select COUNT(differencespeedandlimit) as 'speeding\_voilation\_count', address

FROM mergetable

WHERE differencespeedandlimit > 15

GROUP BY ADDRESS

ORDER BY COUNT(differencespeedandlimit) DESC

LIMIT 50;

+--------------------------+-------------------------------------------------------------------------------+

| speeding\_voilation\_count | address |

+--------------------------+-------------------------------------------------------------------------------+

| 206766 | CBE E TAYLOR AVENUE - CASTLE HILL AVENUE |

| 196921 | GOW S VNB W 92ND STREET - BKLYN GANTRY LOWER LEVEL |

| 195870 | BE N STRATFORD AVENUE - CASTLE HILL AVE |

| 187203 | HRP N LAFAYETTE AVENUE - E TREMONT AVENUE |

| 186051 | TNB N Queens Anchorage - Toll Plaza |

| 184672 | CBE E CASTLE HILL AVE - BE N WATERBURY AVE |

| 183644 | WSE N VICTORY BLVD - SOUTH AVENUE |

| 182811 | BWB S Queens Anchorage - WSE S Exit 14 (Linden Pl) |

| 182095 | WSE N TYRELLAN AVENUE - BLOOMINGDALE ROAD |

| 181794 | BWB N Toll Plaza - HRP N Lafayatte Ave |

| 179418 | WSE S ARDEN AVENUE - BLOOMINGDALE ROAD |

| 178987 | CIP S Hempstead Tpk - Laurelton Pkwy @ SSP |

| 174196 | SIE E BRADLEY AVENUE - CLOVE ROAD |

| 172567 | WSE S VICTORY BOULEVARD - ARDEN AVENUE |

| 171054 | VWE N MP6.39 (Exit 11 Jewel Ave) - MP8.65 (Exit 13 Northern Blvd) |

| 170931 | BE N Castle Hill Avenue - Griswold Ave |

| 170764 | SIE E CLOVE ROAD - FINGERBOARD ROAD |

| 170691 | WSE N-SIE E SOUTH AVENUE - SOUTH AVENUE |

| 169163 | WSE S TYRELLAN AVENUE - 440 S FRANCIS STREET |

| 168255 | SIE W FINGERBOARD ROAD - CLOVE ROAD |

| 167709 | VWE S MP8.65 (Exit 13 Northern Blvd) - MP6.39 (Exit 11 Jewel Ave) |

| 167660 | Whitestone Expwy S Exit 14 (Linden Pl) - VWE S MP8.65 (Exit 13 Northern Blvd) |

| 167207 | CVE NB GCP - WILLETS PT BLVD |

| 165480 | BE S Griswold - Castle Hill Avenue |

| 161570 | WSE S SOUTH AVENUE - VICTORY BOULEVARD |

| 161162 | Belt Pkwy W 182nd St - JFK Expressway |

| 160817 | TNB S Toll Plaza - Queens Anchorage |

| 160683 | GOW S 7TH AVENUE - 92ND STREET |

| 158708 | SIE W RICHMOND AVENUE - SOUTH AVENUE |

| 157768 | WSE N ARDEN AVENUE - VICTORY BLVD |

| 155573 | I-87 NYST S Exit 1 - MDE S Van Cortlandt Park |

| 154651 | SIE W WOOLEY AVENUE - RICHMOND AVENUE |

| 152821 | CIP N ramp to TNB - TNB Queens Anchorage |

| 147107 | Whitestone Expwy N Exit 14 (Linden Pl) - BWB N Queens Anchorage |

| 146346 | SIE E SOUTH AVENUE - RICHMOND AVENUE |

| 144919 | LIE WB LITTLE NECK PKWY - NB CVE |

| 143130 | VNB E-GOWANUS N BROOKLYN GANTRY LOWER LEVEL - 92ND STREET |

| 139305 | Belt Pkwy E 182nd St - Laurelton Pkwy N @ SSP |

| 139251 | VNB W BROOKLYN GANTRY LOWER LEVEL - SI GANTRY LOWER LEVEL |

| 136318 | BE S CASTLE HILL AVENUE - STRATFORD AVENUE |

| 136284 | SIE E RICHMOND AVENUE - WOOLEY AVENUE |

| 134524 | VNB W-SIE W SI GANTRY LOWER LEVEL - FINGERBOARD ROAD |

| 134379 | SIE W BRADLEY AVENUE - WOOLEY AVENUE |

| 130797 | CIP N LIE ramp - TNB |

| 130751 | SIE E WOOLEY AVENUE - BRADLEY AVENUE |

| 128984 | CBE W CASTLE HILL AVENUE - TAYLOR AVENUE |

| 128642 | LIE W 108TH ST - 84TH ST |

| 127269 | CVE NB LIE - WILLETS PT BLVD |

| 125347 | VNB E SI GANTRY LOWER LEVEL - BROOLKYN GANTRY LOWER LEVEL |

| 122853 | VNB W BROOKLYN GANTRY UPPER LEVEL - SI GANTRY UPPER LEVEL |

+--------------------------+-------------------------------------------------------------------------------+

It can be seen that there are several locations, especially ‘Avenues’, where speeding violations are more likely to occur. Moreover, from our query, we can see which ‘Expressway’, ‘bridges’, ‘Toll’, ‘Tunnel’ are most likely to have speeding voilations.

Note that since it is natural to have more avenues than expressway in a city, we do not assume that the data points in our dataset are equal for each category of location. However, we underline the most important locations where traffic violations are most likely to occur.

select COUNT(differencespeedandlimit) as 'speeding\_voilation\_count', month

FROM mergetable

WHERE differencespeedandlimit > 15

GROUP BY month

ORDER BY COUNT(differencespeedandlimit) DESC;

+--------------------------+-----------+

| speeding\_voilation\_count | month |

+--------------------------+-----------+

| 1436103 | SEPTEMBER |

| 1429358 | AUGUST |

| 1281037 | MAY |

| 1197597 | JULY |

| 1168592 | OCTOBER |

| 1083673 | JUNE |

| 1070228 | APRIL |

| 885087 | MARCH |

| 696111 | FEBRUARY |

| 612310 | JANUARY |

| 555209 | DECEMBER |

| 446238 | NOVEMBER |

+--------------------------+-----------+

We note that violations are more likely to occur in September and August. This is interesting because September and August marks the end of Summer vacations in majority of the schools in New York City, increasing the total number of traffic on the road. We also note the the coldest months (November - February) are among the bottom five for speeding voilations.

We would like to narrow down our selection of months by seeing what are the popular days within these months on which traffic speed violations are most likely to occur.

select COUNT(differencespeedandlimit) as 'speeding\_voilation\_count', month, dayofweek

FROM mergetable

WHERE differencespeedandlimit > 15

GROUP BY month, dayofweek

ORDER BY COUNT(differencespeedandlimit) DESC

LIMIT 20;

+--------------------------+-----------+-----------+

| speeding\_voilation\_count | month | dayofweek |

+--------------------------+-----------+-----------+

| 234674 | SEPTEMBER | SUNDAY |

| 227680 | SEPTEMBER | MONDAY |

| 224799 | SEPTEMBER | TUESDAY |

| 224272 | MAY | THURSDAY |

| 218095 | AUGUST | SUNDAY |

| 215394 | AUGUST | FRIDAY |

| 214710 | AUGUST | SATURDAY |

| 214199 | AUGUST | WEDNESDAY |

| 211110 | OCTOBER | TUESDAY |

| 210048 | MAY | WEDNESDAY |

| 201176 | AUGUST | TUESDAY |

| 197303 | SEPTEMBER | SATURDAY |

| 195325 | OCTOBER | MONDAY |

| 194113 | SEPTEMBER | WEDNESDAY |

| 193970 | JULY | THURSDAY |

| 192921 | OCTOBER | SUNDAY |

| 192039 | AUGUST | THURSDAY |

| 184274 | SEPTEMBER | THURSDAY |

| 182778 | MAY | TUESDAY |

| 182490 | JUNE | THURSDAY |

+--------------------------+-----------+-----------+

We note that Sundays and Mondays are among the top days at which violations are likely to take place.

We will also like to explore the relationship between months and the visibility and the effect of these both to the traffic speed violations.

select COUNT(differencespeedandlimit) as 'speeding\_voilation\_count', month, visibility

FROM mergetable

WHERE differencespeedandlimit > 15

GROUP BY month, visibility

ORDER BY COUNT(differencespeedandlimit) DESC

LIMIT 30;

e monitored at: http://compute-1-1.local:25000/query\_plan?query\_id=e34473eb4ecf734d:5863e19700000000

+--------------------------+-----------+------------+

| speeding\_voilation\_count | month | visibility |

+--------------------------+-----------+------------+

| 1084243 | SEPTEMBER | 10 |

| 908204 | AUGUST | 10 |

| 795004 | OCTOBER | 10 |

| 759909 | JULY | 10 |

| 719449 | APRIL | 10 |

| 697568 | JUNE | 10 |

| 645296 | MARCH | 10 |

| 636905 | MAY | 10 |

| 438937 | JANUARY | 10 |

| 389209 | NOVEMBER | 10 |

| 378693 | DECEMBER | 10 |

| 375082 | FEBRUARY | 10 |

| 97054 | AUGUST | 9.1 |

| 91483 | MAY | 3 |

| 88684 | MAY | 2.5 |

| 76009 | JULY | 9.1 |

| 74784 | AUGUST | 8 |

| 67519 | MAY | 9.1 |

| 64915 | AUGUST | 7 |

| 60874 | JULY | 8 |

| 55311 | MAY | 4 |

| 51473 | OCTOBER | 9.1 |

| 51078 | JUNE | 2.5 |

| 50191 | AUGUST | 4 |

| 50145 | AUGUST | 6 |

| 49947 | MAY | 2 |

| 48087 | MAY | 8 |

| 46809 | JULY | 7 |

| 45775 | JUNE | 3 |

| 44749 | SEPTEMBER | 3 |

+--------------------------+-----------+------------+

This is an interesting result. Now, we understand why certain months have higher likelihood of speeding than the others. It is simply because visibility in HOTTER months is higher than visibility in colder months. This implies that a higher visibility pushes the drivers to drive faster, leading to speeding.