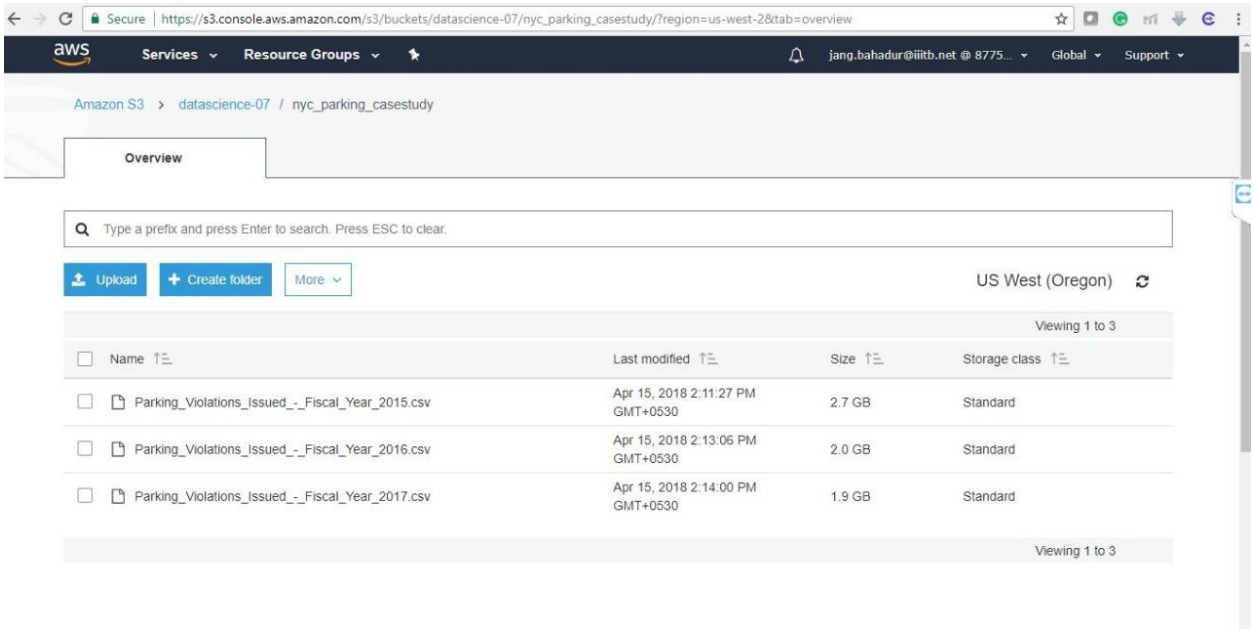
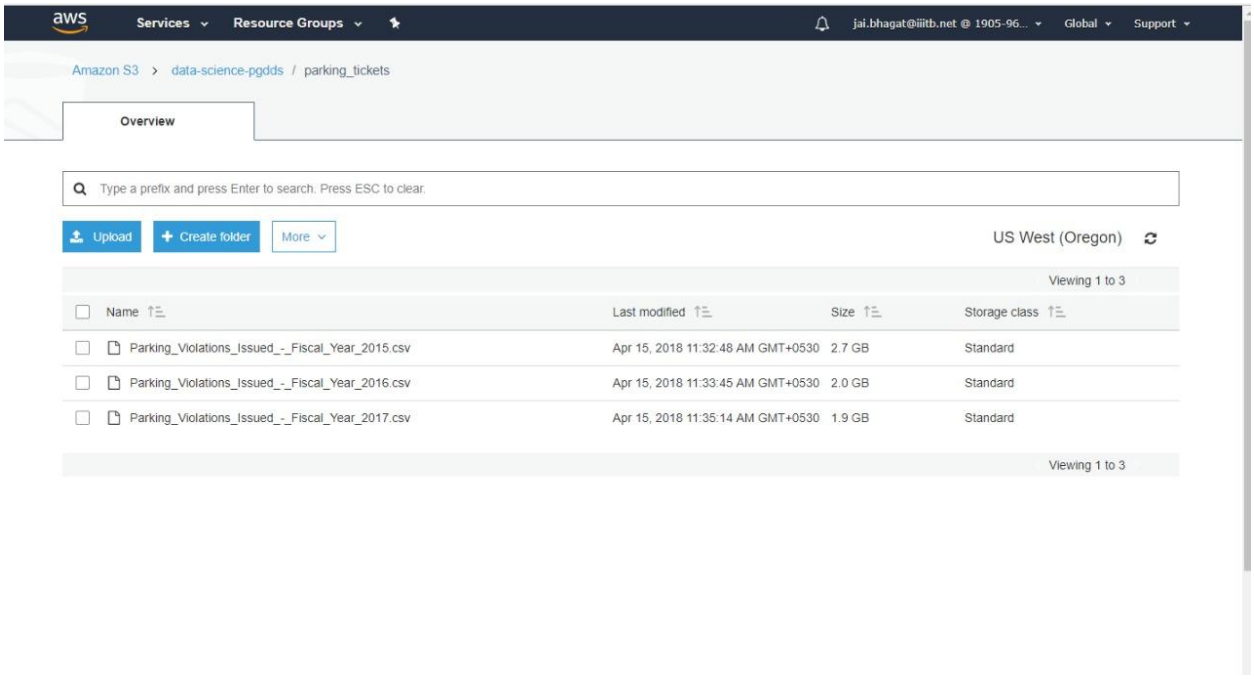
S3 screenshots of four members of the group

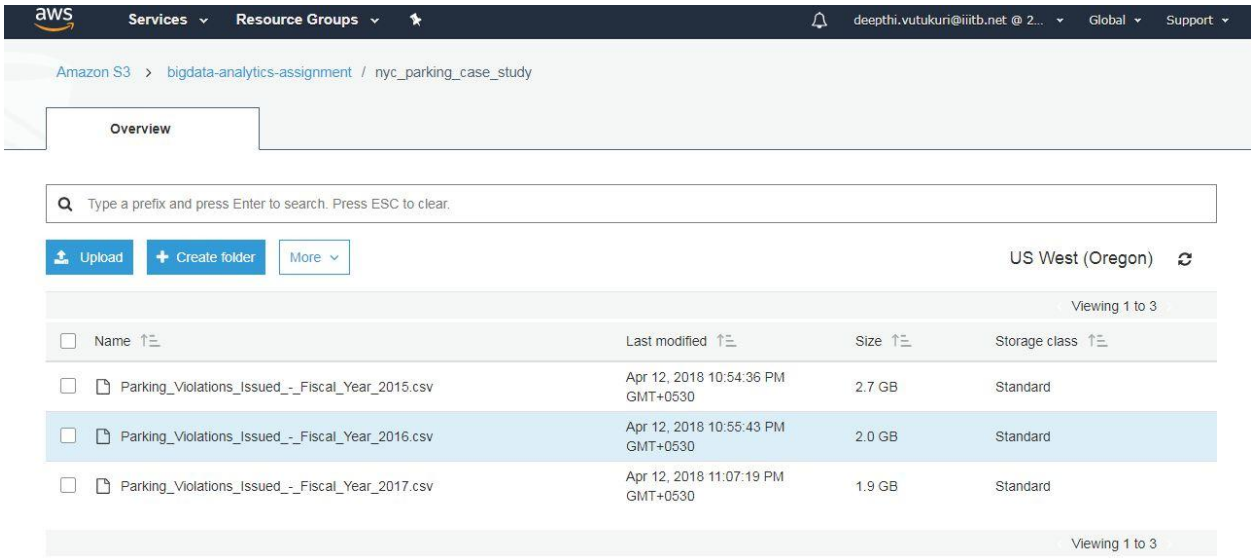
1.Jang Bahadur Umath



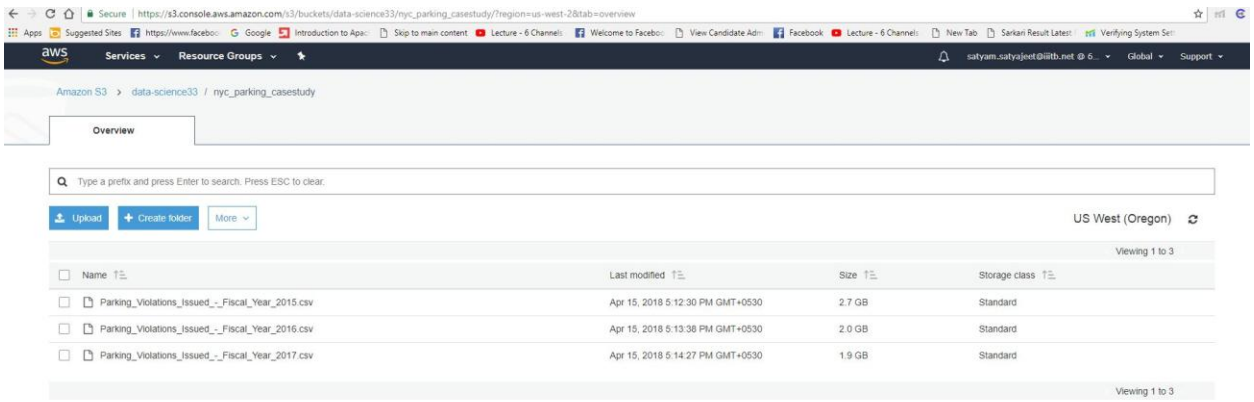
2.Jai Shankar Bhagat



3. Deepthi Vutukuri



4. Satyam Satyajeet



Examine the data - solutions

1. Find total number of tickets for each year.

Ans. After removing duplicate Summon Numbers & filtering for respective years, folloiwng are the number of rows in each dataset

For 2015 – 10903411 Tickets
For 2016 – 10241012 Tickets
For 2017 – 5433018 Tickets

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

Ans. For 2015 – 69 States
For 2016 – 68 States
For 2017 – 65 States

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

Ans. 11029 entries don't have addresses

Aggregation Tasks Solutions

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

Ans. Top 5 violation codes 21,36,38,14,37

#	Violation Code	count
# 1	21	3869197
# 2	36	3111439
# 3	38	2952526
# 4	14	2286502
# 5	37	1699486
# 6	20	1574394

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

Ans. (i) Below are the top 5 vehicle body types

#	Vehicle Body Type	count
# 1	SUBN	9100113
# 2	4DSD	7768379
# 3	VAN	3809438
# 4	DELV	1903396
# 5	SDN	1097903
# 6	2DSD	712067

(ii) Below are the top 5 vehicle make

#	Vehicle Make	count
# 1	FORD	3368392
# 2	TOYOT	2984933
# 3	HONDA	2650193
# 4	NISSA	2213081
# 5	CHEVR	1924180
# 6	FRUEH	1083513

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

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

Ans.

#	Violation Precinct	count
# 1	0	4742644
# 2	19	1410169
# 3	14	894392
# 4	18	854023
# 5	1	805149
# 6	114	756198
# 7	13	695920
# 8	109	582650
# 9	17	520659
# 10	70	481034

3.2. Issuing Precincts (this is the precinct that issued the ticket)

Ans.

#	Issuer Precinct	count
# 1	0	5451818
# 2	19	1372464
# 3	14	870724
# 4	18	831708
# 5	1	781152
# 6	114	742132
# 7	13	680403
# 8	109	589712
# 9	17	507055
# 10	20	474239

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?

Ans. From q3 we concluded that most tickets issued precincts are 19,14,18
so finding the violation code frequency for each of them

For Issuer precinct 19

#	Violation Code	count
# 1	38	201237
# 2	37	193419
# 3	46	188145

4 14 153711
5 21 142611
6 16 116448
For Issuer precinct 14
Violation Code count
1 14 178877
2 69 171058
3 31 98615
4 47 70904
5 42 60347
6 46 29377
For Issuer precinct 18
Violation Code count
1 14 257748
2 69 115832
3 47 63572
4 31 61690
5 42 40750
6 46 38043

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

5.1.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.

5.2 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

Ans. We have divided the time into 6 groups as follows

Early_morning	04 AM to 07 AM
Morning	08 AM to 11 AM
After_Noon	12 PM to 03 PM
Evening	04 PM to 07 PM
Night	08 PM to 11 PM
`Midnight	12 AM to 03 AM

For the above divided groups, the most common occurred violation codes are 21,38,40.

5.3. 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)

Ans. # Time violation_count
1 Morning 4358538
2 After_Noon 1961385
3 Evening 659104

The most common violation codes are occurring mostly in the Morning time i.e. 04 AM to 07 AM.

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

6.1 First, divide the year into some number of seasons, and find frequencies of tickets for each season.

Ans. Dividing the year into four seasons

Summer: June - August

Rainy : September - November

Winter: December - February

Spring: March – May

Frequencies of the tickets for each season is given below

#	seasons	cnt
# 1	Spring	8531399
# 2	Winter	7030570
# 3	Summer	6047215
# 4	Rainy	5803105

6.2. Then, find the 3 most common violations for each of these season

Ans. The most common violations for each season is below

	seasons	vio_code	smn_cnt
1	Spring	21	1211466
2	Spring	36	945698
3	Spring	38	897690
4	Spring	14	725932
5	Spring	37	539357
6	Spring	46	480371
7	Spring	20	480275
8	Summer	21	967365
9	Summer	38	627329
10	Summer	14	519864
11	Summer	36	452576
12	Winter	21	937708
13	Winter	38	840291
14	Winter	36	818799
15	Winter	14	583739
16	Winter	37	439376
17	Rainy	36	894366
18	Rainy	21	752658
19	Rainy	38	587216
20	Rainy	14	456967

7. The fines collected from all the parking violation constitute a revenue source for the NYC police department.

7.1.Let's take an example of estimating that for the 3 most commonly occurring codes.

7.2. Find total occurrences of the 3 most common violation codes

Ans. Total Occurrence = 9933162

Violation Code Count
1 21 3869197
2 36 3111439
3 38 2952526

7.3Then, 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.

Ans. # As per NYC website
Code 21 Avg. Fine = $55
Code 36 Avg. Fine = $50
Code 38 Avg. Fine = $50

7.4. Using this information, find the total amount collected for all of the fines. State the code which has the highest total collection.

Ans. The violation code 21 has the highest collection

#	Violation Code	Count	Fine_Amount	Total_Amount_Collected
# 1	21	3869197	55	212805835
# 2	36	3111439	50	155571950
# 3	38	2952526	50	147626300

Total Amount Collected = 5,16,004,085(in dollars)

7.5. What can you intuitively infer from these findings?

Ans. The most common violation codes for the three years data are 21 , 36 and 38 and among the highest average fine is for code 21 and it is the frequent occurring code(top among 3). So the fines collected from the parking violations is a good source of revenue for NYC Police department and this is happening mostly during Morning time I.e from 04 AM to 07 AM.

Assumptions made for the case study

- As we have given Fiscal Year data from Kaggle, we are asked to do the analysis for the three years 2015,2016 and 2017.So,We have selected the data of 2015, 2016 and 2017 calendar years(i.e from January to December) by filtering it particularly.
- We have removed few columns which are having mostly NA values and are not useful for the data exploration.
- No precinct is named as 0 in NYC .So considering them as blank values in the Violation Precinct column and Issuer Precinct column.