

Traffic Volume Final Report - Group 9

Samantha Feinberg, Samuel Gullion, Ashlynn Sayegh, Alizeh Shaikh, Thomas Veenstra, Emery Whitford, and Eddie Winstanley

Introduction

The objective of this project is to analyze and derive insights from the dataset containing information about traffic detectors deployed at signalized intersections in the City of Austin, Texas. This dataset, maintained by the Arterial Management Division of the City of Austin Transportation & Public Works Department, holds the key to understanding traffic dynamics, congestion patterns, and transportation management within the city. Signalized intersections are critical points in urban transportation systems, where traffic signals regulate the flow of vehicles and pedestrians. The deployment of traffic detectors at these intersections enables the collection of valuable data regarding traffic volume, vehicle speeds, and other relevant parameters. This data can be harnessed to enhance traffic management, optimize signal timings, and address congestion issues. The City of Austin, being a rapidly growing urban center, faces the challenges of evolving traffic patterns, population growth, and urban development. As such, the effective utilization of data from traffic detectors becomes essential for making informed decisions to improve the efficiency and safety of the transportation network. There are a few main motivations we wanted to focus on, traffic optimization, urban planning, safety improvements, resource allocation, and environmental impact. Traffic Optimization includes understanding the data from traffic detectors and how it can lead to more efficient traffic signal timings, reducing congestion, and improving overall traffic flow. This dataset also provides insights into the impact of urban planning and development on traffic patterns. It can aid in planning infrastructure improvements and optimizing transportation systems to accommodate growth. Safety Improvements can be analyzed within this traffic data to identify intersections with higher accident rates, prompting targeted safety measures and infrastructure enhancements. The data can also inform decisions related to the allocation of resources for transportation management, helping the city make strategic investments in improving its infrastructure. We can understand the environmental impact by optimizing traffic flow, the city can potentially reduce emissions and environmental impact associated with traffic congestion.

Data

As previously mentioned, our data set was gathered by the City of Austin Department of Transportation. This dataset contains 5,788 observations on 16 variables related to traffic data in Austin at various key intersections. The 16 variables are as follows: `detector_id` shows the unique ID# for each traffic detector, `detector_type` shows the type of detector such as loop, video or other, `detector_status` shows the status of the detector and whether it was last known to be working or not, `detector_direction` shows the traffic flow direction the detector measures like southbound or northbound, `detector_movement` shows which type of movement the vehicle did at the intersection such as a left turn or right turn, `location_name` shows the name of the intersection the data was taken from, `atd_location_id` shows the unique ID# the city of Austin uses for that intersection, `signal_id` shows the unique ID# used to denote the traffic signal, `created_date` shows the date the data was created, `modified_date` shows the most recent date that the traffic detector data was updated, and `LOCATION` shows the ordered pair of the latitude and longitude of the intersection. The other six variables in the dataset are not described in the primer for the data, so we trimmed them out as the first part of our data cleaning. There were also many columns that had large amounts of missing data so N/A's were considered in the data analysis.

Exploratory Analysis

The exploratory analysis of the Austin Traffic Detector dataset, encompassing 5,788 rows and 16 columns, was an endeavor in meticulous data dissection to garner foundational insights.

Leveraging Python's Matplotlib and Seaborn libraries, we embarked on an extensive visualization journey. These visual tools were pivotal in elucidating key aspects of the dataset, such as detector types, operational statuses, and traffic directions. Each plot, be it a bar graph or scatter plot, served as a window into the dataset's underlying patterns and trends.

Our analysis gave rise to four hypotheses, each sculpted to probe distinct aspects of the traffic detector data. The first hypothesis examined the relationship between detector types (Loop or Gridsmart) and their operational statuses (OK, Broken, Removed, Inactive). This hypothesis was rooted in the assumption that Gridsmart detectors, owing to their advanced technology, might demonstrate superior operational stability. To test this, a Chi-squared test, as shown in the line ``chi2, p, dof, expected = sm.stats.chi2_contingency(contingency_table)``, was used to discern if

there was a significant association between these categorical variables.

The second hypothesis focused on whether the distribution of detector types varied with traffic directions. This was again approached using a Chi-squared test, similar to the first hypothesis, underlining the statistical method's utility in examining associations between categorical variables. The third hypothesis proposed a correlation between the direction of traffic detectors and the frequency of traffic incidents. This was particularly intriguing as it combined categorical and count data. We applied Poisson regression, as indicated by ``poisson_model = smf.glm("incident_count ~ detector_direction", data=Traffic_Counts, family=sm.families.Poisson()).fit()``, to model this relationship.

The fourth hypothesis ventured into exploring the impact of detector locations on the frequency of recorded signal IDs. Here, we turned to linear regression analysis, a method adept at modeling relationships involving continuous data. The Python code ``model = sm.OLS(y, X).fit()`` succinctly encapsulates this approach, linking detector locations (predictor) with signal IDs (response variable).

Modeling

In the modeling phase, our strategy was to meticulously select and apply statistical methods and algorithms that were most suited to the nature of our data and the specificities of our hypotheses. The choice of Chi-squared tests for the first two hypotheses was motivated by the need to explore associations between categorical variables (detector types and operational statuses, and detector types and traffic directions). The Chi-squared test, implemented via ``sm.stats.chi2_contingency(contingency_table)``, served as a robust tool to ascertain the significance of these associations, shedding light on potential dependencies between different categories of detectors and their operational characteristics.

For the third hypothesis, involving the relationship between detector direction and incident frequency, Poisson regression was employed. This choice was informed by the count nature of the incident data. The Poisson regression model, as instantiated in the line ``poisson_model = smf.glm("incident_count ~ detector_direction", data=Traffic_Counts, family=sm.families.Poisson()).fit()``, was adept at modeling this count data, enabling us to explore the potential influence of detector direction on incident occurrences.

The fourth hypothesis, focusing on the relationship between detector location and signal IDs, necessitated a shift to linear regression due to the continuous nature of the signal ID data. The model `sm.OLS(y, X).fit()` was pivotal in quantifying the extent to which detector location could be a predictor of the frequency of signal IDs. This linear regression approach was particularly apt for hypothesizing about and interpreting continuous data relationships.

In each case, the data was partitioned into training and test sets, a standard practice in statistical modeling that ensures the robustness and validity of the models. This partitioning allowed for rigorous model validation and assessment of predictive performance, ensuring that our findings were not only statistically sound but also practically relevant. The careful selection of these statistical methods and regression models, tailored to each specific hypothesis and data type, epitomized our commitment to a nuanced and comprehensive analytical approach.

Discussion

Visualization 1

The objective evaluation of the traffic detector data, visualized through a bar chart, used incident count as the evaluation metric, indicating that detectors in 'OK' status are associated with the highest number of incidents, potentially due to their placement in high-traffic areas. In contrast, 'INACTIVE' and 'REMOVED' detectors also show notable incident figures, hinting at the possibility that non-functioning detectors negatively impact traffic flow and safety. 'BROKEN' detectors correlate with fewer incidents, possibly reflecting less traffic or effective maintenance protocols. The interpretation of these results is constrained by potential limitations, such as data accuracy, unaccounted external variables, and the lack of causative conclusions from the observed correlations. These challenges highlight the need for a more nuanced analysis, taking into account time-series, to fully understand the detectors' performance and their relationship with traffic incidents.

Visualization 2

The second model examines the distribution of different types of traffic detectors across various directions, employing a grouped bar chart as a visualization tool. The evaluation metric here is the count of detectors, segmented by type and direction, which provides a direct measure of

detector prevalence. The visualization suggests a variance in detector type distribution across directions, which could indicate a strategic placement based on traffic flow characteristics or technological requirements of the detectors. However, this model's interpretation comes with limitations. The chart does not reveal why certain detector types are more prevalent in certain directions; it only shows the existence of such a trend. External factors influencing these placements, such as urban planning decisions or historical traffic data, are not considered, which might be crucial for a comprehensive understanding.

Visualization 3

The visualization presents an analysis of the average incident count by traffic detector direction, using mean values as the evaluation metric. Interpreting these findings suggests that detector direction could have a relationship with the incidence of traffic events. In other words, directions with higher averages might be critical areas that require more attention from traffic management authorities. However, there are limitations in the model and the visualization: a warning indicates that two rows with missing values were removed, which could affect the overall accuracy of the average incident counts. Moreover, the model assumes that the average is a robust measure of central tendency, which may not account for outliers or skewed distributions. To draw more comprehensive conclusions, it might be necessary to explore the data further, possibly including other statistical measures like median or mode, and to consider other potential contributing factors to the incidents.

Visualization 4

The visualization titled "Frequency of Signal ID by Location" is a bar chart representing the occurrence of signal IDs across different locations. The bars are colored by location, and the y-axis represents the frequency of each signal ID as a factor, which implies a count of occurrences. From this visualization, we can interpret that some locations have a higher frequency of signal IDs, which might indicate areas with more traffic detection or possibly areas with more traffic flow or incidents. However, the x-axis text has been intentionally blanked out, making it difficult to identify which specific locations correspond to the frequencies shown. The decision to remove axis text and legend could be a limitation if the chart is to be used for detailed analysis or presentation, as it lacks clear labels that would allow for easy identification of

specific locations. To improve this model and visualization, it might be beneficial to include axis labels for clarity and to consider whether treating signal IDs as factors is the most effective approach for the intended analysis.

Ethics

Analyzing the ethical implications of traffic data in Austin, Texas, we can apply three of the 16 ethical framework pieces: respect for privacy, fairness and equity, and social responsibility.

Respect for Privacy

This project involves collecting and analyzing data from traffic detectors, which raises questions about privacy. While traffic data is generally considered non-personal, there's a growing concern over the extent of data collection and its potential for misuse. Ensuring that the data remains anonymous and cannot be traced back to individual drivers is crucial. Ethically, the project must adhere to privacy laws and guidelines, ensuring that the data is used solely for the stated objectives of traffic management, urban planning, and safety improvements. This involves implementing robust data governance policies that define how data is collected, stored, and used, ensuring that the privacy of citizens is not inadvertently compromised.

Fairness and Equity

The project's goal to optimize traffic flow and enhance safety should be equitably beneficial to all city residents, regardless of their location or socioeconomic status. There's a risk that such data-driven initiatives might favor certain areas over others, potentially leading to unequal resource allocation. For instance, prioritizing traffic improvements in wealthier or more developed parts of the city can inadvertently neglect the needs of less affluent neighborhoods. Ethically, it's essential to analyze and use the traffic data in a manner that promotes equitable benefits across the city. This means considering the impact of traffic management decisions on all communities, ensuring that the enhancements in traffic flow, safety measures, and environmental impacts are distributed fairly across different areas of the city.

Social Responsibility

The project holds significant potential for positive societal impact, particularly in enhancing

public safety and reducing environmental harm. Ethically, it's crucial to ensure that the insights derived from the traffic data are used responsibly to serve the broader public interest. This includes making data-driven decisions that improve traffic safety, reduce congestion, and minimize the environmental footprint of urban transportation. The ethical obligation extends to transparent communication with the public about how their data is being used and the benefits it brings. Additionally, there should be a continuous assessment of the societal impacts of the project, ensuring that the outcomes align with the broader goals of public welfare and sustainable urban development.

In summary, while the project presents valuable opportunities for improving urban transportation in Austin, it's imperative to navigate the ethical dimensions of privacy, fairness, and social responsibility to ensure that the project contributes positively and equitably to the community.

Conclusion

After extensive analysis of different traffic detectors and their usage, our group was able to come to various conclusions on the usage of such detectors. Firstly we explored the relationship between detector status and the number of incidents recorded by that detector. It was revealed that detectors labeled 'OK' clearly had the highest count of incidents, though 'BROKEN' or 'REMOVED' detectors did have noticeable counts as well. Further research could be done by contacting the department of transportation to see exactly how they define these terms in order to make inferences on the relationship between status and count. Our second goal was to examine the relationship between detector direction and type. We found that for mostly every direction, both 'LOOP' and 'SENSYS' detector types were the most prevalent. Additional research could be done into how these types of sensors operate to see why they are the most prevalent for most directions, and to see why they are not prevalent at all for certain directions. Our third objective was to look at the traffic flow direction as it pertained to the number of incidents. As could be assumed, each of the four cardinal directions had relatively the same number of incidents, showing an even distribution of traffic flow in all directions. Oddly enough, the direction with the highest number of incidents was 'UNKNOWN', so further research and data collection could be done to try and determine what direction these unknown directions were, though they would likely be spread evenly amongst the four cardinal directions. Lastly we wanted to look at the signal ID distribution across locations. We found that signal ID frequency had a large variation

across location, though with so many different locations included in the data set, it was impossible to label them all on the visualization. Further work could be done by splitting this visualization into many different sub-visualizations so that locations can be displayed in a readable fashion in order for more conclusions to be drawn.

Acknowledgment

NAME	% CONTRIBUTION
Samantha Feinberg	100
Samuel Gullion	100
Ashlynn Sayegh	100
Alizeh Shaikh	100
Thomas Veenstra	100
EmeryWhitford	100
Eddie Winstanley	100

Bibliography

“Advantages of Traffic Signals.” *City of Irvine*, 13 June 2015, www.cityofirvine.org/signal-operations-maintenance/advantages-traffic-signals#:~:text=By%20alternately%20assigning%20right%20of,move%20safely%20through%20an%20intersection.

Shaw, Jeffery. “Signalized Intersections Informational Guide - 2nd Edition.” *Federal Highway Administration*, US Department of Transportation , July 2020, safety.fhwa.dot.gov/intersection/signal/fhwas13027.pdf.

“Arterial Management.” *AustinTexas.Gov*, City of Austin, 20 Nov. 2023, www.austintexas.gov/department/arterial-management.

“Urban Design.” *AustinTexas.Gov*, City of Austin, www.austintexas.gov/department/urban-design.