# ##### --**NYC Parking Case Study:Apache Spark**-----####

**##### This is a documents which is showing our EDA analysis of the data sets and the results ###**

Group Name:**Upgrad Pune- Bhubaneswar Cohorts**

Rishabh Srivastava

Vinod Jha

Saurav Kumar

Amitabha Banerjee

## ## **Problem Statement** ###

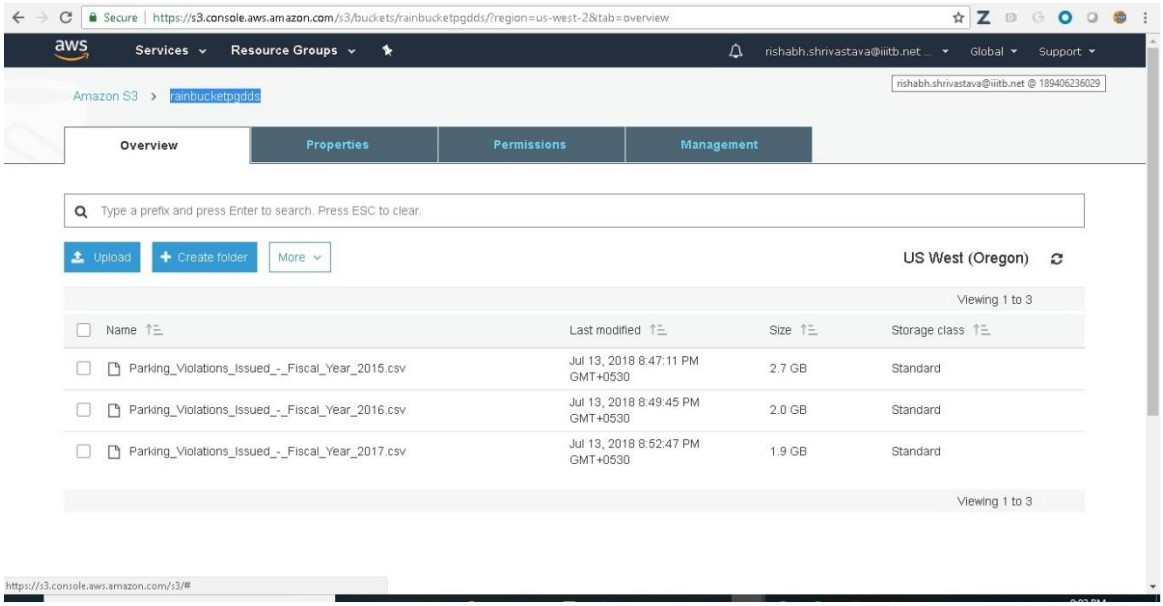
New York City is a thriving metropolis. Just like most other metros of that size, one of the biggest problems it's citizens face, is parking. As it's a busy metropolis due to a huge number of cars lot of parking tickets are issued for multiple reasons . To understand and analyse the reasons behind the parking tickets NYC Police dept has collected data from 2015-17 for a period of 3 years

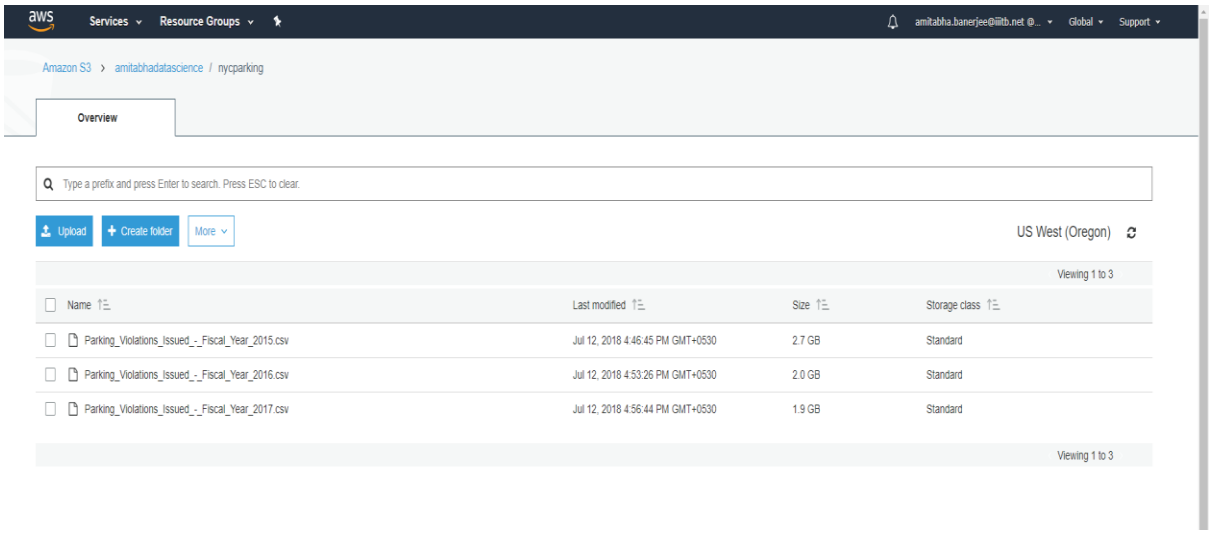
## **##--------- Business Objective ----------###**

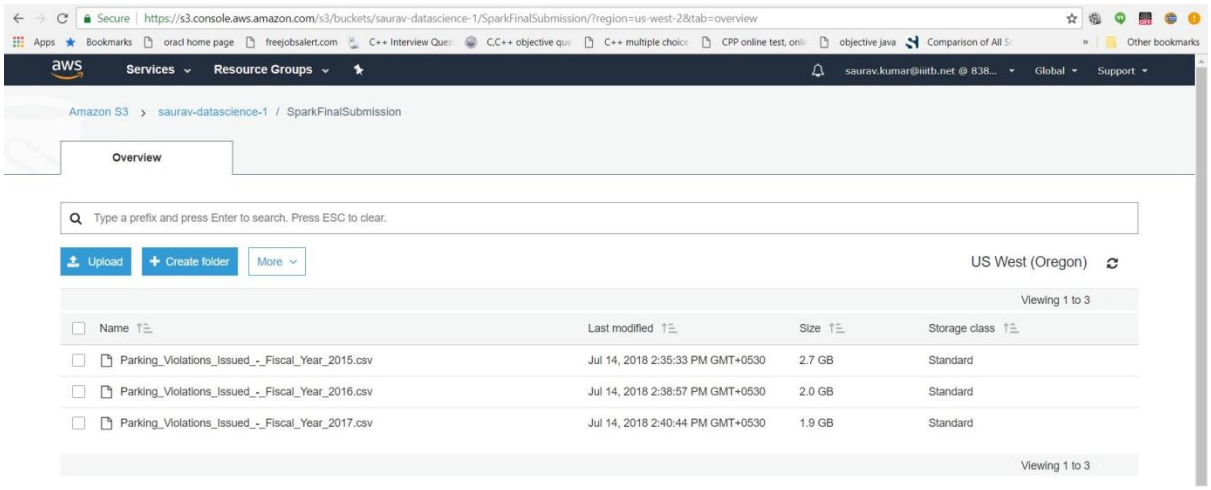
Perform EDA for 3 years that helps in understanding the data and find out the major reasons responsible for so many parking tickets

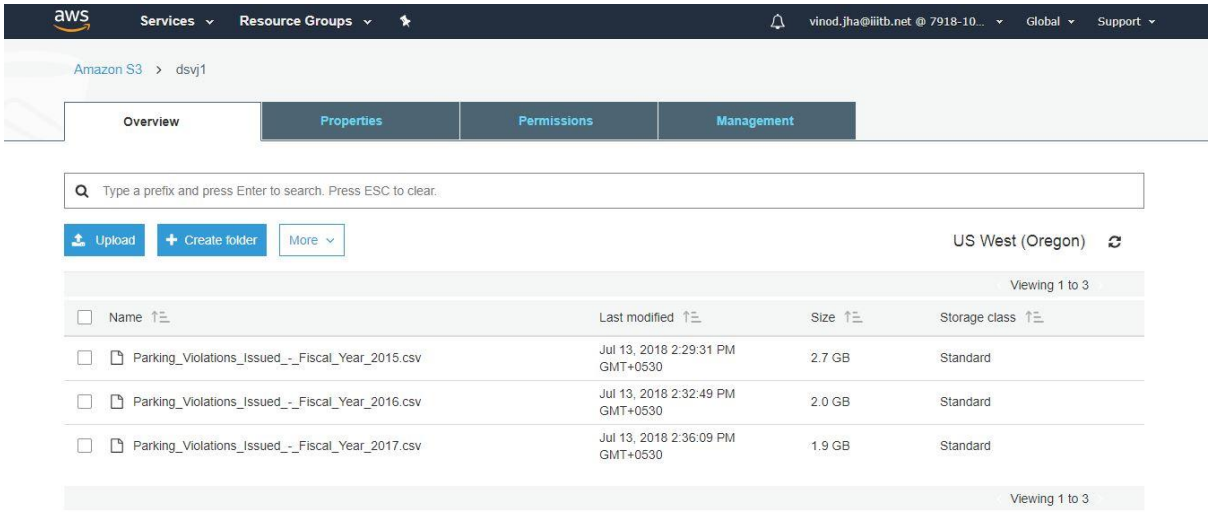
- Start the spark session by loading the SparkR data and read the data
- Check the structure of the dataset

### ----------S3  Bucket Screenshot of Group members---------- ####






**## Data frames for analysis ##**

We have stored the three years data( 2015-17) in three different data frames nyc_tkt15,nyc_tkt16 & nyc_tkt17. Total number of records are mentioned below.

| Years | Number of rows | Number of columns |
|---|---|---|
| 2015 | 10951257 | 51 |
| 2016 | 10626899 | 51 |
| 2017 | 10803028 | 43 |

**###----------DATA CLEANING & VALIDATION----------#######**

- Checked for Duplicate values and removed the duplicate records
- Checked for Missing Values
- Changed the column names in all the data frames to bring uniformity
- Checked for NA's

- Keeping the relevant columns required for analysis
- Checked for outliers in different columns

Post removal of duplicate records the total count is –

| Years | Number of rows |
|---|---|
| 2015 | 10951257 |
| 2016 | 10626899 |
| 2017 | 10803028 |

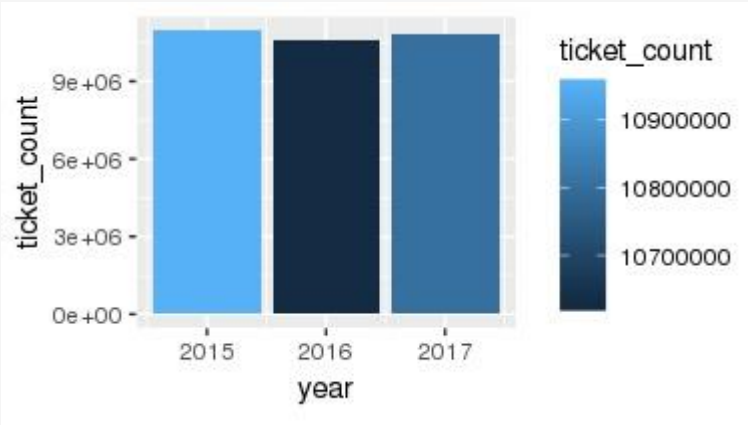
## #######----------Data Analysis Questions----------###############

**Examine the Data**

1. Find total number of tickets for each year.

| Years | Total number of tickets |
|---|---|
| 2015 | 10951257 |
| 2016 | 10626899 |
| 2017 | 10803028 |

## Plotting of Yearly ticket counts




**Observations –The count decreased in 2016 and showed a spike again in 2017 but still lower as compared to 2015. 2015 was the year when maximum number of tickets were issued.**

2. Find out how many unique states the cars which got parking tickets came from.

| Years | Total number of unique states |
|---|---|
| 2015 | 69 |
| 2016 | 68 |
| 2017 | 67 |

Observations –This is showing a decreasing trend.. One interesting fact is that apart from 50 US states we observed few more values for Mexico & Canada as well.

3. Some parking tickets don't have addresses on them, which is a cause for concern. Find out how many such tickets there are.

| Years | Total address missing cases |
|---|---|
| 2015 | 1807864 |
| 2016 | 2035232 |
| 2017 | 2289944 |

Observations –

From the data it seems that the trend of missing addresses on tickets are increasing from 2015-17 which is a case of concern.

Aggregation tasks

1. How often does each violation code occur? (frequency of violation codes - find the top 5)

## Plotting of Top 5 Violation codes across 2015-17


| Top 5 Violation Codes 2015 | |
|---|---|
| Violation Codes | Frequency |
| 21 | 1501614 |
| 38 | 1324586 |
| 14 | 924627 |
| 36 | 761571 |
| 37 | 746278 |

| Top 5 Violation Codes 2016 | |
|---|---|
| Violation Codes | Frequency |
| 21 | 1531587 |
| 36 | 1253512 |
| 38 | 1143696 |
| 14 | 875614 |
| 37 | 686610 |

| Top 5 Violation Codes 2017 | |
|---|---|
| Violation Codes | Frequency |
| 21 | 1528588 |
| 36 | 1400614 |
| 38 | 1062304 |
| 14 | 893498 |
| 20 | 618593 |

Observations –

The top 5 violation codes for 2015 & 16 are similar just we have another entry for 2017 which is 20(lesser number of cases as compared to code 21 and others).

2. How often does each vehicle body type get a parking ticket? How about the vehicle make? (find the top 5 for both)

Vehicle Body & Make Type- number of parking tickets on the basis of vehicle body type and make type

| Top 5 Vehicle body type 2015 | |
|---|---|
| Vehicle Body Type | Frequency |
| SUBN | 3451963 |
| 4DSD | 3102510 |
| VAN | 1605228 |
| DELV | 840442 |
| SDN | 453992 |

| Top 5 Vehicle body type 2016 | |
|---|---|
| Vehicle Body Type | Frequency |
| SUBN | 3466037 |
| 4DSD | 2992107 |
| VAN | 1518303 |
| DELV | 755282 |
| SDN | 424043 |

| Top 5 Vehicle body type 2017 | |
|---|---|
| Vehicle Body Type | Frequency |
| SUBN | 3719802 |
| 4DSD | 3082020 |
| VAN | 1411970 |
| DELV | 687330 |
| SDN | 438191 |

| Top 5 Vehicle Make type 2015 | |
|---|---|
| Vehicle Make Type | Frequency |
| FORD | 1417303 |
| TOYOT | 1123523 |
| HONDA | 1018049 |
| NISSA | 837569 |
| CHEVR | 836389 |

| Top 5 Vehicle Make type 2016 | |
|---|---|
| Vehicle Make Type | Frequency |
| FORD | 1324774 |
| TOYOT | 1154790 |
| HONDA | 1014074 |
| NISSA | 834833 |
| CHEVR | 759663 |

| Top 5 Vehicle Make type 2017 | |
|---|---|
| Vehicle Make Type | Frequency |
| FORD | 1280958 |
| TOYOT | 1211451 |
| HONDA | 1079238 |
| NISSA | 918590 |
| CHEVR | 714655 |

Observations –Top 5 vehicle body type & make type are similar from 2015-17, just a small variation with FORD reporting the maximum number of cases.

3. A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequencies of:

a)Violating Precincts (this is the precinct of the zone where the violation occurred)

| Top 5 Violating Precincts- 2015 | |
|---|---|
| Violating Precinct Code | Frequency |
| 0 | 1633006 |
| 19 | 559716 |
| 18 | 400887 |
| 14 | 384596 |
| 1 | 307808 |

| Top 5 Violating Precincts- 2016 | |
|---|---|
| /iolating Precinct Cod | Frequency |
| 0 | 1868655 |
| 19 | 554465 |
| 18 | 331704 |
| 14 | 324467 |
| 1 | 303850 |

| Top 5 Violating Precincts- 2017 | |
|---|---|
| Violating Precinct Code | Frequency |
| 0 | 2072400 |
| 19 | 535671 |
| 14 | 352450 |
| 1 | 331810 |
| 18 | 306920 |

Observations – The violating precincts are similar from 2015-17 just variation in count. Precinct code 0 showing the maximum number of cases in all the three years and it's showing an increasing trend from 2015-17

b)Issuing Precincts (this is the precinct that issued the ticket)

| Top 5 Issuing Precincts- 2015 | |
|---|---|
| Issuing Precinct Code | Frequency |
| 0 | 1834343 |
| 19 | 544946 |
| 18 | 391501 |
| 14 | 369725 |
| 1 | 298594 |

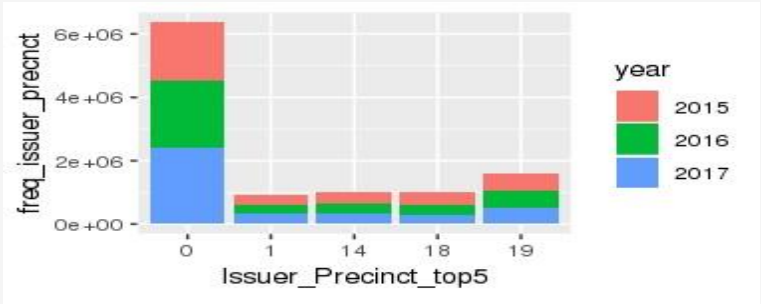
| Top 5 Issuing Precincts- 2016 | |
|---|---|
| Issuing Precinct Code | Frequency |
| 0 | 2140274 |
| 19 | 540569 |
| 18 | 323132 |
| 14 | 315311 |
| 1 | 295013 |

| Top 5 Issuing Precincts- 2017 | |
|---|---|
| Issuing Precinct Code | Frequency |
| 0 | 2388479 |
| 19 | 521513 |
| 18 | 344977 |
| 14 | 321170 |
| 1 | 296553 |

**Observations – It is obvious from the data that precinct code 0 issues the maximum number of tickets. The trend is increasing from 2015-17 and the precinct codes are similar across the 3 years, there is no new entry. Hence kind of conclude that Precinct codes 0,19 and 18 issues the most number of tickets and are situated close to zones where maximum violations took place.**

4. Find the violation code frequency across 3 precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?

## plotting of top 5 issuer precincts across 2015-17




| Three most ticket issuing precincts -2015 | | |
|---|---|---|
| Violation Code | Issuer Precinct | Frequency |
| 36 | 0 | 761571 |
| 7 | 0 | 662201 |
| 5 | 0 | 195352 |
| 21 | 0 | 180481 |
| 14 | 18 | 121004 |
| 38 | 19 | 90437 |
| 37 | 19 | 79738 |
| 14 | 19 | 60589 |
| 69 | 18 | 57218 |
| 21 | 19 | 56416 |

| Three most ticket issuing precincts -2016 | | |
|---|---|---|
| Violation Code | Issuer Precinct | Frequency |
| 36 | 0 | 1253511 |
| 7 | 0 | 492469 |
| 21 | 0 | 237174 |
| 5 | 0 | 112376 |
| 14 | 18 | 99857 |
| 38 | 19 | 77183 |
| 37 | 19 | 75641 |
| 46 | 19 | 73016 |
| 14 | 19 | 61742 |
| 21 | 19 | 58719 |

| Three most ticket issuing precincts -2017 | | |
|---|---|---|
| Violation Code | Issuer Precinct | Frequency |
| 36 | 0 | 1400614 |
| 7 | 0 | 516389 |
| 21 | 0 | 268591 |
| 5 | 0 | 145642 |
| 46 | 19 | 86390 |
| 14 | 14 | 73837 |
| 37 | 19 | 72437 |
| 38 | 19 | 72344 |
| 69 | 14 | 58026 |
| 14 | 19 | 57563 |

**Observations – Precinct code 0 issued most number of tickets for violation code 36 and the frequency has increased from 2015-17**

5. You'd want to find out the properties of parking violations across different times of the day:
    - The Violation Time field is specified in a strange format. Find a way to make this into a time attribute that you can use to divide into groups.
    - Find a way to deal with missing values, if any.
    - Divide 24 hours into 6 equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the 3 most commonly occurring violations

Have divided the 24 hour time into three timeslots- Morning, Late Morning, afternoon, Late afternoon, evening, Late evening & night for our analysis.

**2015**

**Night Top 5 Violations-2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 67430 | 21 | Night |
| 42406 | 40 | Night |
| 39521 | 78 | Night |
| 34081 | 7 | Night |
| 30451 | 14 | Night |
| 25039 | 85 | Night |

**Morning Top 5 Violations-2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 134458 | 14 | Morning |
| 106858 | 21 | Morning |
| 91344 | 40 | Morning |
| 81103 | 20 | Morning |
| 56550 | 36 | Morning |
| 55456 | 7 | Morning |

**Late Morning Top 5 Violations-2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 1192163 | 21 | Late Morning |
| 449070 | 38 | Late Morning |
| 360365 | 36 | Late Morning |
| 297711 | 14 | Late Morning |
| 210978 | 46 | Late Morning |
| 194927 | 71 | Late Morning |

**Afternoon top 5 violatios 2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 568272 | 38 | Afternoon |
| 417613 | 37 | Afternoon |
| 323524 | 36 | Afternoon |
| 267625 | 14 | Afternoon |
| **209828** | 20 | Afternoon |
| **197073** | **46** | **Afternoon** |

**#Evening top 5 violations 2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 241327 | 38 | Evening |
| **175802** | 37 | Evening |
| **168888** | **7** | **Evening** |
| 148538 | 14 | Evening |
| 89709 | 5 | Evening |
| 89337 | 20 | Evening |

**#Late Evening top 5 violations 2015**

| Frequency Of Violation | Violation Code | Time Bin |
|---|---|---|
| 81981 | 7 | Late Evening |
| 62418 | 38 | Late Evening |
| 45821 | 14 | Late Evening |
| 44891 | 40 | Late Evening |
| 31231 | 20 | Late Evening |
| 29324 | 78 | Late Evening |

**Observations –**

**#2015 , The three most commonly occurring violation codes are :-**

**#Violation_Code sum_code_wise**

**#21 1366451**

**#38 1321087**

**#14 924604**

**2016 –**

| #Night top 5 violations 2016 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 72106 | 21 | Night |
| 42098 | 40 | Night |
| 32806 | 78 | Night |
| 31779 | 14 | Night |
| 25076 | 7 | Night |
| **22862** | 20 | Night |
| | | |
| **#Morning top 5 violations 2016** | | |
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 140111 | 14 | Morning |
| 114029 | 21 | Morning |
| 91692 | 40 | Morning |
| 79797 | 36 | Morning |
| 77831 | 20 | Morning |
| 45146 | 71 | Morning |
| | | |
| **#Late Morning top 5 violations 2016** | | |
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 1209243 | 21 | Late Morning |
| 586791 | 36 | Late Morning |
| 388099 | 38 | Late Morning |
| 276273 | 14 | Late Morning |
| 220535 | 46 | Late Morning |
| 185136 | 71 | Late Morning |
| | | |
| | | |
| **#Afternoon top 5 violations 2016** | | |
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| **545717** | **36** | **Afternoon** |
| 488302 | 38 | Afternoon |
| 383361 | 37 | Afternoon |
| 247933 | 14 | Afternoon |
| 216636 | 20 | Afternoon |
| 213713 | 46 | Afternoon |
| | | |
| **Evening top 5 violations 2016** | | |
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 211267 | 38 | Evening |
| 161655 | 37 | Evening |
| 134976 | 14 | Evening |
| 124617 | 7 | Evening |
| 82603 | 20 | Evening |
| 75844 | 46 | Evening |
| | | |
| **Late Evening top 5 violations 2016** | | |
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 60924 | 7 | Late Evening |
| 53174 | 38 | Late Evening |
| 44973 | 40 | Late Evening |
| 44227 | 14 | Late Evening |
| **31163** | 20 | Late Evening |
| **24475** | **78** | **Late Evening** |

**Observations –**

**#2016 , The three most commonly occurring violation codes are :-**

**#Violation_Code sum_code_wise**

**#21 1395378**

**#36 1212305**

## #38 1140842

## 2017 –

| #Night top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 77460 | 21 | Night |
| 50947 | 40 | Night |
| 32243 | 78 | Night |
| 32064 | 14 | Night |
| 25297 | 7 | Night |
| 24611 | 20 | Night |
| | | |
| | | |

| Morning top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| **141276** | 14 | Morning |
| **119469** | **21** | **Morning** |
| 112186 | 40 | Morning |
| 84647 | 20 | Morning |
| 44060 | 7 | Morning |
| 43432 | 71 | Morning |
| | | |

| #Late Morning top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| **1182689** | **21** | **Late Morning** |
| 751422 | 36 | Late Morning |
| 346518 | 38 | Late Morning |
| 274288 | 14 | Late Morning |
| 213696 | 46 | Late Morning |
| 192307 | 71 | Late Morning |
| | | |

| #Afternoon top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 588395 | 36 | Afternoon |
| 462758 | 38 | Afternoon |
| 337075 | 37 | Afternoon |
| 256331 | 14 | Afternoon |
| **229390** | 46 | Afternoon |
| **219207** | **20** | **Afternoon** |
| | | |

| #Evening top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 203232 | 38 | Evening |
| 145784 | 37 | Evening |
| **144749** | 14 | Evening |
| **131768** | **7** | **Evening** |
| 85551 | 46 | Evening |
| 83348 | 20 | Evening |
| | | |

| #Late Evening top 5 violations 2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| **65593** | 7 | Late Evening |
| **47029** | **38** | Late Evening |
| 44779 | 14 | Late Evening |
| 44542 | 40 | Late Evening |
| 31085 | 20 | Late Evening |
| 25133 | 46 | Late Evening |

## Observations –

## #2017 , The three most commonly occurring violation codes are :-

## #Violation_Code sum_code_wise

## #21 1379618

**#36 1339817**

**#38 1073717**

Now, try another direction. For the 3 most commonly occurring violation codes, find the most common times of day (in terms of the bins from the previous part)

#Assumption :- most commonly occurring violation code has maximum count among all the time bins

| Most common Violation codes-2015 | Sum_code_wise |
|---|---|
| 21 | 1366451 |
| 38 | 1321087 |
| 14 | 924604 |

#Using the three most commonly occurring violation codes above to analyse below the time of the day with max violations

| Violation Code 21 -2015 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 1192163 | 21 | Late Morning |
| 133718 | 21 | Afternoon |
| 106858 | 21 | Morning |
| 67430 | 21 | Night |
| 823 | 21 | Evening |
| 587 | 21 | Late Evening |
| | | |

From the above output we can say that during "Late Morning" hours Violation code 21 happens the most

*# code 21 stands for "NO PARKING-STREET CLEANING"*

| Violation Code 38 -2015 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 568272 | 38 | Afternoon |
| 449070 | 38 | Late Morning |
| 241327 | 38 | Evening |
| 62418 | 38 | Late Evening |
| 2852 | 38 | Morning |
| 647 | 38 | Night |
| | | |

From the above output we can say that during "Afternoon" hours Violating code 38 happens the most

*# code 38 stands for "FAIL TO DSPLY MUNI METER RECPT"*

| Violation Code 14-2015 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 297711 | 14 | Late Morning |
| 267625 | 14 | Afternoon |
| 148538 | 14 | Evening |
| 134458 | 14 | Morning |
| 45821 | 14 | Late Evening |
| **30451** | 14 | Night |

From the above output we can say during "Late Morning" hours Violating code 14 happens the most

*# code 14 stands for "FAIL TO DSPLY MUNI METER RECPT"*

| Most common Violation codes-2016 | Sum_code_wise |
|---|---|
| 21 | 1395378 |
| 36 | 1212305 |
| 38 | 1140842 |

| Violation code 21 -2016 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 1209243 | 21 | Late Morning |
| 134329 | 21 | Afternoon |
| 114029 | 21 | Morning |
| 72106 | 21 | Night |
| **601** | 21 | Evening |
| **428** | 21 | Late Evening |

From the above output we can say that during "Late Morning" hours Violating code 21 happens the most

*# code 21 stands for "NO PARKING-STREET CLEANING"*

| Violation Code -36 -2016 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 586791 | 36 | Late Morning |
| 545717 | 36 | Afternoon |
| 79797 | 36 | Morning |
| 41205 | 36 | Evening |
| **1** | 36 | Night |

From the above output we can say during "Late Morning" hours Violating code 36 happens the most

*# code 36 stands for "PHTO SCHOOL ZN SPEED VIOLATION"*

| Violation Code 38 -2016 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 488302 | 38 | Afternoon |
| 388099 | 38 | Late Morning |
| 211267 | 38 | Evening |
| **53174** | 38 | Late Evening |
| **2211** | 38 | Morning |
| **384** | **38** | **Night** |

From the above output we can say during "Afternoon" hours Violating code 38 happens the most

*# code 38 stands for "FAIL TO DSPLY MUNI METER RECPT"*

| Most common Violation codes-2017 | Sum_code_wise |
|---|---|
| 21 | 1379618 |
| 36 | 1339817 |
| 38 | 1073717 |

| Violation Code 21-2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 1182689 | 21 | Late Morning |
| 148013 | 21 | Afternoon |
| **119469** | 21 | Morning |
| **77460** | **21** | **Night** |
| 551 | 21 | Evening |
| 363 | 21 | Late Evening |

From the above output we can say in "Late Morning" time of the day Violating code 21 happens the most
*# code 21 stands for "NO PARKING-STREET CLEANING*"

| Violation Code 36 -2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 751422 | 36 | Late Morning |
| 588395 | 36 | Afternoon |
| **33939** | 36 | Morning |
| **26858** | **36** | **Evening** |

From the above output we can say in "Late Morning" time of the day Violating code 36 happens the most
*# code 36 stands for "PHTO SCHOOL ZN SPEED VIOLATION*"

| Violation Code 38-2017 | | |
|---|---|---|
| **Frequency Of Violation** | **Violation Code** | **Time Bin** |
| 462758 | 38 | Afternoon |
| **346518** | 38 | Late Morning |
| **203232** | 38 | Evening |
| **47029** | **38** | **Late Evening** |
| 2300 | 38 | Morning |
| 464 | 38 | Night |

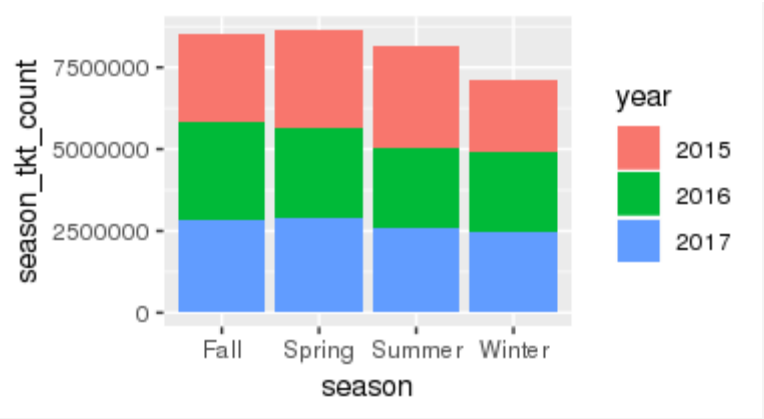
From the above output we can say in "Afternoon" time of the day Violating code 38 happens the most
*# code 38 stands for "FAIL TO DSPLY MUNI METER RECPT"*

6. Let's try and find some seasonality in this data

- First, divide the year into some number of seasons, and find frequencies of tickets for each season.
- Then, find the 3 most common violations for each of these season

**#Assumption: Seasons are defined as: June-August->Summer, September-November->Spring, December-February->Winter, March-May-> Fall**

**## Plotting of seasonal ticket count across 2015-17**




## Data Analysis for 2015

| Violation_Season_bin | Season_ticket_count |
|---|---|
| Spring | 2860987 |
| Summer | 2838306 |
| Fall | 2718502 |
| Winter | 2180241 |

**Observation -Maximum tickets were issued during spring season**

**## Seasonwise count of Violating code**

**##Top 3 violation codes for summer season in 2015**

| Violation Code | Violation Count |
|---|---|
| 21 | 439632 |
| 38 | 344262 |
| 14 | 239339 |

**##Top 3 violation codes for spring season in 2015**

| Violation Code | Violation Count |
|---|---|
| 21 | 425163 |
| 38 | 327048 |
| 14 | 243622 |

**## Top 3 violation codes from winter season in 2015**

| Violation Code | Violation Count |
|---|---|
| 38 | 306997 |
| 21 | 253043 |
| 14 | 193157 |

**# Top 3 violation codes from fall season in 2015**

| Violation Code | Violation Count |
|---|---|
| 21 | 351390 |
| 38 | 326700 |
| 14 | 232300 |

**Overall 21, 38 & 14 are the major violation codes with most number of occurrences in 2015 across all the seasons**

## Data Analysis for 2016

| Violation_Season_bin | Season_ticket_count |
|---|---|
| Spring | 2789066 |
| Summer | 2214536 |
| Fall | 2971672 |
| Winter | 2421620 |

**Most number of tickets were issued during fall season**

**#Top 3 violation codes for summer season in 2016**

| Violation Code | Violation Count |
|---|---|
| 21 | 358896 |
| 38 | 255600 |
| 14 | 200608 |

**#Top 3 violation codes for spring season in 2016**

| Violation Code | Violation Count |
|---|---|
| 21 | 383448 |
| 36 | 374362 |
| 38 | 299439 |

**##Top 3 violation codes from winter season in 2016**

| Violation Code | Violation Count |
|---|---|
| 21 | 359905 |
| 36 | 314765 |
| 38 | 268409 |

**##Top 3 violation codes from fall season in 2016**

| Violation Code | Violation Count |
|---|---|
| 36 | 438320 |
| 21 | 395020 |
| 38 | 303387 |

**Top violating codes across all the seasons are 21,38,36 & 14**

## Data Analysis 2017

| Violation_Season_bin | Season_ticket_count |
|---|---|
| Spring | 2873383 |
| Summer | 2353920 |
| Fall | 2829224 |
| Winter | 2483036 |

**Most number of tickets were issued during spring season**

**#Top 3 violation codes for summer season in 2017**

| Violation Code | Violation Count |
|---|---|
| 21 | 378699 |
| 38 | 235725 |
| 36 | 207495 |

**##Top 3 violation codes for spring season in 2017**

| Violation Code | Violation Count |
|---|---|
| 21 | 402424 |
| 36 | 344834 |
| 38 | 271167 |

**##Top 3 violation codes from winter season in 2017**

| Violation Code | Violation Count |
|---|---|
| 21 | 362016 |
| 36 | 359338 |
| 38 | 259710 |

**##Top 3 violation codes from fall season in 2017**

| Violation Code | Violation Count |
|---|---|
| 36 | 456046 |
| 21 | 357257 |
| 38 | 283816 |

**Most common violation codes are 21,36,38 across all the seasons.**

The fines collected from all the parking violation constitute a revenue source for the NYC police department. Let's take an example of estimating that for the 3 most commonly occurring codes.

```
## Find total occurrences of the 3 most common violation codes
```

| 2015 | |
|---|---|
| Frequency | Violation Code |
| 1501614 | 21 |
| 1324586 | 38 |
| 924627 | 14 |

| 2016 | |
|---|---|
| Frequency | Violation Code |
| 1531587 | 21 |
| 1253512 | 36 |
| 1143696 | 38 |

| 2017 | |
|---|---|
| Frequency | Violation Code |
| 1528588 | 21 |
| 1400614 | 36 |
| 1062304 | 38 |

**Observation – 21,36 & 38 are the most frequent and commonly occurring violation codes for which tickets were issued if we look at the above data across 2015-17**

- Then, search the internet for NYC parking violation code fines. You will find a website (on the nyc.gov URL) that lists these fines. They're divided into two categories, one for the highest-density locations of the city, the other for the rest of the city. For simplicity, take an average of the two.
- Using this information, find the total amount collected for all of the fines. State the code which has the highest total collection.
- What can you intuitively infer from these findings?

**# In 2015 , top three violation codes are 14 , 21 & 38**

| Frequency | Violation Code | Avg Fine | Total Fine |
|---|---|---|---|
| 1501614 | 21 | 54.5 | 81837963 |
| 1324586 | 38 | 50 | 66229300 |
| 924627 | 14 | 115 | 106332105 |

**Total fine collected for top three violation code in $ in 2015 is : $ 25,43,99,368.00**

**# In 2016, top three violation codes are 21, 36 & 38**

| Frequency | Violation Code | Avg Fine | Total Fine |
|---|---|---|---|
| 1531587 | 21 | 54.5 | 83471492 |
| 1253512 | 36 | 50 | 62675600 |
| 1143696 | 38 | 50 | 57184800 |
| | | | |

**Total fine collected for top three violation code in $ in 2016 is : $ 20,33,31,892.00**

**#In 2017, top three violation codes are 21, 36 & 38**

| Frequency | Violation Code | Avg Fine | Total Fine |
|---|---|---|---|
| 1528588 | 21 | 54.5 | 83308046 |
| 1400614 | 36 | 50 | 70030700 |
| 1062304 | 38 | 50 | 53115200 |

**Total fine collected for top three violation code in $ in 2017 is : $ 20,64,53,946.00**

**Observation – Maximum fine amount was collected from 2015 as obvious from the records that most number of tickets were issued during that year.**

**######################## End ############################**

**################### Thank you ############################**