New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

Team Members
- Gaurav Sachdeva
- Sriram Pala
- Parth Samanta
- Meenakshi Lakshmanan

Purpose:
This case study mainly focuses on the usage of Spark and gaining familiarity with how to use SparkR and do exploratory data analysis on Big Data.

Problem Statement:
Parking is one of the biggest problems which New York City faces, combination of a huge number of cars and cramped geography leads to a huge number of parking tickets.

Objectives:
Perform exploratory data analysis on the dataset collected by NYC Police Department for parking tickets.

Dataset:

For the scope of this analysis, we analyze the parking tickets over the year **2017**. The Analysis will be performed on the R-studio mounted on Corestack cluster, using the SparkR library.
The data for this case study is available in HDFS at the following path:
'/common_folder/nyc_parking/Parking_Violations_Issued_-_Fiscal_Year_2017.csv'

Stages in EDA:
1. Setting the Sparkr platform
2. Data Cleaning
3. Examining the Dataset
4. Analyzing the dataset

Pre-requisite:
1. Load the SparkR package with essential commands
2. Add the jar file hive-sql query from RStudio, you need to add a jar file in RStudio
3. sql("ADD JAR /opt/cloudera/parcels/CDH/lib/hive/lib/hive-hcatalog-core-1.1.0-cdh5.11.2.jar")
4. Load the data into the dataset from the HDFS path
   "/common_folder/nyc_parking/Parking_Violations_Issued_-_Fiscal_Year_2017.csv"

Assumption and Data Handling:
1. All data other than the year 2017 was filtered out.
2. Erroneous entries in Registration State, Issuer Precinct, and Violation Precinct were handled.
3. All the NA values in the data columns were checked. None was found.

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

4. The column names which have spaces in the data was handled
5. Seasons were created for analysis of the data
    Spring Season (Bin Number:1) - March, April and May
    Summer Season (Bin Number:2) - June, July and August
    Fall Season (Bin Number:3) - September, October and November
    Winter Season (Bin Number:4) - Dec, Jan and Feb
6. Time was extracted with the date that was provided and the analysis was done after creation of following bins
    WHEN HH >= 0 and HH <= 3 THEN Late Night(Bin Number:1)
    WHEN (HH >= 4 and HH <= 7) THEN Early Morning(Bin Number:1=2)
    WHEN (HH >= 8 and HH <= 11) THEN Morning(Bin Number:3)
    WHEN (HH >= 12 and HH <= 15) THEN Afternoon(Bin Number:4)
    WHEN (HH >= 16 and HH <= 19) THEN Evening(Bin Number:5)
    ELSE Late Evening(Bin Number:6)

Analysis and Findings:

1. A total of 5431918 tickets have been issued by the New York police department for the parking violations. The tickets were issued to vehicles that came from various states, and through the Registration State data that is available we can infer that vehicles from 64 states have been issued tickets by the New York Police Department.

2. Of all the violation codes violation code 21 has the highest frequency with 768087 tickets which is 14% of all the tickets.
The top 5 violation codes identified from the code are below.

Violation Code	COUNT
21	768087
36	662765
38	542079
14	476664
20	319646

21,20,14 is the code for parking in areas not meant for parking,36 for speeding and 38 for parking beyond designated time. Majority tickets are issued for parking related offences.

3. From the data it was found out which vehicle body type and vehicle make has the highest number of tickets from NY Police department. Below are the top 5 vehicle body and make that has got the highest number of tickets.

VehicleBodyType	COUNT
SUBN	1883954
4DSD	1547312
VAN	724029

DELV	358984
SDN	194197

SUBN and 4DSD are majorly pickups for delivery and transportation and are worst offenders.

VehicleMake	COUNT
FORD	636844
TOYOT	605291
HONDA	538884
NISSA	462017
CHEVR	356032

4. The precinct under which the violations occurred the most is measured from the data. And the below are the top 5 precincts with highest violations.

ViolationPrecinct	COUNT
19	274445
14	203553
1	174702
18	169131
114	147444

The precinct which issued the highest number of tickets is also measured from the code and below are the results.

IssuerPrecinct	COUNT
19	266961
14	200495
1	168740
18	162994
114	144054

Offenders are caught majorly in the precinct where violation is committed.

For the precincts which issued the highest number of tickets (top 3) we want to see which violation codes are higher and if there is any resemblance in terms of the violations.

For precinct 19 following are the violation codes

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

ViolationCode	COUNT
46	48445
38	36386
37	36056

For precinct 14 following are the violation codes

ViolationCode	COUNT
14	45036
69	30464
31	22555

For precinct 1 following are the violation codes

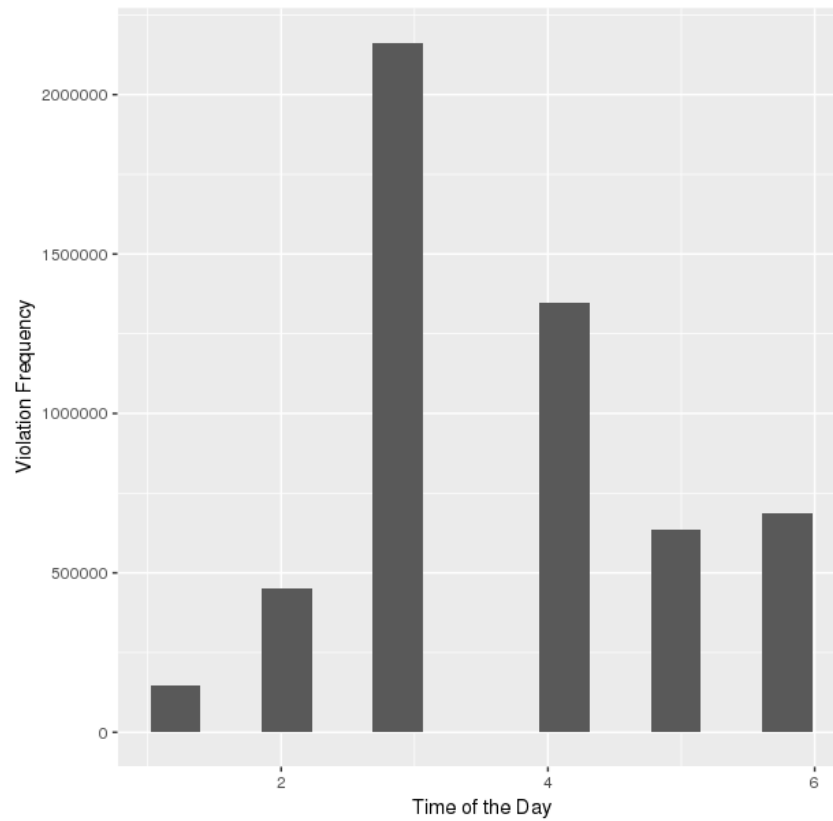| ViolationCode	COUNT
14	38354
16	19081
20	15408

It is observed that violation code 14 is higher in the precinct 14 and 1 and the numbers are also considerably higher than the next highest violation codes. 14 is a general no standing offence.

5. We also analyzed the parking violations across different times of the day.

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

Violations are the most during the third bin which is "Morning". This indicates the time when the maximum vehicles are plying on the road.

New York Parking Case Study

Late night

ViolationCode	COUNT
21	34704
40	23629
14	14168

Early Morning

ViolationCode	COUNT
14	74114
40	60652
21	57897

Morning

ViolationCode	COUNT
21	598070
36	348165
38	176570

Afternoon

ViolationCode	COUNT
38	184829
36	184293
37	130692

Evening

ViolationCode	COUNT
38	102855
14	75902
37	70345

Late Evening

ViolationCode	COUNT
36	101991
38	76314
21	72568

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

We also checked the times of the day during which the top three violation codes (21,36,38) occur

For Violation code 21

Time of the day	COUNT
Morning	598070
Late Evening	72568
Early Morning	57897
Late Night	34704
Afternoon	4589
Evening	259

For violation code 36

Time of the day	COUNT
Morning	348165
Afternoon	184293
Late Evening	101991
Early Morning	14782
Evening	13534

For Violation code 38

Time of the day	COUNT
Afternoon	184829
Morning	176570
Evening	102855
Late Eveninig	76314
Early Morning	1273
Late Night	238

Violations are the most during "Morning". This indicates the time when the maximum vehicles are plying on the road.

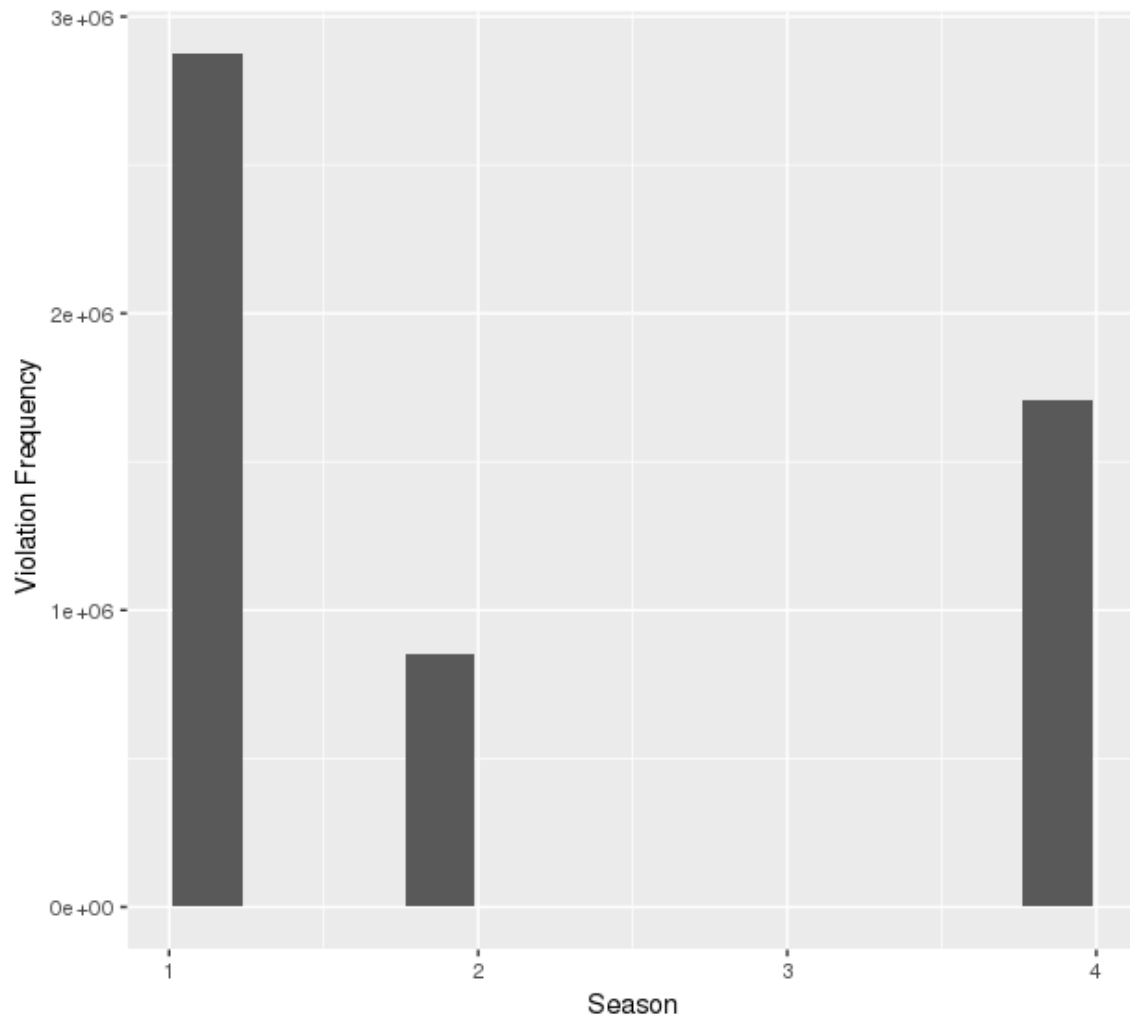6. Analysis on the seasons of the year 2017 during more violations occurred.

    We divided the year to the following 4 seasons and following are the number of tickets issued during these period.

Season	COUNT
Spring	2873383
Winter	1704690
Summer	852866
Fall	979

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

Spring is the season when most violations occur. This is probably the time when the people move outdoors.

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

For each of this season we found the top three violations and below are the results.

Spring

ViolationCode	COUNT
21	402424
36	344834
38	271167

Summer

Violationcode	COUNT
21	127352
36	96663
38	83518

Fall

ViolationCode	COUNT
46	231
21	128
40	116

Winter

ViolationCode	COUNT
21	238183
36	221268
38	187386

It is observed that violation code 21 is the highest across three seasons spring, summer, winter.
And also violation code 36 is the second highest highest across three seasons winter,summer, spring.

We also estimated the revenue that the police department is making with the top three violations.

New York Parking Case Study

Submitted by: DSWarriors   **Date:10/03/2019**

7. Upon checking on NYC website, Violation code 21, 36,38 have separate fines for Manhattan vs other areas. We calculated the average of the fines

   Avg fine for Violation Code 21=(65+45)/2=55
   Avg fine for Violation Code 36=(50+50)/2=50
   Avg fine for Violation Code 38=(65+35)/2=50

   Total fine collected for violation code 21 is 42244785 dollars (approx $42.25 Mn)
   Total fine collected for violation code 36 is 33138250 dollars(approx $33.13 Mn)
   Total fine collected for violation code 38 is 27103950 dollars(approx $27.1 Mn)

   Since they are parking related offences improper parking also brings in significant funds to the NYC