NYC Parking Tickets: An Exploratory Analysis

(Big Data)

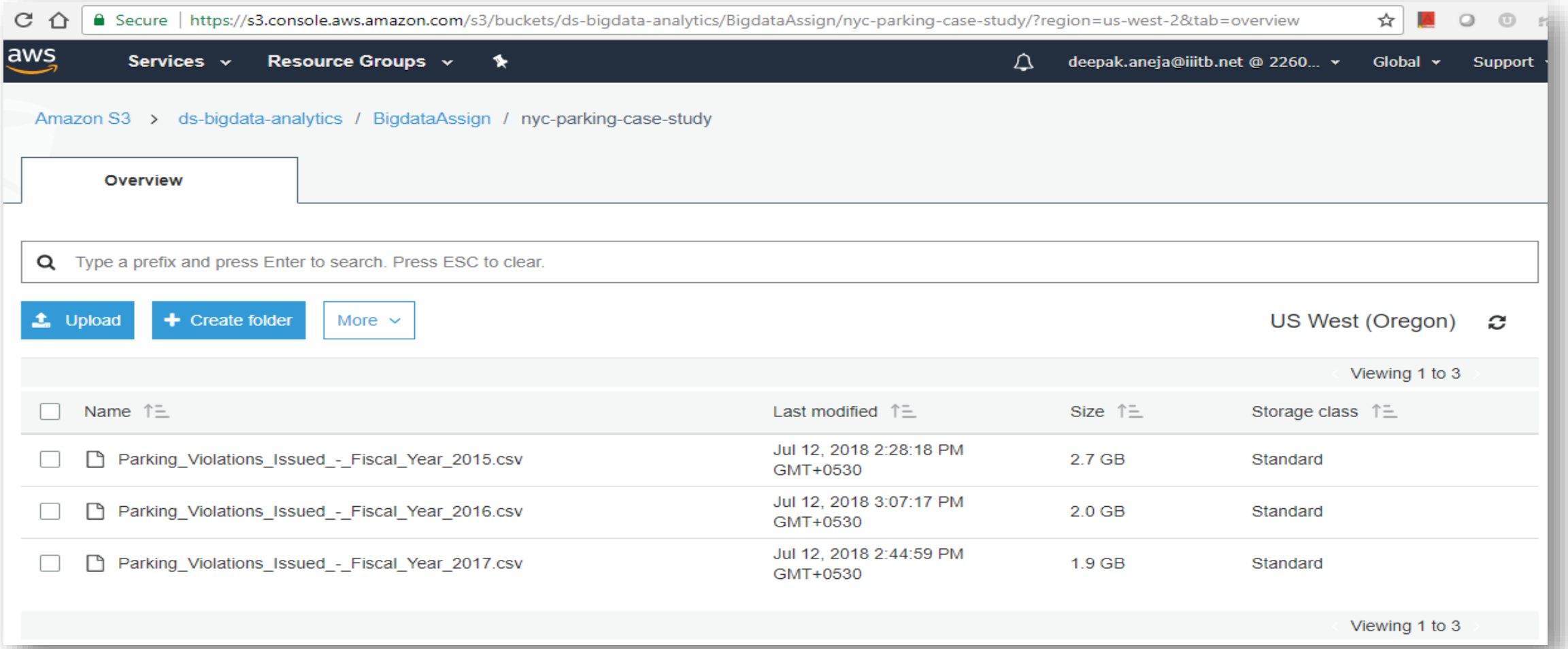
Group Name:

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

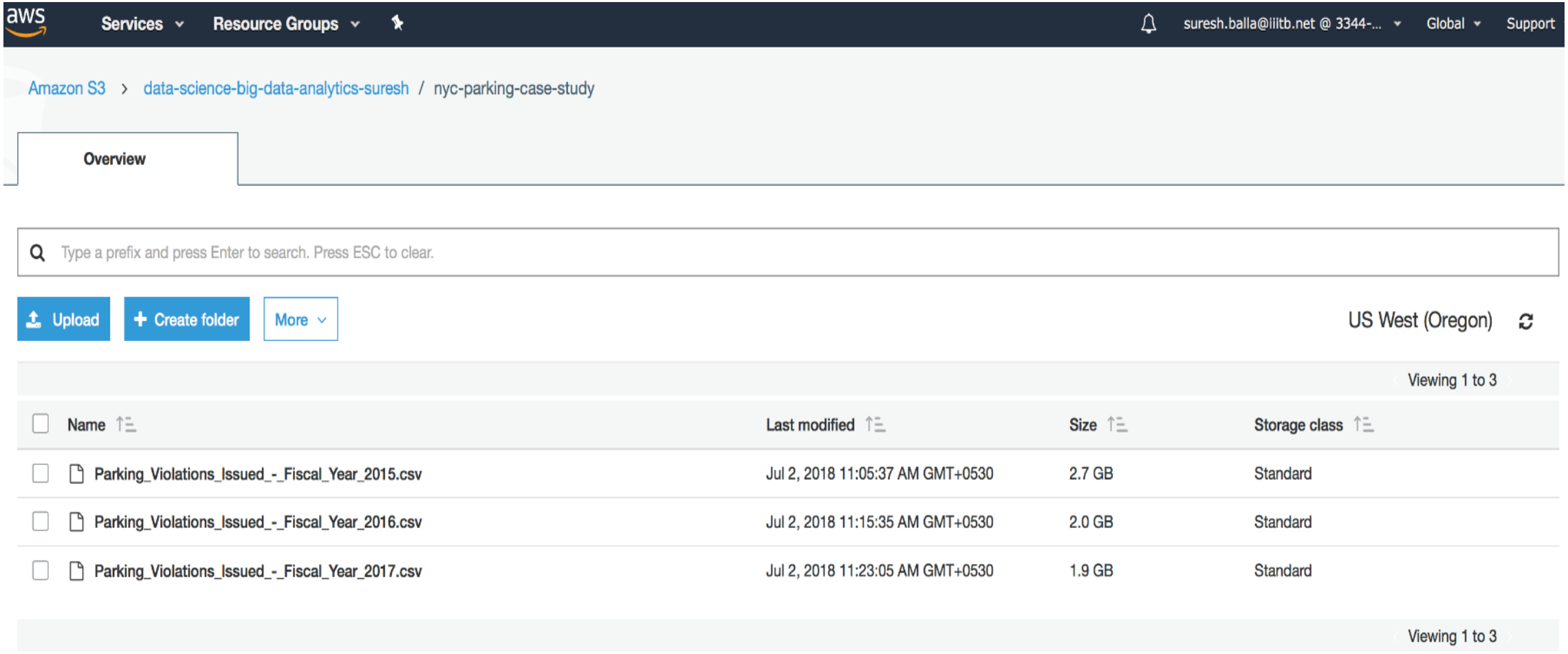
15 Jul 2018

Table of Contents

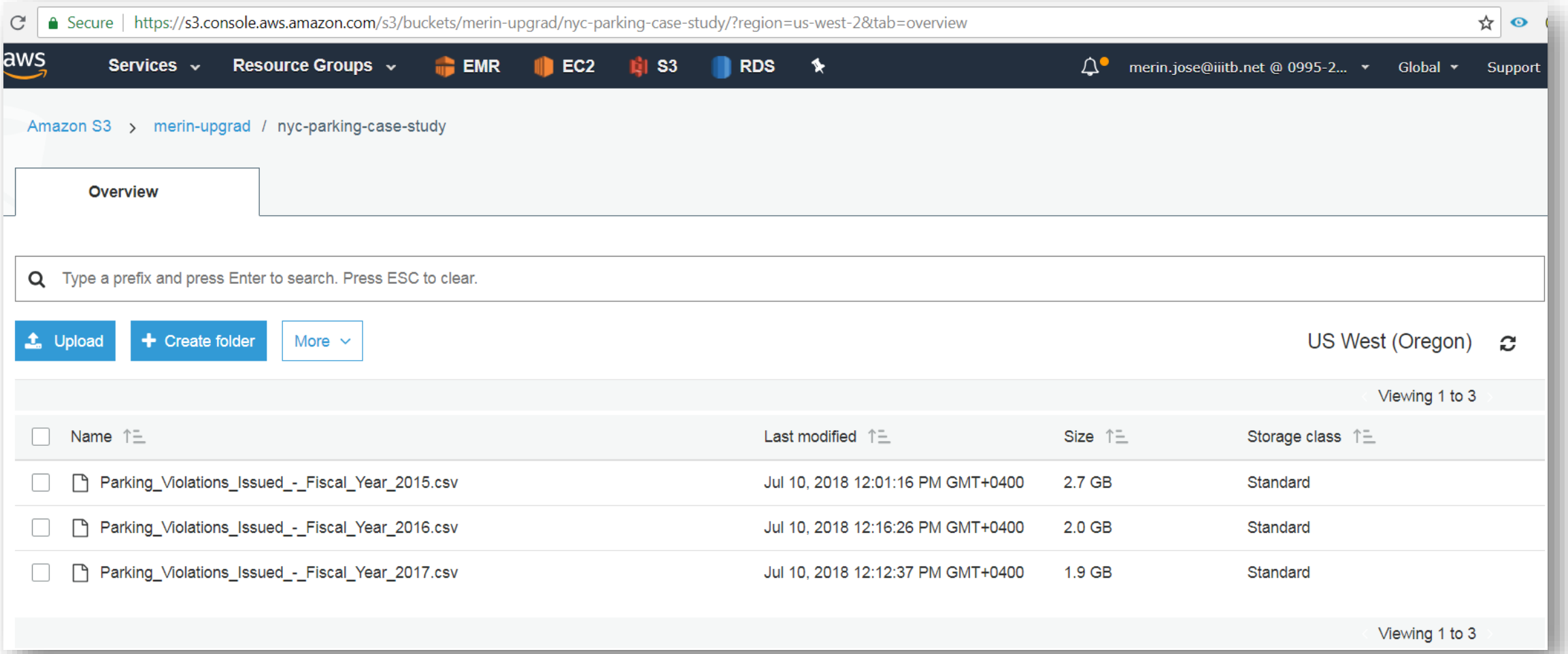
S3 Bucket: Deepak Aneja (Bucket Path: s3://ds-bigdata-analytics/BigdataAssign/nyc-parking-case-study/)



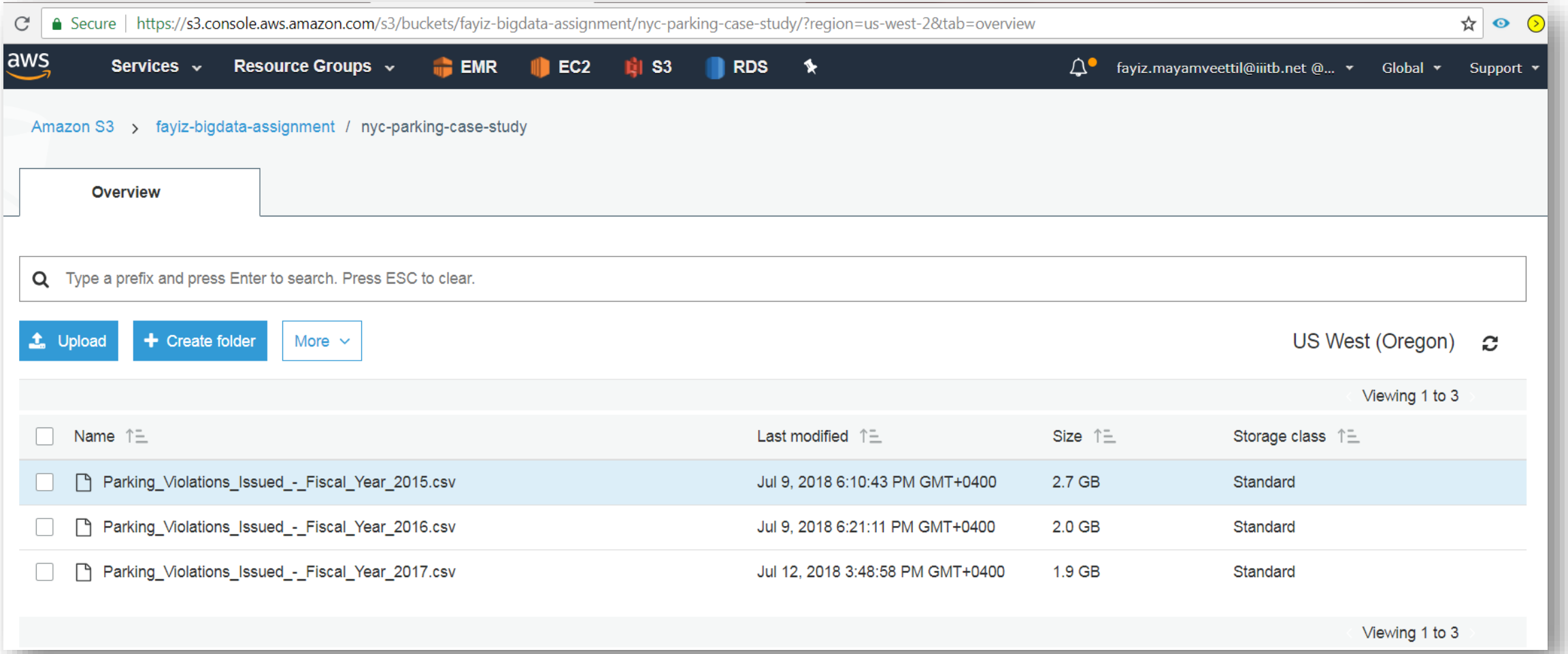
S3 Bucket: Suresh Balla (Bucket Path: s3://data-science-big-data-analytics-suresh/nyc-parking-case-study/)



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
  - Check for NA values in the columns used for analysis
- Status of data cleaning
  - --

Examine the data

1. Find total number of tickets for each year.

Year	No. Of Tickets
2015	11,809,233
2016	10,626,899
2017	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	1,992,401
2016	2,035,232
2017	2,035,232

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	1,630,912
2	38	1,418,627
3	14	988,469
4	36	839,197
5	37	795,918

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	3,729,346
2	4DSD	3,340,014
3	VAN	1,709,091
4	DELV	892,781
5	SDN	524,596

Top Violation By Vehicle Make - 2015		
#	Body Tpe	No. Of Violation
1	FORD	1,521,874
2	TOYOT	1,217,087
3	HONDA	1,102,614
4	NISSA	908,783
5	CHEVR	897,845

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	1,799,170
2	19	598,351
3	18	427,510
4	14	409,064
5	1	329,009

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	2,037,745
2	19	579,998
3	18	417,329
4	14	392,922
5	1	318,778

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?**

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.
   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
   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)

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

Aggregation Tasks

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.**
   a) **Find total occurrences of the 3 most common violation code**
   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.**
   c) **Using this information, find the total amount collected for all of the fines. State the code which has the highest total collection.**
   d) **What can you intuitively infer from these findings?**