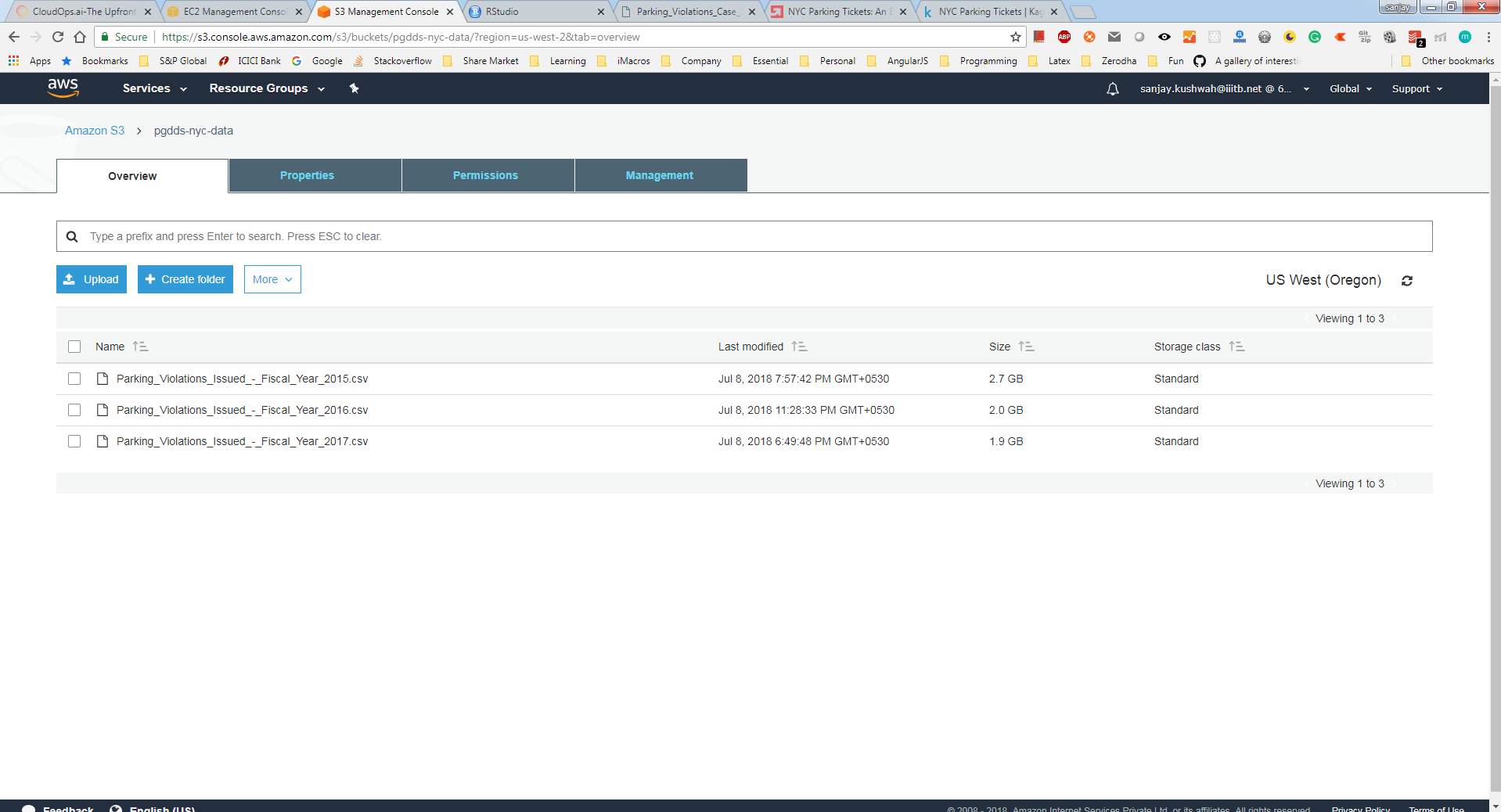
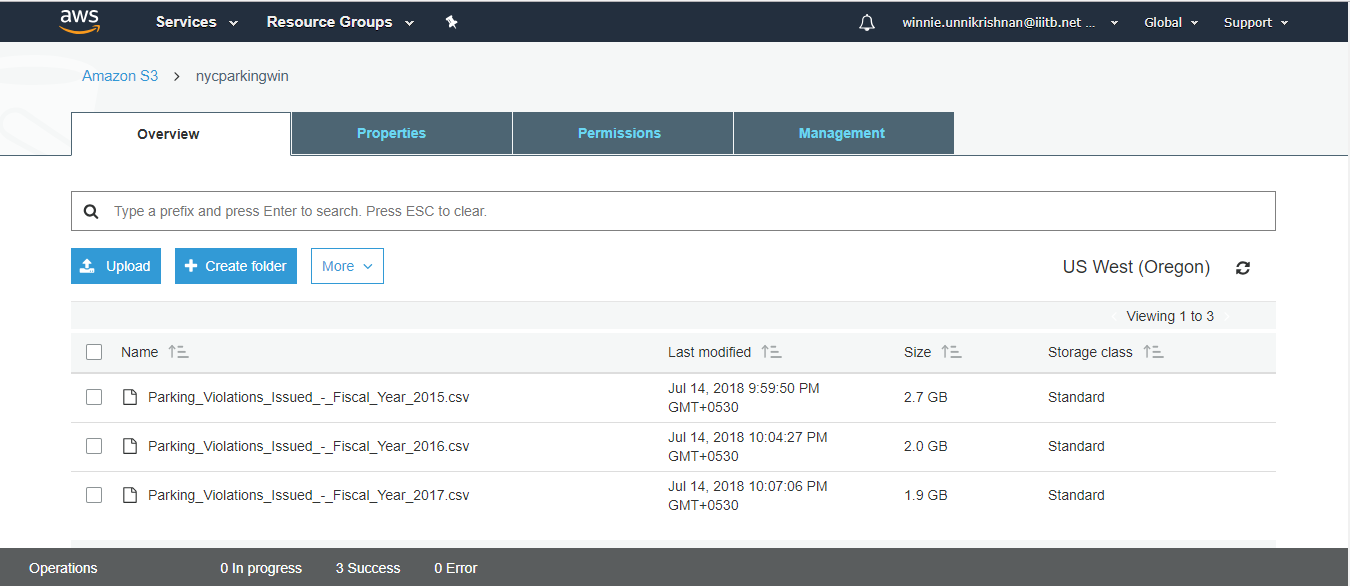
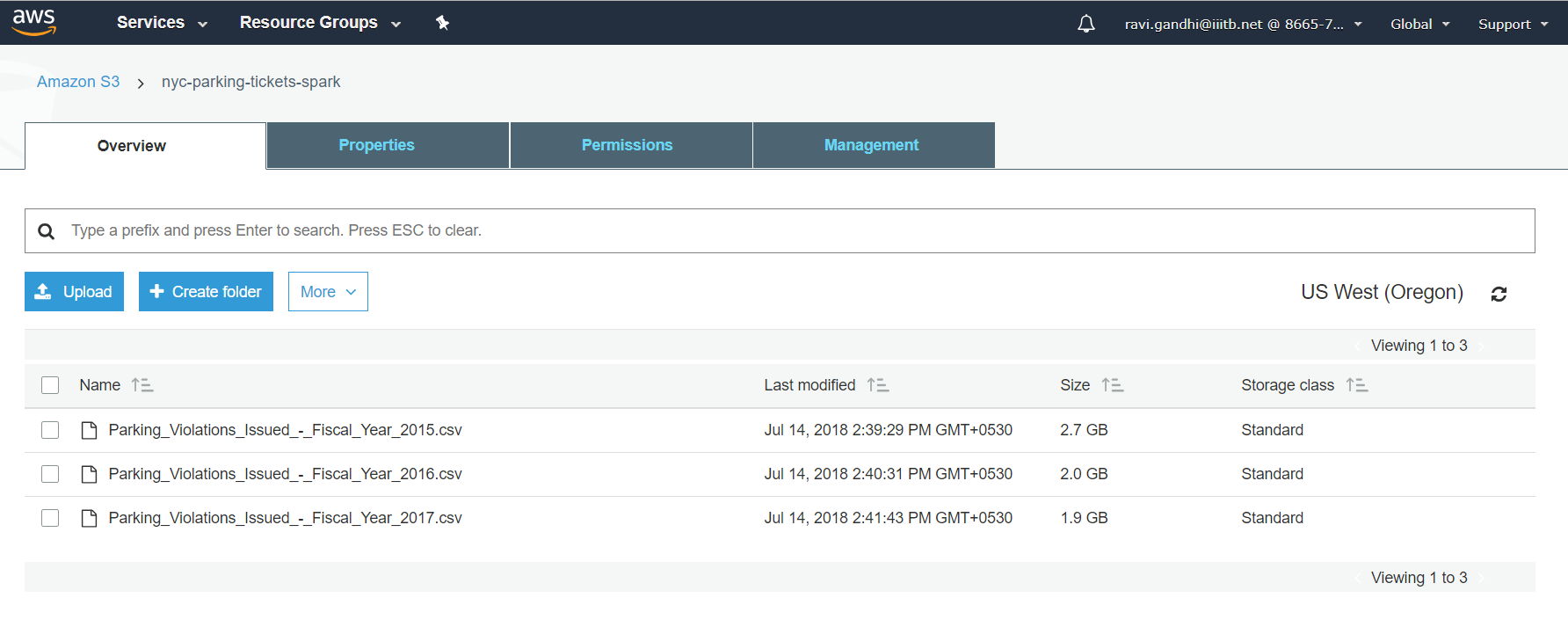
***Screen Shot of S3 Bucket:***

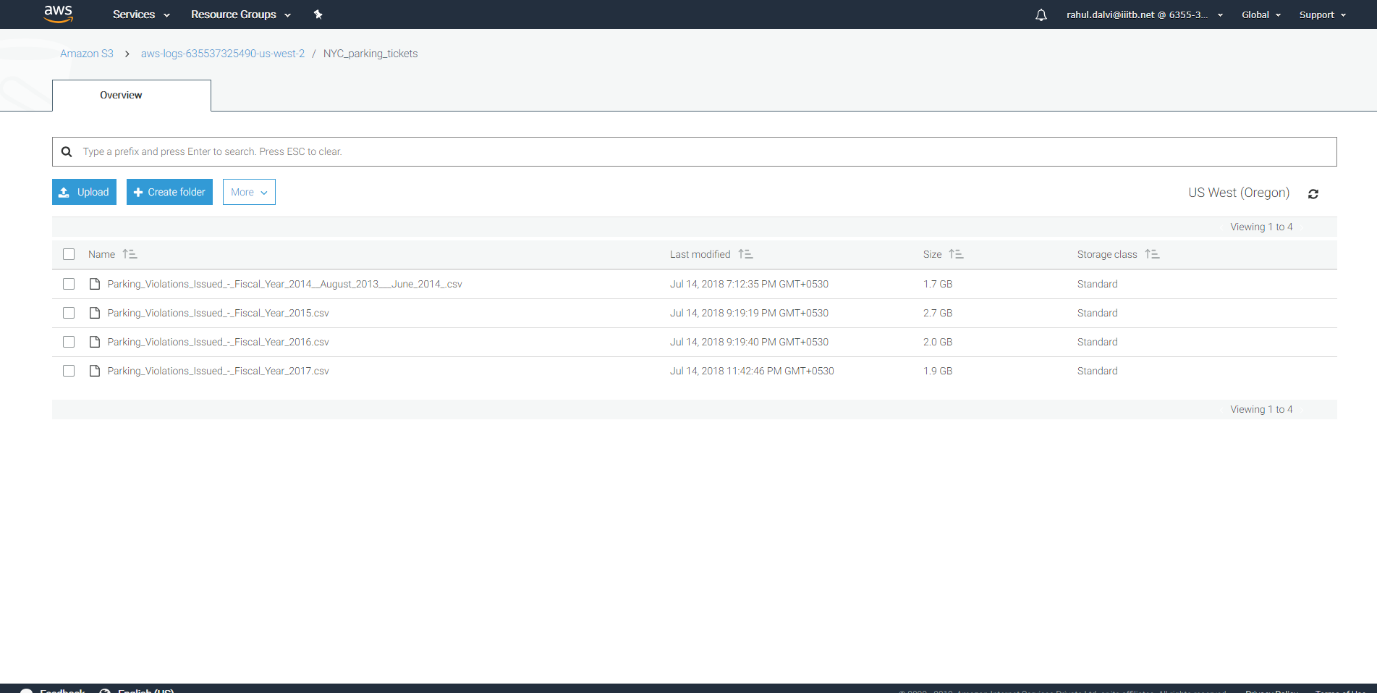
**Sanjay khuswah:** ****

**Winnie Unnikrishnan:**



**Ravi Gandhi:**

**Rahul Dalvi:**



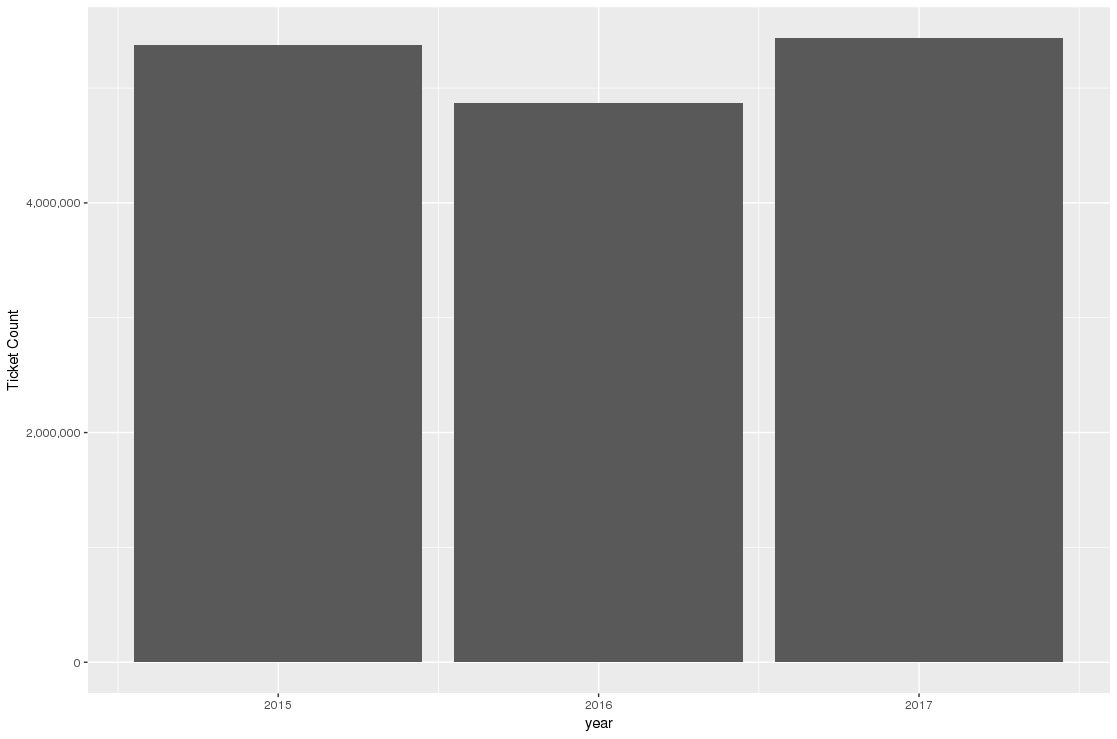
**Assumptions:**

**1) Each CSV file has data from other years as well, we have taken only the records with respective year provided as per file name.**

**2) Tickets with missing address are considered using the column violation location. i.e. It is considered where the violation location is missing**

**Examine the data :**

1. **Find total number of tickets for each year**



|  |  |
| --- | --- |
| **Year** | **Tickets** |
| 2015 | 5373971 |
| 2016 | 4872621 |
| 2017 | 5431918 |

**Insights:**

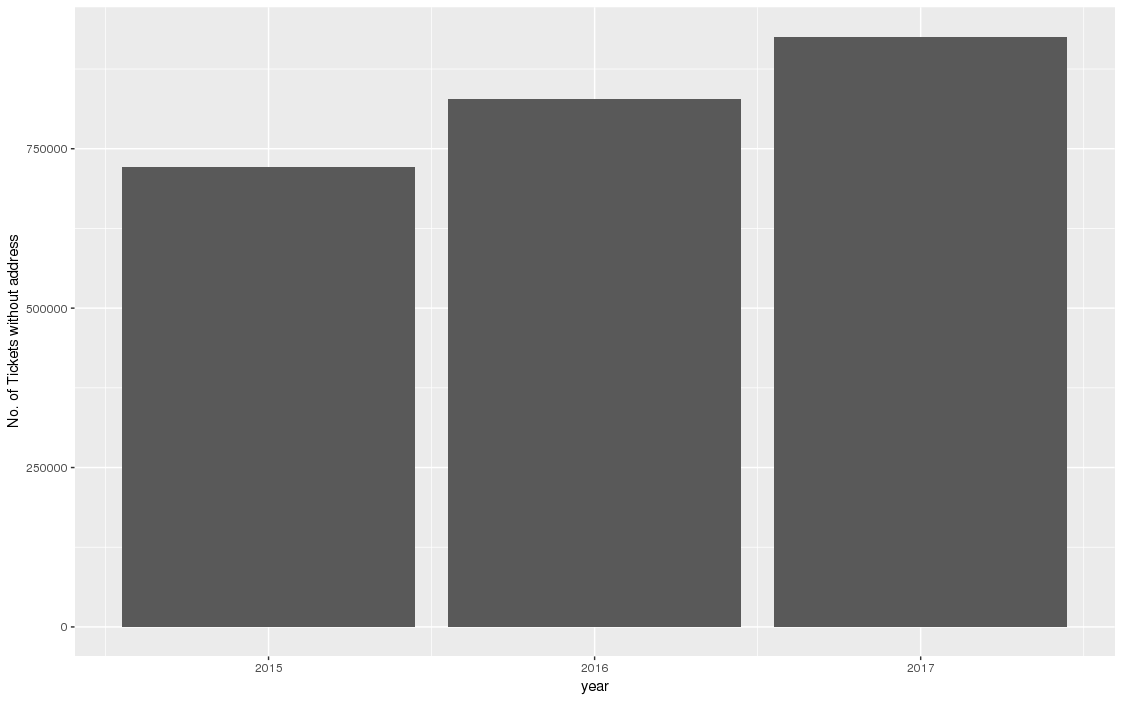
* 2017 has max number of tickets issued and 2016 has the least number of tickets issued.

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



|  |  |
| --- | --- |
| **Year** | **No of states** |
| 2015 | 68 |
| 2016 | 67 |
| 2017 | 65 |

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



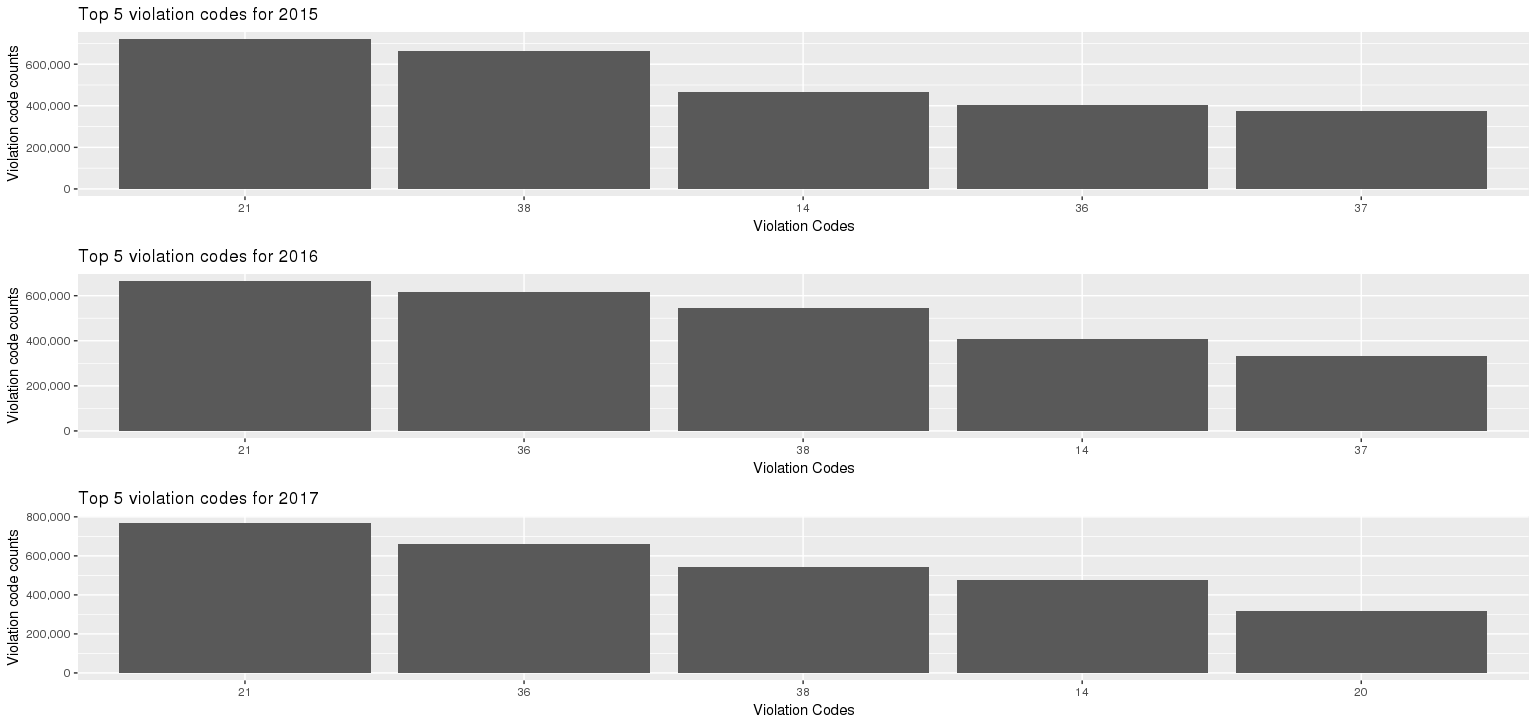
|  |  |
| --- | --- |
| **Year** | **No. of tickets without address** |
| 2015 | 721275 |
| 2016 | 828348 |
| 2017 | 925596 |

**Insights:**

* There seems to be an issue in the way the tickets are issued, as the tickets without address are increasing day on day

**Aggregation tasks**

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



|  |  |
| --- | --- |
| **Year 2015** | |
| **Violation Code** | **Ticket Numbers** |
| 21 | 720902 |
| 38 | 663904 |
| 14 | 466488 |
| 36 | 406249 |
| 37 | 373229 |
| **Year 2016** | |
| **Violation Code** | **Ticket Numbers** |
| 21 | 664947 |
| 36 | 615242 |
| 38 | 547080 |
| 14 | 405885 |
| 37 | 330489 |
| **Year 2017** | |
| **Violation Code** | **Ticket Numbers** |
| 21 | 768087 |
| 36 | 662765 |
| 38 | 542079 |
| 14 | 476664 |
| 20 | 319646 |

**Insights:**

* The violation codes 21, 38, 14, 36 are the most common violation codes that occurs across all three years

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

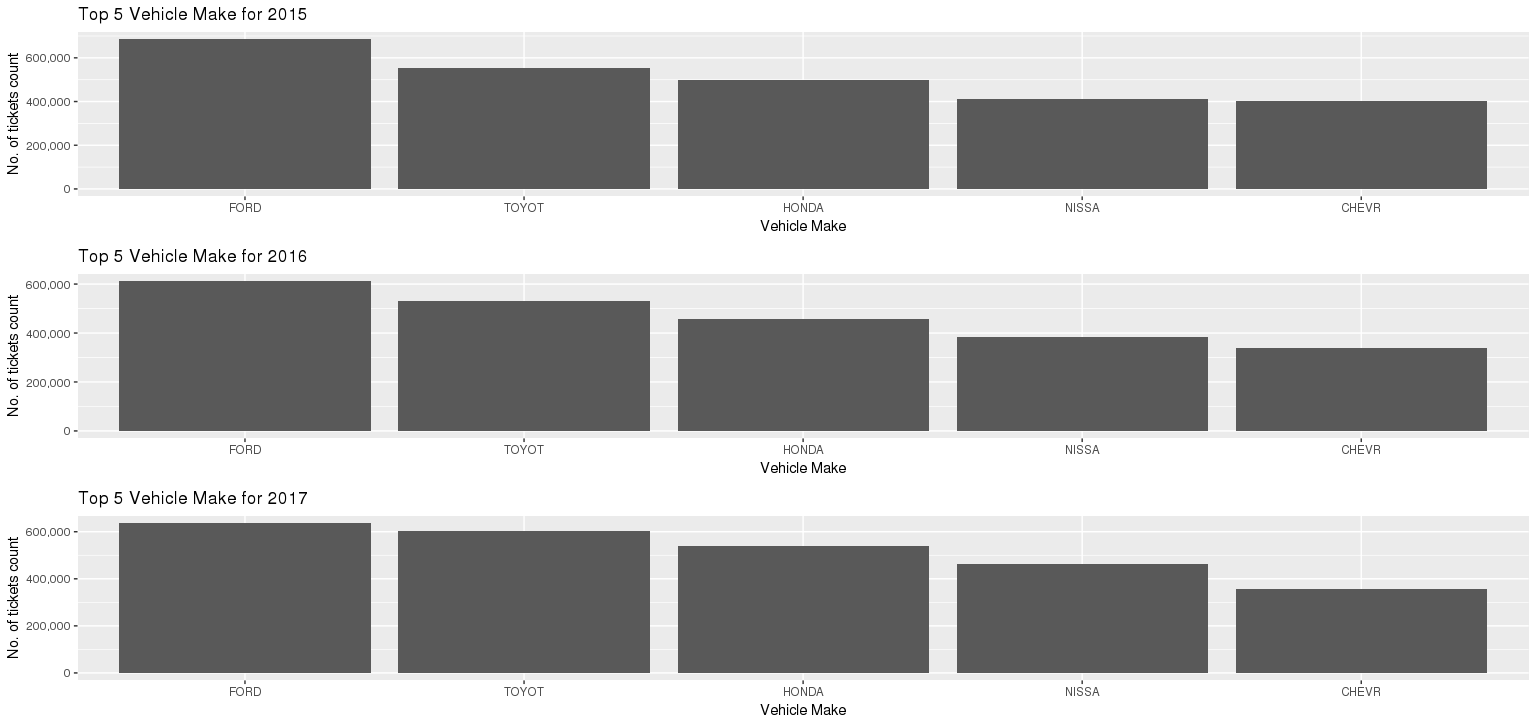


|  |  |
| --- | --- |
| **Year 2015** | |
| **Vehicle Body Type** | **Ticket Numbers** |
| SUBN | 1715517 |
| 4DSD | 1514580 |
| VAN | 795457 |
| DELV | 419548 |
| SDN | 209381 |
| **Year 2016** | |
| **Vehicle Body Type** | **Ticket Numbers** |
| SUBN | 1596326 |
| 4DSD | 1354001 |
| VAN | 722234 |
| DELV | 354388 |
| SDN | 178954 |
| **Year 2017** | |
| **Vehicle Body Type** | **Ticket Numbers** |
| SUBN | 1883954 |
| 4DSD | 1547312 |
| VAN | 724029 |
| DELV | 358984 |
| SDN | 194197 |

**Insights:**

* SUBN, 4DSD, VAN, DELV, SDN are the most common Vehicle body type to get max tickets consistently across all 3 years. The body type needs to be considered about to avoid parking tickets

**Vehicle Make:**

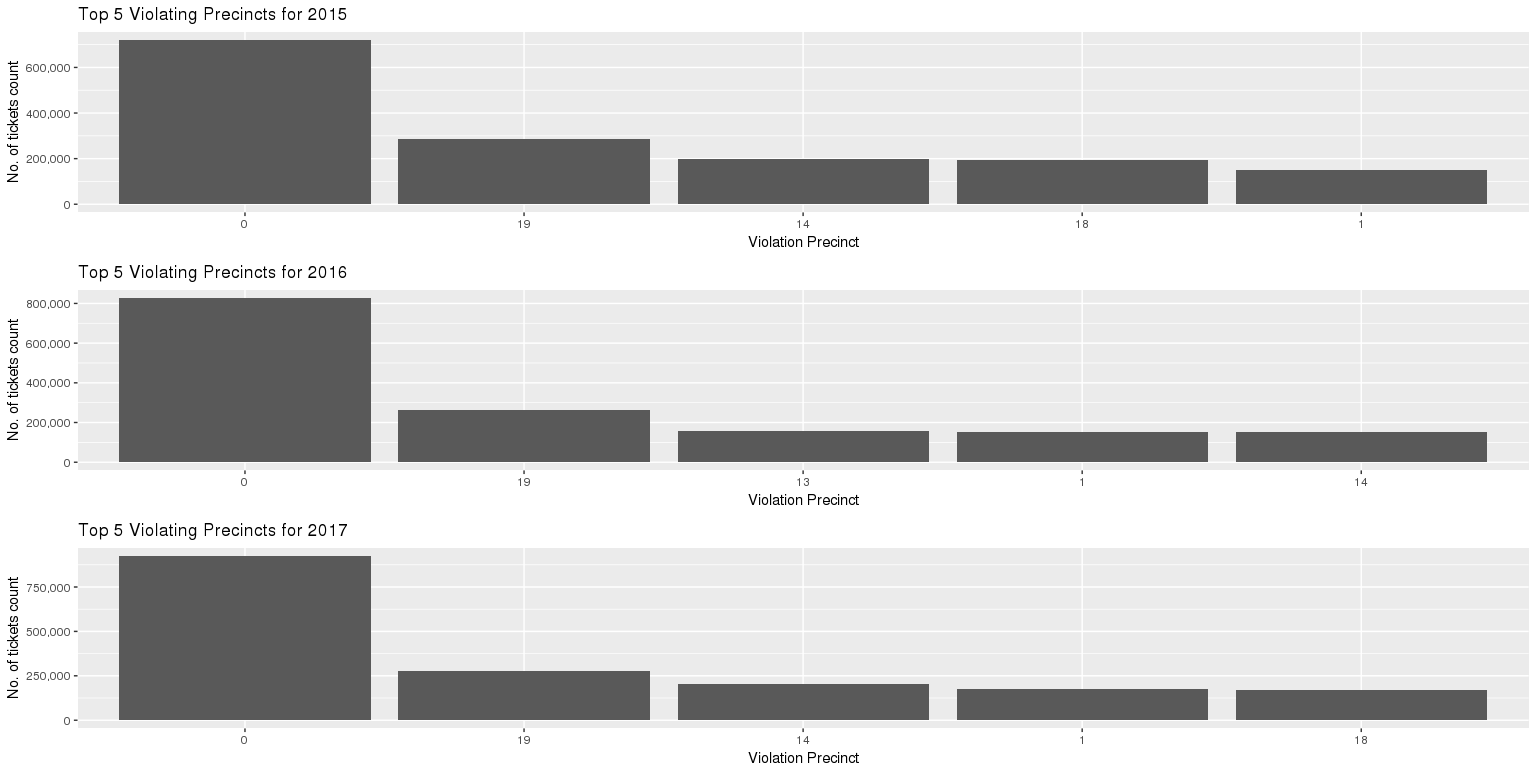


|  |  |
| --- | --- |
| **Year 2015** | |
| **Vehicle Make** | **Ticket Numbers** |
| FORD | 685900 |
| TOYOT | 554392 |
| HONDA | 498858 |
| NISSA | 411857 |
| CHEVR | 404841 |
| **Year 2016** | |
| **Vehicle Make** | **Ticket Numbers** |
| FORD | 612276 |
| TOYOT | 529115 |
| HONDA | 459469 |
| NISSA | 382082 |
| CHEVR | 339466 |
| **Year 2017** | |
| **Vehicle Make** | **Ticket Numbers** |
| FORD | 636844 |
| TOYOT | 605291 |
| HONDA | 538884 |
| NISSA | 462017 |
| CHEVR | 356032 |

**Insights:**

* FORD, TOYOT, HONDA, NISSA, CHEVR are the most common Vehicle Make to get max tickets consistently across all 3 years. The make needs to be considered about to avoid parking tickets

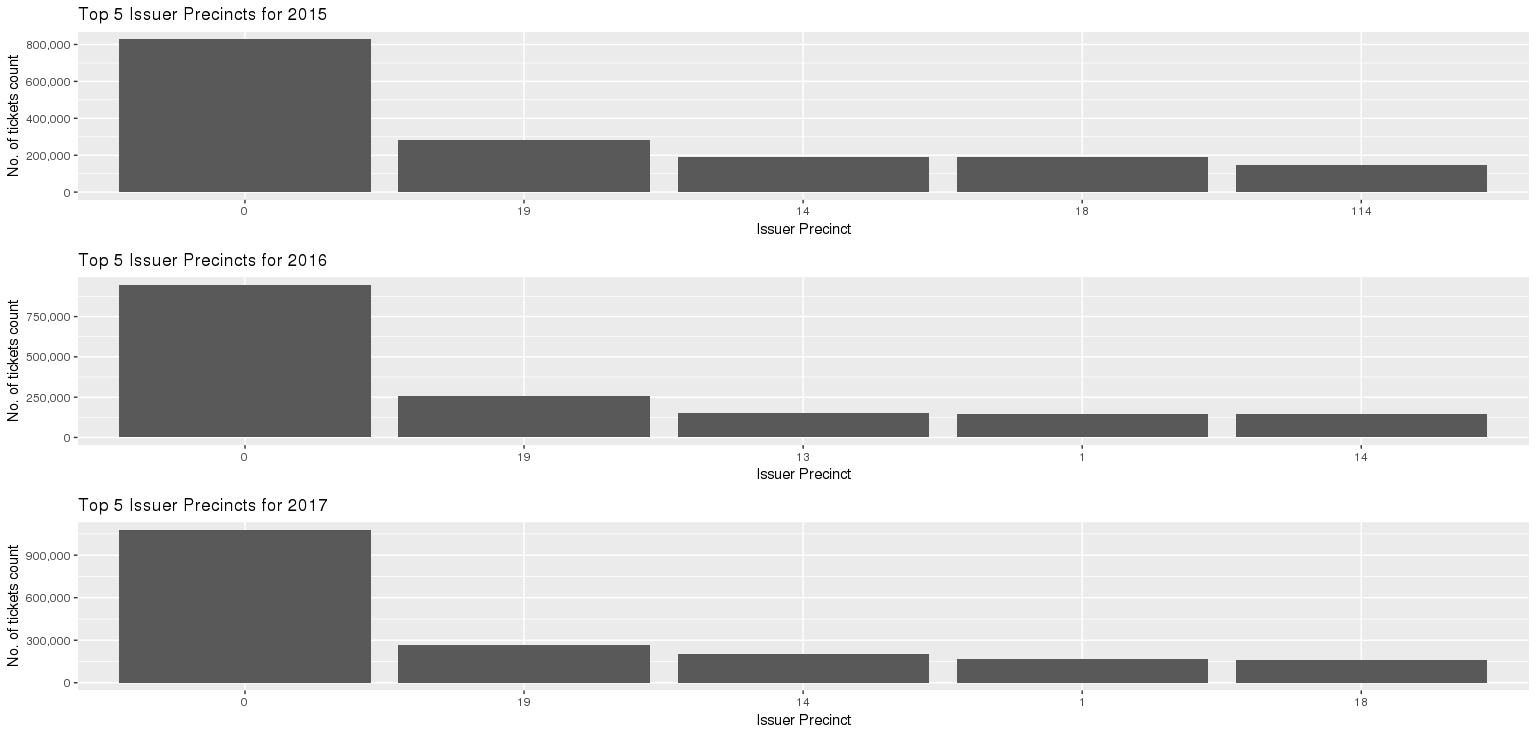
1. **A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequencies of:**
   1. **Violating Precincts (this is the precinct of the zone where the violation occurred)**



|  |  |
| --- | --- |
| **Year 2015** | |
| **Violation Precinct** | **Ticket Numbers** |
| 0 | 721275 |
| 19 | 287403 |
| 14 | 197011 |
| 18 | 193593 |
| 1 | 152040 |
| **Year 2016** | |
| **Violation Precinct** | **Ticket Numbers** |
| 0 | 828348 |
| 19 | 264299 |
| 13 | 156144 |
| 1 | 152231 |
| 14 | 150637 |
| **Year 2017** | |
| **Violation Precinct** | **Ticket Numbers** |
| 0 | 925596 |
| 19 | 274445 |
| 14 | 203553 |
| 1 | 174702 |
| 18 | 169131 |

**Insights:**

* The most common Violation Precinct are 0,19,14 across all three years
  1. **Issuing Precincts (this is the precinct that issued the ticket)**



|  |  |
| --- | --- |
| **Year 2015** | |
| **Issuer Precinct** | **Ticket Numbers** |
| 0 | 828570 |
| 19 | 279931 |
| 14 | 190403 |
| 18 | 190337 |
| 114 | 149532 |
| **Year 2016** | |
| **Issuer Precinct** | **Ticket Numbers** |
| 0 | 948438 |
| 19 | 258049 |
| 13 | 153478 |
| 1 | 146987 |
| 14 | 146165 |
| **Year 2017** | |
| **Issuer Precinct** | **Ticket Numbers** |
| 0 | 1078406 |
| 19 | 266961 |
| 14 | 200495 |
| 1 | 168740 |
| 18 | 162994 |

**Insights:**

* The most common Issuer Precinct are 0,19,14 across all three years

1. **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?**

|  |  |
| --- | --- |
| **Year 2015 - Issuer Precinct - 0** | |
| **Violation Code** | **Ticket Numbers** |
| 36 | 406249 |
| 7 | 253730 |
| 21 | 96218 |
| 5 | 55192 |
| 66 | 2343 |
| **Year 2015 - Issuer Precinct - 19** | |
| **Violation Code** | **Ticket Numbers** |
| 38 | 45647 |
| 37 | 40665 |
| 14 | 31295 |
| 16 | 29738 |
| 46 | 27049 |
| **Year 2015 - Issuer Precinct - 14** | |
| **Violation Code** | **Ticket Numbers** |
| 69 | 41004 |
| 14 | 38696 |
| 31 | 20676 |
| 47 | 14480 |
| 42 | 14446 |
| **Year 2016 - Issuer Precinct - 0** | |
| **Violation Code** | **Ticket Numbers** |
| 36 | 615242 |
| 7 | 165111 |
| 21 | 104351 |
| 5 | 43467 |
| 66 | 3821 |
| **Year 2016 - Issuer Precinct - 19** | |
| **Violation Code** | **Ticket Numbers** |
| 37 | 38052 |
| 38 | 37855 |
| 46 | 36442 |
| 14 | 28772 |
| 21 | 25588 |
| **Year 2016 - Issuer Precinct - 13** | |
| **Violation Code** | **Ticket Numbers** |
| 69 | 23356 |
| 47 | 17532 |
| 38 | 16447 |
| 14 | 15812 |
| 37 | 13589 |
| **Year 2017 - Issuer Precinct - 0** | |
| **Violation Code** | **Ticket Numbers** |
| 36 | 662765 |
| 7 | 210175 |
| 21 | 126053 |
| 5 | 48076 |
| 66 | 5258 |
| **Year 2017 - Issuer Precinct - 19** | |
| **Violation Code** | **Ticket Numbers** |
| 46 | 48445 |
| 38 | 36386 |
| 37 | 36056 |
| 14 | 29797 |
| 21 | 28415 |
| **Year 2017 - Issuer Precinct - 14** | |
| **Violation Code** | **Ticket Numbers** |
| 14 | 45036 |
| 69 | 30464 |
| 31 | 22555 |
| 47 | 18364 |
| 42 | 10027 |

**Insights:**

* There doesn’t seem to be any common Violation codes for top 3 Issuer Precinct

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

24 hours has been split into 6 time-bins.

On analysis of count of violations vs violation codes against the bins, below pattern is found.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2015 | | 2016 | | 2017 | |
| Time bin | Violation Code | count | Violation Code | count | Violation Code | count |
| 1 | 21 | 30663 | 21 | 31956 | 21 | 36957 |
| 40 | 20613 | 40 | 19078 | 40 | 25866 |
| 78 | 17198 | 78 | 14706 | 78 | 15528 |
| 2 | 14 | 68994 | 14 | 65347 | 14 | 74114 |
| 21 | 49098 | 21 | 48240 | 40 | 60652 |
| 40 | 46783 | 40 | 42306 | 21 | 57897 |
| 3 | 21 | 573741 | 21 | 525280 | 21 | 598069 |
| 38 | 235956 | 36 | 284279 | 36 | 348165 |
| 36 | 189347 | 38 | 185395 | 38 | 176570 |
| 4 | 38 | 287564 | 36 | 273581 | 36 | 286284 |
| 37 | 212536 | 38 | 234221 | 38 | 240721 |
| 36 | 177439 | 37 | 183854 | 37 | 167026 |
| 5 | 38 | 111178 | 38 | 105657 | 38 | 102855 |
| 37 | 83676 | 37 | 79991 | 14 | 75902 |
| 14 | 73424 | 14 | 63778 | 37 | 70345 |
| 6 | 7 | 29936 | 38 | 20851 | 7 | 26293 |
| 38 | 27571 | 7 | 20246 | 40 | 22337 |
| 40 | 22491 | 40 | 20030 | 14 | 21045 |

Highest violations happen in time slots 3 and 4. i.e. 9AM -4 PM

On Analysis of most occurred violation code and the hours these violations occur, below pattern is found

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2015 | | | 2016 | | | 2017 | | |
| Violation Code | time\_bin | count | Violation Code | time\_bin | count | Violation Code | time\_bin | count |
| 21 | 3 | 573741 | 21 | 3 | 525280 | 21 | 3 | 598069 |
| 4 | 66820 | 4 | 58989 | 4 | 74695 |
| 2 | 49098 | 2 | 48240 | 2 | 57897 |
| 38 | 4 | 287564 | 36 | 3 | 284279 | 36 | 3 | 348165 |
| 3 | 235956 | 4 | 273581 | 4 | 286284 |
| 5 | 111178 | 2 | 39129 | 2 | 14782 |
| 37 | 4 | 212536 | 38 | 4 | 234221 | 38 | 4 | 240721 |
| 5 | 83676 | 3 | 185395 | 3 | 176570 |
| 3 | 68283 | 5 | 105657 | 5 | 102855 |

Highest number violations happen for code 21 and 38 and in time slots 3 and 4. i.e. 9AM -4 PM

1. **Let’s try and find some seasonality in this data**
   1. **First, divide the year into some number of seasons, and find frequencies of tickets for each season**

Divided the year into seasons of 3 months each

Season 1 : Jan – Mar

Season 2 : Apr - Jun

Season 3 : Jul – Sep

Season 4 : Oct – Dec

|  |  |
| --- | --- |
| **Year 2015** | |
| **Season** | **Ticket Numbers** |
| 2 | 2907331 |
| 1 | 2466640 |
| **Year 2016** | |
| **Season** | **Ticket Numbers** |
| 1 | 2668423 |
| 2 | 2202295 |
| 3 | 1057 |
| 4 | 846 |
| **Year 2017** | |
| **Season** | **Ticket Numbers** |
| 2 | 2760833 |
| 1 | 2669069 |
| 3 | 1046 |
| 4 | 970 |

* 1. **find the 3 most common violations for each of these season**

|  |  |
| --- | --- |
| **Year 2015 - Season 1** | |
| **Season** | **Ticket Numbers** |
| 38 | 336746 |
| 21 | 281386 |
| 14 | 219828 |
| **Year 2016 - Season 1** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 349297 |
| 36 | 341787 |
| 38 | 308987 |
| **Year 2017 - Season 1** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 373874 |
| 36 | 348240 |
| 38 | 287000 |
| **Year 2015 - Season 2** | |
| **Season** | **Ticket Numbers** |
| 21 | 439516 |
| 38 | 327158 |
| 14 | 246660 |
| **Year 2016 - Season 2** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 315234 |
| 36 | 273455 |
| 38 | 238083 |
| **Year 2017 - Season 2** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 393885 |
| 36 | 314525 |
| 38 | 255064 |
| **Year 2016 - Season 3** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 249 |
| 46 | 214 |
| 40 | 89 |
| **Year 2017 - Season 3** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 228 |
| 46 | 219 |
| 40 | 109 |
| **Year 2016 - Season 4** | |
| **Violation Codes** | **Ticket Numbers** |
| 21 | 167 |
| 46 | 164 |
| 40 | 80 |
| **Year 2017 - Season 4** | |
| **Violation Codes** | **Ticket Numbers** |
| 46 | 219 |
| 40 | 121 |
| 21 | 100 |

**Insights:**

* We can see that there is not much data in season 3 and 4, as file is a fiscal year file and we have assumed initially only the data with the year mentioned in the file name, so the maximum data it has for the year is only till June (only for 2 quarters)
* The Most common violation for each season can be seen in the table above

**7) Total fine collected by violation code**

Below is the plot of fine collected vs violation code. Most fine is collected for code 14 , 21 and 46

