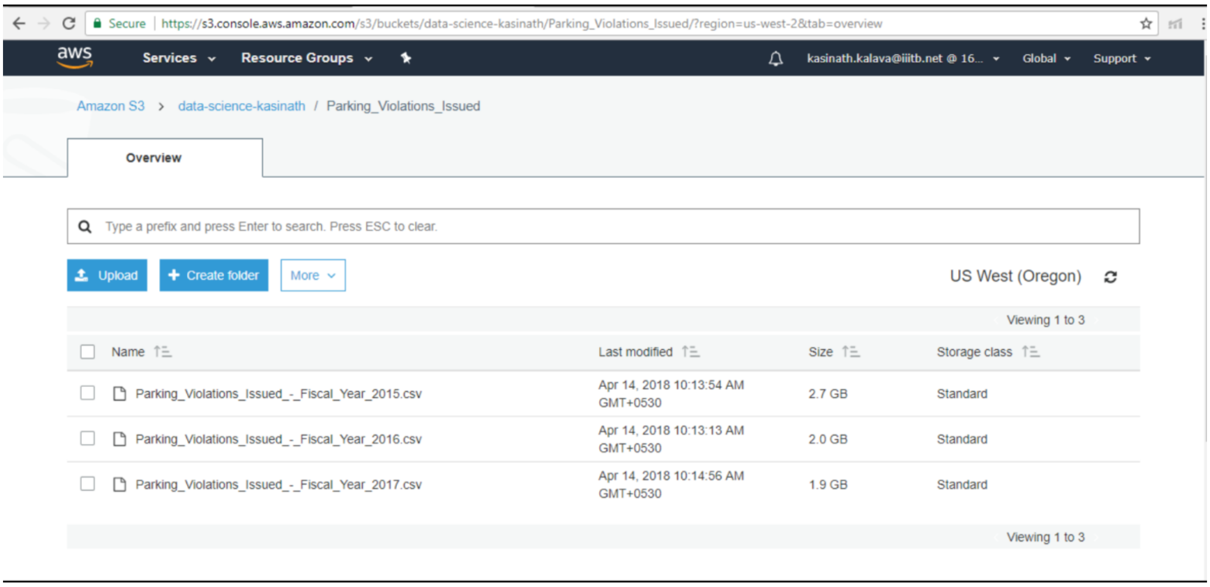
S3 BUCKET SCREENSHOTs of group members



https://s3.console.aws.amazon.com/s3/buckets/data-science-kasinath/Parking_Violations_Issued/?region=us-west-2&tab=overview

Amazon S3 > data-science-kasinath / Parking_Violations_Issued

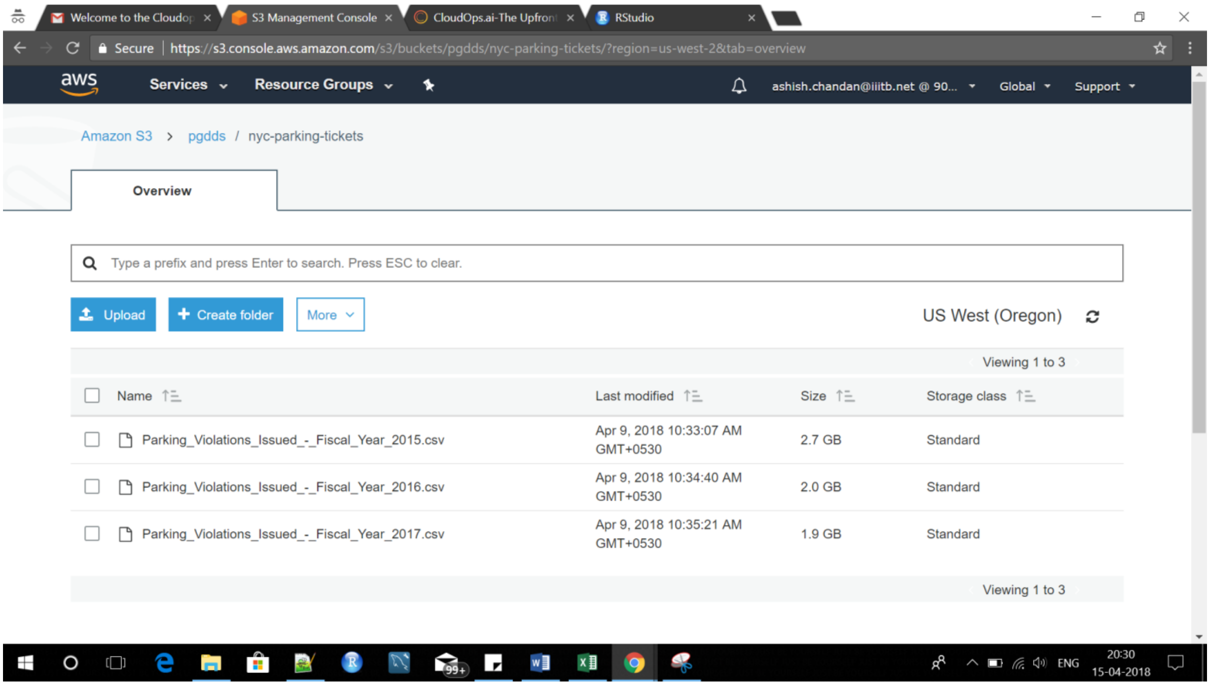
Type a prefix and press Enter to search. Press ESC to clear.

US West (Oregon)

Viewing 1 to 3

	Name	Last modified	Size	Storage class
□	Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Apr 14, 2018 10:13:54 AM GMT+0530	2.7 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Apr 14, 2018 10:13:13 AM GMT+0530	2.0 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Apr 14, 2018 10:14:56 AM GMT+0530	1.9 GB	Standard

Viewing 1 to 3



https://s3.console.aws.amazon.com/s3/buckets/pgdds/nyc-parking-tickets/?region=us-west-2&tab=overview

Amazon S3 > pgdds / nyc-parking-tickets

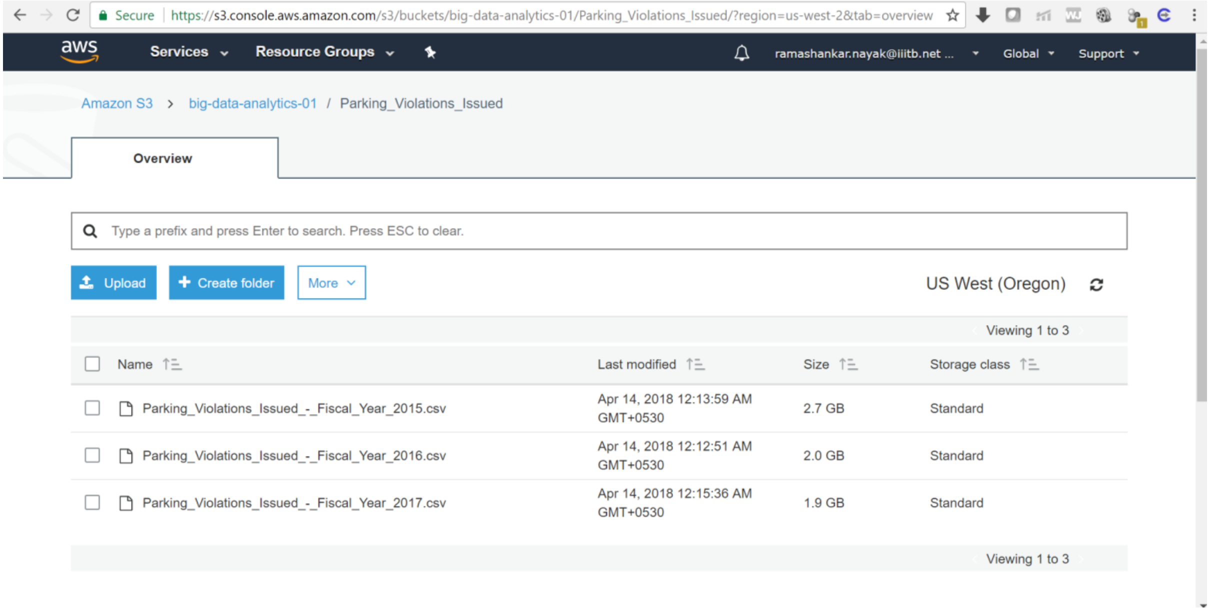
Type a prefix and press Enter to search. Press ESC to clear.

US West (Oregon)

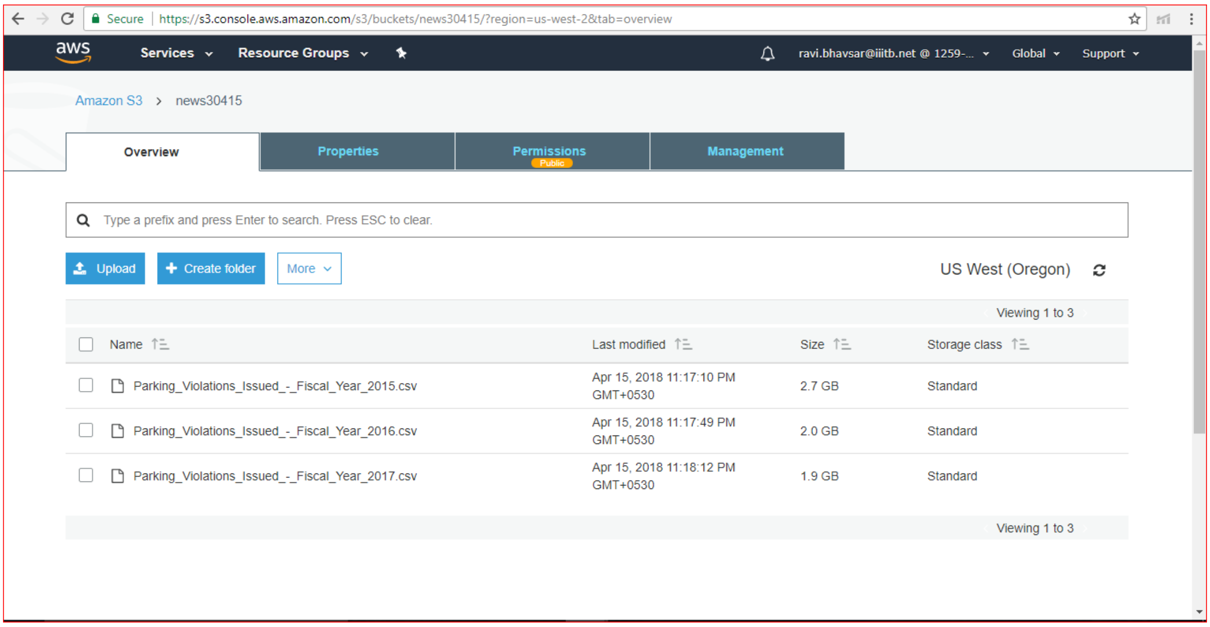
Viewing 1 to 3

	Name	Last modified	Size	Storage class
□	Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Apr 9, 2018 10:33:07 AM GMT+0530	2.7 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Apr 9, 2018 10:34:40 AM GMT+0530	2.0 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Apr 9, 2018 10:35:21 AM GMT+0530	1.9 GB	Standard

Viewing 1 to 3


Secure
https://s3.console.aws.amazon.com/s3/buckets/big-data-analytics-01/Parking_Violations_Issued/?region=us-west-2&tab=overview
aws
Services
Resource Groups
ramashankar.nayak@iiitb.net ...
Global
Support
Amazon S3 > big-data-analytics-01 / Parking_Violations_Issued
Overview
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
More
US West (Oregon)
Viewing 1 to 3

□	Name	Last modified	Size	Storage class
□	Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Apr 14, 2018 12:13:59 AM GMT+0530	2.7 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Apr 14, 2018 12:12:51 AM GMT+0530	2.0 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Apr 14, 2018 12:15:36 AM GMT+0530	1.9 GB	Standard

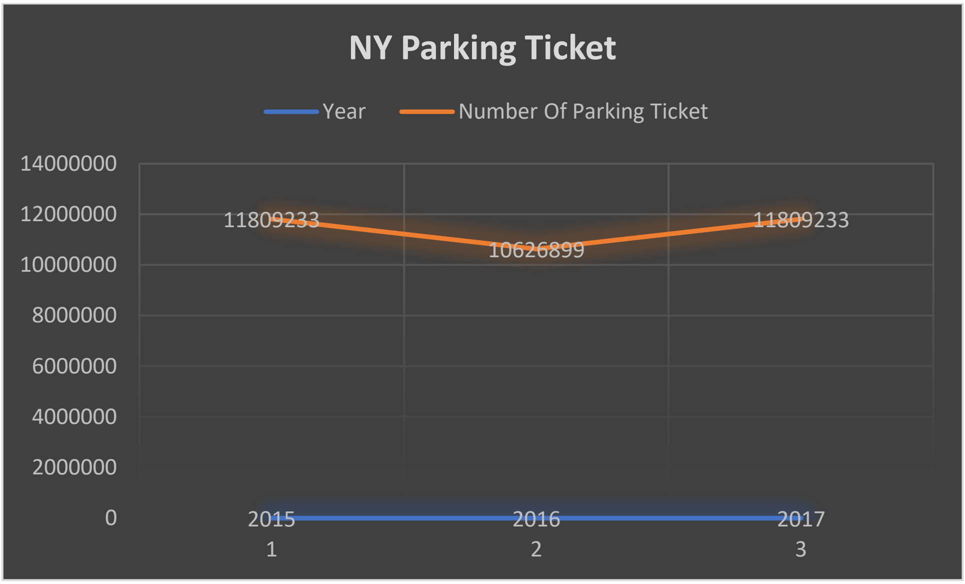

Secure
https://s3.console.aws.amazon.com/s3/buckets/news30415/?region=us-west-2&tab=overview
aws
Services
Resource Groups
ravi.bhavsar@iiitb.net @ 1259-...
Global
Support
Amazon S3 > news30415
Overview
Properties
Permissions
Public
Management
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
More
US West (Oregon)
Viewing 1 to 3

□	Name	Last modified	Size	Storage class
□	Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Apr 15, 2018 11:17:10 PM GMT+0530	2.7 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Apr 15, 2018 11:17:49 PM GMT+0530	2.0 GB	Standard
□	Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Apr 15, 2018 11:18:12 PM GMT+0530	1.9 GB	Standard

Examine the data.

1. Find total number of tickets for each year.

Year	Number of Parking Ticket
2015	11809233
2016	10626899
2017	11809233



We have assumed that all the data in files Parking_Violations_Issued_-_Fiscal_Year_2015.csv, Parking_Violations_Issued_-_Fiscal_Year_2016.csv and Parking_Violations_Issued_-_Fiscal_Year_2017.csv are of 2015,2016 and 2017 respective years.

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

Year	Unique States Count
2015	69
2016	68
2017	67

Registration State variable is takin to determine the state of a car.

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

Year	Missing Address Count	% of total
2015	1992401	16.87155
2016	2035232	19.1517
2017	2289944	19.39113

We have considered null/empty value in house number or Street Name as missing address.

Aggregation tasks

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

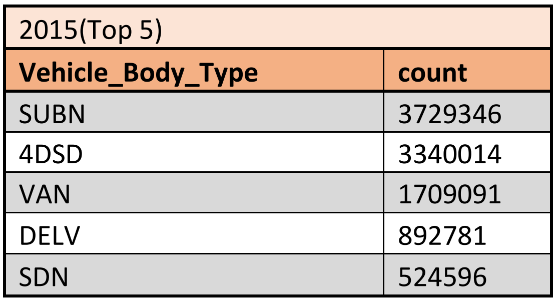
2015(Top 5)	
Violation_Code	**Count**
21	1630912
38	1418627
14	988469
36	839197
37	795918

2016(Top 5)	
Violation_Code	**Count**
21	1531587
36	1253512
38	1143696
14	875614
37	686610

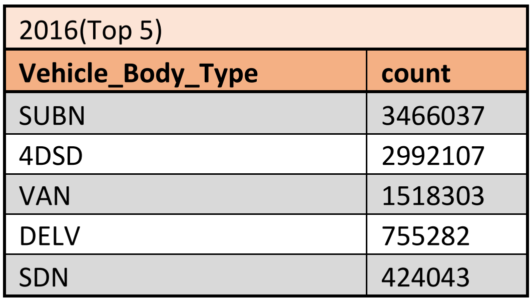
2017(Top 5)	
Violation_Code	**Count**
21	1528588
36	1400614
38	1062304
14	893498
20	618593

Violation_Code 21,36,38 and 14 are in top 5 violation code for the three years.

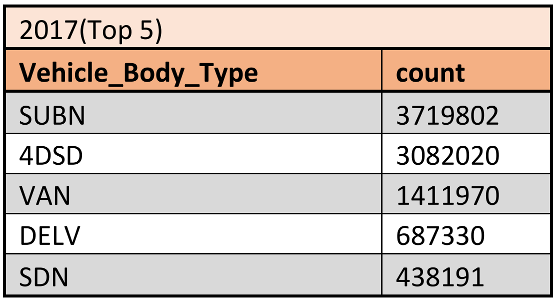
2. How often does each vehicle body type get a parking ticket? How about the vehicle make? (find the top 5 for both)



2015(Top 5)	
Vehicle_Body_Type	**count**
SUBN	3729346
4DSD	3340014
VAN	1709091
DELV	892781
SDN	524596

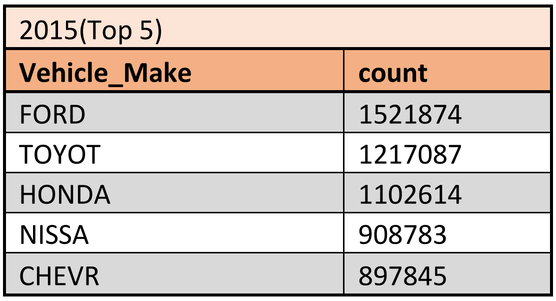


2016(Top 5)	
Vehicle_Body_Type	**count**
SUBN	3466037
4DSD	2992107
VAN	1518303
DELV	755282
SDN	424043

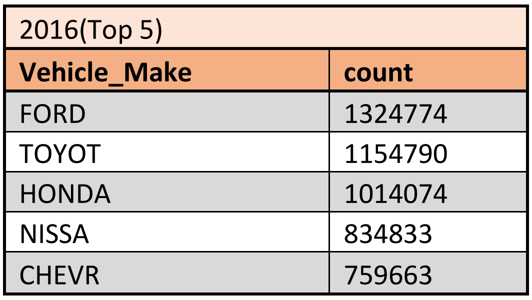


2017(Top 5)	
Vehicle_Body_Type	**count**
SUBN	3719802
4DSD	3082020
VAN	1411970
DELV	687330
SDN	438191

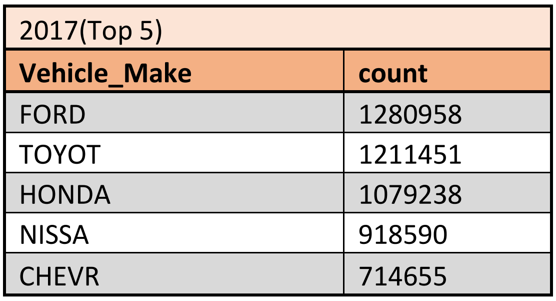
Vehicle body type SUBN, 4DSD, VAN, DELV and SDN occupies the top five positions for all three years



2015(Top 5)	
Vehicle_Make	**count**
FORD	1521874
TOYOT	1217087
HONDA	1102614
NISSA	908783
CHEVR	897845



2016(Top 5)	
Vehicle_Make	**count**
FORD	1324774
TOYOT	1154790
HONDA	1014074
NISSA	834833
CHEVR	759663



2017(Top 5)	
Vehicle_Make	**count**
FORD	1280958
TOYOT	1211451
HONDA	1079238
NISSA	918590
CHEVR	714655

Vehicle make FORD, TOYOT, HONDA, NISSA and CHEVR occupies the top five positions for all three years

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

1. Violating Precincts (this is the precinct of the zone where the violation occurred)
2. Issuing Precincts (this is the precinct that issued the ticket)

2015(Top 5)	
Violation_Precinct	**count**
0	1799170
19	598351
18	427510
14	409064
1	329009

2016(Top 5)	
Violation_Precinct	**count**
0	1868655
19	554465
18	331704
14	324467
1	303850

2017(Top 5)	
Violation_Precinct	**count**
0	2072400
19	535671
14	352450
1	331810
18	306920

Violation precint 0 and 19 are in top two

2015(Top 5)	
Issuer_Precinct	**count**
0	2037745
19	579998
18	417329
14	392922
1	318778

2016(Top 5)	
Issuer_Precinct	**count**
0	2140274
19	540569
18	323132
14	315311
1	295013

2017(Top 5)	
Issuer_Precinct	**count**
0	2388479
19	521513
14	344977
1	321170
18	296553

# Issuer precint 0 and 19 are in top two

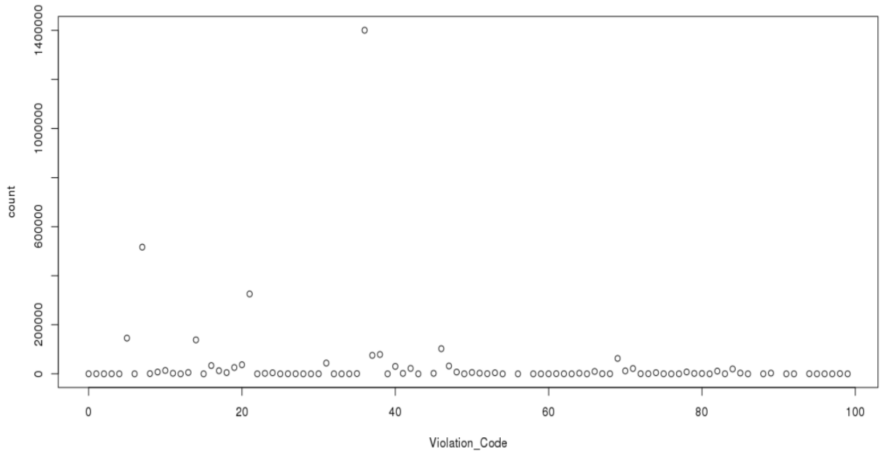
4. Find the violation code frequency across 3 precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?

For 2015, top 3 precincts which has issued most number of tickets are 0,19, and 18.

Across these three precincts, top 5 violation codes are:

2015(Top 5)	
Violation_Code	**count**
36	839197
7	719747
21	276205
5	224517
14	198228

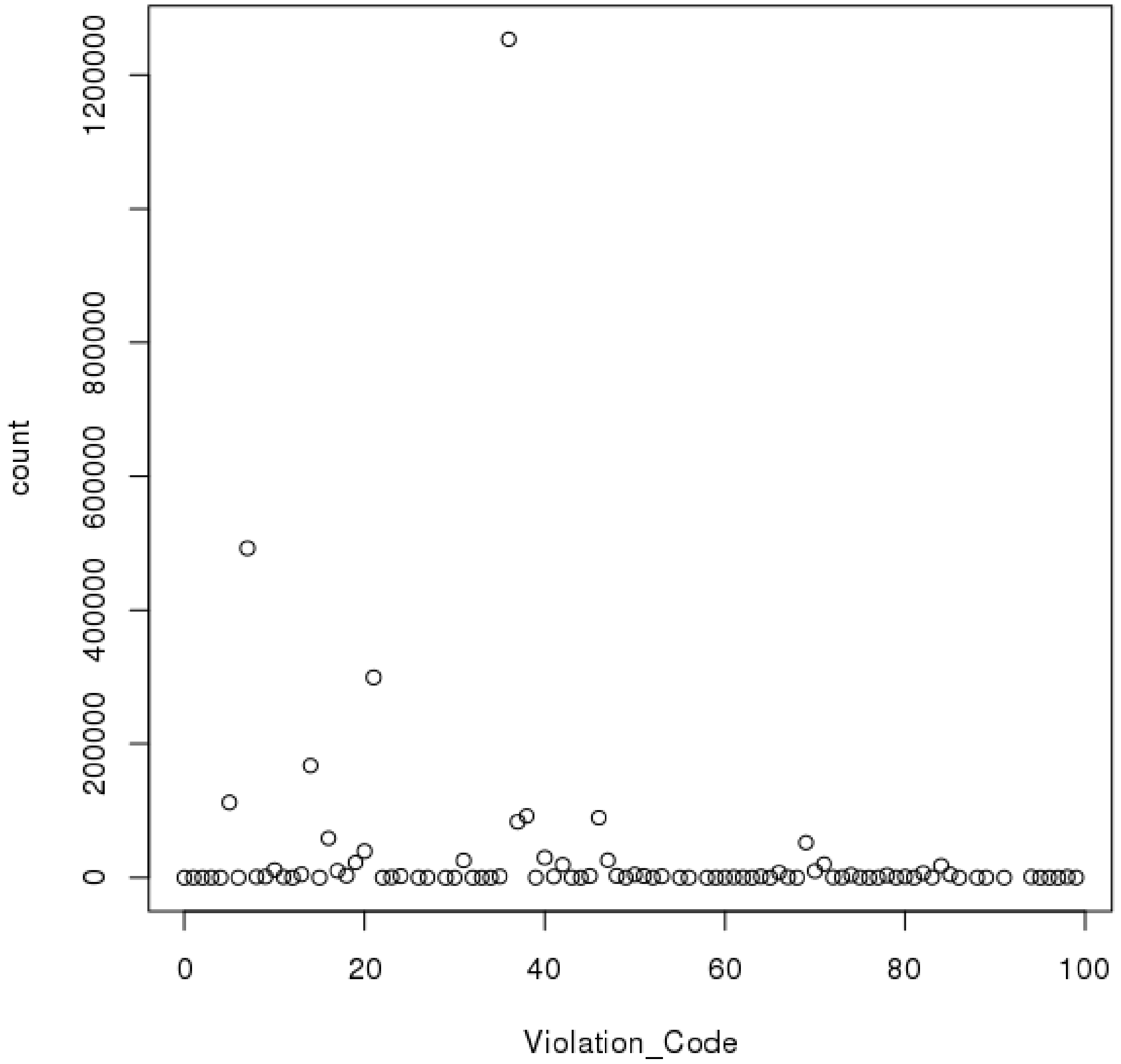
# Below plot is for Issuer Precincts 0,19 and 18



For 2016, top 3 precincts which has issued most number of tickets are 0,19, and 18.

Across these three precincts, top 5 violation codes are:

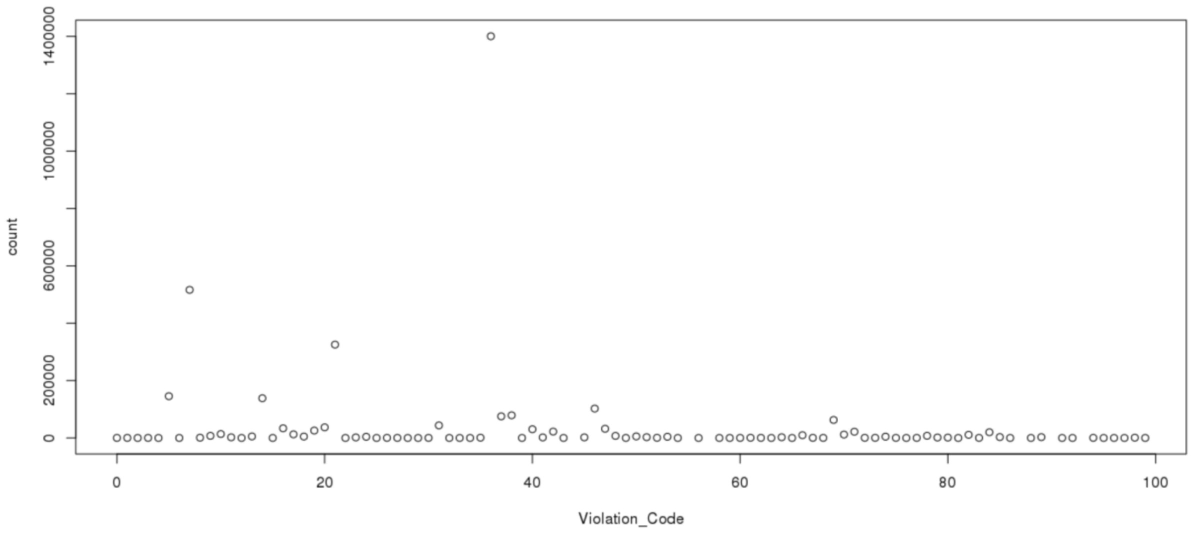
2016(Top 5)	
Violation_Code	**count**
36	1253511
7	492469
21	299409
14	167587
5	112376



For 2017, top 3 precincts which has issued most number of tickets are 0,19, and 14.

Across these three precincts, top 5 violation codes are:

2017(Top 5)	
Violation_Code	**count**
36	1400614
7	516390
21	325435
5	145643
14	138488



2015: For top 3 Issuer_Precinct (0,19,18)			in all Issuer_Precinct	
Violation_Code	count	percentage	count	percentage
7	719747	23.71433	719753	6.094833
36	839197	27.64999	839197	7.106279

2016: For top 3 Issuer_Precinct (0,19,18)			in all Issuer_Precinct	
Violation_Code	count	percentage	count	percentage
7	492469	16.39391	492478	4.634259
36	1253511	41.72841	1253512	11.795652

2017: For top 3 Issuer_Precinct (0,19,18)			in all Issuer_Precinct	
Violation_Code	count	percentage	count	percentage
7	516390	15.86467	516395	4.780095
36	1400614	43.03003	400614	12.965013

# All three years both Violation code 7 and 36 are the most frequent for top 3 issuer precinct and after comparing the percentage of frequency of the two violation code with overall data we can clearly state that the two violation codes are more frequent then overall data

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
- 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)

Day Time in hours	Label
00:00 to 03:59	1
04:00 to 07:59	2
08:00 to 11:59	3
12:00 to 15:59	4
16:00 to 19:59	5
20:00 to 23:59	6

Year	Missing or incorrect date	% of total
2015	1911	0.016182253
2016	4288	0.040350435
2017	84	0.000711308

Missing value percentage is very low, so will just ignore the records for data analysis.

2015		
Violation_Code	**Violation_Time_bin**	**count**
21	1	74053
40	1	47141
78	1	42724
14	2	143264
21	2	118316
40	2	98134
21	3	1291540
38	3	480358
36	3	396838
38	4	609518
37	4	446469
36	4	357306
38	5	258838
37	5	187186
7	5	182347
7	6	89813
38	6	66023
40	6	49928

2016		
Violation_Code	**Violation_Time_bin**	**count**
21	1	72106
40	1	42098
78	1	32806
14	2	140111
21	2	114029
40	2	91692
14	2	140111
21	2	114029
40	2	91692
36	4	545717
38	4	488302
37	4	383361
38	5	211267
37	5	161655
14	5	134976
7	6	60924
38	6	53174
40	6	44973

2017		
Violation_Code	**Violation_Time_bin**	**count**
21	1	77460
40	1	50947
78	1	32243
14	2	141276
21	2	119469
40	2	112186
21	3	1182689
36	3	751422
38	3	346518
36	4	588395
38	4	462758
37	4	337075
38	5	203232
37	5	145784
14	5	144749
7	6	65593
38	6	47029
14	6	44779

For 2015, Violation code 21, 38 and 14 have the maximum frequency.

2015		
Violation_Code	**Violation_Time_bin**	**count**
21	3	1291540
21	4	145374
21	2	118316
38	4	609518
38	3	480358
38	5	258838
14	3	317009
14	4	284944
14	5	160432

For 2016, Violation code 21, 36 and 38 have the maximum frequency.

2016		
Violation_Code	**Violation_Time_bin**	**count**
21	3	1209243
21	4	134329
21	2	114029
36	3	586791
36	4	545717
36	2	79797
38	4	488302
38	3	388099
38	5	211267

For 2017, Violation code 21, 36 and 38 have the maximum frequency.

2017		
Violation_Code	**Violation_Time_bin**	**count**
21	3	1182689
21	4	148013
21	2	119469
36	3	751422
36	4	588395
36	2	33939
38	4	462758
38	3	346518
38	5	203232

- 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

Month	Lable(season)
`01-03	1
`04-06	2
`07-09	3
`10-12	4

Assumption: All the records present in a file belongs to the respective year of file name irrespective of year of issue date of the records.

2015		
season	**Violation_Code**	**count**
1	38	419424
1	21	370713
1	14	271353
2	21	471586
2	38	346719
2	14	262602
3	21	412078
3	38	352481
3	14	240742
4	21	376535
4	38	300003
4	14	213772

2016		
season	**Violation_Code**	**count**
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

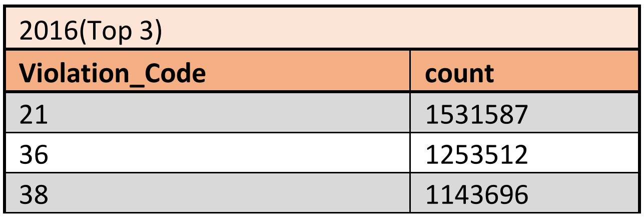
2017		
season	**Violation_Code**	**count**
1	21	374202
1	36	348240
1	38	287017
2	21	421184
2	36	369902
2	38	266909
3	21	385774
3	38	244985
3	36	239879
4	36	442593
4	21	347428
4	38	263393

6. 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.

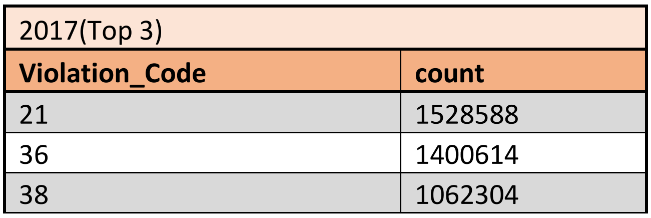
- Find total occurrences of the 3 most common violation codes
- Then, search the internet for NYC parking violation code fines. You will find a website (on the nyc.gov URL) that lists these fines. They’re divided into two categories, one for the highest-density locations of the city, the other for the 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?

Three most frequent Violation Code with count:

2015(Top 3)	
Violation_Code	**count**
21	1630912
38	1418627
14	988469

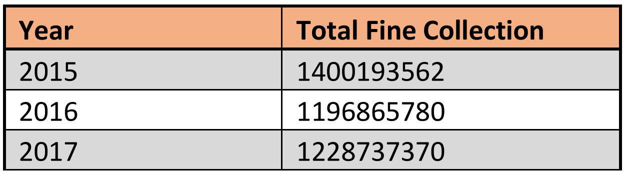


2016(Top 3)	
Violation_Code	**count**
21	1531587
36	1253512
38	1143696



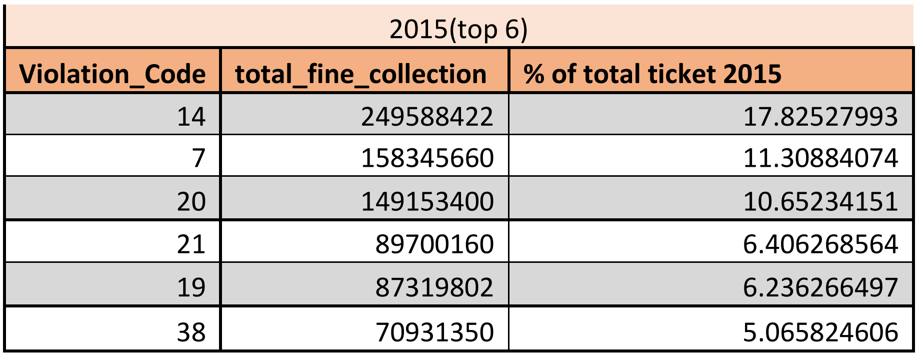
2017(Top 3)	
Violation_Code	**count**
21	1528588
36	1400614
38	1062304

Total Fine collected:

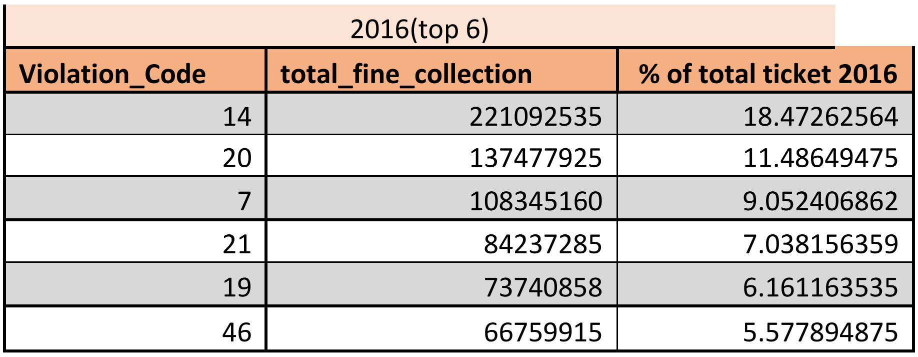


Year	Total Fine Collection
2015	1400193562
2016	1196865780
2017	1228737370

Total fine collected for each Violation code (top 6):



2015(top 6)		
Violation_Code	**total_fine_collection**	**% of total ticket 2015**
14	249588422	17.82527993
7	158345660	11.30884074
20	149153400	10.65234151
21	89700160	6.406268564
19	87319802	6.236266497
38	70931350	5.065824606



2016(top 6)		
Violation_Code	**total_fine_collection**	**% of total ticket 2016**
14	221092535	18.47262564
20	137477925	11.48649475
7	108345160	9.052406862
21	84237285	7.038156359
19	73740858	6.161163535
46	66759915	5.577894875

2017(top 6)		
Violation_Code	**total_fine_collection**	**% of total ticket 2017**
14	225608245	18.36098181
20	139183425	11.32735346
7	113606900	9.245824435
21	84072340	6.842173279
19	73457300	5.978275081
36	70030700	5.699403445

# Violation code 14 has the highest total fine collection

# Violation codes 14,20,7,21,19 and 36 are among the top five for having the most total fine collection and also the percentage fine collection is almost same for all the three years.

################################## END ##