NEW YORK PARKING TICKETS YEARLY ANALYSIS

AAKANKSHA TASGAONKAR
AMOGH MAHESH
RUPA MAHENDRAN
SRIGANESH BASKARAN
MSIS graduate students
California State University, Los Angeles

Abstract: This project will demonstrate the usage of Hadoop, MapReduce, and Hive on big data. We will apply the knowledge learned during the lecture, extensive researches and development of HiveQL in order to generate data and visualize it on Power BI, Tableau and 3D maps. We are using NYC Parking Tickets of past four years as the foundation to generate results. The Department of Finance is responsible for collecting and processing payments for all parking tickets and camera violations. The NYC Department of Finance collects data on every parking ticket issued in NYC (~10M per year). Fundamentals of this project include a report paper, a tutorial on the queries, and one group presentation.

URL: https://www.kaggle.com/new-york-city/nyc-parking-tickets
Dataset size: 8GB
Cluster version: IOP4.2
No of nodes: 5
Memory size: 32GB
CPU Speed: 2.195 GHz

1. Introduction

Based on the list of data provided by our instructor, we have done some researches and exclusively decided which data we are using for this project. We are going to manipulate and filter the datasets below following with step:

- New York City Parking Tickets; data size is 8GB.
- Cleaning down the information to have a detailed comparison between the years.
- From each dataset, sorted out the type of violations, to see which locations those violations happened often, type of the vehicle which created violations, and also analyzed the year when the violation was high.
- The tools we are using is HiveQL, Putty, Oracle Cloud, tableau, excel 3D maps and Power BI.

2. Manipulating datasets

2.1 Tools and data processing

- We extracted our parking ticket data from the corresponding website in .csv format: New York parking tickets from Kaggle.com.

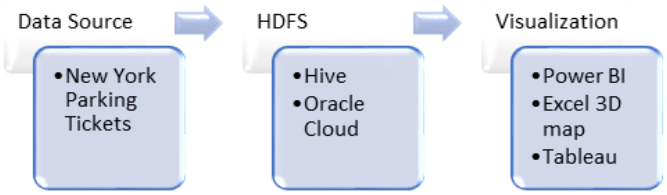

Figure 1: Data Processing

- We use basic commands to connect to the data source such as pscp, mkdir, and also uploaded our data to HDFS and used Hive queries to create external tables on the .csv data.
- Then, we used Hive queries to select the desired data from the external table and filtered out unimportant data. (cleaned the data)
- We used hive queries to analyse the data
- Lastly, we used Power BI, Tableau and 3D maps to reproduce the selected data in the form of information by generating the appropriate graphs, maps and chat.

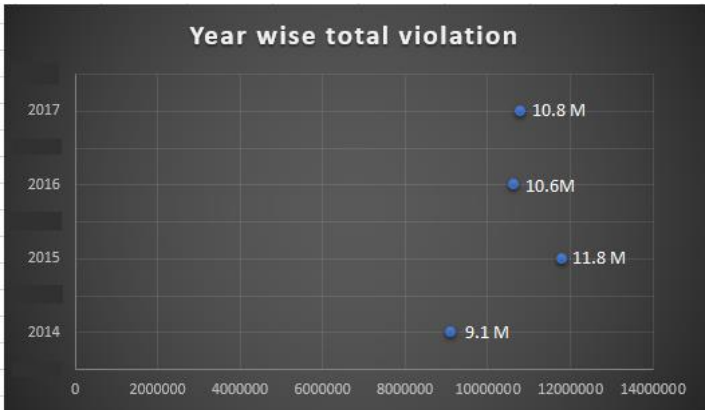

Figure 2: Year Wise Overall Violation

The above graph shows the year wise overall violation. The maximum number of violations was in the year 2015 with the count of 11.8M whereas, the minimum was in the year 2014 with the count of 9.1 M. Last year, the violation was approximately same as 2016.

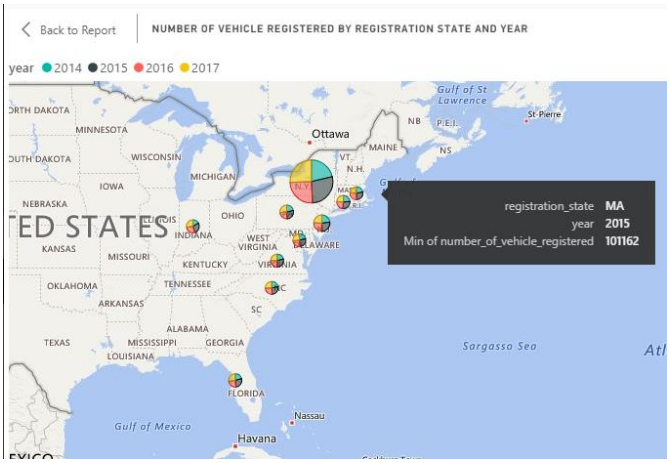

Figure 3: Vehicle registration

By analyzing the above pie chart, the state which acquired the most number of tickets based on vehicle registrated state is New York. We have also analyzed on the yearly basis that it is in the year 2015 New York has attained the maximum vehicle registrations.

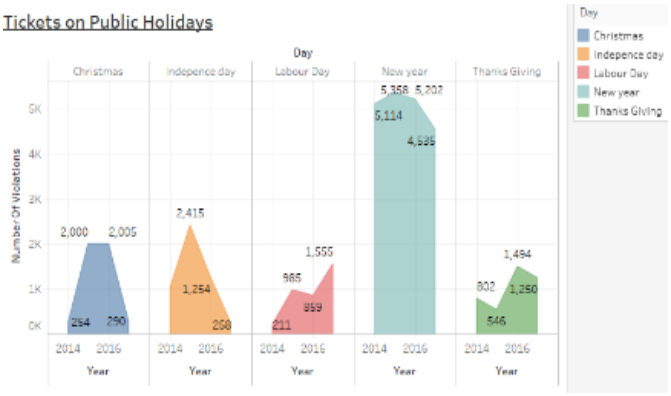

Figure 4: Tickets on public holiday

As we observe from the above area graph, we can conclude that the maximum number of violations are seen in the New Year period among all the public holidays. If we compare the number of violations tickets, we get to know that it has been doubled on the New Year in comparison to the Independence Day.

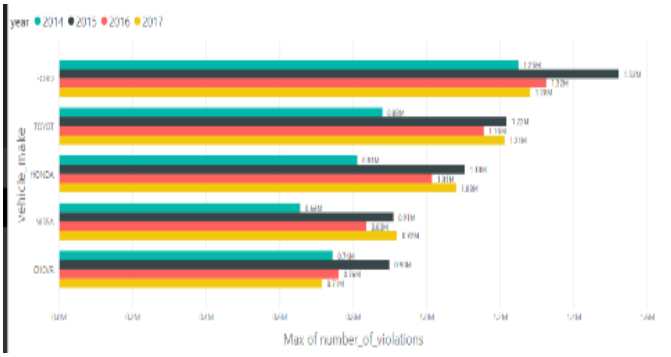

Figure 5: Tickets according to the vehicle companies

We have also analyzed the violation rate yearly to discover the trend of violation among the vehicle types. Ford has the highest number of violations in all the years. Since the number of ford customers is high therefore the violation is at its peak.

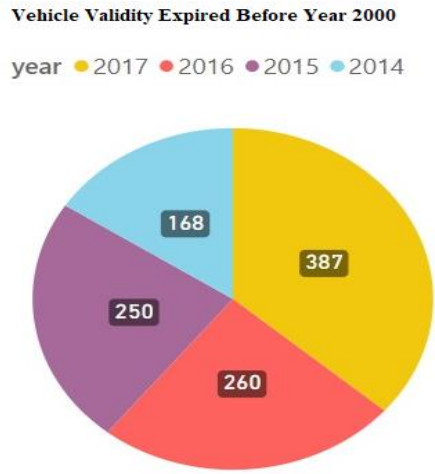

Figure 6: Vehicle expired bedore year 2000

This above pie chart represents the vehicles that are expired before the year 2000. Based on the analysis, it is noticed that a greater number of expired vehicles were used in the year 2017.On the same, the least number of expired vehicles were used in the year2014.

Figure 7: Most violation street 2014

With respect to the analysis based on the top 5 violated locations in the year 2014, it is visualized that 16E 58th street, NY is the top most violated street followed by 55E 59th street, NY.

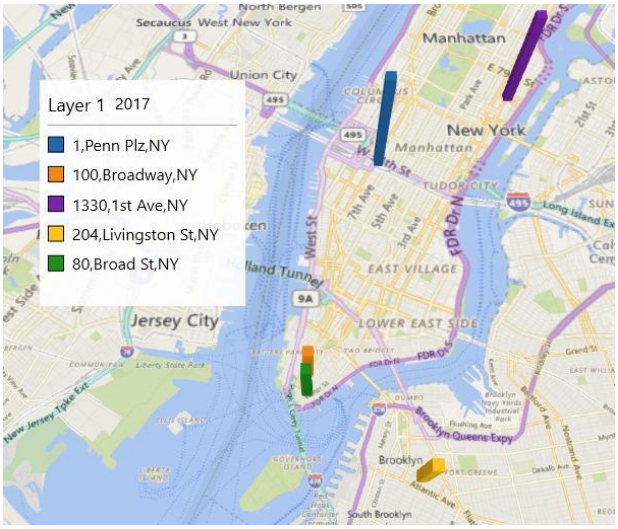

Figure 8: Most violated street 2017

To put the whole result in a deeper perspective, we wrote queries to analyze that 1, Penn Plz, NY has been marked as the most violated street in the year 2017.

2.2 Comparing the analysis

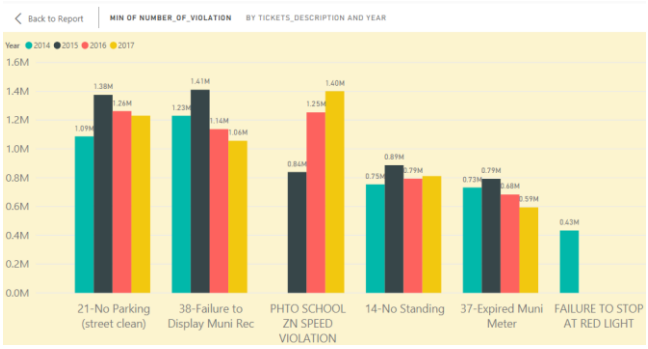

Figure 9: Year wise violation and ticket description

Until this point of the data analyzing process, we realized that in each individual violation listed during the year 2015, No parking and Failure to display muni record are having approximately same number of violation tickets. We observed that the number of violation tickets for the failure to stop at red light can only be seen for the year 2014. The number of violations got doubled for the year 2017 as compared to 2015 for the phto school zn speed violation.

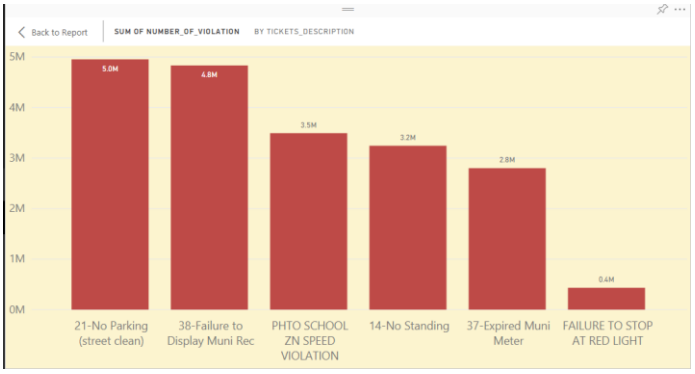

Figure 10: Sum of violation by ticket description

The above analysis illustrates the sum of the number of violations for different categories of tickets. The highest sum of number of violations are seen in no parking street whereas the least has been seen for failure to stop at red light.

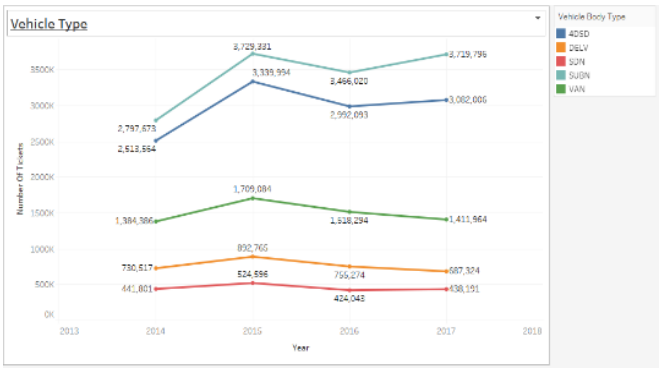

Figure 11:Vehicle type

As observed, there was a minute change in the number of tickets between the consecutive years for the sedan vehicles. On the contrary, there was a drastic difference in the number of tickets during 2014-2015 caused by the suburban vehicles. At the same time, the sedan vehicles made the minimum number of tickets (424,043) in 2016, whereas the suburban vehicles made the maximum number of tickets (3,729,331) in 2015. To conclude, the suburban vehicles got the highest number of tickets in all the consecutive years.

3. Summary

- We successfully used many tools learned in class such as HiveQL, IBM Bluemix, and Tableau to use and manipulate data.
- The parking violation has been gradually reduced as compared to previous years.
- The suburban vehicles has been marked high for its parking violations.

4. Github URL

https://github.com/amoghmahesh/hiveanalysisonnycparkingticket

5.Reference

https://www.kaggle.com/new-york-city/nyc-parking-tickets

http://www.nyc.gov/html/dof/html/pdf/faq/stars_codes.pdf

https://www.kaggle.com/donyoe/exploring-42-3m-nyc-parking-tickets/notebook

https://www1.nyc.gov/site/finance/vehicles/services-violation-codes.page