NYC Parking Tickets Assignment

Descriptive Answers

Group Members:
Ishan Savio Kerketta
Devanshi Kulshreshtha
Paromita Sarkar
Aniket Mitra

Problem Statement:

New York is one of the world's largest metros and has a very large population using cars as a primary medium of travel. The classic combination of a huge number of cars and cramped geography is the exact recipe that leads to a huge number of parking tickets.

This data pertains to parking tickets collected by the NYC Police Department over the year 2017.
The attempt is to perform some exploratory analysis on this data.

Data Preparation:

We take a subset of the NYC Parking Tickets dataset, such that it only contains data of the year 2017.

Questions:

Examine the data

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

Answer: 5,431,918 tickets issued in the year

2. Find out the number of unique states from where the cars that got parking tickets came from.
(Hint: Use the column 'Registration State')

Answer: 65 unique states (without data cleaning)

3. There is a numeric entry '99' in the column which should be corrected.
Replace it with the state having maximum entries.
Give the number of unique states again.

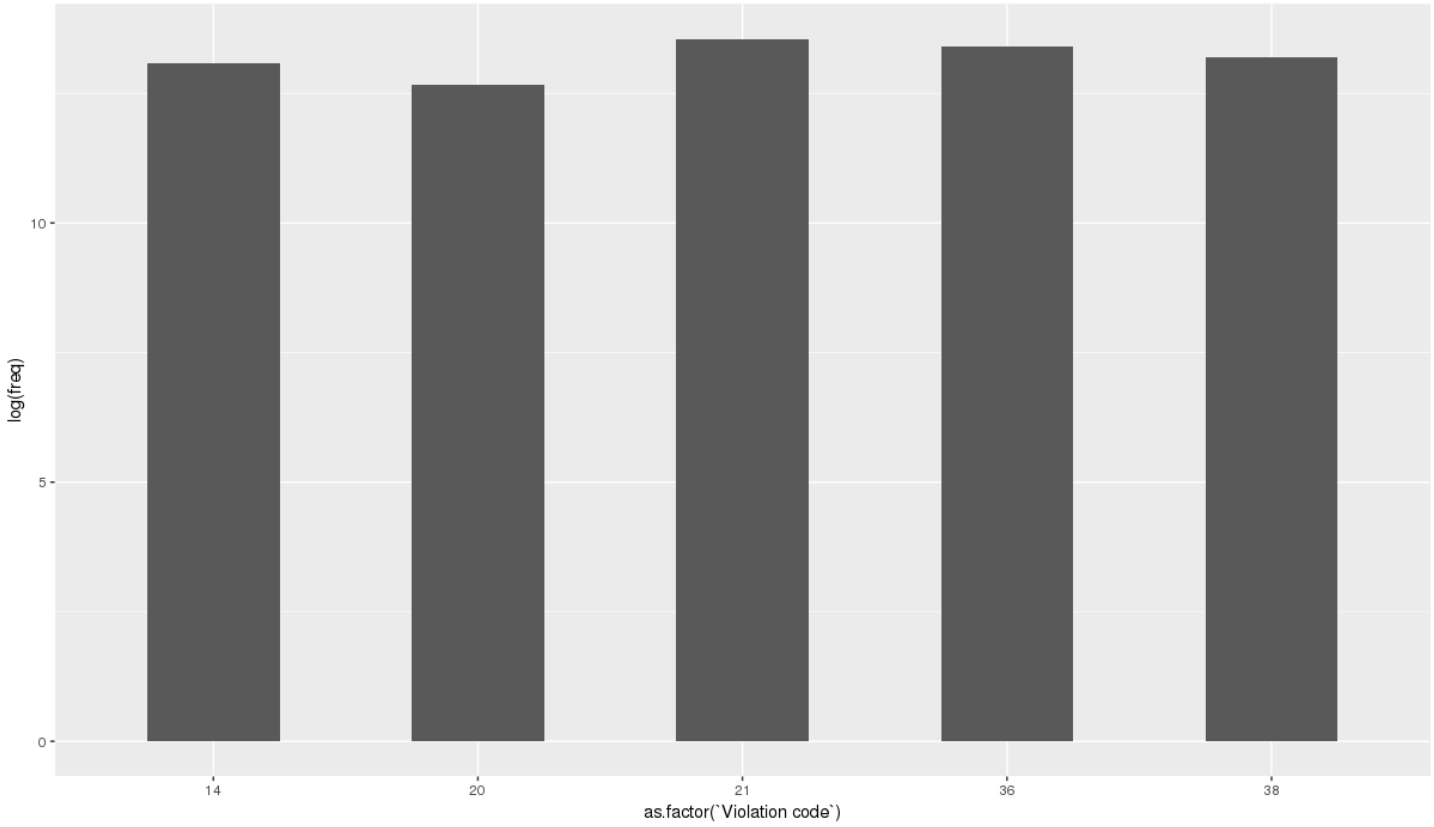
Answer: NY has maximum entries. Replacing the erroneous numeric entry '99' with NY.
Now the number of unique states = 64 (with data cleaning)

Aggregation tasks

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

Answer:

Violation Code	Frequency of parking tickets
21	768087
36	662765
38	542079
14	476664
20	319646



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

Answer:

Vehicle Body Type	Frequency of parking tickets
SUBN	1883954
4DSD	1547312
VAN	724029
DELV	358984
SDN	194197

How about the 'vehicle make'? (Hint: find the top 5 for both)

Answer:

Vehicle Make	Frequency of parking tickets
FORD	636844
TOYOT	605291
HONDA	538884
NISSAN	462017
CHEVR	356032

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

Find the (5 highest) frequency of tickets for each of the following:

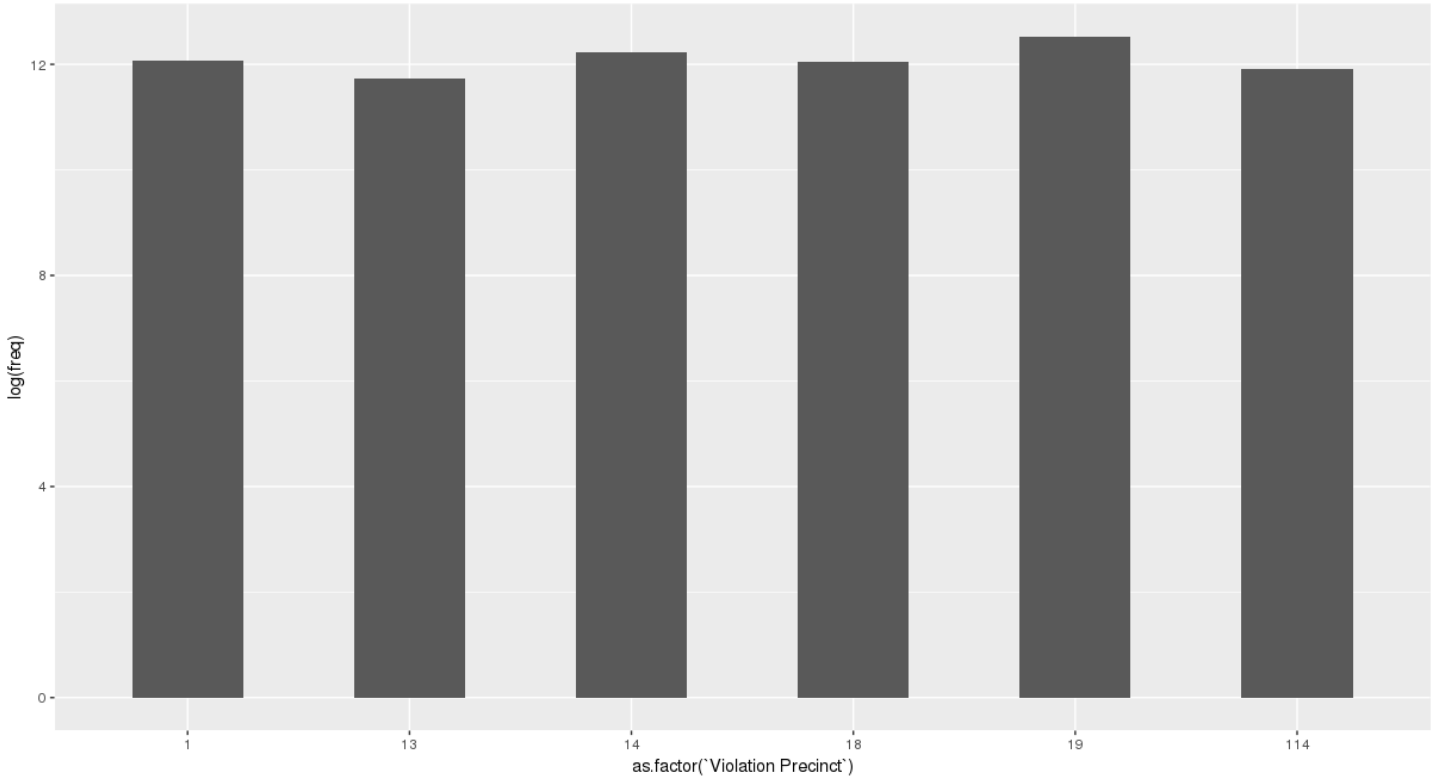
i. 'Violation Precinct' (this is the precinct of the zone where the violation occurred).

Using this, can you make any insights for parking violations in any specific areas of the city?

Answer:

Frequency of Top 6 Violation Precinct after removing Violation Precinct = 0 is as follows:

Violation Precinct	Frequency of parking tickets
19	274445
14	203553
1	174702
18	169131
114	147444
13	125113



ii. **'Issuer Precinct' (this is the precinct that issued the ticket)**
Here you would have noticed that the dataframe has 'Violating Precinct' or 'Issuing Precinct' as '0'. These are the erroneous entries. Hence, provide the record for five correct precincts.
(Hint: Print top six entries after sorting)

Answer:

Issuer Precinct	Frequency of parking tickets
19	266961
14	200495
1	168740
18	162994
114	144054
13	122490

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?
Hint: In the SQL view, use the 'where' attribute to filter among three precincts.

Answer:
Top 5 Violation code frequency across precinct 19 which have issued the most number of tickets

Issuer Precinct	Violation Code	Frequency of parking tickets
19	46	48445
19	38	36386
19	37	36056
19	14	29797
19	21	28415

Top 5 Violation code frequency across precinct 14 which have issued the most number of tickets

Issuer Precinct	Violation Code	Frequency of parking tickets
14	14	43036
14	69	30464
14	31	22555
14	47	18364
14	42	10027

Top 5 Violation code frequency across precinct 1 which have issued the most number of tickets

Issuer Precinct	Violation Code	Frequency of parking tickets
1	14	38354
1	16	19081
1	20	15048
1	46	12745
1	38	8535

5. You'd want to find out the properties of parking violations across different times of the day:
i. **Find a way to deal with missing values, if any.**
Hint: Check for the null values using 'isNull' under the SQL.
Also, to remove the null values, check the 'dropna' command in the API documentation.

Answer:
There are 16 Null values in Violation Time.
There are 38509 Null values in Vehicle Make.

After removing all Null Values from the dataset, the number of records is 5378917

ii. **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.**

Answer:

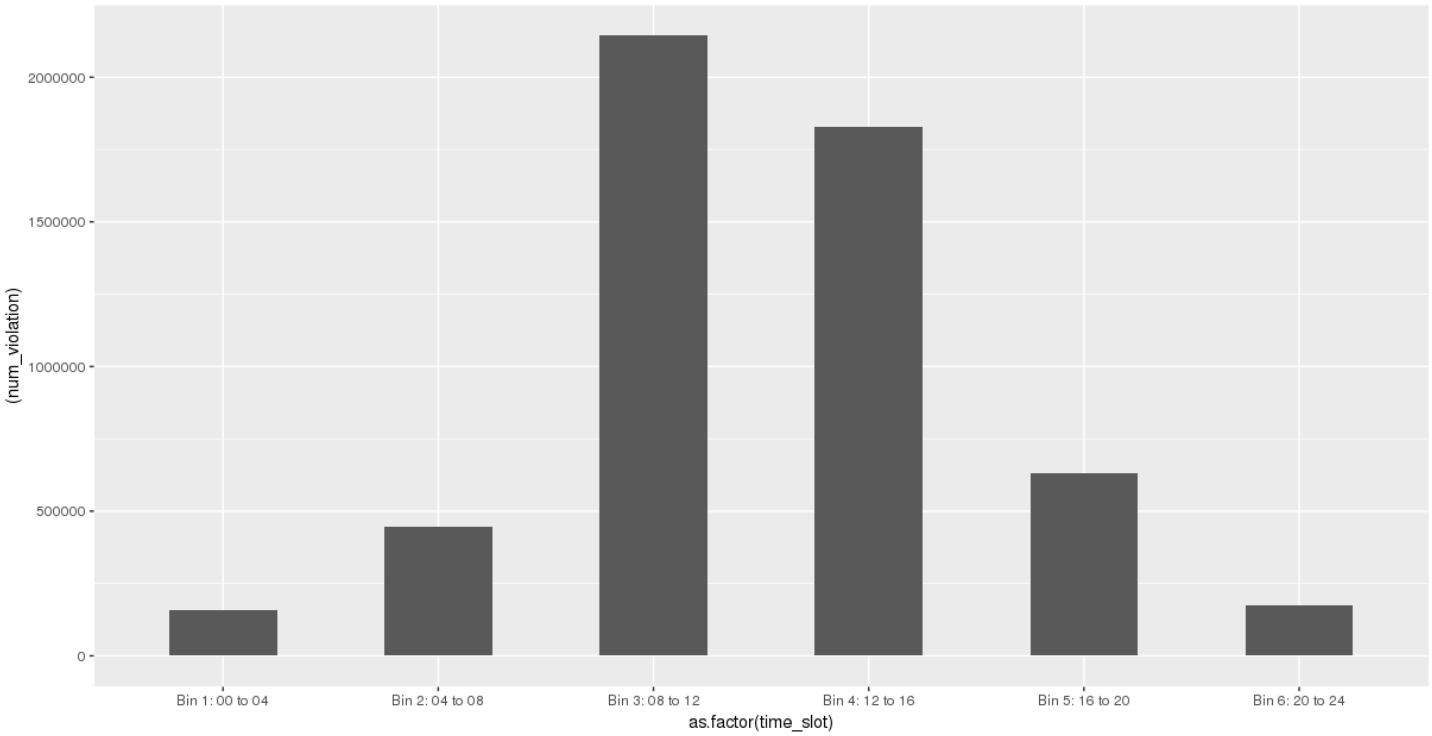
The codes used to convert the string into a time attribute is shown below:

```
#Adding 'M' to the string 'Violation Time'
park_tix_2017$M_char <- "M"
park_tix_2017$viol_time_string <- concat(park_tix_2017$`Violation Time`, park_tix_2017$M_char)

#Rectifying time values (changing values like 0015AM to 1215AM) for proper conversion
park_tix_2017$time_12_hr <- ifelse(substr(park_tix_2017$viol_time_string,1,2) == "00", "12",
substr(park_tix_2017$viol_time_string,1,2))
park_tix_2017$viol_time_string <- concat(park_tix_2017$time_12_hr, substr(park_tix_2017$viol_time_string,3,6))

#Converting time string to timestamp
park_tix_2017$viol_time <- to_timestamp(park_tix_2017$viol_time_string, format = 'hhmma')

#Extracting hour from timestamp
park_tix_2017$viol_hour <- hour(park_tix_2017$viol_time)

#Dropping unnecessary columns
park_tix_2017$time_12_hr <- NULL
park_tix_2017$M_char <- NULL
park_tix_2017$`Violation Time` <- NULL

#Creating Temporary view again to update changes
createOrReplaceTempView(park_tix_2017, "park_tix_2017_tbl")
```

iii. **Divide 24 hours into six equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups,**
Find the three most commonly occurring violations.
Hint: Use the CASE-WHEN in SQL view to segregate into bins. For finding the most commonly occurring violations, a similar approach can be used as mention in the hint for question 4.

Answer:

Violation Precinct	Frequency of parking tickets
Bin 1: 00 to 04	159103
Bin 2: 04 to 08	444965
Bin 3: 08 to 12	2143476
Bin 4: 12 to 16	1828779
Bin 5: 16 to 20	629397
Bin 6: 20 to 24	173197



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

Answer:

Three most common Violation Codes (Frequency of Top 3 violation codes):

Violation Code	Frequency of parking tickets
21	759775
36	661827
38	540128

Most common time of the day for Violation code 21:

Time Slots	Frequency of parking tickets
8 to 12	592582
12 to 16	73917
4 to 8	56723
0 to 4	36134
16 to 20	210
20 to 24	209

Most common time of the day for Violation code 21 is 8 AM to 12 PM

Most common time of the day for Violation code 36:

Time Slots	Frequency of parking tickets
8 to 12	347650
12 to 16	285895
4 to 8	14766
16 to 20	13516

Most common time of the day for Violation code 36 is 8 AM to 12 PM

Most common time of the day for Violation code 38

Time Slots	Frequency of parking tickets
12 to 16	239909
8 to 12	175779
16 to 20	102581
20 to 24	20311
4 to 8	1252
0 to 4	296

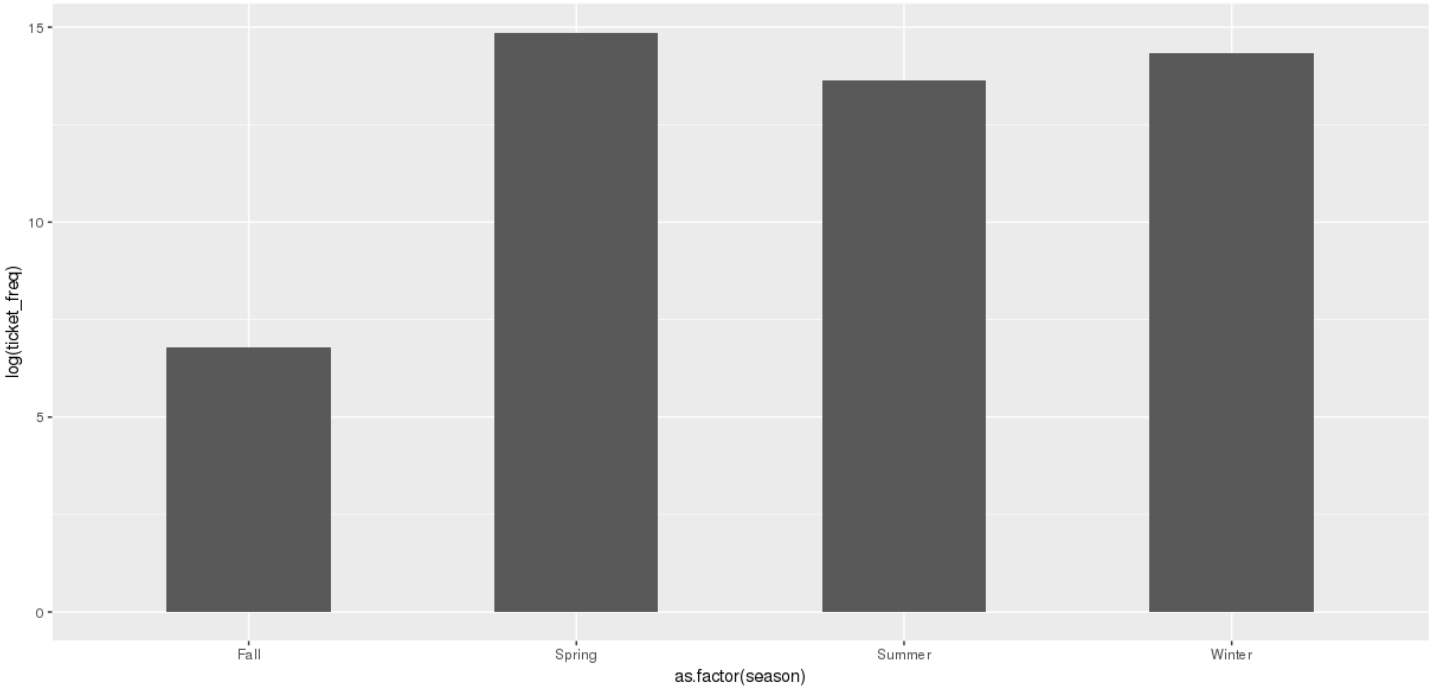
Most common time of the day for Violation code 38 is 12 PM to 4 PM

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

i. **First, divide the year into some number of seasons, and find frequencies of tickets for each season. (Hint: Use Issue Date to segregate into seasons)**

Answer:

Season	Frequency of parking tickets
Spring	2843495
Sumer	846864
Fall	889
Winter	1687669



ii. Then, find the three most common violations for each of these seasons.
(Hint: A similar approach can be used as mention in the hint for question 4.)

Answer:

Finding the three most common violations for Spring

Season	Frequency of parking tickets
Spring	397844
Spring	344366
Spring	270062

Finding the three most common violations for Summer

Season	Frequency of parking tickets
Summer	126376
Summer	96505
Summer	83190

Finding the three most common violations for Fall

Season	Frequency of parking tickets
Fall	203
Fall	118
Fall	114

Finding the three most common violations for Winter

Season	Frequency of parking tickets
Winter	235437
Winter	220956
Winter	186869

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 three most commonly occurring codes.

i. Find total occurrences of the three most common violation codes

Answer:
Total occurrences of the three most common violation codes

Violation Code	Frequency of parking tickets
21	759775
36	661827
38	540128

The fines collected from all the parking violation constitute a revenue source for the NYC police department.
Visit the website: http://www1.nyc.gov/site/finance/vehicles/services-violation-codes.page
It lists the fines associated with different violation codes.
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.

Violation Code 21
Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.
Manhattan 96th St. & below: $65 # All other areas $45
Average fine = $(65 + 45)/2 = $55

Violation Code 36

Exceeding the posted speed limit in or near a designated school zone.

Manhattan 96th St. & below: $50 # All other areas $50

Average fine = $50

Violation Code 38

Parking Meter -- 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.

Manhattan 96th St. & below: $65 # All other areas $35

Average fine = $(65 + 35)/2 = $50

Using this information, find the total amount collected for the three violation codes with maximum tickets.

Violation Code	No. of Tickets	Average Fee	Total Revenue
21	759,775	$55	**$41,787,625**
36	661,827	$50	$33,091,350
38	540,128	$50	$27,006,400

State the code which has the highest total collection.

Violation Code 21 has the highest total collection of **$41,787,625.**

Inference:

Drivers in New York City tend to violate 'No Parking' signs the most, out of all traffic violations. NYC Police Department collects maximum revenue for this violation compared to all other traffic violations.

This could mean that there are a great number of vehicles in New York City and finding a parking space is difficult. Hence, people tend to take the risk of parking in 'No Parking' areas and often commit a traffic violation as a result of this.