Assessing Vision Zero Impact Through Data Science

Rafael Garcia Cano da Costa, Jirou Xu
Center For Data Science
New York University
New York, USA

Lucien Banks Wilson
Urban Interface
Kohn Pedersen Fox
New York, USA

***Abstract*—Vision Zero is a policy initiative to reduce deaths and serious injuries on traffic. Through data analysis, this study evaluates the impact of the program in New York City (NYC). The program's main success factors are driver's and pedestrian's education, but the adoption of actions on high density areas and increase of law enforcement would make the program achieve better results. The presented framework can be used to verify policies in other cities, and is optimizable with new datasets, softwares, algorithms, information, and granular data analysis.**

Keywords—analysis; enforcement; training; traffic; injuries

I. Introduction

Vision Zero was first introduced in Sweden in 1997 to bring the number of deaths and serious injuries on traffic to zero, and was later introduced in other countries, such as Canada and United Kingdom, as a multinational approach to traffic safety [1]. In NYC, the model's goal received a special attention from Mayor Bill De Blasio, who launched Vision Zero in 2014. Multiple entities adopted initiatives to minimize deaths and injuries on traffic in the city. For example, in partnership with the NYC's Taxi and Limousine Commission (TLC), registered drivers train every three years for license renewal. The training covers orientation on yielding in favor of pedestrians, and on respecting speed limits and red lights. Streets have also been redesigned to keep pedestrians safer, including features such as new crossing points, and waiting areas between curbs. Additionally, new laws reduced the citywide speed limit from thirty to twenty five miles per hour, strengthened sanctions for hit and run cases, and changed automobile design to minimize blind spots. This paper applies data analysis techniques to quantify the impact of Vision Zero, to indicate how current aspects of the program can be improved and to suggest new approaches that may enhance its effectiveness.

II. Related Work

Papers and reports [2] indicate that drivers, pedestrians, laws, and urban engineering are important players when it comes to traffic accidents [3]. One of the steps taken by the city in connection with Vision Zero was the establishment of a training program on traffic education for taxi and for-hire drivers. Another example is the implementation of workshops through which citizens share their concerns and suggestions to improve street safety [4]. Another instance is the enforcement of legislation like the October 2014 law that reduced the citywide speed limit from thirty to twenty fives miles per hour. In case someone is hit by a car at the new speed limit, the chance of death is fifty percent lower as compared to the previous speed limit [5]. Private and public sectors' participation is imperative to cause impact [6]. On the industry side, the TLC is spreading training on safe driving among professional drivers [7]. Those responsible for urban engineering are also contributing through street redesign, which leads to safer roads [5].

III. Methodology, Data, and Analysis

NYC makes useful datasets available to the public through the internet [8]. Once relevant datasets are identified, they should be further explored in order to assess if their content is helpful for the analysis being conducted and which features (columns) should be preserved during the data cleaning process. Deleting irrelevant data mitigates memory issues when manipulating large contents. Visually exploring the data through plotting may lead to useful insights. These data insights can be validated through hypothesis testing. Additionally, predictive techniques like machine learning can identify new factors that can amplify the impact of a study. Many times this approach requires finding a common factor among disparate datasets for integrating them in order to gauge the interaction and contribution of each feature towards the desired outcome of a study. Those steps can be summarized as data collection, data cleaning, data exploration, data engineering, and data analysis.

The datasets used in this paper can be divided into three categories. First, Taxi data, made available by the TLC, was used to evaluate if NYC's new speed limit had any impact on reducing traffic accidents (Citi Bike data supplemented some findings). Second, the New York Police Department (NYPD) Collision data provided records on traffic injuries and fatalities before and after the enacted program. Finally, Subway Entrance Locations, Bus Stop Locations, and Public Schools Locations were used to integrate the collision dataset to identify features contributing to cases of collision.

The initial approach was to narrow down the data available on the five boroughs to mitigate the computational cost of working on the whole data at once. The data on Manhattan was used first. After filtering the data on Manhattan using polygon query through Hadoop, the dataset was partitioned by year. The data from 2014 and 2015 had their month of September filtered for hypothesis testing. These steps were implemented on the Taxi, Citi Bike, and NYPD Collision datasets.

The data on collisions were preprocessed on Hadoop, which counted the number of collisions per month between 2011 and 2016. The expectation was to verify a new trend just after 2014. However, Fig. 1 showed change just in June, 2016. Enhancing the previous picture contribution, a heat map showing the density of occurrences could lead to new ways of mitigating the number of future cases through initiatives focusing on the higher density areas. The darker the area, the higher the density of cases. Fig. 2 has both graphs displaying the highest concentration in Midtown, and on specific intersections in lower Manhattan and in Harlem. This kind of map can point out to areas where new actions can be introduced by the city to lower even more the number of traffic collisions.

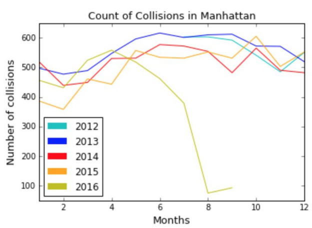

Fig. 1

The picture on the left depicts 2014 while the other one 2015

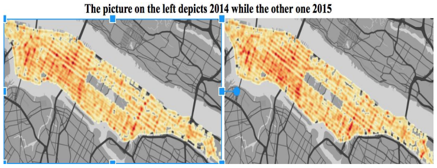

Fig. 2

The data on taxis and for-hire vehicles were used to evaluate the trend of traffic speed after the new law. The expectation was to see lower average speeds. The distribution of the data was plotted on September 2014 and 2015, and the speeds over seventy miles per hour were disconsidered. The descriptive statistics show a shift of the data points to the left of the x-axis when the median dropped from 10.10 to 9.84 on yellow cabs, and from 11.68 to 11.26 on green cabs. Also, the IQR (Interquartile Range) shows that fifty percent of the data is more concentrated since the distance between its extreme values went from 5.9 to 5.66 on yellow cabs, and from 5.35 to 5.16 on green cabs. Another good indicator is the skewness that went from 1.30 to 1.21 on yellow cabs, and from 1.41 to 1.37 on green cabs, meaning that the data is less spread to the right tail. Therefore, it is expected that a hypothesis testing the difference of means among the data points from 09-2014 and 09-2015 will indicate a drop on the average speed of taxi trips in Manhattan after the enactment of the law. The chosen approach was a paired *t* test for the mean difference in related populations. With level of significance of five percent the average speed on 09-2014 and 09-2015 are not the same, and in fact there was a probable decrease on the average speed in 09-2015, indicating the new speed limit effect. Moreover, the percentage of average speeds above twenty five miles per hour to those below or equal was 1.7 on September, 2014, and 1.3 on September, 2015 for yellow cabs. Respectively, 2.8 and 2.3 for green cabs. Fig. 10 shows the new trend after 2014, having 2016 with the lowest one. In addition, it is important to look at the fact that the city is growing which might increase the number of cars on the streets, and could lead to a drop on the average speed. To clarify this point, Fig. 3 has the count of trips in Manhattan. One interesting fact is the deep valley on February, 2014, caused by a severe winter with deadly snowing storms [9]. Also, the regular annual trend of decline around August seems related to the school vacation period in NYC [10]. Fig. 3 shows that the number of trips in Manhattan is decreasing for yellow and green cabs. This raises some questions because the decrease on the number of trips is not very compatible with the drop on the average speed. Clarifying, Fig. 4 was plotted using the for-hire dataset which includes trips from Lyft and Uber among others. The FHV (For Hire Vehicle) graph indicates an increase on the number of trips among the FHV category. Even though there were no data prior to 2015, the increase was significant from 3,000,000 to 11,000,000 trips. This can explain the drop on the number of trips on yellow and green cabs. As a result, there must be happening a shift of customers to the FHV service. Also, to corroborate this finding, a research done by FiveThirtyEight concluded on October 2015 that the number of regular cab and for-hire trips was stable over previous years. Hence, it seems reasonable to conclude that the average speed is decreasing over the past years [11]. However, to have more certainty, it is useful to know the number of licensed cars over years to see if it increased even though the total amount of pickups has been being stable. Also, the number of Citi Bike users is increasing, which may represent less customers for cabs.

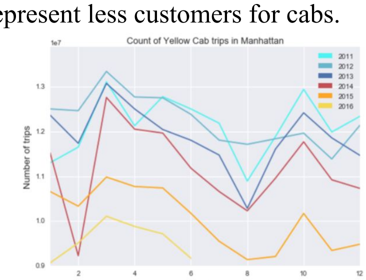

Fig. 3

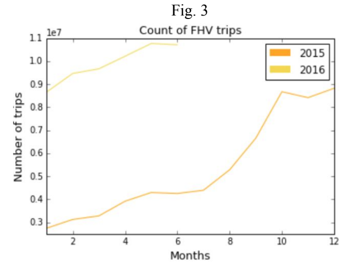

Fig. 4

For future actions, the higher density of overspeed (above 25mph) among yellow cabs on trips shorter than or equal to 0.1 of the mile happens in Midtown. This information could be exploited by choosing areas where devices can be installed to identify drivers incurring in a wrong behaviour which could lead to a later punishment. Another aspect is the possibility to allocate NYPD personal to the higher density areas for law enforcement. Also, it raises questions on the relationship between the higher number of collisions and overspeed in Midtown, because Midtown is an area with high density of people walking around. Among the people, tourists account for a big group that can be vulnerable to traffic accidents for being distracted while crossing roads.

The trend on the number of collisions before and after a specific street is redesigned was picked as a way to test Vision Zero's effectiveness. Among some redesigns, the number of collisions per year was plotted, and the average was computed to measure their impact. The first analysed was completed on February 2014 comprising Bowery, Kenmare, Delancey, and Spring Streets. The central islands on Bowery close to Kenmare were redesigned allowing easier turns to Spring to improve pedestrian safety. The available collision data on the intersections from 07-2012 to 09-2016 was plotted on Fig. 5. In 2014 there was a decrease on the number of collisions as opposed to 2015. It is possible that people were getting used to the new changes, and being more cautious when driving or crossing the street in 2014. In sequence, people already familiar with the redesign in 2015 repeated numbers from 2013. Finally in 2016, there was a huge drop on that number. This was an interesting fact because it matched the trend observed on the first collision line plot. Therefore, something is driving the positive impact of Vision Zero with more power than just the decreasing on the average speed or the street redesign. The second redesign analysed was completed on January 2014 on Broadway and 96th St. The main changes are two left turn that were banned, new phasing of signals, and expanded pedestrian space. The available collision data on the intersections from 01-2013 to 09-2016 are on Fig. 6. Contrasting with the previous redesign, the modifications on Broadway and 96th seem to have less impact for no impact in 2014 against 2013. Maybe, drivers were not aware of changes, and could speedup before the light turned green for expecting a faster phasing as before, or they could turn left when not supposed to do so. However, there was progress in 2015 as opposed to the previous redesign. Maybe, the drop in 2015 was due to familiarity with the new redesign. Also, in 2016 the drop was significant, and it is worth mentioning that there was no occurrence on the dataset from March to September, 2016.

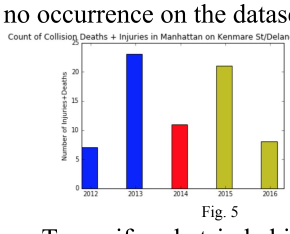

Fig. 5

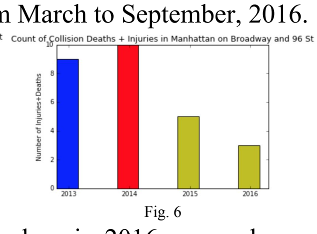

Fig. 6

To verify what is behind the drop in 2016, research was done on the annual reports released by the NYC on Vision Zero. The reports mention training on defensive driving and yielding in favor of pedestrians as a requirement to license renewal. Maybe, most of the drivers would have to renew their licenses by the end of 2015 or beginning of 2016. The number of drivers receiving this kind of training over the years should be verified as a next step. Moreover, many initiatives raising drivers' and pedestrians' awareness to traffic safety through training took place in 2015. The city reached out to people in schools as a way to educate them on how to behave on streets. The TLC awarded and recognized drivers with outstanding driving practices, and issued 283% more summonses for overspeeding in 2015 than in 2014 [7]. Another possible explanation to this surprising change in 2016 could be that there are plenty of street redesign at the beginning of 2016 when compared to 2014. In conclusion, Vision Zero is comprised by different action plans that work in an integrated manner. Each action plan is making a contribution, but it seems that education of drivers and pedestrians through training and awareness adds a substantial improvement to the effectiveness of the program.

Looking for new aspects correlated with collisions to improve Vision Zero, a dataset was engineered spatially joining collisions data and features like bus stop location. If a bus stop were placed up to twenty five meters from a collision point, the new feature would present one or zero otherwise. The same was done for subway entrance locations and wi-fi hotspot locations. After that, Decision Tree and Random Forest models were applied to extract the feature importance for predicting the occurrence or not of any death (target variable) on each collision. Fig. 7 shows that the presence of a bus stop (5) is highly correlated with deaths. Also, subway entrance presence (6) has some influence as well. This calls for actions around bus stop sites. Trying to optimize the model, a new dataset was engineered. The presence of public schools (14) was introduced on Fig. 8. Another change was in the distance between the place where the collision took place and the other features. If the bus stop (11), subway entrance (12), and wifi spot (13) were up to fifty meters from a collision point, the new feature would present one or zero otherwise. For the schools, the range was a hundred meters for schools being big buildings with latitude and longitude centered in the middle of the construction. The wifi hotspot importance is interesting, although these hotspots became available in the middle of 2015. They are located in areas with high density of people like Midtown [12], and these areas need actions from the program.

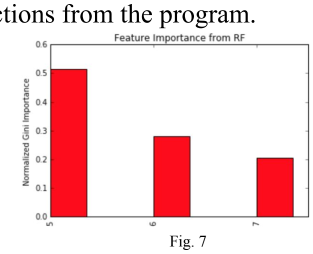

Fig. 7

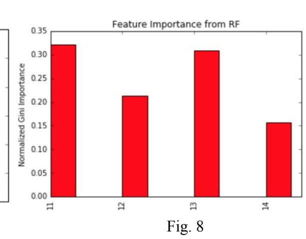

Fig. 8

IV. Conclusions

An important aspect noticed is that Vision Zero is transforming the lives of pedestrians and drivers in NYC, successfully decreasing the number of traffic collisions. Our research found three key conclusions. First, the action plan resorting to speed law enactment and enforcement is being effective. The great outcome is going to be the fifty percent higher chance of survival if someone is hit by a car at the new speed when compared to the old one. Continued law enforcement and education may achieve even lower rates of death on traffic. Second, the street redesign analysis gave a mix of information, but it made possible to understand that a key factor is the people's behaviour. Apparently, the best way to correct bad behaviour is through education. NYC already has an education program in place, but the more widespread the program, the better results for Vision Zero. Third, the correlation discovery signaled that the city should create policies towards high density areas like Midtown and that bus stops should receive special attention from drivers. One way to mitigate accidents close to these locations would be through the enactment of severe laws for drivers involved in cases close to bus stops or a set of supposed high density areas. More details are on an appendix called *long_report.pdf.*

V. Discussion and Future Work

This work has some limitations. Most of them are linked to available datasets, chosen algorithms, how datasets were split, collected information, area of coverage, and software. Additional datasets could strengthen this paper, such as datasets on drivers who went through training for driving safer, on the number of cabs and for-hire cars in NYC, on deployment of traffic officers, or on distribution of tourists and pedestrians from Flickr in Manhattan. The algorithm used for calculating the distance between collision sites and other features was based on *Haversine*, but a Google Map API could be more precise leading to a better modeling. Another detail is the way some data were split. For instance, the data analysed under the street redesign approach had the street changed in February 2014, but the data on 2014 was combined while it should be split between before and after that month. Further, redesigns not analysed due to time constraints need the same study for comparing goal and effectiveness of each intervention strategy leading to more consistent findings. In addition, data on the average speed should be split into different slots of time such as day or night and seasons which would provide more insights. Calls to the DOT (Department of Transportation) to gather further information were not successful. The idea was to collect new data like the frequency of repaving the streets, which could be used as a new feature on building a dataset since drivers can get distracted by avoiding pavement defects on streets and hit a person or another car as a consequence. However, the DOT did not have information other than what was available online. This work should be developed on the datasets of the other NYC boroughs to derive broader insights about the effectiveness of the program. On the software side, future work will involve developing tools for visually exploring the analysis in addition to mapping it as it was done with heat maps. Also, the models used to discover correlations are not resistant to imbalanced datasets, which do not have a target variable around fifty percent of the whole dataset. This problem could be minimized with ArcGis, a spatial analytical tool which applies Moran's I Index for feature correlation. Moreover, a cluster analysis should be done to reinforce the current insights or discover new ones. Despite the referenced limitations, this work can help NYC evaluate the Vision Zero program and find ways to improve it. Moreover, other cities can follow this paper as a guideline for verifying the level of success of their implemented policies.

Acknowledgment

The team behind this work thanks Professor Claudio T. Silva for all effort in bringing to class an outstanding team of advisors and lecturers that made available resourceful information to the success of this study. Also, a special thank to the direct advisor, Lucien Banks Wilson for all invaluable insights, guidance, and contribution. Finally, a thank to Fabio Miranda who gave access to his heat map visualizer.

References

1. "Vision Zero." *Wikipedia*. Wikimedia Foundation, n.d. Web. 07 Dec. 2016.
2. "Library." *Library - Vision Zero*. NYC.gov, n.d. Web. 07 Dec. 2016.
3. "Vision Zero - An Ethical Approach to Safety and Mobility." *Vision Zero - An Ethical Approach to Safety and Mobility - Accident Research Centre*. N.p., n.d. Web. 07 Dec. 2016. 4. Pavelski, Joel. "Deadly Winter Storm Hits Northeast: The Latest." *New York Post*. N.p., 13 Feb. 2014. Web. 07 Dec. 2016. 5. *Vision Zero 1 Year Report*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016
6. *Vision Zero Action Plan*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016.
7. *Vision Zero 2 Year Report*. Rep. NYC.gov, n.d. Web. 07 Dec.2016.
8. "NYC Open Data." *NYC Open Data*. NYC.gov, n.d. Web. 07 Dec.2016.
9. *Pedestrian Safety Action Plan Manhattan*. Rep. NYC.gov, n.d.Web.07Dec.2016.
10. "Holidays and Key Dates." *Holidays and Key Dates*. New York City Department of Education, 2016. Web. 07 Dec. 2016.
11. "Uber Is Taking Millions Of Manhattan Rides Away From Taxis." *Uber Is Taking Millions Of Manhattan Rides Away From Taxis*. Ed. Reuben Fischer-Baum and Carl Bialik. FiveThirtyEight, 08 Mar. 2016. Web. 07 Dec. 2016. 12. Intersection. "CityBridge | LinkNYC." *CityBridge | LinkNYC*. N.p., n.d. Web. 11 Dec. 2016.