the submitter
Cross Street 1	First Cross street based on the geo validated incident location
Cross Street 2	Second Cross Street based on the geo validated incident location
Intersection Street 1	First intersecting street based on geo validated incident location
Intersection Street 2	Second intersecting street based on geo validated incident location
Address Type	Type of incident location information available.
City	City of the incident location provided by geovalidation.
Landmark	If the incident location is identified as a Landmark the name of the landmark will display here
Facility Type	If available, this field describes the type of city facility associated to the SR
Community Board	Provided by geovalidation.
BBL	Borough Block and Lot, provided by geovalidation. Parcel number to identify the location of location of buildings and properties in NYC.
Borough	Provided by the submitter and confirmed by geovalidation.
X Coordinate (State Plane)	Geo validated, X coordinate of the incident location.
Y Coordinate (State Plane)	Geo validated, Y coordinate of the incident location.
Open_Data_Channel_Type	Indicates how the SR was submitted to 311. i.e. By Phone, Online, Mobile, Other or Unknown.
Latitude	Geo based Lat of the incident location
Longitude	Geo based Long of the incident location
Location	Combination of the geo based lat & long of the incident location
Park Facility Name	If the incident location is a Parks Dept facility, the Name of the facility will appear here
Park Borough	The borough of incident if it is a Parks Dept facility
Vehicle Type	If the incident is a taxi, this field describes the type of TLC vehicle.
Taxi Company Borough	If the incident is identified as a taxi, this field will display the borough of the taxi company.
Taxi Pick Up Location	If the incident is identified as a taxi, this field displays the taxi pick up location
Bridge Highway Name	If the incident is identified as a Bridge/Highway, the name will be displayed here.
Bridge Highway Direction	If the incident is identified as a Bridge/Highway, the direction where the issue took place would be displayed here.
Road Ramp	If the incident location was Bridge/Highway this column differentiates if the issue was on the Road or the Ramp.
Bridge Highway Segment	Additional information on the section of the Bridge/Highway were the incident took place.

Tools

Big data is high volume of data from multiple sources which cannot be managed by traditional data processing software. However, this massive amount of data can be used to answer a lot complex questions and business problems. There are three v's in big data and they are volume, velocity and variety.

Big data is high volume of unstructured data. For some organizations, the data might be in terabytes and for some it may be in petabytes. Velocity is the fast rate at which data is coming in and acted upon. Some products act in real time or near real time and they require real time actions. Variety refers to the many types of data that is available. Data can be in the form of text, numbers, images or videos. Traditional data was structured and could be easily stored in a dataset. But the big data these days is unstructured or semi structured and it needs metadata to be worked upon.

To handle such big data any power tools have been launched in the market and new products are getting launched very frequently, more powerful than the other. There are many free sources as well as paid services available in the market to use big data and derive meaningful insights from it.

Some of the tools that are being used in this project are Google Cloud Platform, ELK stack, log stash configuration file and Kibana. These tools are used to work with big data set that will fail on the traditional systems. These tools are used to store, analyse and create insights from the dataset. These are some of the best tools available in the market.

Google Cloud Platform (GCP) is a suite of cloud computing services offered by Google. It runs on the same infrastructure as other google services such as google search or google mail/ gmail. GCP provides services for computing, data storage,

data analytics and machine learning apart from some management tools. Registration requires login with the credit card details. However, the user is not charged immediately. User gets 300 credits and once the credits are exhausted then the user will start getting charged for the GCP services.

Google Cloud Platform provides infrastructure as a service, platform as a service, and serverless computing environments. Google Cloud Platform is a part of Google Cloud, which includes the Google Cloud Platform public cloud infrastructure, as well as Google Workspace (G Suite), enterprise versions of Android and Chrome OS, and application programming interfaces (APIs) for machine learning and enterprise mapping services.

GCP allows the user to create a cluster of their desired size. GCP contains a tool called Dataproc. Dataproc is a managed Spark and Hadoop service that lets you take advantage of open-source data tools for batch processing, querying, streaming, and machine learning. Dataproc automation helps you create clusters quickly, manage them easily, and save money by turning clusters off when you don't need them.

ELK stack is one of the technologies that is being used in the project. ELK is the acronym for three open-source projects: Elasticsearch, Logstash and Kibana. Elasticsearch is a search and analytical engine. Logstash is a configuration or processing file which ingests data from multiple sources, simultaneously transforming it and storing it to a storing system like Elasticsearch. Kibana is the tool which is used to visualize the data with charts and graphs in Elasticsearch. Together these three tools makeup the ELK stack.

Elastic search is an open-source search and analytical search engine. Elasticsearch allows to store, search, and analyze huge volumes of data quickly and in near real-time and give back answers in milliseconds. It's able to achieve fast search responses

because instead of searching the text directly, it searches an index. It uses a structure based on documents instead of tables and schemas and comes with extensive REST APIs for storing and searching the data. At its core, Elasticsearch is a server that can process JSON requests and gives back JSON data. Some of the concepts of Elastic search are documents, indexes and inverted indexes.

Documents are the basic unit of information that can be indexed in Elasticsearch expressed in JSON, which is the global internet data interchange format. An index is a collection of documents that have similar characteristics. An index is the highest-level entity that you can query against in Elasticsearch. You can think of the index as being similar to a database in a relational database schema.

Kibana is a tool in Elastic search which is used for data visualization and management. Kibana provides real time line graphs, histograms, pie charts and maps. It supports in creating visuals for the Elasticsearch data and navigate through the Elastic Stack. However, one drawback of Kibana is that every visualization can only work against one a single index pattern. Thus, if there are indices with different data then user needs to create separate visualizations for each index.

Logstash is used to aggregate and process data and send it to Elasticsearch. It is an open-source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to collect. It also transforms and prepares data regardless of format by identifying named fields to build structure, and transform them to converge on a common format.

Analysis

After loading the dataset in Elasticsearch successfully, the next step is to use Kibana and make visualizations in Elasticsearch. This analysis is performed by a team of three data analyst. Each member has created two visuals each.

The goal of this project is to create visualizations in Kibana for better analysis purpose. It is difficult to draw insights from a huge dataset just by looking at it. Thus, visualizations are an excellent way of display large volume of data consolidated together in different types of graphs. Also, visuals are an easy way to explain things to the non technical audience. Colourful visualization is also more appealing to the eyes than tables.

Below are the various visuals created by each data analyst and their descriptions.

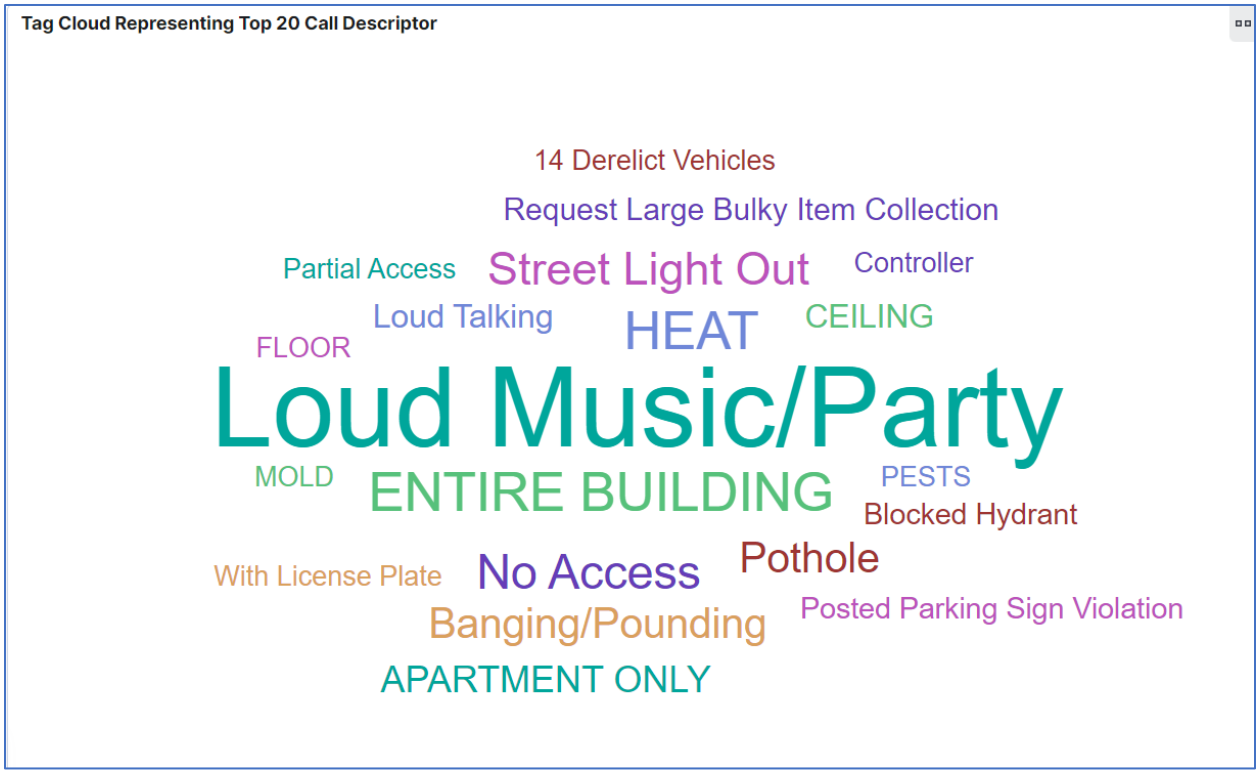
The following two visuals are created by **Amjad El Baba:**

1. Tag cloud representing the top 20 call descriptors.

One of the best visuals used to describe categorical data is Tag Cloud, because it captures relative sizes of data categories and it's a visual representation of the most popular words.

Obviously, the Loud Music/Party is the Top1 descriptor, and most of the calls was due to that. This is reasonable since New York is one of the highest growing population cities & it's highly probable that a lot of people would enjoy partying which will cause a headache for other people or neighbors.

Then the list continues as Entire Building, Heat, Street Light On, No Access, and so on. Please refer the tag cloud figure below.



2. Horizontal bar chart of the top 5 boroughs with 311 service call requests.

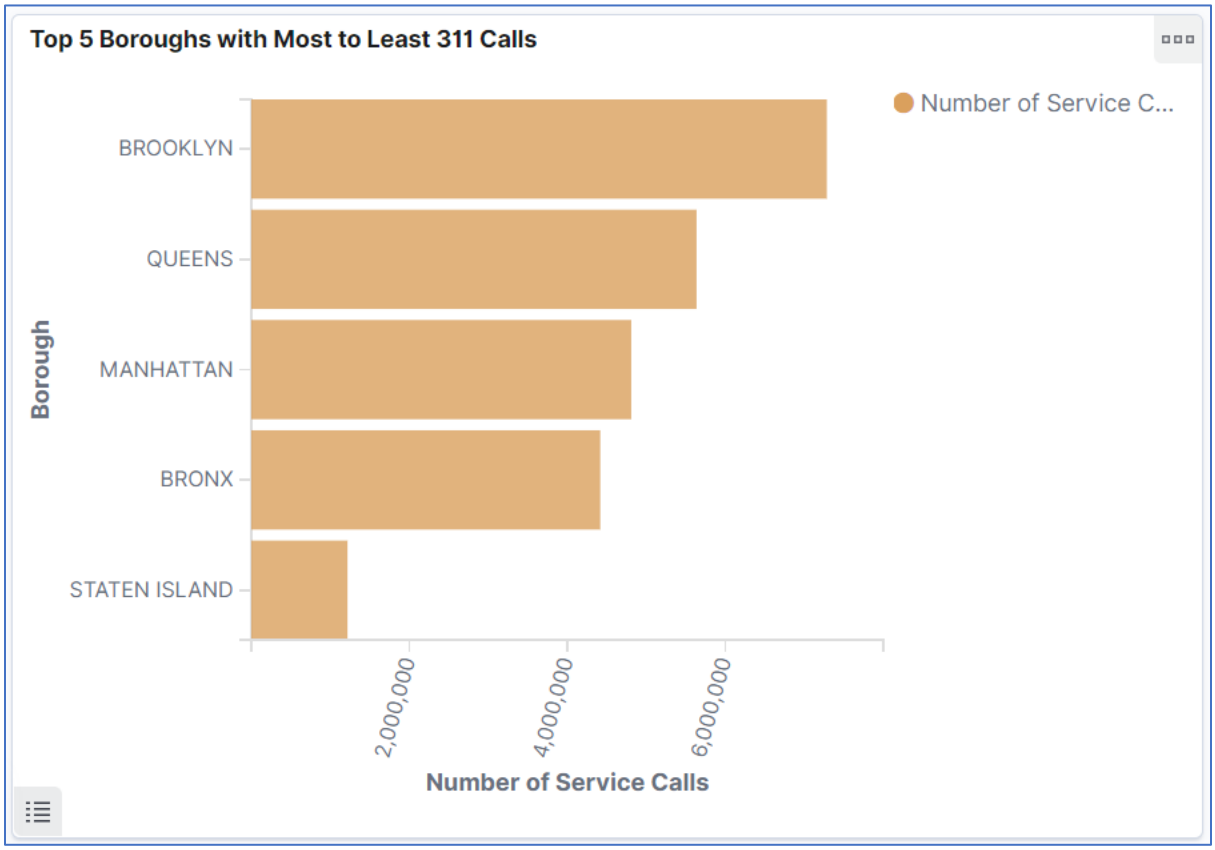
The following Bar Chart demonstrates the Top 5 boroughs calling 311, the boroughs are ordered in a descending order.

They are listed as follow:

1. Brooklyn
2. Queens
3. Manhattan
4. Bronx
5. Staten Island

There are three reasons of the large number of calls in these boroughs, either the large population number, or the people living there are facing some issues more than

other people living in the other boroughs, or both reasons. Please refer to the bar chart below.

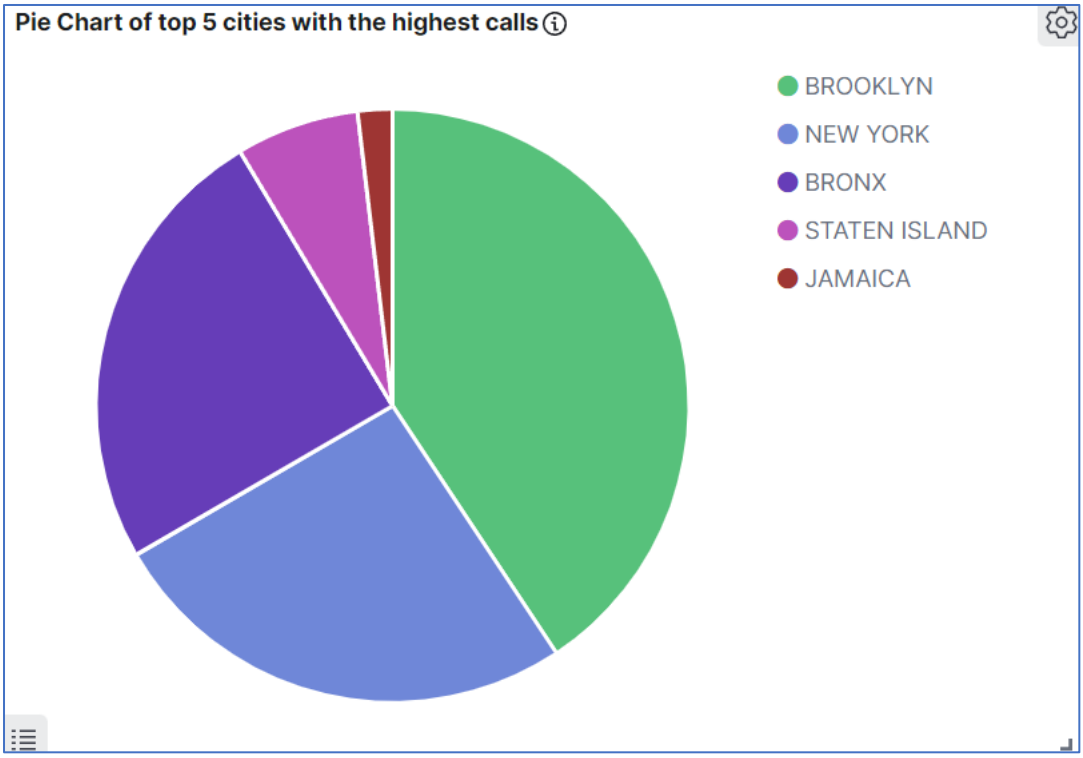


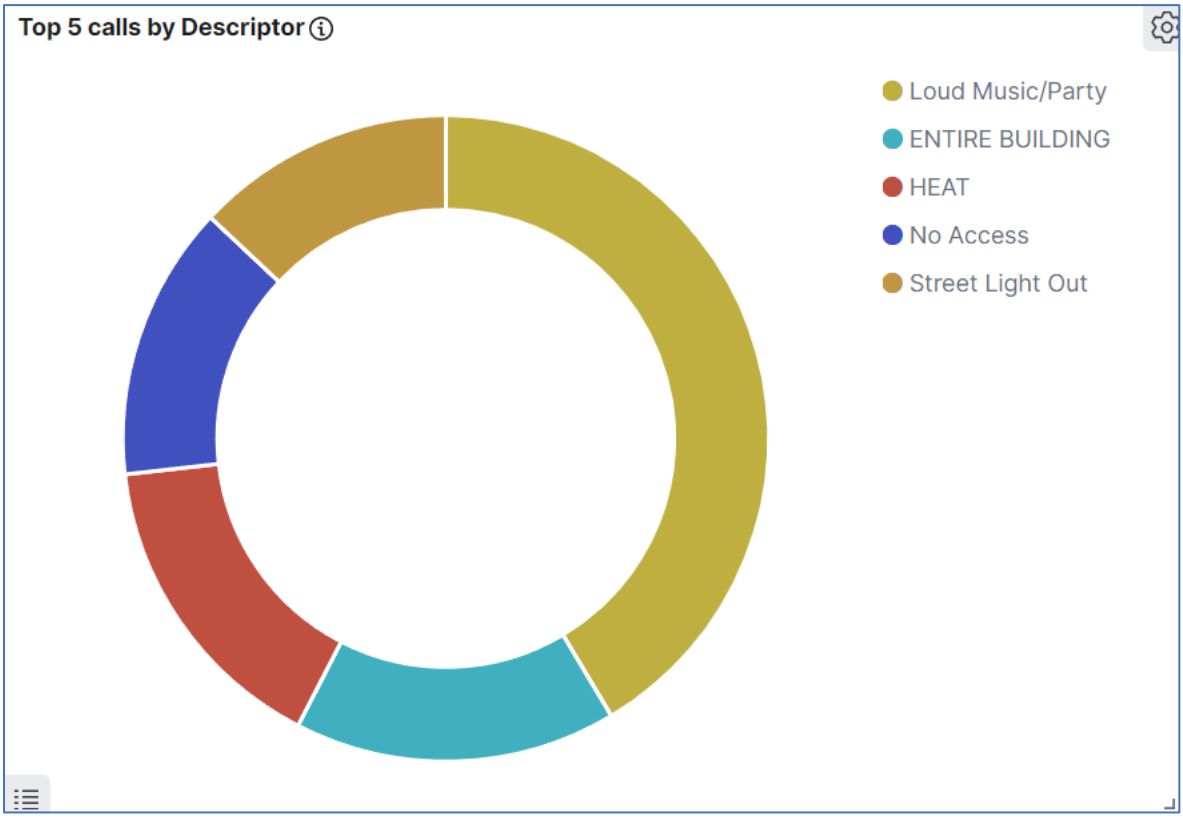
The following analysis is performed by **Aanandita Chavan,**

1. A pie chart showing the top 5 cities with the highest calls alongside the top five calls (Descriptor) in each city

Pie charts shows the parts to whole relationship. For example, it can show the percentage of each category in a dataset. It is useful in seeing which category has the highest percentage of value.

In this pie chart, the top 5 cities with the highest calls are depicted along with the top 5 descriptor calls in each city. Please refer the figure below.

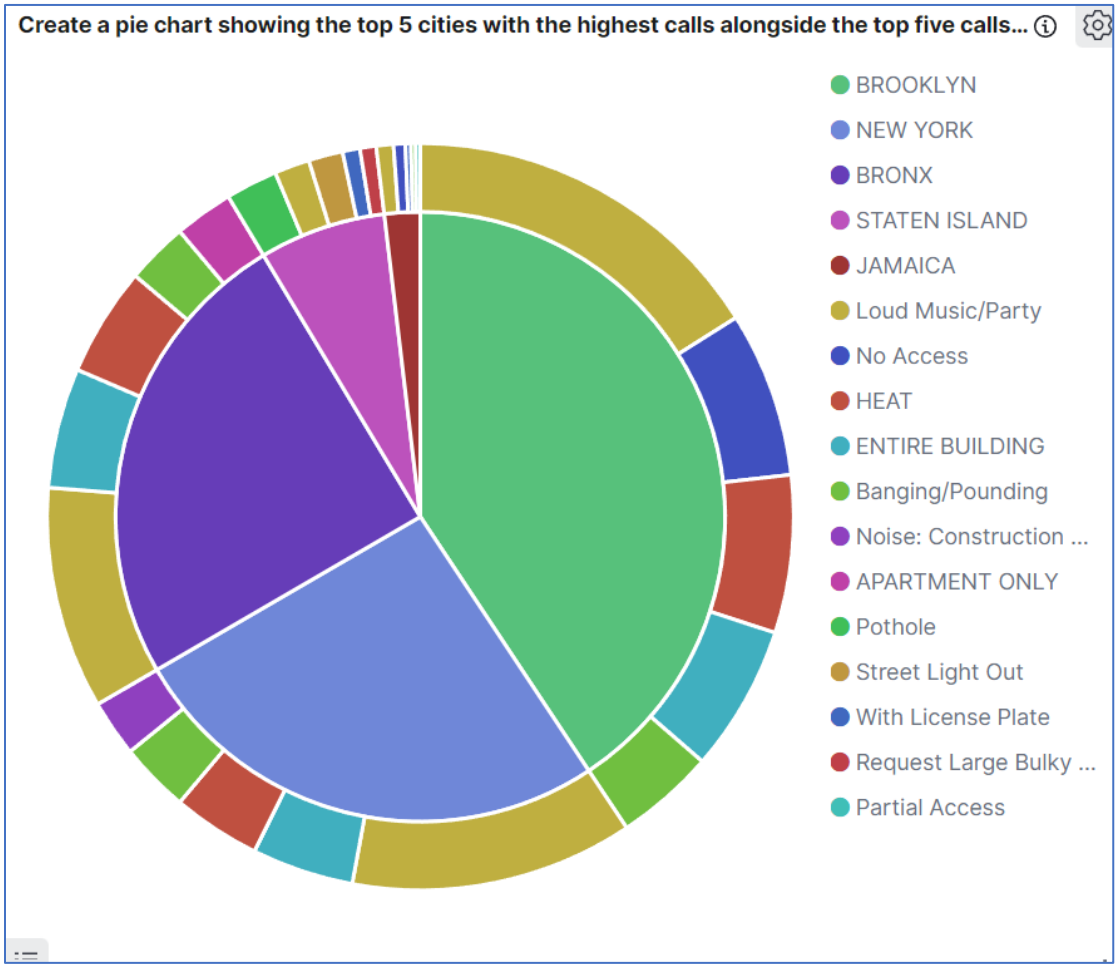

Pie Chart of top 5 cities with the highest calls
BROOKLYN
NEW YORK
BRONX
STATEN ISLAND
JAMAICA


Top 5 calls by Descriptor
Loud Music/Party
ENTIRE BUILDING
HEAT
No Access
Street Light Out

The first pie chart shows the top 5 cities with the highest calls. It is observed that Brooklyn has the greatest number of calls, followed by New York City and then Bronx. Brooklyn and New York city are metro cities and are always busy. Thus, it is justified that these cities get the most 311 calls.

The second pie chart shows the top 5 calls by their descriptor. Descriptor is a column in the data set which describes the call request. From the pie chart it can be seen that Loud Music/Party is the most popular complaint.

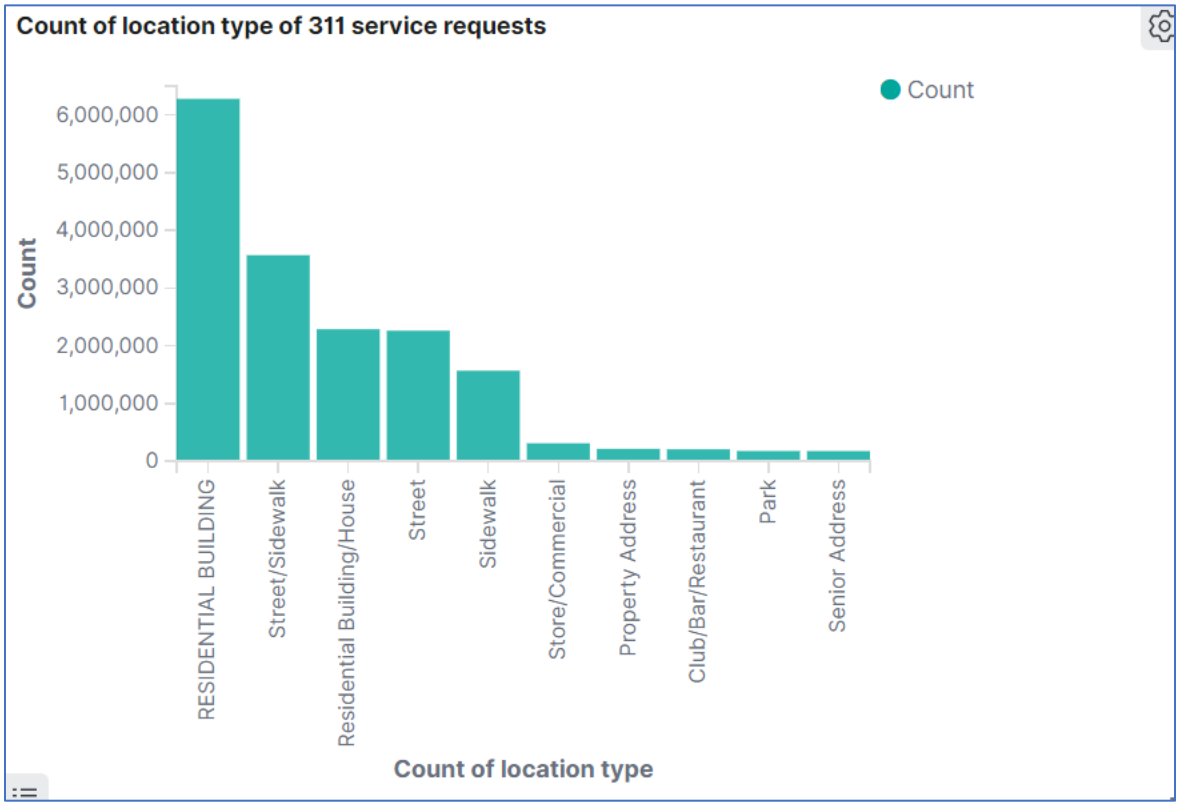
Below pie chart shows the top 5 cities with the most calls and the top 5 calls in each city. Brooklyn is the city with the greatest number of calls and in Brooklyn Loud Music/ Party is descriptor with the greatest number of calls.



2. Bar chart of count of location type of the 311-service request.

Bar chart are useful in comparing different categorical or discrete variables as long as there are not too many categories.

Below bar chart shows the count of the location type from which the 311-service request was made. Residential buildings are the location type with the greatest number of calls and senior address has the lowest number of 311 calls.



The following analysis is performed by **Prateek Singh,**

1. Table of top 10 cities with the highest number of calls and top 10 calls by descriptor

Top 10 cities with the greatest number of 311 calls

Top 10 Cities with the greatest number of 311 call requests	Count
BROOKLYN	7,289,314
NEW YORK	4,649,918
BRONX	4,439,012
STATEN ISLAND	1,202,716
JAMAICA	338,565
FLUSHING	258,404
Jamaica	249,527
ASTORIA	237,893
RIDGEWOOD	184,776
Flushing	175,148

Top 10 311 service call requests

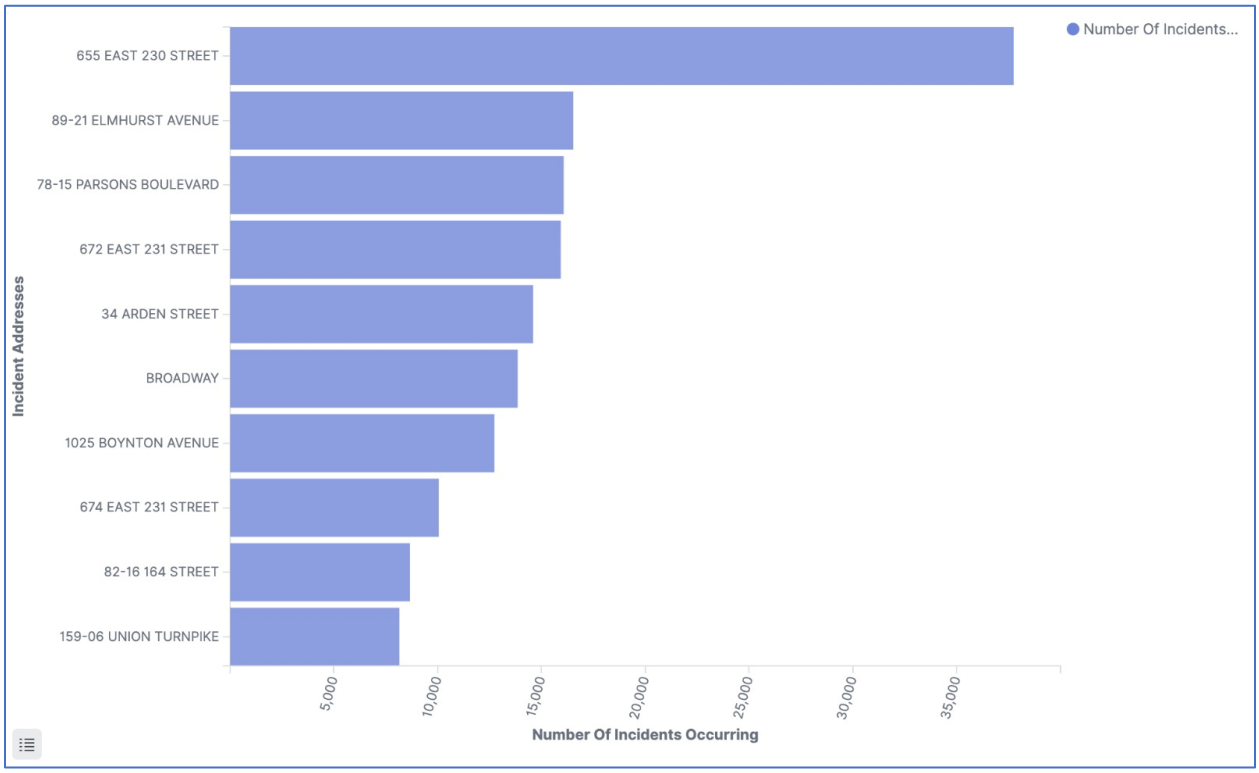
Top 10 311 calls by their descriptors	Count
Loud Music/Party	2,309,476
ENTIRE BUILDING	898,326
HEAT	871,935
No Access	767,415
Street Light Out	724,710
Pothole	613,932
Banging/Pounding	602,384
APARTMENT ONLY	479,866
CEILING	355,652
Loud Talking	355,531

Table is useful when user wants to see the exact value of each item. The above table display the top 10 cities and top 10 calls in the 311 service request dataset. The first table show the top 10 cities with the highest 311 service request calls

along with the count of each city occurring the dataset. The city with the greatest number of calls is Brooklyn. Brooklyn got more than 7 million 311 calls. Following Brooklyn are New York and Bronx and both have over 4 million calls.

The second table shows the top 10 311 calls by their descriptor along with the count of each call. The top most complain is loud music/party and it was reported more than 2 million times. Following that is Entire Building and Heat calls.

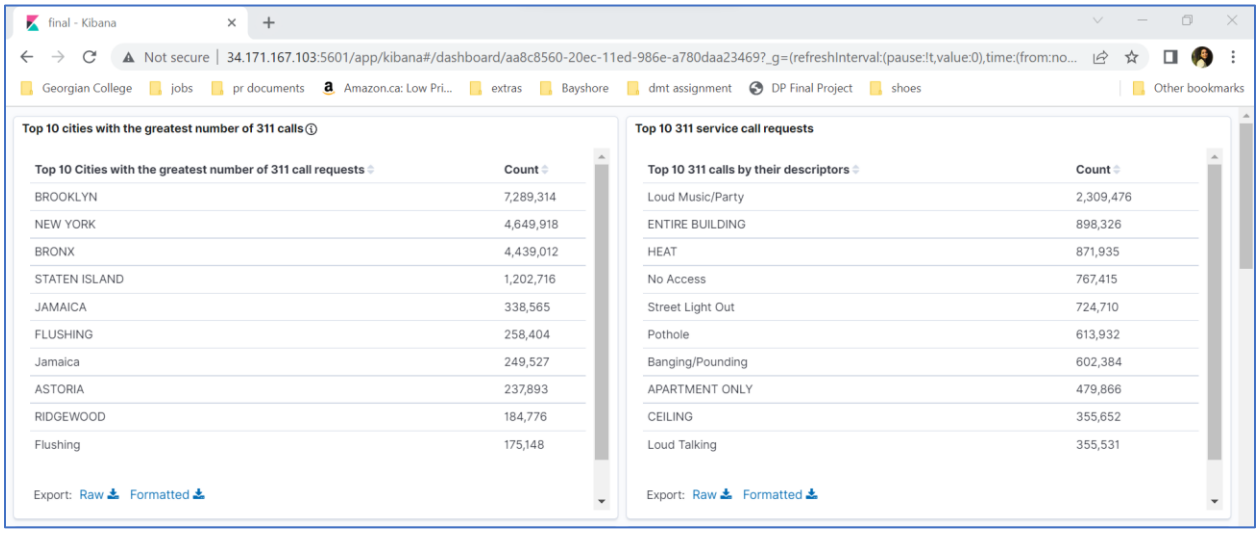
2. Most common addresses with greatest number of service calls

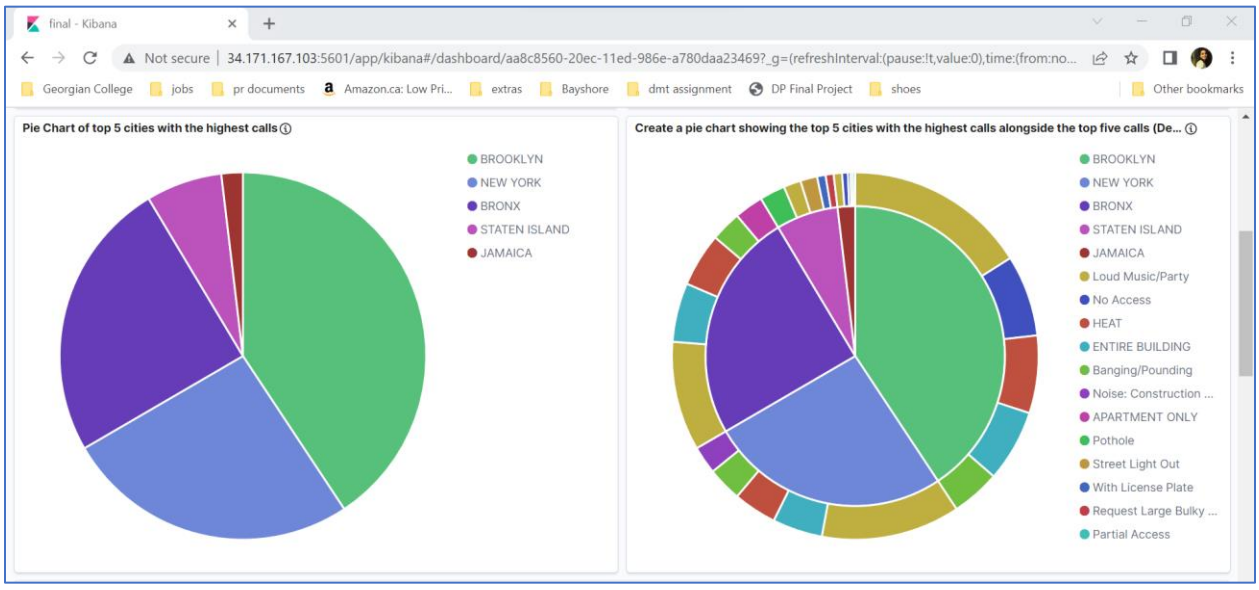


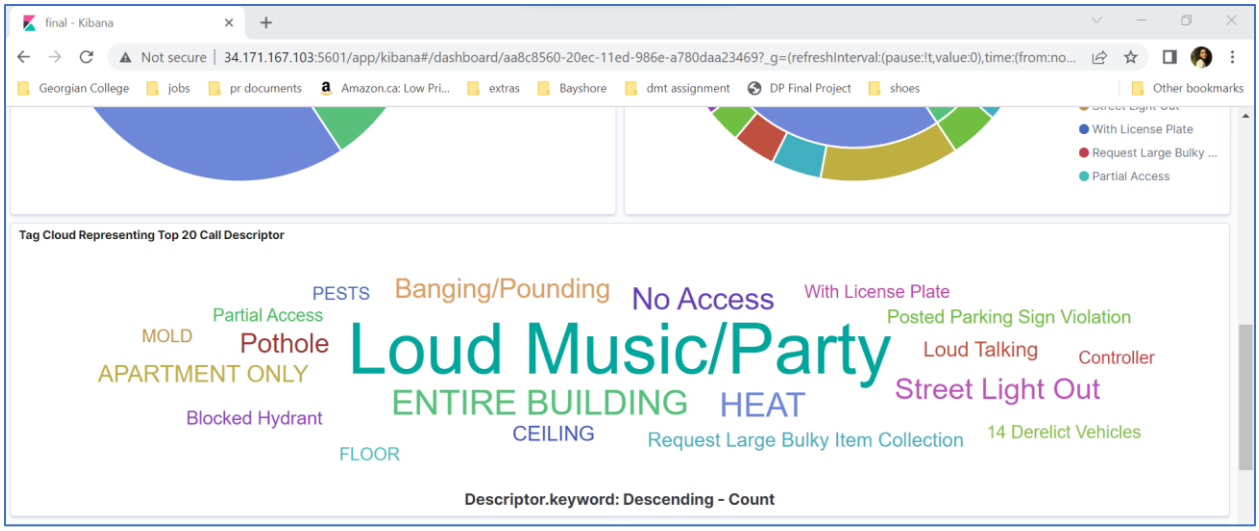
The above horizontal bar chart shows the address from which the greatest number of service calls were made. “655 East 230 Street” got the greatest number of service calls. Whereas, 159-06 Union Turnpike address got the least number of 311 service calls.

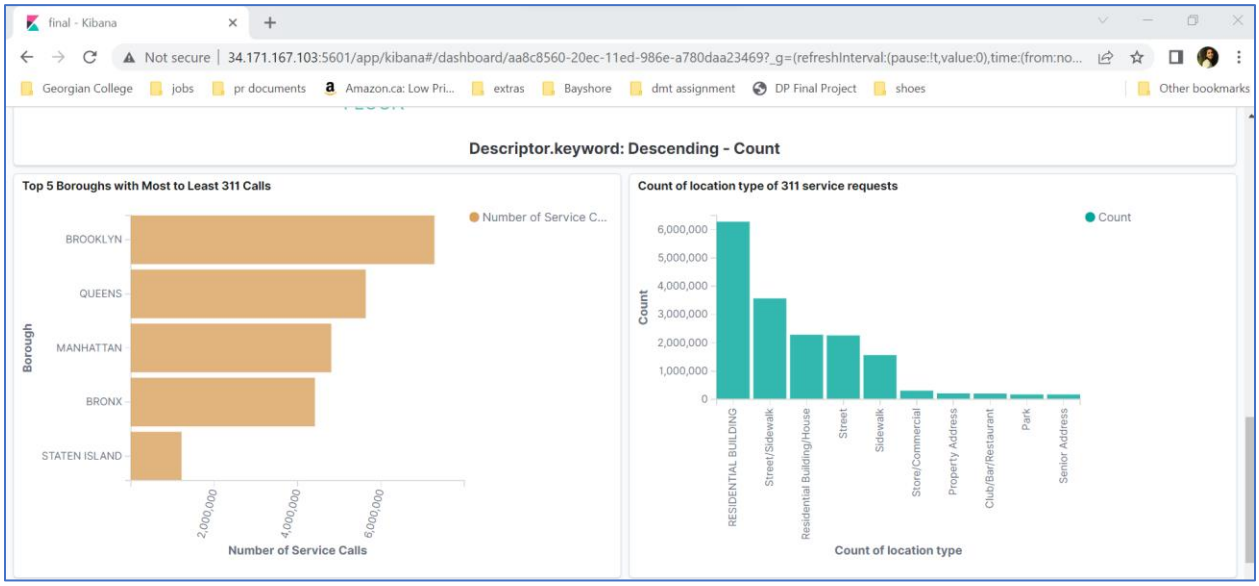
Dashboard

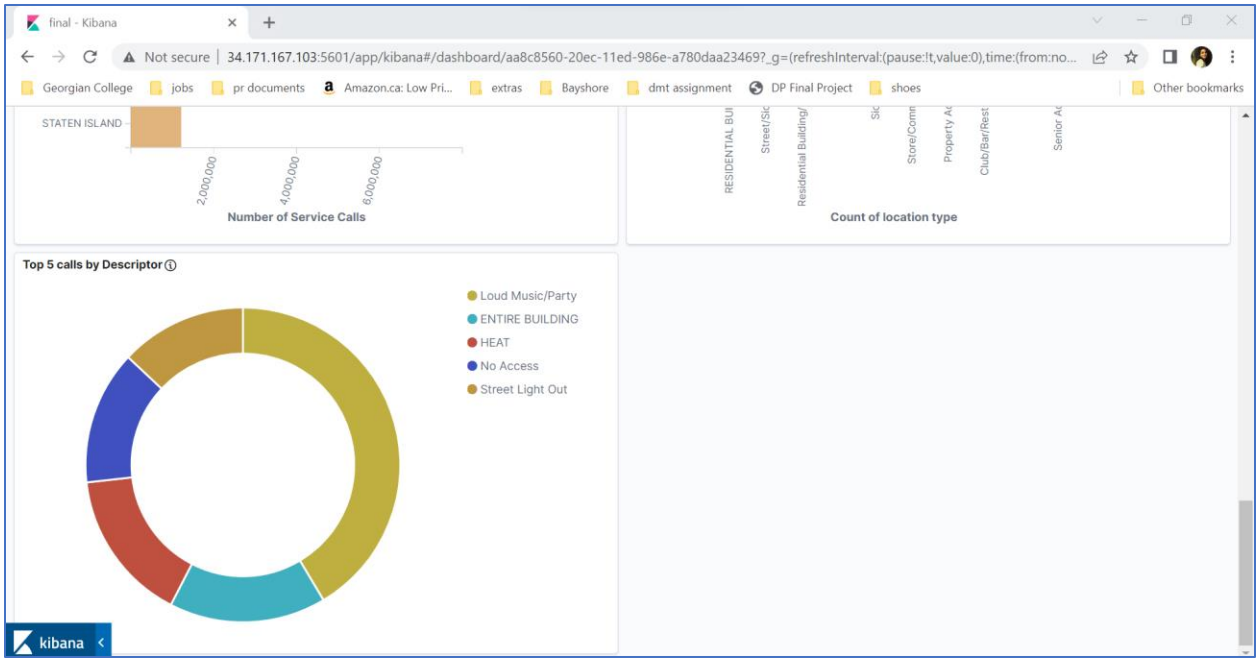
Once the visuals are made in Kibana, they can be added to the Elasticsearch dashboard. Below is the image on the dashboard created for this project.






final - Kibana
Not secure | 34.171.167.103:5601/app/kibana#/dashboard/aa8c8560-20ec-11ed-986e-a780daa23469?_g=(refreshInterval:(pause:!t,value:0),time:(from:no...
Georgian College
jobs
pr documents
Amazon.ca: Low Pri...
extras
Bayshore
dmt assignment
DP Final Project
shoes
Other bookmarks
With License Plate
Request Large Bulky ...
Partial Access
Tag Cloud Representing Top 20 Call Descriptor
PESTS
Banging/Pounding
No Access
With License Plate
Partial Access
Posted Parking Sign Violation
MOLD
Pothole
Loud Music/Party
Loud Talking
Controller
APARTMENT ONLY
Street Light Out
ENTIRE BUILDING
HEAT
Blocked Hydrant
CEILING
Request Large Bulky Item Collection
14 Derelict Vehicles
FLOOR
Descriptor.keyword: Descending - Count

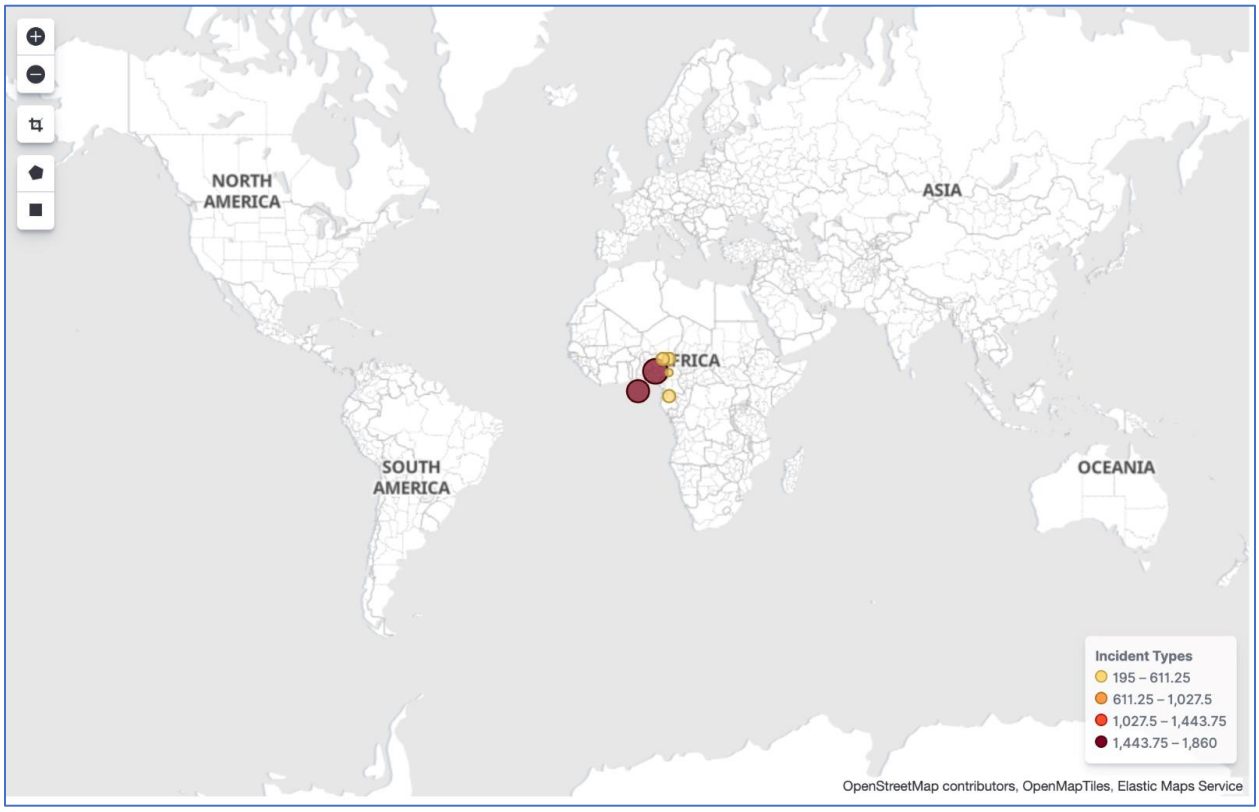

final - Kibana
Not secure | 34.171.167.103:5601/app/kibana#/dashboard/aa8c8560-20ec-11ed-986e-a780daa23469?_g=(refreshInterval:(pause:!t,value:0),time:(from:no...
Georgian College
jobs
pr documents
Amazon.ca: Low Pri...
extras
Bayshore
dmt assignment
DP Final Project
shoes
Other bookmarks
Descriptor.keyword: Descending - Count
Top 5 Boroughs with Most to Least 311 Calls
Number of Service C...
BROOKLYN
QUEENS
MANHATTAN
BRONX
STATEN ISLAND
Borough
2,000,000
4,000,000
6,000,000
Number of Service Calls
Count of location type of 311 service requests
Count
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
Count
RESIDENTIAL BUILDING
Street/Sidewalk
Residential Building/House
Street
Sidewalk
Store/Commercial
Property Address
Club/Bar/Restaurant
Park
Senior Address
Count of location type

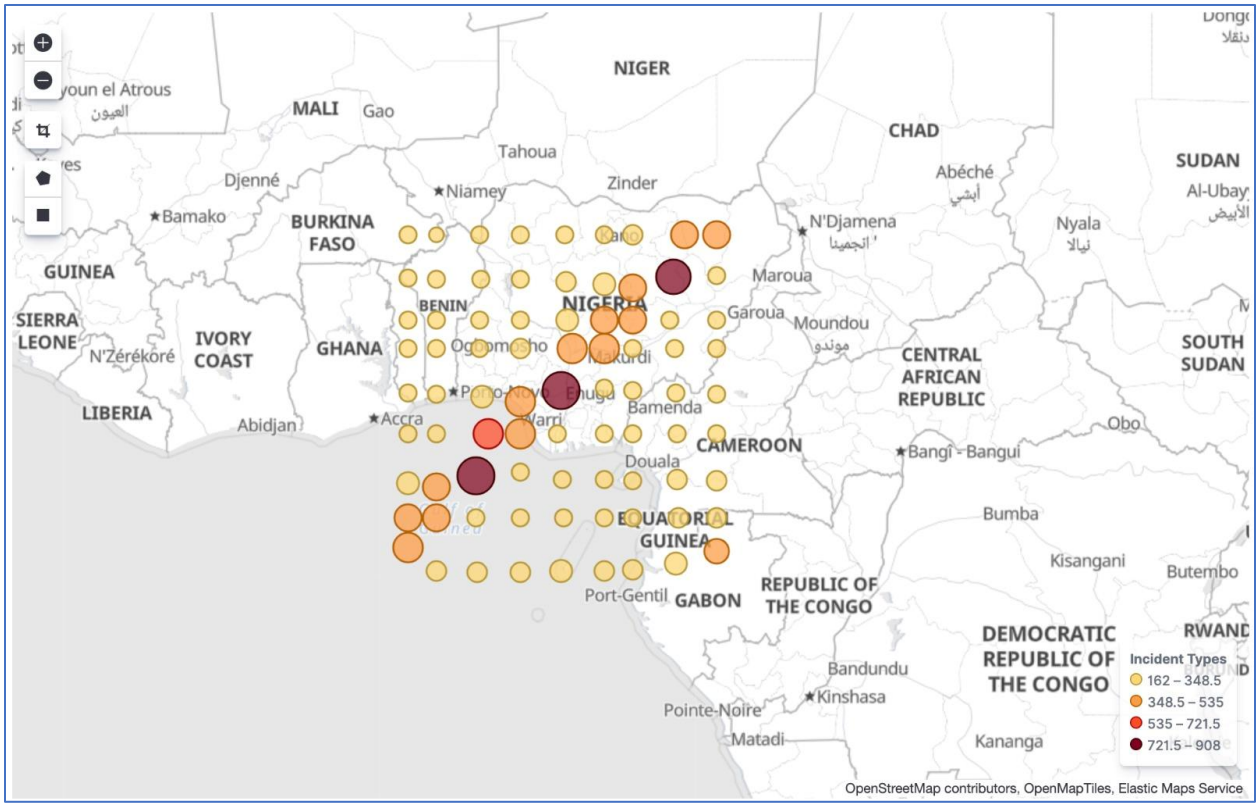

final - Kibana
Not secure | 34.171.167.103:5601/app/kibana#/dashboard/aa8c8560-20ec-11ed-986e-a780daa23469?_g=(refreshInterval:(pause:!t,value:0),time:(from:no...
Georgian College
jobs
pr documents
Amazon.ca: Low Pri...
extras
Bayshore
dmt assignment
DP Final Project
shoes
Other bookmarks
STATEN ISLAND
2,000,000
4,000,000
6,000,000
Number of Service Calls
Count of location type
Top 5 calls by Descriptor
Loud Music/Party
ENTIRE BUILDING
HEAT
No Access
Street Light Out
kibana

Issues Faced

1. The size of the dataset – the total size of the data set is 30.5 GB. Uploading the dataset in Kibana took a lot of time and sometimes up to two and hour hours. The PUT reindex api also took similar amounts of time to reindex.

2. The dataset was such that the location format was incorrect in the given dataset. Therefore, a new column was created called “loc” which took the latitude and longitude fields as parameters. However, for some reason the values of latitude and longitude kept getting corrupted when parsed into the new field “loc” and therefore an accurate co ordinate map could not be made.




NIGER
MALI
Gao
Tahoua
Zinder
Niamey
Djenné
Bamako
BURKINA FASO
CHAD
Abéché
SUDAN
N'Djamena
Nyala
GUINEA
Maroua
BENIN
NIGERIA
Garoua
Moundou
SIERRA LEONE
N'Zérékoré
IVORY COAST
GHANA
CENTRAL AFRICAN REPUBLIC
SOUTH SUDAN
LIBERIA
Abidjan
Accra
Bamenda
CAMEROON
Douala
Bangi - Bangui
Obo
EQUATORIAL GUINEA
Bumba
Kisangani
Butembo
Port-Gentil
GABON
REPUBLIC OF THE CONGO
DEMOCRATIC REPUBLIC OF THE CONGO
Bandundu
Kinshasa
Pointe-Noire
Matadi
Kananga
Incident Types
162 – 348.5
348.5 – 535
535 – 721.5
721.5 – 908
OpenStreetMap contributors, OpenMapTiles, Elastic Maps Service

Conclusion

Thus, from the above analysis we can see that Brooklyn is the city with the greatest number of 311 calls and in top 5 cities, Loud Music/Party is the common reason of 311 service calls.

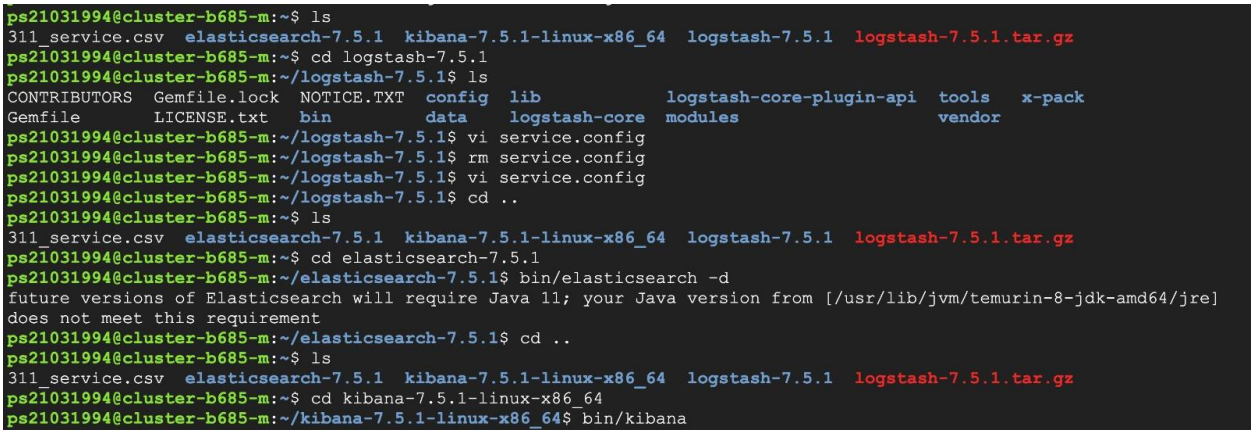
With the help of the ELK stack, by using the log stash file the dataset was loaded into Elasticsearch. Using Kibana, the dashboard was created displaying the various charts.

References

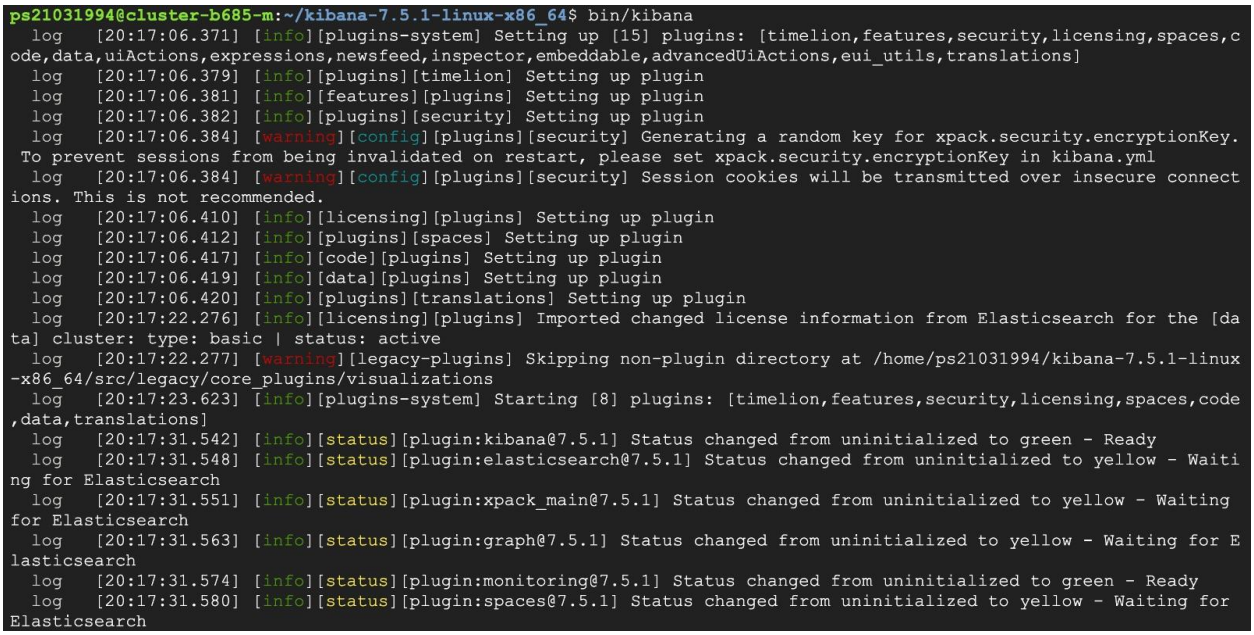
- https://en.wikipedia.org/wiki/3-1-1
- https://portal.311.nyc.gov/
- https://www.safewise.com/blog/what-is-311/
- https://settlement.org/ontario/daily-life/communication/phone/what-services-can-i-get-if-i-call-211-311-or-411-is-it-free-to-call/
- https://www.toronto.ca/home/311-toronto-at-your-service/311-frequently-asked-questions/.
- https://www.oracle.com/ca-en/big-data/what-is-big-data/
- https://www.knowi.com/blog/what-is-elastic-search/
- https://www.elastic.co/what-is/elk-stack
- https://www.elastic.co/guide/en/logstash/current/introduction.html

Appendix

Launching Elastic Search and Kibana



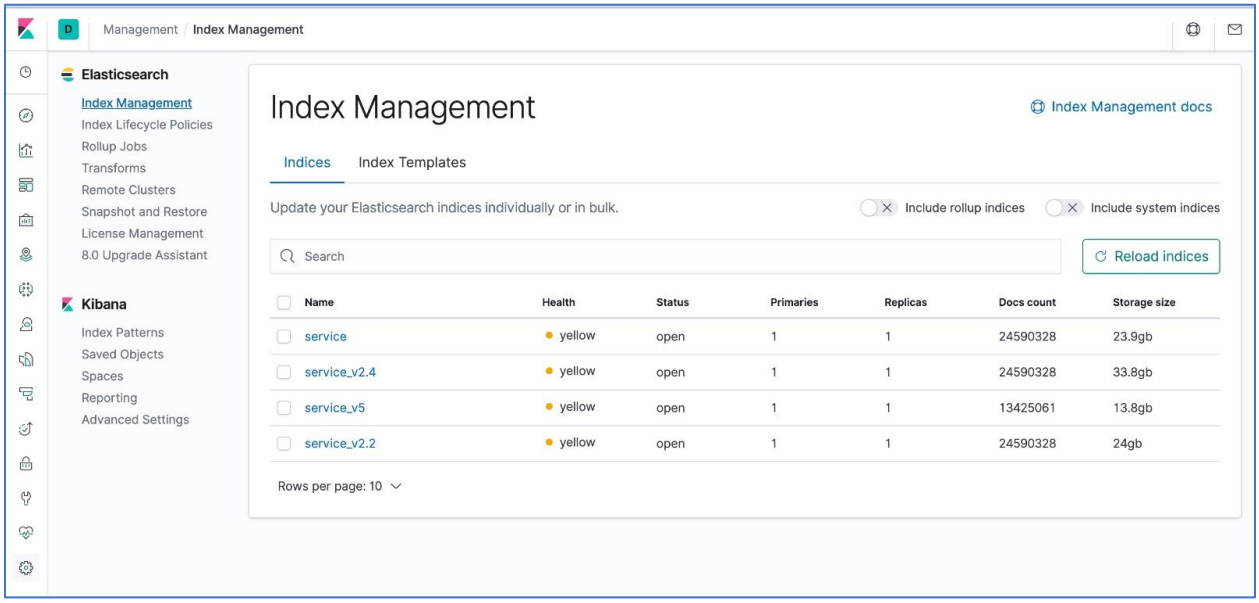
In the above image elastic search is launched successfully in daemonize mode. See the command “bin/elasticsearch -d”



In the above image Kibana is successfully launched.

Below image shows data getting uploaded into Elastic Search.

Below image shows that the index management in Kibana. The data from the log stash is loaded in Elastic Search



Config File Appendix

- Config/ original upload of all data into Kibana using log stash

```
input {


file {


path => "/home/ps21031994/311_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']

        }
```

```
        mutate {convert => ["X Coordinate (State Plane)", "integer"]}
        mutate {convert => ["Y Coordinate (State Plane)", "integer"]}
        mutate {convert => ["Latitude", "float"]}
        mutate {convert => ["Longitude", "float"]}


        date {
      match => ["Created Date", "MM/dd/YYYY HH:mm:ss a"]
    }
        date {
      match => ["Closed Date", "MM/dd/YYYY HH:mm:ss a"]
    }
        date {
      match => ["Due Date", "MM/dd/YYYY HH:mm:ss a"]
    }
        date {
      match => ["Resolution Action Updated Date", "MM/dd/YYYY HH:mm:ss a"]
    }
}
output {
        elasticsearch {
                hosts => "localhost"
                index => "service"
        }
```

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

- second config/ creating a separate column "loc" to mutate to float and then typecast to geo point.

```
input {

file {

path => "/home/ps21031994/311_service.csv"
start_position => "beginning"
sincedb_path => "/dev/null"

}

}

filter {
        csv {
                separator => ","
                columns => ['Descriptor', 'Latitude', 'Longitude']
        }

        mutate {convert => ["Latitude", "float"]}
        mutate {convert => ["Longitude", "float"]}
  mutate {
```

```
    add_field => {
    "loc" => ["%{[Latitude]}","%{[Longitude]}"]
    }
  }
  mutate {
    convert => {
    "loc" => "float"
    }
  }

}
output {
        elasticsearch {
                hosts => "localhost"
                index => "service_v5"
        }
        stdout {codec => dots}

}
```

- PUT api for creating an index to typecast **loc**(float) to **loc(geo point)**

```
PUT /service_v5.1
{
   "mappings" : {
     "properties" : {
       "@timestamp" : {
```

```
  "type" : "date"
},
"@version" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"Descriptor" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"Latitude" : {
  "type" : "float"
},
"Longitude" : {
  "type" : "float"
```

```
},
"column10" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column11" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column12" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
```

```
      }
    }
  },
  "column13" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  },
  "column14" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  },
  "column15" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
```

```
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column16" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column17" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column18" : {
  "type" : "text",
```

```
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column19" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column20" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
```

```
"column21" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column22" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column23" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
```

```
    }
  },
  "column24" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  },
  "column25" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  },
  "column26" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
```

```
      "ignore_above" : 256
    }
   }
},
"column27" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column28" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column29" : {
  "type" : "text",
  "fields" : {
```

```
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column30" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column31" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column32" : {
```

```
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column33" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column34" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
```

```
},
"column35" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column36" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column37" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
```

```
    }
   }
},
"column38" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column39" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column4" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
```

```
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column40" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column41" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column5" : {
  "type" : "text",
```

```
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column6" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column7" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
```

```
"column8" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"column9" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"host" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
```

```
        }
      },
      "loc" : {
        "type" : "geo_point"
      },
      "message" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "path" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      }
    }
  }
```

```
}
```

- post reindex api to move data between indices

```
POST _reindex
{
  "source": {
    "index": "service_v5"
  },
  "dest": {
    "index": "service_v5.1"
  }
}
```