**CSDA1020**

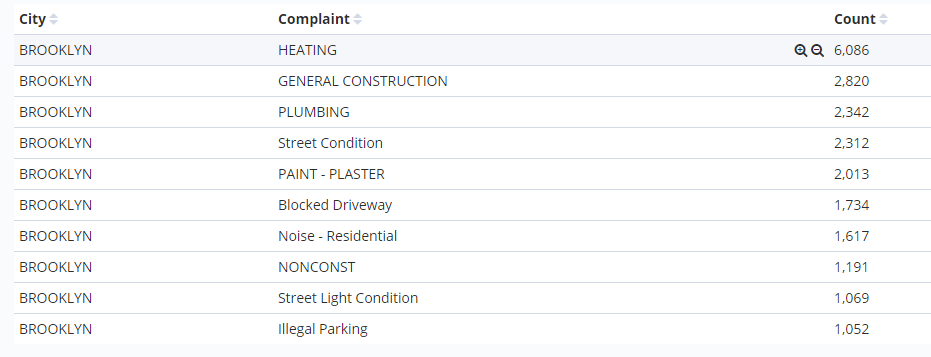
**Big Data Analytics Tools**

**Project 4, Hadoop and Hive Project**

**Group 1: Ioana Moca, Emily Dang, Karly Freiburger, Kashif Mirza, Sushma**

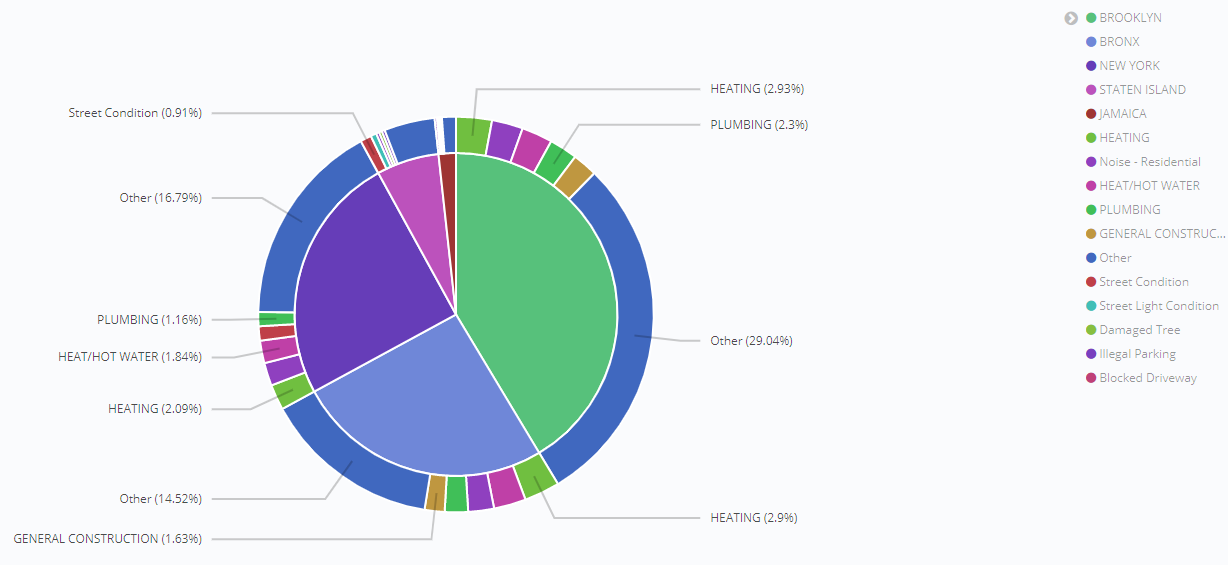
1. Create a table showing the top 10 cities with the highest calls alongside the count of top 10 complaint calls (by Descriptor) in each city.

Method: Visualizing using a data table, the metric is Count. There are two buckets. The first splits rows by City.keyword, ordered by Count, size 10. The second splits rows by Complaint Type.keyword, ordered by Count, size 10



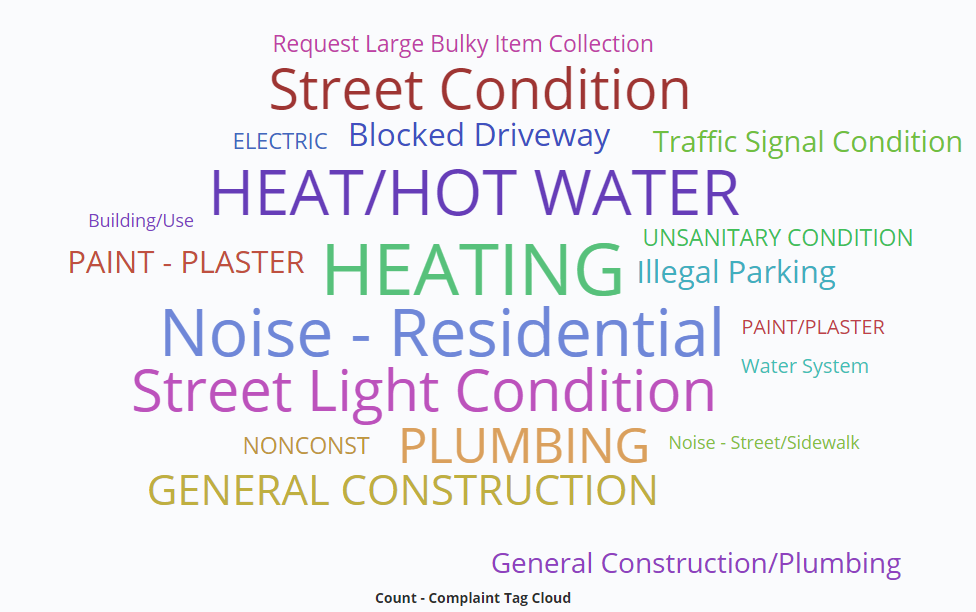
1. Create a pie chart showing the top 5 cities with the highest calls alongside the top five calls (Descriptor) in each city

Method: Visualize using the pie chart option. Select split slices and select the aggregation, Term and the field as City.keyword. Add a sub-bucket, split slices and select aggregation as Term. This time the field is the Complaint Type.keyword and the size is five.



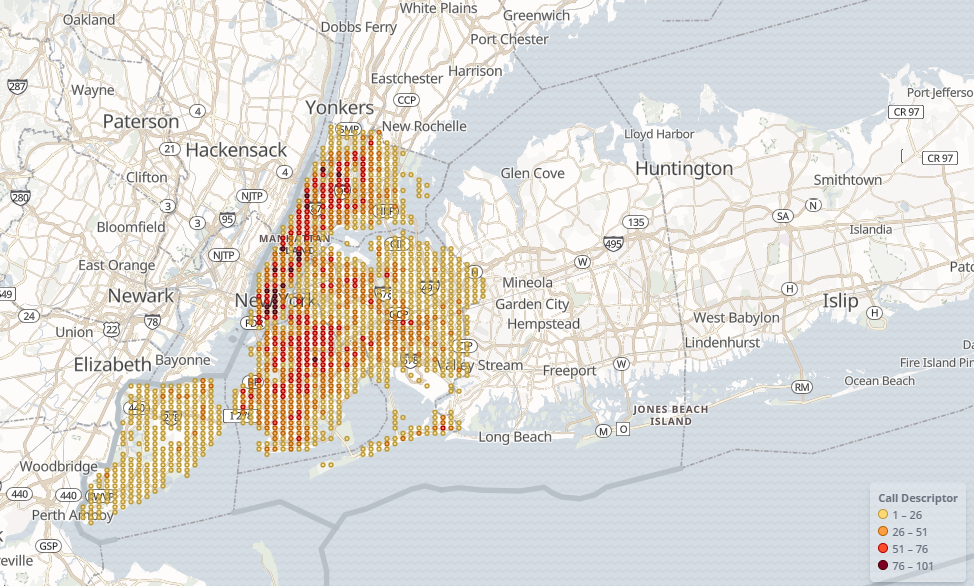
1. Create a tag cloud representing the top 20 call descriptors

Method: Visualize using the tag cloud option. Under buckets, the aggregation selected is Terms and the field is Complaint Type.keyword. The size is 20.



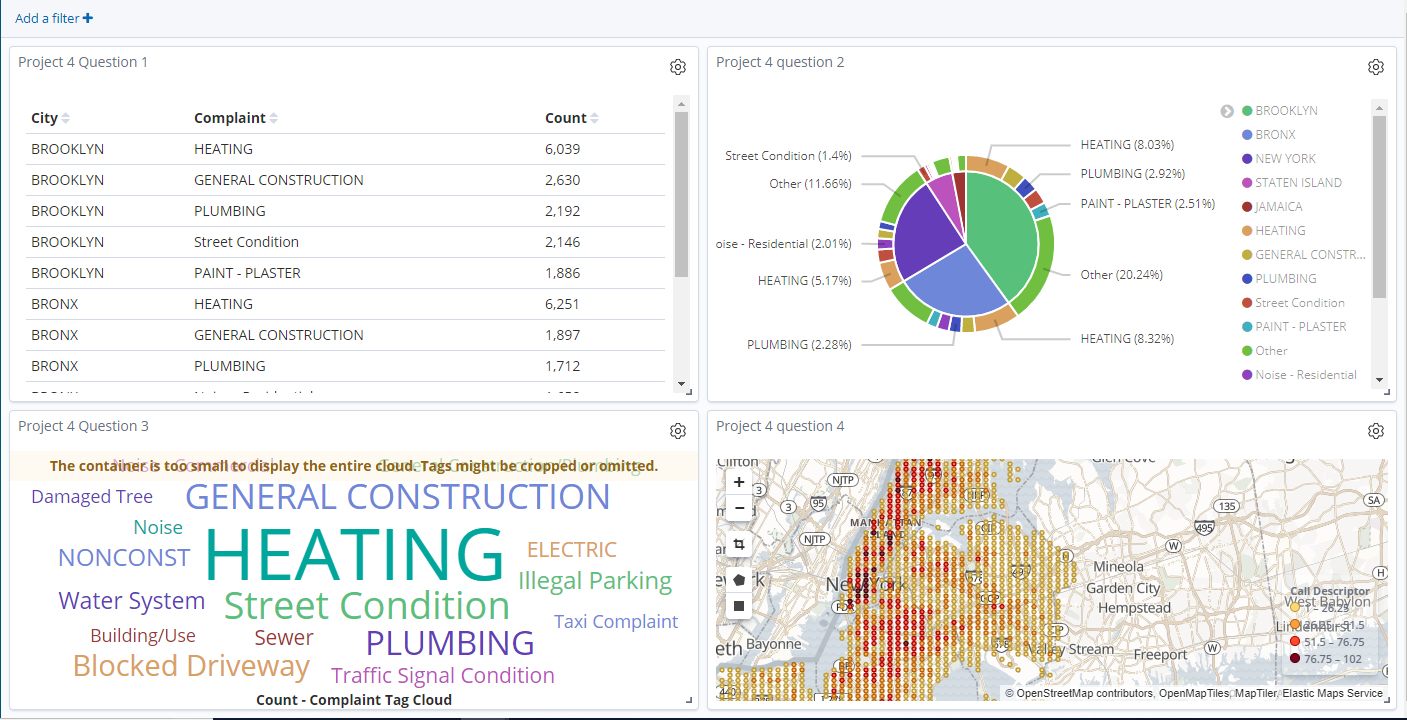
1. Create a coordinated map of all the major call descriptors in each city

Method: Create a new index and map it manually. Include a mapping for th geopoint. (note: ES mapping is available in the appendix). Update the logstash file to reflect the new index name. Visualize using coordinate map. Select Unique Count of Descriptor.keyword. Under buckets, keep geo coordinates and aggregation by geohash, and specify location as the field. The precision was set to 6.



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

Method: Save all visualizations. Select dashboard in the leftmost menu and add each of the four saved visualizations. Output is below.



Appendix 1: ES Mapping

PUT /project4map/

PUT /project4map/\_mapping

{

"project4map" : {

"mappings" : {

"311\_services" : {

"properties" : {

"@timestamp" : {

"type" : "date"

},

"@version" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Address Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Agency" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Agency Name" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"BBL" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Borough" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Bridge Highway Direction" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Bridge Highway Name" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Bridge Highway Segment" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"City" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Closed Date" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Community Board" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Complaint Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Created Date" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Cross Street 1" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Cross Street 2" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Descriptor" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Due Date" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Facility Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Incident Address" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Incident Zip" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Intersection Street 1" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Intersection Street 2" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Landmark" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Location" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Location Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Open Data Channel Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Park Borough" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Park Facility Name" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Resolution Action Updated Date" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Resolution Description" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Road Ramp" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Status" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Street Name" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Taxi Company Borough" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Taxi Pick Up Location" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Unique Key" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"Vehicle Type" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"X Coordinate (State Plane)" : {

"type" : "long"

},

"Y Coordinate (State Plane)" : {

"type" : "long"

},

"host" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"location" : {

"properties" : {

"lat" : {

"type" : "float"

},

"lon" : {

"type" : "float" }

},

"message" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

},

"path" : {

"type" : "text",

"fields" : {

"keyword" : {

"type" : "keyword",

"ignore\_above" : 256

}

}

}

}

}

}

}

}

PUT /project4map/\_mapping/311\_services

{

"311\_services" : {

"properties": {

"location": {

"type": "geo\_point",

"index": false

}

}

}

}

Appendix 2: Logstash Configuration

input {

file {

path => "/home/imoca\_econ/311.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"

"Y Coordinate (State Plane)" => "integer"

"Latitude" => "float"

"Longitude" => "float"

}

}

mutate {rename => {

"Latitude" => "[location][lat]"

"Longitude" => "[location][lon]"

}

}

}

output {

elasticsearch {

hosts => "localhost"

index => "project4map"

document\_type => "311\_services"

}

#stdout {}

}