

|  |  |
| --- | --- |
|  | Exploring Traffic Violations Data in United States |
|  |  |
|  | Krishna Bhimraj Ghorpade  CIN: 306603510  Email: kghorpa@calstatela.edu |

1. **Dataset URL’s:**

[**https://catalog.data.gov/dataset/traffic-violations-56dda**](https://catalog.data.gov/dataset/traffic-violations-56dda)

[**https://www.kaggle.com/felix4guti/traffic-violations-in-usa**](https://www.kaggle.com/felix4guti/traffic-violations-in-usa)

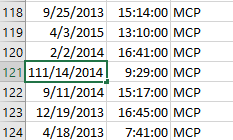
**Objective of the study:**

This dataset contains information about the traffic violation which occurred in the United States. This dataset was published by the Montgomery County of Maryland in order to analyze the traffic violations and selected incidents within the United States. Every year, thousands of accidents occur due to a multitude of contributing factors like countless citations have been issued for traffic violations across the country, and states have reaped untold billions of dollars of revenue from violators [1]. Traffic violations are generally divided into major and minor types of violations. The most minor type are parking violations, which are not counted against a driving record, though a person can be arrested for unpaid violations [2]. Most of the reasons for the traffic violation was like over speeding, Driving with a suspended license, Violation of license restrictions, Committing a hit-and-run accident. Ignoring traffic signs and traffic cops, running red lights and stop signs, not pulling to the side for an emergency vehicle etc. [3] and most traffic offenses are infractions, which are minor crimes, they can still have negative consequences. Traffic violations can result in expensive tickets, higher insurance rates, and possibly suspension of your driver's license [4]. This data accumulated provides a platform to investigate the various details of the violations like the location of incident, fatal injury, alcohol consumption, it’s severity, violation type, property damage, personal injury and vehicle details like the its type, make, color etc. It contains 1018634 rows and 35 columns which covers the whole United States.

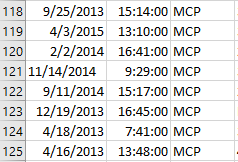
1. **Data Cleaning:**
2. **Illegal Values:**

The date entered is incorrect. It is entered as 111/14/2014. The month value 111 is out of range and cannot be considered. So, it is corrected and changed to 11/14/2014. It is shown in the following table.

**Before:**



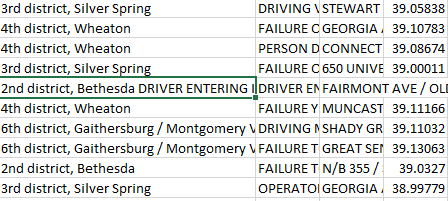
**After:**



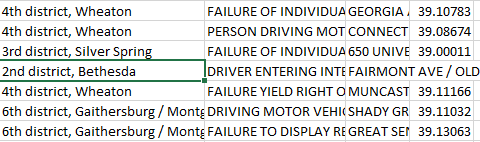
1. **Embedded Values:**

The two columns shown are for the location and the description. But the two data are entered in the same column instead of one. This can be seen in the following table. It is corrected, and the two entries are being entered in their respective fields.

**Before:**



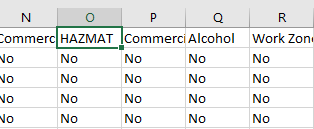
**After:**



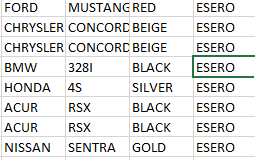
1. **Cryptic values, Abbreviations:**

There are 3 cryptic values and Abbreviations found in the data set. They are as follows.

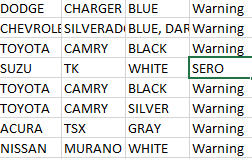
**HAZMAT:** Hazardous Materials



**ESERO:** Electronic Safety Equipment Repair Order



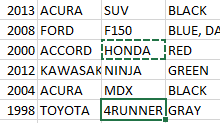
**SERO:** Safety Equipment Repair Order



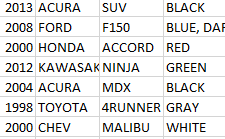
1. **Misfielded value:**

ACCORD is the name of the car and HONDA is the Maker’s name. These 2 values were misfielded. They were assigned it to their respective column after.

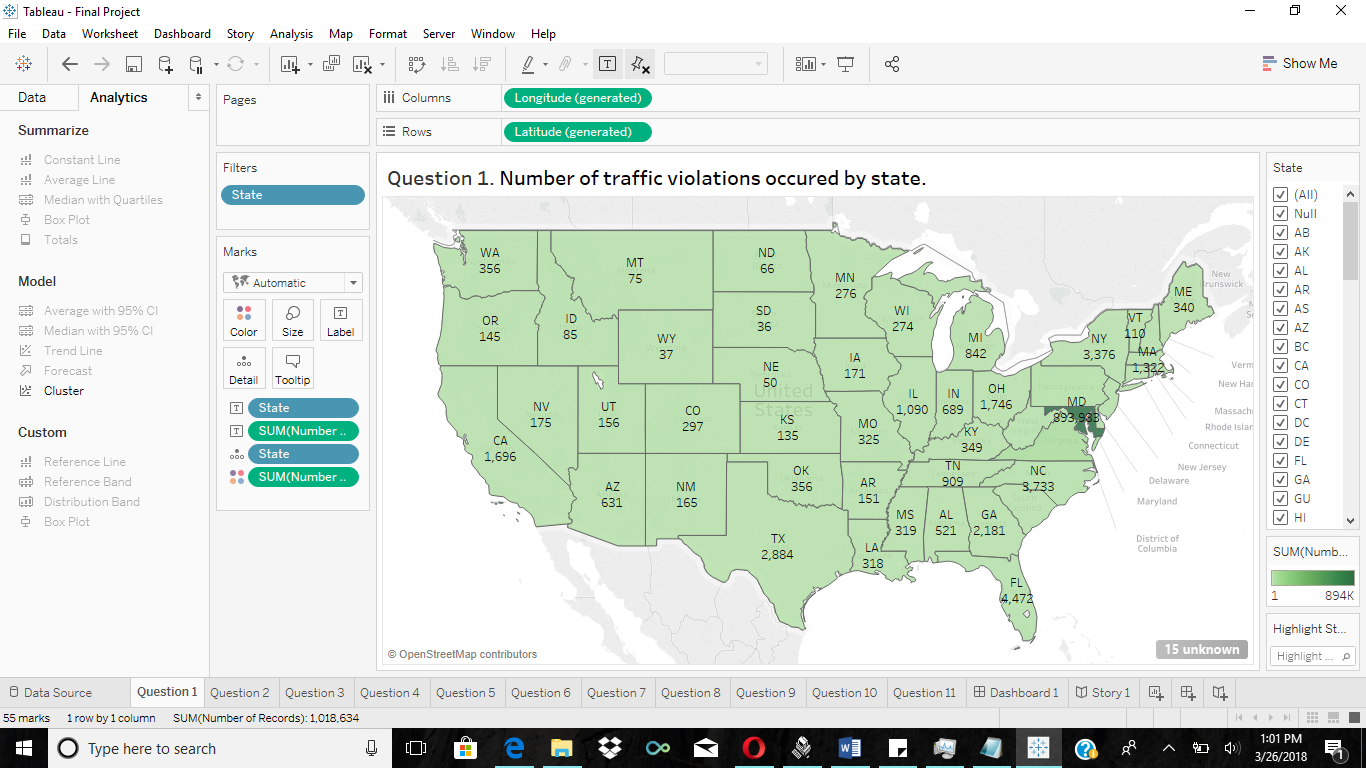
**Before:**



**After:**



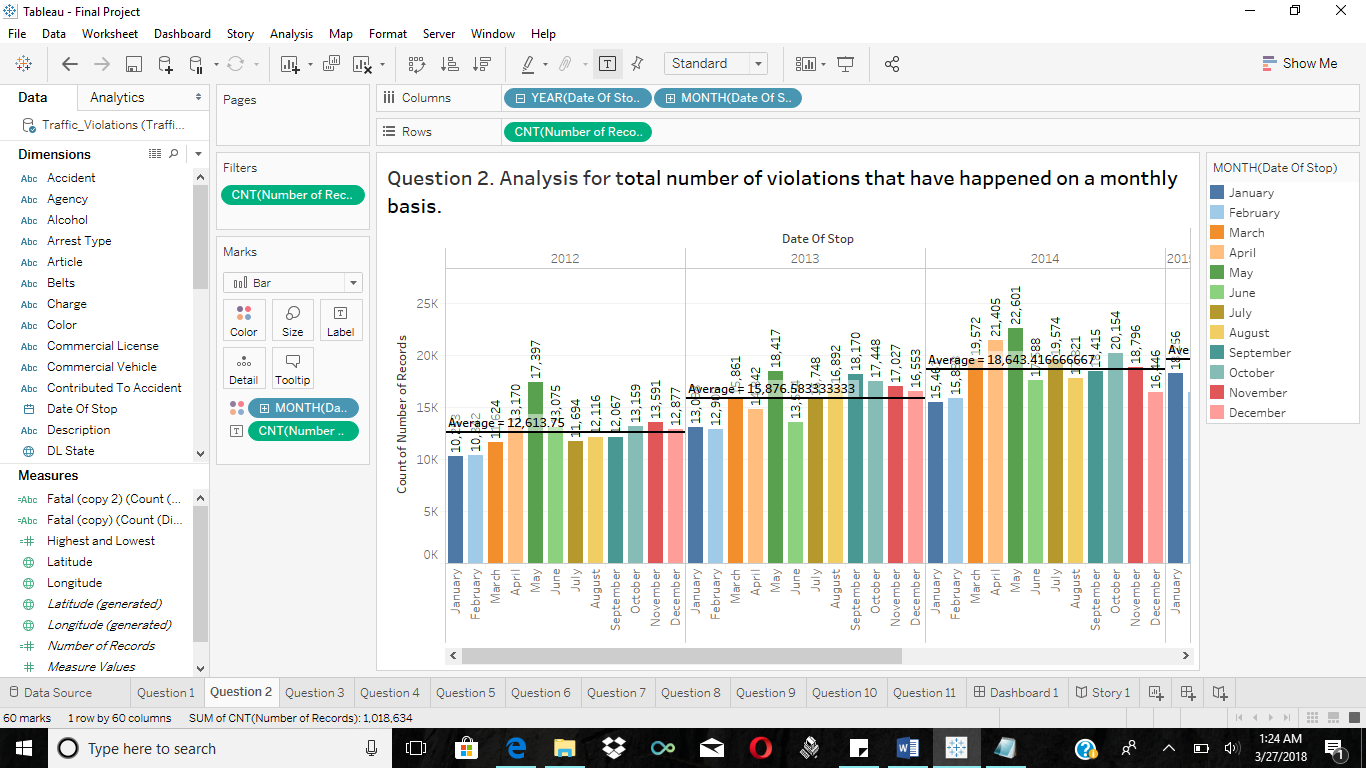
1. **Data Visualizations**
2. **What is the count of Number of traffic violations occurred by state?**



**[Tools used: Geographic map]**

This visualization illustrates the states where the traffic violations were recorded. As seen in the area of United States, the States like Maryland(MD), Florida (FL) etc., recorded the maximum and State like Wyoming (WY), South Dakota (SD) etc., recorded the minimum traffic violation. The State of Maryland (MD) is marked dark green in color for better understanding to indicate the maximum recorded traffic violations. It also describes the severity of traffic violations by state. To visualize the spatial question, using geographic map helps to understand the trends or patterns in data.

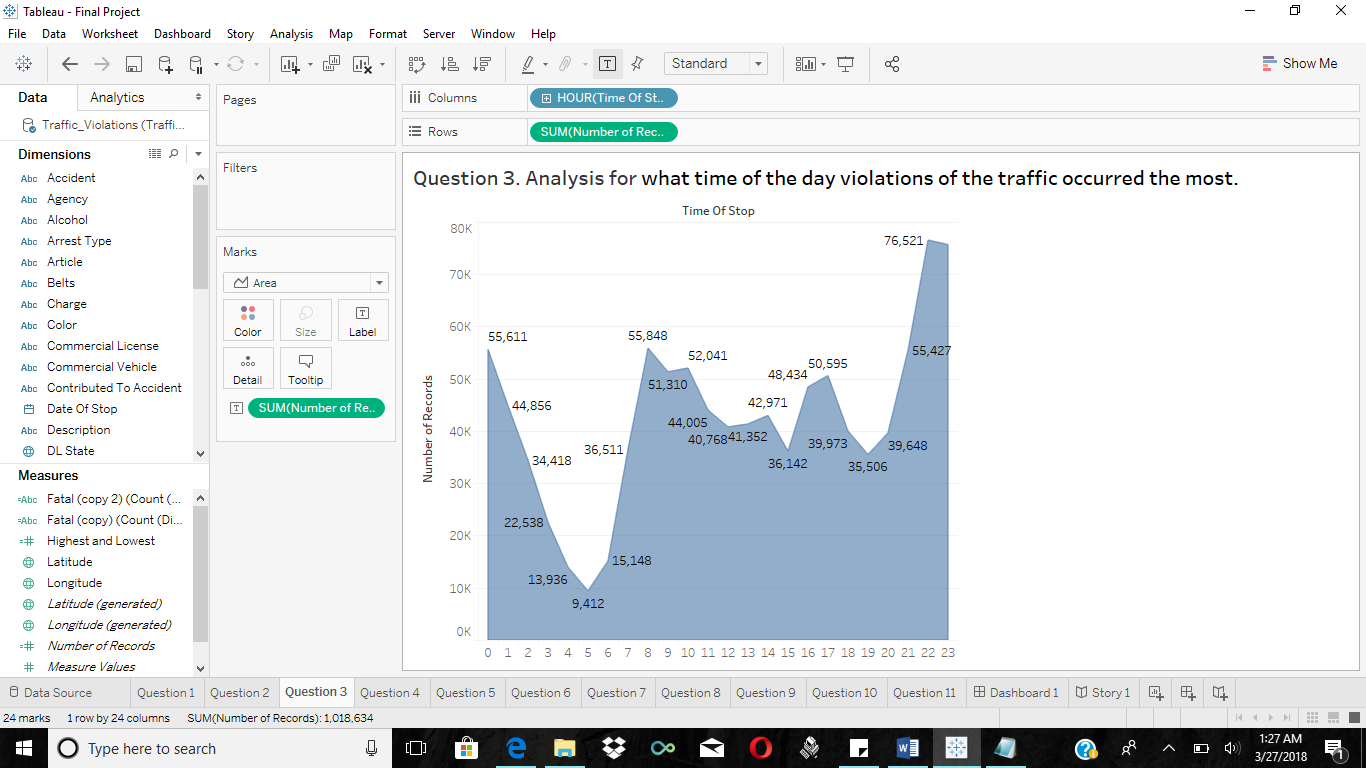
1. **Analysis for total number of violations that have happened on monthly basis.**



**[Tools used: Calculated fields, Dates, Reference Line]**

The above analysis using a side-by-side bar chart shows the count of the total number of traffic violations that happened per month of all the years considered with the help of a calculated field COUNT ([Number of Records]). The different months are all displayed in different colors as shown in the filter label. The average of the number of records per year has also been displayed using the reference line to have a rough idea about the number of violations occurred. The average number of violations was the highest in 2014 and lowest in 2012 which says that there is a considerable increase in the number of traffic violations. It can also be seen that the lowest number of traffic violations was in January 2012 where there were just 10,253 violations and the highest was in May 2014 where there were 22,601 violations which were recorded.

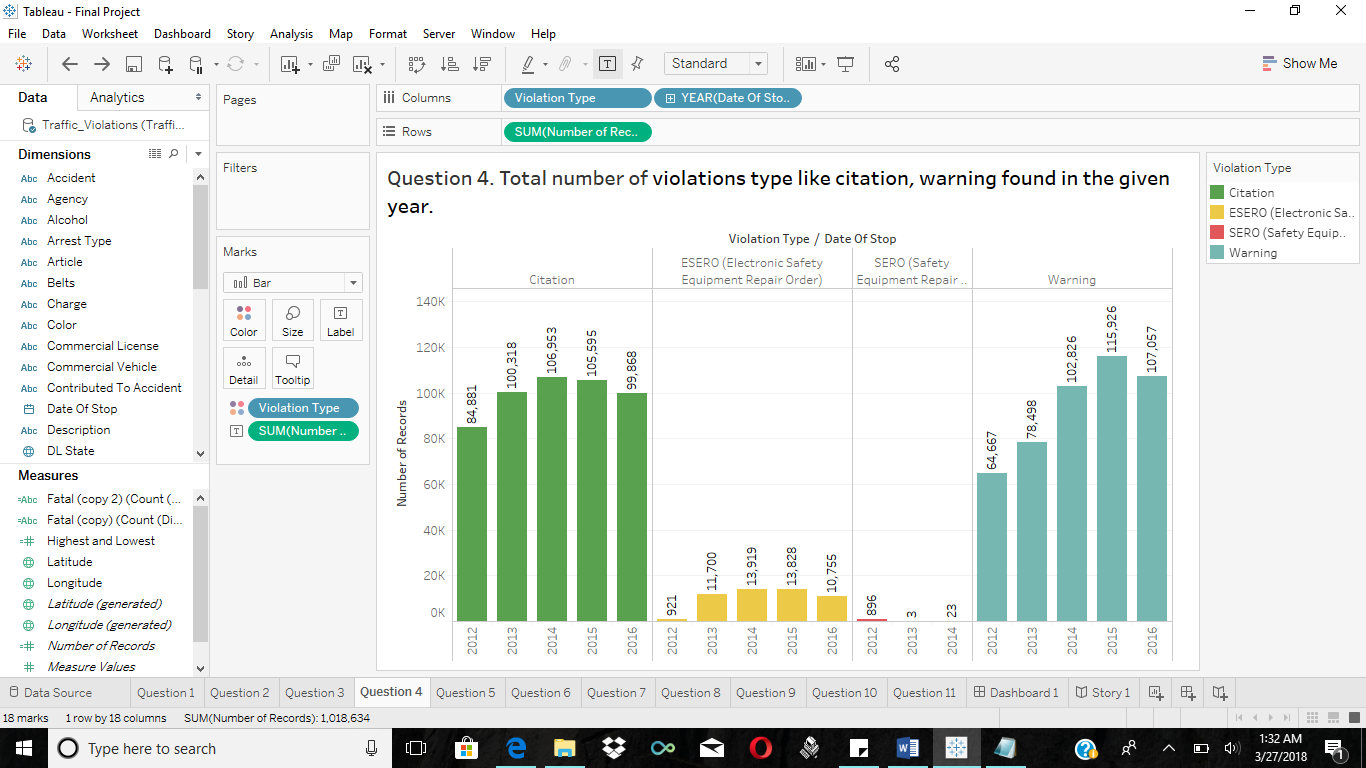
1. **Analysis for what time of the day violations of the traffic occurred the most.**



**[Tools used: Calculated Fields, Time, Area Charts]**

This analysis is done by using an area chart which shows the count of the total number of traffic violations that happened at what time of the day in the whole 24 hours with the help of Number of Records and the Time of Stop. It can be seen from the above visualization that the majority of the traffic violations are occurring around day time between 8:00 hours to 17:00 hours. The maximum traffic violations were recorded at 22:00 hours which is 76,521 and minimum traffic was found at 05:00 hours which is 9,412.

1. **What was the count for the total number of violations type like citation, warning found in the given year?**

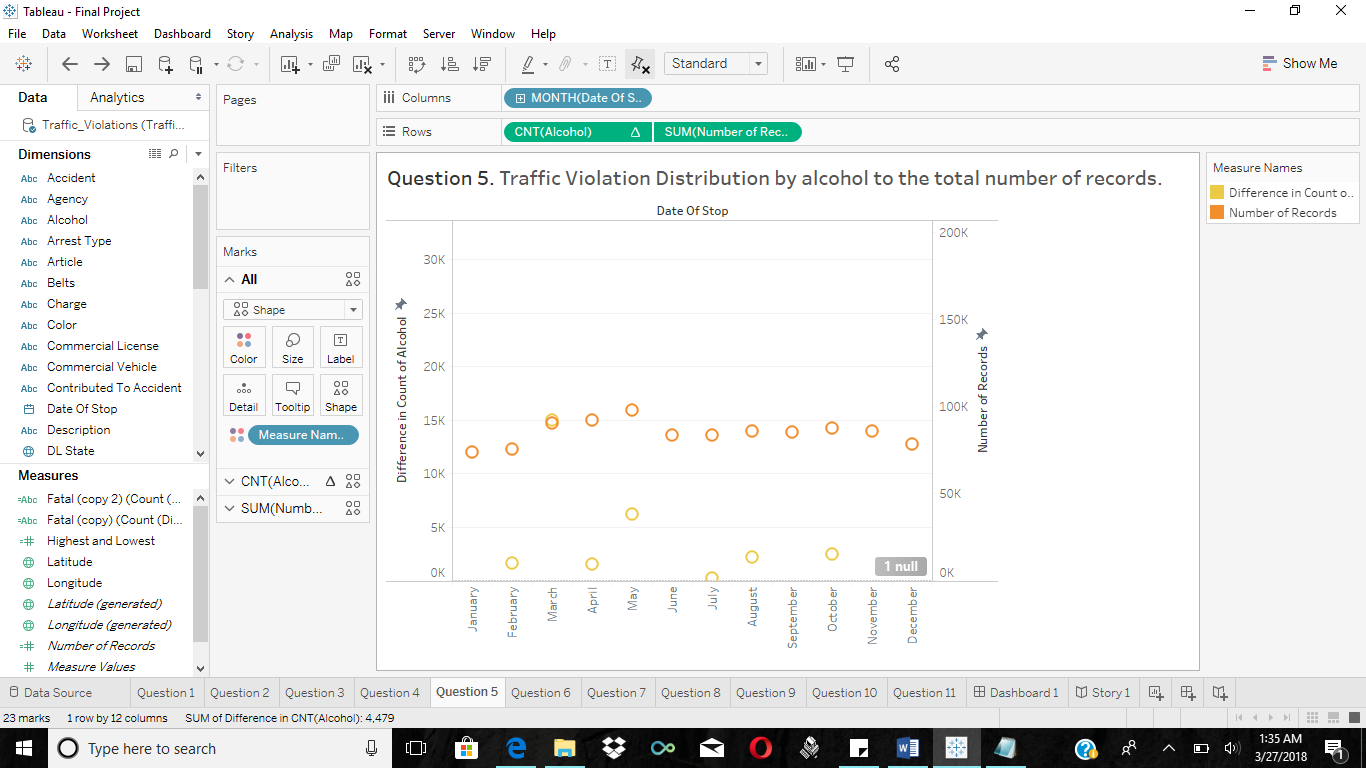


**[Tools used: Calculated Fields, Side-by-side bars]**

This visualization using a side-by-side bar chart shows the total number of violations type like Citation, EESRO (Electronic Safety Equipment Repair Order), SERO (Safety Equipment Repair Order) & Warning with the help of a field Violation Type / Date of Stop and Number of Records

The different violations types are all displayed in different colors as shown in the filter label. It can be seen that violation type Warning is recorded the most which is 115,926 in 2015 and violation type SERO is recorded the minimum which is 3 in 2013. Violation type Citation and Warning have recorded the followed compared to ESERO and SERO which are minimum.

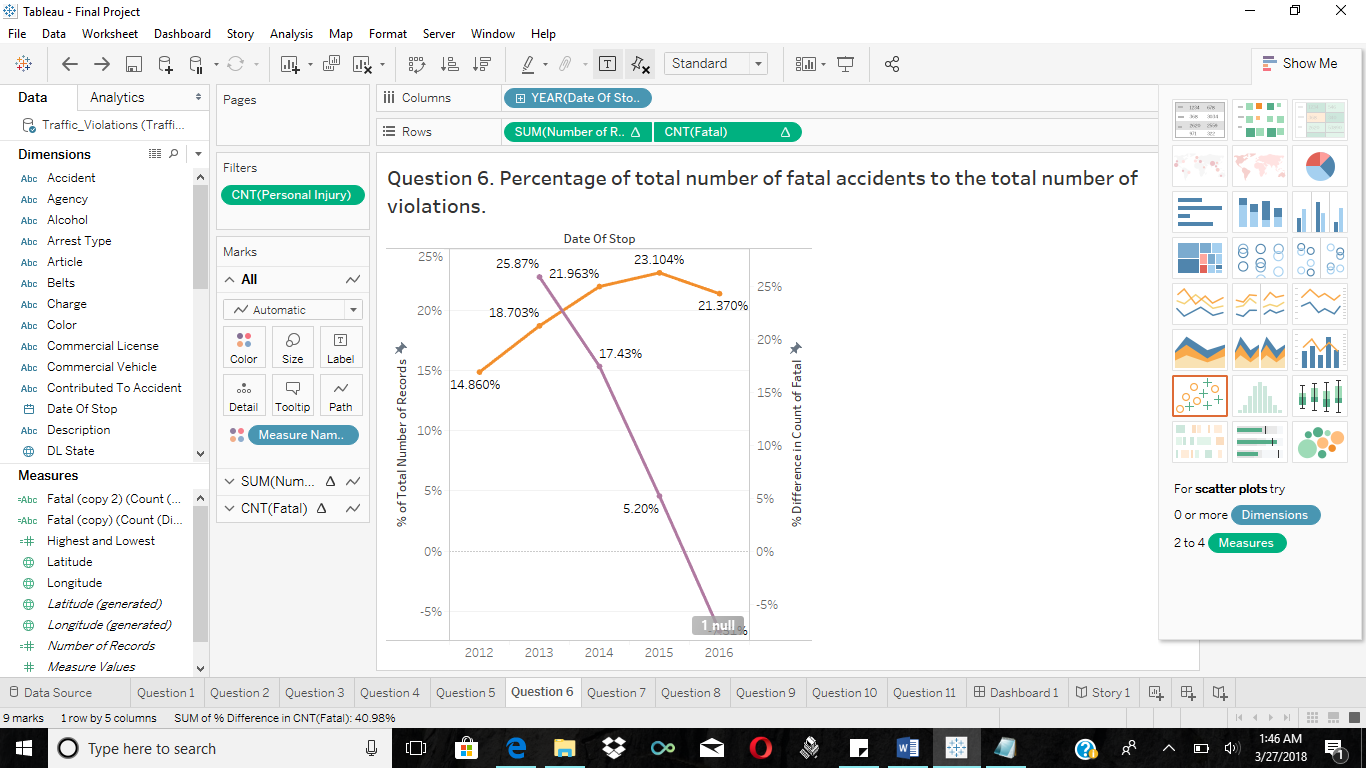
1. **How many Traffic Violation Distribution by alcohol were found to the total number of records?**



**[Tools used: Calculated Fields, Scatter plots]**

In addition to the above bubble chart this above visualization uses two calculated fields, Number of Records and Difference in count of alcohol in a scatter plot to find out the exact count of traffic violations in which the violator had alcohol consumption. It can be found that in the month of march the traffic violation was maximum which is 14,985 along with total number of violations which is 90,342 and the minimum is in July which is 251 along with number of violations which are 83,683. From this we can infer that alcohol has played major role in traffic violations and the government should make new strict rules for people driving car with alcohol consumption.

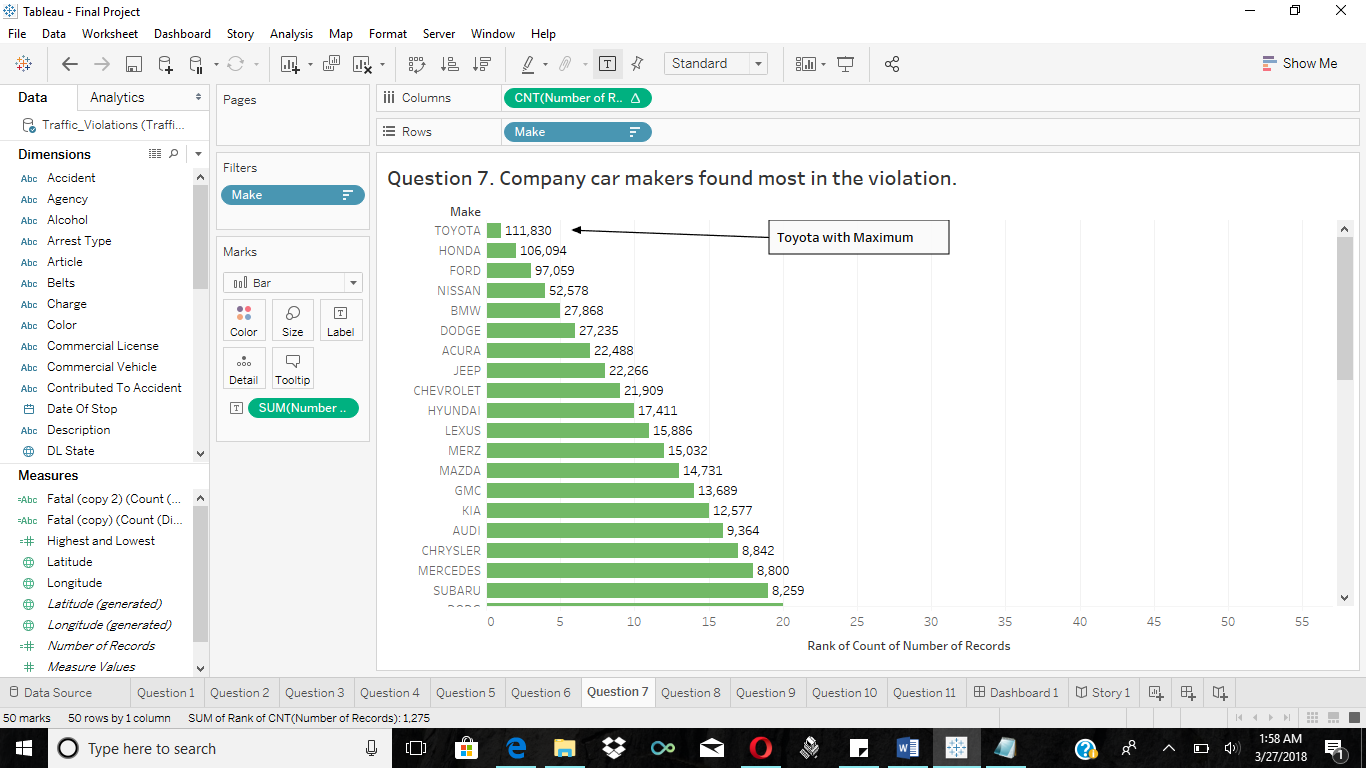
1. **Percentage analysis of total number of fatal accidents to the total number of violations.**



**[Tools used: Calculated Fields, Dual Axis Chart]**

This Visualization shows the percentage of total number of fatal accidents to the total number of violations. This is done considering the fields Total Number of Records which is orange in color on the left-hand side of the X-axis and Count of Fatal is purple in color on the right-hand side of the X-axis. The percentage total number of records and percentage difference in count of fatal is considered for understanding the statistics. It can observed that in the year 2013 recorded the maximum fatal accidents which is 25.87 % along with total number of records which is 18.703%. The year 2015 recorded the minimum fatal accidents which is 5.20% along with total number of records 23.104%. The year 2016 recorded the negligence of fatal accidents. It can be observed that as the number of traffic violations are increasing per year, the number of fatal accidents are decreasing.

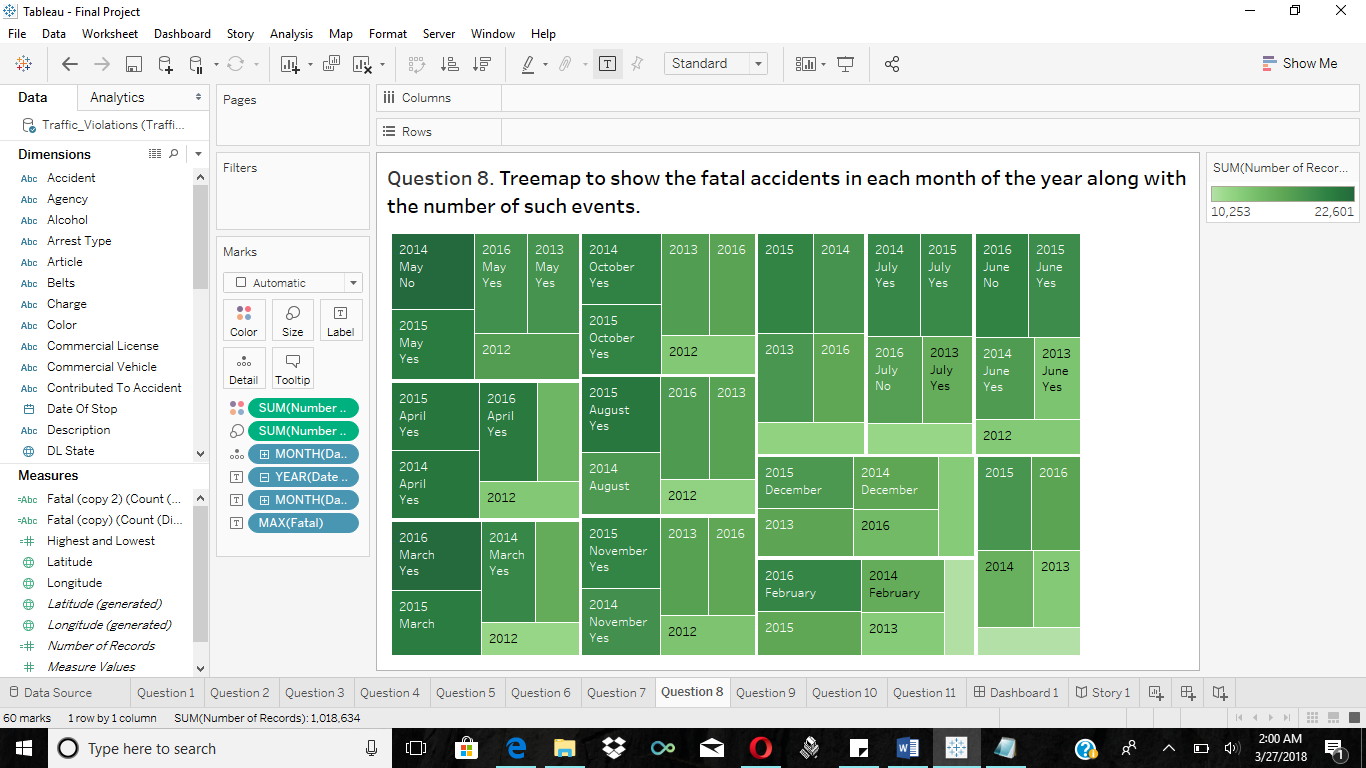
1. **Which company car makers were found most in the violation?**



**[Tools used: Calculated Fields, Rank]**

The above analysis shows the which Company car makers found in the traffic violations. The classification is done considering the total number of records and the rank for the car maker company. The car maker company Toyota is caught maximum times i.e.111,830 and has ranked 1 for the traffic violations followed by Honda which is ranked 2 and has 106,094 records, Ford which is ranked 3 and has 97,059 etc. This clearly explains which manufactures cars are found mostly in the violation.

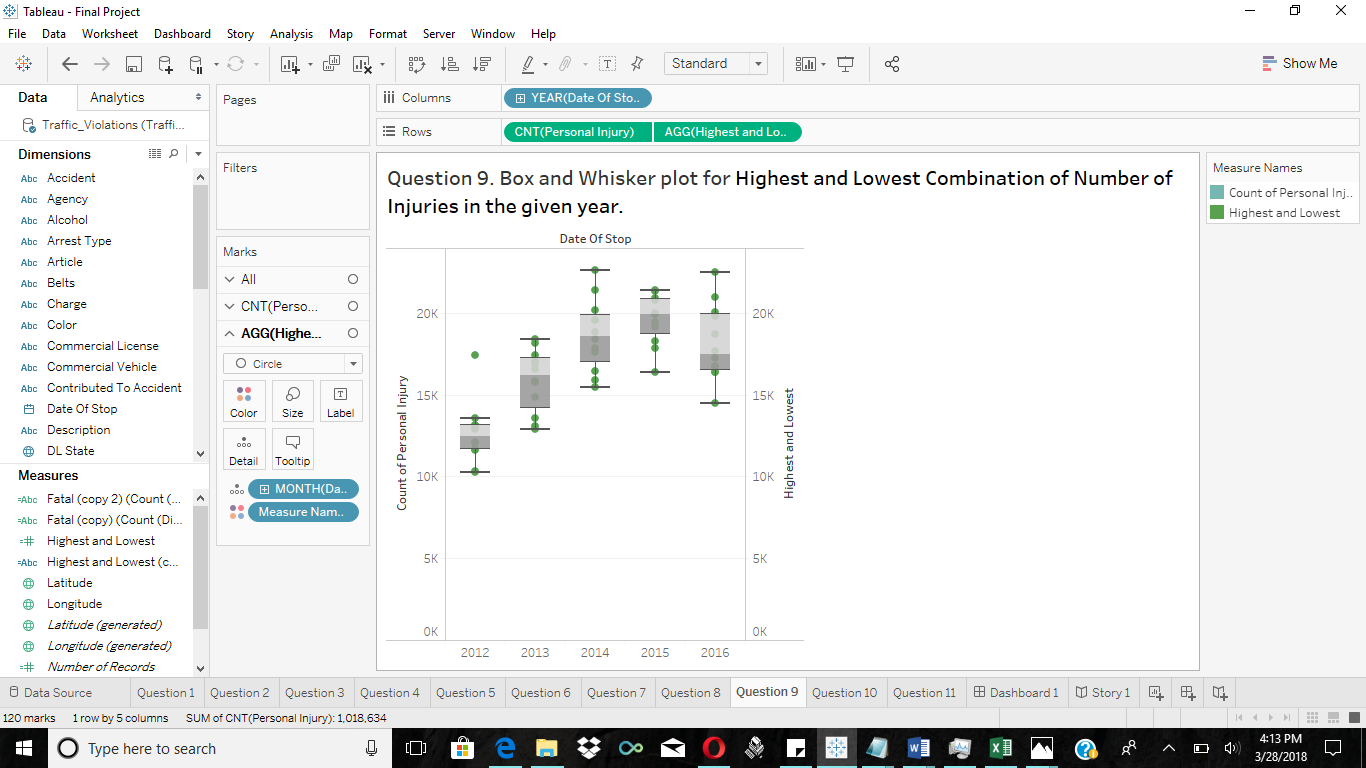
1. **Which month of the given year recorded fatal accidents along with the number of given events?**



**[Tools used: Calculated Fields, Dates, Tree Map]**

The above visualization demonstrates which month of the particular recorded the fatal accidents of the total traffic violations with the help of tree-map. The ones which says yes are the ones which recorded the fatal accidents and the others didn’t record the fatal accidents. The color contrast from dark to light demonstrates the count of fatal accidents to the total number of records. If color is dark green, then maximum violations occurred along with fatal accidents. If the color is light green, then minimum violations occur along with fatal accidents. This tree-map visualization expands from dark green from the left edge to light green color to the right edge. The dark green color indicates the total number of records in the given month of particular year along with the total number fatal accidents. Demonstration of number of such events that have occurred and that can be distinguished by the size of the blocks in the tree-map. More the number of incidents, bigger is the block of the tree-map. I have also labeled the respective years on them to make it more understandable. The number of such violations that have occurred is again calculated by a calculated field which counts the number of such violations. This helps us analyze the severity of the fatal accidents and to see if they have reduced over time.

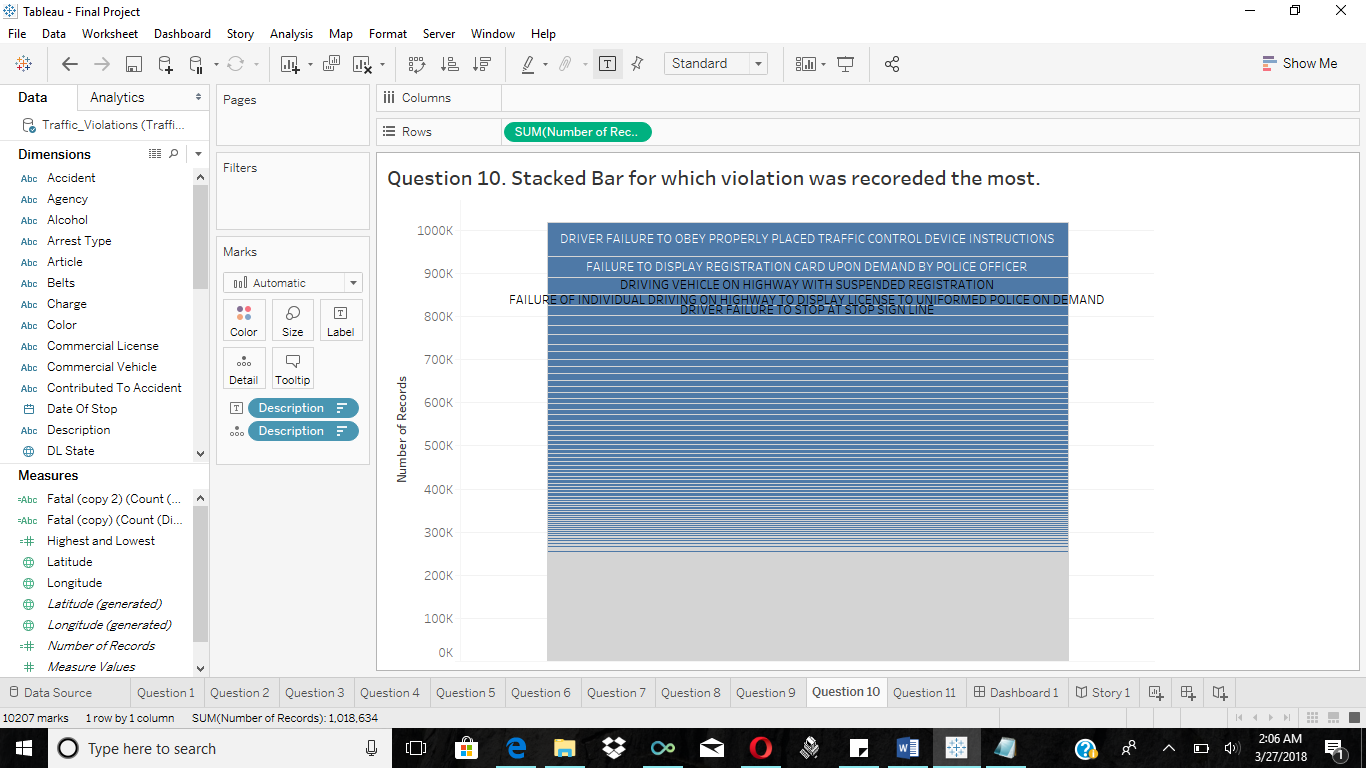
1. **Analysis for Highest and Lowest Combination of Number of Injuries in the given year.**



**[Tools used: Calculated Fields, Dates, Box & Whisker plot]**

This visualization refers to a period for Highest and Lowest Combinations of Number of Injuries in the given year. It helps to investigate the injuries caused to the driver and see what should be done to avoid this in the future. As shown above, the highest number of injuries, i.e. 22,601 were recorded in the year 2014 which makes 2014 the crucial year for the maximum injuries followed by the second highest in row with 22,521 injuries in the year 2016. The year 2012 is considered as the safe because minimum injuries were recorded i.e. 10,253 followed by 12,909 in the year 2013. The total number of maximum and minimum injuries has been got from a calculated field AGG (Highest and Lowest) = STR (COUNT ([Date of Stop])). The Whisker plot is a convenient way of graphically depicting groups of the total injuries through their quartiles. The lines extending vertically from the boxes (whiskers) indicates variability outside the upper and lower quartiles. The bottom and top of the box are the first and third quartiles, and the band inside the box is the second quartile (the median). The entire summary of every plot is listed out on the right in the summary plot when the cursor is pointed to the respective observation.

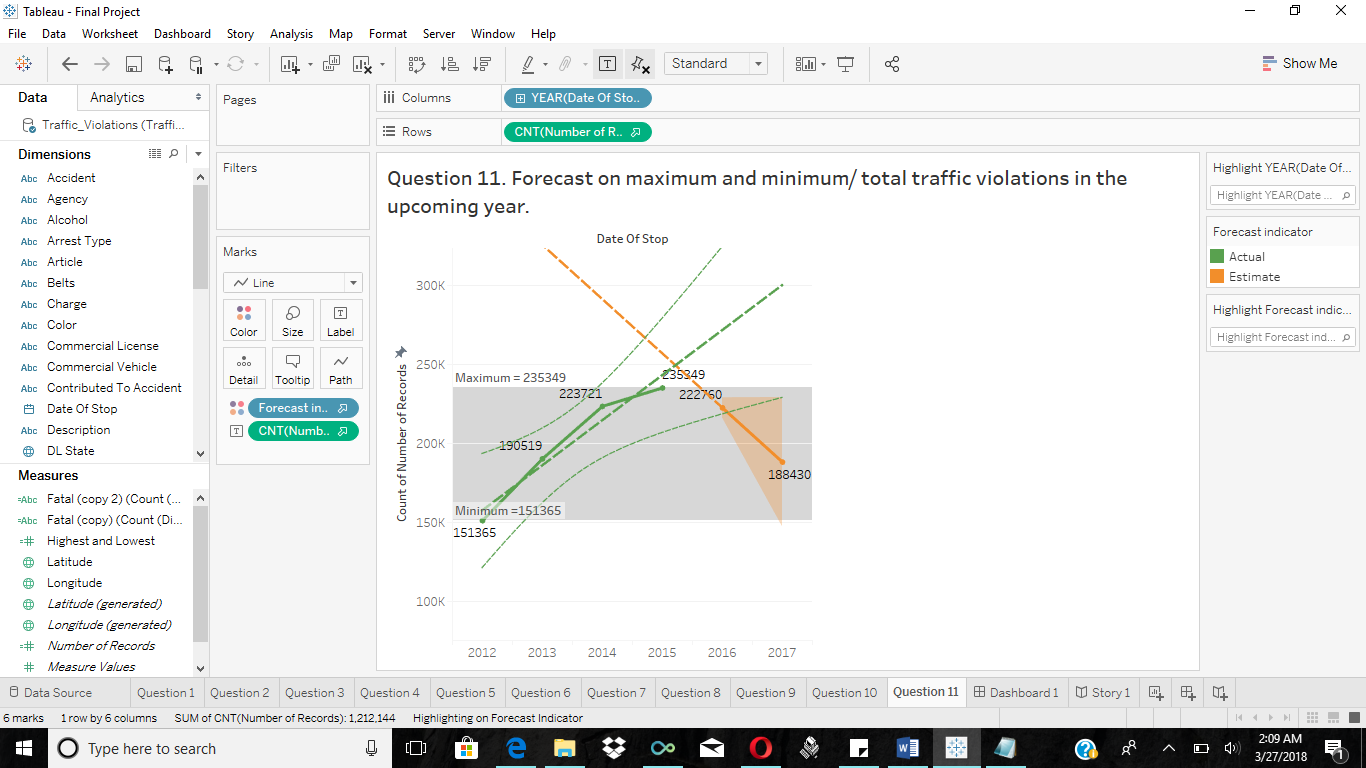
1. **Which type of violation was recorded the most?**



**[Tools used: Calculated Fields, Stacked Bar]**

This Visualization shows that which violation was recorded the most. The most common violation was the DRIVER FAILURE TO OBEY PROPERLY PLACED TRAFFIC CONTROL DEVICE INSTRUCTIONS which was counted 79,659 followed by FAILURE TO DISPLAY REGISTRATION CARD UPON DEMAND BY POLICE OFFICER which is 48,701. The minimum was FOR THE PURPOSE OF AVOIDING TRAFFIC CONTROL DEVICE which is 2.

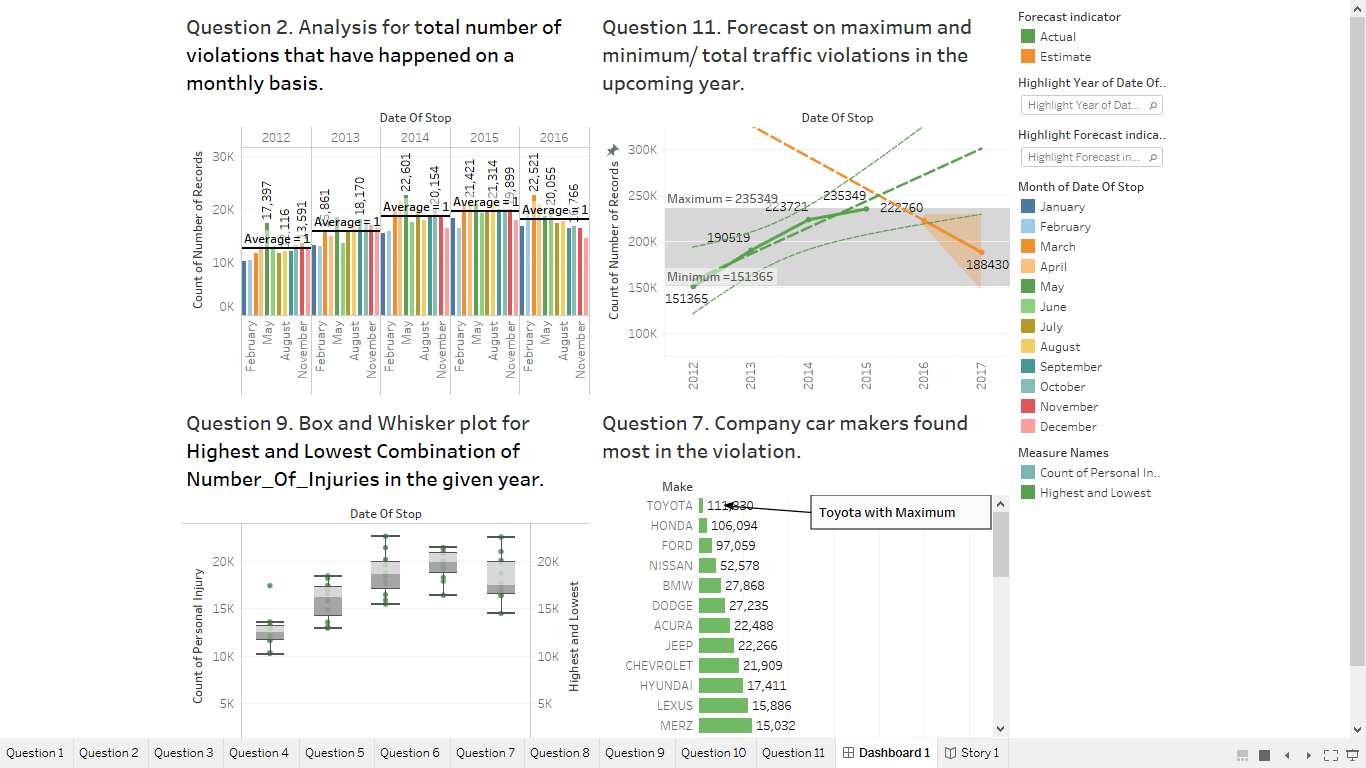
1. **Forecast on maximum and minimum/ total traffic violations in the upcoming year.**



**[Tools used: Calculated Fields, Dates, Forecast Trend Lines]**

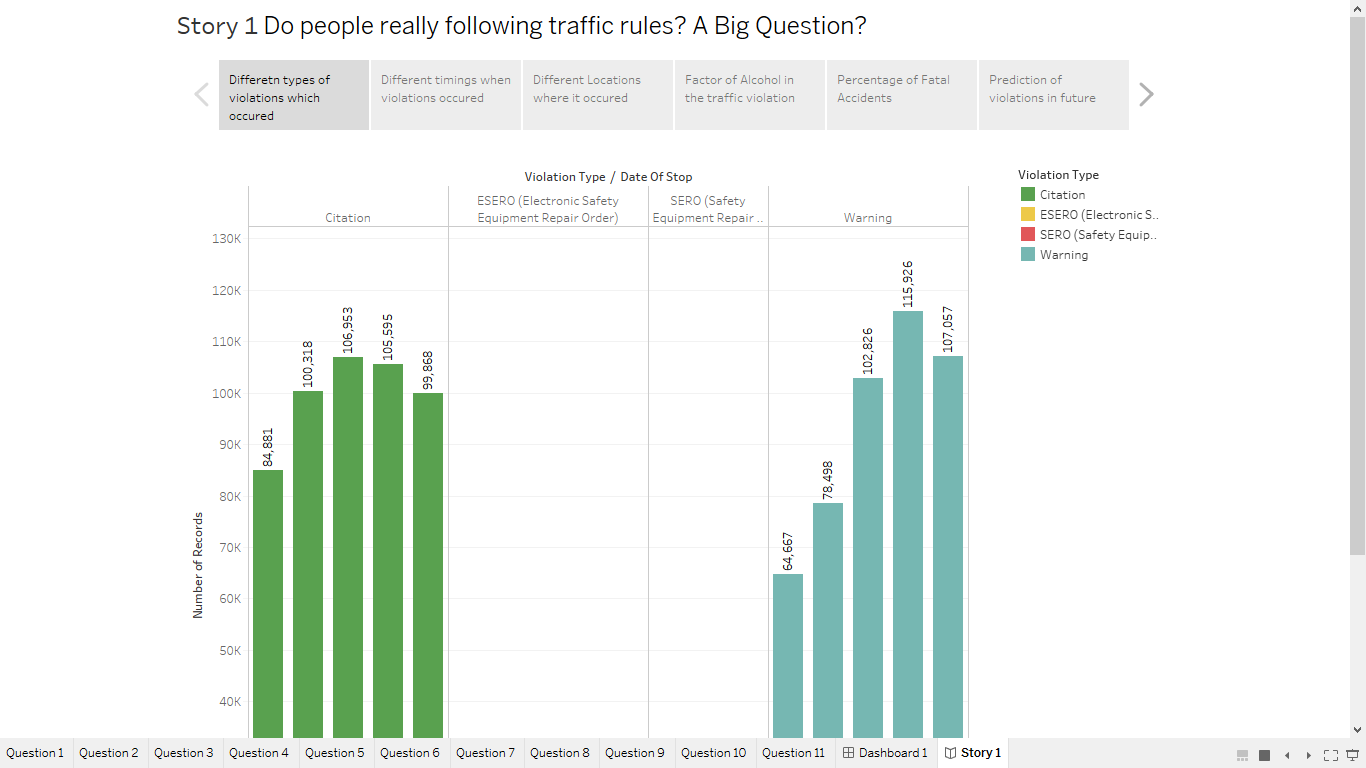
This Visualization shows the predicted values of the total number of traffic violations in the upcoming years. The field Number of records is considered to get the future prediction. The reference band is used above to show the maximum and minimum traffic violations which occurred in the given years. The minimum was in the year 2012 which had a count of 151,365 and the maximum was in the year 2015 which had 235,349 violations. But in future the count decreases. The forecast line is shown orange in color which is decreased to 188,430. The upcoming year 2017 will have less number of traffic violations compared to the previous years.

1. **Dashboard:**



1. **Story Telling**

Using car is considered as one of the best and preferred ways of travel these days for short distances and to commute within 40 to 50 miles of radius. It helps to avoid the time-consuming part like waiting, halting etc. of public transport like buses and metros. Travelling by bus for 10 miles takes 40 to 45 minutes by bus but if we have a car the time is reduced to 15 to 20 minutes. So, traveling by car is most convenient and preferred way. It’s easy to think of car as the easy mode of transportation but following the traffic rules while driving is more of all of these. If you don’t follow the traffic rules, the driver is caught by police, considered as violent, ticket is being given, lands in trouble and even can be fatal. So, while driving, the rider should follow the traffic rules to avoid violations. The number of traffic violations have been increasing per year by considering that the number of cars and passengers have increased in the last recent years.



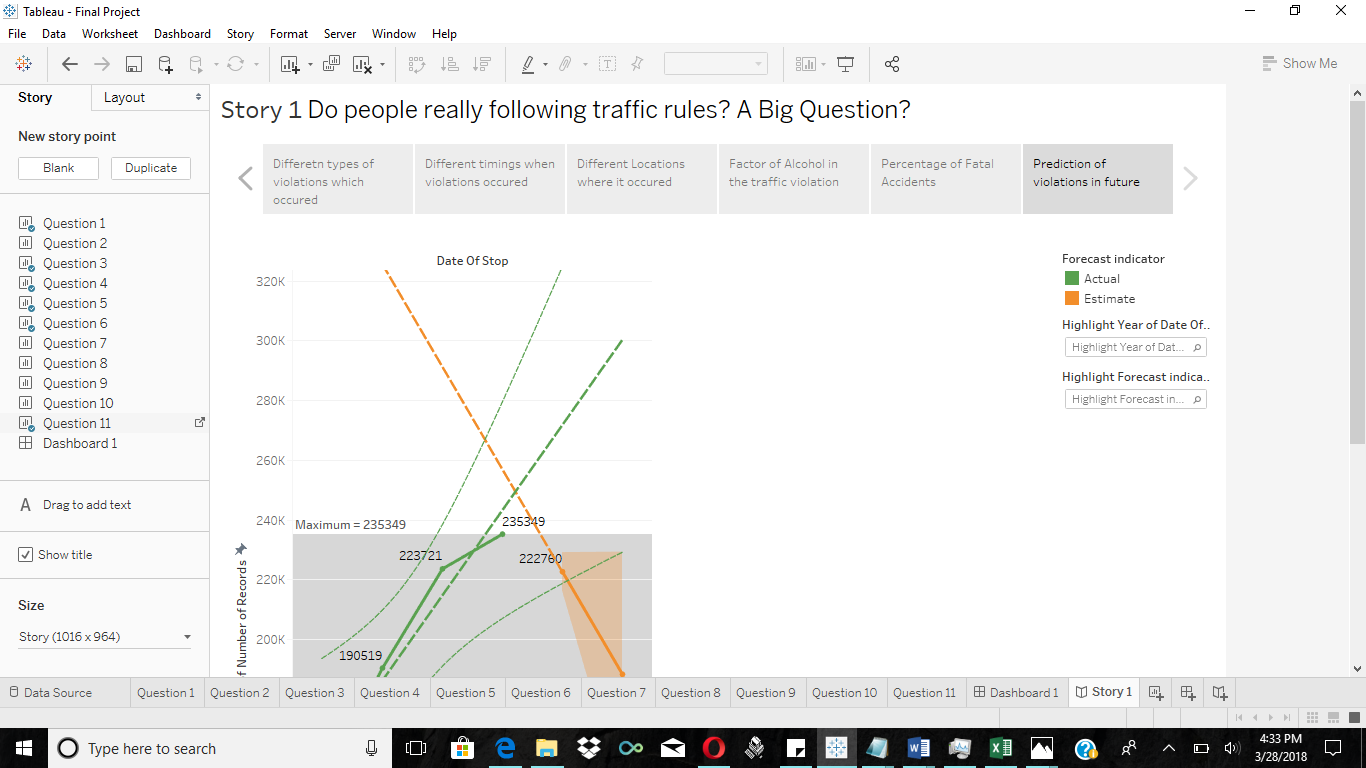
Each state's Department of Motor Vehicles or Bureau of Motor Vehicles maintains a database of drivers, including their convicted traffic violations. Upon being ticketed, a driver is given the option to mail into the local court or the court for the jurisdiction in which the violation is alleged—a plea of guilty, not guilty or nolo contendere within a certain time frame usually ten to fifteen days, although courts generally provide leniency in this regard. Additionally, the driver can request a mitigation hearing, which acknowledges that the driver is guilty of a moving violation but is requesting a hearing with a judge to reduce the fines associated with the ticket. If the driver pleads guilty, the outcome is equivalent to a conviction after the hearing. Upon conviction, the driver is generally fined a monetary amount and, for moving violations, is additionally assessed a penalty under each state's point system. If a driver is convicted of a violation in a state other than the state in which the motorist is registered, information about the ticket is relayed in accord with state policy and agreements between the two states, including the Non-Resident Violator Compact. If the ticket information is not abstracted to the state in which the driver is licensed, then the record of the conviction remains local to the state where the violation took place. [1]

According to the data from National Highway Traffic Safety Administration (NHTSA), there were 1,018,634 traffic violations recorded in the past recent years. The traffic violations are divided into types like citation, ESERO (Electronic Safety Equipment Repair Order), SERO (Safety Equipment Repair Order) and Warning. Citation was recorded maximum in all of the traffic violations and was found maximum 106,953 the year 2014 which was followed by ESERO which has maximum record 13,919 in the year 2014 and SERO was recorded maximum 896 in the year 2012. SERO. Warning was recorded maximum 115,926 in the year 2015 which clearly states that warnings was considered as most recorded violation. [5]

According to the DMV (Department of Motor Vehicles) the count of the total number of traffic violations that happened in the whole 24 hours of a particular day with the help of Number of Records and the Time of Stop. It can be seen from the given visualization that the majority of the traffic violations are occurring around day time between 8:00 hours to 17:00 hours. The maximum traffic violations were recorded at 22:00 hours which is 76,521 and minimum traffic was found at 05:00 hours which is 9,412. [1]

The States like Maryland (MD) recorded the maximum traffic violations 893,983 followed by Florida (FL) with a count of 4,472 etc. and State like Wyoming (WY) recorded the minimum traffic violations 37 followed by South Dakota (SD) 36 etc. The State of Maryland (MD) is marked dark green in color for better understanding to indicate the maximum recorded traffic violations and describes the severity of traffic violations by state. [4]

The Driving Citation Statistics states that in the month of march the traffic violation was maximum which is 14,985 along with total number of violations which is 90,342 and the minimum is in July which is 251 along with number of violations which are 83,683. From this we can infer that alcohol has played major role in traffic violations and the government should make new strict rules for people driving car with alcohol consumption [6]. The year 2013 recorded the maximum fatal accidents which is 25.87 % along with total number of records which is 18.703%. The year 2015 recorded the minimum fatal accidents which is 5.20% along with total number of records 23.104%. The year 2016 recorded the negligence of fatal accidents. It can be observed that as the number of traffic violations are increasing per year, the number of fatal accidents is decreasing. If we consider the above facts and assumptions, we can say that the experts determine what went wrong and work out how to prevent these terrible violations from happening again. They examine the wrecks and official records, and hear from eyewitnesses, passengers and road safety experts as they reconstruct some of the most tragic disasters in road crash history.



If we compare all of the data which has been explained above states that the predicted values of the total number of traffic violations in the upcoming years. The minimum was in the year 2012 which had a count of 151,365 and the maximum was in the year 2015 which had 235,349 violations. But in future the count decreases. The forecast line is shown orange in color which is decreased to 188,430. The upcoming year 2017 will have less number of traffic violations compared to the previous years. With all these analysis, investigations and predictions hopefully better measures are taken to reduce traffic violations and avoid fatal accidents in the future.

**References:**

[1] "Using Vehicle Code to Determine Traffic Ticket Penalties & Points | DMV.ORG." DMV.ORG: The DMV Made Simple. <http://www.dmv.org/articles/using-vehicle-code-to-determine-traffic-ticket-penalties-and-points/>

[2] "Traffic Citation Law and Legal Definition | USLegal, Inc.." Legal Definitions Legal Terms Dictionary | USLegal, Inc.. Web. <http://definitions.uslegal.com/t/traffic-citation/> .

[3] "Types of Traffic Tickets - FindLaw." Traffic Laws - FindLaw. <http://traffic.findlaw.com/traffic-tickets/types-of-tickets.html> .

[4] "Driving Citation Statistics – Statistic Brain." Statistic Brain – Market Research, Rankings, Financials, Percentages <http://www.statisticbrain.com/driving-citation-statistics/>

[5] "Research & Data | NHTSA." *NHTSA | National Highway Traffic Safety Administration.*

Web. <http://www.nhtsa.gov/research-data> .

[6] "Traffic Tickets." *Traffic Ticket USA — Legal Help for Traffic Tickets.* <http://www.trafficticketusa.com/traffic-tickets/>.