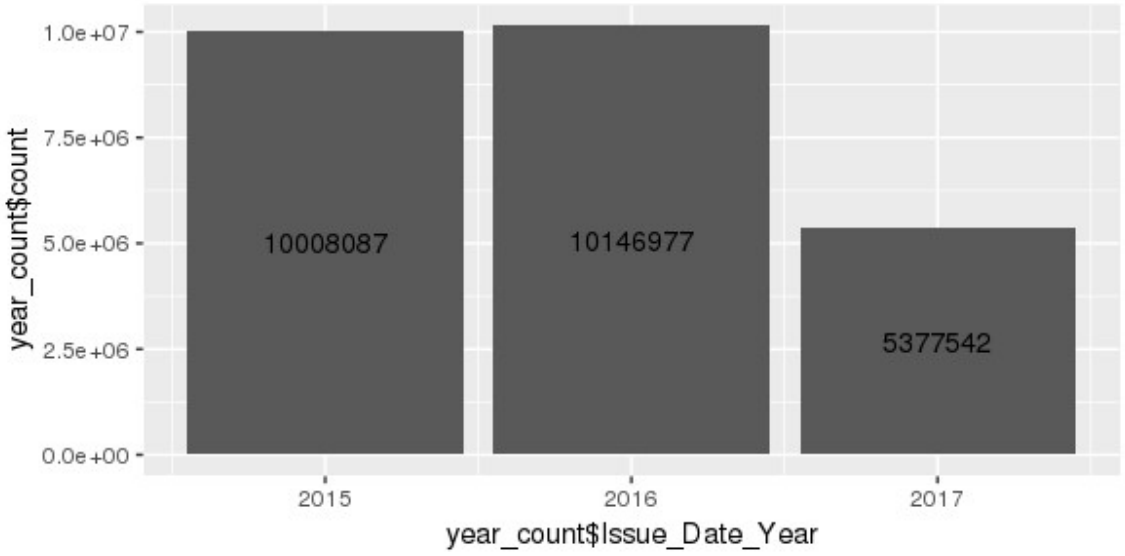
NYC PARKING TICKETS: AN EXPLORATORY ANALYSIS

Background:
New York City is a thriving metropolis. Just like most other metros that size, one of the biggest problems its citizens face is parking. The classic combination of a huge number of cars and a cramped geography is the exact recipe that leads to a huge number of parking tickets.
In an attempt to scientifically analyse this phenomenon, the NYC Police Department has collected data for parking tickets.

For the scope of this analysis, we wish to compare the phenomenon related to parking tickets over three different years - 2015, 2016, 2017.

Questions

Q1. Find the total number of tickets for each year



- Number of tickets over 2015-17 decreased.
- 2015 has 10008087tickets
- 2016 has 10146977 tickets
- 2017 has 5377542 tickets

Q2. Find out the number of unique states from where the cars that got parking tickets came from. *(Hint:*

Use the column 'Registration State')

There is a numeric entry in the column which should be corrected. Replace it with the state having

maximum entries. Give the number of unique states for each year again.

- Number of unique states from where cars got parking tickets = **69 states**

	Registration State	Record_count		Registration State	Record_count
1	NY	20024087	43	NM	7013
2	NJ	2266677	44	AR	6928
3	PA	638387	45	WV	6172
4	CT	339772	46	NV	5161
5	FL	330976	47	SD	4462
6	MA	214649	48	NE	3610
7	IN	193528	49	KS	3442
8	VA	172940	50	UT	2746
9	MD	142111	51	NS	2502
10	NC	130781	52	MT	2208
11	IL	87417	53	AK	2056
12	GA	84077	54	GV	1895
13	TX	80532	55	ND	1511
14	99	73154	56	HI	1148
15	AZ	62025	57	WY	877
16	OH	57891	58	AB	583
17	CA	54938	59	PR	425
18	SC	51051	60	BC	365
19	OK	47120	61	PE	306
20	ME	41600	62	NB	292
21	MI	39463	63	MB	82
22	MN	38854	64	SK	53
23	TN	38680	65	FO	26
24	DE	37643	66	MX	13
25	RI	27074	67	NT	10
26	NH	21780	68	YT	5
27	AL	17090	69	NF	2
28	VT	16976			
29	WA	15428			
30	OR	13828			
31	ON	12838			
32	QB	11166			
33	IA	10182			
34	DC	9982			
35	KY	9800			
36	MO	9703			
37	ID	9585			
38	CO	9484			
39	WI	9411			
40	MS	9175			
41	DP	9173			
42	LA	7685			

- Number of NY entries 20024087and Number of 99 entries 73154

- Numeric entry in the column ‘Registration State’ is 99 and NY is the maximum entries and hence replacing Registration State 99 with ‘NY’

- After replacement, Number of unique states in parking tickets data year wise :

Issue_Date_Year	Count (# of unique states)
2015	68
2016	67
2017	64

Q3. 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.

The values should not be deleted or imputed here. This is just a check.

- `4448923` entries do not have address for violation location
- `17.4% records have violation location missing from the records`
- Year 2016 has highest records with no violation location

Issue_Date_Year	Count (#)
2015	1555016
2016	1970527
2017	923380
Total	**4448923**

Aggregation tasks

Q1. How often does each violation code occur? Display the frequency of the top five violation codes.

Violation Code	Total
21	3595056
36	2964148
38	2749510
14	2138952
37	1594604

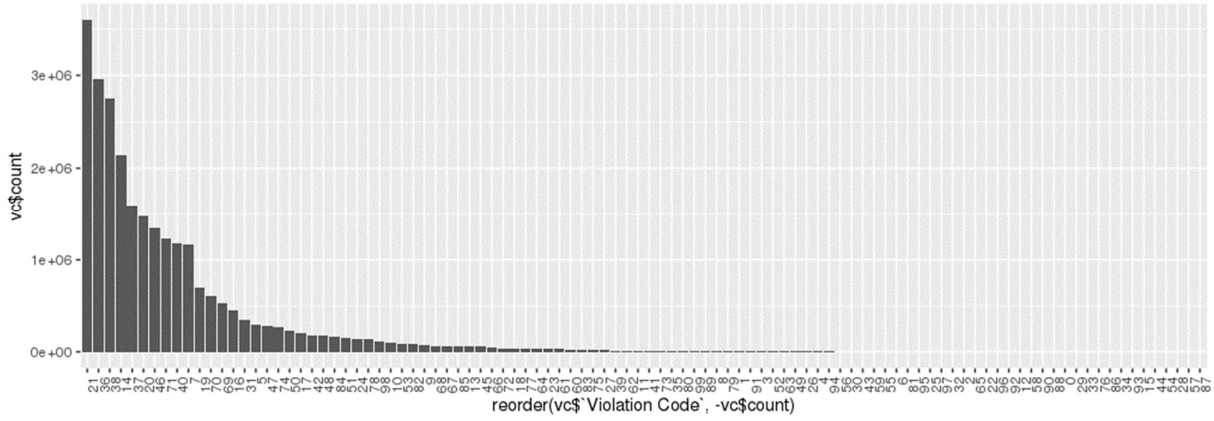
- 21 is most frequent violation code

Year 2015	
Violation Code	Total
21	1425779
38	1143343
36	951024
14	851119
37	668394

Year 2016	
Violation Code	Total
21	1409997
36	1351297
38	1066339
14	815800
37	633584

Year 2017	
Violation Code	Total
21	759280
36	661827
38	539828
14	472033
20	317551

Plots highlighting above data



Q2. How often does each 'vehicle body type' get a parking ticket? How about the 'vehicle make'? (***Hint***: *find the top 5 for both)*

Vehicle Body Type	Count	Vehicle Make	Count
SUBN	8551193	FORD	3155810
4DSD	7296843	TOYOT	2801819
VAN	3571103	HONDA	2486750
DELV	1754513	NISSA	2077591
SDN	994615	CHEVR	1799535

- SUBN, 4DSD, VAN, DELV and SDN are most common vehicle body type to get a parking ticket
- Ford, Toyota, Honda, Nissan and Chevrolet are most common vehicle makes to get a parking ticket

Year-wise Analysis (Top 5 Vehicle Body Type)

Year 2015	
Vehicle Body Type	Count
SUBN	3245663
4DSD	2862109
VAN	1448572
DELV	734937
SDN	390372

Year 2016	
Vehicle Body Type	Count
SUBN	3425471
4DSD	2888670
VAN	1403728
DELV	667754
SDN	413483

Year 2017	
Vehicle Body Type	Count
SUBN	1880059
4DSD	1546064
VAN	718803
DELV	351822
SDN	190760

Year-wise Analysis (Top 5 Vehicle Make)

Year 2015	
Vehicle Make	Count
FORD	1267693
TOYOT	1065459
HONDA	952177
NISSA	780454
CHEVR	748049

Year 2016	
Vehicle Make	Count
FORD	1252931
TOYOT	1132494
HONDA	997047
NISSA	836343
CHEVR	696322

Year 2017	
Vehicle Make	Count
FORD	635186
TOYOT	603866
HONDA	537526
NISSA	460794
CHEVR	355164

Q3. 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:

1. '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?

Violation Precinct	Count
0	4448923
19	1320716
14	830512
18	790596
1	750644
114	712183

- Violation Precinct as ‘0’ are erroneous entries
- 19,14,18,1,114 are top 5 Violation precincts where violations occur the most with 19 being consistently highest

Yearwise Analysis (Violation Precinct)

Year 2015	
Violation Precinct	Count
19	526252
18	340438
14	334275
114	286258
1	273800

Year 2016	
Violation Precinct	Count
19	522311
1	304737
14	295558
18	284146
114	279142

Year 2017	
Violation Precinct	Count
19	272153
14	200679
1	172107
18	166012
114	146783

- 19 precinct is the highest across year with most number of violations, however, number of violations have been decreasing
- Precinct 18 violations have improved (decreased significantly) over the years

2. 'Issuer Precinct' (this is the precinct that issued the ticket)

*Here you would have noticed that the data frame 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)*

Issuer Precinct	Count
0	5061166
19	1288533
14	811162
18	771257
1	731608
114	699980

- Issuer Precinct as ‘0’ are erroneous entries
- 19,14,18,1,114 are top 5 Issuer precincts where violations occur the most with 19 being consistently highest
- **Yearwise Analysis (Issuer Precinct)**

Year 2015	
Issuer Precinct	Count
19	513583
18	334355
14	325592
114	282353
1	268316

Year 2016	
Issuer Precinct	Count
19	510002
1	296457
14	287561
18	276548
114	274101

Year 2017	
Issuer Precinct	Count
19	264948
14	198009
1	166835
18	160354
114	143526

Q4. 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**: You can analyse the three precincts together using the 'union all' attribute in SQL view. In the SQL view, use the 'where' attribute to filter among three precincts and combine them using 'union all'.*

- 19, 14 and 18 are top three precincts which have issued the most number of tickets

Issuer Precinct wise violation code

Issuer Precinct 19	
Violation Code	Count
38	186785
37	182166
46	177442
14	145387
21	133729

Issuer Precinct 14	
Violation Code	Count
14	166912
69	160205
31	92628
47	66760
42	56359

Issuer Precinct 18	
Violation Code	Count
14	239395
69	107960
47	59707
31	56661
42	38060

- Violation code 14 is common across top three precincts 19,14 and 18
- Top 5 Violation codes 14,69,31,47 and 42 are common between 14 and 18 issuer precincts

Year wise analysis - Issuer Precinct wise violation code

Year 2015					
Issuer Precinct 19		Issuer Precinct 14		Issuer Precinct 18	
Violation Code	Count	Violation Code	Count	Violation Code	Count
38	76862	69	71005	14	104232
37	71766	14	67082	69	50557
14	59317	31	35147	47	26188
46	57692	47	24683	31	24848
21	53514	42	23738	42	17451

Year 2016					
Issuer Precinct 19		Issuer Precinct 14		Issuer Precinct 18	
Violation Code	Count	Violation Code	Count	Violation Code	Count
37	74379	69	58999	14	85542
38	73734	14	55300	69	37369
46	72431	31	34980	31	19959
14	56463	47	23770	47	19469
21	51844	42	22608	42	14442

Year 2017					
Issuer Precinct 19		Issuer Precinct 14		Issuer Precinct 18	
Violation Code	Count	Violation Code	Count	Violation Code	Count
46	47319	14	44530	14	49621
38	36189	69	30201	69	20034
37	36021	31	22501	47	14050
14	29607	47	18307	31	11854
21	28371	42	10013	46	7595

Q5 You'd want to find out the properties of parking violations across different times of the day:

- 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.*
- 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.
- 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.*

- Now, try another direction. For the 3 most commonly occurring violation codes, find the most common time of the day (in terms of the bins from the previous part)
    - No null entries in Violation time
    - Violation time field is mentioned in “0942A”, “0318P”, “0410P”, “0259P”,”0459P” etc.,
    - Violation time “0942A” can be read as 9(hr) – 42(min) in the morning (A should be read as AM)
    - After converting to 24-hour time format, we try to use max and min hour to check on values
    - There are 2343248 entries which are not meaningful ie has hour value over 23
    - Total 24 hour time is divided into 6 time bins as
        - Hour 0 – Hour 3 as 1
        - Hour 4 – Hour 7 as 2
        - Hour 8 – Hour 11 as 3
        - Hour 12 – Hour 15 as 4
        - Hour 16 – Hour 19 as 5
        - Hour 20 – Hour 23 as 6
- **Most common violation codes are 21, 36 and 38**
- **All three common violations have Violation Time Bucket as 3 which is Morning 8 to 11, which indicate that major violations happen during this time bucket**

Top 5 violation code by time bucket and code		
Violation Code	Violation Time Bucket	Count
21	3	2825191
36	3	1483548
38	3	928803
38	4	899625
36	4	848969

Year wise Top violation code by time bucket and code

Top 5 violation code by time bucket and code (year 2015)		
Violation Code	Violation Time Bucket	Count
21	3	1133017
36	3	449311
38	3	398052
38	4	367090
37	4	289704

Top 5 violation code by time bucket and code (year 2016)		
Violation Code	Violation Time Bucket	Count
21	3	1099915
36	3	686587
36	4	382783
38	3	355058
38	4	348430

Top 5 violation code by time bucket and code (year 2017)		
Violation Code	Violation Time Bucket	Count
21	3	592259
36	3	347650
38	4	184105
36	4	184050
38	3	175693

- **21, 36 and 38 are most common violations across years and happen during time bucket 3 which is between 8 to 11 am**

Q6 . 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. *(**Hint**: Use Issue Date to segregate into seasons)*
- 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.)*

Divided whole year into 4 seasons as

Month 1 – Month 3 as 1

Month 4 – Month 6 as 2

Month 7 – Month 9 as 3

Month 10 – Month 12 as 4

Month_Bucket	Count
1	6489169
2	7108573
3	4692555
4	4897743

- **January to March Month Bucket has highest number of violations**

Month Bucket wise violation codes

Month_Bucket = 1	Count
21	824866
38	759550
36	688665

Month_Bucket = 2	Count
21	1033621
36	740325
38	727504

Month_Bucket = 3	Count
21	713757
38	493050
14	405934

Month_Bucket = 4	Count
36	751179
21	699299
38	482932

- **21, 38, 36 and 14 are most common violations across month buckets**

Year wise analysis (Month bucket wise violation codes)

Year	Month Bucket	Violation Code	Count
2015	1	38	226022
2015	1	21	175707
2015	1	14	159197
2015	2	21	369613
2015	2	38	276722
2015	2	14	214645
2015	3	21	367950
2015	3	38	272166
2015	3	14	217383
2015	4	21	388242
2015	4	36	375921
2015	4	38	245213
2016	1	21	315896
2016	1	36	294616
2016	1	38	276697
2016	2	21	309932
2016	2	36	285092
2016	2	38	223546
2016	3	21	345564
2016	3	38	220875

2016	3	36	204171
2016	4	36	375258
2016	4	21	310941
2016	4	38	237713
2017	1	21	333263
2017	1	36	293799
2017	1	38	256831
2017	2	21	354076
2017	2	36	266183
2017	2	14	232710
2017	3	21	243
2017	3	46	202
2017	3	40	112
2017	4	46	209
2017	4	40	132
2017	4	21	116

- **Violation codes analysis**
- **21, 38 are common for first 3 months across years**
- **14 for first year followed by 36 for remaining months are the norm**

- **21 is the category for 2nd set of months in all the three years**
- **with difference being remarkably high in 2015 & 2017**
- **38,36,14 share 2 slot each across the years**

- **Year 2017 have very low counts for this month**
- **21,38 again feature as dominating the for years 2015 and 2016**
- **14, 36 are in 3rd ranking for these two years respectively**

- **46 and 40 are the odd one for 2017**
- **2017 observation for 4th set is same as for 3rd month**
- **21,36,38 are the are present in both 2015 and 2016**

Q7 . 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.

- Find total occurrences of the three most common violation codes
- Then, 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.

- Using this information, find the total amount collected for the three violation codes with maximum tickets. State the code which has the highest total collection.
- What can you intuitively infer from these findings?

Top three violations are 21, 36 and 38

Violation Code	Total
21	3271543
36	2544207
38	2463036

Violation Code	Manhattan_fine	all_others	avg_fine	Revenue	Definition of Violation Code
21	65	45	55	179934865	Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.
36	50	50	50	127210350	Exceeding the posted speed limit in or near a designated school zone.
38	65	35	50	123151800	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.

- **Revenue calculated as Number of violations multiplied by Average fine**
- **Top Violation 21 has generated ~180 million dollars of revenue followed by 36 (127million) and 38 (123 million)**
- **No parking is most common violation and largest source of revenue for NYC**

Issue Year	Violation Code	REC_COUNT	Avg Fine	**Revenue**
2015	21	1301512	55	71583160
2015	38	1020123	50	51006150
2015	36	825088	50	41254400
			Total	**163843710**
2016	21	1282333	55	70528315
2016	36	1159137	50	57956850
2016	38	958831	50	47941550
			Total	**176426715**
2017	21	687698	55	37823390
2017	36	559982	50	27999100
2017	38	484082	50	24204100

			Total	**90026590**

- **Overall amount derived from fines is decreasing (in line with fall in number of tickets)**
- **2015 has highest revenue amount at 164 million which has nearly halved by 2017 to 90 million dollars only**