

# Apache Hive Analysis On New York City Parking Tickets

Jihyun Moon, Hao Chen, James Logue  
Department of Information Systems, California State University  
Los Angeles

[jmoon14@calstatela.edu](mailto:jmoon14@calstatela.edu), [hchen53@calstatela.edu](mailto:hchen53@calstatela.edu), [jlogue@calstatela.edu](mailto:jlogue@calstatela.edu)

**Abstract:** New York being a very hectic urban state in terms of vehicle traffic will have its fair share of violations associated with parking. Similar to Los Angeles, the dataset will show the commonalities most individuals face when they violate the law. With the usage of data analytics and our datasets, visualization will be possible to pinpoint precise locations of all matter. All vehicles are processed with their violation code as well as other information such as time, location, and which officer was responsible for handing out these tickets. Our goal is with NYC parking ticket data, we can help examine some common reasons and factors to getting a parking violation. Through analysis, we feel we can provide a few things to watch out for those who plan on visiting or going to New York City.

## 1. Introduction

The topic of New York City parking was picked for the interest of finding out if there are any patterns in getting a parking ticket. Many people view parking tickets as an irritating and unpleasant to receive. In this paper, the objective is to find out the conditions that make it likely to receive these tickets. We will look for patterns as to help others understand what to avoid while at New York City or any other metropolitan areas.

## 2. Related work:

In this project we were inspired by many past works for New York Parking tickets such as articles on how someone found fire hydrants that generated a lot of tickets every year and how tickets were distributed throughout New York City.

### 2.1 Article on Ben Wellington

Our work to find the information on where parking tickets were given in New York City was based on the article, Open data saves New York City drivers from parking tickets, by Phil Johnson [2]. This article talks about a professor named Ben Wellington looked through the data from NYC open data to see which fire hydrants generate the most tickets. He found two hydrants that generated more than \$50,000 a year in tickets because the lines weren't drawn to indicate it was illegal to park there and

when his founding became more popular the Department of Transportation quickly marked that the spots were illegal.

### 2.2 Article on Open Data, and "I Quant NY"

Both our group and Ben Wellington used Big Data as stated in another article, Ben Wellington, Open Data, and 'I Quant NY', by Farah Halime [1]. The article stated that, "a big data provider that Wellington has mined several times for insight on legal claims against the city".<sup>1</sup> Both of our projects used big data to analyze parking tickets in New York City. The only difference is that we used it to find different data. In our project we used it to find patterns in getting a parking ticket in NY and he used it to find what areas generated the most parking tickets in NY.

### 2.3 Article on Visualizing NYC Parking Violation Data

Another article we found that had a similar study as us is, Visualizing New York City's Parking Violation Data [3], by Steven Ginzberg stated that, "I set out to visualize the Parking Violations issued in New York City during 2016."<sup>2</sup> Our projects are different because instead of using Big Data Steven used R and ggplot to visualize the parking violations in New York City. R and ggplot are platforms for data analysis that are capable of using data to create almost any type of graph. He visualized his data using mostly histograms and bar charts while we used a 3D map to help visualize our data. Both of our projects have similar goals which is to find patterns in parking violations and to help people avoid those parking spots so they don't end up getting a ticket.

## 3. Background/existing work:

While writing this paper and topic we would like to explore other existing works to the topic of our paper, but will begin with some background.

---

<sup>1</sup> He is using big data to help shape conversations and promote data-drive policy in America's largest city.

<sup>2</sup> Steven is using his knowledge in R to help people see what the patterns in parking ticket issuances and find perfect parking spots for the people.

In 2018, New York City's annual revenue from parking tickets, which is the largest in America is \$545 million a year [8]. The city after New York, Chicago, with the second largest revenue is \$264 million. This shows the astonishing amount of revenue that this city gets and it's almost twice as much as the runner up. This is a city that has complex laws that often confuse even though he or she may have lived in the city some time [5].

For this reason this topic seems to hold weight in the city of New York city more than any other major city in the world. The data set we used was from NYC Open Data, four data sets from 2014 to 2017 [9][10][11][12]. Now we will go into some similar and existing work.

### 3.1 Article on LA Parking Citations

This article written by Ethan Ward and Coco Huang explores the data that they analyzed given from the City of LA website. Their finding is that the number of citations from this year compared to last year was less [4]. Another thing they researched is what was the most common reasons for parking citation which was street cleaning, expired meters, red zone, preferential parking, and finally display of tabs. We did something similar when looking for patterns in citation reasons.

In the paper they find out the most common places that are for citations are Downtown, Hollywood, Westlake, Sawtelle, and Venice. This article is on revenue, the data we got didn't have that information. Our goal is similar this paper in some ways but we researching New York City and are interested to see if there are any similarities in the top violation shown in this paper.

### 3.2 Journal from Annals of GIS

Our second paper one where the authors use NYC parking tickets open data and learn ways to predict spatiotemporal legality of on-street parking [5]. Four types of spatial analysis units tested are point, street, census tract, and grid. This study was done because regulation rules make the search for on-street legal parking hard. Other data used is points of interest data and human mobility data.

There is pretty much no real-time information on street parking spots whether it is legal or not. This is a case for even the most local residents. They used the coordinates of each ticket using Google Maps API. Using and testing 6 different machine learning models, they learn in conclusion that the random forest (RF) produced the best spatial resolution with the least errors.

Though they also used data like we did but for different years, our paper will be looking more at the patterns of violation and circumstances then only specifically looking at locations.

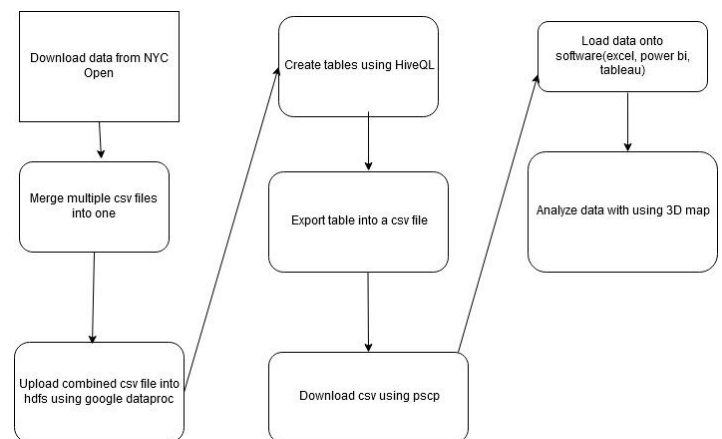
### 3.3 Article from Towards Data Science

Another article that is similar in some ways from our topic is the study from the author Colin Sinclair named *Avoid Parking Tickets in San Francisco Using Data Analytics*. This article is on finding out where is the place that one can have the best odds of not getting a parking ticket in San Francisco. His reason for this is him being a transient person who moves to different locations in the city every year and he finds that permits change often and would lead him to gambling on whether he is violating a parking rule or not.

He wanted to see if the higher volume are less susceptible than residential areas compared to city streets. He contacted the city for data and got a year's worth of SF parking ticket data. He then used openaddresses.io to reverse geocoded addresses, and with any he couldn't generate, he used OpenStreetMap API [6]. He got traffic volume from the County Transportation Authority with forecasting tools. He also got street cleaning routes from openDataSF.com. He used a relational database called SQLite (which the data was stored from San Francisco City) to create a Schema. He paired each ticket with street id and find or match similar addresses. Then with the street volume data and street cleaning data he saw that both use street names so he decided to filtered only on matches of the same street name. With this he was able to do the analysis. In the end he finds that his theory was right and found ways to reduce tickets by 30% and concludes that there's lower volume areas have less parking tickets.

In our paper we too look for some factors that increase the chances of getting parking tickets to help them decrease their chances of getting a ticket.

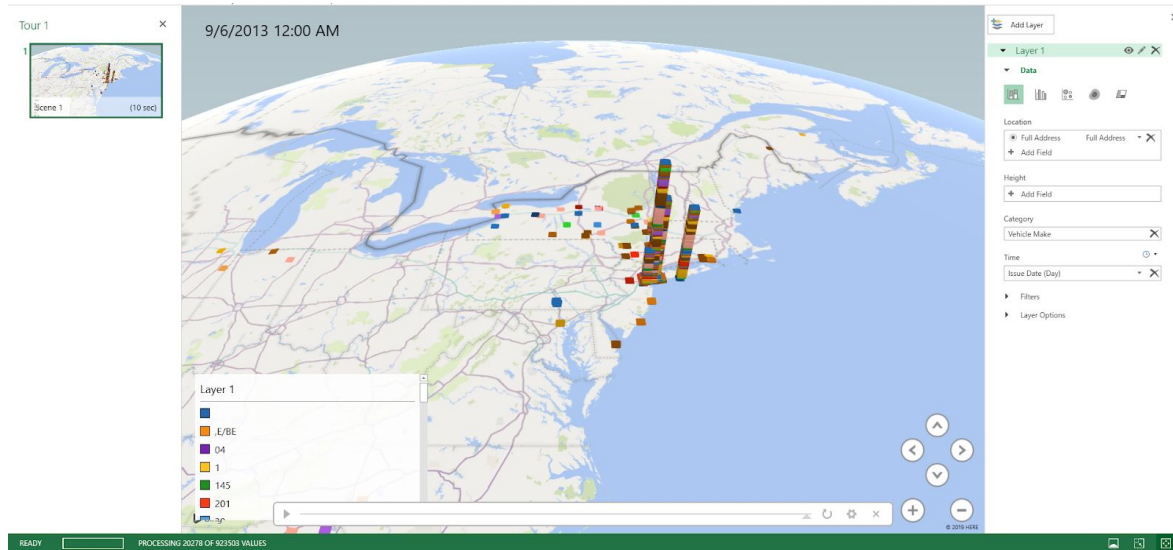
## 4. Our Work:



**Figure 1:** This illustrates our process on how we created used data from NYC Open to create a visualization using 3D maps to see patterns in NYC parking tickets.

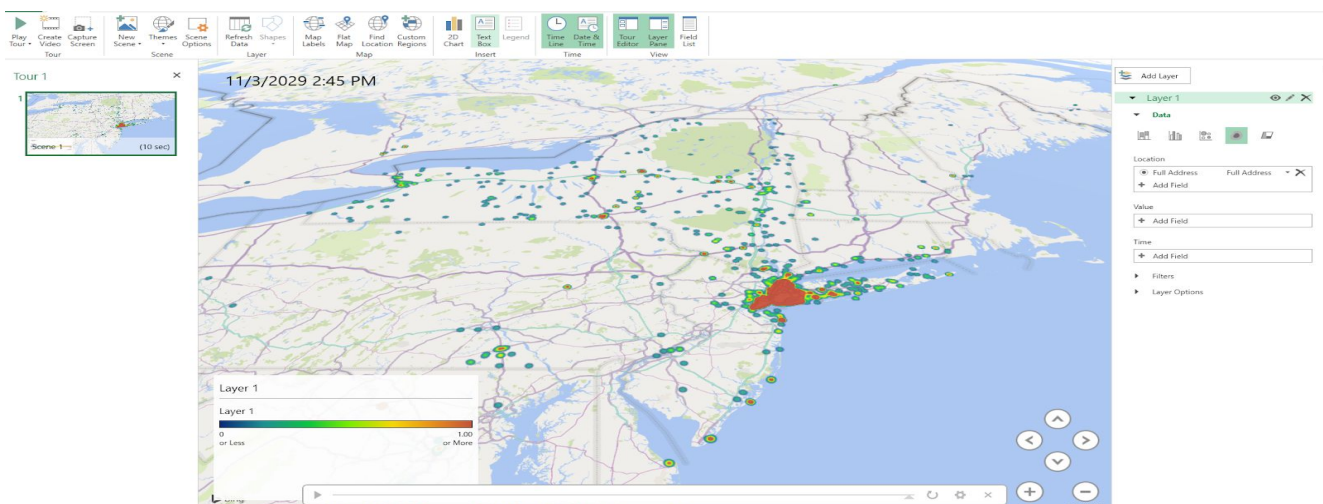
We downloaded data from the NYC open and from there put got all the csv files into one by merging them. Once that is finished we uploaded the file into hdfs using data proc. We use HiveQL to create tables then export the table into a csv file. Using pscp download the file and load into software like excel and analyze that data using 3D map.

We were also able to use the SQL in Hive and find out what times have it's most likely to get a ticket which was from 8am to 12pm, and the least likely which is around 4am. We also found out the most common reasons for ticket is failure to display meter ticket, street cleaning, speed near school zone, and failure to stop at a red light.



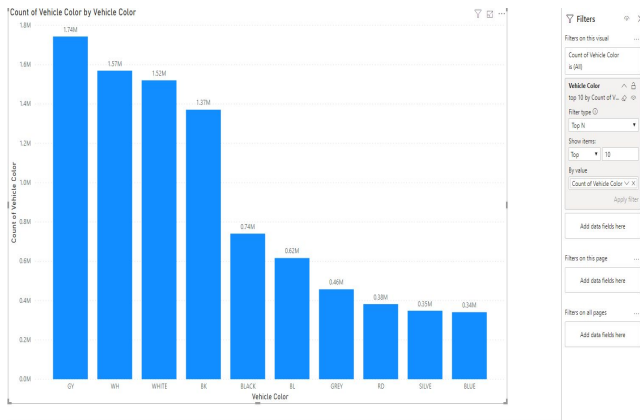
**Figure 2:** This illustrates our visual 3D map we created so that we can see the area to where there are the most tickets from the data set.

In *Figure 2* we see that most of the addresses in the map were in the west part of New York but we also see that some of the addresses are coming from more eastern areas.



**Figure 3:** This illustrates our visual 3D map but this time with heat map so that we can see the area to where there are the most tickets from the data set in a different way.

With this second map (*Figure 3*) we can see clear the red spot which is New York and see there are a few addresses around it but not the majority of the tickets.



**Figure 4:** Count for vehicle color

With Power BI (*figure 4*) charts we were able to visualize what car color was ticketed the most which was gray.

## 5. Conclusion:

In conclusion, we found a lot of interesting information in our project such as where tickets were given throughout NYC and be able to tell the patterns on where and why the tickets were given in certain locations in NYC. We learned a lot through research on how different people used the NYC open database to see information such as two fire hydrants generating more than \$50,000 a year in tickets. Besides using big data people have used R and ggplot to help visualize the parking tickets in NYC.

We learned that there are times where drivers need to be more careful. We found similar reasons that drivers in Los Angeles get tickets like expired meters and street cleaning are also one of the most common violations that people who park in New York are experiencing [4]. We displayed two 3D maps using excel to give a visual of where the majority of the ticket addresses are which is something similar to what Sinclair did in his research [6]. This paper is to help increase the awareness of what to look out for when parking and to share that there are ways we can find and learn information on this topic ourselves.

## References:

[1] Halime, Farah. "Ben Wellington, Open Data, and 'I Quant NY'." *OZY*, 1 Sept. 2015, <https://www.ozy.com/rising-stars/ben-wellington-open-data-and-i-quant-ny/37479/>.

[2] Johnson, Phil, and Phil Johnson. "Open Data Saves New York City Drivers from Parking Tickets." *ITworld*, ITworld, 4 June 2014, <https://www.itworld.com/article/2695542/open-data-saves-new-york-city-drivers-from-parking-tickets.html>.

[3] Ginzberg, Steven "Visualizing New York City's Parking Violation Data." *NYC Data Science Academy*, <https://nycdatascience.com/blog/student-works/visualizing-new-york-citys-parking-violation-data/>.

[4] Ward, E., & Huang, C. (2019, September 5). "Parking citations dipping, but tell that to folks who park Downtown." Retrieved from <https://xtown.la/2019/08/28/parking-citations-down-but-tell-that-to-folks-that-parked-downtown/>.

[5] Song Gao, Mingxiao Li, Yunlei Liang, Joseph Marks, Yuhao Kang & Moying Li (2019) "Predicting the spatiotemporal legality of on-street parking using open data and machine learning", *Annals of GIS*, 25:4, 299-312, DOI: 10.1080/19475683.2019.1679882

[6] Sinclair, C. (2019, January 21). "Avoiding Parking Tickets in San Francisco Using Data Analytics." Retrieved from <https://towardsdatascience.com/avoiding-parking-tickets-in-san-francisco-using-data-analytics-33ff5fbc2153>.

[7] Mays, J. (2016, March 24). "New York City Collects Record \$1.9 Billion in Fines and Fees." Retrieved from <https://www.dnainfo.com/new-york/20160324/civic-center/new-york-city-collects-record-19-billion-fines-fees/>.

[8]Smith, C. (2019, September 4). Parking Ticket Revenues At 16 Major U.S. Cities Tops \$1.4B Annually. Retrieved from <https://www.motor1.com/news/353183/parking-ticket-revenues-major-cities/>.

[9] Department of Finance. (2017, November 15). Parking Violations Issued - Fiscal Year 2014: NYC Open Data. Retrieved from <https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2014/jt7v-77mi>.

[10] Department of Finance. (2017, September 14). Parking Violations Issued - Fiscal Year 2015: NYC Open Data. Retrieved from <https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2015/c284-tqph>.

[11]Department of Finance. (2017, September 14). Parking Violations Issued - Fiscal Year 2016: NYC Open Data. Retrieved from <https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2016/kiv2-tbus>.

[12] Department of Finance. (2017, August 10). Parking Violations Issued - Fiscal Year 2017: NYC Open Data. Retrieved from <https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2017/2bnn-yakx>.