Conestoga College

School of Applied Computer Science & Information Technology

PROG8450 - Big Data Integration and Storage

Analyzing New York City 311 Calls Using Apache Hive

Kizhakepura Velayudhan Geedhu

November 26, 2023

Table of Contents

Table of Contents

Introduction

In this assignment, the dataset of New York City 311 calls is explored and analyzed using Hive, a powerful data warehouse infrastructure built on top of Hadoop. The dataset provides valuable information about the various complaints and service requests received by the city's 311 call center. By leveraging the capabilities of Hive, the data set is loaded into a table, and a thorough analysis is done to gain meaningful insights.

The purpose of this report is to present our findings and offer recommendations to the Mayor of New York based on our analysis. While the appendices will provide detailed documentation of the Hive setup and queries, the primary focus of the report will be on the actionable recommendations derived from our analysis.

To effectively convey our findings, tables, graphs, and other visualizations are utilized aiming to persuade the mayor to take necessary actions based on the insights. By presenting compelling visual representations along with recommendations, to create a persuasive case for the proposed changes that New York City should consider for its own betterment.

Data Acquisition and Analyzation

Data Source

The data Source is the Dropbox link provided below which contains the smaller version of the entire dataset which is about 570MB.

https://www.dropbox.com/s/nmz1zd2bw2n5ora/nyc_311_sample.csv

Analysis and Recommendation

a. Top 10 Complaint Types

The Analysis revealed that out of the total 326 distinct Complaint Types, the topmost complaint types and their count value are as shown below:

Complaint Type	Request Count
Noise - Residential	87656
HEAT/HOT WATER	53796
Illegal Parking	53313
Blocked Driveway	39628
Street Condition	34947
Street Light Condition	32542
Request Large Bulky Item Collection	32309
Noise - Street/Sidewalk	29949
HEATING	26939
PLUMBING	25551

Table 1: Top 10 Complaint types and their respective Number of Count Value

Recommendation: To address the identified top 10 complaints, implement awareness campaigns or community programs to mimic the recurring issues.

b. Identify the top 5 Boroughs that generate the most complaints.

In total, there are 66014 boroughs of which the top 5 boroughs that have the most complaint count are as shown below in Table 2.

Borough	Complaint Count
BROOKLYN	227772
QUEENS	182421
MANHATTAN	143468
BRONX	141352
STATEN ISLAND	39093

Table 2: Top Boroughs that generate the most Complaints.

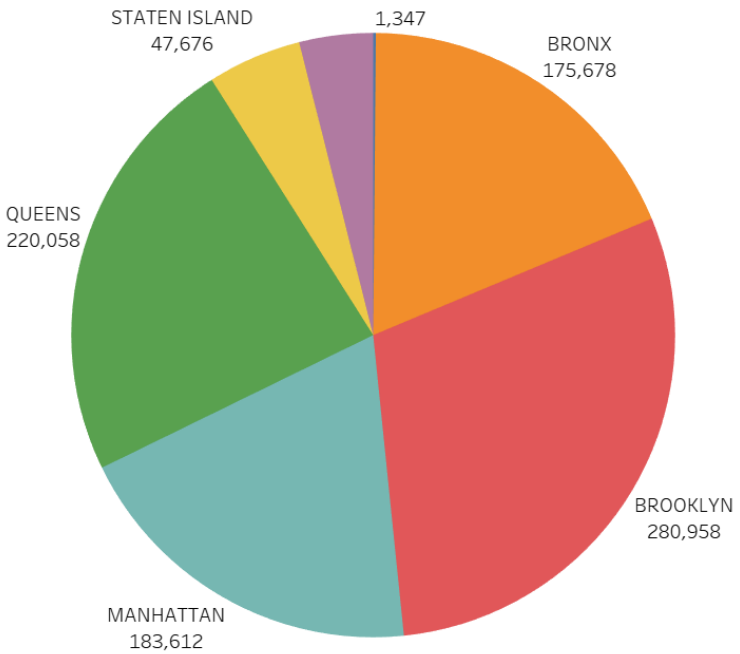

Fig 1: Borough and their respective Complaint Count

Recommendation: Speak to the administrative officials of the borough with the highest complaint count and identify the reasons for the high count. Allocate resources to identify the issues and provide awareness program and hardware resources to meet all the requirements and have the complaint count reduced.

c. Trends in the types of complaints and requests being made over time:

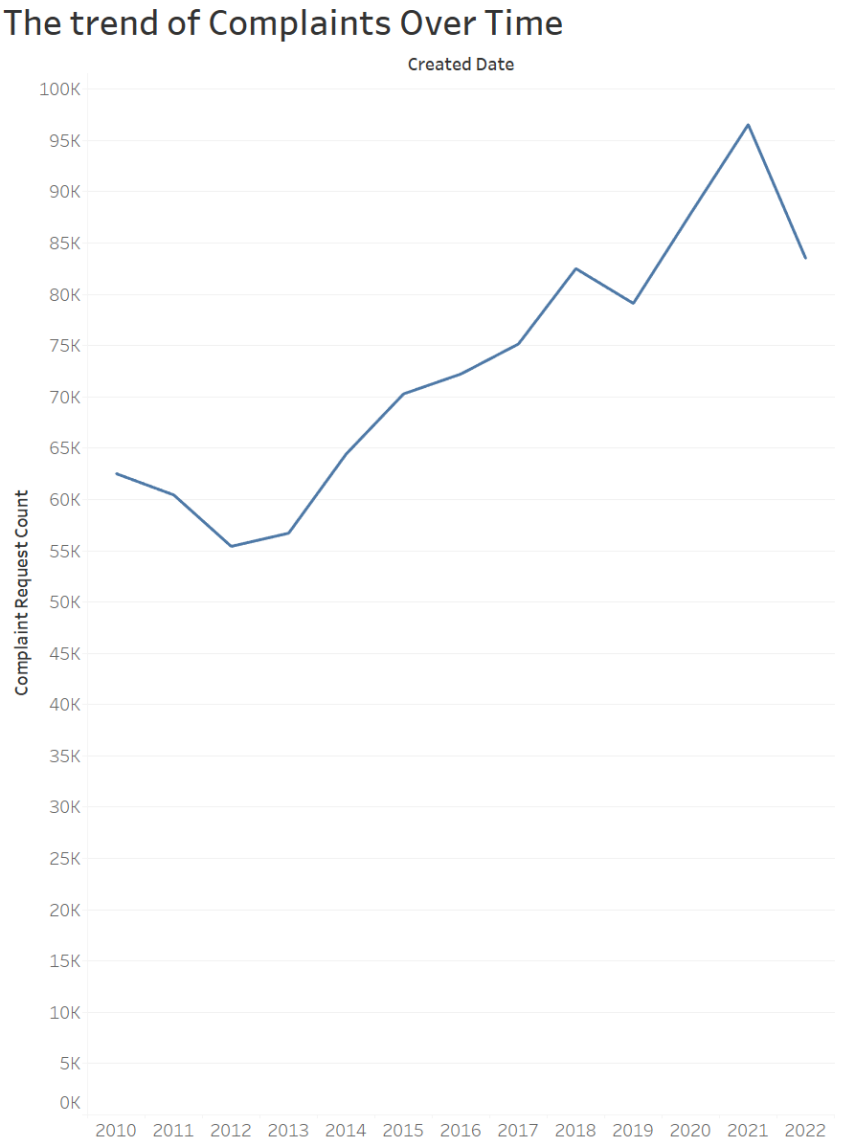

Fig 2: The trend of complaint counts from the year 2010 to 2022.

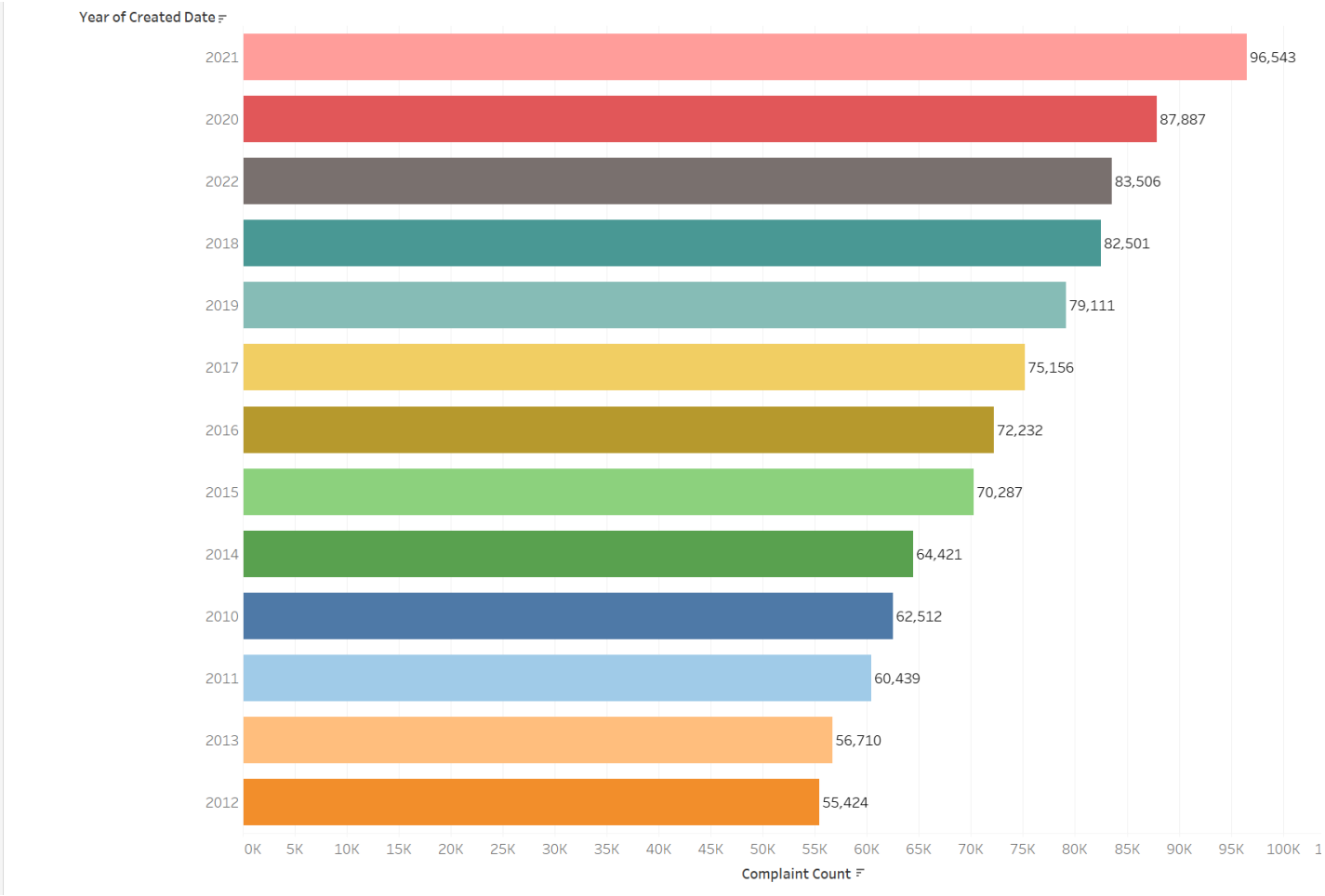

Fig 3: Complaint count of each year in descending order.

From the Fig 2, it is very much clear that the complaint count kept drasticaly increasing from the year 2012 to 2021 approximately from the range 55K to 95K. Inspite of this much hike it has decreased the complaint count by around 10k in an year that is 2022. As the count has significantly reduced, should make sure that the complaints that are being completely knocked out as shown in Table 3 are identified and the same procedure is followed so that it doesn't occur again.

Complaint Type	Complaint Count
COVID-19 Non-essential Construction	13
Collection Truck Noise	5
Cooling Tower	1
Dept of Investigations	1
Derelict Bicycle	39
Dirty Conditions	1087
Home Delivered Meal - Missed Delivery	8
Homeless Street Condition	38
Litter Basket / Request	74
Mass Gathering Complaint	30
Missed Collection (All Materials)	1264
Non-compliance with Phased Reopening	440
Other Enforcement	153
Overflowing Litter Baskets	52
Peeling Paint	1
Posting Advertisement	1
Recycling Enforcement	7
Sanitation Condition	773
Snow	193
Snow Removal	61
Storm	120
Sweeping/Inadequate	11
Sweeping/Missed	51
Sweeping/Missed-Inadequate	2
Vacant Lot	53
Vending	2
Window Guard	1
X-Ray Machine/Equipment	1

Table 3: Complaint Types that existed in 2021 but did not exist in 2022.

Complaint Type	Complaint Count
Building Drinking Water Tank	3
FHV Licensee Complaint	1
Internal Code	1
Private School Vaccine Mandate Non-Compliance	6
Radioactive Material	1
Recycling Basket Complaint	3
Transfer Station Complaint	1
ZTESTINT	1

Table 4: Newly added Complaint Type in 2022 and not listed in 2021.

Recommendation:

1. Understand the steps taken to get rid of complaint types that were listed in 2021 and not in 2022 and continue practicing the same.
2. Understand the reason for the newly added complaint type and do the needful.
3. After which identify other complaint types and the most highly rated one to avoid them by imposing necessary rules and regulations and also by creating awareness.

Conclusion

The report presents a detailed analysis of the 311 service center complaints. The proposals aim to serve as a roadmap for the mayor's office to implement focused activities, enhance overall public satisfaction, and improve service delivery. The appendices contain screenshots of the Hive setup and queries, providing a comprehensive understanding of the analysis process.

Appendices

Appendix A: Hive Setup

1. Create a directory in the Local directory and navigate to that directory.

```
[cloudera@quickstart ~]$ mkdir Mid_Term_Assignment
[cloudera@quickstart ~]$ cd Mid_Term_Assignment/
```

2. Download the data set using the below command.

```
[cloudera@quickstart Mid_Term_Assignment]$ wget https://www.dropbox.com/s/nmz1zd2bw2n5ora/nyc_311_sample.csv
--2023-11-22 14:51:25--  https://www.dropbox.com/s/nmz1zd2bw2n5ora/nyc_311_sample.csv
Resolving www.dropbox.com... 162.125.11.18, 2620:100:6050:18::a27d:b12
Connecting to www.dropbox.com|162.125.11.18|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: /s/raw/nmz1zd2bw2n5ora/nyc_311_sample.csv [following]
--2023-11-22 14:51:25--  https://www.dropbox.com/s/raw/nmz1zd2bw2n5ora/nyc_311_sample.csv
Connecting to www.dropbox.com|162.125.11.18|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://uc380f4a77631f765a2369043f5a.dl.dropboxusercontent.com/cd/0/inline/CIDwQJmDqfo0oULyUoEgg5UwISytLOTy2m7mA2fZmwz0N1jnz6x6T9Oso1pLHtByySdJHwKy85Xv07LjItQmwG6Sap932mYqaLa413uCmoV6Kd6ubw7iyh_3A8jo2
T7UPmI/file# [following]
--2023-11-22 14:51:26--  https://uc380f4a77631f765a2369043f5a.dl.dropboxusercontent.com/cd/0/inline/CIDwQJmDqfo0oULyUoEgg5UwISytLOTy2m7mA2fZmwz0N1jnz6x6T9Oso1pLHtByySdJHwKy85Xv07LjItQmwG6Sap932mYqaLa413uCmoV6Kd
6ubw7iyh_3A8jo2T7UPmI/file
Resolving uc380f4a77631f765a2369043f5a.dl.dropboxusercontent.com... 162.125.11.15, 2620:100:6050:15::a27d:b0f
Connecting to uc380f4a77631f765a2369043f5a.dl.dropboxusercontent.com|162.125.11.15|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 562137645 (536M) [text/plain]
Saving to: "nyc_311_sample.csv"

100%[==========================================================================================================================================================>] 562,137,645 37.5M/s   in 15s

2023-11-22 14:51:41 (36.5 MB/s) - "nyc_311_sample.csv" saved [562137645/562137645]
```

3. Create a directory in Hadoop.

```
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -mkdir -p /user/Geedhu/311_calls
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -ls /user
Found 9 items
drwxr-xr-x   - cloudera supergroup          0 2023-11-26 07:36 /user/Geedhu
drwxr-xr-x   - cloudera cloudera            0 2023-11-25 08:03 /user/cloudera
drwxr-xr-x   - mapred   hadoop              0 2017-10-23 09:15 /user/history
drwxrwxrwx   - hive     supergroup          0 2017-10-23 09:17 /user/hive
drwxrwxrwx   - hue      supergroup          0 2017-10-23 09:16 /user/hue
drwxrwxrwx   - jenkins  supergroup          0 2017-10-23 09:15 /user/jenkins
drwxrwxrwx   - oozie    supergroup          0 2017-10-23 09:16 /user/oozie
drwxrwxrwx   - root     supergroup          0 2017-10-23 09:16 /user/root
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user/spark
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -ls /user/Geedhu
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-11-26 07:36 /user/Geedhu/311_calls
[cloudera@quickstart Mid_Term_Assignment]$
```

4. Copy the csv file from local directory to Hadoop's Directory.

```
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -copyFromLocal nyc_311_sample.csv /user/Geedhu/311_calls
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -ls /user/Geedhu
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-11-26 07:39 /user/Geedhu/311_calls
[cloudera@quickstart Mid_Term_Assignment]$ hadoop fs -ls /user/Geedhu/311_calls
Found 1 items
-rw-r--r--   1 cloudera supergroup  562137645 2023-11-26 07:39 /user/Geedhu/311_calls/nyc_311_sample.csv
[cloudera@quickstart Mid_Term_Assignment]$
```

5. Launch Hive and then create the database named '311_calls'.

```
hive> CREATE DATABASE 311_calls;
OK
Time taken: 8.628 seconds
hive> SHOW DATABASES;
OK
311_calls
default
stocks_db
Time taken: 0.859 seconds, Fetched: 3 row(s)
hive>
```

6. Use Database '311_calls'.

```
hive> USE 311_calls;
OK
Time taken: 0.306 seconds
```

7. Create a table with the name ‘nycDataGeedhu’.

```
hive> CREATE TABLE IF NOT EXISTS nycDataGeedhu (
    > Unique_Key STRING,
    > Created_Date STRING,
    > Closed_Date STRING,
    > Agency STRING,
    > Agency_Name STRING,
    > Complaint_Type STRING,
    > Descriptor STRING,
    > Location_Type STRING,
    > Incident_Zip STRING,
    > Incident_Address STRING,
    > Street_Name STRING,
    > Cross_Street_1 STRING,
    > Cross_Street_2 STRING,
    > Intersection_Street_1 STRING,
    > Intersection_Street_2 STRING,
    > Address_Type STRING,
    > City STRING,
    > Landmark STRING,
    > Facility_Type STRING,
    > Status STRING,
    > Due_Date STRING,
    > Resolution_Description STRING,
    > Resolution_Action_Updated_Date STRING,
    > Community_Board STRING,
    > BBL STRING,
    > Borough STRING,
    > X_Coordinate_State_Plane STRING,
    > Y_Coordinate_State_Plane STRING,
    > Open_Data_Channel_Type STRING,
    > Park_Facility_Name STRING,
    > Park_Borough STRING,
    > Vehicle_Type STRING,
    > Taxi_Company_Borough STRING,
    > Taxi_Pick_Up_Location STRING,
    > Bridge_Highway_Name STRING,
    > Bridge_Highway_Direction STRING,
    > Road_Ramp STRING,
    > Bridge_Highway_Segment STRING,
    > Latitude STRING,
    > Longitude STRING,
    > Location STRING
    > )
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LOCATION '/user/Geedhu/311_calls'
    > TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.628 seconds
```

8. Check if table created or not.

```
hive> show tables;
OK
nycdatageedhu
Time taken: 0.027 seconds, Fetched: 1 row(s)
```

9. Load the data into the table created and check whether the data is loaded to the table.

```
hive> LOAD DATA INPATH '/user/Geedhu/311_calls/nyc_311_sample.csv' INTO TABLE nycDataGeedhu;
Loading data to table 311_calls.nycdatageedhu
Table 311_calls.nycdatageedhu stats: [numFiles=1, numRows=0, totalSize=562137645, rawDataSize=0]
OK
Time taken: 1.56 seconds
hive> SELECT * FROM nycDataGeedhu LIMIT 2;
OK
36921360        08/11/2017 06:40:58 AM  08/11/2017 09:34:57 PM  NYPD    New York City Police Department Blocked Driveway        Partial Access  Street/Sidewalk 11105   23-57 33 STREET 33 STREET       23 ROAD 24
 AVENUE ""      ""      ADDRESS ASTORIA ""      Precinct        Closed  08/11/2017 02:40:58 PM  The Police Department responded and upon arrival those responsible for the condition were gone. 08/11/2017 09:34:5
7 PM    01 QUEENS       4008340013      QUEENS  1008279 220690  ONLINE  Unspecified     QUEENS  ""      ""      ""      ""      ""      ""      ""      40.77238675404456       -73.91324635425228      "(40.77238
675404456
54845917        07/20/2022 06:12:00 PM  07/20/2022 06:12:00 PM  DEP     Department of Environmental Protection  Water System    Fire Hydrant Emergency (FHE)    ""      10453   1571 UNDERCLIFF AVENUE  UNDERCLIFF
 AVENUE W 174 ST        W 175 ST        ""      ""      ADDRESS BRONX   ""      ""      Closed  ""      The Department of Environmental Protection investigated this complaint and shut the running hydrant.    07
/20/2022 06:12:00 PM    05 BRONX        2028800135      BRONX   1005548 248170  ONLINE  Unspecified     BRONX   ""      ""      ""      ""      ""      ""      ""      40.847818695682015      -73.9230190877889"
(40.847818695682015
Time taken: 0.113 seconds, Fetched: 2 row(s)
```

Appendix B: SQL Queries Used

Query 1: Top 10 most common complaints and service requests made to the 311 call center.

```
hive> SELECT Complaint_Type, COUNT(*) AS Num_Complaints FROM nycDataGeedhu GROUP BY Complaint_Type ORDER BY Num_Complaints DESC LIMIT 10;
Query ID = cloudera_20231126100606_e4d6c332-d0c6-40d0-a1a7-8427c7d2a70d
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0001
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 10:07:10,247 Stage-1 map = 0%,  reduce = 0%
2023-11-26 10:07:45,323 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 2.68 sec
2023-11-26 10:07:46,656 Stage-1 map = 56%,  reduce = 0%, Cumulative CPU 9.78 sec
2023-11-26 10:07:50,194 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 11.24 sec
2023-11-26 10:08:04,484 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.19 sec
MapReduce Total cumulative CPU time: 15 seconds 190 msec
Ended Job = job_1701012338336_0001
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0002
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-11-26 10:08:17,853 Stage-2 map = 0%,  reduce = 0%
2023-11-26 10:08:23,111 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.76 sec
2023-11-26 10:08:30,649 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.78 sec
MapReduce Total cumulative CPU time: 1 seconds 780 msec
Ended Job = job_1701012338336_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 15.19 sec   HDFS Read: 562192883 HDFS Write: 12908 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.78 sec   HDFS Read: 18505 HDFS Write: 245 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 970 msec
OK
Noise - Residential     87656
HEAT/HOT WATER  53796
Illegal Parking 53313
Blocked Driveway        39628
Street Condition        34947
Street Light Condition  32542
Request Large Bulky Item Collection     32309
```

```
Noise - Residential     87656
HEAT/HOT WATER  53796
Illegal Parking 53313
Blocked Driveway        39628
Street Condition        34947
Street Light Condition  32542
Request Large Bulky Item Collection     32309
Noise - Street/Sidewalk 29949
HEATING 26939
PLUMBING        25551
Time taken: 109.53 seconds, Fetched: 10 row(s)
hive>
```

Query 2: Total number of Distinct complaint types

```
hive> SELECT COUNT(DISTINCT Complaint_Type) AS total_distinct_complaints FROM nycDataGeedhu;
Query ID = cloudera_20231126103434_b5ee0924-1810-4329-a928-6087cfb26e19
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0003
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1
2023-11-26 10:34:43,629 Stage-1 map = 0%,  reduce = 0%
2023-11-26 10:35:05,271 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 2.77 sec
2023-11-26 10:35:12,091 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 11.28 sec
2023-11-26 10:35:18,708 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 12.81 sec
MapReduce Total cumulative CPU time: 12 seconds 810 msec
Ended Job = job_1701012338336_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 12.81 sec   HDFS Read: 562184029 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 810 msec
OK
326
Time taken: 47.452 seconds, Fetched: 1 row(s)
```

Query 3: Total number of Unique Borough in the New York City

```
hive> SELECT COUNT(DISTINCT Borough) AS total_distinct_Borough FROM nycDataGeedhu WHERE BOROUGH IS NOT NULL;
Query ID = cloudera_20231126110101_59e96fa2-dab8-4f87-bee0-5e982348774c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0006
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1
2023-11-26 11:01:35,850 Stage-1 map = 0%,  reduce = 0%
2023-11-26 11:01:53,885 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 7.8 sec
2023-11-26 11:01:55,538 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 11.87 sec
2023-11-26 11:01:56,570 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 14.32 sec
2023-11-26 11:02:01,737 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 16.16 sec
MapReduce Total cumulative CPU time: 16 seconds 160 msec
Ended Job = job_1701012338336_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 16.16 sec   HDFS Read: 562182083 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 160 msec
OK
66014
Time taken: 35.627 seconds, Fetched: 1 row(s)
```

Query 4: To identify which Borough generates the greatest number of complaints.

```
hive> SELECT Borough, COUNT(*) AS Num_Complaints FROM nycDataGeedhu WHERE Borough IS NOT NULL GROUP BY Borough ORDER BY Num_Complaints DESC LIMIT 5;
Query ID = cloudera_20231126111010_8b842291-6fe8-485c-b9c4-f1ae86515f06
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0009
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 11:10:37,980 Stage-1 map = 0%,  reduce = 0%
2023-11-26 11:10:52,557 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 3.7 sec
2023-11-26 11:10:56,085 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 7.83 sec
2023-11-26 11:10:57,139 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.93 sec
2023-11-26 11:11:06,432 Stage-1 map = 100%,  reduce = 33%, Cumulative CPU 11.55 sec
2023-11-26 11:11:07,465 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 17.4 sec
MapReduce Total cumulative CPU time: 17 seconds 400 msec
Ended Job = job_1701012338336_0009
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-11-26 11:11:16,663 Stage-2 map = 0%,  reduce = 0%
2023-11-26 11:11:21,808 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.05 sec
2023-11-26 11:11:28,007 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.21 sec
MapReduce Total cumulative CPU time: 3 seconds 210 msec
Ended Job = job_1701012338336_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 17.4 sec   HDFS Read: 562190351 HDFS Write: 2003474 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.21 sec   HDFS Read: 2009044 HDFS Write: 80 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 610 msec
OK
BROOKLYN        227772
QUEENS  182421
MANHATTAN       143468
BRONX   141352
STATEN ISLAND   39093
Time taken: 58.36 seconds, Fetched: 5 row(s)
```

Query 5: Trends in the types of complaints and requests being made over time:

1. Recent data with latest date.

```
hive> SELECT DATE_FORMAT(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a')), 'yyyy-MM-dd') AS date,
    > COUNT(*) AS total_complaints
    > FROM nycDataGeedhu
    > GROUP BY DATE_FORMAT(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a')), 'yyyy-MM-dd')
    > ORDER BY date DESC LIMIT 10;
Query ID = cloudera_20231126121212_df13ab6a-cad7-4f5d-aaea-de55941b5d5f
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0020
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 12:12:15,716 Stage-1 map = 0%,  reduce = 0%
2023-11-26 12:12:34,015 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 3.88 sec
2023-11-26 12:12:36,487 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU 14.36 sec
2023-11-26 12:12:40,862 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 16.04 sec
2023-11-26 12:12:41,915 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 17.6 sec
2023-11-26 12:12:51,854 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 23.1 sec
MapReduce Total cumulative CPU time: 23 seconds 100 msec
Ended Job = job_1701012338336_0020
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0021
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-11-26 12:13:05,763 Stage-2 map = 0%,  reduce = 0%
2023-11-26 12:13:12,403 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.58 sec
2023-11-26 12:13:19,808 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.83 sec
MapReduce Total cumulative CPU time: 2 seconds 830 msec
Ended Job = job_1701012338336_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 23.1 sec   HDFS Read: 562196756 HDFS Write: 143020 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.83 sec   HDFS Read: 148847 HDFS Write: 148 SUCCESS
Total MapReduce CPU Time Spent: 25 seconds 930 msec
OK
2022-11-16      2
2022-11-15      283
```

```
2022-11-16      2
2022-11-15      283
2022-11-14      331
2022-11-13      250
2022-11-12      212
2022-11-11      187
2022-11-10      263
2022-11-09      290
2022-11-08      217
2022-11-07      266
Time taken: 75.045 seconds, Fetched: 10 row(s)
hive>
```

Recent trend from the initial start of the date:

```
hive> SELECT DATE_FORMAT(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a')), 'yyyy-MM-dd') AS date,
    > COUNT(*) AS total_complaints
    > FROM nycDataGeedhu
    > GROUP BY DATE_FORMAT(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a')), 'yyyy-MM-dd')
    > ORDER BY date ASC LIMIT 10;
Query ID = cloudera_20231126121515_901af5d1-6066-4bc3-a3f3-3b42a528ca0f
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0022
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 12:15:36,705 Stage-1 map = 0%,  reduce = 0%
2023-11-26 12:15:53,318 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 3.35 sec
2023-11-26 12:15:56,542 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU 8.48 sec
2023-11-26 12:15:57,571 Stage-1 map = 56%,  reduce = 0%, Cumulative CPU 13.86 sec
2023-11-26 12:15:59,658 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 14.94 sec
2023-11-26 12:16:00,694 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 16.13 sec
2023-11-26 12:16:10,391 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 20.76 sec
MapReduce Total cumulative CPU time: 20 seconds 760 msec
Ended Job = job_1701012338336_0022
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0023
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-11-26 12:16:23,275 Stage-2 map = 0%,  reduce = 0%
2023-11-26 12:16:30,825 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.51 sec
2023-11-26 12:16:38,148 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.82 sec
MapReduce Total cumulative CPU time: 2 seconds 820 msec
Ended Job = job_1701012338336_0023
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 20.76 sec   HDFS Read: 562196756 HDFS Write: 143020 SUCCESS
```

```
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.82 sec   HDFS Read: 148847 HDFS Write: 149 SUCCESS
Total MapReduce CPU Time Spent: 23 seconds 580 msec
OK
2010-01-01      83
2010-01-02      115
2010-01-03      178
2010-01-04      284
2010-01-05      256
2010-01-06      216
2010-01-07      204
2010-01-08      211
2010-01-09      125
2010-01-10      149
Time taken: 72.992 seconds, Fetched: 10 row(s)
hive>
```

Query 6: To identify complaints that existed in 2021 but do not exist in 2022, you can use a Hive query with a

```
hive> SELECT a.Complaint_Type AS Complaint_Type_2021,COUNT(a.Complaint_Type) AS Total_Complaints_2021
    > FROM nycDataGeedhu a
    > LEFT JOIN (
    >     SELECT DISTINCT Complaint_Type FROM nycDataGeedhu
    >     WHERE YEAR(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a'))) = 2022
    >   ) b
    > ON
    >   a.Complaint_Type = b.Complaint_Type
    > WHERE
    >   YEAR(FROM_UNIXTIME(UNIX_TIMESTAMP(a.Created_Date, 'MM/dd/yyyy hh:mm:ss a'))) = 2021
    >   AND b.Complaint_Type IS NULL
    > GROUP BY
    >   a.Complaint_Type;
Query ID = cloudera_20231126140707_aa5a78b6-2030-4ef5-bcd3-00ae3733dbdc
Total jobs = 4
Launching Job 1 out of 4
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0024, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0024/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0024
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 14:07:28,196 Stage-1 map = 0%,  reduce = 0%
2023-11-26 14:07:51,038 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 11.53 sec
2023-11-26 14:07:54,753 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU 12.94 sec
2023-11-26 14:07:55,918 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 13.31 sec
2023-11-26 14:07:57,040 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 14.9 sec
2023-11-26 14:07:58,879 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.85 sec
2023-11-26 14:08:10,881 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 19.16 sec
MapReduce Total cumulative CPU time: 19 seconds 160 msec
Ended Job = job_1701012338336_0024
Stage-7 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Execution log at: /tmp/cloudera/cloudera_20231126140707_aa5a78b6-2030-4ef5-bcd3-00ae3733dbdc.log
2023-11-26 02:08:16     Starting to launch local task to process map join;      maximum memory = 932184064
2023-11-26 02:08:17     Dump the side-table for tag: 1 with group count: 173 into file: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-07-17_547_1028631114663359358-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile01--.hashtable
2023-11-26 02:08:17     Uploaded 1 File to: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-07-17_547_1028631114663359358-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile01--.hashtable (6436 bytes)
2023-11-26 02:08:17     End of local task; Time Taken: 0.597 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 3 out of 4
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1701012338336_0025, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0025/
```

```
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0025
Hadoop job information for Stage-5: number of mappers: 3; number of reducers: 0
2023-11-26 14:08:29,097 Stage-5 map = 0%,  reduce = 0%
2023-11-26 14:08:43,416 Stage-5 map = 33%,  reduce = 0%, Cumulative CPU 2.9 sec
2023-11-26 14:08:47,809 Stage-5 map = 67%,  reduce = 0%, Cumulative CPU 8.4 sec
2023-11-26 14:08:48,951 Stage-5 map = 100%,  reduce = 0%, Cumulative CPU 14.74 sec
MapReduce Total cumulative CPU time: 15 seconds 710 msec
Ended Job = job_1701012338336_0025
Launching Job 4 out of 4
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0026, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0026/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0026
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2023-11-26 14:09:03,393 Stage-3 map = 0%,  reduce = 0%
2023-11-26 14:09:09,583 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 0.67 sec
2023-11-26 14:09:15,782 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 1.56 sec
MapReduce Total cumulative CPU time: 1 seconds 560 msec
Ended Job = job_1701012338336_0026
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 19.16 sec   HDFS Read: 562195643 HDFS Write: 6474 SUCCESS
Stage-Stage-5: Map: 3   Cumulative CPU: 15.71 sec   HDFS Read: 562186742 HDFS Write: 2122 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 1.56 sec   HDFS Read: 7693 HDFS Write: 662 SUCCESS
Total MapReduce CPU Time Spent: 36 seconds 430 msec
OK
COVID-19 Non-essential Construction     13
Collection Truck Noise  5
Cooling Tower   1
Dept of Investigations  1
Derelict Bicycle        39
Dirty Conditions        1087
Home Delivered Meal - Missed Delivery   8
Homeless Street Condition       38
Litter Basket / Request 74
Mass Gathering Complaint        30
Missed Collection (All Materials)       1264
NonCompliance with Phased Reopening     440
Other Enforcement       153
Overflowing Litter Baskets      52
Peeling Paint   1
Posting Advertisement   1
Recycling Enforcement   7
Sanitation Condition    773
Snow    193
```

```
Snow    193
Snow Removal    61
Storm   120
Sweeping/Inadequate     11
Sweeping/Missed 51
Sweeping/Missed-Inadequate      2
Vacant Lot      53
Vending 2
Window Guard    1
X-Ray Machine/Equipment 1
Time taken: 121.416 seconds, Fetched: 28 row(s)
```

Query 7: To find complaint types that are newly added in 2022 (not listed in 2021)

```
hive> SELECT b.Complaint_Type AS Newly_Added_Complaint_Type_2022, COUNT(b.Complaint_Type) AS Total_Complaints_2022
    > FROM nycDataGeedhu b LEFT JOIN
    >   (
    >     SELECT DISTINCT Complaint_Type FROM nycDataGeedhu
    >     WHERE YEAR(FROM_UNIXTIME(UNIX_TIMESTAMP(Created_Date, 'MM/dd/yyyy hh:mm:ss a'))) = 2021
    >   ) a
    > ON
    >   a.Complaint_Type = b.Complaint_Type WHERE
    >   YEAR(FROM_UNIXTIME(UNIX_TIMESTAMP(b.Created_Date, 'MM/dd/yyyy hh:mm:ss a'))) = 2022
    >   AND a.Complaint_Type IS NULL GROUP BY b.Complaint_Type;
Query ID = cloudera_20231126142222_d0dd8c97-8009-422d-9b61-b8adfc3854ab
Total jobs = 4
Launching Job 1 out of 4
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0027, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0027/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0027
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 14:22:12,090 Stage-1 map = 0%,  reduce = 0%
2023-11-26 14:22:27,179 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 2.71 sec
2023-11-26 14:22:32,670 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 13.26 sec
2023-11-26 14:22:33,818 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 14.29 sec
2023-11-26 14:22:42,169 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 16.57 sec
2023-11-26 14:22:43,196 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 18.03 sec
MapReduce Total cumulative CPU time: 18 seconds 30 msec
Ended Job = job_1701012338336_0027
Stage-7 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Execution log at: /tmp/cloudera/cloudera_20231126142222_d0dd8c97-8009-422d-9b61-b8adfc3854ab.log
2023-11-26 02:22:48     Starting to launch local task to process map join;      maximum memory = 932184064
2023-11-26 02:22:49     Dump the side-table for tag: 1 with group count: 193 into file: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-22-04_496_2271638440160154910-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable
2023-11-26 02:22:49     Uploaded 1 File to: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-22-04_496_2271638440160154910-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable (7163 bytes)
2023-11-26 02:22:49     End of local task; Time Taken: 0.71 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 3 out of 4
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1701012338336_0028, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0028/
```

```
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0027
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 3
2023-11-26 14:22:12,090 Stage-1 map = 0%,  reduce = 0%
2023-11-26 14:22:27,179 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 2.71 sec
2023-11-26 14:22:32,670 Stage-1 map = 78%,  reduce = 0%, Cumulative CPU 13.26 sec
2023-11-26 14:22:33,818 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 14.29 sec
2023-11-26 14:22:42,169 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 16.57 sec
2023-11-26 14:22:43,196 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 18.03 sec
MapReduce Total cumulative CPU time: 18 seconds 30 msec
Ended Job = job_1701012338336_0027
Stage-7 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Execution log at: /tmp/cloudera/cloudera_20231126142222_d0dd8c97-8009-422d-9b61-b8adfc3854ab.log
2023-11-26 02:22:48     Starting to launch local task to process map join;      maximum memory = 932184064
2023-11-26 02:22:49     Dump the side-table for tag: 1 with group count: 193 into file: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-22-04_496_2271638440160154910-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable
2023-11-26 02:22:49     Uploaded 1 File to: file:/tmp/cloudera/faa939fc-105f-452d-9a9b-67f6fcd564bb/hive_2023-11-26_14-22-04_496_2271638440160154910-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable (7163 bytes)
2023-11-26 02:22:49     End of local task; Time Taken: 0.71 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 3 out of 4
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1701012338336_0028, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0028/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0028
Hadoop job information for Stage-5: number of mappers: 3; number of reducers: 0
2023-11-26 14:22:57,825 Stage-5 map = 0%,  reduce = 0%
2023-11-26 14:23:14,019 Stage-5 map = 33%,  reduce = 0%, Cumulative CPU 3.15 sec
2023-11-26 14:23:15,192 Stage-5 map = 67%,  reduce = 0%, Cumulative CPU 7.7 sec
2023-11-26 14:23:16,224 Stage-5 map = 100%,  reduce = 0%, Cumulative CPU 13.47 sec
MapReduce Total cumulative CPU time: 13 seconds 470 msec
Ended Job = job_1701012338336_0028
Launching Job 4 out of 4
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1701012338336_0029, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1701012338336_0029/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1701012338336_0029
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2023-11-26 14:23:28,338 Stage-3 map = 0%,  reduce = 0%
```

```
2023-11-26 14:23:28,338 Stage-3 map = 0%,  reduce = 0%
2023-11-26 14:23:32,516 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 0.67 sec
2023-11-26 14:23:37,637 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 1.5 sec
MapReduce Total cumulative CPU time: 1 seconds 500 msec
Ended Job = job_1701012338336_0029
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 18.03 sec   HDFS Read: 562195643 HDFS Write: 7198 SUCCESS
Stage-Stage-5: Map: 3   Cumulative CPU: 13.47 sec   HDFS Read: 562186742 HDFS Write: 737 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 1.5 sec   HDFS Read: 6308 HDFS Write: 212 SUCCESS
Total MapReduce CPU Time Spent: 33 seconds 0 msec
OK
Building Drinking Water Tank    3
FHV Licensee Complaint  1
Internal Code   1
Private School Vaccine Mandate Non-Compliance   6
Radioactive Material    1
Recycling Basket Complaint      3
Transfer Station Complaint      1
ZTESTINT        1
Time taken: 94.803 seconds, Fetched: 8 row(s)
hive>
```