

NYC Parking Ticket - EDA - Case Study

Submitted By : *Srinivasan. G, Naveed.J, Denny.J and Kumar.A*

Date of Submission : *24-June-2019*

Batch : *September – 2018*

Objectives

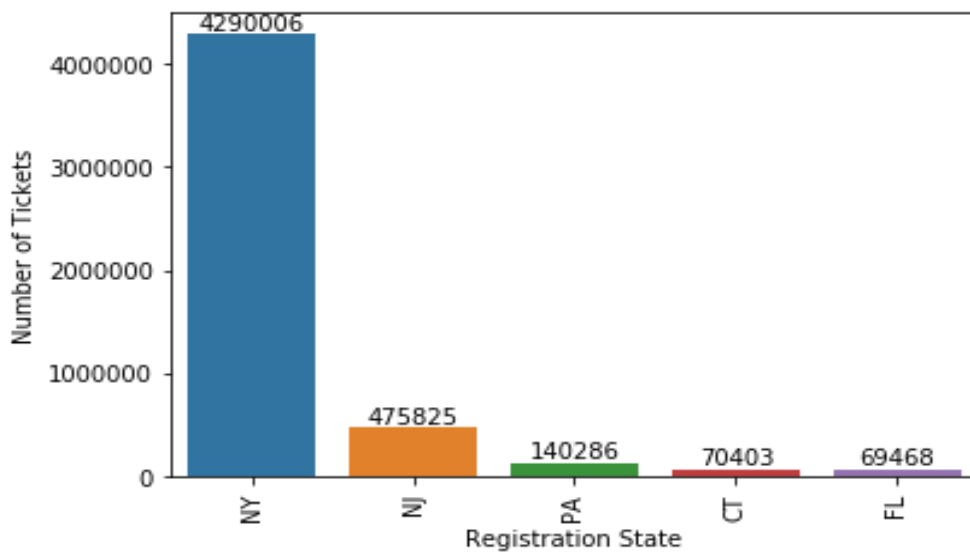
The objective of this case study is to do exploratory data analysis using pyspark and drawing inference from the data, in big data clustered distributed environment. Also one of the main objective is to gain familiarity with how data analysis works in pyspark.

Our Analysis is focussed on performing exploratory data analysis to understand the parking violation dataset. We will compare metrics/phenomenon related to Parking Tickets for the fiscal year 2017. The Analysis will be performed using Pyspark running in corestack with data in Hadoop HDFS Cluster.

A. Examine the data

1. The number of parking tickets issued during year 2017 is approximately 5,431,918 Million Tickets. (i.e, 5.4 Million Tickets).
2. The number of unique states of the registered plate of cars receiving parking tickets in the year are 65. However, there are 16055 records with state code 99. This implies that there is several parking tickets issued to vehicle less with an unknown registration state. For the sake of case study, we considered these unknown state code as 'NY'. So, the maximum parking tickets received by vehicles having registered plates are 64 and the state is 'NY'.

Plot-1 - Top 5 Registration State with Highest Number of Tickets

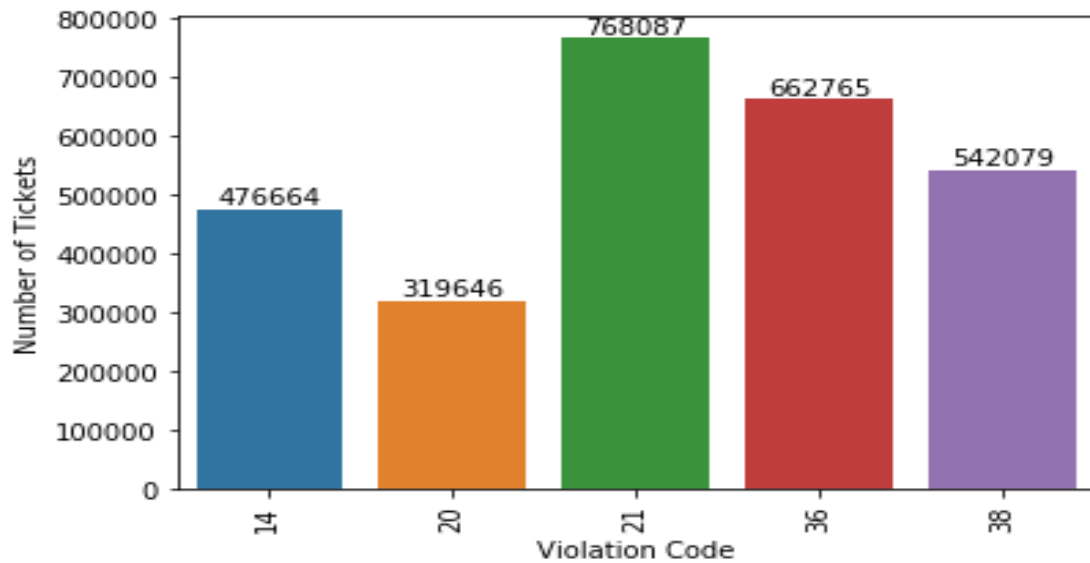


B. Aggregation tasks

1. The frequency of top 5 Violation Codes high during the year 2017 are for ;

- 21 - No parking where parking is not allowed by sign, street marking or traffic control device.
- 36 - Exceeding the posted speed limit in or near a designated school zone.
- 38 - 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 - Standing or parking where standing is not allowed by sign, street marking or; traffic control device.
- 20 - No parking where parking is not allowed by sign, street marking or traffic control device.
- Therefore, the highest frequency of tickets through the years is issued for No-parking violations, Exceeding Speed limit near school zones and Failing to show tag in the windsheild.

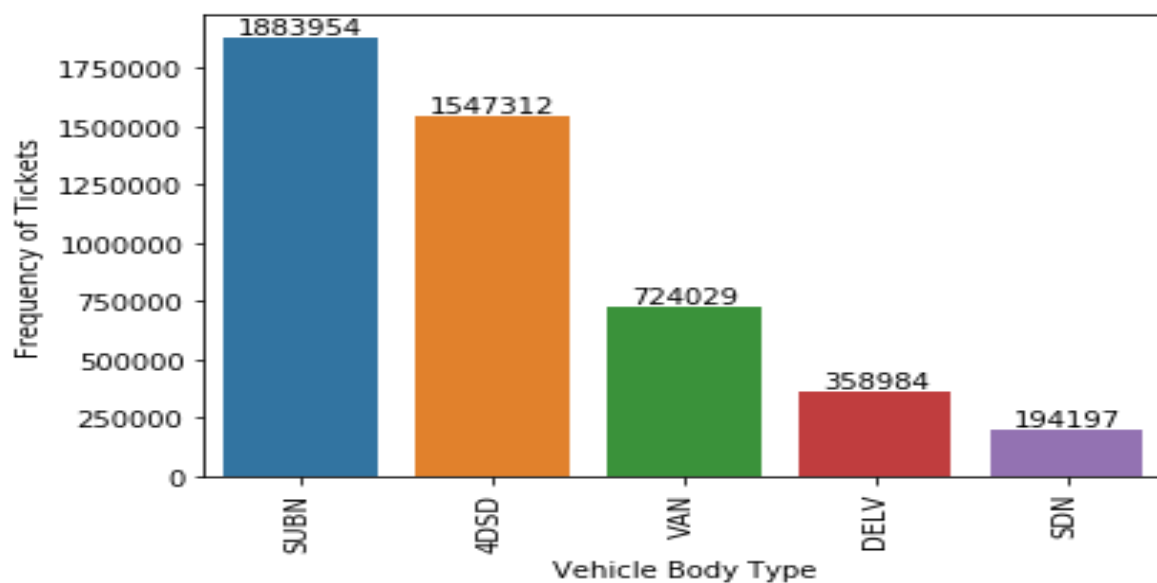
Plot-2 - Top 5 Violation Codes in Year 2017



2. The top 5 Vehicle Body Types issued with parking tickets in 2017 are ;

- SUBN - Vehicles that can be registered with the suburban body type include station wagons, sport utility vehicles, hearses and ambulances.
- 4DSD - 4 door Sedans
- VAN - Type of vehicle used for transporting goods or people.
- DELV - Delivery Truck
- SDN - Civilian Sedan

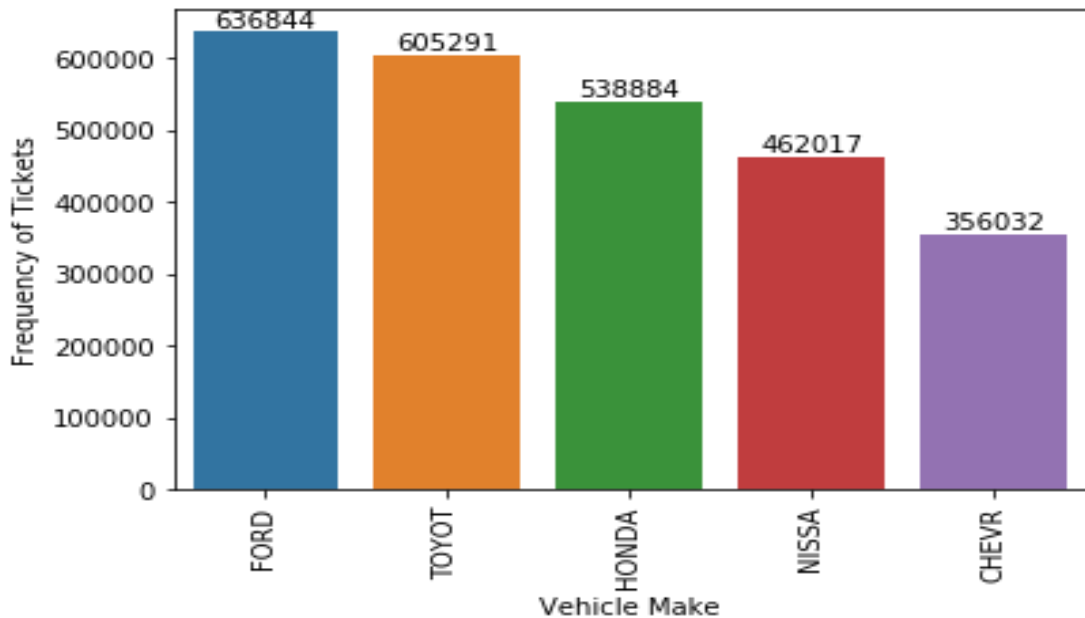
Plot-3 - Top 5 Vehicle Body Type Violations in Year 2017



- The top 5 Vehicle Make having issued with parking tickets in 2017 are ;

- Ford, Toyota, Honda , Nissan and Cheverlet.

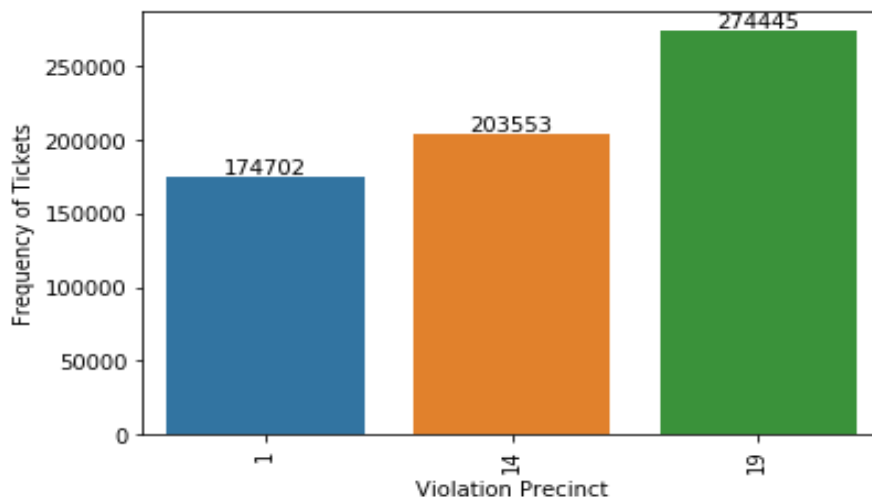
Plot-3 - Top 5 Vehicle Make Violations in Year 2017



3.1. From the analysis of violation of Precinct against the number of parking tickets ;

- it is clear the highest number of tickets were issued with Violation Precinct 0. Since 0 is unknown precinct it may be issued with faulty tickets. The Violation Precinct codes are 19,14 and 1.

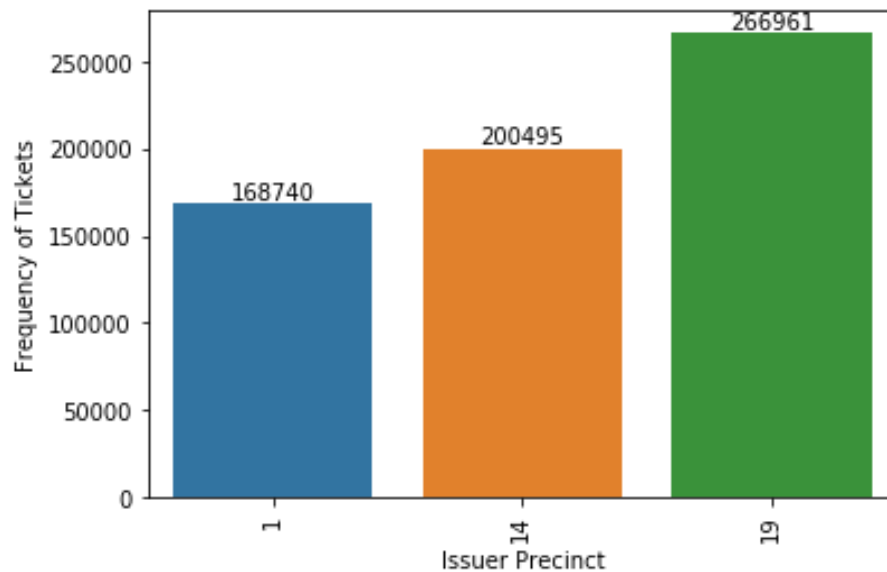
Plot-5 - Top 3 Police Station Zones where Violations Occured in Year 2017



3.2. From the analysis of Issues Precinct against the number of parking tickets ;

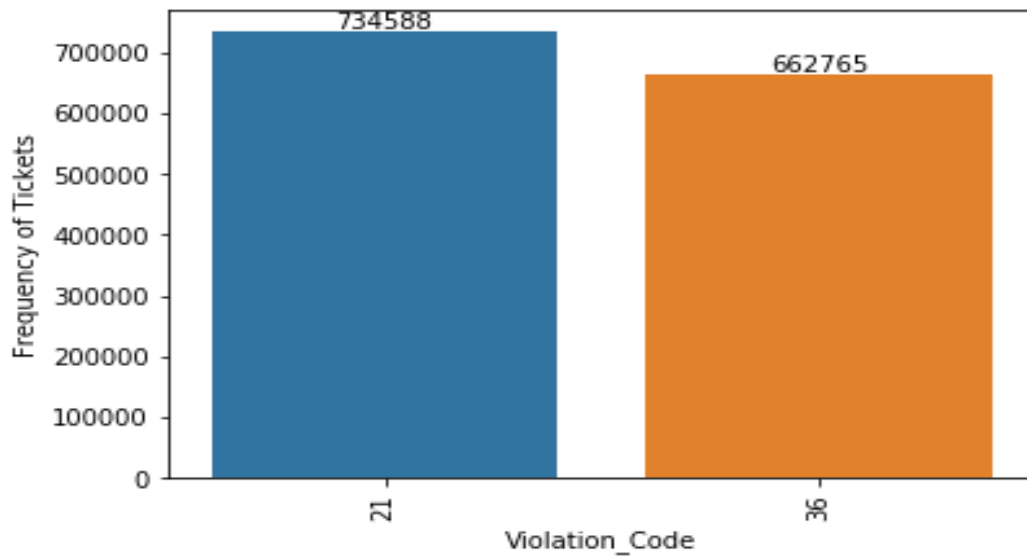
- Reveals the highest number of tickets are issued for code 0. Again, there is no code as 0 , the validity of the tickets raised are questionable. Other top issue Precinct codes are 19,14 and 1 .

Plot-6 - Top 3 Police Station Zones where Tickets where Issued in Year 2017



4. While observing the Top-5 violation codes for which tickets were issued in each of the above 3 Issuing Precincts (19,14 and 1) and comparing the results with all other Issuing Precincts for the year 2017, the violation codes 21 and 36 have exceptionally high count of tickets.

Plot-7 - Top 2 Violation Codes Other than 19,14,1 in Year 2017



5.1. There are no NULL values in the data.

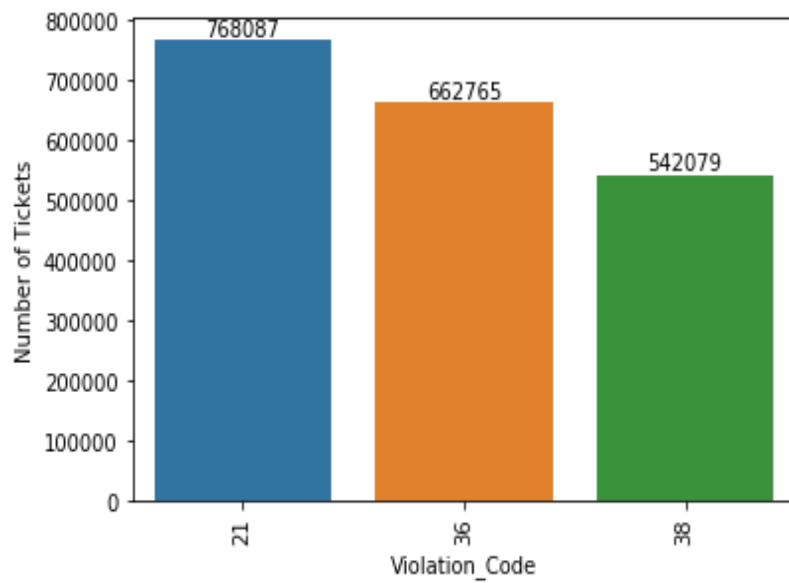
5.2. The Violation Time has been transformed into a suitable timestamp format and Divide 24 hours into 6 equal discrete bins of time.

5.3. While observing the Frequency of Top-3 Violation Codes in each Violation Time Bin it is clear that the majority of the Tickets are issued between 0800-1100 Hrs and 1200-1500 Hrs. It is also important to note, the exceptionally high number of tickets for Violation Code 21. This is expected as Code 21 stands for No- Parking Zone Tickets. Majority of the public might park inappropriately during the morning rush, issued between 0800-1100 Hrs.

5.4. The Top-3 Violation Codes for 2017 are;

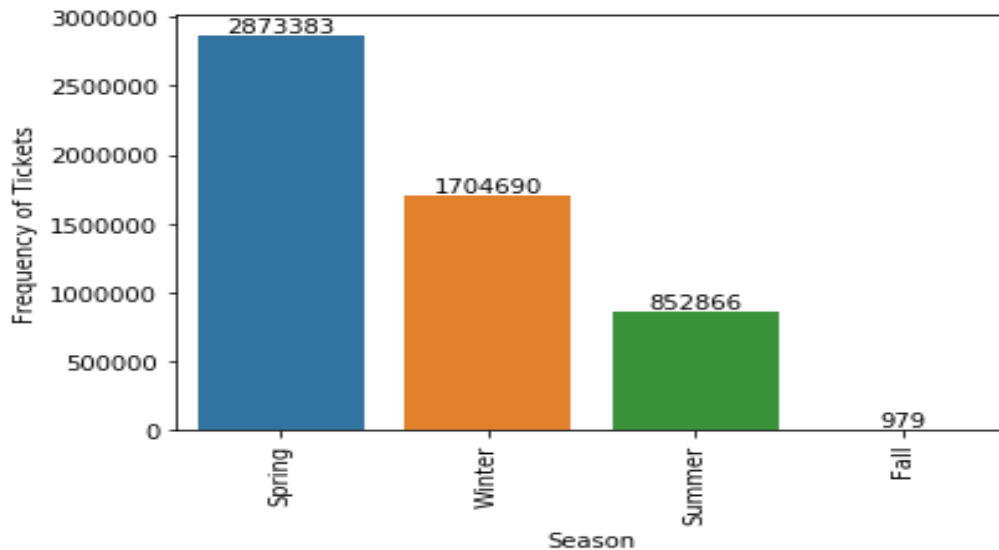
- 21: No Parking Violation
- 36: Exceeding speed limit near School Zone
- 38: Failing to show a receipt or tag in the windshield
- While looking at the distribution of Violation Code 21, there is noticeably a significant peak between 0800-1100 Hrs and dips between 1600-2300 Hrs. For Violation Code 36 there are almost equal peaks between 0800-1100 Hrs and 1200-1500 Hrs this may be as a result of people rushing to drop and pickup children to and from school.

Plot-8 - Top 3 Violation Codes in Most Common Violation Time Period 8 AM to 11 AM in Year 2017



6. After dividing the Issue Month into number of seasons, for each season and for the year 2017, Spring and Winter account for the highest number of parking tickets.

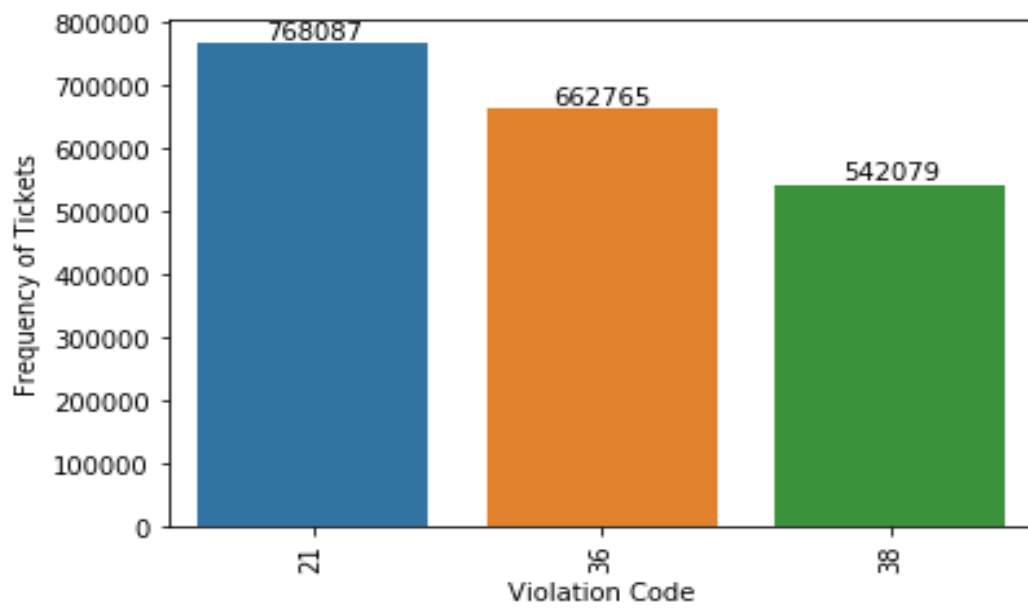
Plot-9 - Number of Violations Per Season in Year 2017



7.1. The Top three most Violation Codes are ;

- 21: No Parking Violation,
- 36: Exceeding speed limit near School Zone,
- 38: Failing to show a receipt or tag in the windshield

Plot-10 - Top 3 Most Commonly Occurring Violations in Year 2017



7.2. In the year 2017, Violation Code 21 brought the highest total fine amount \$42,244,785.00 Millions(42.2 Millions).