

Analysis of the San Fransico TOLE Program

Michael Arango, Yeshwant Chillakuru, Jackson Crum

March 6, 2017

Background on Dataset

General information and Data Collection

The Transit-Only Lane Enforcement (TOLE) is run by the San Francisco Municipal Transportation Agency (SFMTA) from March 2008 to February 2015 as a pilot program to reduce obstruction of transit-only lanes through increased enforcement. Currently, private and commercial vehicles that drive and park in transit-only lanes can cause delays in bus service, increase traffic jams, and inconveniencing customers.

Under the TOLE program, all public buses are equipped with video cameras to capture the license plates of vehicles parked or stopped in a dedicated transit-lane. The vehicle's registered owner will then receive the citation in the mail in aims to deter future transit-only lane violations. These citations and their details form the dataset we are examining. The data were obtained from the US government's open data site at <https://catalog.data.gov/dataset/muni-driver-reported-transit-only-lane-violations>.

San Francisco currently operates 26 miles transit-only lanes with 160,000 daily passengers and planned future expansion of 22 miles, so the TOLE program has strong potential for long-term impact by increasing the efficiency of the municipal transit system and potentially increasing municipal revenue temporarily as drivers adjust to the enforcement scheme.

Source: https://www.sfmta.com/sites/default/files/projects/2015/TOLE%20one%20pager_v3.pdf

```
print("Cite the above source properly")
```

```
## [1] "Cite the above source properly"
```

Background on dataset

The unit of observation was an individual citation for a transit-only lane violation reported by municipal drivers under the TOLE program, and the raw dataset contains the following variables regarding the transit-only lane violations:

- **Object.ID** – Unique key that corresponds to each observation
- **Ticket.Number** – Unique number that corresponds to each ticket cataloged with the Superior Court of San Francisco
- **Citation.Issue.Date** – The date (MM/DD/YY) that the traffic only lane violation citation was issued
- **Citation.Issue.Month** – The month (MM) that the traffic only lane violation citation was issued; This column was completely blank when downloaded
- **Citation.Issue.Time** – The time of day (HH:MM) PST that the traffic only lane violation citation was issued
- **Location** – The street address where the traffic only lane violation citation was issued
- **Violation.Code** – A code that corresponds to the type of traffic only lane violation that was committed to warrant issuing a citation
- **Violations** – Description of the traffic only lane violation issued in the citation
- **Fine.Amount** – The dollar amount of to be paid for the traffic only lane violation
- **Citation.Status** – Status of the citation (Open, Closed, Unapplied)
- **Amount.Paid** – The dollar amount paid for the traffic only lane violation
- **Amount.Due** – Outstanding balance for the traffic only lane violation

- **Suspend.Code** – Code that represents reason for license suspension (Suspended until the driver goes to driving school, pays fine, too many points on license, etc.)
- **Suspend.Process.Date** – Date the license suspension goes into effect
- **Suspend.Until.Date** – Date the license suspension is over
- **Disposition.Code** – The city of San Francisco does not release disposition codes
- **Last.Edited.Date** – Date that represents the last date the corresponding observation was edited; This column was completely blank when downloaded
- **Geom** – The latitude and longitude (latitude, longitude) of the traffic only lane violation

Limitations of Dataset and Data Cleaning

The limitations of the data include the timing of the implementation of the video camera citation system. While SFMTA claims 100% of their buses are equipped with cameras, they fail to indicate if the 100% implementation was at the start or end of the program. We believe there was a trial period at the beginning, so we decided to drop variables in 2008 when the program first started because there were only 154 citations issued that year, but typically several thousand citations issued in the following years. In addition our dataset lacked demographic controls for the regions where citations were issued. This means demographic factors could cause unobserved biases in our data. Similarly to the domestic controls, the dataset lacks observations of citations issued before the program was put in place, which prevents a before and after analysis. While out of the scope of this initial analysis, future analysis can map to the **Geom** variable in the dataset to census tracts to derive demographic information and could be used to compare to other citations.

Some observations were missing key data. We decided to drop all observations with no **Geom** data, so that future geospatial and demographic analysis could be done (12 observations). We dropped all observations where **Citation.Status** was not specified or “unapplied”, as we determined “unapplied” was out of the scope of this project (84 observations). We dropped **Violations** where the listed violation was determined to be “No Violation” (1 observation) or occurred less than or equal to 70 times (77 observations). We also dropped citations in the year 2015 to avoid biases in monthly and yearly analysis because only 2 months were included for 2015.

Overall, we dropped 374 of the original 17,178 observations in the dataset. This only accounts for 2.18 percent of the dataset. Thus, we are confident that we did not introduce any inherent biases to the data by dropping the aforementioned observations.

Background Prior External Analysis Related to Content of Dataset

Most of the analysis compared TOLE data to traffic flow rates. According to SFMTA, certain corridors experienced reduction in delays after TOLE implementation. Westbound Geary Street corridor saw a 3% daily and 7% afternoon reduction in delays. Westbound Stutter Street saw a 15% daily and 20% afternoon reduction in delays. However, this analysis lacks significant meaning because it cherry picks certain areas and also lacks a control comparison for other areas that may not have been in the TOLE program. Additionally, their analysis has some clear flaws as they state that of the 1700 different vehicles received a citation from March to August of 2012, and only 2% received a citation during the “same three-month period in 2013.” Again this analysis lacks a control comparison and suffers a clear inconsistency as March to August is a six-month period, not three month as SFMTA stated.

According to Michael Rhodes, San Francisco’s transit-only lane violations are generally unenforced. While SFMTA had not released citation data prior to Rhodes’ article in 2009, he cites anecdotal evidence from interviews with municipal drivers and San Francisco Police Department. Based on the municipality’s weak analysis and the anecdotal evidence provided by Rhodes of the need for an effective transit-only lane violation enforcement scheme, a further analysis of municipality data must be conducted. While we cannot analyze a before and after comparison of the TOLE program on transit-only lane violations or traffic flow improvement, we can examine the distribution of violations by violation type, time of day, day of the week, and revenue potential.

Source1: https://www.sfmta.com/sites/default/files/projects/2015/TOLE%20one%20pager_v3.pdf Source2: <http://sf.streetsblog.org/2009/08/11/violations-in-sfs-transit-only-lanes-rampant-and-rarely-enforced/>

```
print("Cite the above sources properly")
```

```
## [1] "Cite the above sources properly"
```

Broad Questions

The TOLE program was implemented in 2008 as a pilot set to expire in 2012, unless extended upon an evaluation in 2011. In 2011, the City of San Francisco decided to vote to implement the TOLE program permanently. Therefore, understanding the effects and dimensions of the program are vital to constantly improving it. We believed in an exploratory analysis that looking at revenue generation would be in ideal means of understanding this program. Revenue optimization both increases funding for the state and creates economic incentive for drivers' to avoid citations. Under properly placed incentives and fine structures, revenue optimization can lead to ideal traffic outcomes.

Broad Question: *Under what parameters does the TOLE program optimize TOLE program revenues?*

We started with this broad question before choosing a full SMART (Specific, Measurable, Answerable, Relevant, Time bound) question to allow EDA to inform the development of the SMART question.

Source: http://www.leginfo.ca.gov/pub/11-12/bill/asm/ab_1001-1050/ab_1041_cfa_20110429_113836_asm_floor.html

```
print("Cite the above sources properly")
```

```
## [1] "Cite the above sources properly"
```

- Research on program from "Description_of_TOLE_Program.pdf"
- Introduce 3 "Broad Questions"

EDA & SMART Question

- Show EDA
- Show refinement into SMART Questions

Answering the SMART Question

-

NOTES:

-

TO-DO:

- CITATIONS!!!!!!!!!!!!!!!!!!!!
- add summary stats/dimensions to the background of the dataset
- Bill introduced in 2007, reevaluated in 2011, made permanent in 2015