ANALYTICS CODE -

Project topic : Analysis of building violations and the probability of it causing fire accidents

Tasks done so-far:

- Creation of a database table in Impala
- Querying the tables individually to understand useful patterns from selected columns
- Querying the tables taken together to understand useful patterns from selected columns

DATASET 1 : Fire Department Calls for Service
DATASET 2 : Notices of Violation Issued by the Department of Building Inspection
DATASET 3 : Fire Incidents

ANALYTICS-1 : Analysing the number of fire calls & building violations for areas based on their zipcodes and over years and get a pattern [Building_violations, Fire calls]

a)Analysing total calls and violation over area
impala-shell -i compute-1-1:21000 -q 'select zipcode, count(*) from pa1373.fire_incidents as fi join pa1373.firecalls as fc on fi.incident_no = fc.incident_number where year >= "2008" and year<= "2018" group by zipcode order by zipcode;' -B -o analytics11.csv

impala-shell -i compute-1-1:21000 -q 'select zipcode, count(*) from pa1373.viola where year >= "2008" and year<= "2018" group by zipcode order by zipcode;' -B -o analytics12.csv

```
Query submitted at: 2018-12-09 18:49:09 (Coordinator: http:/
Query progress can be monitored at: http://compute-1-1.local
+---------+----------+
| zipcode | count(*) |
+---------+----------+
| 94102   | 29047    |
| 94115   | 16531    |
| 94108   | 13094    |
| 94124   | 9209     |
| 94134   | 5476     |
+---------+----------+
```

Analysing violations and calls over years:

```
ar<='2018' group by year order by year desc;
Query: select year, count(*) from viola where year>='2008' and year<='2018' grou
p by year order by year desc
Query submitted at: 2018-12-11 09:45:40 (Coordinator: http://compute-1-1.local:2
5000)
Query progress can be monitored at: http://compute-1-1.local:25000/query_plan?qu
ery_id=c243f22264475528:4ceed0b800000000
+------+----------+
| year | count(*) |
+------+----------+
| 2018 | 11461    |
| 2017 | 16023    |
| 2016 | 15973    |
| 2015 | 14971    |
| 2014 | 13616    |
| 2013 | 11533    |
| 2012 | 9253     |
| 2011 | 15200    |
| 2010 | 18783    |
| 2009 | 16947    |
| 2008 | 20195    |
+------+----------+
Fetched 11 row(s) in 0.21s
```

ANALYTICS-2 Analysing the number of violations over different areas and years according to nov_category [Building_violations]

// getting top 5 areas(zipcodes) where max violations are reported

select zipcode,count(*) from building_violations group by zipcode order by count(*) desc limit 5;

// for each area: find the change in the count of violations of different types (nov_category description) over the past 10 years

impala-shell -i compute-1-1:21000 -q 'Select * from (Select nov_category_description,year,count(*) from (Select * from sg5783.building_violations where zipcode=94110) sub1 where year> "2008" and year< "2018" group by sub1.nov_category_description, sub1.year) sub2 where sub2.nov_category_description="fire section" or sub2.nov_category_description="smoke

detection section" or sub2.nov_category_description="building section" or sub2.nov_category_description="plumbing and electrical section" or sub2.nov_category_description="interior surfaces section" order by sub2.nov_category_description asc, sub2.year asc' -B -o analytics21.csv;

impala-shell -i compute-1-1:21000 -q 'Select * from (Select nov_category_description,year,count(*) from (Select * from sg5783.building_violations where zipcode=94109) sub1 where year> "2008" and year< "2018" group by sub1.nov_category_description, sub1.year) sub2 where sub2.nov_category_description="fire section" or sub2.nov_category_description="smoke detection section" or sub2.nov_category_description="building section" or sub2.nov_category_description="plumbing and electrical section" or sub2.nov_category_description="interior surfaces section" order by sub2.nov_category_description asc, sub2.year asc' -B -o analytics22.csv;

impala-shell -i compute-1-1:21000 -q 'Select * from (Select nov_category_description,year,count(*) from (Select * from sg5783.building_violations where zipcode=94117) sub1 where year> "2008" and year< "2018" group by sub1.nov_category_description, sub1.year) sub2 where sub2.nov_category_description="fire section" or sub2.nov_category_description="smoke detection section" or sub2.nov_category_description="building section" or sub2.nov_category_description="plumbing and electrical section" or sub2.nov_category_description="interior surfaces section" order by sub2.nov_category_description asc, sub2.year asc' -B -o analytics23.csv;

impala-shell -i compute-1-1:21000 -q 'Select * from (Select nov_category_description,year,count(*) from (Select * from sg5783.building_violations where zipcode=94102) sub1 where year> "2008" and year< "2018" group by sub1.nov_category_description, sub1.year) sub2 where sub2.nov_category_description="fire section" or sub2.nov_category_description="smoke detection section" or sub2.nov_category_description="building section" or sub2.nov_category_description="plumbing and electrical section" or sub2.nov_category_description="interior surfaces section" order by sub2.nov_category_description asc, sub2.year asc' -B -o analytics24.csv;

impala-shell -i compute-1-1:21000 -q 'Select * from (Select nov_category_description,year,count(*) from (Select * from sg5783.building_violations where zipcode=94103) sub1 where year> "2008" and year< "2018" group by sub1.nov_category_description, sub1.year) sub2 where sub2.nov_category_description="fire section" or sub2.nov_category_description="smoke detection section" or sub2.nov_category_description="building section" or sub2.nov_category_description="plumbing and electrical section" or sub2.nov_category_description="interior surfaces section" order by sub2.nov_category_description asc, sub2.year asc' -B -o analytics25.csv;

ANALYTICS-3 :Srishti Building section && interior surface section === structural collapse relating according to time and area.[Fire_calls_service, Building_violations]

// finding the common top 5 zipcodes having maximum number of calls and violations
Select zipcode,count(*) from building_violations where zipcode in (Select zipcode_of_incident from (Select zipcode_of_incident, count(*) from fire_calls_service group by zipcode_of_incident order by count(*) desc) sub1 limit 10) group by zipcode order by count(*) desc limit 5;

```
Query submitted at: 2018-12-09 18:49:09 (Coordinator: http:/
Query progress can be monitored at: http://compute-1-1.local
+---------+----------+
| zipcode | count(*) |
+---------+----------+
| 94102   | 29047    |
| 94115   | 16531    |
| 94108   | 13094    |
| 94124   | 9209     |
| 94134   | 5476     |
+---------+----------+
```

impala-shell -i compute-1-1:21000 -q 'Select year, count(*) from (Select * from sg5783.building_violations where (nov_category_description="building section" or nov_category_description="interior surfaces section") and (year>="2008" and year<="2018") and zipcode = 94102)sub1 group by sub1.year order by sub1.year asc' -B -o analytics511.csv;

impala-shell -i compute-1-1:21000 -q 'Select year, count(*) from (Select * from sg5783.fire_calls_service where (call_type="Confined Space / Structure Collapse") and

(year>="2008" and year<="2018") and zipcode_of_incident = 94102)sub1 group by sub1.year order by sub1.year asc' -B -o analytics512.csv;

impala-shell -i compute-1-1:21000 -q 'Select year, count(*) from (Select * from sg5783.building_violations where (nov_category_description="building section" or nov_category_description="interior surfaces section") and (year>="2008" and year<="2018") and zipcode = 94115)sub1 group by sub1.year order by sub1.year asc' -B -o analytics521.csv;

impala-shell -i compute-1-1:21000 -q 'Select year, count(*) from (Select * from sg5783.fire_calls_service where (call_type="Confined Space / Structure Collapse") and (year>="2008" and year<="2018") and zipcode_of_incident = 94115) sub1 group by sub1.year order by sub1.year asc' -B -o analytics522.csv;