Predicting Crime Rate in Cleveland Census Tracts Using Permit Data

Eric Geisler

John Carroll University

December 6th, 2024

Executive Summary

The history of the city of Cleveland is complex—where there has been urban advancement, there has been redlining, economic progress has come alongside widespread gentrification, and cultural growth has been underlined by the prevalence of poverty and crime. While local government has put a more intensive focus in recent years on investing into areas with high rates of poverty and crime, the pervasiveness of crime in the city of Cleveland still remains a massive problem—one that costs the city millions of dollars per year. Thus, having a method of predicting the effect of infrastructure investment on crime rate in certain areas could allow policymakers to better understand the relationship between cost of investment and cost of crime, and therefore devise more efficient investment strategies. After constructing a long short-term memory model and inputting permit data to predict crime rate per 1,000 residents in Cleveland census tracts (geographic subdivisions of a county that facilitate data collection), the ability to predict the effect of infrastructure investment on crime rate varies on a regional and temporal basis. When compared to a baseline of predicting the crime rate per 1,000 residents in census tracts with the overall mean crime rate per 1,000 residents of the previous year, model predictions were 17% more accurate overall and 48% more accurate for 2024 data exclusively. Predictions greatly varied between census tracts; therefore, the model should only be employed on select census tracts with accurate predictions. Policymakers should redesign the investment strategy for census tracts with high spending on investment and a predicted increase in the cost of crime. Policymakers should additionally redesign the investment strategy for census tracts where the predicted savings in the cost of crime do not justify the amount being invested into infrastructure.

Problem Background

A drive throughout the city of Cleveland and its suburbs is a whirlwind experience. The transitions between grandiose neighborhoods with upscale housing and beautiful landscaping and declining areas with abandoned buildings and deteriorating infrastructure are seamless. While pockets of high-end development foster a surface level impression of splendor, the truth remains: Cleveland is a dangerous city—one of the most dangerous cities in the entire United States. According to the 2024 crime report on United States medium-sized cities, Cleveland ranks top three in rates of murder, rape, aggravated assault, and robbery (Gabriele).

This crime is expensive. In 2023, crime cost the city of Cleveland $7,397 per capita, resulting in a total expense of $2,675,317 (Boldizar). This societal cost of crime was devised through methodology brought forth by research from the University of Miami, which involves calculating the cost of factors such as economic loss of any victims, criminal justice system processes, policing, and the opportunity cost of committing specific crimes (McCollister et al.). Recently, more investment has been funneled into the city through the "15-minute city" initiative, where access to all major amenities is reasonable and easy, in order to alleviate poverty and crime (Luscher). Increasing investment into the infrastructure of target areas can help the city save money through long-term, recurring savings from decreased crime rates. This initiative has put a spotlight on the lackluster infrastructure in and around the city, and while being a good starting point for change, the plan has primarily focused on transportation infrastructure thus far and not infrastructure as a whole (City of Cleveland).

The goal of this project is to fill this gap. By providing an analysis of how general infrastructure investment will affect the crime rates in given census tracts moving forward, this increased emphasis on infrastructure improvements can be targeted more effectively as the plan shifts away from exclusively transportation. Specifically, the ability to predict crime rates given infrastructure investment is presented as a long-term solution to analyze the cost of investment infrastructure in relation to the cost of crime. By optimizing investment into infrastructure improvements in problematic areas, the city can save money going forward that otherwise would have been spent fighting high crime rates and undertaking inefficient projects. This would cultivate a safer city with a better overall quality of life for its residents, and allow the city to reinvest these savings back into the community. Obviously, while the aforementioned focus on infrastructure investment has come through the City of Cleveland specifically, there will need to be collaboration between city officials, county officials, and the private sector.

Related Work

Existing literature can help provide a background on the modern understanding of the relationship between investment into infrastructure and crime rate. In a broad sense, previous research suggests that increased investment into the infrastructure of a given area reduces the crime rate of said area in some capacity. The Catalonia region of Spain experienced a significant uptick in investment in the late 2000s through the FEIL initiative. This government-led funding campaign resulted in a decrease in economically motivated crimes in the region (Montolio). Within the United States, investment campaigns in the city of Pittsburgh also fostered a decrease in criminal activity. From the 1990s into the 2000s, broad investment ($468 million) into neighborhoods around the city of Pittsburgh resulted in a decrease in the violent crime rate of these areas (Baird et al.). The effects of these initiatives in Spain and the United States on crime, however, were not translated into economic savings.

Additionally, similar research has been conducted regarding the influence of specific types of investment on crime. Between 2011 and 2015, the city of Portland planted thousands of trees in various neighborhoods in an effort to increase green infrastructure around the city. This investment led to a decrease in violent crime in these neighborhoods (Burley). In San Diego, an increase in affordable housing, facilitated through easier access to home mortgage money, was associated with a lower overall crime rate (Bunting). Across cities in the United States, alteration of abandoned buildings and parking lots has an indirect relationship with firearm violence (Branas et al.). Lastly, in certain areas of China, investment into educational spaces resulted in a significant lowering of the regional crime rate (Yin et al.).

Finally, existing research has explored the predictability of crime. In the cities of Philadelphia and Seattle, machine learning was utilized to predict criminal instances based on building and road layouts of specific urban areas (He and Zheng). More research exists that attempts to predict crime in certain areas, however this is primarily based on predicting specific crimes using past crime data.

While previous research delves into the relationship between investment and crime, and even explores methodology behind predicting crime, there remains a gap in the literature. Firstly, no relevant research on the topic has been conducted within the city of Cleveland specifically. Additionally, previous works examine this relationship mainly by analyzing investment through strictly monetary values. No research looks at investment through street-level building and demolition permit data. As exemplified above, some research discusses the relationship between specific types of investment and crime, but no works explore the nuances between types of investment. Lastly, no existing literature has used permits and investment types to predict crime.

Overview

Again, the ability to predict the effects of infrastructure investment on crime rates will allow for a more efficient investment strategy in the city of Cleveland. Generally, this problem will be addressed using machine learning through a long short-term memory (LSTM) model. Data on building permits and demolition permits from the city of Cleveland will be inputted to the model to predict the crime rate in a given area. The time span in question is 2016-2024. Data will be grouped by census tract and by month. Thus, the final dataset will consist of building and demolition permit counts for each census tract per month dating back to 2016. Records will be evaluated in six-month sliding windows in order to analyze trends over time. The model will predict a given tract's crime rate per 1,000 residents per window, which will be examined using accuracy measures to determine the viability of implementing the model into local policy. Census tracts were chosen as the geographic unit for analysis in order to find a middle ground between large neighborhoods and individual census blocks. As the typical census tract holds roughly a few thousand people, these units will facilitate capturing specific regional trends while also providing a broad enough area to have meaningful implementation. Accuracy measures will be converted to economic terms to facilitate model application.

Data

All data used for this research comes from the Cleveland Open Data Portal, organized by the Office of Urban Analytics and Innovation. Specifically, five datasets were used: Crime Incidents, Building Permits, Demolition Permits, Census Tract 2010 to 2020 Net Change, and City of Cleveland Census. The Crime Incidents dataset has about 700,000 rows, organizing reported crimes in the greater Cleveland area from 2015 to 2024 (some earlier data was filtered out). The dataset contains information about date, location, and type of offenses. The Building Permits dataset has about 170,000 rows. This dataset contains information about issued building permits in the Cleveland area during this same time period, such as date, location, permit type, value, and use type. Similarly, the Demolition Permits dataset captures the date, location, permit type, and value of demolition permits in the Cleveland area during this span. The dataset contains slightly less than 10,000 rows. The Census Tract 2010 to 2020 Net Change dataset tells the population change of each census tract from 2010 to 2020, covering 178 tracts. Lastly, the City of Cleveland Census dataset was used to get the 2020 populations for each census tract in the Cleveland area. With the Cleveland Open Data Portal being recently organized by the Office of Urban Analytics and Innovation, the datasets were all clean and ready for analysis.

Methods

Starting with the outcome variable, the column labeled "UCRdesc" in the Crime Incidents dataset, giving the offense type, was transformed into a new, bucketed column "CrimeType". This column classified all crimes as "Violent", "Nonviolent", or "Vice". From the "OffenseDate" column, a new column "YearMonth" was created giving the year and month a crime was committed. Then, the data was grouped by census tract ("DW_Tract2020") and "YearMonth", and the values of "CrimeType" were summed. Thus, the new dataset featured rows that provided the counts of total, violent, nonviolent, and vice crimes for each month where data was available in each census tract.

Next, the population datasets were joined based on census tract. The population change from 2010 to 2020 ("Total Population Change") was divided by the number of years in the time period to get a new "ChangePerYear" column. This column was divided by months in a year and added/subtracted from the year 2020. A new dataset was created with the (estimated) population for each month from 2015 to 2024 per census tract. The year 2015 was later filtered out of the final data due to an abundance of missing values. The new dataset was then joined with the Crime Incidents dataset to form the dataset “CrimeStatistics”. Each population was joined with the proper census tract and month. Crime totals were divided by census tract populations and multiplied by 1,000 to get a crime rate per 1,000 residents statistic for total crimes and each crime type.

A similar process was used to manipulate the Building Permits and Demolition Permits datasets. Within Building Permits, the column “USE_GROUP_1” was transformed into a new “Use” column which categorized the permit uses (business, recreation, etc.). This was additionally done for the “PERMIT_CATEGORY” column, which was transformed into “Permit Category” with values “New”, “Maintenance”, and “Alteration” delineating the permit type. Both permit types and permit uses were counted across months to get a new dataset of these counts grouped by census tract and month. Additionally, the job value of permits was summed with the column “TotalJobValue” in the new dataset. For Demolition Permits, the values of the column “PERMIT_SUBTYPE” (“Residential” or “Commercial”) were counted and put into a new dataset grouped by census tract and month. Again, the job value of permits was summed with the column “TotalDemolitionJobValue” in the new dataset.

Finally, all datasets were merged together. This dataset contained the count of building permit types, count of building permit uses, sum of building permit job value, count of demolition permit types, sum of demolition permit job value, and aforementioned crime statistics for each census tract broken down by year and month. Total counts for new building permits and new demolition permits in general were also included. This dataset was then further grouped into sliding six-month windows (ie. January through June, February through July, etc.) to get counts and sums for all features (and averages for the crime statistics) across these spans. This was the final dataset.

All variables were analyzed, and twelve total features were chosen as inputs for the model. These are as follows: total building permits ("TotalPermits"), total building permit job value ("TotalJobValue"), total housing permits ("Housing_Permits"), total business permits ("Businesses_Permits"), total institutional care permits, including hospitals and retirement homes ("Institutional Care_Permits"), total food permits, including grocery stores and restaurants ("Food_Permits"), total recreation permits, including parks and gyms ("Recreation_Permits"), total education permits ("Education_Permits"), total hazard permits ("Hazard_Permits"), total other permits ("Other_Permits"), total demolition permits ("TotalDemolitionPermits"), and total demolition permit job value ("TotalDemolitionJobValue"). The total crime rate per 1,000 residents in a given census tract ("Total_per_1000") was chosen as the dependent variable for prediction.

Before constructing a model, a correlation analysis was conducted to examine features on a cursory level. For this analysis, six census tracts were studied—the two lowest populated census tracts, the two highest populated census tracts, and two randomly chosen census tracts—as the relationship between features and crime is likely to vary on a tract-by-tract basis. In order to account for temporal effects, features were compared with crime statistics from 12 months later. Firstly, the correlation matrices of census tract 39035111700 (Figure 1) and census tract 39035114501 (Figure 2), the two lowest populated census tracts, are shown below:

Figure 1

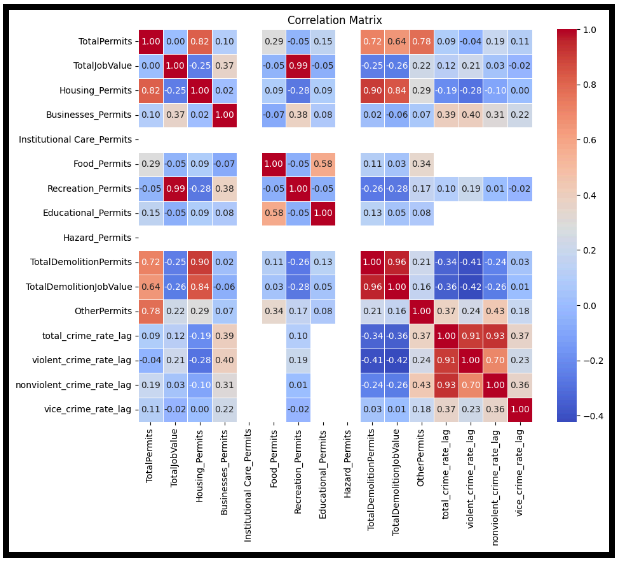
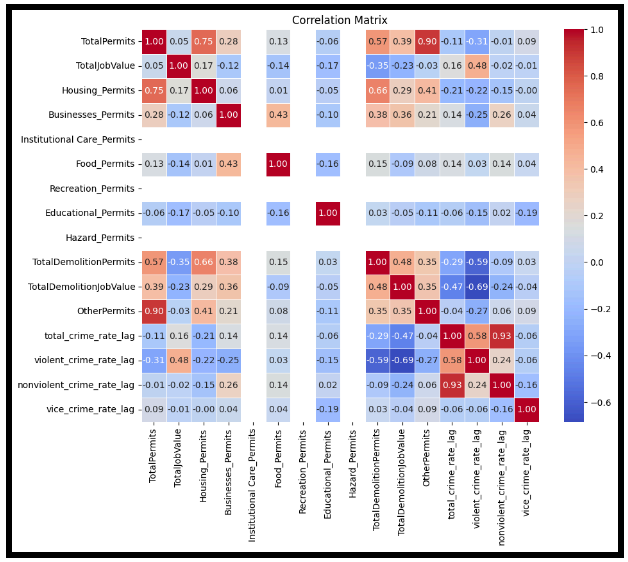


Figure 2



Based on these correlation matrices, there is already evidence that the relationship between permit data and crime statistics will vary between different census tracts. The correlation between total building permits and crime rate per 1,000 residents was positive for census tract 39035111700, but negative for census tract 39035114501. The correlation between total demolition permits and crime rate per 1,000 residents, however, was negative for both census tracts, which suggests that tearing down deteriorating structures in general decreases criminal activity. Next, the correlation matrices for the two highest populated census tracts, census tract 39035117700 (Figure 3) and census tract 39035124100 (Figure 4), are shown below:

Figure 3

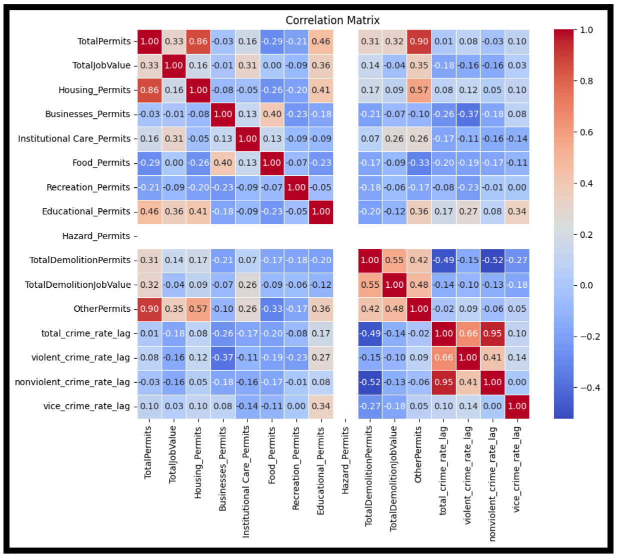
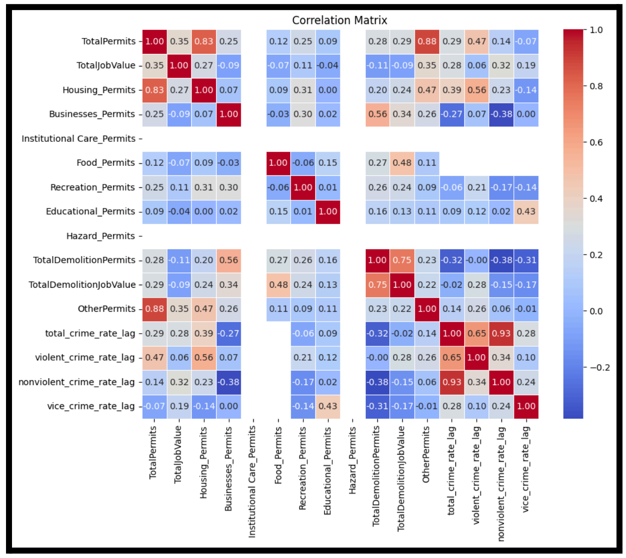


Figure 4



In both census tracts, the correlation between total demolition permits and crime rate per 1,000 residents is again negative. While both census tracts additionally have positive correlations between total building permits and crime statistics, that of census tract 39035124100 is much stronger. The correlations of individual permit types and crime statistics appear to vary heavily between the census tracts as well. Lastly, the correlation matrices for the two random census tracts, census tract 39035117700 (Figure 5) and census tract 39035124100 (Figure 6), are shown below:

Figure 5

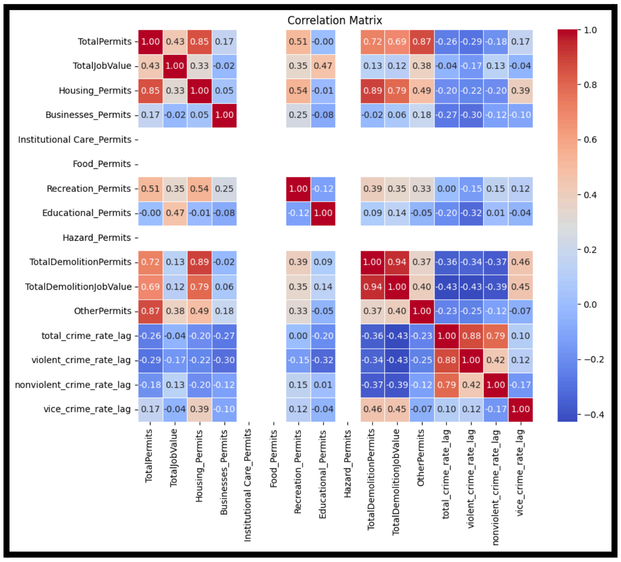
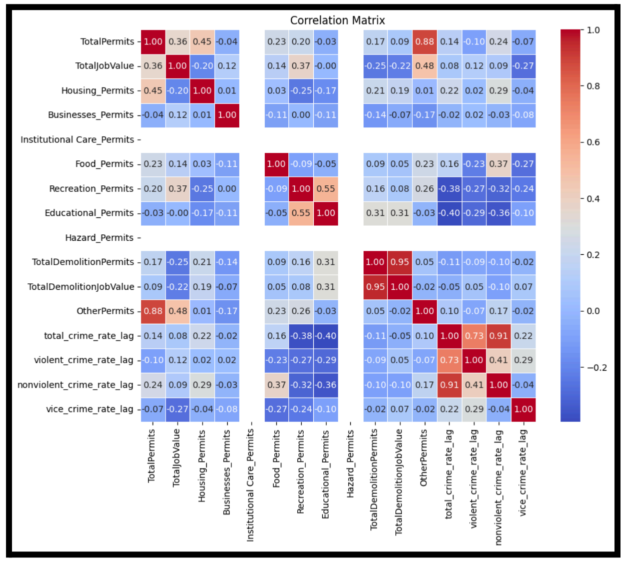


Figure 6

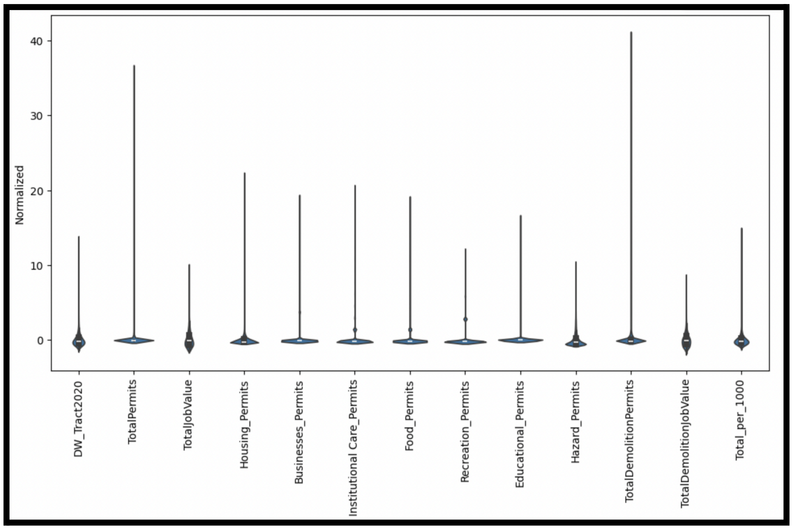


For the random census tracts, there is again a split in the correlations between total building permits and crime rate per 1,000 residents, as one is positive (census tract 39035124100) and one is negative (census tract 39035117700). Again, the correlations between total demolition permits and crime rate per 1,000 residents are both negative.

There is clearly significant variance in the correlations between permit data and crime statistics. Thus, census tracts must be analyzed independently in the model. Overall, the variance in the effects of the selected features on crime statistics supports an approach of obtaining and implementing results for each census tract individually. Additionally, the variance in missing data (specifically with regards to permit types) shows that the investment strategy differs from tract to tract. Likely, this signifies that there are confounding variables that are not accounted for in this data. Thus, any model results must be evaluated by policymakers within the context of regional differentiators.

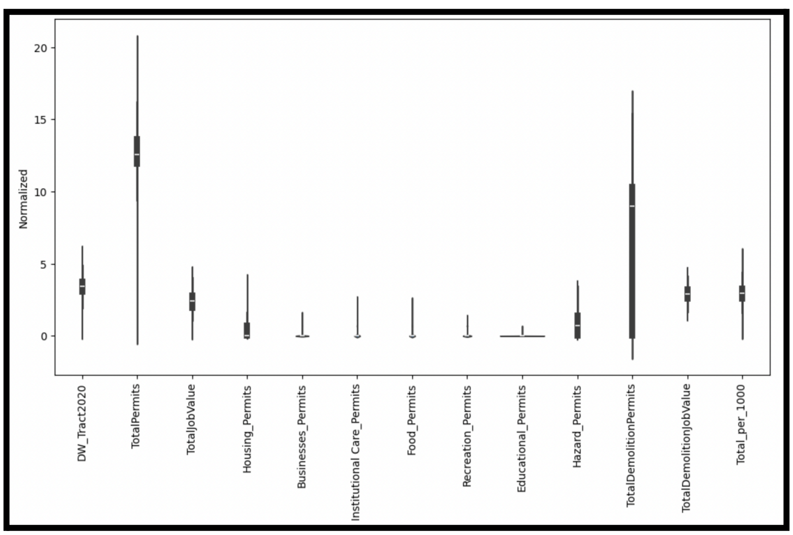
The selected features were normalized before being used as inputs into a model. The distributions of the raw data for each feature (Figure 7) were extremely skewed, seen below:

Figure 7



To normalize the data, the natural log of each feature was taken. In order to account for values of zero, a universal constant was added as well. The final distribution of each feature in the training data is less but still considerably skewed. Ultimately, as evidenced by the distributions of each feature, the data itself is a limitation of this research, as it is difficult to model. The distributions (Figure 8) are shown below:

Figure 8



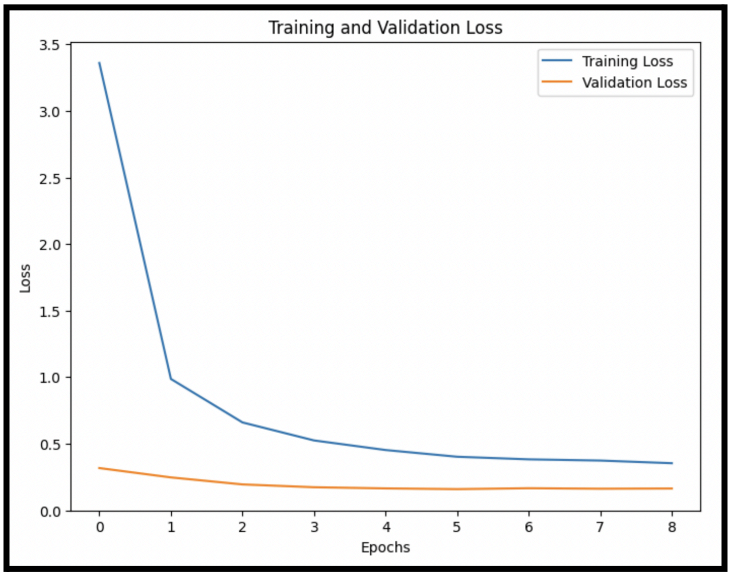
The normalized distributions of each feature in the validation and testing data were representative of those shown in the training data.

In order to analyze this data, a long short-term memory model was utilized, as the research goal centers around discovering patterns with sequential data. The model was built through the TensorFlow machine learning package. It contains three layers: two bidirectional LSTM layers and an output layer. The bidirectional LSTM layers contained sixty-four and thirty-two units, both with ReLU activation, and the output layer contained one unit. Each bidirectional LSTM layer was followed by a 30% dropout to prevent overfitting, where models memorize training data and cannot make accurate predictions on new data. The model allowed for a maximum of twenty epochs, with early stopping enabled with a patience of three epochs. An initial learning rate of .0001 was utilized, dropping by 50% every three epochs. The batch size used was 128. The model was run along a data split of 70% for training, 20% for validation, and 10% for testing. The aforementioned twelve features served as inputs. The model outputted predictions for total crime rate per 1,000 residents, “Total_per_1000”, in individual Cleveland census tracts. The model used six windows of data to make a singular prediction one year in advance to capture the assumed lag effects. For example, if the last window used for prediction contained July to December in 2018, the model would predict the total crime rate per 1,000 residents for this same range of months in 2019.

Results

The model performance was highly variable, as prediction accuracy heavily fluctuated between different census tracts and years. This variance suggests that temporal and regional factors have an immense influence on overall crime rate. Consequently, model results must be interpreted separately for different census tracts and different years. Overall, the model appeared to learn reasonably well with training and validation data, shown below with the loss plot (Figure 9):

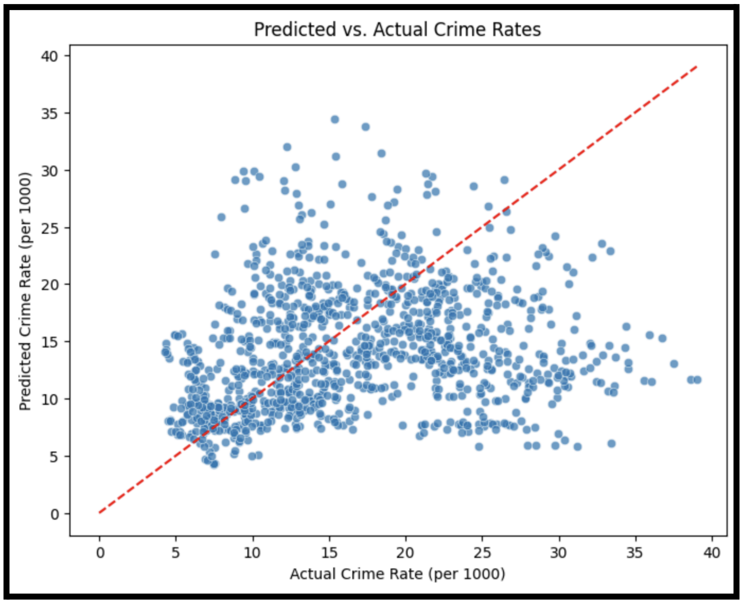
Figure 9



The model produced a final training loss of .3550 with a mean absolute error of .4699, and a final validation loss of .1637 with a mean absolute error of .3155. Since the training and validation loss curves are somewhat comparable and plateau at similar values, the model appears to be fitting the data in an adequate fashion.

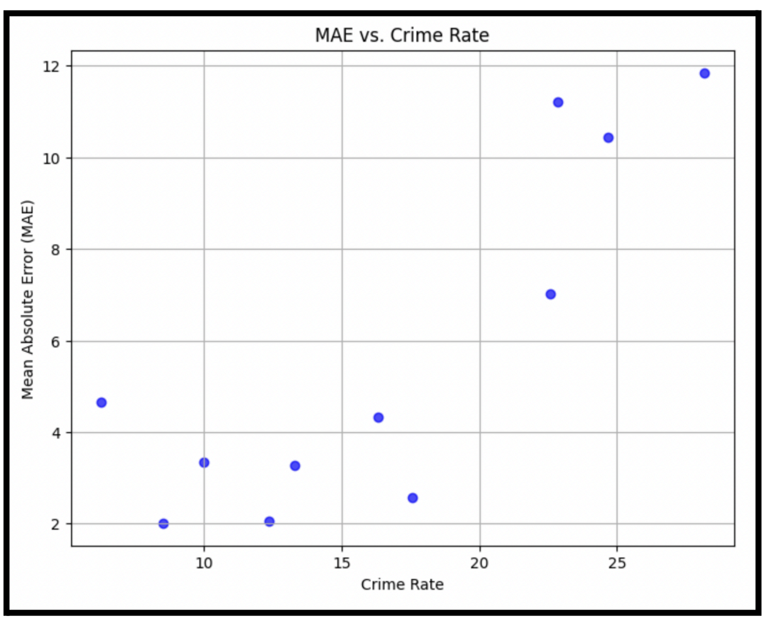
Moving onto the testing data, the model produced a final test loss of .1535 and a mean absolute error of .3217. After undoing the normalization on the predictions, this became an actual mean absolute error of 5.64. In simplified terms, each model prediction was off by an average of 5.64 from the actual total crimes per 1,000 residents. In the context of the research, this means that each model prediction was off by about .6%. The scatterplot of the predictions compared to the actual values for all census tracts in the testing data (Figure 10) is shown below:

Figure 10



Evidently, the model predictions were highly scattered. The model seems to struggle with predicting higher crime rates per 1,000 residents specifically. This trend is additionally apparent when analyzing the relationship between prediction accuracy and crime rate per 1,000 residents for each census tract in the testing data, shown below (Figure 11):

Figure 11



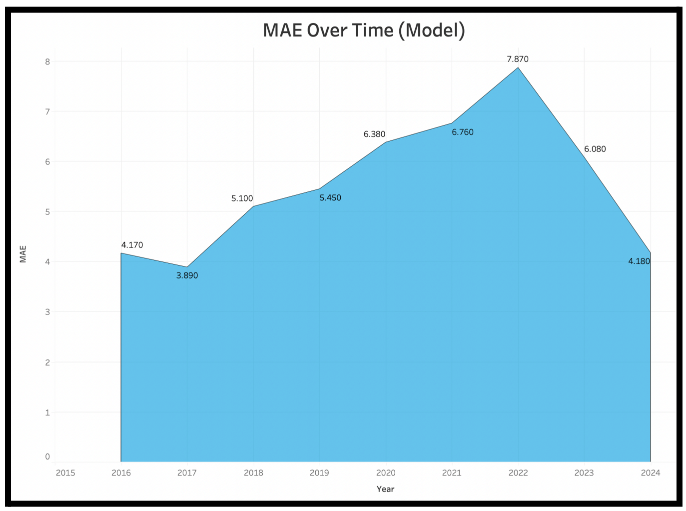
This problem can be addressed in future work by employing a method of stratified sampling to ensure that census tracts with high crime rates are adequately represented in the training data.

However, there are confounding factors that appear to alter model performance. Mainly, the model seems to predict far worse during 2020 through 2023, the years associated with the pandemic (the federal public health emergency for COVID-19 from the United States Department of Health and Human Services expired on May 11th, 2023). This variance is displayed below (Table 1 and Figure 12):

Table 1

	Year	mae
0	2016	4.176537
1	2017	3.891389
2	2018	5.099633
3	2019	5.446554
4	2020	6.377977
5	2021	6.761600
6	2022	7.866912
7	2023	6.081592
8	2024	4.178387

Figure 12



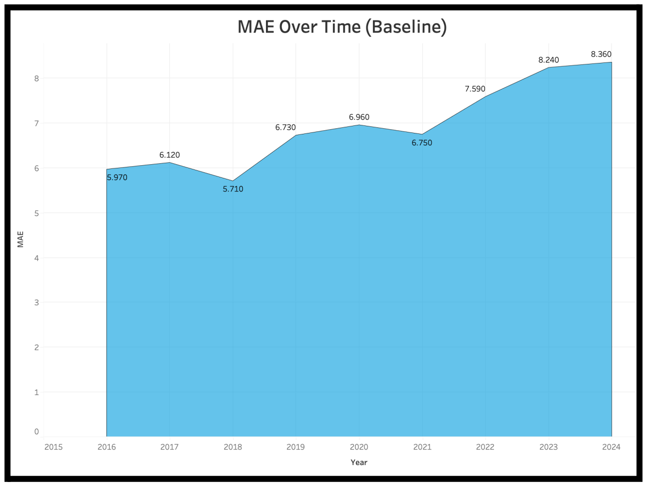
Clearly, the pandemic years significantly altered the overall mean absolute error of the model.

Regarding comparative model performance, a baseline was employed of using the average crime rate per 1,000 residents across all Cleveland census tracts to predict crime rates in census tracts for the next calendar year. This baseline had an overall mean absolute error of 6.79. Thus, accounting for all data, the model performed about 17% better than the baseline—the average prediction from the model was 17% closer to the actual crime rate per 1,000 residents than that of the baseline. The model provides an overall better estimate of future crime rate, and thus can be used to generate a more accurate analysis of the relationship between the cost of infrastructure investment and the predicted cost of crime in census tracts. Circling back to temporal discrepancies, the baseline also had significant variance by year, shown below (Table 2 and Figure 13):

Table 2

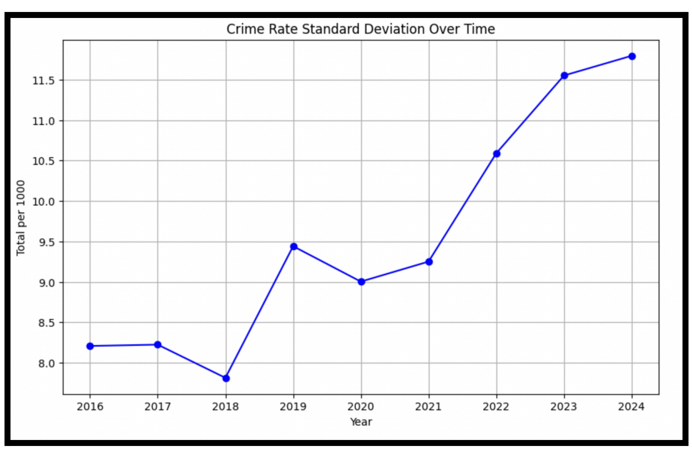
	Year	mae
0	2017	5.972548
1	2018	5.524321
2	2019	6.375341
3	2020	6.720327
4	2021	6.460021
5	2022	7.273322
6	2023	7.909644
7	2024	8.104681

Figure 13



The baseline mean absolute error increasing by year suggests that crime rate across Cleveland is generally becoming more variable. This is proven by examining the standard deviation of the overall Cleveland crime rate per 1,000 residents by year (Figure 14), shown below:

Figure 14



Thus, crime is getting more difficult to predict, presumably related to lasting effects from the pandemic. When taking only data from 2024 into account, the model performance becomes much better compared to the baseline. For 2024 predictions, the model had a mean absolute error of 4.18, compared to the mean absolute error of the baseline of 8.10. This constitutes an increase in performance of 48% for the model over the baseline, compared to the 17% increase overall.

Discussion

As aforementioned, the model accuracy varies greatly between individual census tracts. Thus, the model cannot be implemented on a broad level, but rather must be implemented on a tract-by-tract basis. A simple way to determine which tracts are viable options for the model to be implemented for is baseline comparison. The model performed better than the baseline (had a lower mean absolute error) in around half of census tracts in the testing data. However, there were also multiple census tracts where the model, while having a higher mean absolute error than the baseline, still performed better than the model's overall mean absolute error for all tracts in the testing data. Namely, model predictions for census tract 39035110901 had a mean absolute error of 3.13, compared to the baseline mean absolute error for that specific tract of 2.16, and model predictions for census tract 39035119502 had a mean absolute error of 3.97, compared to the baseline mean absolute error for that specific tract of 3.03. Thus, determination on if the model can be employed for a given census tract can come from a combination of model performance for that specific tract being better than baseline performance for said tract and being better than the overall model performance across all tracts.

Mainly, model application will revolve around cost analysis. Analyzing only the census tracts where model use is deemed to be permissible according to the aforementioned criteria, the predictions can be directly compared to the total cost of investment. This can be done through comparing the sum of job value of building and demolition projects in a given year to the model predictions for the following year. For example, census tract 39035122200 (near Garfield Heights) saw a total investment of $1,344,331 in 2023. For 2024, the model predicts the crime rate per 1,000 residents of the tract to drop by .057% from its figure in 2023. When translated into financial terms using the per capita crime cost in Cleveland (calculated through the aforementioned research brought forth by the University of Miami) and the population of the tract, this results in $80,259 in total savings. Obviously, making determinations on the value of investment in census tracts through this context is outside of the scope of this specific research. These figures must be analyzed by relevant policymakers to determine if investment into a given area is worth it from an economic standpoint regarding potential savings on crime. Recurring effects must also be taken into consideration.

However, this research can directly facilitate better understanding of the value of certain investment types in Cleveland through comparison of similar census tracts. Mainly, comparisons can be drawn from census tracts with reasonably similar investment figures. For example, census tract 39035119600 (East Shaker Heights) had a total investment of $275,494 in 2023, and a predicted total savings in crime cost (from a predicted decrease in crime rate per 1,000 residents) of $285,748 in 2024. Census tract 39035110901 (Newburgh Heights) had a total investment of $350,587 in 2023, yet was predicted to save only $49,695 from a lower decrease in crime rate. Notably, census tract 39035119600 had three issued building permits related to education in 2023, while Census tract 39035110901 had zero. Otherwise, the breakdown of investment type was similar for both census tracts. Census tract 39035117600 (West Euclid) had a total investment of $411,332 in 2023, and a predicted total savings in crime cost of $27,017 in 2024. Conversely, census tract 39035118800 (Little Italy) had a total investment of $439,701 in 2023, but a predicted increase in crime cost of $131,760 (from a predicted increase in crime rate per

1,000 residents) in 2024. Census tract 39035117600 had almost double the amount of issued building permits related to housing and significantly more issued demolition permits than census tract 39035118800 in 2023. Again, the breakdown of investment type was otherwise similar for both census tracts. These examples point to the potential value of educational and housing building permits and demolition permits in relation to reducing regional crime rates, and consequently saving money on the cost of crime.

Overall, the model can be implemented on a tract-by-tract basis to examine the relationship between cost of investment, as well as investment type, and cost of crime. This analysis can be used to formulate a more efficient strategy for investment. While most census tracts did not have a positive return on investment for 2024, model predictions can be extrapolated (and regenerated with new data) to determine the timeline for a positive return on investment given potential long-term savings on crime.

Limitations

The data inputted into the model presented limitations in that there are great discrepancies amongst census tracts. Census tracts vary based on population, and thus investment. Smaller census tracts in the data do not have large counts of permits and have high variance in their crime rates, fostering an issue in terms of predictability. Again, this problem can potentially be solved with stratified sampling, where an adequate number of small census tracts are ensured to be in the training data. Additionally, the pandemic presents challenges for predicting trends in investment and crime rate over time. There are additional concerns with potential confounding factors. For example, more investment in areas has been linked to an increased police presence, which could result in a lowering of crime rate (Zhou et al.). As aforementioned, the model specifically also has limitations. The model must be implemented on a tract-by-tract basis, as the accuracy varies depending on the tract. There are clearly regional factors not accounted for in the data that affect the relationship between infrastructure investment and crime rate in census tracts.

Further Research

Future research in this area must explore these regional factors that influence the relationship between infrastructure investment and crime rate. With a better understanding of these confounding factors, more accurate models can be constructed in the future. Additionally, future research must analyze the temporal influence of the pandemic on this relationship. Specifically, research must uncover if this influence will affect this relationship moving forward, or if there will be a reversion of the relationship between infrastructure investment and crime rate to pre-pandemic trends.

References

Baird, M. D. et al. “Does Large-Scale Neighborhood Reinvestment Work? Effects of Public–Private Real Estate Investment on Local Sales Prices, Rental Prices, and Crime Rates.” *Housing Policy Debate*, vol. 30, no. 2, 2019, pp. 164-190, https://doi.org/10.1080/10511482.2019.1655468.

Boldizar, Ava. “Where Columbus, Cleveland and Cincinnati rank among the most dangerous US cities.” *NBC4i*, 29 Dec. 2013, https://www.nbc4i.com/news/data-desk-ohio/where-columbus-cleveland-and-cincinnati-rank-among-the-most-dangerous-us-cities/.

Branas, Charles C et al. “Urban Blight Remediation as a Cost-Beneficial Solution to Firearm Violence.” *American Journal of Public Health*, vol. 106, no. 12, 2016, pp. 2158-2164, doi:10.2105/AJPH.2016.303434.

Bunting, W.C. “Does increased access to home mortgage money reduce local crime rates? Evidence from San Diego County?” *Regional Science and Urban Economics*, vol. 84, Sep. 2020, https://doi.org/10.1016/j.regsciurbeco.2020.103570.

Burley, Blair. “Green infrastructure and violence: Do new street trees mitigate violent crime?” *Health & Place*, vol. 54, Nov. 2018, pp. 43-49, https://doi.org/10.1016/j.healthplace.2018.08.015.

City of Cleveland. “Cleveland takes big steps forward toward Mayor Bibb’s vision for a 15-minute city.” City of Cleveland, Jul. 2023, https://www.clevelandohio.gov/news cleveland-takes-big-steps-toward-mayor-bibbs-vision-15-minute-city.

Gabriele, Rob. “2024 Crime Rates in U.S. Cities Report.” *SafeHome*, 20 Sept. 2024, https://www.safehome.org/resources/crime-statistics-by-state/.

He, Jingyi, and Hao Zheng. “Prediction of crime rate in urban neighborhoods based on machine learning." *Engineering Applications of Artificial Intelligence*, vol. 106, Nov. 2021, https://doi.org/10.1016/j.engappai.2021.104460.

Luscher, Dan. "Cleveland's 15-minute city journey." *The 15-Minute City Project*, 25 Apr. 2024, https://www.15minutecity.com/blog/cleveland.

McCollister, Kathryn et al. "The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation." *Drug Alcohol Depend*, vol. 108, Apr. 2010, https://doi.org/10.1016/j.drugalcdep.2009.12.002.

Montolio, Daniel. "The effects of local infrastructure investment on crime." *Labour Economics*, vol. 52, June 2018, pp. 210-310, https://doi.org/10.1016/j.labeco.2018.02.006.

Yin, Zongyi et al. "Public Security Expenditure, Education Investment, and Social Stability: An Empirical Analysis Based on Provincial Panel Data from China." *Revista De Cercetare Si Interventie Sociala*, vol. 59, Dec. 2017, pp. 239-260, http://jcu.ohionet.org/login?url=https://www.proquest.com/scholarly-journals/public-security-expenditure-education-investment/docview/1980708046/se-2?accountid=35810.

Zhou, Yilin et al. "Does Neighborhood Investment Actually Affect Crime? New Evidence from LIHTC and Smartphone-based Measures of Policing?" *The Review of Economics and Statistics*, Nov. 2023, https://www.anderson.ucla.edu/sites/default/files/document/2024-02/2024-04WP.pdf.