Parking Violations Issued - New York 2020

CIS4560 Introduction To Big Data

Group 1: Francisco Cortés, Arturo Pena, Kevin Anaya, Luigi Olmos, Robert Saravia

Data Set & Source

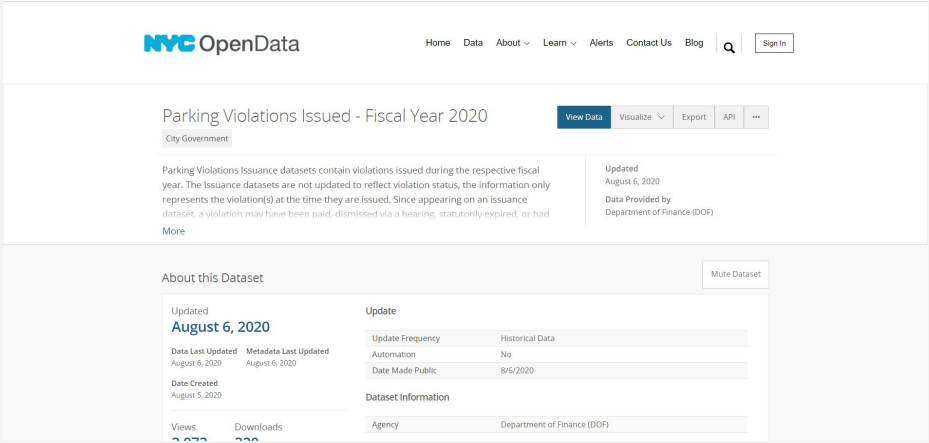
DataSource:https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2020/p7t3-5i9s

File Size

Data Set Uncompressed: 2GB

Data Set Compressed: 0.34 GB

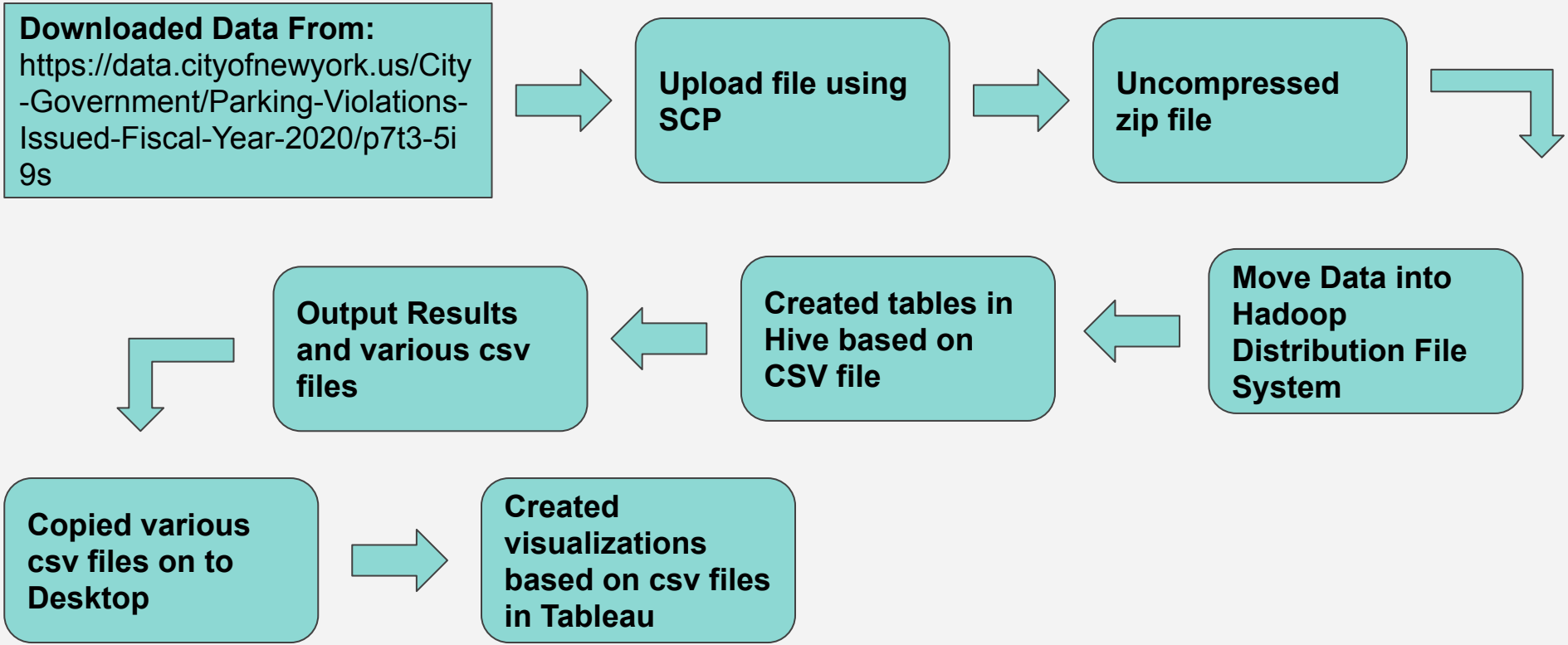
GitHub Link:
https://github.com/fcortes19/CIS-4560-Project



Specifications

- **Number of nodes:** 3
- **Memory:**192 GB
- **Storage :** 24TB/18TB (RAID)
- **Cluster Version:** 3.1.4
- **Applications:** Hive & Tableau

Implementation Flow Chart



Abstract & Overview

This project aims to provide insights and data analysis based on the data set we found using the 2020 Parking Tickets dataset provided by the state of New York. We will be utilizing a dataset that contains statistics such as registration state, violation code, vehicle make, street code. The project will utilize Hadoop, Hive, Tableau. One of the visualizations we will use in Tableau is a bar chart which tells us NY county receives the most parking tickets. Additionally, we will create a treemap in Tableau for which the parking violation code is the most violated. We also use a bar chart for which parking issuing agency gives the most tickets. The last visual we use is GEO spatial mapping for plate registration, given the most parking tickets.

Overview (Cont)

- Dataset has 43 column and over 12 million entries
- Create Table Statement (0.197 seconds)

```
0: jdbc:hive2://bigdata3.iscu.ac.kr:2181,bigd> CREATE EXTERNAL TABLE IF NOT EXISTS Parking_Violations_Issued(Summons_Number BIGINT, Plate_Id STRING, Registr
ation_State STRING, Plate_Type String, Issue_Date STRING, Violation_Code INT, Vehicle_Body_Type STRING, Vehicle_Make STRING, Issuing_Agency STRING, Street_C
ode1 INT, Street_Code2 INT, Street_Code3 INT, Vehicle_Expiration_Date BIGINT, Violation_Location INT, Violation_Precinct INT, Issuer_Precinct INT, Issuer_Co
de BIGINT, Issuer_Command STRING, Issuer_Squad BIGINT, Violation_Time STRING, Time_First_Observed STRING, Violation_County STRING, Violation_In_Front_Of_Or_
Opposite STRING, House_Number STRING, Street_Name STRING, Intersecting_Street STRING, Date_First_Observed BIGINT, Law_Section INT, Sub_Division STRING, Viol
ation_Legal_Code STRING, Days_Parking_In_Effect STRING, From_Hours_In_Effect STRING, To_Hours_In_Effect STRING, Vehicle_Color STRING, Unregistered_Vehicle I
NT, Vehicle_Year INT, Meter_Number STRING, Feet_From_Curb INT, Violation_Post_Code STRING, Violation_Description STRING, No_Standing_or_Stopping_Violation S
TRING, Hydrant_Violation STRING, Double_Parking_Violation STRING)
. . . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LOCATION '/user/fcortes6/ParkingViolationsIssued20/' TBLPROPERT
IES ('skip.header.line.count'='1');
```

NY county that receive the most parking tickets

```
0: jdbc:hive2://bigdata3.iscu.ac.kr:2181,bigd> select violation_county, COUNT(violation_county) AS Sum_of_Violation_County from parking_violations_issued G
roup By violation_county Order By Sum_of_Violation_County Desc LIMIT 13;
```

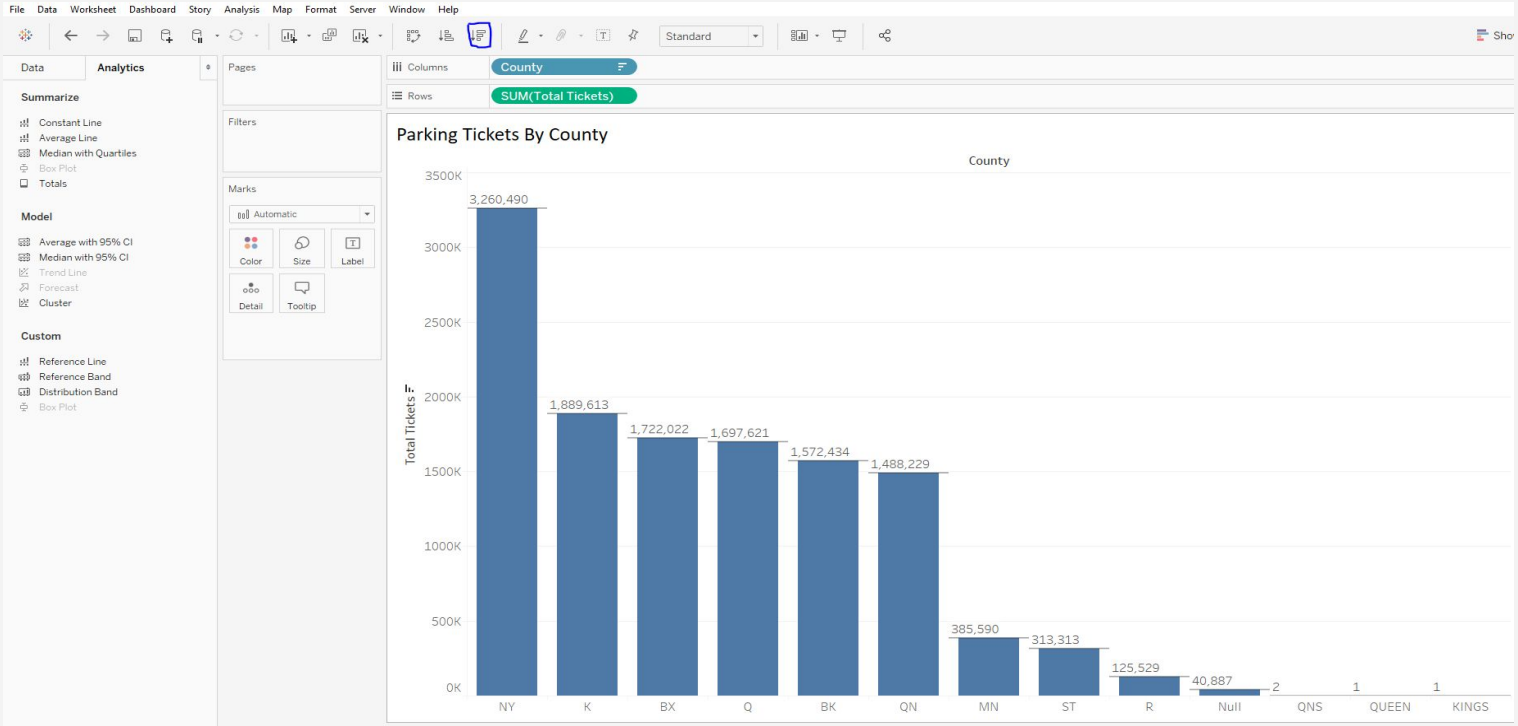
13 rows

(34.042 seconds)

violation_county	sum_of_violation_county
NY	3260490
K	1889613
BX	1722022
Q	1697621
BK	1572434
QN	1488229
MN	385590
ST	313313
R	125529
	40887
QNS	2
KINGS	1
K F	1

Analysis and Visualization

For the NY county that receive the most parking tickets in a bar chart



Parking Violation Code most violated

```
0: jdbc:hive2://bigdata3.iscu.ac.kr:2181,bigd> select violation_code, COUNT(violation_code) AS Sum_Of_Violation_Code from parking_violations_issued Group B
y violation_code Order By Sum_Of_Violation_Code Desc LIMIT 13;
```

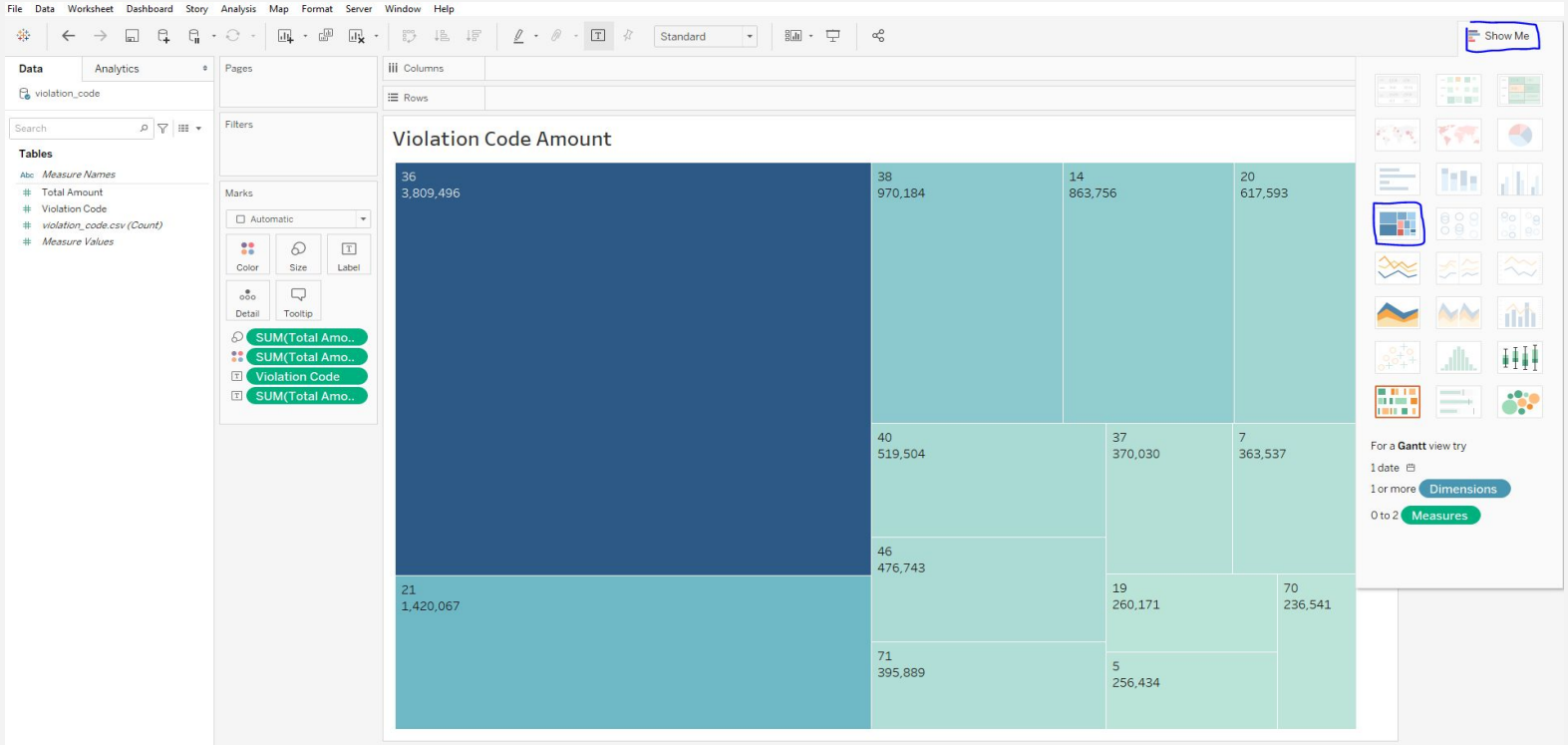
13 rows

(23.48 seconds)

violation_code	sum_of_violation_code
36	3809496
21	1420067
38	970184
14	863756
20	617593
40	519504
46	476743
71	395889
37	370030
7	363537
19	260171
5	256434
70	236541

Analysis and Visualization

For which Parking Violation Code is the most violated on a Treemaps.



Parking Violation Code (Cont)

- 36: $50 Exceeding the posted speed limit in or near a designated school zone.
- 21: $65 Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.
- 38: $65 Parking Meter -Failing to show a receipt or tag in the windshield. Drivers get a 5-minute grace period past the expired time on parking meter receipts.
- 14: $115 General No Standing: Standing or parking where standing is not allowed by sign, street marking or; traffic control device.
- 20: $65 General No Parking: No parking where parking is not allowed by sign, street marking or traffic control device.
- 40: $115 Stopping, standing or parking closer than 15 feet of a fire hydrant. Between sunrise and sunset, a passenger vehicle may stand alongside a fire hydrant as long as a driver remains behind the wheel and is ready to move the vehicle if required to do so.

Parking Issuing Agency

```
0: jdbc:hive2://bigdata3.iscu.ac.kr:2181,bigd> select issuing_agency, count(issuing_agency) AS Sum_Of_Issuing_Agency from parking_violations_issued Group B
y issuing_agency Order By Sum_Of_Issuing_Agency Desc LIMIT 13;
```

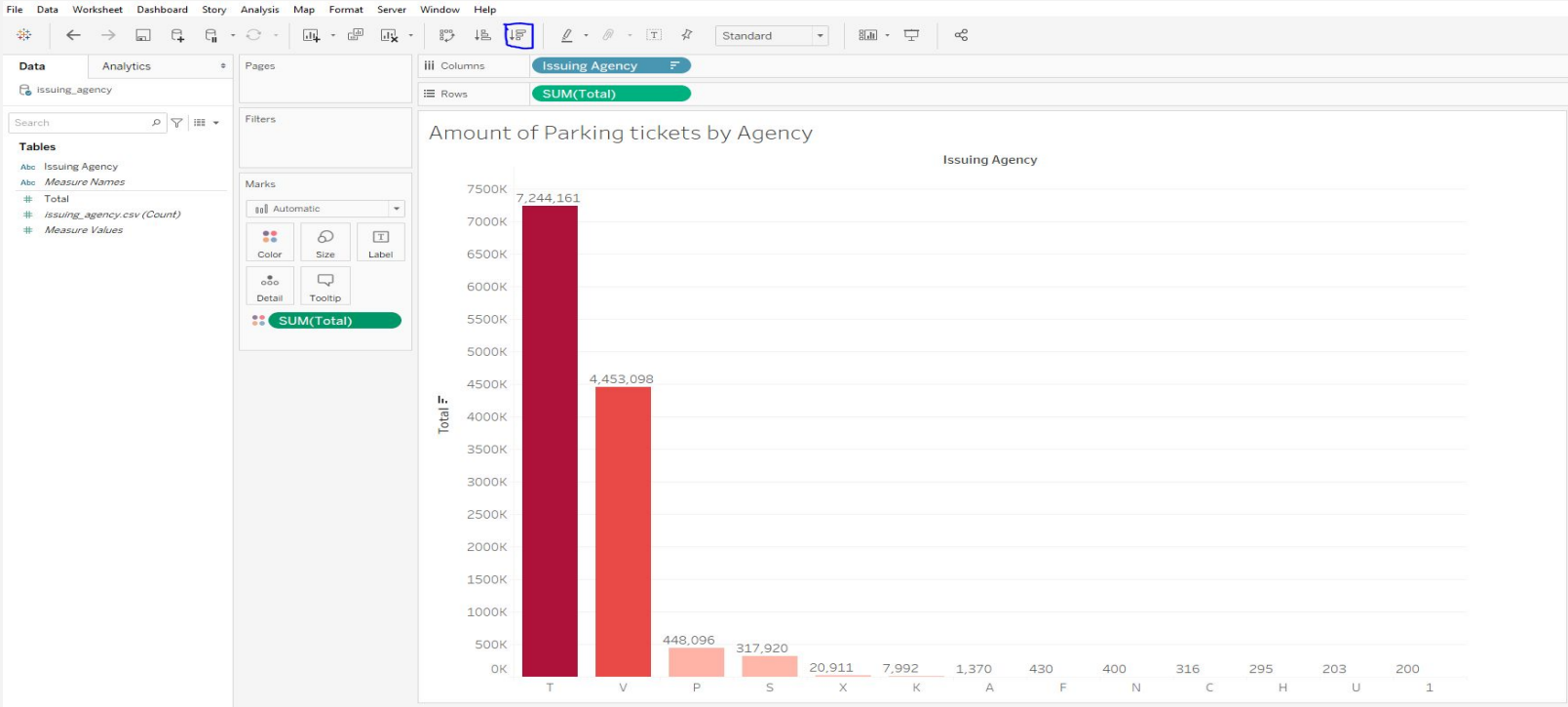
13 rows

(24.863 seconds)

issuing_agency	sum_of_issuing_agency
T	7244161
V	4453098
P	448096
S	317920
X	20911
K	7992
A	1370
F	430
N	400
C	316
H	295
U	203
1	200

Analysis and Visualization

For which parking issuing agency give the most tickets



Issuing agency

- T stands for Traffic Enforcement Agents
- V stands for New York City Parking Violation Bureau
- S stands for Sheriff Department
- F stands for Department of Finance’s
- B stands for Department of Business Services

STATE OF PLATE REGISTRATION

```
0: jdbc:hive2://bigdata3.iscu.ac.kr:2181,bigd> select Registration_State, count(Registration_State) AS Sum_Of_Registration_State from parking_violations_is
sued GROUP By Registration_State Order By Sum_of_Registration_State Desc;
```

69 rows

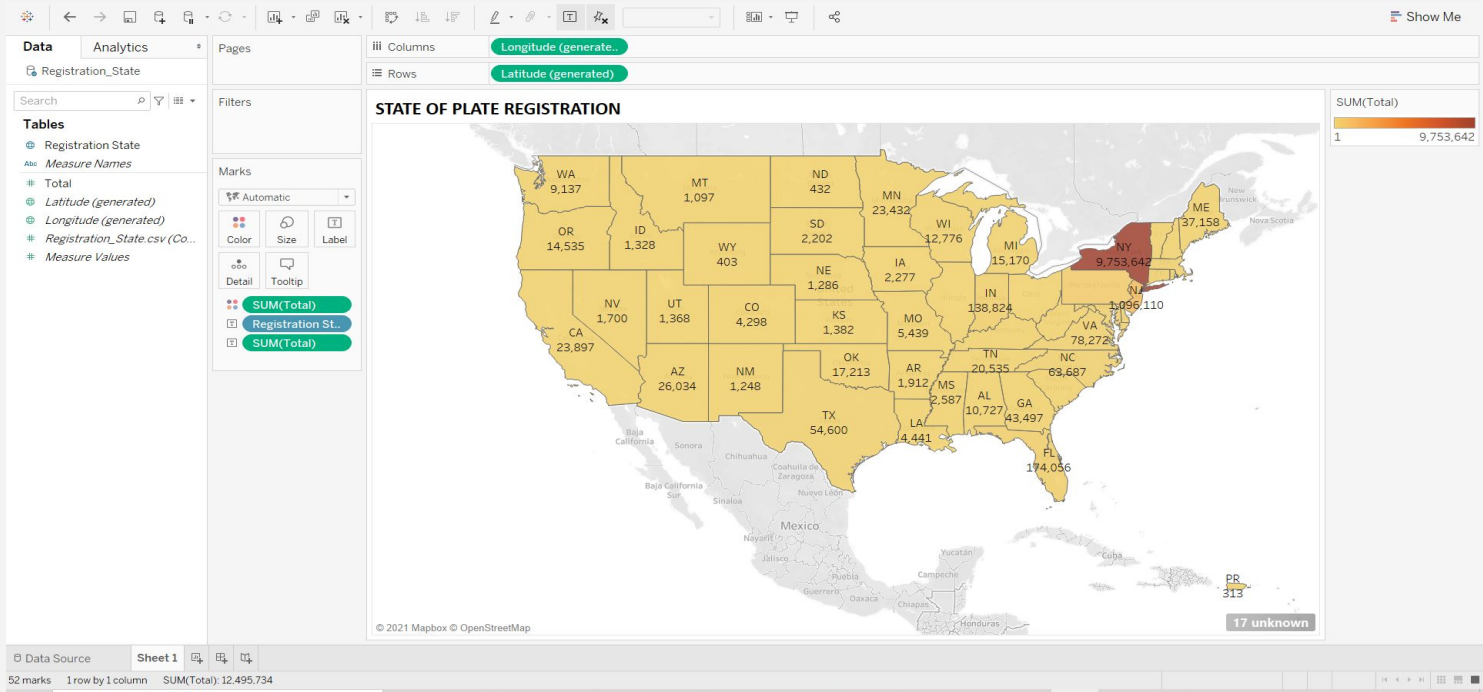
(24.652 seconds)

registration_state	sum_of_registration_state
NY	9753642
NJ	1096110
PA	338779
FL	174056
CT	165205
IN	138824
MA	83726
VA	78272
NC	63687
MD	59940
TX	54600
GA	43497
ME	37158
IL	32855
OH	27309
SC	26997
AZ	26034
99	24654
CA	23897
MN	23432
GV	21283
TN	20535
OK	17213
DE	15673
MI	15170
OR	14535
WI	12776
RI	11239
AL	10727
WA	9137
NH	8944
VT	8097
ON	5548
MO	5439
LA	4441
CO	4298
KY	4285
QB	3551
DC	3448
WV	3032
DP	2904
MS	2587
IA	2277
SD	2202
AR	1912
NV	1700
KS	1382
UT	1368
ID	1328
NE	1286
NM	1248
MT	1097
AK	465
ND	432
WY	403
PR	313
HI	182
AB	175
NS	112
BC	99
NB	94
FO	55
MB	29
PE	18
SK	8
MX	7
NT	3
YT	2
MM"	1

Analysis and Visualization

For which STATE OF PLATE REGISTRATION is given the most parking tickets

States	
NY	9,753,642
NJ	1,096,110
PA	338,779
FL	174,056
CT	165,205
IN	138,824
MA	83,726
VA	78,272
NC	63,687
MD	59,940
TX	54,600
GA	43,497
ME	37,158
IL	32,855
OH	27,309
SC	26,997
AZ	26,034
99	24,654
CA	23,897
MN	23,432



Conclusion

In conclusion, we used a multitude of data provided by the state of New york. The data included information on the 2020 parking violations. Details that were recorded contained various information such as registration state, violation code, vehicle make, street code, etc. From this dataset, we were able to analyze it using Hadoop, Hive, and Tableau. Using Hive we created a table named Parking Violation which included all the data. From there we sorted the data by creating queries and used Tableau to visualise certain aspects of the data. Overall we found the data to very interesting, the sheer amount of violations which occurred and the amount of money which comes from it was surprising.