NYC Parking Tickets: An Exploratory Analysis

(Big Data)

Group Name:

1. Deepak Aneja
2. Suresh Balla
3. Merin Jose
4. Fayiz Mayam Veettil

15 Jul 2018

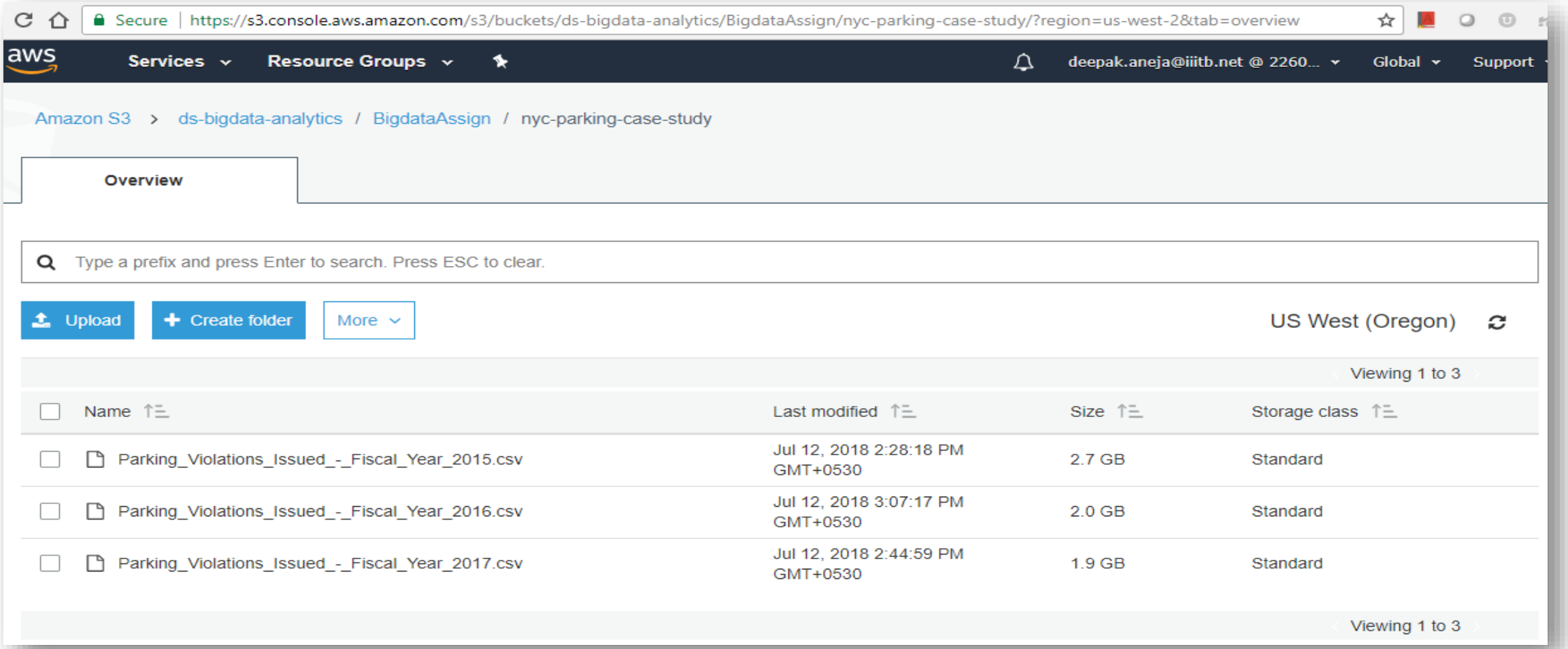
Table of Contents

About the data

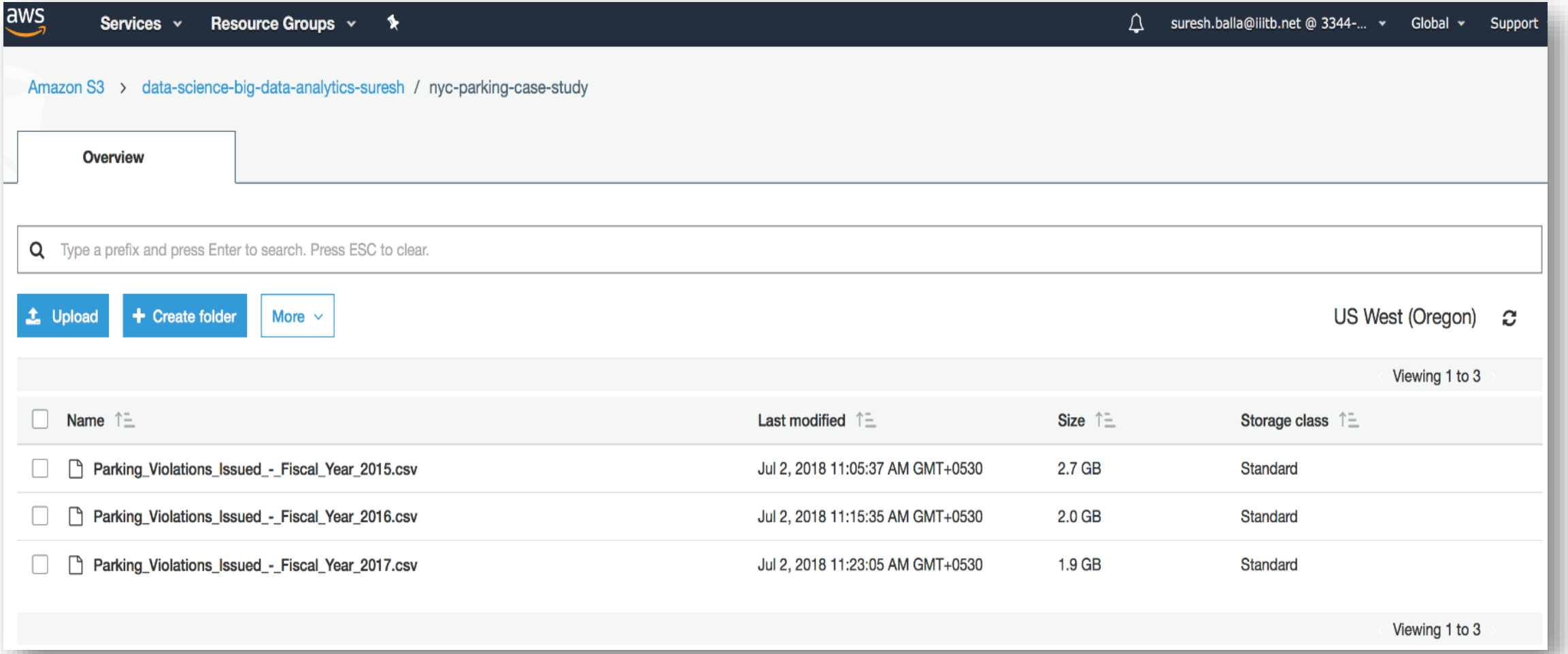
- Data on Parking ticket issued in NYC, collected by NYC Department of Finance
- Data Source: Kaggle

  https://www.kaggle.com/new-york-city/nyc-parking-tickets
- 2015, 2016 and 2017 data sets are used for analysis
- Total 6.6 GB of data across all three years
- The files are roughly organized by fiscal year (July 1 - June 30) of NYC government, hence all the analysis done based on fiscal year and each metric compared across the 3 years.
- External data from nyc.gov for Violation Code Fines are used for revenue analysis

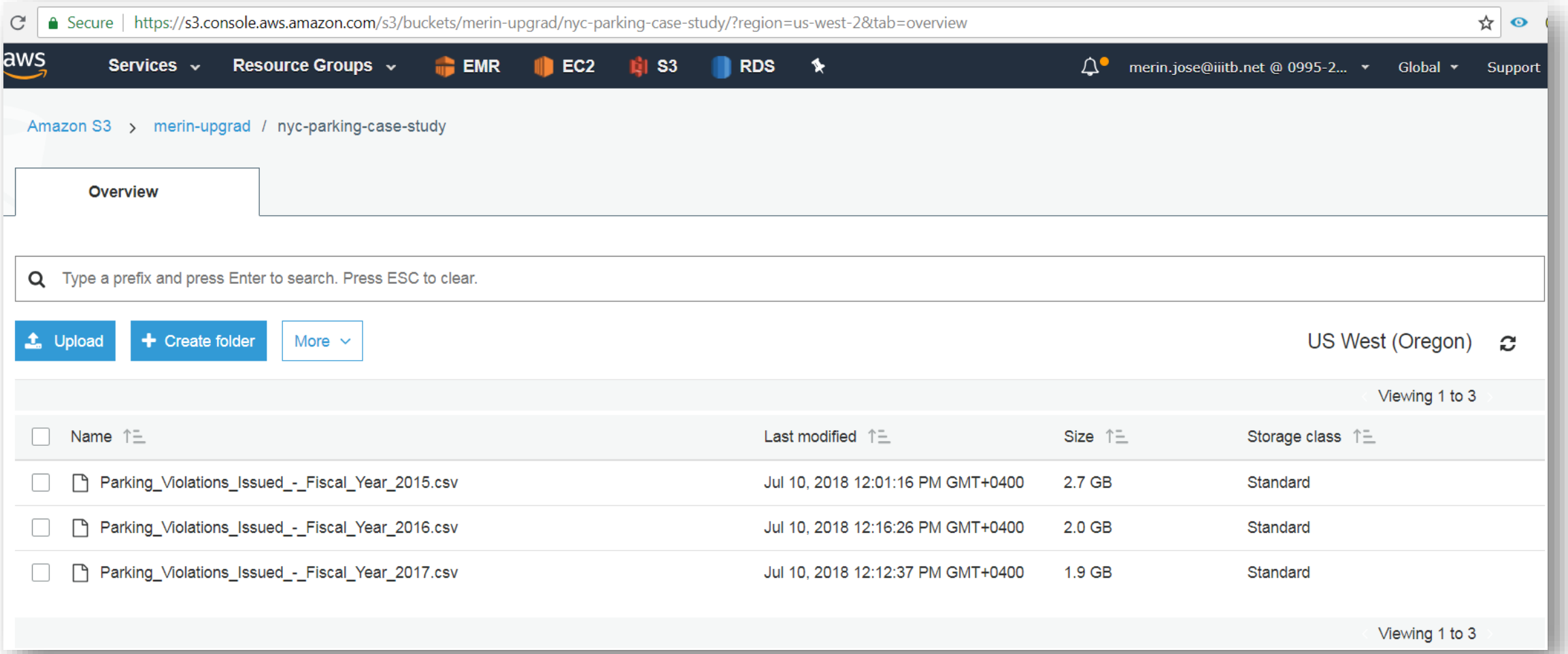
S3 Bucket: Deepak Aneja (Bucket Path: s3://ds-bigdata-analytics/BigdataAssign/nyc-parking-case-study/)



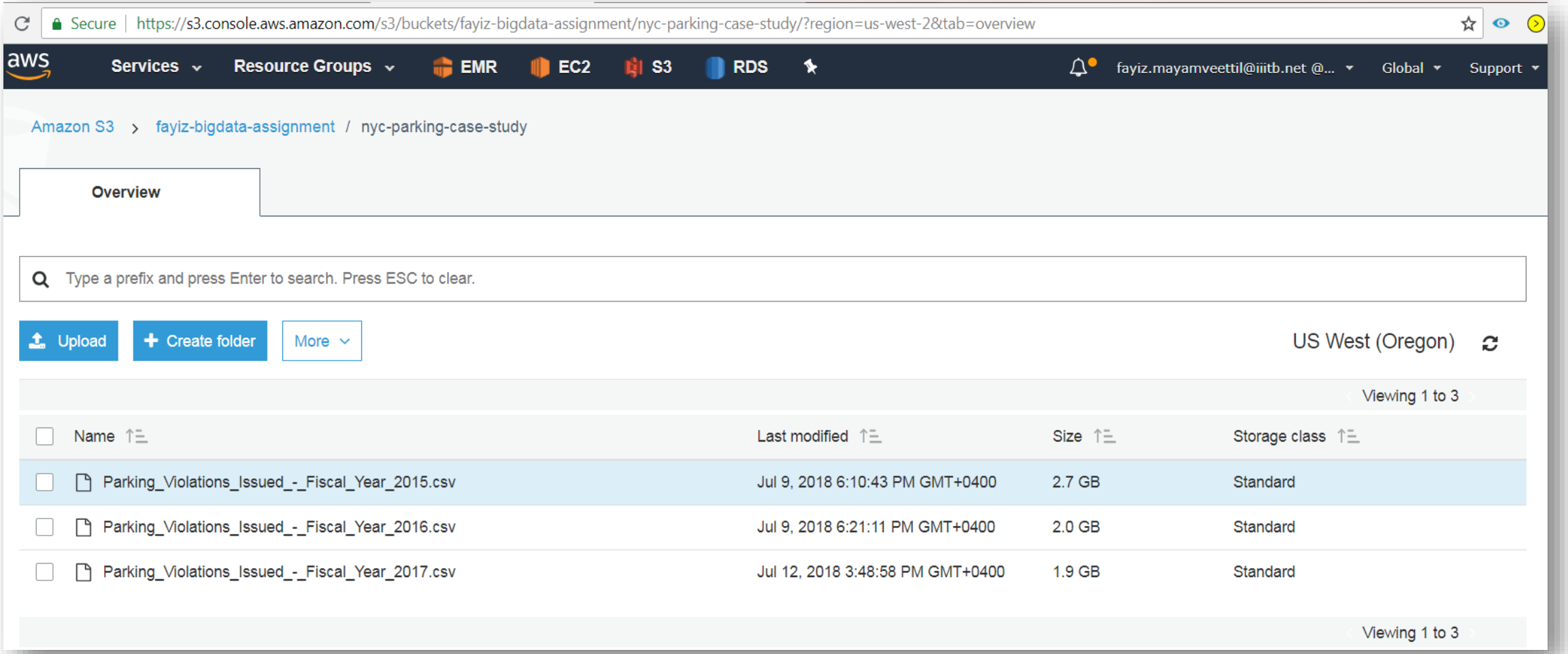
S3 Bucket: Suresh Balla (Bucket Path: s3://data-science-big-data-analytics-suresh/nyc-parking-case-study/)



S3 Bucket: Merin Jose (Bucket Path: s3://merin-upgrad/nyc-parking-case-study/)



S3 Bucket: Fayiz Mayam Veettil (Bucket Path: s3://fayiz-bigdata-assignment/nyc-parking-case-study/)



Exploratory Data Analysis

- Check for possible data inconsistencies
  - Check for duplicate records on Summons Number
  - Check for NA values in the columns used for analysis
- Status of data cleaning
  - 2017 dataset lacks 8 columns that are included in the other 2 datasets
  - Each files contains data from current year July 1st to June 30th next year
  - ~10 M records per year
  - 857,977 number of records found duplicated and cleaned in 2015 data file

Examine the data

1. Find total number of tickets for each year.

Year	No. Of Records	No. Of Tickets after removing duplicates
2015	11,809,233	10,951,256
2016	10,626,899	10,626,899
2017	10,803,028	10,803,028

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

Year	States
2015	69
2016	68
2017	67

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

Year	Ticket without addr.
2015	1,807,864
2016	2,035,232
2017	2,289,944

Aggregation Tasks

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

Top Violation in 2015		
#	Violation Code	No. Of Violation
1	21	1,501,614
2	38	1,324,586
3	14	924,627
4	36	761,571
5	37	746,278

Top Violation in 2016		
#	Violation Code	No. Of Violation
1	21	1,531,587
2	36	1,253,512
3	38	1,143,696
4	14	875,614
5	37	686,610

Top Violation in 2017		
#	Violation Code	No. Of Violation
1	21	1,528,588
2	36	1,400,614
3	38	1,062,304
4	14	893,498
5	20	618,593

Aggregation Tasks

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

Top Violation By Body Type - 2015		
#	Body Tpe	No. Of Violation
1	SUBN	3,451,963
2	4DSD	3,102,510
3	VAN	1,605,228
4	DELV	840,441
5	SDN	453,992

Top Violation By Vehicle Make - 2015		
#	Body Tpe	No. Of Violation
1	FORD	1,417,303
2	TOYOT	1,123,523
3	HONDA	1,018,049
4	NISSA	837,569
5	CHEVR	836,389

Top Violation By Body Type - 2016		
#	Body Tpe	No. Of Violation
1	SUBN	3,466,037
2	4DSD	2,992,107
3	VAN	1,518,303
4	DELV	755,282
5	SDN	424,043

Top Violation By Vehicle Make - 2015		
#	Body Tpe	No. Of Violation
1	FORD	1,324,774
2	TOYOT	1,154,790
3	HONDA	1,014,074
4	NISSA	834,833
5	CHEVR	759,663

Top Violation By Body Type - 2017		
#	Body Tpe	No. Of Violation
1	SUBN	3,719,802
2	4DSD	3,082,020
3	VAN	1,411,970
4	DELV	687,330
5	SDN	438,191

Top Violation By Vehicle Make - 2015		
#	Body Tpe	No. Of Violation
1	FORD	1,280,958
2	TOYOT	1,211,451
3	HONDA	1,079,238
4	NISSA	918,590
5	CHEVR	714,655

Aggregation Tasks

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

Violating Precincts - 2015		
#	Violating Precincts	No. Of Violation
1	0	1,633,006
2	19	559,716
3	18	400,887
4	14	384,596
5	1	307,808

Violating Precincts - 2016		
#	Violating Precincts	No. Of Violation
1	0	1,868,655
2	19	554,465
3	18	331,704
4	14	324,467
5	1	303,850

Violating Precincts - 2017		
#	Violating Precincts	No. Of Violation
1	0	2,072,400
2	19	535,671
3	14	352,450
4	1	331,810
5	18	306,920

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)**

Issuing Precincts - 2015		
#	Issuing Precincts	No. Of Violation
1	0	1,834,343
2	19	544,946
3	18	391,501
4	14	369,725
5	1	298,594

Issuing Precincts - 2016		
#	Issuing Precincts	No. Of Violation
1	0	2,140,274
2	19	540,569
3	18	323,132
4	14	315,311
5	1	295,013

Issuing Precincts - 2017		
#	Issuing Precincts	No. Of Violation
1	0	2,388,479
2	19	521,513
3	14	344,977
4	1	321,170
5	18	296,553

Aggregation Tasks

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?**
   - Assumption: Precincts Code 0 is considered as one of the Precincts
   - Precincts 0, 19, 18 and 14 are the top Precincts across all 3 years
   - No consistency in violation codes however code 14, 46 are most common in all year/precincts

Top 3 Precincts which issued most tickets - 2015		
#	Precinct Code	No. Of Tickets
1	0	1,633,006
2	18	400,887
3	19	559,716

Top 3 Precincts which issued most tickets - 2016		
#	Precinct Code	No. Of Tickets
1	0	1,868,655
2	18	331,704
3	19	554,465

Top 3 Precincts which issued most tickets - 2017		
#	Precinct Code	No. Of Tickets
1	0	2,072,400
2	19	535,671
3	14	352,450

Top Violation Code for Precinct 0 - 2015		
#	Violation Code	No. Of Violation
1	36	761,571
2	7	662,202
3	5	195,353
4	46	2,178
5	14	1,777

Top Violation Code for Precinct 18 - 2015		
#	Violation Code	No. Of Violation
1	14	123,639
2	69	57,255
3	31	30,632
4	47	29,909
5	42	19,867

Top Violation Code for Precinct 19 - 2015		
#	Violation Code	No. Of Violation
1	38	92,499
2	37	80,426
3	14	61,556
4	21	57,592
5	16	56,509

Top Violation Code for Precinct 0 - 2016		
#	Violation Code	No. Of Violation
1	36	1,253,511
2	7	492,469
3	5	112,376
4	46	1,506
5	21	1,292

Top Violation Code for Precinct 18 - 2016		
#	Violation Code	No. Of Violation
1	14	102,281
2	69	47,948
3	47	24,600
4	31	22,922
5	42	17,679

Top Violation Code for Precinct 19 - 2016		
#	Violation Code	No. Of Violation
1	38	79,275
2	46	77,751
3	37	76,466
4	14	62,758
5	21	60,096

Top Violation Code for Precinct 0 - 2017		
#	Violation Code	No. Of Violation
1	36	1,400,614
2	7	516,388
3	5	145,642
4	46	1,621
5	14	1,254

Top Violation Code for Precinct 19 - 2017		
#	Violation Code	No. Of Violation
1	46	90,530
2	38	74,926
3	37	73,359
4	14	58,640
5	21	56,516

Top Violation Code for Precinct 14 - 2017		
#	Violation Code	No. Of Violation
1	14	75,850
2	69	58,032
3	31	40,150
4	47	31,201
5	42	20,666

Aggregation Tasks

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

Percentages of Null Values			
Column Name	2015	2016	2017
Summons Number	0	0	0
Plate ID	1.82628E-05	9.41008E-06	9.25666E-06
Registration State	0	0	0
Plate Type	0	0	0
Issue Date	0	0	0
Violation Code	0	0	0
Vehicle Body Type	0.3761852	0.3695434	0.3952133
Vehicle Make	0.6234536	0.5982742	0.6761715
Issuing Agency	0	0	0
Street Code1	0	0	0
Street Code2	0	0	0
Street Code3	0	0	0
Vehicle Expiration Date	0	9.41008E-06	0
Violation Location	14.91159	17.58421	19.18351
Violation Precinct	0	9.41008E-06	0
Issuer Precinct	0	9.41008E-06	0
Issuer Code	0	9.41008E-06	0
Issuer Command	14.78482	17.48729	19.09321
Issuer Squad	14.78615	17.492	19.10151
Violation Time	0.0142632	0.04027515	0.00058317
Time First Observed	89.48479	89.38105	92.21749
Violation County	15.23991	16.13865	0.3660733
Violation In Front Of Or Opposite	15.57431	18.28207	20.00583
House Number	16.49486	19.13465	21.18497
Street Name	0.04988469	0.07785903	0.03710997
Intersecting Street	72.74455	70.65415	68.82767

Percentages of Null Values			
Column Name	2015	2016	2017
Date First Observed	9.13137E-06	9.41008E-06	0
Law Section	9.13137E-06	9.41008E-06	0
Sub Division	0.003424265	0.03597475	0.007155401
Violation Legal Code	85.20115	82.5058	80.90621
Days Parking In Effect	23.85113	26.98262	25.10792
From Hours In Effect	43.98682	46.82596	50.45758
To Hours In Effect	43.98682	46.82596	50.45755
Vehicle Color	1.091966	1.222473	1.410086
Unregistered Vehicle?	88.82023	89.3083	89.56222
Vehicle Year	2.73941E-05	9.41008E-06	0
Meter Number	81.25123	82.12376	83.47248
Feet From Curb	2.73941E-05	9.41008E-06	0
Violation Post Code	25.95143	28.20427	29.53049
Violation Description	11.18008	10.72126	10.43661
No Standing or Stopping Violation	99.99999	100	100
Hydrant Violation	99.99999	100	100
Double Parking Violation	99.99999	100	100
Latitude	100	100	
Longitude	100	100	
Community Board	100	100	
Community Council	100	100	
Census Tract	100	100	
BIN	100	100	
BBL	100	100	
NTA	100	100	

Aggregation Tasks

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

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

Violation By Hours Bin - 2015		
Bin	Violation Code	No. Of Violation
0-4	21	18,881
	14	3,918
	20	697
4-8	69	187,924
	14	184,213
	20	125,938
8-12	14	12,861
	18	4,459
	38	2,474
12-16	21	1,294,084
	38	1,245,388
	37	706,248
16-20	14	63,510
	38	17,080
	20	6,888
20-24	78	43,604
	14	18,717
	21	2,283
<NA>	36	761,571
	7	662,208
	14	609,248

Violation By Hours Bin - 2016		
Bin	Violation Code	No. Of Violation
0-4	21	19,662
	14	5,707
	38	469
4-8	14	176,301
	69	163,335
	20	114,656
8-12	21	1,314,774
	38	1,076,854
	37	651,231
12-16	14	12,876
	18	2,840
	38	2,725
16-20	14	58,716
	38	13,924
	20	5,207
20-24	78	38,755
	14	14,709
	21	1,903
<NA>	36	1,253,512
	46	580,417
	14	577,649

Violation By Hours Bin - 2017		
Bin	Violation Code	No. Of Violation
0-4	21	22,860
	14	3,843
	38	546
4-8	14	180,105
	69	130,723
	21	114,804
8-12	21	1,298,898
	38	999,735
	37	567,134
12-16	14	14,722
	38	3,811
	18	2,732
16-20	14	60,804
	38	12,386
	20	4,147
20-24	78	32,668
	14	13,505
	21	2,441
<NA>	36	1,400,614
	46	599,904
	14	592,911

*** NA's in the result sets are kept separately as it's not logical to add them in any of the time bin, unless otherwise specified**

Aggregation Tasks

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

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)

Top Violations By Time Bin - 2015		
Violation Code	Time Bin	No. Of Violation
14	<NA>	609,248
	4-8	184,213
	16-20	63,510
	8-12	32,160
	20-24	18,717
	12-16	12,861
	0-4	3,918
21	8-12	1,294,084
	4-8	103,654
	<NA>	82,552
	0-4	18,881
	20-24	2,283
	16-20	148
	12-16	12
38	8-12	1,245,388
	4-8	43,950
	16-20	17,080
	<NA>	13,235
	12-16	2,474
	20-24	2,030
	0-4	429

Top Violations By Time Bin - 2016		
Violation Code	Time Bin	No. Of Violation
21	8-12	1,314,774
	4-8	110,287
	<NA>	84,819
	0-4	19,662
	20-24	1,903
	16-20	137
	12-16	5
36	<NA>	1,253,512
38	8-12	1,076,854
	4-8	35,880
	16-20	13,924
	<NA>	12,013
	12-16	2,725
	20-24	1,831
	0-4	469

Top Violations By Time Bin - 2017		
Violation Code	Time Bin	No. Of Violation
21	8-12	1,298,898
21	4-8	114,804
21	<NA>	89,317
21	0-4	22,860
21	20-24	2,441
21	16-20	232
21	12-16	36
36	<NA>	1,400,614
38	8-12	999,735
38	4-8	31,931
38	16-20	12,386
38	<NA>	12,341
38	12-16	3,811
38	20-24	1,554
38	0-4	546

*** NA's in the result sets are kept separately as it's not logical to add them in any of the time bin, unless otherwise specified**

Aggregation Tasks

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.

Violation By Season - 2015		
Bin	Season	No. Of Violation
1	Winter	2,182,331
2	Fall	2,718,868
3	Spring	2,951,328
4	Summer	3,098,729

Violation By Season - 2016		
Bin	Season	No. Of Violation
1	Winter	2,424,488
2	Summer	2,438,069
3	Spring	2,790,946
4	Fall	2,973,396

Violation By Season - 2017		
Bin	Season	No. Of Violation
1	Winter	2,485,331
2	Summer	2,606,208
3	Fall	2,830,802
4	Spring	2,880,687

Aggregation Tasks

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

b) Then, find the 3 most common violations for each of these season

Common Violations By Season - 2015		
Season	Violation Code	No. Of Violation
Fall	21	351,423
	38	326,702
	14	232,339
Spring	21	425,350
	38	327,057
	14	243,769
Summer	21	471,627
	38	363,815
	14	255,182
Winter	38	307,012
	21	253,214
	14	193,337

Common Violations By Season - 2016		
Season	Violation Code	No. Of Violation
Fall	36	438,320
	21	395,357
	38	303,397
Spring	21	383,757
	36	374,362
	38	299,459
Summer	21	392,205
	38	272,419
	14	215,683
Winter	21	360,268
	36	314,765
	38	268,421

Common Violations By Season - 2017		
Season	Violation Code	No. Of Violation
Fall	36	456,046
	21	357,479
	38	283,828
Spring	21	402,807
	36	344,834
	38	271,192
Summer	21	405,961
	38	247,561
	36	240,396
Winter	21	362,341
	36	359,338
	38	259,723

Aggregation Tasks

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.**
   a) **Find total occurrences of the 3 most common violation code**

Top Violation in 2015		
#	Violation Code	No. Of Violation
1	21	1501614
2	38	1324586
3	14	924627

Top Violation in 2015		
#	Violation Code	No. Of Violation
1	21	1531587
2	36	1253512
3	38	1143696

Top Violation in 2015		
#	Violation Code	No. Of Violation
1	21	1528588
2	36	1400614
3	38	1062304

Aggregation Tasks

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.**

   b) **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.**

Top violation codes in terms of fines collected and ticket issued	
Violation Code	**Violation Name**
14	General No Standing: Standing or parking where standing is not allowed by sign, street marking or; traffic control device.
21	Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.
36	Exceeding the posted speed limit in or near a designated school zone.
38	Meter Issue - Failing to show a receipt or tag in the windshield.

Aggregation Tasks

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.

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

Total Fine Collected	
Year	Collection
2015	758,131,492
2016	715,966,840
2017	732,576,992

Violation Code with Most Collection		
Year	Violation Code	Collection
2015	14	106,332,105
2016	14	100,695,610
2017	14	102,752,270

d) What can you intuitively infer from these findings?

Top Violation by count in 3 years is with code 21 which is "Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device. but max revenue is with code 14 which has higher fine for issue "Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device."

Thanks

Merin Jose

Suresh Balla

Deepak Aneja

Fayiz Mayam Veettil