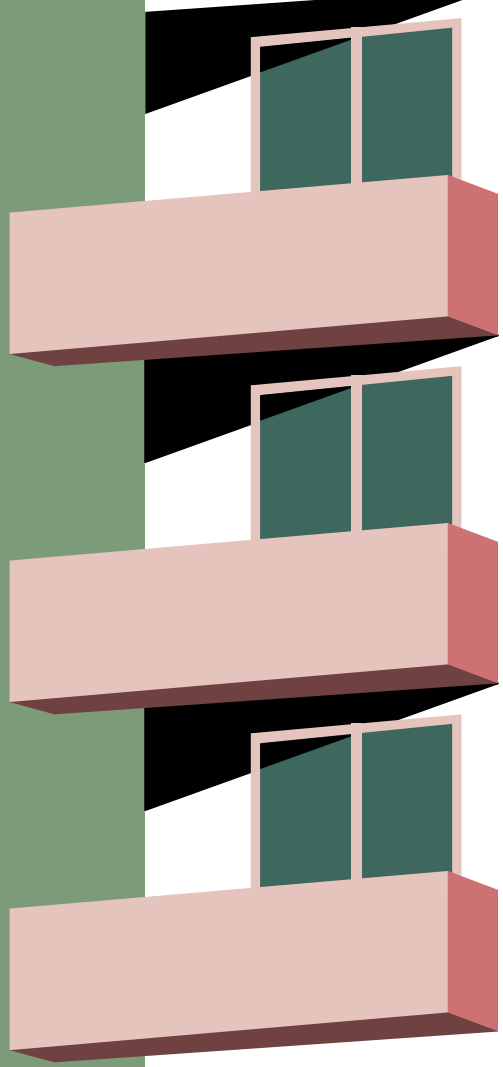
CITY OF CLEVELAND

ADRIANA SCHERMAIER, CAM ROBERSON, ERIC GEISLER, TRYNA RODMAN

01

THE PROBLEM



THE PROBLEM

QUESTION:

How does infrastructure investment in Cleveland census tracts affect crime rates?

THE PROBLEM

CRIME

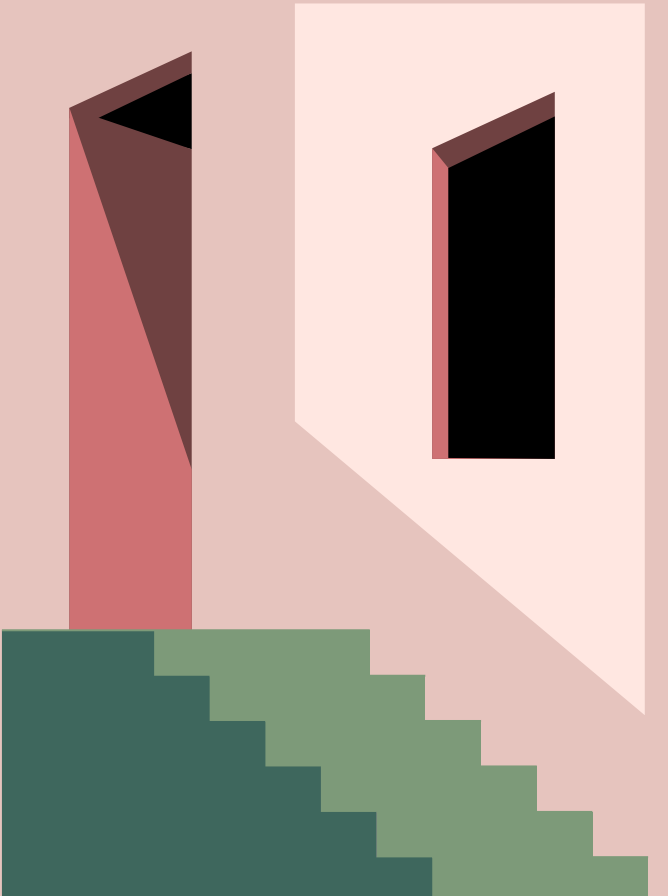
Cleveland has the 3rd highest rate of murder, 2nd highest rate of rape, and 2nd highest rate of robbery out of all medium-sized cities in the United States.

MONEY

The city of Cleveland spent $7,397 per capita on costs associated with crime in 2023, equating to a total of over $2 million.

PLANNING

The local “15-minute city” initiative has yet to expand into general infrastructure, opening the door for a cursory analysis for optimization.



02

THE DATA

DATASETS

CRIME INCIDENTS

~700,000 rows
2015 to 2024

BUILDING PERMITS

~170,000 rows
2015 to 2024

DEMOLITION PERMITS

~9,700 rows
2015 to 2024

2020 CENSUS TRACTS

TRACT POPULATION CHANGE

FEATURE ENGINEERING

Created crime statistics by adding in dimension of population.

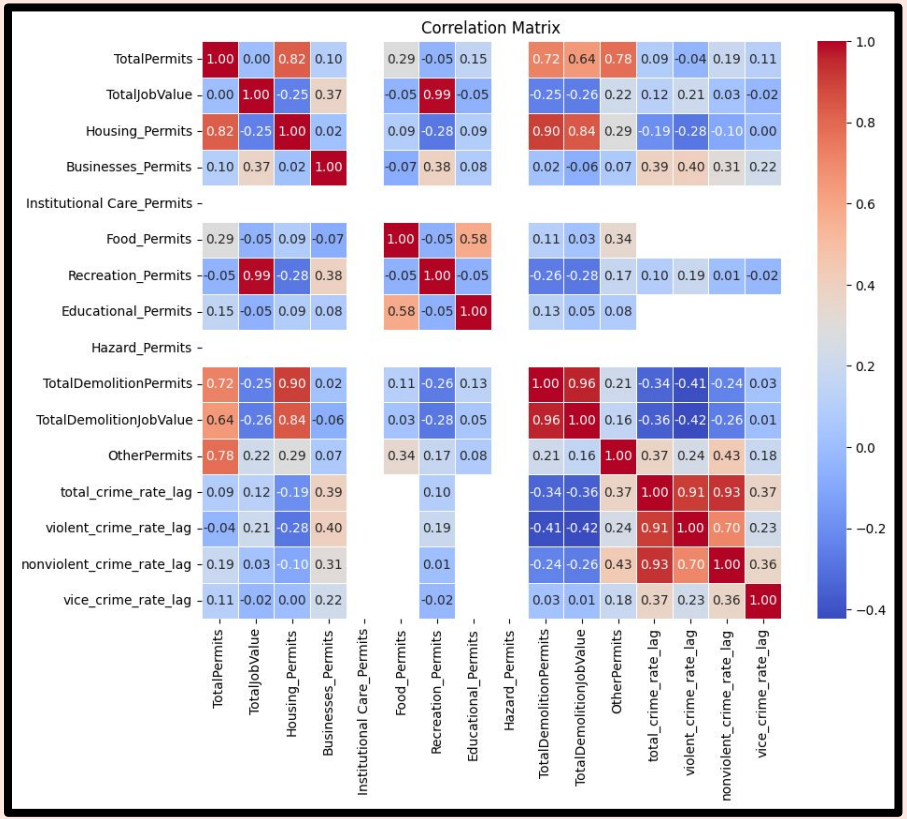
Engineered count of permit data, sum of value data, and average of crime data.

Grouped variables into 6-month sliding windows.

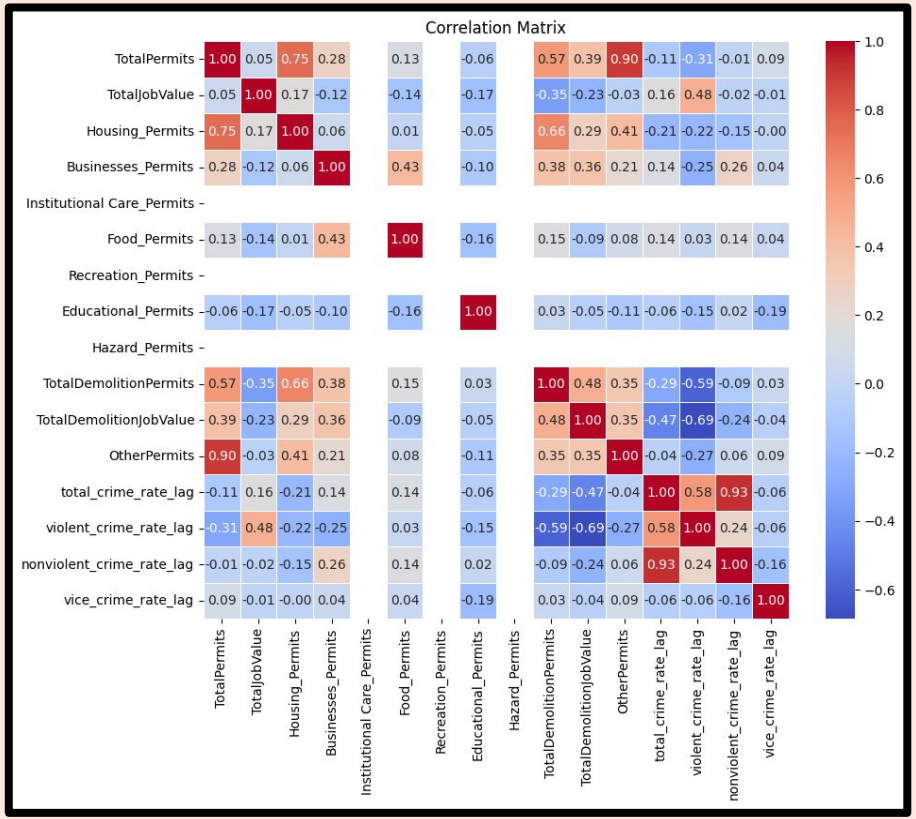
Finalized predictor and outcome variables with correlation analyses.

CORRELATION ANALYSIS: LOW POPULATION

TRACT 39035111700

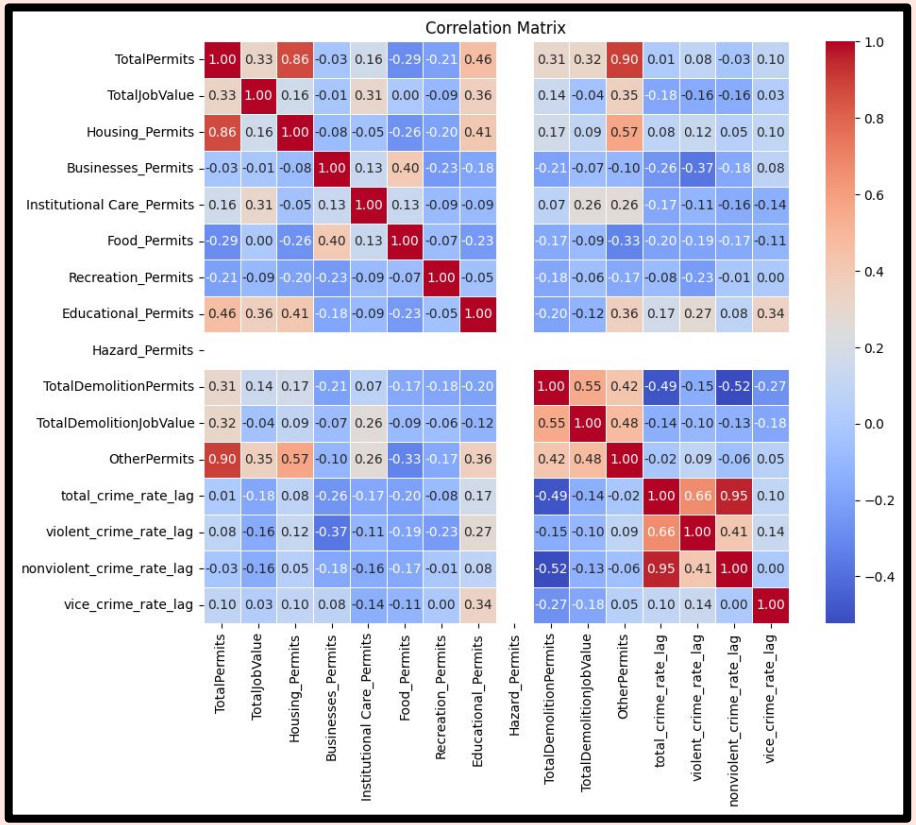


TRACT 39035114501

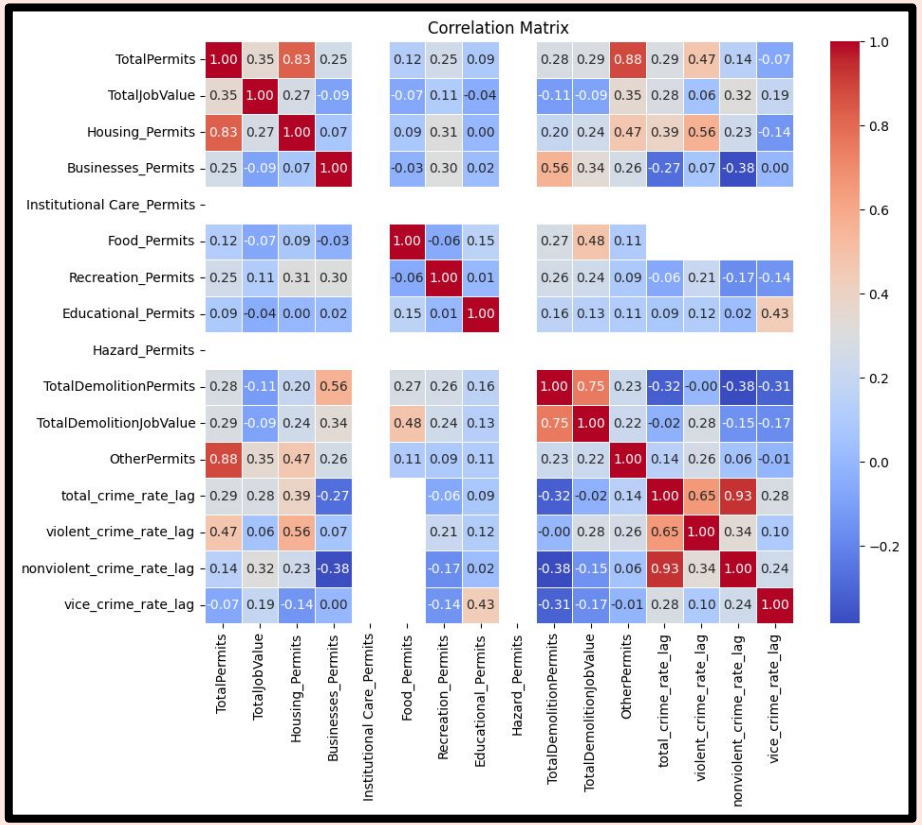


CORRELATION ANALYSIS: HIGH POPULATION

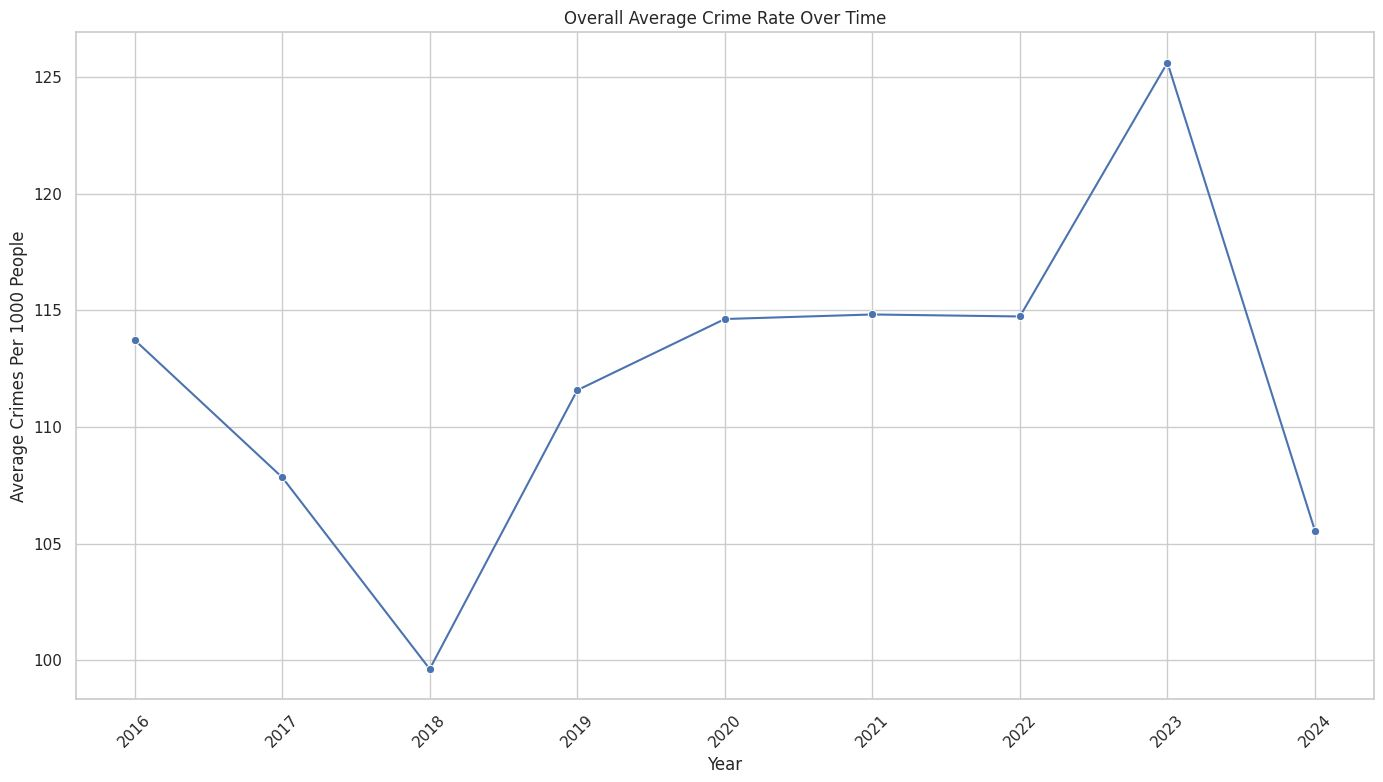
TRACT 39035117700



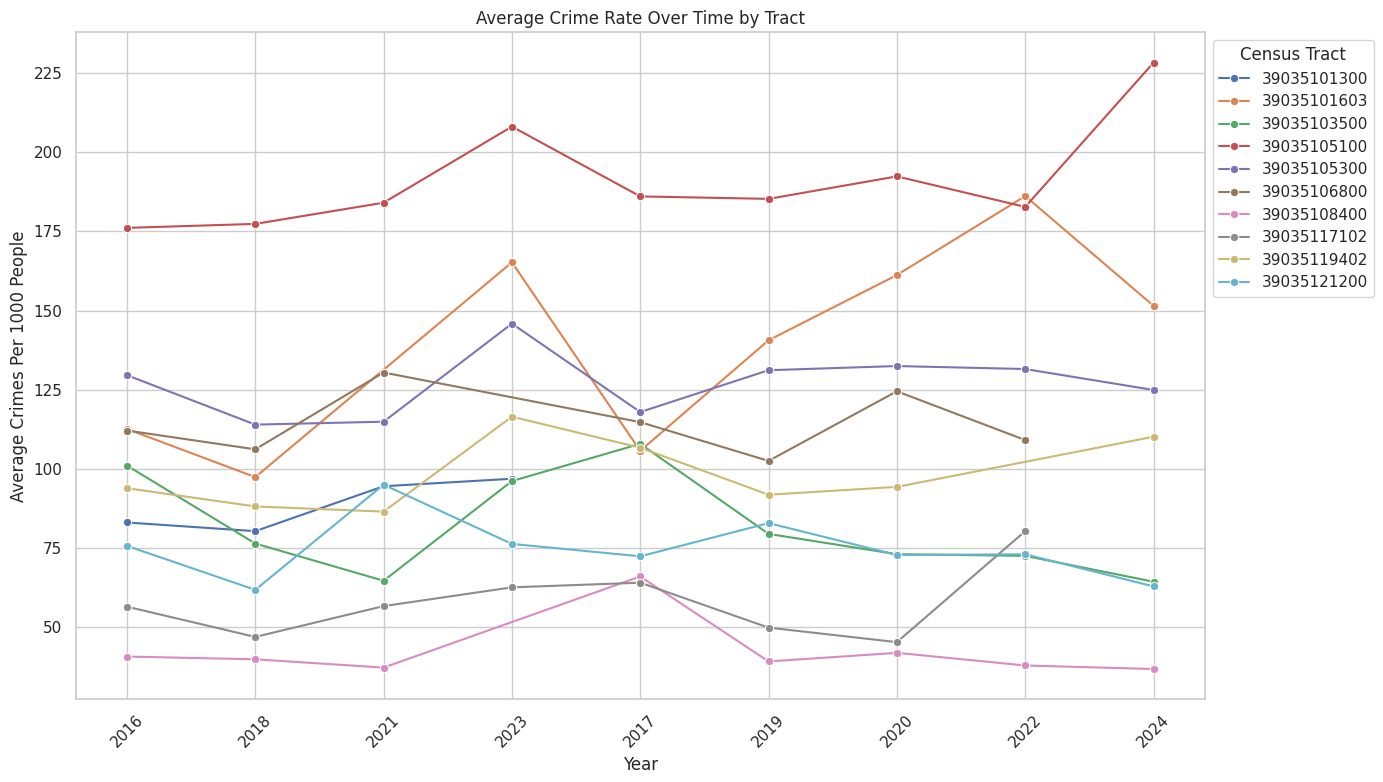
TRACT 39035124100



EXPLORATORY ANALYSIS

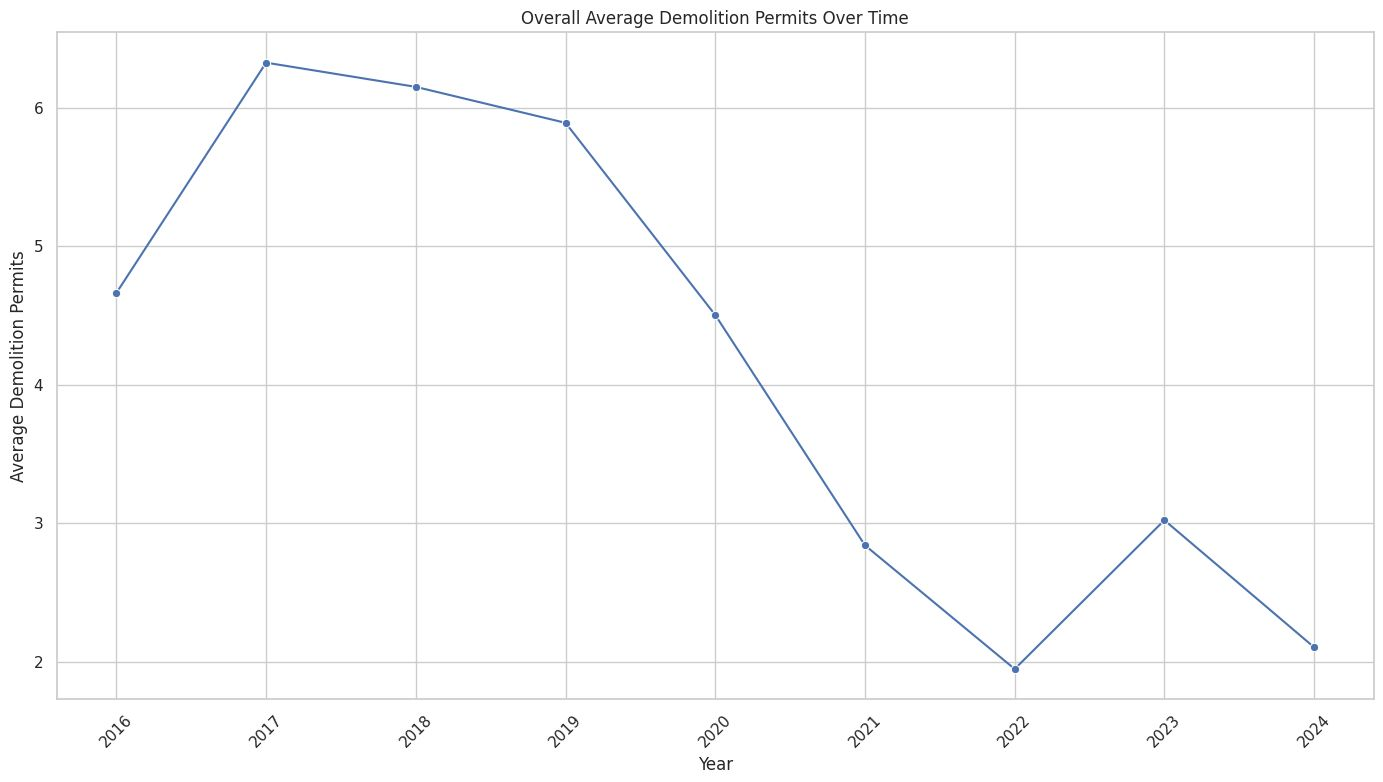


EXPLORATORY ANALYSIS

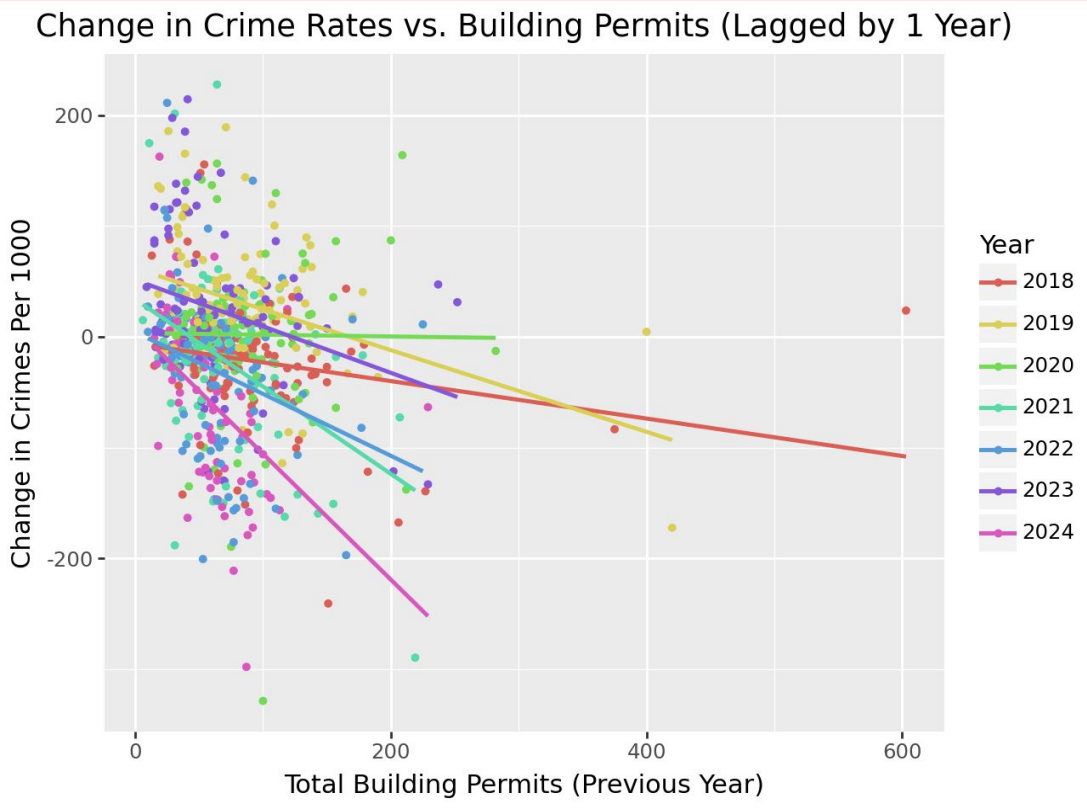


EXPLORATORY ANALYSIS

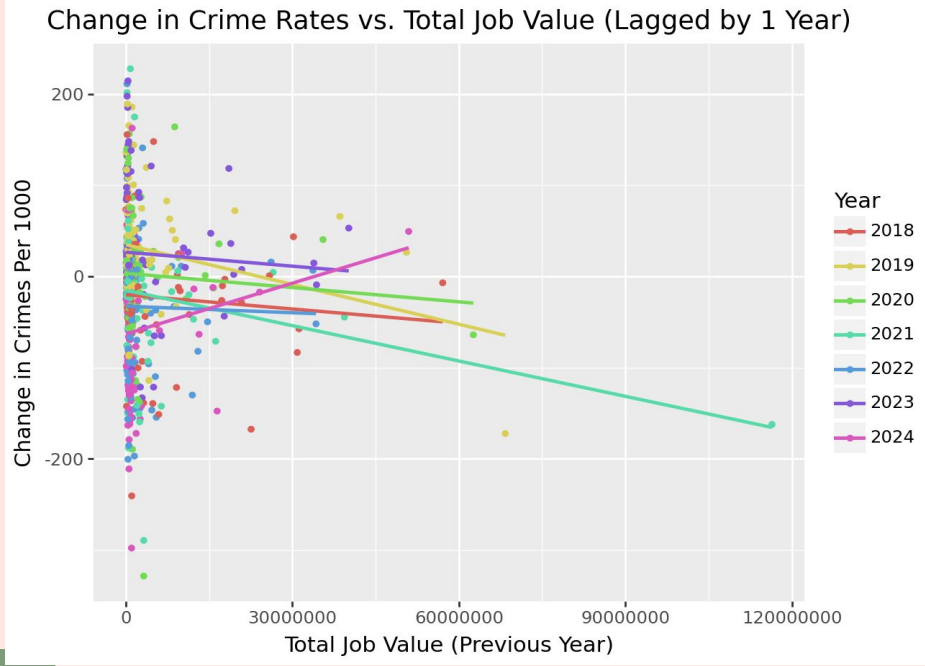
EXPLORATORY ANALYSIS

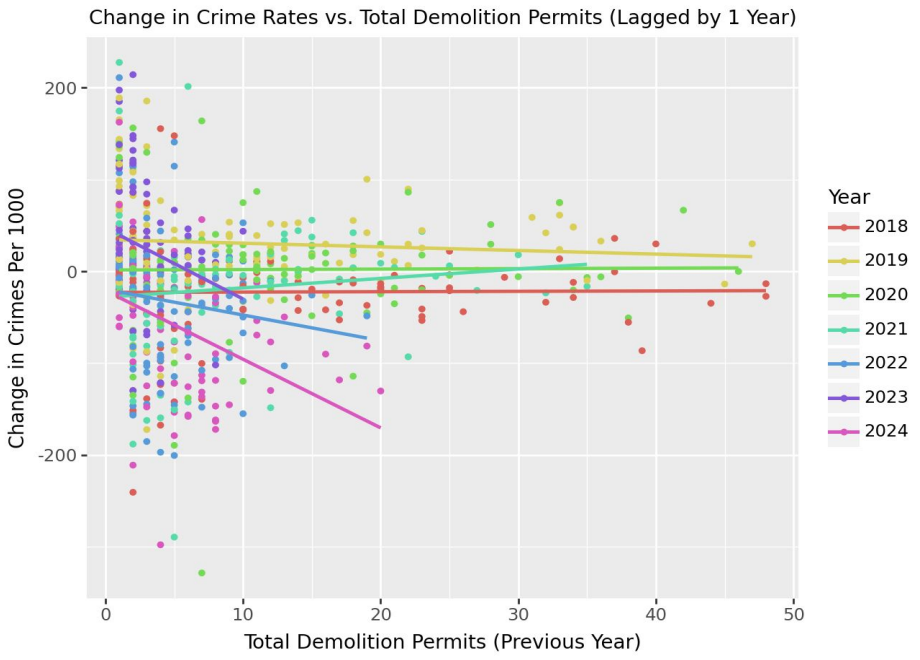


EXPLORATORY ANALYSIS



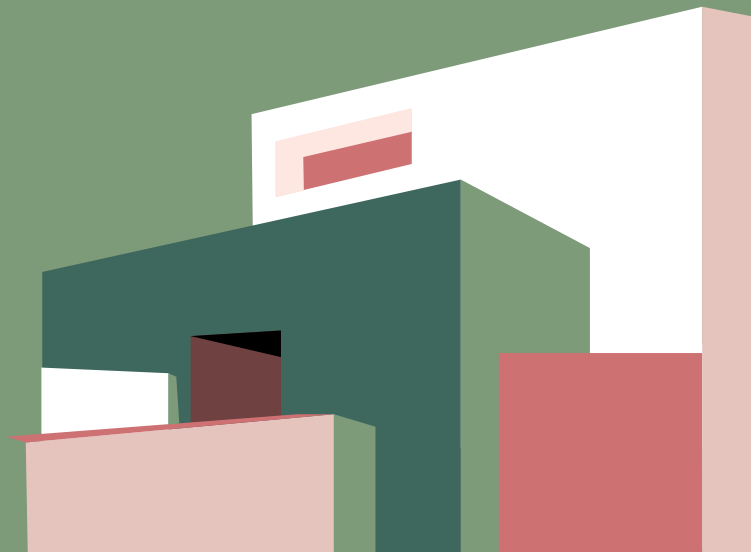
EXPLORATORY ANALYSIS





03

THE MODEL

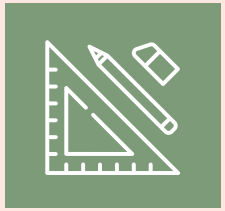


BASELINE

Using the overall crime rate per 1,000 residents of the previous year in Cleveland to predict the crime per rate per 1,000 residents in census tracts.

SPLITTING THE DATA

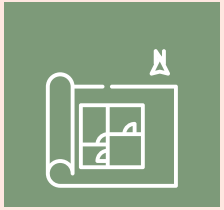
Split by unique census tracts



TRAINING

70%

(79 tracts)



VAL

20%

(23 tracts)

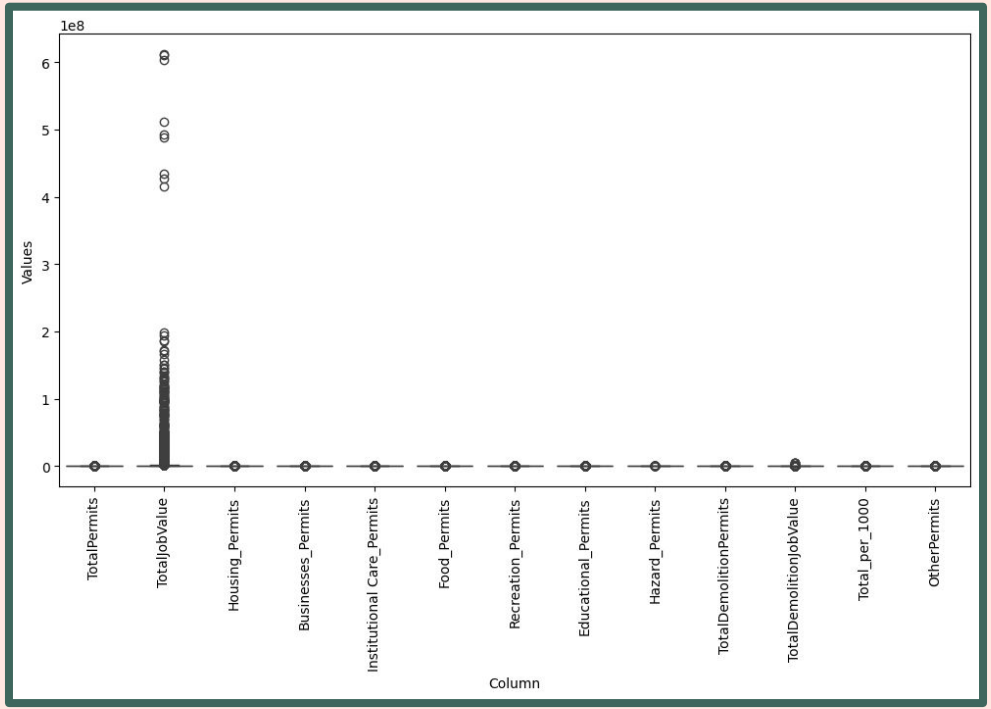
TEST

10%

(11 tracts)

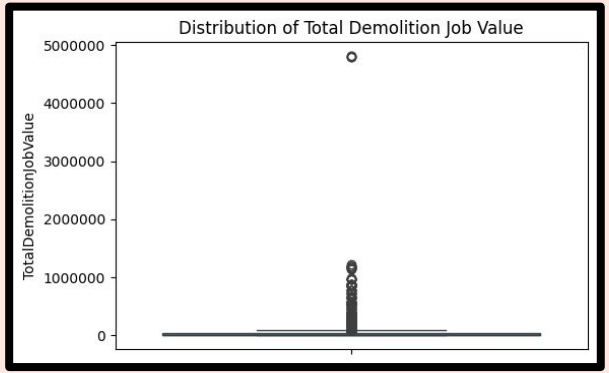
NORMALIZATION

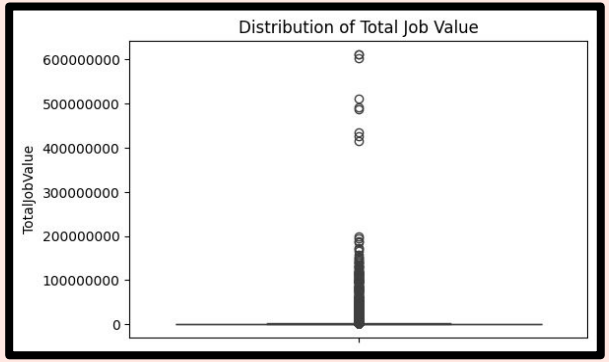
BEFORE:

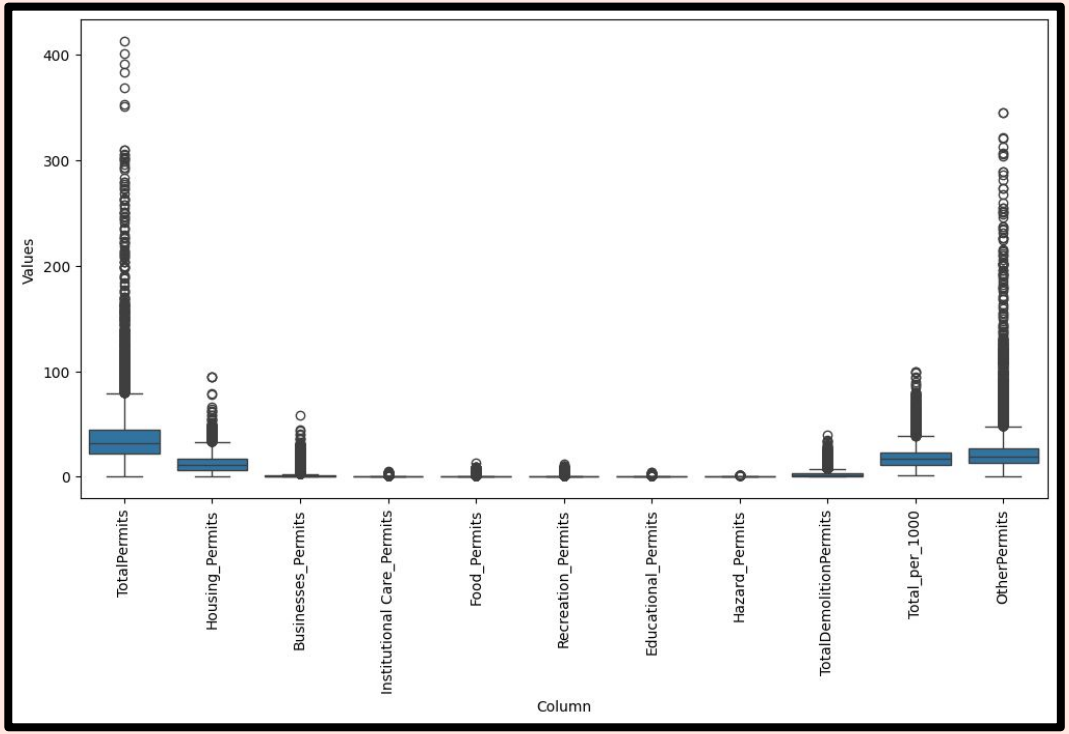


NORMALIZATION

BEFORE:

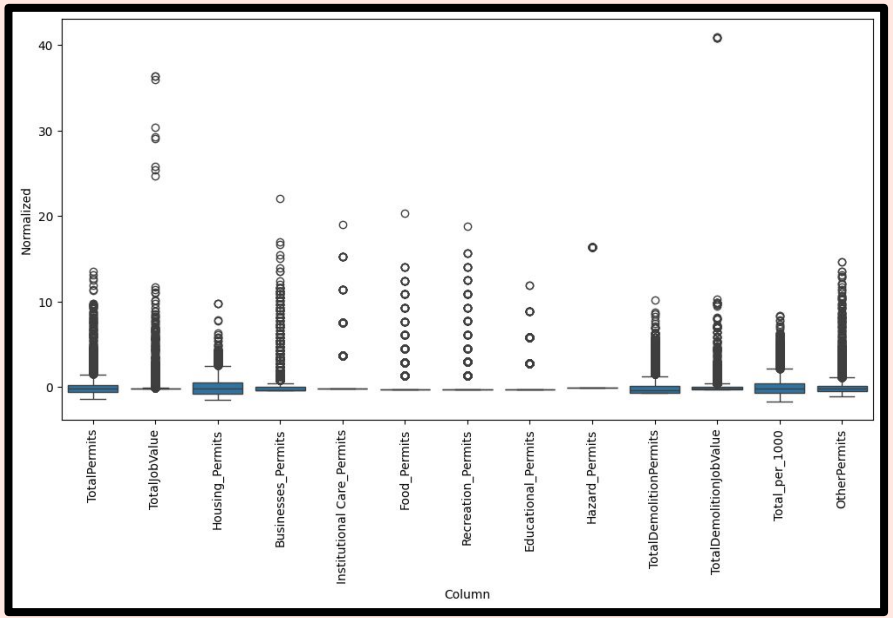




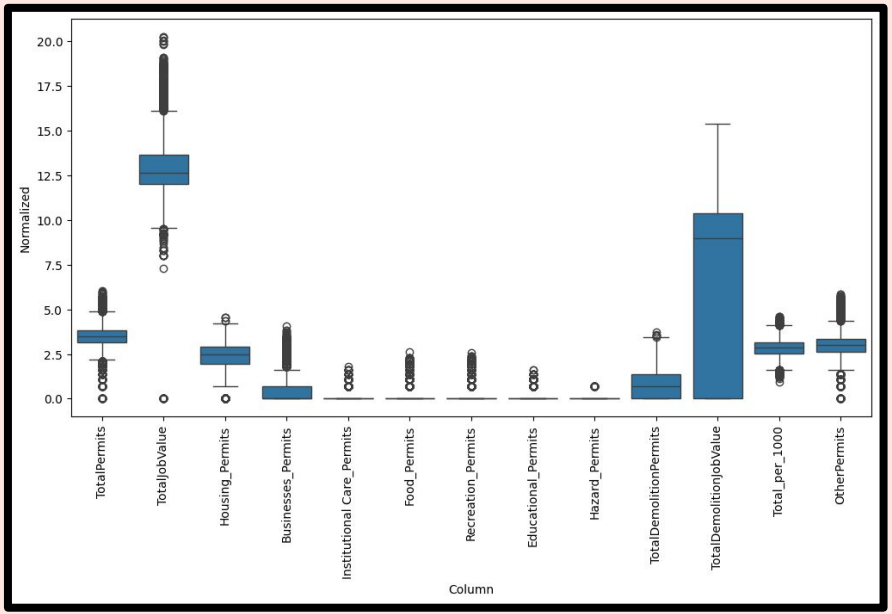


NORMALIZATION

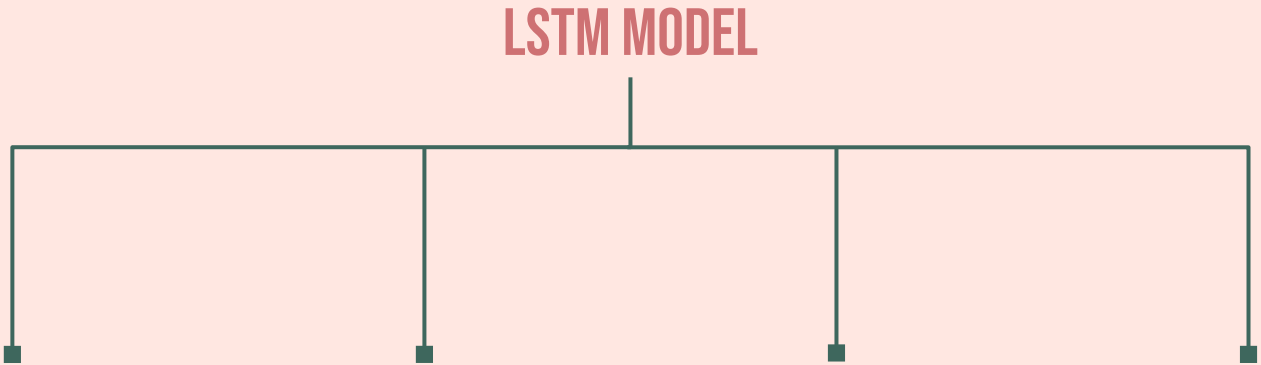
Z-SCORE:



LOG:



MODEL STRUCTURE



LSTM LAYERS

- 2 bidirectional layers
- 64 and 32 units
- 0.3 dropout rate
- ReLU activation
- 1 unit output layer

OPTIMIZER & LOSS

- Adam optimizer
- MSE loss function
- MAE metric

HYPERPARAMETERS

- batch size of 128
- 20 epochs
- learning rate of .0001

ADAPTATIONS

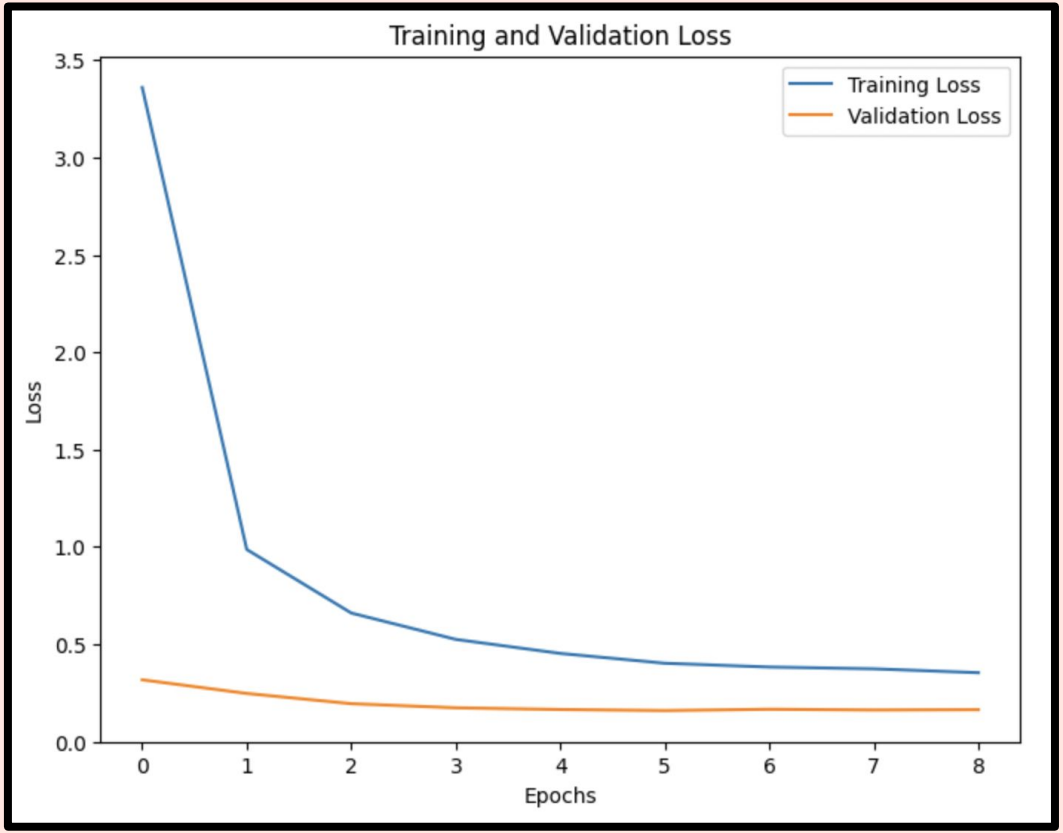
- learning rate scheduler
- early stopping

EVALUATION

TRAINING LOSS:
.3550

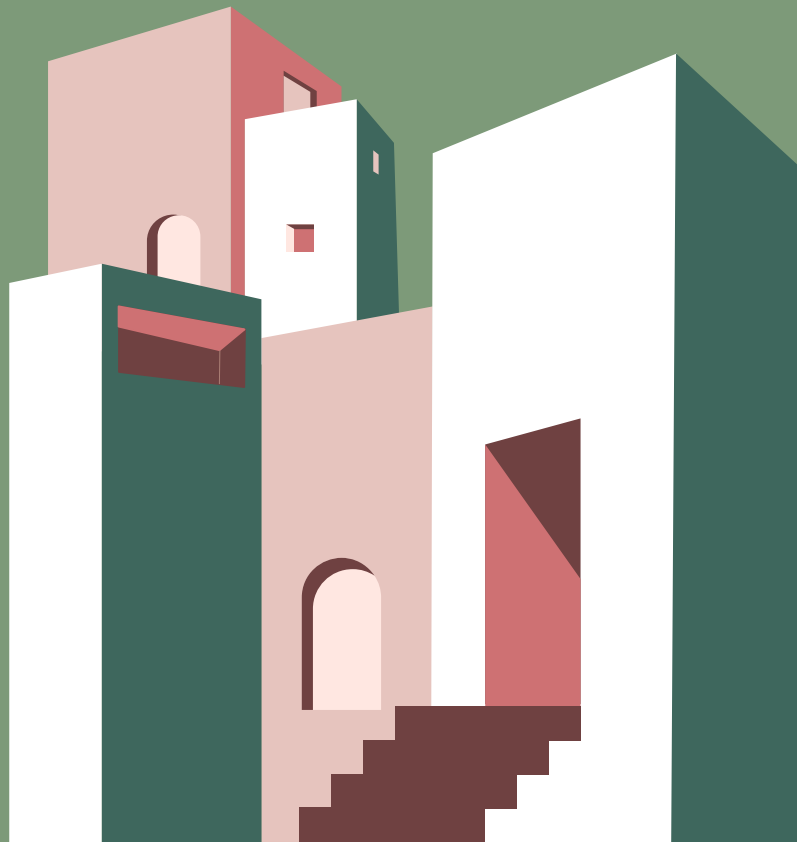
VALIDATION LOSS:
.1637

TEST LOSS:
.1535

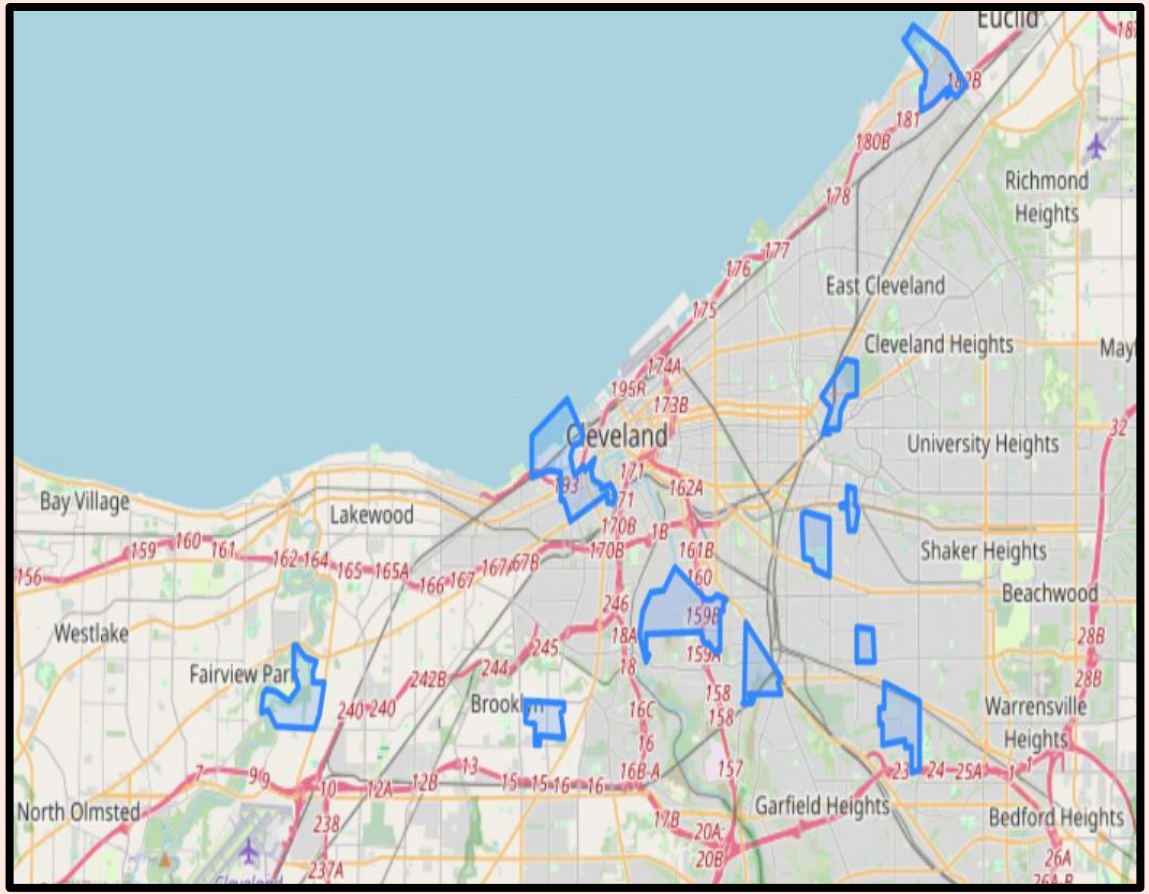


04

THE RESULTS

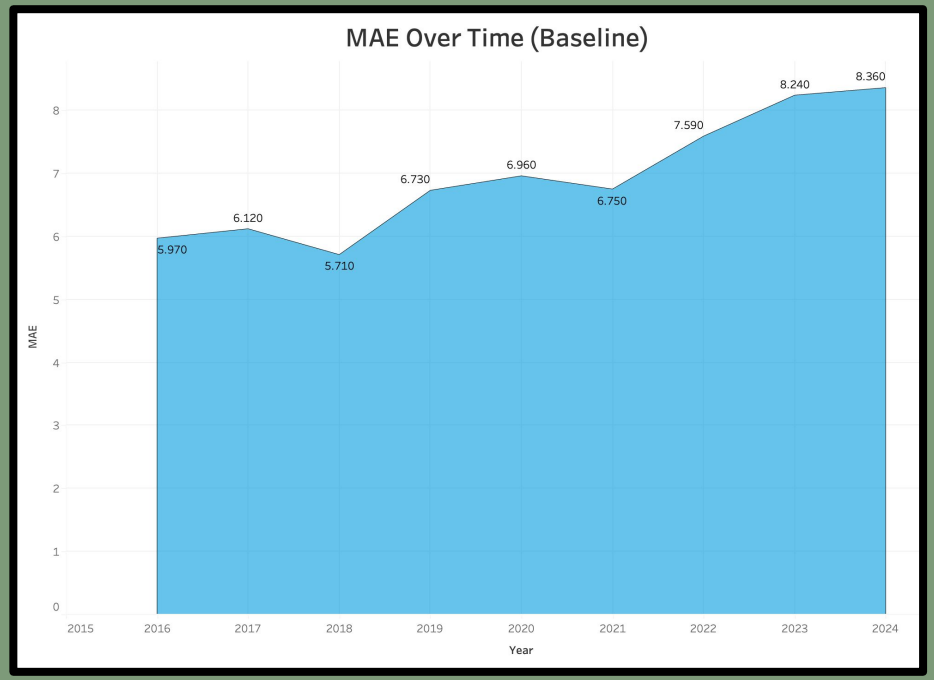


TEST CENSUS TRACTS

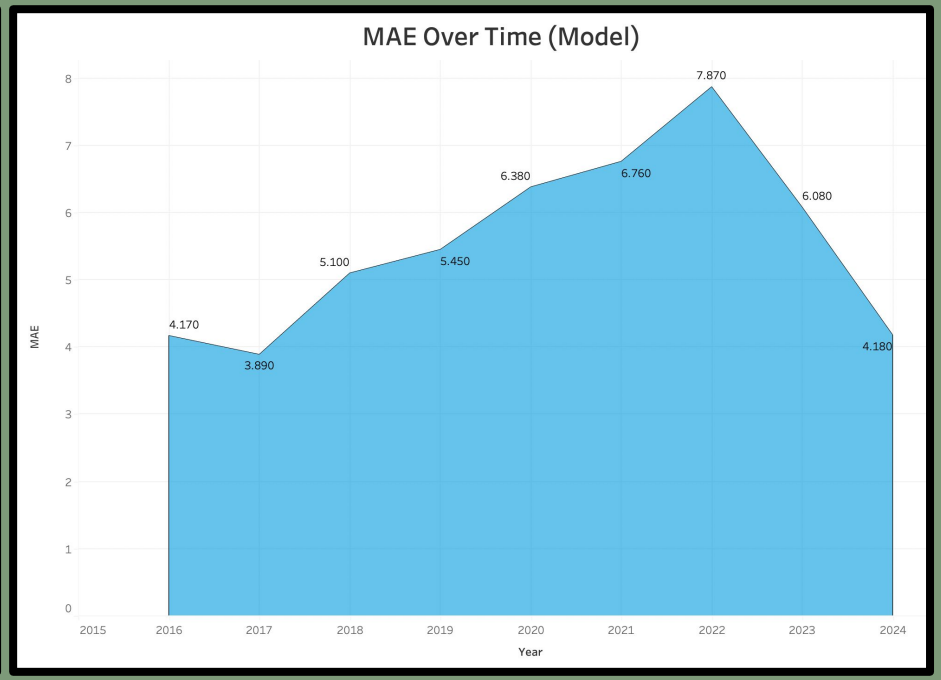


ACCURACY

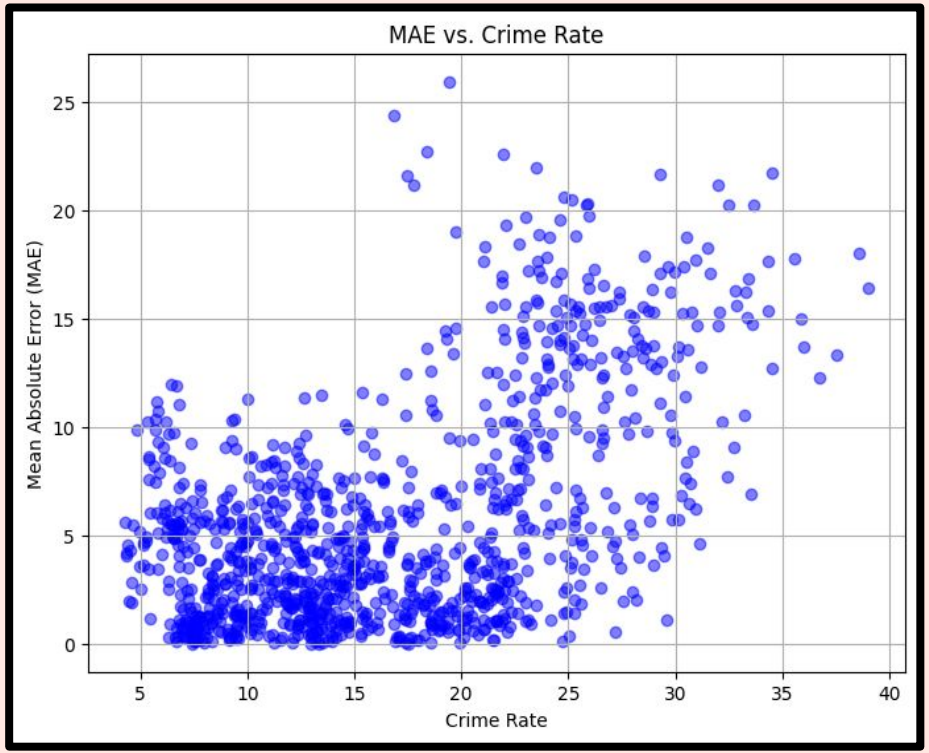
BASELINE
MAE: 6.8 MAPE: 49.1

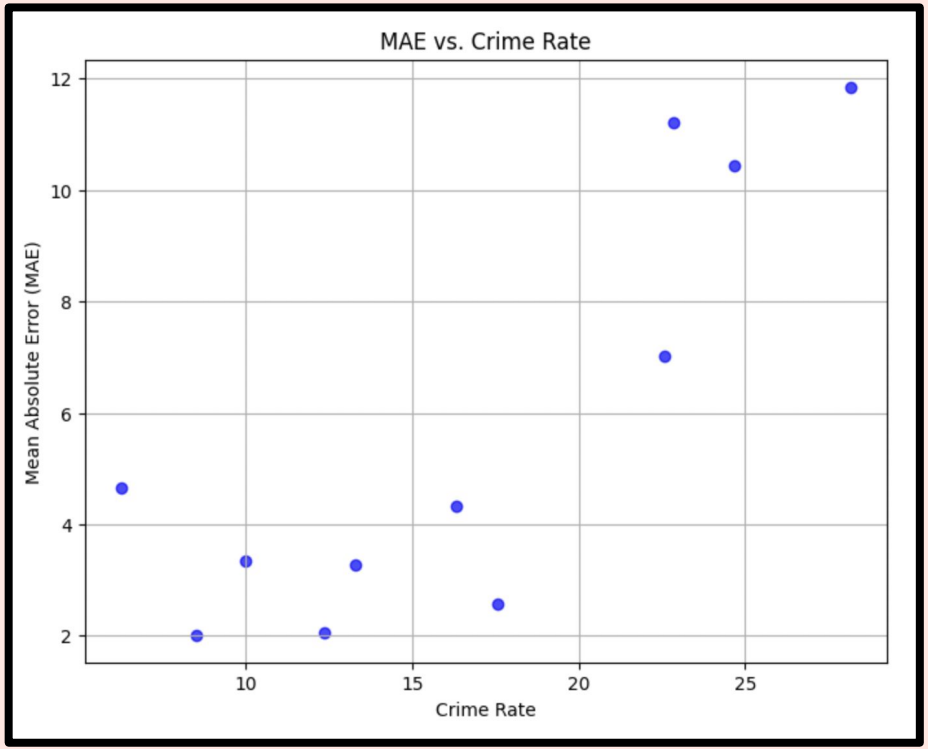


MODEL
MAE: 5.6 MAPE: 37.3

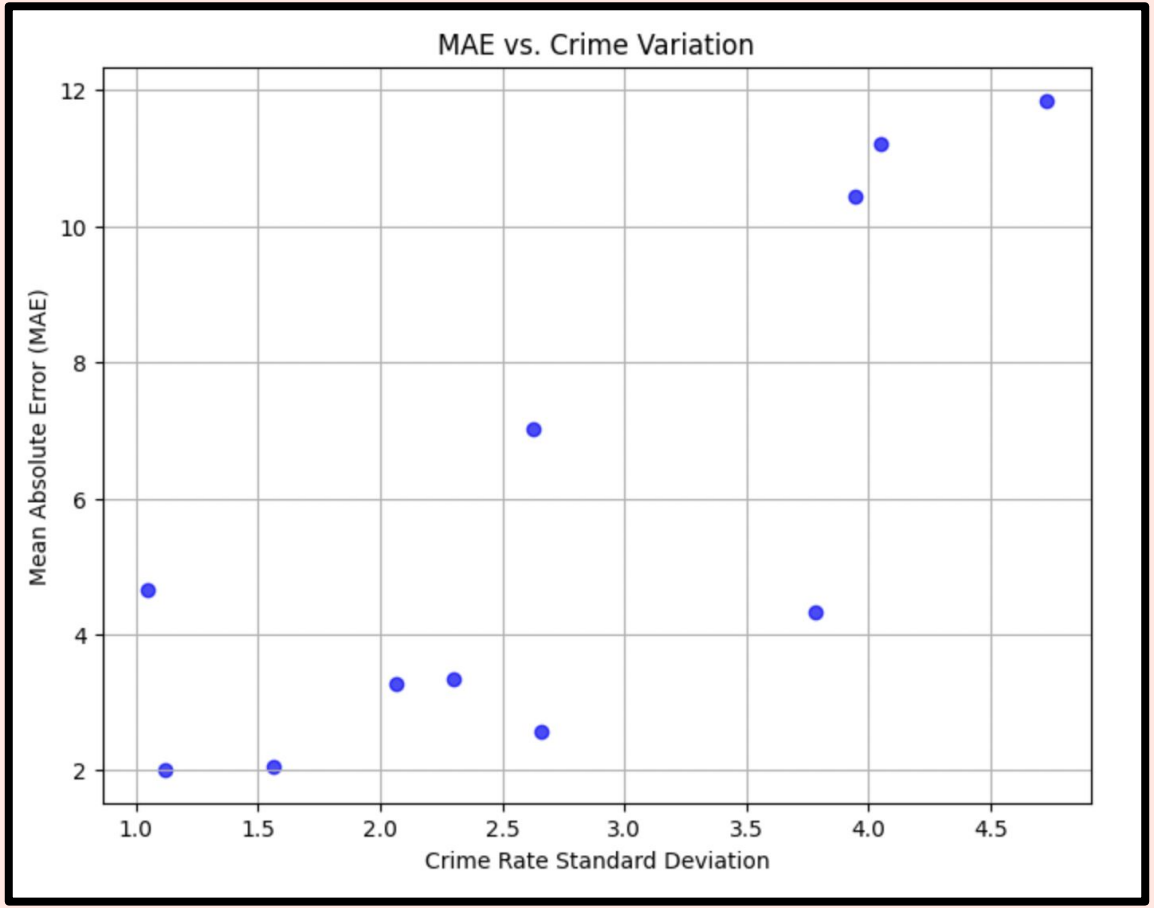


ACCURACY VERSUS CRIME RATE

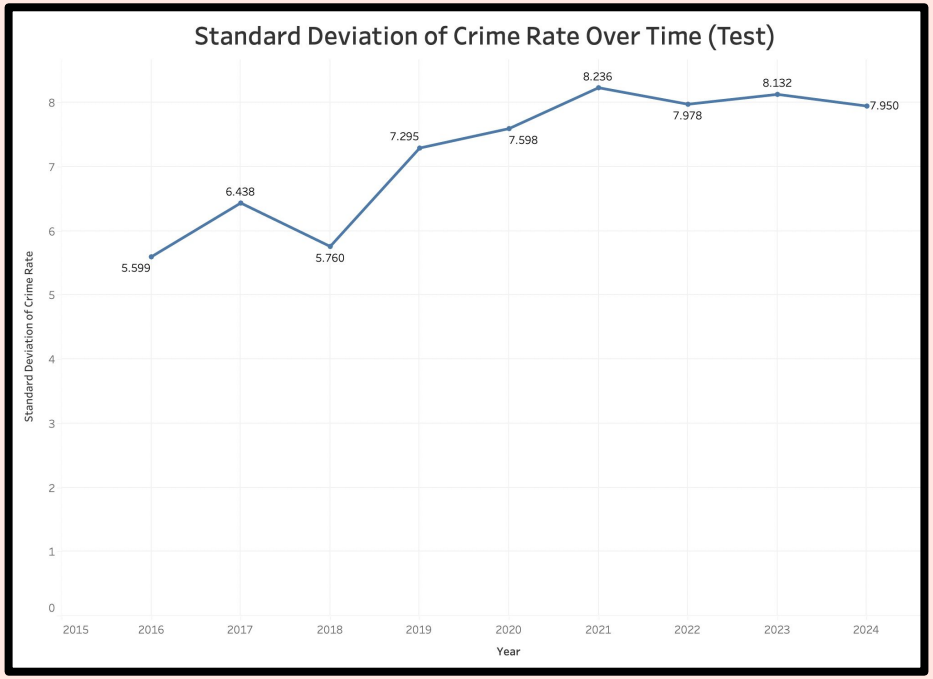


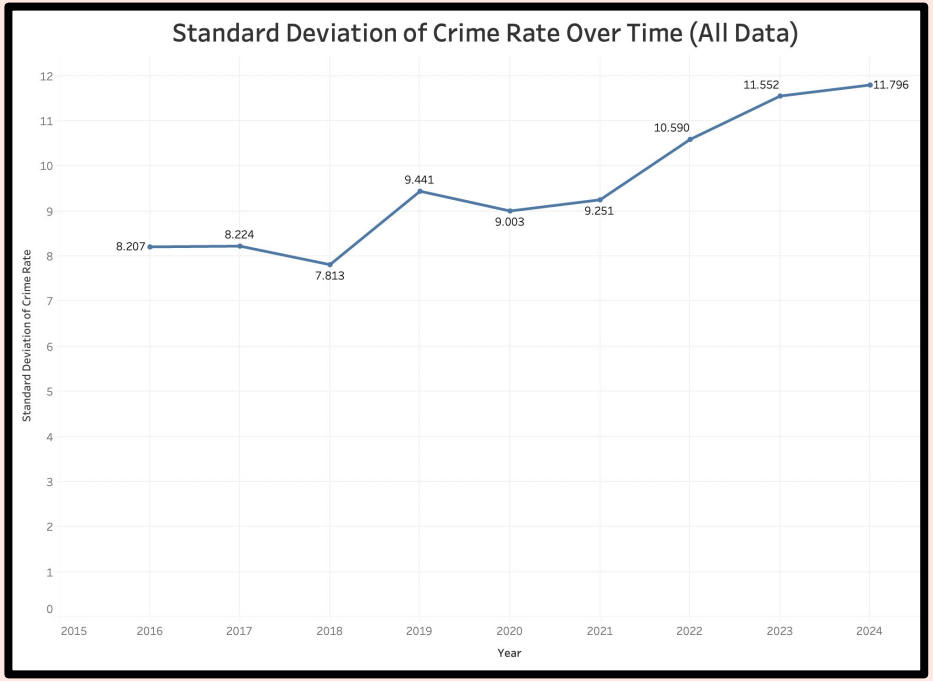


ACCURACY VERSUS CRIME VARIANCE

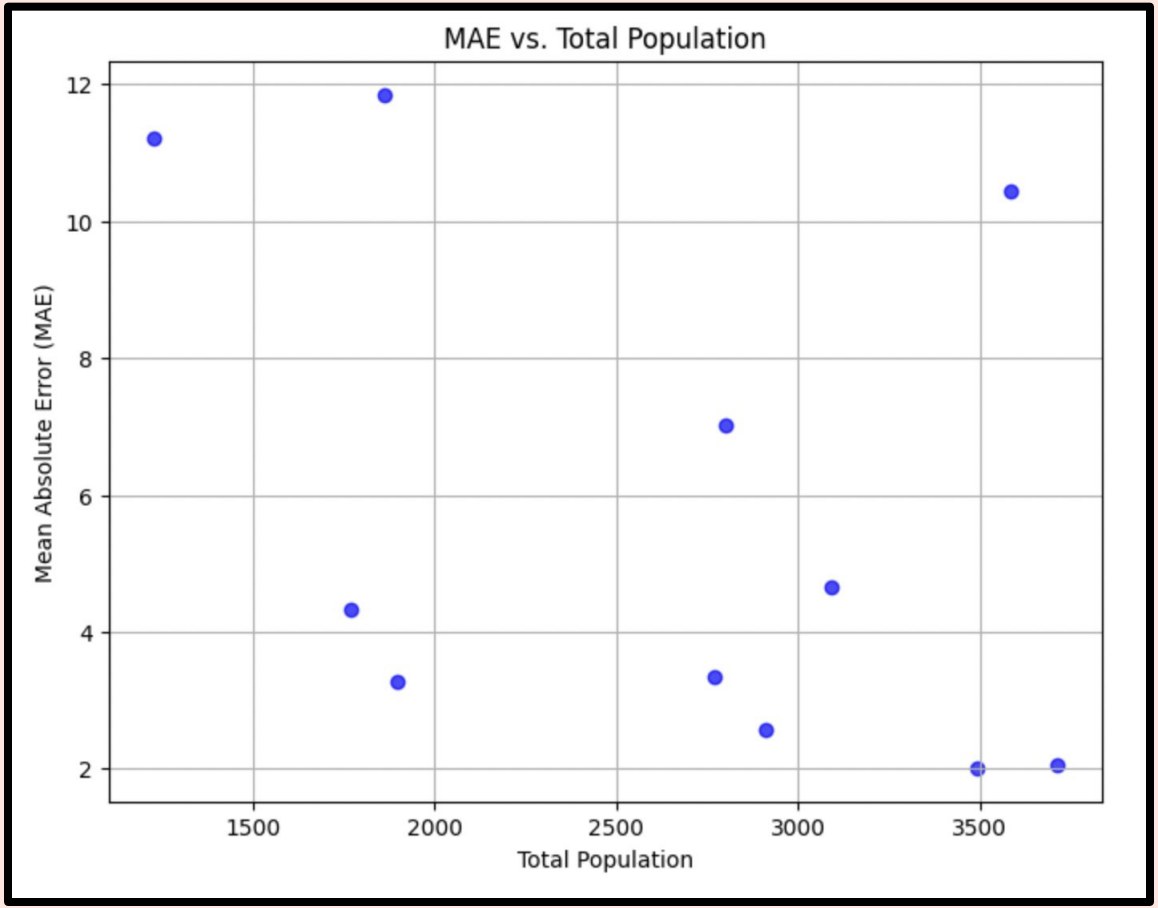


CRIME IS BECOMING MORE VARIABLE

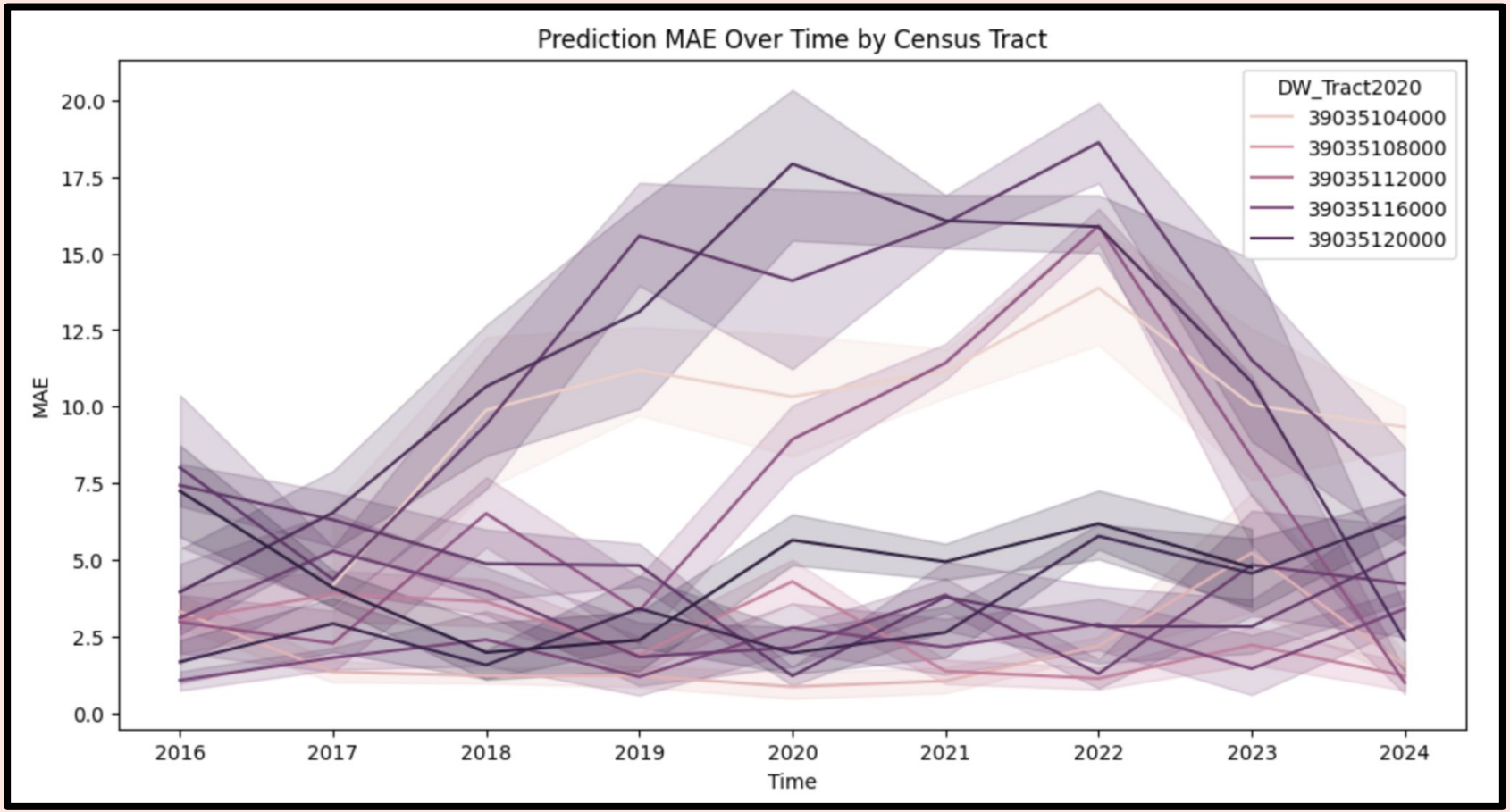




ACCURACY VERSUS POPULATION

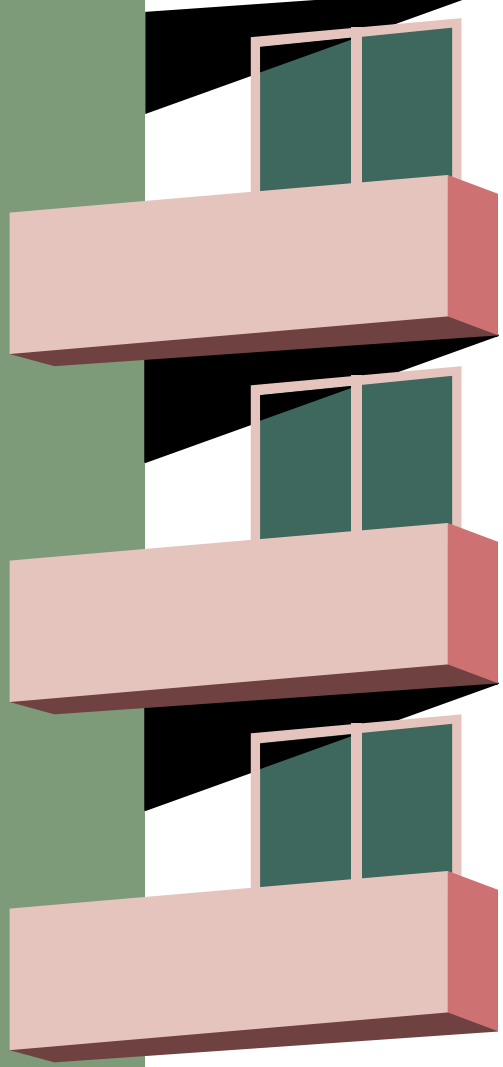


ACCURACY BY TRACT

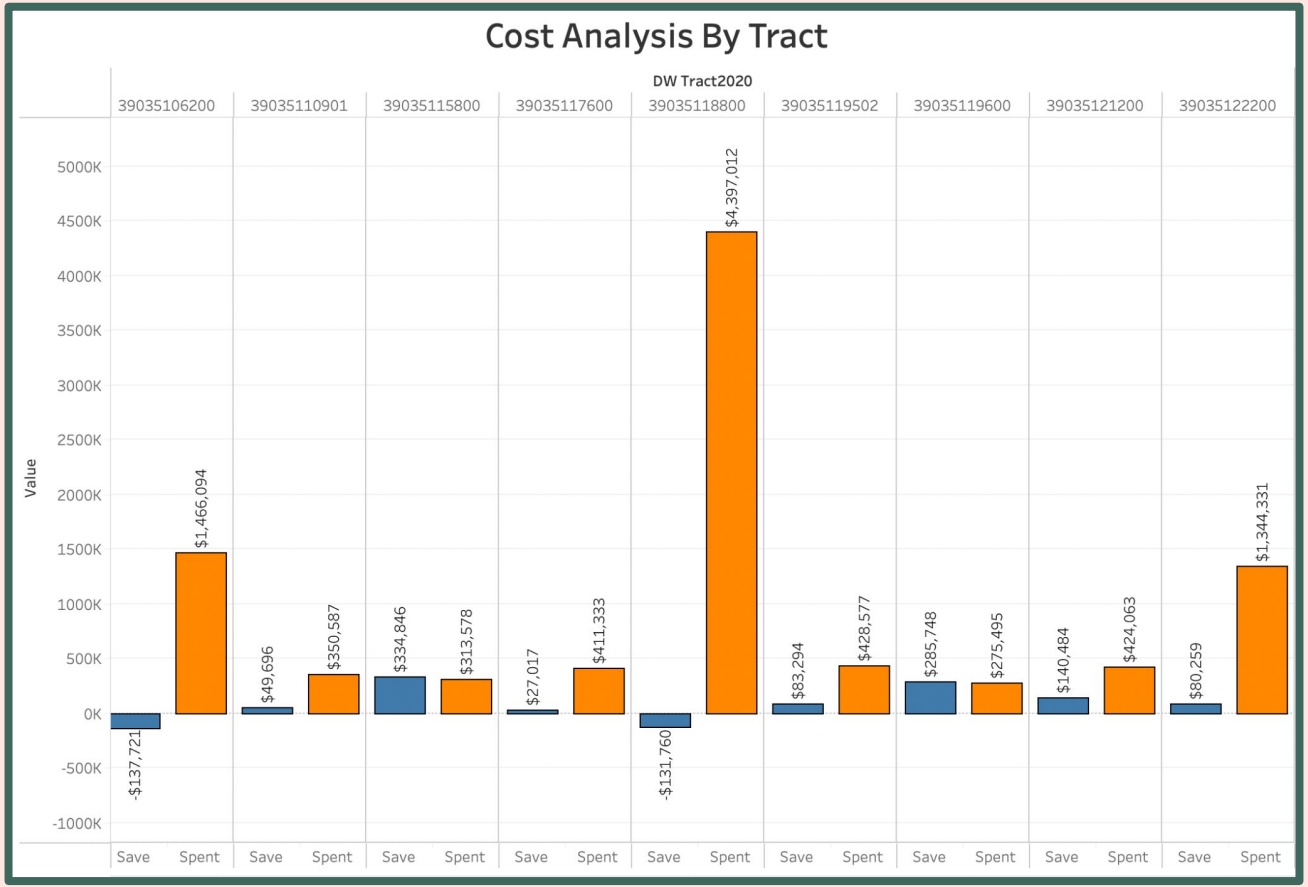


05

IMPLEMENTATION



IMPLEMENTATION VIA COST ANALYSIS



IMPLEMENTATION ON A TRACT BY TRACT BASIS

TRACT 39035117600
(WEST EUCLID)

2023 INVESTMENT:

$411,332

2024 CRIME SAVINGS:

$27,017

TOTAL:

-$384,315

TRACT 39035118800
(LITTLE ITALY)

2023 INVESTMENT:

$439,701

2024 CRIME SAVINGS:

-$131,760

TOTAL:

-$571,461

2023 BREAKDOWN OF INVESTMENT TYPE SAW WEST EUCLID INVEST MORE INTO HOUSING AND DEMOLITION THAN LITTLE ITALY

IMPLEMENTATION ON A TRACT BY TRACT BASIS

TRACT 39035119600
(EAST SHAKER HEIGHTS)

2023 INVESTMENT:

$275,494

2024 CRIME SAVINGS:

$285,748

TOTAL:

$10,254

TRACT 39035110901
(NEWBURGH HEIGHTS)

2023 INVESTMENT:

$350,587

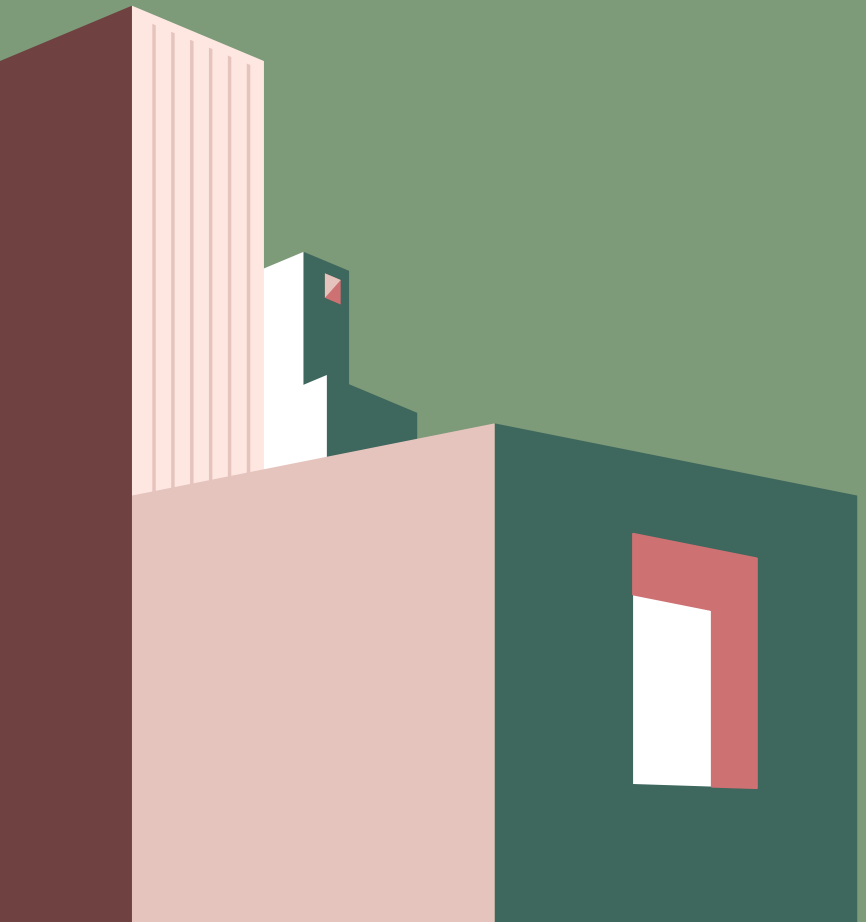
2024 CRIME SAVINGS:

$49,695

TOTAL:

-$300,892

2023 BREAKDOWN OF INVESTMENT TYPE SAW EAST SHAKER HEIGHTS INVEST MORE INTO EDUCATION THAN NEWBURGH HEIGHTS



06

CONCLUSION

LIMITATIONS

- Limited knowledge of marks a “good” figure for savings on crime
- Combining datasets sacrificed some explanatory data (demographic data in crime incidents)
- Census data not available for every year - unreliable population data and no demographics
    - Confounding variables resulted in high variance in model accuracy between tracts
- Data was heavily skewed and difficult to input into a model
- Only building data available - lacking other infrastructure
- Continuous output variable - harder to assess
- Incorporates time - complicates result analysis

TAKEAWAYS

- Crime rate, crime variance, and population appear to affect the model
- The pandemic altered trends for both permit and crime data
  - Further analysis is needed with more data
- Demographics are likely an important factor and should be included
- Investment strategy does appear to influence the economics of census tracts
  - Tracts with similar investment but different breakdowns differed in projected savings
- No apparent correlation between amount invested and amount saved

QUESTIONS