Making of Individual Project

Vidyashree Venkatesh

2019-05-05

Contents

Chapter 1

Project Objective:

To explore the data set associated with the city of Chicago's Automated Speed Enforcement Program and to transform the data into interesting visualizations by exploring and analyzing the data. And to draw three interesting insights from the data to be presentable to the mayor of Chicago in the form visualization. Stages of project development.

Stage 1: Data Exploration

Stage 2: Extracting interesting aspects from the data

Stage 3: Improvising the stage 2 output with advanced features of Tableau.

Chapter 2

Summary of the data:

The Data set provides an accurate view into the Automated Speed Enforcement Program violations taking place in Children's Safety Zones. Each row in the data set represents speed light camera violation

Data Source: (dat, 2019b)

Chapter 3

Overview of the 'Automated Speed Enforcement' program:

It is a program undertaken by the city of Chicago to protect children and other pedestrians by speeding motor vehicles. The program uses enhanced safety measures such as Pedestrian refuge islands, Safety zone signage, and street stencils, High-visibility crosswalk markings, Speed feedback signs and many more. The automated speed enforcement cameras are one among the program's safety measure, these automated safety cameras to identify and ticket motorists who are breaking the law by exceeding the speed limits. The registered owner of the speeding vehicle will be issued the speeding ticket. (pro, 2019)

Chapter 4

Stage 1: Observations from data exploration

As part of the data exploration developed distinct visualizations in Tableau some of the key observations were that there existed a total of 5,159,847 violations captured through the speed cameras between 2014 to present (at the time of data collection). Although the violations are declining crosswise over time, the violations across month are fluctuating. The violations captured across the cameras in different locations and positions vary significantly, and certain areas have higher violations recorded consistently.

Learnings gained after this stage of the project by self-critiquing the visualizations.

- With the intended audience in mind, visualizations must be self-exploratory and straightforward and must involve less effort.

- Use of additional colors to be avoided.

- Use of complex visualizations to convey a simple message to be avoided.

- Beautification of the visualization with the attractive features of the tableau to be resisted.

Source: (dat, 2019a)

Chapter 5

# Stage 2:First version,development of insights from the data

The goal of this stage of the project is to draw appealing aspects from data to present the mayor of Chicago and also to incorporate the lessons learned in the data exploration stage. Firstly, it is essential to understand what sort of insight will the mayor be interested in or what ideas from the data be helpful for the mayor to take further actions on the automated speed enforcement program. Also, most importantly it is to be understood that the objective of such a program is not to capture more and more violations instead to enforce speed limit with various measures and bring in public awareness and reduce the mishaps from over speed. With this in consideration, the following questions pervade to explore data and are the focus of the analysis made.

- What impact has the speed enforcement program made on the violation activity over time since enforced? Does the event involve a trend or pattern?

- How efficient are the camera locations in capturing the violations? Is it something good or bad to learn that specific camera in particular areas consistently capture significant number of violations?

- In terms of violation breaches collectively at zip code level are there a sufficient number of cameras placed or are they more than necessary?

5.1   Finding 1:

The basis of this finding is to understand how the traffic safety with respect to speed violations has changed after the camera installation and how is the trend followed across months.
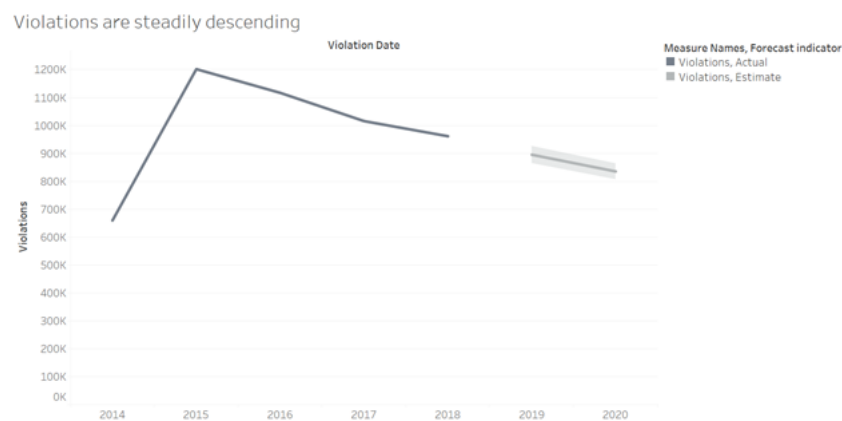
Associated Visualizations: Below visualization was the initial version of violations across years.

Violations across Years

The trend of sum of Violations for Violation Date Year.
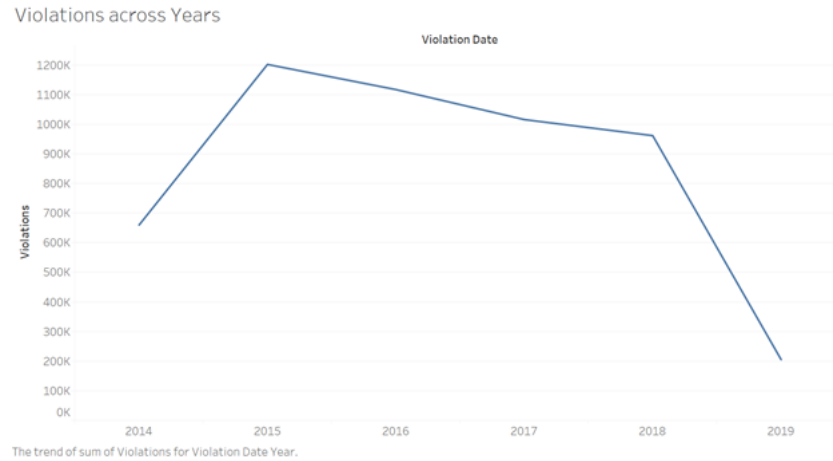
With the above visualization the trend line follows a steep decline from year 2018 to 2019, that's because the dataset does not consist of data for entire year of 2019. This trend would present a deceptive notion of drastic decline of violations.

The visualization was improved with the help of filters in tableau where years from 2014 to 2018 were selected and the forecast feature was incorporated to estimate the trend proceeding 2018.

Below is the improvised version of the violation trend. Line graph are more suitable to represent the trend lines hence the below graph is depicted in the line chart form.


Violations are steadily descending

Source:(tab, 2019)   The trend of Violations for Violation Date Year. Color shows details about Violations and Forecast indicator.

Further the analysis how the violations varied across months was observed from below visualization. A timeline chart is incorporated here to stretch the time axis and will let us zoom in and out to access more details from the overall trend.

Trend of Monthly Violations
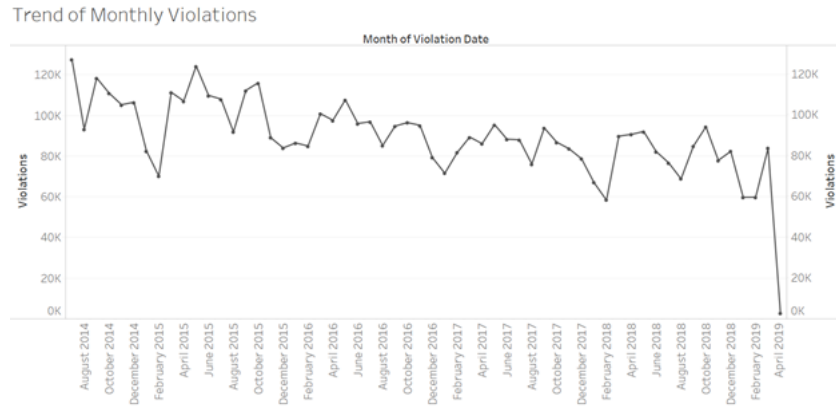
Source:(tab, 2019) The trends of sum of Violations and sum of Violations for Violation Date Month.

The decline in the trend is further investigated to learn that there exists an average of 20% drop in violations from 2015 to 2018 across month. In deducing this visualization data for the year 2014 and 2019 are not chosen as there is no data availability for a year long. The line graphs are chosen as they best depict the trends. And the difference in % is shown through a down ward column chart to show negative percentage.
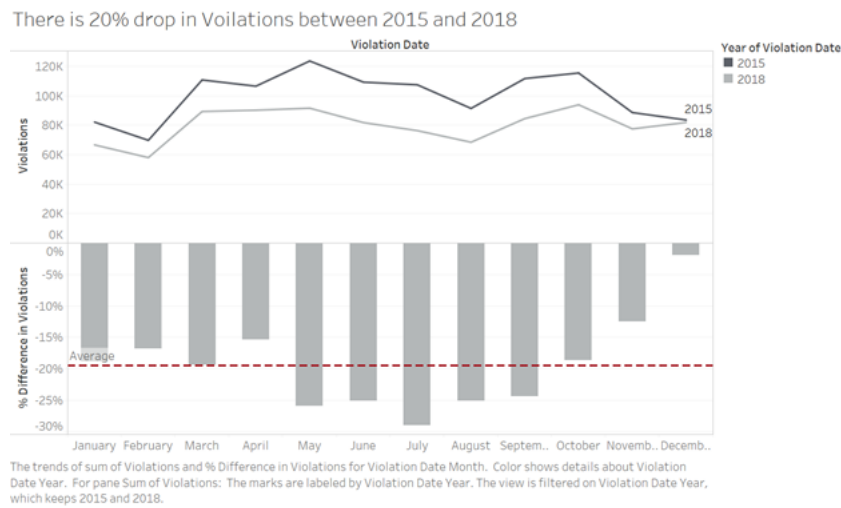
Source:(tab, 2019) The trends of sum of Violations and % Difference in Violations for Violation Date Month. Color shows details about Violation Date Year. For pane Sum of Violations: The marks are labeled by Violation Date Year. The view is filtered on Violation Date Year, which keeps 2015 and 2018.

Result:

It can be inferred that there is ~20% reduction in violations at safety zones after camera installation and the trend would continue to decline, this strongly conveys a message that the speed camera installation is driving awareness among public and helping the city of Chicago public be cautious and wise with the speed in safety zones.
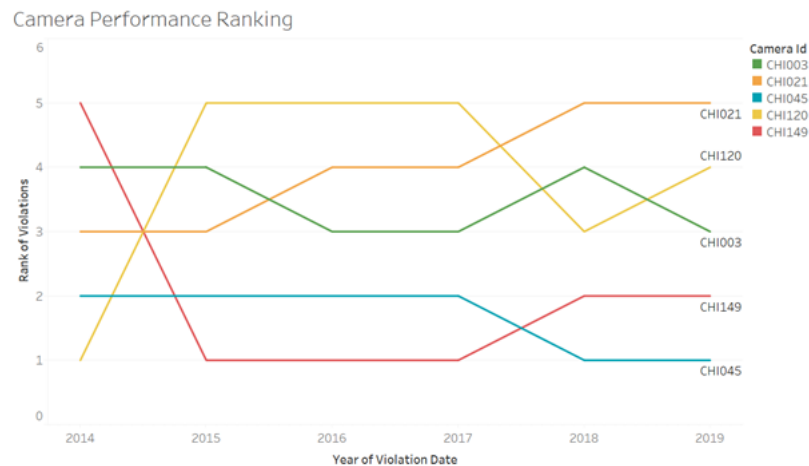
5.2 Finding 2:

Performance of a speed camera analyzed in terms of the total violations captured by them and made an attempt to learn the hidden insight from consistent top performing cameras.

Associated Visualizations: Firstly, to examine the number of violations captured by each camera below visualization was created as a histogram as it is the more suitable form to represent the distribution of violations across cameras. Here the camera ids are arranged in decreasing order of their captured violations. It can be observed that few of the cameras have outperformed the rest in terms of capturing the highest violations. To ease the identification of top performing cameras as set was created in tableau to group the top 5 cameras with respect to sum(violations) and a color code is applied to differentiate the rest.
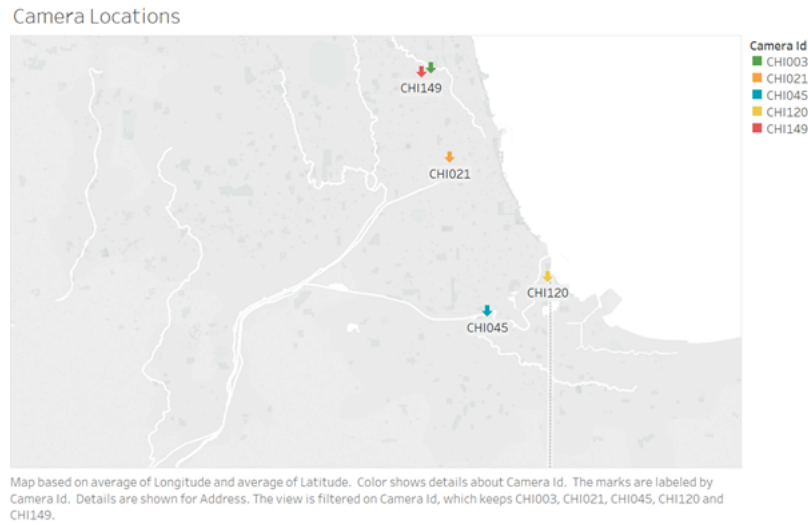
Source:(tab, 2019) Sum of Violations for each Camera Id. Color shows details about In / Out of Top Performing Cameras.
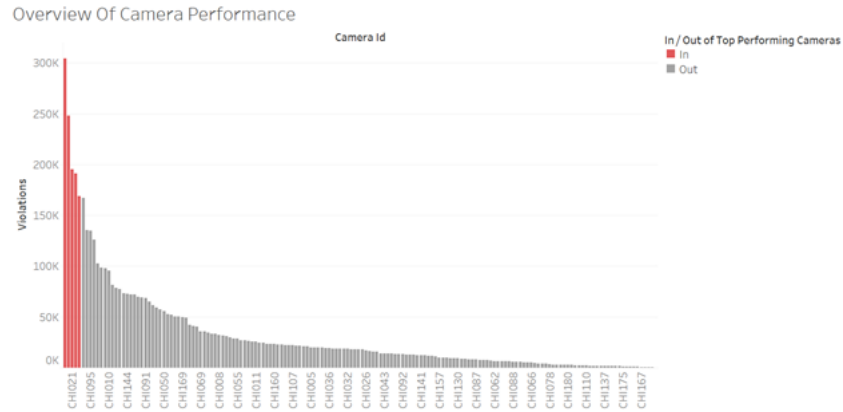
Although the visualization targets to gain the overview of cameras it also helps to identify the sensitive locations associated with those camera positions where there are high violations recorded. Further to learn more about how consistent are the cameras in capturing high violations across time, below visualization was created for the trend of rank violations associated to the top 5 cameras. On observation it can be inferred that the cameras with ids CHI045 and CHI149 have consistently been the cameras to capture highest violations across years.


Source:(tab, 2019) The trend of Rank of Violations for Violation Date Year. Color shows details about Camera Id. The marks are labeled by Camera Id. The view is filtered on Camera Id, which keeps CHI003, CHI021, CHI045, CHI120 and CHI149.

Another visualization is created to pin down the locations of the above cameras on the map.

Camera Locations

Map based on average of Longitude and average of Latitude. Color shows details about Camera Id. The marks are labeled by Camera Id. Details are shown for Address. The view is filtered on Camera Id, which keeps CHI003, CHI021, CHI045, CHI120 and CHI149.

Source:(tab, 2019)

It is essential to analyze the underlying aspects with the consistency of top performing cameras. The speed cameras were installed with purport to protect school children and pedestrians; the overall objective of the program is to utilize various safety measures to enforce the speed limit. Conflicting conclusions exist in theory several studies consider only speed reduction and many overlook the fundamental measure of safety, the violation rate from the finding one even though there is a significant decline in the overall rate of violations. Coming down to specific location or camera level the consistent position of cameras for higher violations captured raises a concern that there exist some flaws in the safety program which fails to impose more safety measures in those particular locations. And, it drives to conclude that the safety program fails to consider public safety as the primary concern. Below is an attempt to summarize above visualizations in a dashboard.
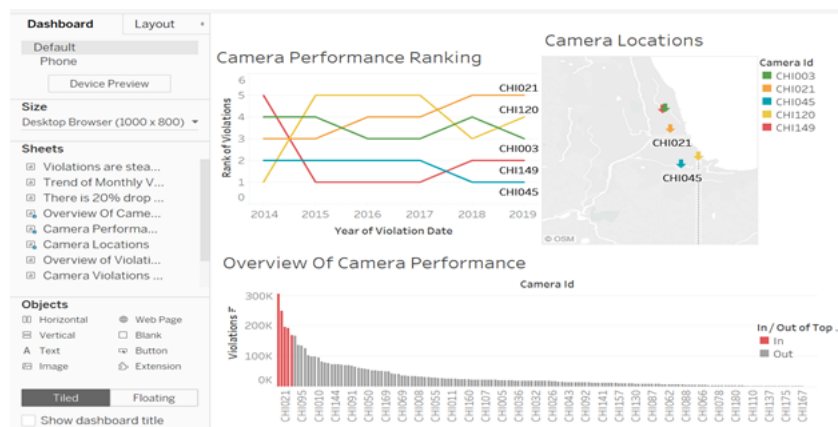
Source:(tab, 2019)

Result:

The fact to be conveyed to the Mayor is that there exist high sensitive locations in the city which have consistently recorded more violations across the years. It is crucial for the concerned authorities to investigate the underlying causes and improve safety measure like crossing guards and police officers around schools; and infrastructure improvements, such as signs, crosswalk markings and other traffic safety improvements in the affected locations. With this exercise, a significant number of violations can be reduced.

5.3   Finding 3:

There are a total of 162 cameras installed across the city of Chicago distributed in 40 different zip codes. 12 among those 162 cameras are newly installed in 2018.(cam, 2019) The program claims that Camera locations

are chosen based on available data regarding traffic, speeding, and crashes.

Another distribution strategy consistent with public safety would be to have the cameras distributed in areas (zip codes) known to have a higher than average violation rate. If speed camera deployment deviates from these distributions raises concern that public safety was undermined by speed camera program.

Associated Visualizations: The following visualization summed the number of cameras in each zip code and the number of violations from all cameras in each zip code through color concentration. The histogram form of bar chart was chosen to present the distribution of camera counts over zip codes. The inference from this visualization is that distribution of cameras is quite uneven, there are certain zip codes that have high number of cameras installed where the record of violations is lower than the zip codes with fewer cameras.
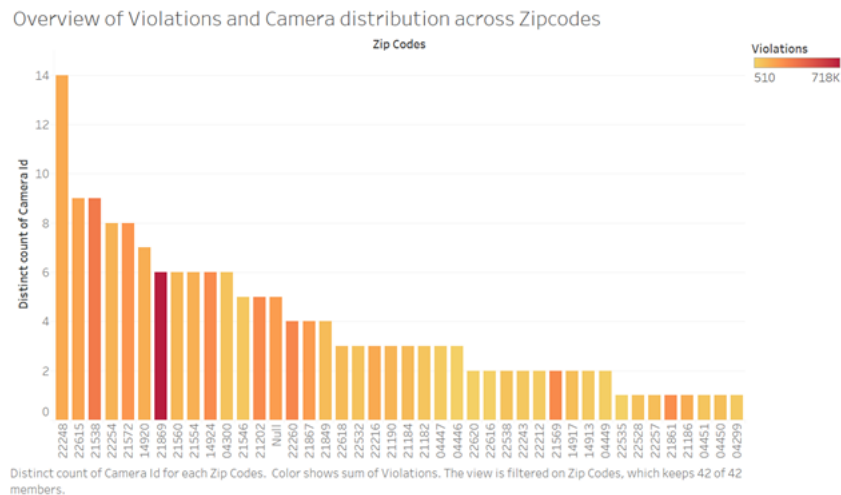
Source:(tab, 2019)

Further, to investigate the implication of violations recorded in bottom ten zip codes to the overall violations the below visualization was created with a tableau filter to group bottom ten zip codes to a set as 'low violations zip code' and a bar graph is chosen to present the comparison of two categories ,the percentage contribution of violations from in and out of the set. It can be noticed that ~25% of the zip codes in the analysis contribute deficient % of violations, i.e. 3.58%. This provides additional evidence to the claim that some camera installations are ineffective

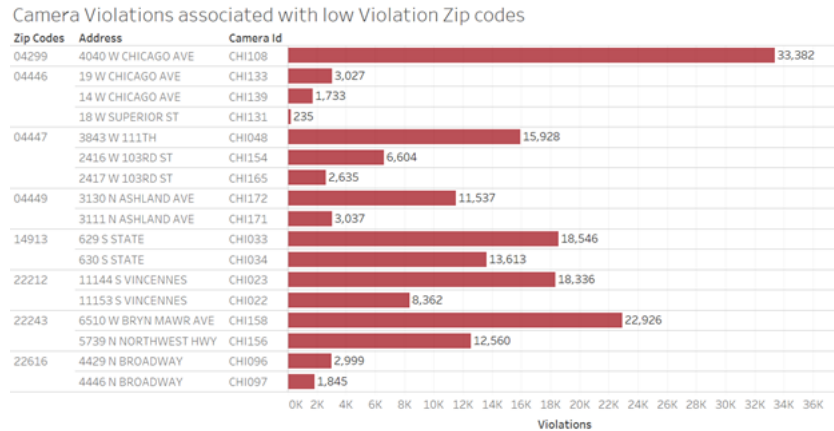% of Violations from low violation zipcodes and the rest
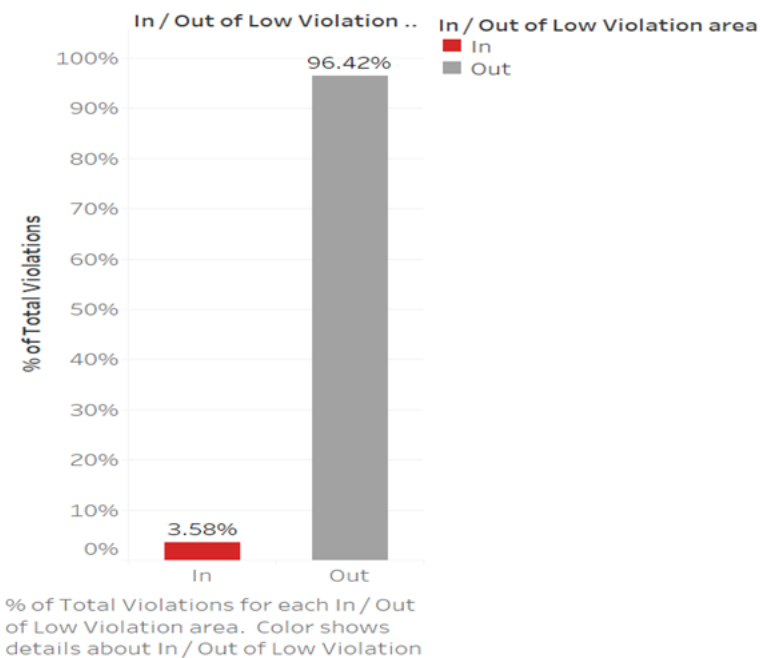
Source:(tab, 2019)

I then created a visualization to deduce some of the low-efficiency camera locations, more like a table format, using the same set of low violation zip codes, associated addresses, and camera ids. Here the 12 newly installed cameras in 2018 were excluded by the filter as they would naturally have low violations captured.

The same visualizations was improvised to a bar plot to make it visually more appealing and that it would help to obtain the viewers' attention to the variation in total violations to respective locations.

Source:(tab, 2019)

The exciting aspect here is that across the city there are about 1500 safety zones and legally only 20 % of the speed camera installation is allowed. Hence the identification of camera installation locations is a crucial aspect in this speed camera program. With the above visualization analysis, it can be inferred that there exists un even distribution of the cameras. The present low-efficiency camera locations can be safeguarded by other infrastructure improvements, such as signs, crosswalk markings, and other traffic safety improvements. And these cameras can be installed in different safety zones for upholding the goal of public safety.

Result:

The Camera distribution across zip codes are uneven in terms of their efficiency. It is required for the Mayor of the city of Chicago to insist on a resurvey to identify critical locations for camera installation.

5.4   Summary:

The lessons learned by data exploration stage, such as to avoid the use of too many colors and complex charts to convey a simple intent were incorporated to the best in this phase. The selection of chart type for the visualizations was made with careful consideration, such as a line chart to represent trends, bar graphs for comparisons and histograms for distribution. The key aspects extracted from this stage to present the Mayor of the city of Chicago are as follows 1. A decline of violation rate of ~20% from 2015 to 2018 signifies the automated speed enforcement program is impactful to some extent. 2. The consistent high volume of violations captured by cameras across years calls for further safety improvements. 3. The locations for camera installations require a new study by concerned authorities to identify critical safety zones and improve the efficiency of the program.

Road map for enhancements form stage2 to stage 3: This stage of the project will incorporate improvising the visualizations created in the previous phase and creation of meaningful dashboards by using advanced and interactive features of Tableau.

Chapter 6

Stage 3: Revised Version, implementation of improvisations.

The objective of this stage is to improvise the first version of the project based on the roadmap previously defined. Concerning the purpose, a sequence of steps was undertaken to make the visualizations better and improve the context of each visualization.

As a primary step, the first version was reviewed to identify typos and scope of improvements in renaming titles and legends. All the visualizations were revised to exclude typos and further, the charts were analyzed If they are in the context of their argumentation or if there exists a room for improvisation.

In terms of data wrangling in the first version, data were filtered on a worksheet, but as part of improvisation, the data has been filtered globally at the data source level particularly the data of the years 2014 and 2019 has been purposefully excluded from the analysis as they do not contain data for an entire year. Below is the documentation of revisions made with individual findings in an attempt to better present the insight.

6.1   Revision to finding 1 visualizations:

The intended insight to be conveyed here is that although the overall violations decline across years, across months show fluctuations, and there are few months like November and December and others that show less impact from the speed enforcement program hence require the authorities examine underlying causes.

Associated revised visualizations:

Below is one of the visualizations in finding1, 'Violations are steadily descending' this chart was improvised in terms of:

- Editing axis to give a more meaningful name to it.
- Removal of unnecessary gridlines.
- The graph line is faded by decreasing the opacity under the color legend leaving behind the trend line.
- Thus the graph is revised to portray the intended information more straightforwardly.
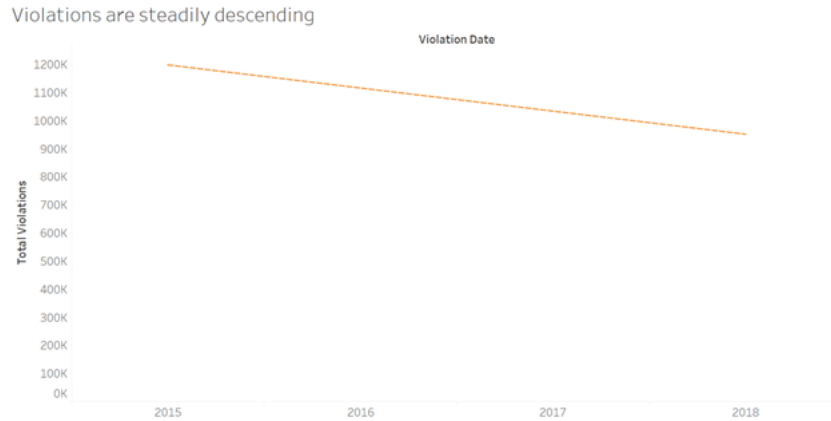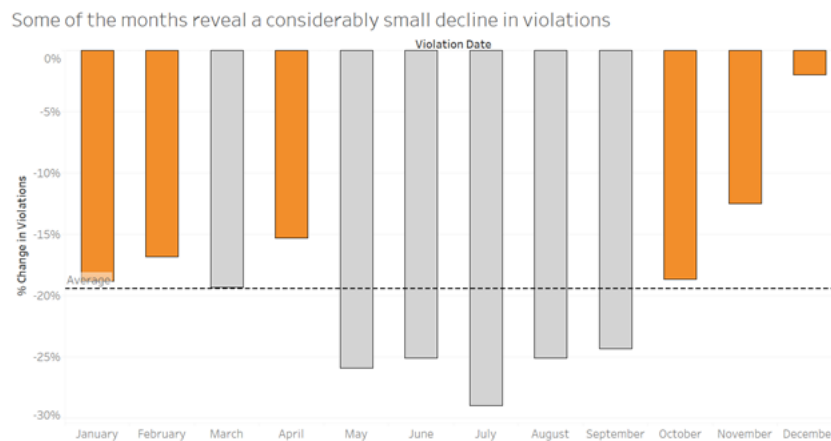
Source:(rev, 2019)

The next visualization 'Monthly violations trend' is the trend across months which has been improvised in terms of: * Renaming the y-axis. * An advanced feature of the tableau is employed to display the pattern across months in an animated fashion. * The name of the control panel is changed. * Control panel provides the viewer an opportunity to auto play the change in trend, in a way to make it more visually appealing.
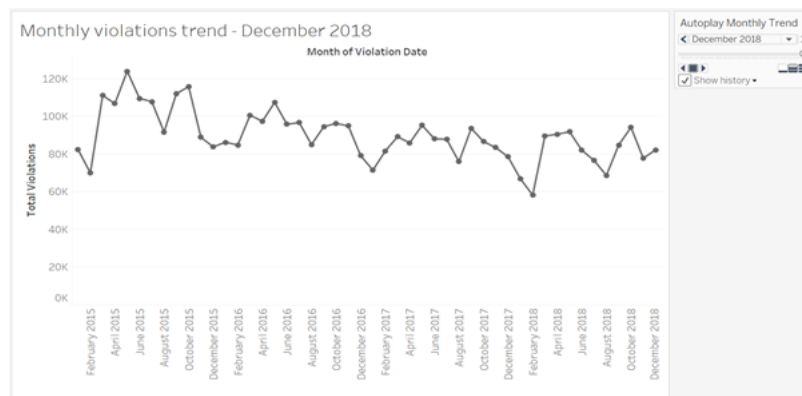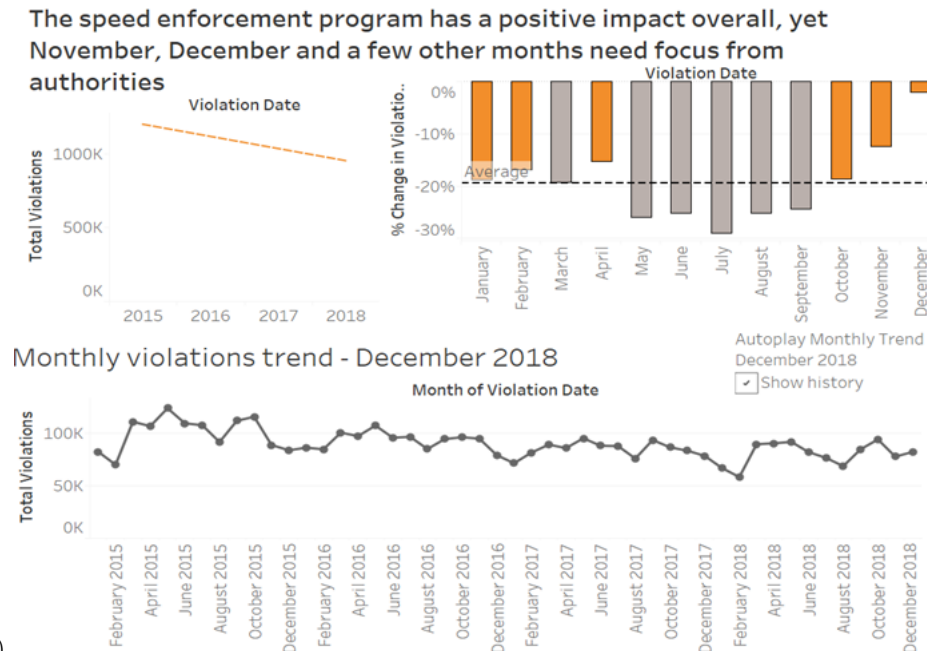

Source:(rev, 2019)

Another visualization part of the finding1 has been improved on the below lines. * The first version consisted of a dual axis chart which did not convey any extra information; hence it was excluded in the revised version. * The months are categorized by creating a set of impactful months and less impact months. * To seek attention towards the months which have shown lower percentage drop the color code has been enhanced to highlight the same. * The title was updated to convey a strong sense of the message.

Source:(rev, 2019)

The corresponding story and dashboards were revised to include all the improvised charts. Below is the dashboard for revised version.
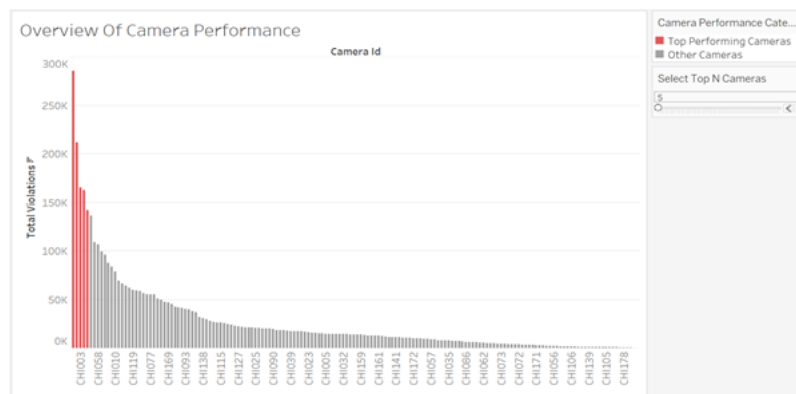
Source:(rev, 2019)

6.2 Revision of finding 2 visualization:

Here the insight being conveyed is that the locations of the cameras CHI045 and CHI149 with their consistent high violations captured raise concerns on the effectiveness of the speed enforcement program.

Associated revised visualizations:

Initially, the visualizations were designed to focus on top 5 performing cameras, in the revised version a parameter control is set to provide flexibility to the viewer to select top 'N' cameras. The parameter dynamically alters the camera performance set based on the value delivered through the control panel. The parameter control has a minimum of 5 and maximum value of 50 set to limit the range. Also, a geographical visualization is excluded in the revised version as it did not possess significant play in conveying the insight.

The first visualization under the finding 2 is 'Overview of Camera Performance'
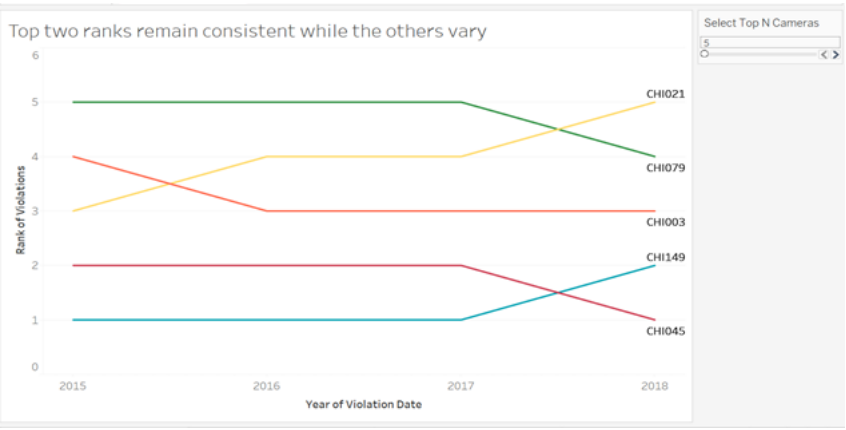

Source:(rev, 2019)

The above chart is revised in terms of * Renaming y-axis, filter aliases, and filter titles. * A parameter controls to dynamically set-top N cameras of focus is created.

Next visualization displays the ranking positions of top 'N' performing cameras. The control of deciding the number of cameras for analysis is provided to the viewer to enable the user to visualize the different ranking

pattern as the number of cameras is increased. The title has been edited to make it more appropriate.
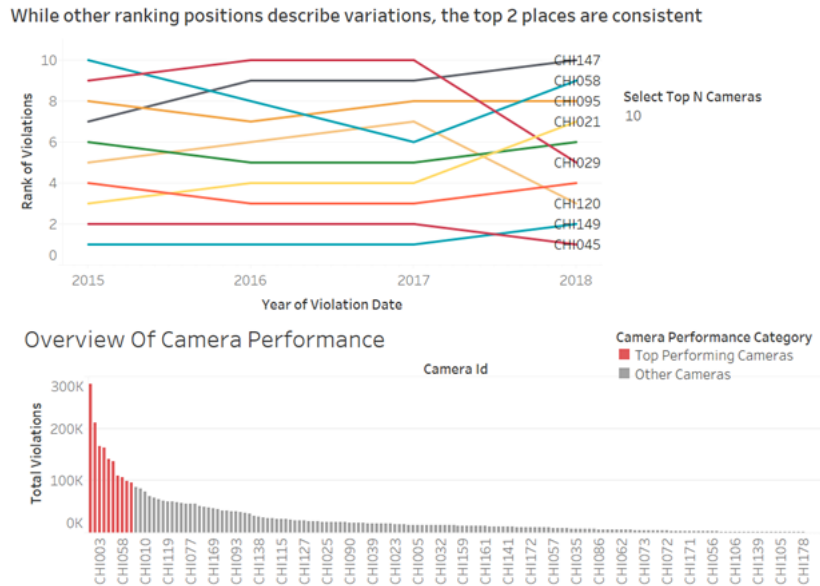

Source:(rev, 2019)

The corresponding story and dashboards were revised to include all the improvised charts. Below is the dashboard for revised version. One of the key feature included in this dashboard is that the parameter control is applied to both of the visualization which makes it more meaningful to show the number of cameras under focus as well as the corresponding rankings at the same time.

Dashboard for top 5 cameras:
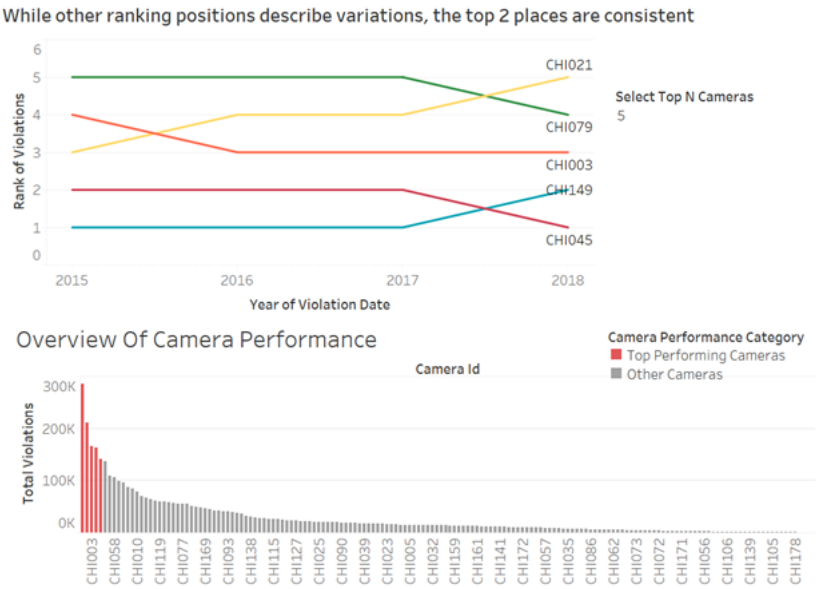
Source: (rev, 2019)

Dashboard for top 10 cameras:

Source(rev (2019))

Thus in this dashboard one of the interactive feature of tableau is incorporated to make the claim more evident.

6.3 Revision of finding 3 visualization:

This finding intends that to convey there is variation in performance of the program across zip codes with the influence of the number of cameras installed and importantly to claim to have more cameras installed restricts the violations to a greater extent. In the first version, the attempt was made to drill down to identify less impactful cameras and present them to be replaced. But the approach seemed to diverge away from the insight analysis. Hence few of the visualizations have been wholly revised to convey the argument in a better way.

Associated revised visualizations:

The first visualization here is to represent the existence of variation of violations with the impact of the total cameras across zip codes. The revisions made here are as follows. * Titles, color legends, and axis labels were changed appropriately. * The grid line is left intact as they make the readability of bar plots easier. * One of the data points on zip codes was null which has been excluded in the revised version.
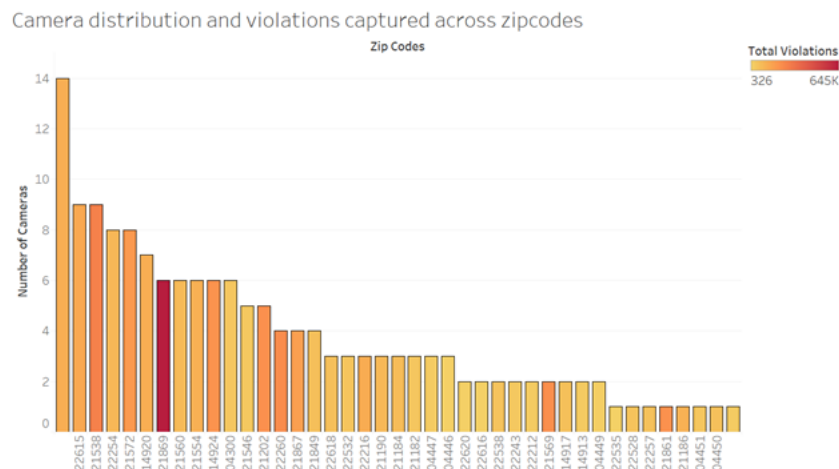
Source:(rev, 2019)

To further strengthen the impact of having more cameras at a location the following two visualizations were created as part of a revision to the previous version. Firstly, the zip codes were categorized into 2 sets under the set name 'Zipcode Category' where one set 'Camera Count>=5' contains the data of all the zip codes with more than or equal to 5 cameras (4 being the average number of cameras) and the other set 'Camera Count <5' contains the rest of the zip codes. The below visualization serves a basis of the analysis to provide the actual count of cameras that are being compared for their impact being grouped under different categories. Here the aliases and coloring conventions have been maintained in consistent to related visualization.
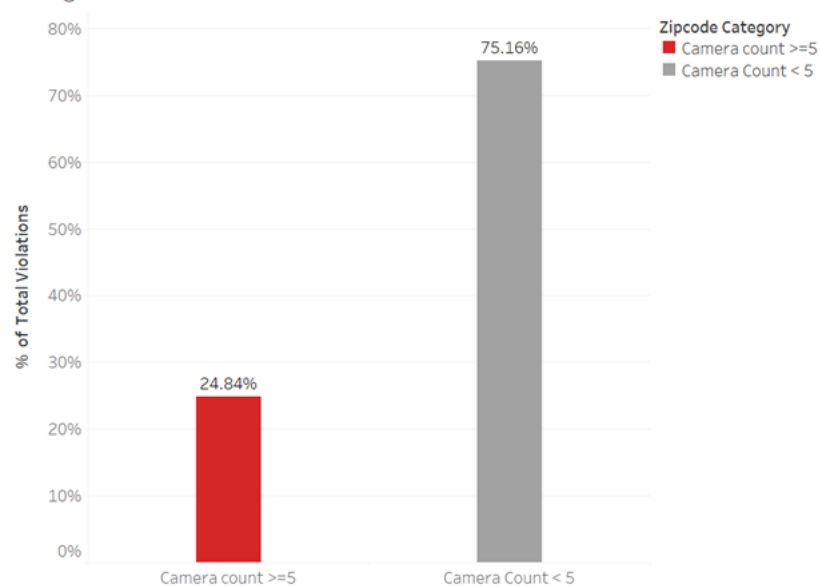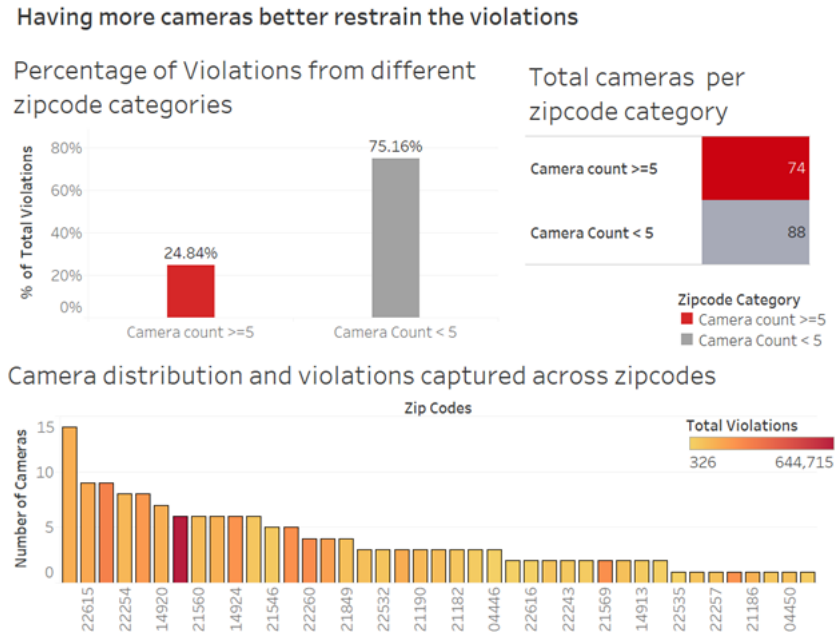

Source: (rev, 2019)

Next visualization, with the help of previous charts stands to claim the argument that zip codes with higher cameras restrict violations to greater extent. Here the title, color legend aliases and the axis labels are edited appropriately.

Source: (rev, 2019)

The corresponding story and dashboards include the revised charts. Below is the dashboard for the finding 3.

Having more cameras better restrain the violations

Percentage of Violations from different zipcode categories

Total cameras per zipcode category

Camera distribution and violations captured across zipcodes

Source: (rev, 2019)

6.4   Summary :

Without hampering the insights recognized in the first version of the project, an attempt has been made to improvise the visualizations in terms of appropriate titles, excluding typos, the addition of meaningful axis and control panel labels. Also, the visualizations from the first version were criticized in terms of their ability to present the argumentation, and necessary additions and deletions were incorporated. Some of the advanced features of tableau were employed such as autoplay (animation charts), parameter control. Overall the revision is made to powerfully convey the below three insights to the Mayor of Chicago. 1. The speed enforcement program has a positive impact overall, yet November, December and a few other months need focus from authorities. 2. The consistent high volume of violations captured by particular cameras across years calls for further safety improvements in those locations. 3. More installation of cameras, better control of violations.

Bibliography

(2019). Chicago automated speed enforcement camera analysis.

(2019). Children's safety zone program and automated speed enforcement.

(2019a). Individual project vidya2.

(2019). Revised version.

(2019b). Speed camera violations.

(2019). Version1.