# Parking Violations Issued - FY 2020 using Hadoop and Tableau

Authors: Francisco Cortés, Arturo Pena, Kevin Anaya, Luigie Olmos, Robert Saravia

Department of Information Systems, California State University Los Angeles
CIS4560 Introduction To Big Data
fcortes6@calststela.edu, kanaya8@calstatela.edu, lolmos11@calstatela.edu ,
apena69@calstatela.edu, rsaravi3@calstatela.edu

**Abstract:** This project aims to provide insights and data analysis based on the data set we found using the 2020 Parking Tickets dataset provided by 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 to tell us NY county that receives the most parking tickets. Additionally, we will create a treemap in Tableau for which 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 a GEO spatial mapping for state of plate registration that is given the most parking tickets.

## 1. Introduction

This project uses Hadoop and Hive, and tableau to keep and process and 2020 Parking Tickets dataset provided by New York. The dataset mainly consists of information statistics such as registration state, violation code, vehicle make, street code. We have chosen this dataset because it reflects the parking ticket problems in New York County, which we people would face in the city of New York. NYC Open Data is a popular platform for datasets that includes many different kinds of datasets in New York about almost anything about the city. The important thing is they try to make it the most recent. City departments are interested in using such information in order to understand and improve to prevent people from violating the law.

## 2. Related Work

New York is one of the biggest cities. The data set has an incredible amount of information regarding parking violations issued. NYC's open data is committed to making city data available for anyone to access online. They also have datasets by different agencies. In another dataset, they analyze response behavior for 6,646,540 tickets issued to 1,980,698 unique passenger vehicle license plates, totaling $424 million in fines and $85 million in late penalties.[1]

The census counts every resident in the United States. Its data collected by the census determine the number of seats each state has in the U.S. House. Census tracts evaluated in a study are located in zip codes with estimated occupancy rates between 101% and 844%.[2] This shows how other open data websites have many data for parking violations issued in New York.

Although NYC has multiple and continuous datasets, our use of the data is far less in comparison. An example of another study done in New York uses traffic violations to see the inequality and effects between white people who receive tickets and Black and Latinos who receive tickets. The study is titled *Driving While Black and Latinx: Stops, Fines, Fees, and Unjust Debt* by New York Law School. What the study finds is that there is a disproportion of violations in areas where there are predominantly Black or Latino residents. They were able to analyze data from parking violations and the corresponding zip code to cross-reference the demographics of the area. They found other information that in zip codes where the demographic is predominantly Black, they were issued several parking tickets at once, unlike in zip codes where the demographic is primarily white.

Our group also came across another study done by Lefevre Drucila titled *The Rise in Traffic Violation Tickets in New York*. Drucila goes over the recent increase in traffic violations through the years in New York. Druclia also finds that the number one cause of Traffic Citation is over exceeding the speed limit. He goes further into the top 10 reasons New York residents are given traffic violations and suspects the simple reason is that technology has gotten more advanced over the ages. Not only are officers well equipped with technology to help them catch criminals. However, also they are able now to track car movements easier than it was before. "It has become difficult for cops to lose you on the road. Once they have detected your vehicle, your license plate is automatically listed on their system." (Drucila) Drucila discusses the ease it is for an officer to give a citation with just a view of a license plate from a vehicle.

The use of traffic violations is vast. We can determine not only the causes of the violation but also the effects it has on those who receive them.

## 3. Specifications

The dataset comprises parking violations issued in New York. NYC open data is publicly open and data is shared upon everyone. The dataset is of the size 2GB and covers several years in May (May-08, 1972 – Dec-31, 2020) Table 1 shows files and size of the files from the dataset.

*Table 1 Data Specification*

| Data Set | Size (Total 2GB) |
|---|---|
| irregularities_head | 13742 MB |
| jams_head[1] | 2032 MB |
| alerts_head | 1923 MB |

The below table shows the specification for the Oracle cluster we are using and Hadoop specification for our project.

*Table 2 H/W Specification*

| Number of nodes | 3 |
|---|---|
| OCPUs | 10 |
| CPU speed | 2000.084 MHz |
| Memory | 182 GB |
| Storage | 82GB |
| HDFS Capacity | 120 |

## 4. Implementation Flowchart

Initially, the raw dataset, which comprises the detail of New York Parking Tickets, the whole process of date manipulation is shown in the below flowchart (Figure 1). There are three data logs in csv format that were uploaded to the Hadoop File System. After that, HiveQL is used as a querying language to create the tables' schema, clean data. Once the output file has been downloaded and opened in Tableau, we use Tableau to obtain the visualizations.

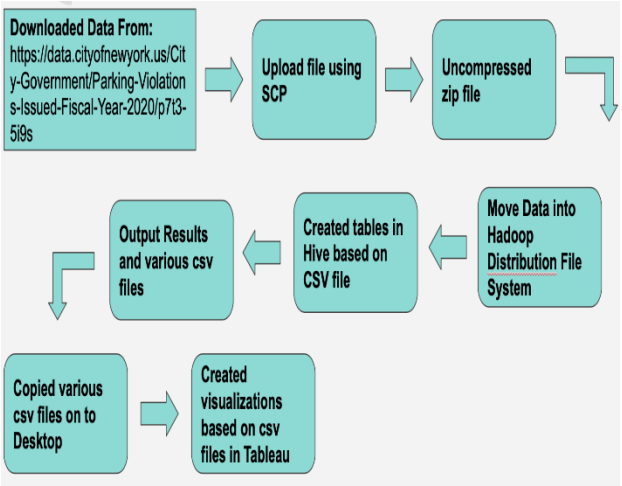


*Figure 1* - Implementation Flowchart

## 5. Data Cleaning

Once we were able to download the data files locally, we had to upload them to the Hadoop Cluster in hdfs using gitbash. We then had to verify that the file was correctly imported using the ls command. After verifying that the data file was correctly imported, we had to unzip the folder and create a directory to place the file in. Following adding the uncompressed file into a directory, we began to use hive to create tables and queries. The subsequent command was utilized to create the table:

```
DROP TABLE IF EXISTS
Parking_Violations_Issued;
---Create Table Parking_Violations_Issued
CREATE EXTERNAL TABLE IF NOT EXISTS
Parking_Violations_Issued(Summons_Number
BIGINT, Plate_Id STRING, Registration_State
STRING, Plate_Type String, Issue_Date STRING,
Violation_Code INT, Vehicle_Body_Type STRING,
Vehicle_Make STRING, Issuing_Agency STRING,
Street_Code1 INT, Street_Code2 INT, Street_Code3
INT, Vehicle_Expiration_Date BIGINT,
Violation_Location INT, Violation_Precinct INT,
Issuer_Precinct INT, Issuer_Code 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,
Violation_Legal_Code STRING,
Days_Parking_In_Effect STRING,
From_Hours_In_Effect STRING,
To_Hours_In_Effect STRING, Vehicle_Color
STRING, Unregistered_Vehicle INT, Vehicle_Year
INT, Meter_Number STRING, Feet_From_Curb
INT, Violation_Post_Code STRING,
Violation_Description STRING,
No_Standing_or_Stopping_Violation STRING,
Hydrant_Violation STRING,
Double_Parking_Violation STRING)
ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' LOCATION
'/user/fcortes6/ParkingViolationsIssued20/'
TBLPROPERTIES ('skip.header.line.count'='1');
```

Right after the creation of the table, four queries were made. These queries were made for the visualization of the data. This in turn allowed us to display a deeper understanding of the data set.

## 6. Analysis and Visualization

After data cleaning and preparation for further analysis, The first visualization (Figure 2), a bar chart , was made in Tableau. Move County to columns and Total Tickets to rows. Sort by descending order button shown with blue square.

We used the bar chart to show the count of parking tickets in each county. By using the total tickets column, we were able to come up with a bar with a count for each country.

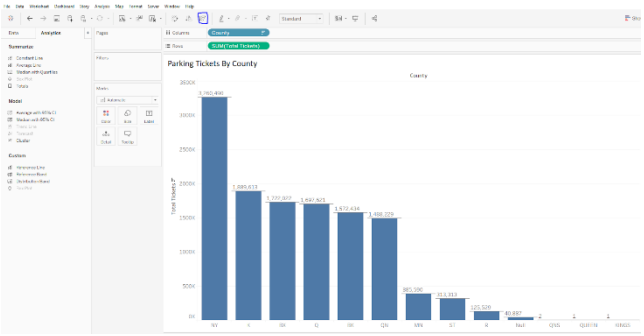


*Figure 2* - Parking Tickets By County in Tableau Bar Chart

### 6.1 Tableau

The tree map from Tableau has been used in the next map (Figure 3), We move Violation Code to columns and Total Amount to rows. Select Treemaps on the top right corner of the page marked with blue square.

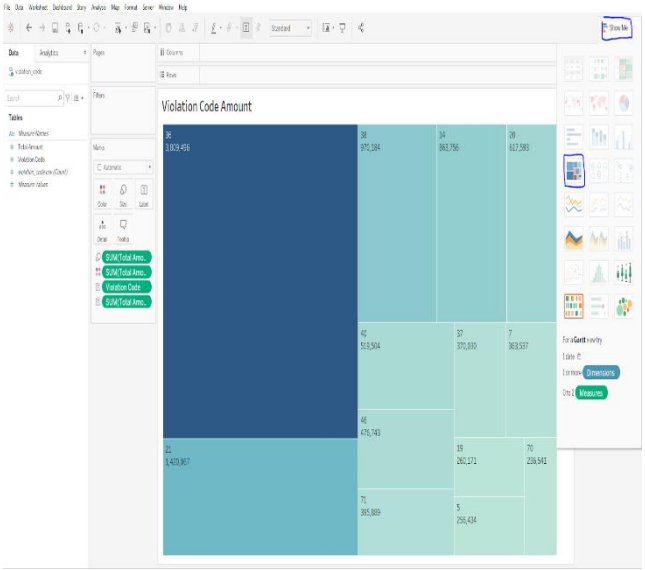


*Figure 3* - For which Parking Violation Code is the most violate.

Another Figure we created using Tableau was another bar chart (Figure 4). This bar chart however analyzed the agency in the dataset which issued the most violations in an ascending to descending order. What we found was that the traffic enforcement agency was the highest with around 7 million issues violations. The agency with the second highest issued violations was the New York City Parking Violation Bureau. In Figure 4 we can see the highest is denoted with darker red while the lowest is a lighter shade of red.

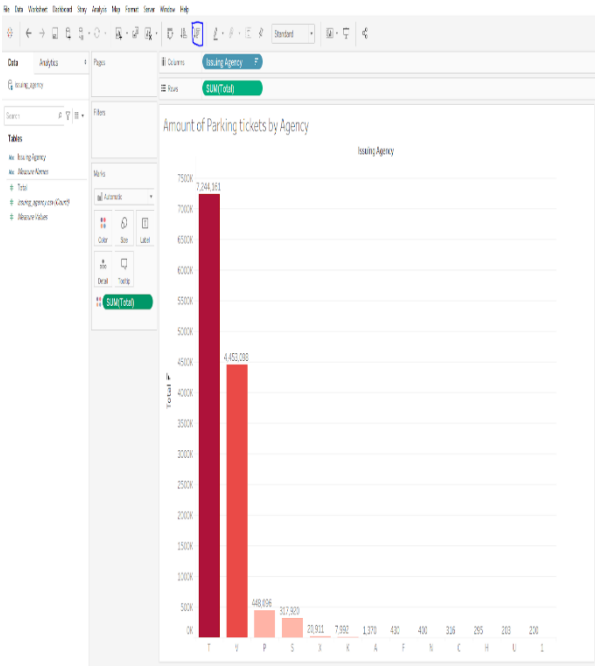


*Figure 4* - For which parking issuing agency give the most tickets.

The last figure we created (Figure 5) using Tableau was a Geospatial Mapping for the state license registration for the vehicle issued a citation. According to the dataset, the map highlights which state in the U.S. has the most parking tickets in New York. Our team realized that the state with citations for most parking violations was from New York, which comes as no surprise. However, at least every state has received a parking violation from the state of New York. While looking at Figure 5, we see that as you move towards the west coast, the number of parking violations decreases, nonetheless it is surprising that California and Arizona have one of the higher issued parking violations compared to the rest of the west coast since they are the furthest away from the state of New York. There are also two other states that frequently get Parking Violations from N.Y. The two are Minnesota and Florida. This is also peculiar since Minnesota has a lower population than

California, yet it has a significant amount of parking violations compared to the states around it. Another outlier we see is Florida, which is the state with the greatest amount of parking violations which isn't near N.Y.

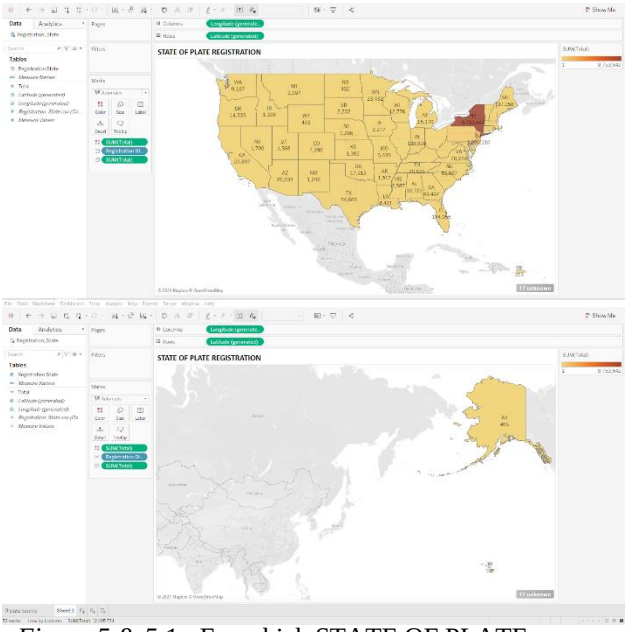


*Figure 5 & 5.1* - For which STATE OF PLATE REGISTRATION is given the most parking tickets.

## 7. 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 visualize certain aspects of the data. Overall, we found the data to be very interesting, the sheer amount of violations which occurred and the amount of money which comes from it was surprising.

For more information, dashboards and code visit project's GitHub link.

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

## References

[1] Heffetz, O., O'Donoghue, T., & Schneider, H. S. (2017). Forgetting and heterogeneity in task delay: Evidence from new york city parking-ticket recipients (No. w23012). National Bureau of Economic Research.Waze (October, 2016) https://www.nber.org/system/files/working_papers/w23012/w23012.pdf

[2] Chen, Q., Conway, A., & Cheng, J. (2017). Parking for residential delivery in New York City: Regulations and behavior. Transport Policy, 54, 53-60. https://www.researchgate.net/profile/Quanquan-Chen/publication/312457000_Parking_for_residential_delivery_in_New_York_City_Regulations_and_behavior/links/59c04f62a6fdcca8e56fde23/Parking-for-residential-delivery-in-New-York-City-Regulations-and-behavior.pdf

[3] New York Law School, "DRIVING WHILE BLACK AND LATINX: Stops, Fines, Fees, and Unjust Debts" (2020). Racial Justice Project. 8. https://digitalcommons.nyls.edu/racial_justice_project/8

[4] Drucila, L. (2020, January 17). The Rise in Traffic Violation Tickets in New York. Medium. https://medium.com/swlh/the-rise-in-traffic-violation-tickets-in-new-york-10e2b51ba8ee