

NYC Parking Case Study

Data Cleaning:

- Retain only Distinct Rows
- Rename columns by removing spaces and replacing with Underscore
- Create new column from Issue Date column in Date Format

Examine the data:

1. Find the total number of tickets for each year.

Assumption:

Summons Number is Unique Identifier of Each Ticket

| Year | Number of Tickets |
|------|-------------------|
| 2015 | 10951256 |
| 2016 | 10626899 |
| 2017 | 10803028 |

2. Find out the number of unique states from where the cars that got parking tickets came from.

Data Cleaning:

Registration State contained numeric value 99 which was replaced with state having maximum entries (NY)

| Year | Unique states before replacing value "99" | Unique states after replacing value "99" |
|------|---|--|
| 2015 | 69 | 68 |
| 2016 | 68 | 67 |
| 2017 | 67 | 66 |

3. Some parking tickets don't have the address for violation location on them, which is a cause for concern. Write a query to check the number of such tickets.

Assumption:

House Name and Street Name indicate if address is present in ticket or not.

| Year | Number of Tickets without Address |
|------|-----------------------------------|
| 2015 | 1807864 |
| 2016 | 2035232 |
| 2017 | 2289944 |

Aggregation tasks:

1. How often does each violation code occur? Display the frequency of the top five violation codes.

- 2015

| Violation Code | Frequency of violation codes |
|----------------|------------------------------|
| 21 | 1501614 |
| 38 | 1324586 |
| 14 | 924627 |
| 36 | 761571 |
| 37 | 746278 |

- 2016

| Violation Code | Frequency of violation codes |
|----------------|------------------------------|
| 21 | 1531587 |
| 36 | 1253512 |
| 38 | 1143696 |
| 14 | 875614 |
| 37 | 686610 |

- 2017

| Violation Code | Frequency of violation codes |
|-----------------------|-------------------------------------|
| 21 | 1528588 |
| 36 | 1400614 |
| 38 | 1062304 |
| 14 | 893498 |
| 20 | 618593 |

OBSERVATION:

Violation Code “21” identified as top Violation code for all the 3 years based on frequency.

2. How often does each 'vehicle body type' get a parking ticket?
How about the 'vehicle make'?

- 2015

| Vehicle Body Type | Frequency of parking ticket |
|--------------------------|------------------------------------|
| SUBN | 3451963 |
| 4DSD | 3102510 |
| VAN | 1605228 |
| DELV | 840442 |
| SDN | 453992 |

| Vehicle Make | Frequency of parking ticket |
|---------------------|------------------------------------|
| FORD | 1417303 |
| TOYOT | 1123523 |
| HONDA | 1018049 |
| NISSA | 837569 |
| CHEVR | 836389 |

- 2016

| Vehicle Body Type | Frequency of parking ticket |
|-------------------|-----------------------------|
| SUBN | 3466037 |
| 4DSD | 2992107 |
| VAN | 1518303 |
| DELV | 755282 |
| SDN | 424043 |

| Vehicle Make | Frequency of parking ticket |
|--------------|-----------------------------|
| FORD | 1324774 |
| TOYOT | 1154790 |
| HONDA | 1014074 |
| NISSA | 834833 |
| CHEVR | 759663 |

- 2017

| Vehicle Body Type | Frequency of parking ticket |
|-------------------|-----------------------------|
| SUBN | 3719802 |
| 4DSD | 3082020 |
| VAN | 1411970 |
| DELV | 687330 |
| SDN | 438191 |

| Vehicle Make | Frequency of parking ticket |
|--------------|-----------------------------|
| FORD | 1280958 |
| TOYOT | 1211451 |
| HONDA | 1079238 |
| NISSA | 918590 |
| CHEVR | 714655 |

OBSERVATION:

For all 3 years, Vehicle Body Type and Vehicle Make got Parking tickets in same order of frequency.

3. A precinct is a police station that has a certain zone of the city under its command.

Assumption:

Excluded Precinct “0” as not Valid and erroneous entries as specified.

- a. Find the (5 highest) frequency of tickets for 'Violation Precinct' (this is the precinct of the zone where the violation occurred).

•2015

| Violation Precinct | Frequency |
|--------------------|-----------|
| 19 | 559716 |
| 18 | 400887 |
| 14 | 384596 |
| 1 | 307808 |
| 114 | 300557 |

•2016

| Violation Precinct | Frequency |
|--------------------|-----------|
| 19 | 554465 |
| 18 | 331704 |
| 14 | 324467 |
| 1 | 303850 |
| 114 | 291336 |

•2017

| Violation Precinct | Frequency |
|--------------------|-----------|
| 19 | 535671 |
| 14 | 352450 |
| 1 | 331810 |
| 18 | 306920 |
| 114 | 296514 |

- b. Find the (5 highest) frequency of tickets for 'Issuer Precinct' (this is the precinct that issued the ticket)

•2015

| Issuer Precinct | Frequency |
|-----------------|-----------|
| 19 | 544946 |
| 18 | 391501 |
| 14 | 369725 |
| 1 | 298594 |
| 114 | 295601 |

•2016

| Issuer Precinct | Frequency |
|-----------------|-----------|
| 19 | 540569 |
| 18 | 323132 |
| 14 | 315311 |
| 1 | 295013 |
| 114 | 286924 |

•2017

| Issuer Precinct | Frequency |
|-----------------|-----------|
| 19 | 521513 |
| 14 | 344977 |
| 1 | 321170 |
| 18 | 296553 |
| 114 | 289950 |

OBSERVATION:

For all 3 years, The Precinct Code remain the same although they move up and down based on Frequency.

4. Find the violation code frequency across three 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?

- 2015

In 2015, Top 3 Issuer Precincts are 19,18,14

| Frequency | Violation Code | Issuer Precinct |
|-----------|----------------|-----------------|
| 121004 | 14 | 18 |
| 90437 | 38 | 19 |
| 80368 | 69 | 14 |
| 79738 | 37 | 19 |
| 77269 | 14 | 14 |
| 60589 | 14 | 19 |
| 57218 | 69 | 18 |
| 56416 | 21 | 19 |
| 56318 | 16 | 19 |
| 44082 | 46 | 19 |

Observation:

Violation code 14 is common among all Top 3 Issuer Precincts and has high frequency.

- 2016

In 2016, Top 3 Issuer Precincts are 19,18,14

| Frequency | Violation Code | Issuer Precinct |
|-----------|----------------|-----------------|
| 99857 | 14 | 18 |
| 77183 | 38 | 19 |
| 75641 | 37 | 19 |
| 73016 | 46 | 19 |
| 67932 | 69 | 14 |
| 62426 | 14 | 14 |
| 61742 | 14 | 19 |
| 58719 | 21 | 19 |
| 52354 | 16 | 19 |
| 47881 | 69 | 18 |

Observation:

Violation code 14 is common among all Top 3 Issuer Precincts and has high frequency.

- 2017

In 2017, Top 3 Issuer Precincts are 19,14,1

| Frequency | Violation Code | Issuer Precinct |
|-----------|----------------|-----------------|
| 86390 | 46 | 19 |
| 73837 | 14 | 14 |
| 73522 | 14 | 1 |
| 72437 | 37 | 19 |
| 72344 | 38 | 19 |
| 58026 | 69 | 14 |
| 57563 | 14 | 19 |
| 54700 | 21 | 19 |
| 39857 | 31 | 14 |
| 38937 | 16 | 1 |

Observation:

Violation code 14 is common among all Top 3 Issuer Precincts and has high frequency.

5. You'd want to find out the properties of parking violations across different times of the day:
 - a. Find a way to deal with missing values, if any.
 - b. 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.
 - c. 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

- d. 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:

Removed NULL Values even though they were insignificant in size.

Considered Hour bins as below

Bin 1 -- 0 to 3 hrs

Bin 2 -- 4 to 7 hrs

Bin 3 -- 8 to 11 hrs

Bin 4 -- 12 to 15 hrs

Bin 5 -- 16 to 19 hrs

Bin 6 -- 20 to 23 hrs

- 2015

| Hour Bin | Violation Code | Frequency |
|----------|----------------|-----------|
| 1 | 21 | 63574 |
| 1 | 40 | 36491 |
| 1 | 78 | 34842 |
| | | |
| 2 | 40 | 22154 |
| 2 | 14 | 17387 |
| 2 | 7 | 15746 |
| | | |
| 3 | 21 | 1294480 |
| 3 | 38 | 451531 |
| 3 | 36 | 416910 |
| | | |
| 4 | 38 | 568341 |
| 4 | 37 | 417626 |
| 4 | 36 | 323526 |
| | | |
| 5 | 38 | 241327 |
| 5 | 37 | 175802 |
| 5 | 7 | 168888 |
| | | |
| 6 | 7 | 81981 |
| 6 | 38 | 62418 |
| 6 | 14 | 45821 |

Most Common times of day when Top 3 Violation Codes occur

| Hour Bin | Frequency |
|----------|-----------|
| 3 | 2160793 |
| 4 | 977448 |
| 5 | 390688 |
| 6 | 108826 |
| 1 | 90697 |
| 2 | 22319 |

- 2016

| Hour Bin | Violation Code | Frequency |
|----------|----------------|-----------|
| 1 | 21 | 67798 |
| 1 | 40 | 37261 |
| 1 | 78 | 29473 |
| | | |
| 2 | 40 | 21763 |
| 2 | 14 | 19307 |
| 2 | 20 | 11651 |
| | | |
| 3 | 21 | 1318487 |
| 3 | 36 | 666586 |
| 3 | 14 | 397077 |
| | | |
| 4 | 36 | 545717 |
| 4 | 38 | 488363 |
| 4 | 37 | 383379 |
| | | |
| 5 | 38 | 211267 |
| 5 | 37 | 161655 |
| 5 | 14 | 134976 |
| | | |
| 6 | 7 | 60924 |
| 6 | 38 | 53174 |
| 6 | 40 | 44973 |

Most Common times of day when Top 3 Violation Codes occur

| Hour Bin | Frequency |
|----------|-----------|
| 3 | 2375082 |
| 4 | 1172721 |
| 5 | 253073 |
| 1 | 68123 |
| 6 | 53602 |
| 2 | 5088 |

- 2017

| Hour Bin | Violation Code | Frequency |
|----------|----------------|-----------|
| 1 | 21 | 73160 |
| 1 | 40 | 45960 |
| 1 | 14 | 29312 |
| | | |
| 2 | 40 | 26225 |
| 2 | 14 | 15630 |
| 2 | 20 | 13724 |
| | | |
| 3 | 21 | 1296759 |
| 3 | 36 | 785361 |
| 3 | 14 | 399934 |
| | | |
| 4 | 36 | 588395 |
| 4 | 38 | 462859 |
| 4 | 37 | 337096 |
| | | |
| 5 | 38 | 203232 |
| 5 | 37 | 145784 |
| 5 | 14 | 144749 |
| | | |
| 6 | 7 | 65593 |
| 6 | 38 | 47029 |
| 6 | 14 | 44779 |

Most Common times Of day when Top 3 Violation Codes occur

| Hour Bin | Frequency |
|----------|-----------|
| 3 | 2430699 |
| 4 | 1203568 |
| 5 | 230641 |
| 1 | 73523 |
| 6 | 47392 |
| 2 | 5638 |

OBSERVATION:

For the all three years, majorly top 3 violations occurred in between 8 AM to 11 AM (Hour Bin 3) and next following was in between 12 PM to 3 PM (Hour Bin 4).

6. Let's try and find some seasonality in this data
 - a. First, divide the year into some number of seasons, and find frequencies of tickets for each season.
 - b. Then, find the 3 most common violations for each of these season

ASSUMPTION:

Considered Seasons as below

Season 1 -- January to March

Season 2 -- April to June

Season 3 -- July to September

Season 4 -- October to December

- 2015

Number of Tickets in each Season

| Season Bin | Number of Tickets |
|------------|-------------------|
| 1 | 2475936 |
| 2 | 3250418 |
| 3 | 2789425 |
| 4 | 2435478 |

Common Violations in every season of 2015

| Season Bin | Violation Code | Total Tickets |
|------------|----------------|---------------|
| 1 | 38 | 336762 |
| 1 | 21 | 281600 |
| 1 | 14 | 220029 |
| | | |
| 2 | 21 | 471580 |
| 2 | 38 | 346719 |
| 2 | 14 | 262595 |
| | | |
| 3 | 21 | 397871 |
| 3 | 38 | 348466 |
| 3 | 14 | 234606 |
| | | |
| 4 | 21 | 350563 |
| 4 | 38 | 292639 |
| 4 | 14 | 207397 |

- 2016

Number of Tickets in each Season

| Season Bin | Number of Tickets |
|------------|-------------------|
| 1 | 2671331 |
| 2 | 2425877 |
| 3 | 2728663 |
| 4 | 2801028 |

Common Violations in every season of 2015

| Season Bin | Violation Code | Total Tickets |
|------------|----------------|---------------|
| 1 | 21 | 349644 |
| 1 | 36 | 341787 |
| 1 | 38 | 308999 |
| | | |
| 2 | 21 | 348473 |
| 2 | 36 | 294015 |
| 2 | 38 | 254909 |
| | | |
| 3 | 21 | 403720 |
| 3 | 38 | 305360 |
| 3 | 14 | 234943 |
| | | |
| 4 | 36 | 433966 |
| 4 | 21 | 429750 |
| 4 | 38 | 274428 |

- 2017

Number of Tickets in each Season

| Season Bin | Number of Tickets |
|------------|-------------------|
| 1 | 2671332 |
| 2 | 3018840 |
| 3 | 2463936 |
| 4 | 2648920 |

Common Violations in every season of 2015

| Season Bin | Violation Code | Total Tickets |
|------------|----------------|---------------|
| 1 | 21 | 374202 |
| 1 | 36 | 348240 |
| 1 | 38 | 287017 |
| | | |
| 2 | 21 | 421184 |
| 2 | 36 | 369902 |
| 2 | 38 | 266909 |
| | | |
| 3 | 21 | 385774 |
| 3 | 38 | 244985 |

| | | |
|---|----|--------|
| 3 | 36 | 239879 |
| | | |
| 4 | 36 | 442593 |
| 4 | 21 | 347428 |
| 4 | 38 | 263393 |

OBSERVATION:

For the all three years, Violations 21 and 38 occurred highly compared to others in every season.

7. 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
- 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?

- 2015

Top 3 Violation Codes in 2015

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

Total Fine Amount collected from the Top 3 Violations

| Violation Code | Total Fine Amount |
|-----------------------|--------------------------|
| 14 | 106332105 |
| 21 | 82588770 |
| 38 | 66229300 |

- 2016

Top 3 Violation Codes in 2016

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

Total Fine Amount collected from the Top 3 Violations

| Violation Code | Total Fine Amount |
|-----------------------|--------------------------|
| 21 | 84237285 |
| 36 | 62675600 |
| 38 | 57184800 |

- 2017

Top 3 Violation Codes in 2016

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

Total Fine Amount collected from the Top 3 Violations

| Violation Code | Total Fine Amount |
|-----------------------|--------------------------|
| 21 | 84072340 |
| 36 | 70030700 |
| 38 | 53115200 |

OBSERVATIONS:

- While Violation Code 14 had brought in the most amount in 2015 as the fine amount is more than double the fine (\$115 compared to \$50 and \$55) of other 2 Violation Codes.
- In 2016 and 2017, Violation Code 21 had brought in the most revenue compared to other Violations
- Violation Code 14 has not been in the Top 3 Violations in 2016 and 2017
- Violation Codes 21, 36 and 38 have constantly been the top Violations for two consecutive years (2016-2017) and the revenue is also proportional to the number of Violation tickets in 2016 and 2017 as the Fine amount is more or less the same for the top 3 Violation Codes 21,36,38 (~\$50)