For the Attention of Citadel LLC

A STATISTICAL APPROACH TO THE FIGHT AGAINST CRIME IN PHILADELPHIA

April 2023

Andrea Bizzoni
Ruige Kong
Yiyong Luo
Yutong Wang

Contents

1 Executive Summary **2**
- 1.1 Background . . . 2
- 1.2 Summary of Findings . . . 3

2 Technical Exposition **4**
- 2.1 General Approach . . . 4
- 2.2 Data Collection and Cleaning . . . 5
  - 2.2.1 Description of Datasets . . . 5
  - 2.2.2 Handling Missing Values . . . 5
  - 2.2.3 Data Transformation . . . 5
- 2.3 Exploratory Data Analysis . . . 6
- 2.4 Summary of Methodology . . . 10
- 2.5 Knox Text . . . 11
  - 2.5.1 Time Complexity Analysis . . . 12
  - 2.5.2 Results . . . 13
- 2.6 Spatial Model . . . 13
  - 2.6.1 Results . . . 14
- 2.7 Temporal Model . . . 16
  - 2.7.1 Model Specification . . . 16
  - 2.7.2 Implementation Details . . . 17
  - 2.7.3 Results . . . 18

3 Conclusion **21**

4 Appendix **22**
- 4.1 Theory in Poisson Regression . . . 22
  - 4.1.1 Data Generating Process . . . 22
  - 4.1.2 Estimation . . . 22
  - 4.1.3 Inference . . . 23
- 4.2 Negative Binomial Regression . . . 23
- 4.3 Cost of Crime Analysis . . . 24
- 4.4 Additional Graphs and Visualisations . . . 24

1 Executive Summary

In this report, we analyse crime data related to the city of Philadelphia in the United States. We implement useful data visualisation techniques to identify crime trends across two dimensions - space and time. Subsequently, we develop two rigorous statistical models, a spatial model and a temporal model, with the purpose of predicting future criminal activity across the city. This analysis will better inform the Philadelphia Police Department (PPD) and the City Council as they allocate resources and create policies to tackle urban crime.

1.1 Background

Crime is a significant and long-standing issue in Philadelphia. The city consistently ranks in the top 8% of the most dangerous cities in the United States, scoring well above the national average in terms of criminal activity [8]. The situation worsens when we consider violent crimes such as homicide and assault, where the numbers are strikingly high. With a violent crime rate of 8.11 per 1000 residents, Philadelphia takes first place as the most violent city in the U.S. with a population over 1 million residents [7]. In 2021, a staggering 562 homicides were recorded in Philadelphia, an increase of 13% from the previous year and the highest number since 1960, when police first began tracking killings in the city. Comparatively, New York City, which has a population more than five times that of Philadelphia, recorded 488 homicides in 2021 [10]. Similarly, Los Angeles, with a population more than double that of Philadelphia, recorded 397 homicides in the same year [1].

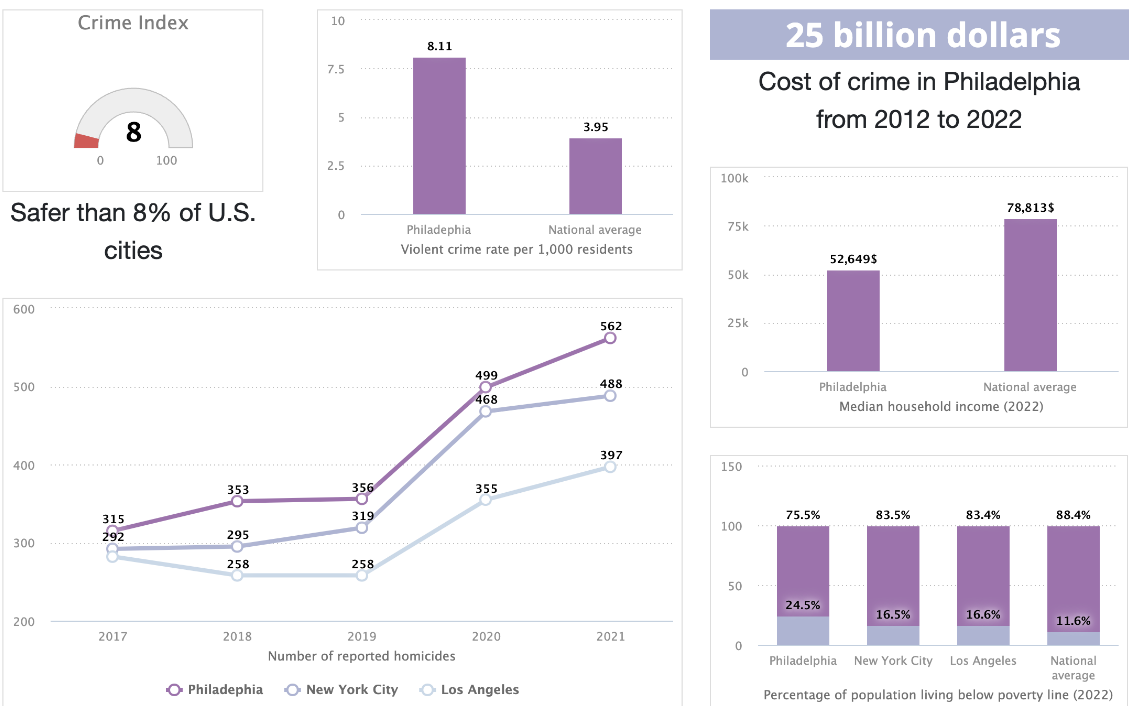


Figure 1: Dashboard of useful statistics regarding crime and wealth in Philadelphia, PA

However, crime does not only hurt the victims. Intense criminal activity has a significant impact on a city's economy and development, as it generally discourages businesses from investing and drives residents away, leading to a decline in property values and tax revenues. The cost of crime is also substantial and stunts economic growth. As estimated by the Cost of Crime Analysis (Appendix 4.3), criminal activities from the past ten years have left the city of Philadelphia short of nearly 25 billion dollars. It should not come as a surprise, therefore, to see that Philadelphia lags behind several other U.S. cities of similar size when it comes to median household income, scoring below the national median. Similar sub-par performance can be observed across a variety of other economic metrics [8].

It is clear, in essence, that crime in Philadelphia represents an important challenge for the PPD and the City Council, as it greatly impacts the safety and well-being of their residents and slows down the economic growth of their city. For this reason, we have decided to focus on analysing crime data from the past ten years (2012-2022).

Our study first utilises a visualisation approach to examine the intensity of crimes in terms of their spatial and temporal dimensions, providing insights into the locations, time, and frequency of criminal activity. Following the visualisations, a hypothesis test is performed to determine the feasibility of analysing the spatial and temporal information independently. Based on the hypothesis test results, two models are constructed to investigate the spatial and temporal effects, respectively, aiming to provide useful guidance to the PPD.

1.2 Summary of Findings

1. **Findings from Data Exploratory Analysis**: The exploratory data analysis (EDA) was conducted to understand the structure and characteristics of the crime data in Philadelphia. The EDA involved various data visualisation techniques, such as:

   - a tract map with a colour gradient showing crime incidence (3)
   - a heatmap of crime incidence at different periods of the day (4)
   - a visualisation of the top 10 crime-ridden tracts (5)

   The findings indicate that, for instance, time of day (6) is a crucial variable for the time-series analysis, whereas days of the week (7) do not show significant variations across the week in terms of crime incidence. The EDA helped identify high crime incidences areas, such as Lower and Upper North Philadelphia, West and Southwest Philadelphia, and Center City. The results provide insights into the patterns and trends in the crime data that guided the development of the time-series model and the spatial model.

2. **Findings from Knox test**: The Knox test is a statistical test that helps to check whether there is a pattern in the arrangement of data over time. It is used to detect any possible temporal dependence or correlation in a sequence of data.

   - We noted that the time for computing the test statistic would increase very fast. Thus, we use Monte-Carlo methods to compute the test statistic for a randomly selected subset, repeat several times, and take the majority vote.
   - For thresholds we select, we do not reject the null hypothesis. This means that there is non-sufficient evidence of temporal dependence or correlation in the data. In other words, the data appears to be randomly distributed over time, and there is no significant pattern or structure in the sequence of data.

3. **Findings from the spatial model**: The spatial model aggregates crimes in each census tract and uses the demographic information in each census tract to explain the number of crimes. The results show that:

   - Among all features in 'philadelphia_population_metrics.csv', only the population size is statistically significant in explaining the crime count. However, statistical significance does not necessarily imply practical significance or causality. It only suggests that there is a statistically significant association between the independent and dependent variables.

- The critical difference between the spatial model and the temporal model is that, here, we do not separate the dataset into the training set and the testing set. This is because we realise without considering the time-dependent nature of the data, it is hard to have high accuracy. It could be understood as a correlation analysis by controlling the other factor's effect.

4. **Findings from the temporal model**:

- It is advised that the manpower of a police station remain flexible when crime counts are predicted to be moderately high, i.e. more than one crime incidence will happen in a 4-hour time frame because crime counts tend to have a higher variation relative to their predictions.
- The temporal model effectively identifies tracts with high daily and 4-hourly crime counts, see Figure 14 and 15 for details.
- At least for the five most dangerous tracts, crime counts exhibit 24-hour seasonality, meaning that the crime count of a tract in a specific time frame (e.g. 4 pm - 8 pm time frame) is significantly associated with the count on the previous day.
- For the five most dangerous tracts, crime counts are negatively related to past counts from 12 hours prior.

2 Technical Exposition

In this section, we provide a detailed technical exposition of both our temporal and spatial analysis of crime incidence in Philadelphia. We first present a series of informative visualisations that illustrate the patterns and trends we observed in the data. We then describe the mathematical foundations and algorithms that underpin our model, including the derivation of key equations and the implementation of key procedures. Finally, we detail our evaluation metrics and present our results, which demonstrate the effectiveness of our approach.

2.1 General Approach

The aim of this study is to investigate spatial-temporal patterns of crime in Philadelphia using a hypothesis-testing framework. Our general approach consists of two main steps: 1) application of the Knox test to detect spatial-temporal clustering of crimes, and 2) application of an autoregressive model for count time series with excess zeros [12], to model the time evolution of crime counts.

In the first step, we employ the Knox test, a commonly used spatial statistic, to determine if there are significant spatial-temporal clusters of crime events in the study area. This test is based on the null hypothesis that the spatial and temporal distribution of crime events are independent. A rejection of this null hypothesis indicates that there are significant spatial-temporal clusters of crime.

In the second step, we apply the autoregressive model for counting time series with excess zeros to capture the temporal dependence of crime counts. This model is suitable for count data with over-dispersion and excess zeros, which are common features of crime data. We estimate the model parameters and conduct hypothesis testing to evaluate the goodness of fit of the model and assess its ability to explain the observed crime patterns.

Overall, this approach allows us to investigate the spatial-temporal patterns of crime in Philadelphia and identify any underlying temporal dynamics that may be contributing to the observed patterns. By using hypothesis testing, we aim to provide a rigorous and objective analysis of the data, which can inform future policy and decision-making in the field of crime prevention and control.

2.2 Data Collection and Cleaning

In this part, we introduce how we pre-process the data and perform exploratory data analysis, along with some simple visualisations to guide the next step of the analysis.

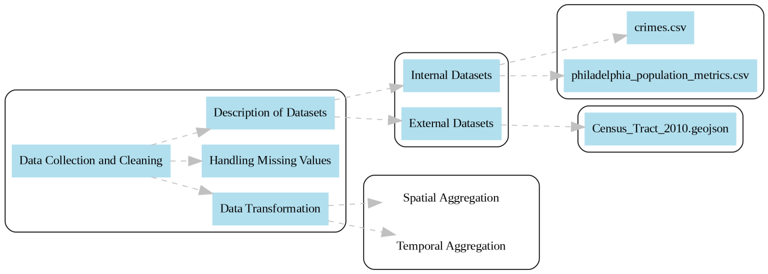


Figure 2: An overview diagram of the data preparation process

2.2.1 Description of Datasets

For this analysis, we used both internal datasets (as provided by the Datathon Correlation One Team) and external sources (mostly from OpenDataPhilly):

1. Internal Datasets
   - **crimes.csv** - this dataset contains the full record of crimes committed between 2012-2022 in Philadelphia. This information is the primary source of information used in our study.
   - **philadelphia_population_metrics.csv** - this dataset contains the demographic information for Philadelphia. As this is quite invariant in time, the dataset is primarily used in the spatial marginal modelling section of our study.
2. External Datasets
   - **Census_Tract_2010.geojson** - this dataset contains the shapefile and boundaries for the census tracts in Philadelphia. This is used for data visualisation purposes and mapping crime intensity in different tracks.

2.2.2 Handling Missing Values

In handling missing data, we adopted the listwise deletion technique, which involves removing any observation that has missing values in any variable of interest. While this method is easy to implement, it has some limitations. One major drawback is the potential for data loss, especially if the missing pattern is related to specific covariates. Consequently, using listwise deletion can result in biased estimates for model parameters, making it crucial to evaluate the randomness of the missing pattern.

In this study, we removed all data related to homicides since the time variable, which is a critical component for our spatial-temporal analysis, was missing. Consequently, we excluded this type of crime from our analysis to avoid any potential distortion in our findings.

2.2.3 Data Transformation

In order to gain insights from our spatial-temporal crime data, we employed data aggregation in our analysis. Aggregation involves combining the raw data into larger units, such as by averaging values over time intervals or grouping data by geographic regions. We chose to aggregate our data at the tract level, as we believe that this is the most appropriate unit for our analysis due to its geographical and demographic homogeneity. We chose to aggregate data both over time and space.

Aggregating data over time can have several benefits, including reducing noise and variability in the data and making it easier to identify larger-scale patterns and trends over time. It can also help to reveal longer-term patterns and trends that may not be immediately apparent in the raw data. However, it is essential to note that aggregating data over time can also lead to a loss of information, mainly if the aggregation interval is too large. Furthermore, it may mask short-term patterns and trends that could be important for understanding the data.

Given the analysis above, it is vital to consider the trade-offs between reducing noise and variability in the data and losing important information or masking short-term or local patterns and trends. The detailed procedure is shown below:

- For the **temporal dimension**: we select four hours as a smallest time interval. This is because the patrol arrangement is changed to another group every 8 hours on weekdays and every 12 hours at weekends [2]. We wanted to keep each time unit within the same patrol circulation. As our results show, '4 hours' worked very well, while we also tried to use '1 hour' as the unit, in which case we did not have enough data points.
- For the **spatial dimension**: we select 'census tract' as the smallest region (map shown in Appendix 4.4). We also notice that there are 'census block' level data and 'census block group' level data. The reasons we choose 'census tract' are because:
  1. Tracts tend to be larger than blocks but smaller than districts. It strikes a balance between having enough granularity to capture meaningful spatial variation in crime and having enough observations to estimate reliable models. In fact, for practical purposes, it can help police officers navigate crime more precisely.
  2. Furthermore, tracts, in our case, are more homogeneous in terms of demographic and socioeconomic characteristics than blocks but less homogeneous than districts. This is important for modelling, as it allows for the inclusion of spatially-varying covariates that are relevant to crime (e.g., poverty rates, racial composition).

2.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in any data science project as it helps us understand the structure and characteristics of the data.

In our study, we aim to model crime incidence in different tracts of Philadelphia using a time-series approach and a spatial marginal model. To achieve this, we have conducted extensive EDA and visualisation on the dataset to analyse the spatial and temporal crime distribution in Philadelphia. Our motivation behind conducting such a thorough EDA and visualisation is to identify patterns and trends in the data that can guide us in developing the most promising approach for predicting crime incidence.

Through this process, we have determined that time of day is a critical variable in our time-series analysis, whereas days of the week did not show significant variation across the week in terms of crime incidence. The insights we have gained from EDA and visualisation have helped us develop a robust time-series model and a spatial model that considers relevant demographic and geographical information.

In particular, we performed the following data visualisation:

1. **Tract map with colour gradient showing crime incidence**(3): This map shows the crime incidence for each tract in Philadelphia using a colour gradient. The darker the colour, the higher the crime incidence. This map provides a visual representation of the crime incidence patterns across the city, helping us to identify the high crime incidence areas. In particular, we can see that crime

incidence is very much concentrated in Lower and Upper North Philadelphia, West and Southwest Philadelphia, and Center City.

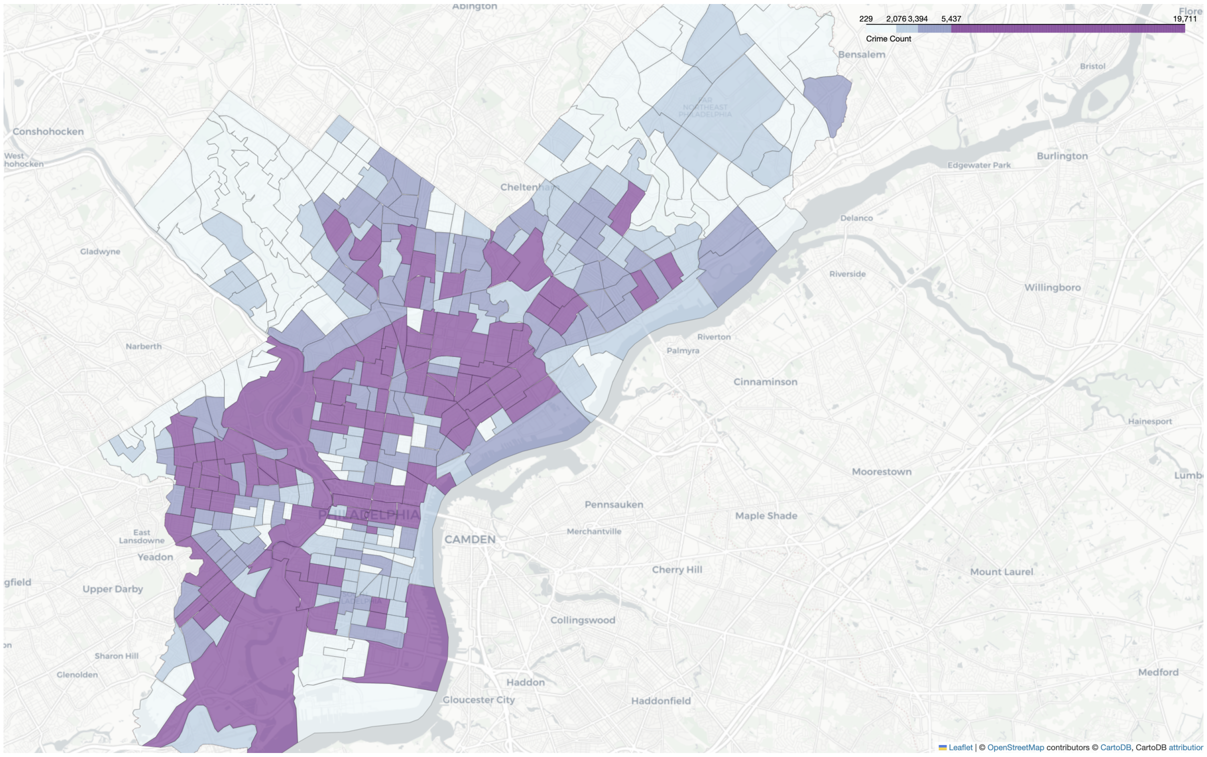
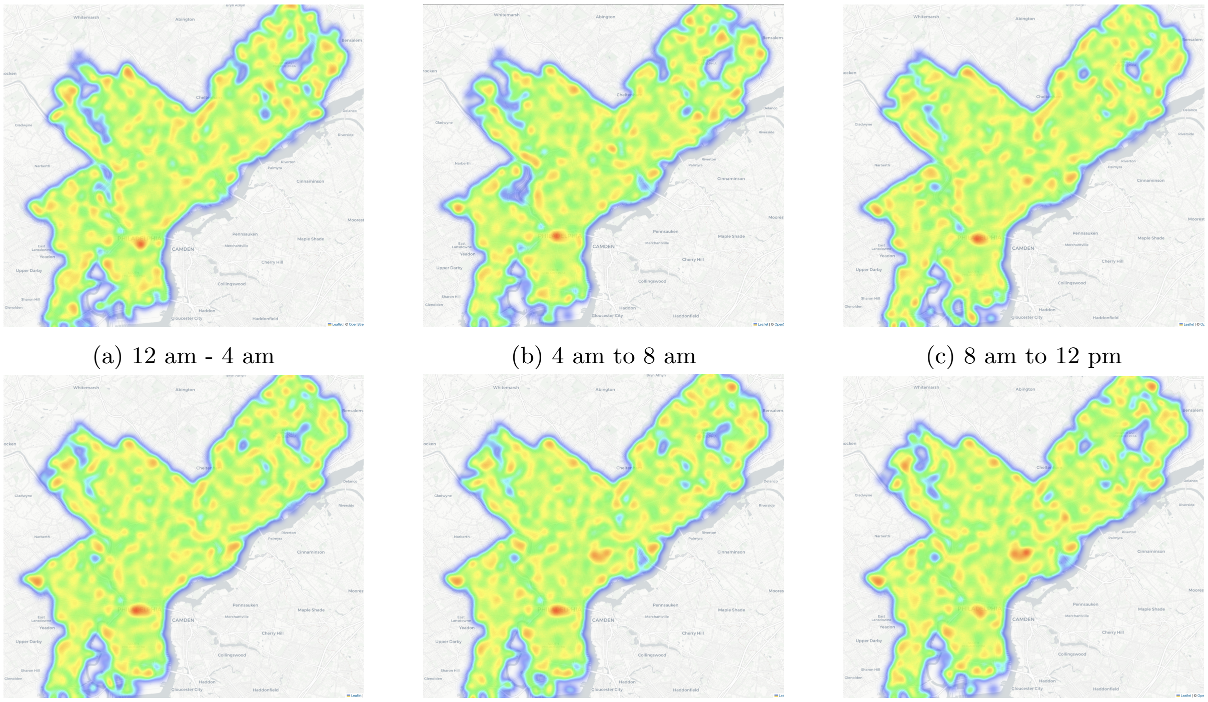


Figure 3: A map with the crime intensity in different tracts of Philadelphia. The colour gradient shows the differential frequency of crime incidence. The quantiles are 0%, 25%, 50%, 75%, 100%

2. **Heat map of crime incidence at different periods of the day**(4): This heat map shows the crime incidence at different periods of the day (early morning - 4 am to 8 am, morning - 8 am to 12 pm, noon - 12 pm to 4 pm, afternoon - 4 pm to 8 pm, evening - 8 pm to 12 am, midnight - 12 am to 4 am) across the city of Philadelphia. It provides a visual representation of the time-of-day patterns of crime incidence across the city and helps us to identify any trends or patterns in the crime data. In this case, we notice that some crime hotspots near Center City and West Philadelphia change intensity drastically throughout different times of the day.



(a) 12 am - 4 am

(b) 4 am to 8 am

(c) 8 am to 12 pm

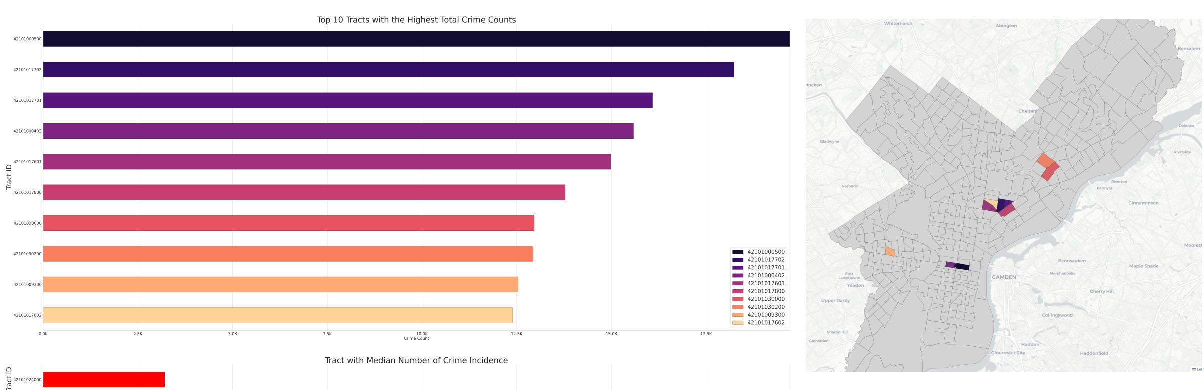
(d) 12 pm to 4 pm

(e) 4 pm to 8 pm

(f) 8 pm to 12 am

Figure 4: Heatmaps showing the evolution of total crime frequency throughout different times of the day.

3. **Visualisation of the top 10 crime-ridden tracts**(5): The combination of histogram and tract map help us to identify the tracts that have the highest crime incidence and to compare the distribution of crime incidence across these tracts. Compared with the crime heatmaps, it is apparent that these top crime-ridden tracts are concentrated near crime hotspots in Center City and Lower and Upper North Philadelphia.



(a) Histogram of the top 10 most frequent crime-ridden tracts

(b) Tract map with the top 10 most frequent crime-ridden tracts

Figure 5: A histogram and a geographical visualisation of the top 10 most crime-ridden tracts in Philadelphia. The histogram and the tract are colour-matched for ease of navigating the crime-ridden tracts' locations

4. **Histograms of crime distribution throughout the day**(6): These histograms show the distribution of crime incidence throughout different time frames of the day across the city of Philadelphia. They help us to identify any trends or patterns in the time-of-day distribution of crime incidence, which also facilitates the time-series modelling of crime incidence.

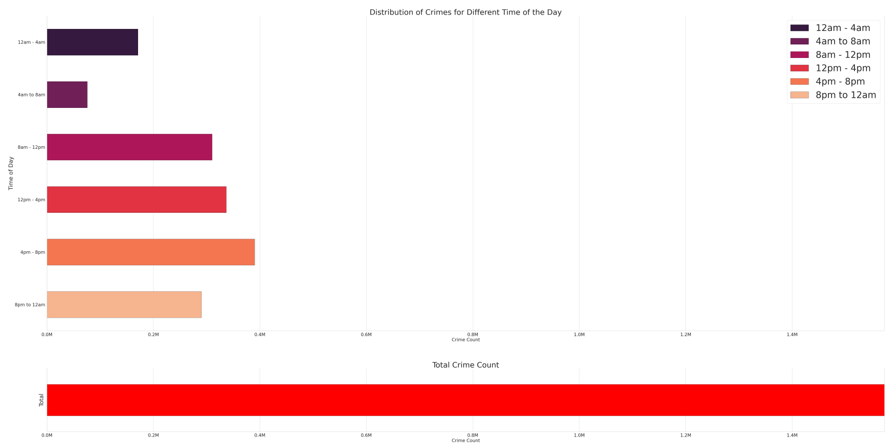


Figure 6: Histogram displaying the distribution of total crime incidence throughout different periods of the day.

5. **Histograms of crime distribution at different days of the week**(7): These histograms show the distribution of crime incidence across different days of the week (Monday through Sunday) across the city of Philadelphia. They help us identify trends or patterns in the day-of-week distribution of crime incidence.

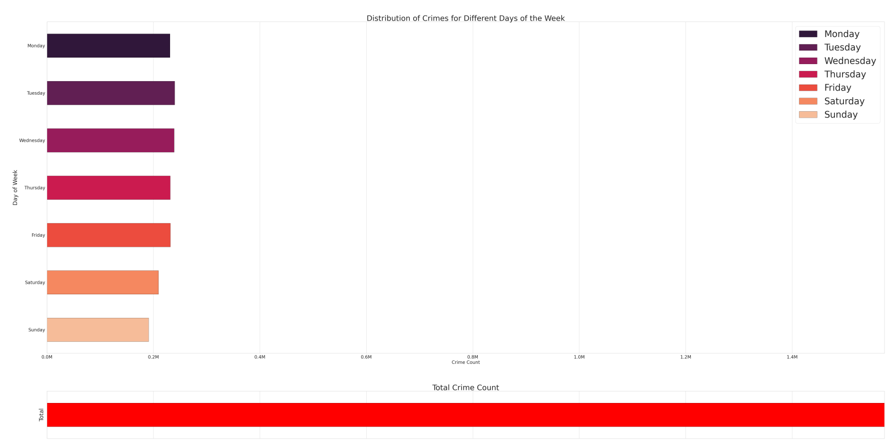


Figure 7: Histogram displaying the distribution of total crime incidence throughout different days of the week.

6. **Correlation Analysis for Crime Counts and Demographic information**(1):

Variable Name	Correlation Coefficient
crime_count	1.000
count_all_races_ethnicities	0.344
percent_white_nh	-0.417
percent_black_nh	0.260
percent_asian_nh	-0.059
percent_hispanic	0.318
shape__area	-0.057
shape__length	-0.108

Table 1: Correlation between Crime Counts and Demographic Information

These graphs and tables help us to identify patterns and trends in the crime data and to visualise the distribution of crime incidence across time and space. This helps us to determine which variables are important in predicting crime incidence, which in turn guides the development of our time-series model and spatial marginal model. The visualisations help us to discard or incorporate certain variables based on the significance of their impact on crime incidence, leading us to the most promising approach for predicting and modelling crime incidence in Philadelphia. In our case, we noticed that, in contrast to our expectation, the total crime incidence actually does not vary significantly throughout different days of the week. The day-of-the-week variable is, therefore, not included in our time-series analysis.

2.4 Summary of Methodology

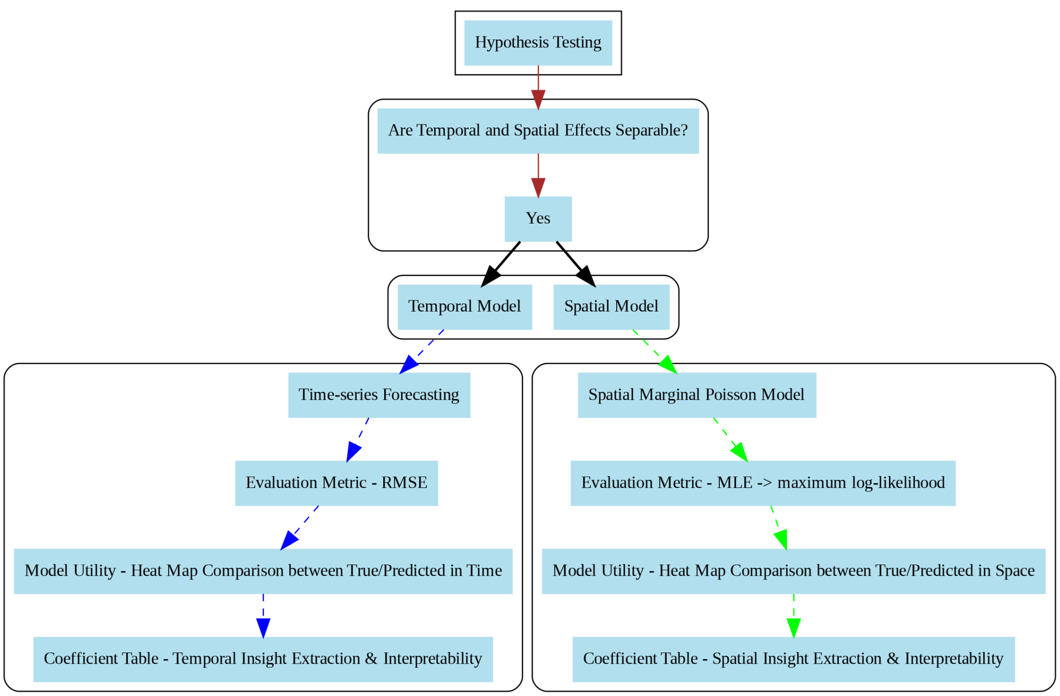


Figure 8: An overview diagram of our methodology and model design

Our goal is to extract the probability of a crime happening at a given position and time. From our datasets, crime.csv resembles typical spatial-temporal data. Each row of the 'crimes.csv' contains:

- a crime event
- the position it happened (tract ID, longitude and latitude)
- the dispatch date and time of the police officers

For a position (tract s in this study), we have:

Assumption 1. *The counts of crimes within time period* $[t, t+\Delta t]$*, denote as* $y_{s,t}$*, follow a time-varying distribution with parameter* $\boldsymbol{\theta}_{s,t}$

Note that we have selected the time unit Δt to be 4 hours.

Under this model setting, we are interested in:

1. Whether there is any spatial-temporal interaction.

2. How $\boldsymbol{\theta}_{s,t}$ varies with respect to s and t.

To answer the first question, we apply the Knox test; while for the second, we estimate parameters and do hypothesis testing to see whether it is significant.

This is the restriction we put on the function class. In statistical learning theory, if we do not put enough restrictions on the feasible function class, then the fitted model would suffer from overfitting with high probability. It is also reasonable to have the assumption.

2.5 Knox Text

The Knox test [4] is originally developed to detect spatial–temporal interaction for disease events. It is a statistical method used to detect spatio-temporal interaction in a set of geo-coded events with corresponding time stamps. Due to the similarity of the data structure between disease events data and crime events data, it is also applicable in crime data analysis.

The test compares the observed spatial and temporal distances between pairs of events with the expected distances if the events were randomly distributed. The null hypothesis of the Knox test is that the events are randomly distributed, while the alternative hypothesis is that there is spatio-temporal interaction.

For a dataset containing n number of crime events in the study area, there are $\frac{n(n-1)}{2}$ possible pairs of events to test. For each pair, the test checks whether the spatial and temporal distances between the events fall within the specified critical values [9], which define the neighbourhood around each event. By comparing the observed number of pairs that fall within the neighbourhood to the expected number under random distribution, the test can detect spatio-temporal interaction between crime events.

The test statistic X can be computed as follows:

$$X = \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mathbb{I}\{l_{i,j}^s < \delta\} \mathbb{I}\{l_{i,j}^t < \tau\},$$

where $l_{i,j}^s$ represents the space distance between crime i and crime j, and $l_{i,j}^t$ represents the space distance between crime i and crime j. δ and τ represent the critical space distance and critical time distance, which allow us to choose. Although the subjectivity in the selection of thresholds δ and τ can not be abandoned, it could also be meaningful. For example, [3] suggest that repeat burglaries tend to cluster within one to 2 months and within $300 - 400$ meters of a prior burglary.

If the Knox test statistic exceeds the critical value, it indicates that the observed number of pairs that fall within the neighbourhood is greater than expected under random distribution, providing evidence

Table 2: Contingency Table for Knox Testing

		SPACE: Close	SPACE: Not Close
Time	Close	Space and Time	Time Only
	Not Close	Space Only	Not Close

of spatio-temporal interaction between crime events. In this case, the null hypothesis of a random distribution of crime events can be rejected in favour of the alternative hypothesis of spatio-temporal interaction.

By setting the threshold δ and τ, we apply the contingency table (as table 2 shown) chi-squared test:

H_0: There is no association between spatial dimension and temporal dimension. Versus

H_1: There is some association between spatial and temporal dimensions.

Under the null, we have:

$$\sum_{r=1}^{2}\sum_{c=1}^{2}\frac{(E_{r,c}-O_{r,c})^2}{E_{r,c}} \sim \chi^2(1).$$

For a general contingency table, the summation follows $\chi^2_{(r-1)\times(c-1)}$. In this special case, the contingency table is 2-by-2, thus $(r-1)\times(c-1)=1$. Following the standard approach: (1) find the critical value for a given significant level, (2) compute the above statistic, and (3) compare the statistic with the critical value, we could finish the hypothesis testing.

2.5.1 Time Complexity Analysis

To calculate the test statistic, we need to calculate the distance between each pair of crime events and compare them with the threshold. Thus, if the sample size is n, then the time complexity is $O(n^2)$.

We have $n > 1.5$ million, hence it is not possible to calculate the exact test statistic in 7 days. To encounter this problem, we compute the test statistic for a random subset of all rows. The decision for rejection of the null based on the subset could be different from the decision based on the entire sample with a certain probability. If we could ensure the probability that these two decisions agree is larger than $\frac{1}{2}$, then we could repeatedly select a random subset several times, and get a sequence of decisions. Eventually, take a majority vote in the sequence of decisions.

In this alternative way, there are several flexible parameters: (1) the size of the small sample m, and (2) the times we repeat k. If we denote the true decision as R, and the sub-sample decision as R_i, rejection as 1, none rejection as 0, then:

$$\begin{aligned}
\text{P(sub-sample decision agrees with true decision—true is not to reject)} &= \text{P}(R_i = R|R=0) = f(n,m) \\
\text{P(sub-sample decision agrees with true decision—true is rejection)} &= \text{P}(R_i = R|R=1) = g(n,m).
\end{aligned}$$

It is essential to see that the population-level probability is fixed, once given n and m. As a result, if we repeat the procedure k times, and each time we i.i.d. sampling from the set of IDs of rows, and take a majority vote:

$$
\begin{aligned}
& \mathrm{P}(\text{majority vote of sub-sample decisions agrees with true decision}) \\
= \; & \mathrm{P}\Big(\mathbb{I}\{R_1 + \cdots + R_n \geq \frac{n}{2}\} = R\Big) \\
= \; & \mathrm{P}\Big(\mathbb{I}\{R_1 + \cdots + R_n \geq \frac{n}{2}\} = R | R = 1\Big)\mathrm{P}(R = 1) + \mathrm{P}\Big(\mathbb{I}\{R_1 + \cdots + R_n \geq \frac{n}{2}\} = R | R = 0\Big)\mathrm{P}(R = 0) \\
= \; & \sum_{i=0}^{[\frac{n}{2}]+1} \binom{n}{i} f^i (1-f)^{n-i} \phi + \sum_{i=0}^{[\frac{n}{2}]+1} \binom{n}{i} g^i (1-g)^{n-i} (1-\phi).
\end{aligned}
$$

Through calculation, $\forall \varepsilon > 0$, we could find m, k such that:

$$
\mathrm{P}\Big(\mathbb{I}\{R_1 + \cdots + R_n \geq \frac{n}{2}\} = R\Big) \geq 1 - \varepsilon.
$$

Note that this approach yields time complexity $O(km^2)$. This shows that, if we want a ε-approximation result, we could set km^2 as the objective function and solve a minimisation problem to minimise the time complexity for the approximation algorithm.

2.5.2 Results

As in [3], we select $\delta = 400$ meters, and $\tau = 60$ days. We set $k = 10$, and $m = 150$. The p-values in 10 times are: $(0, 0.740, 1.000, 0.573, 1, 1, 0.813, 0.949, 1, 1)$. This is the evidence that we do not reject the null. We also tried different (δ, τ) pairs, the decisions are not to reject the null.

We understand that if we want to conclude there is no association (interaction) between spatial dimension and temporal dimension, we need to check for all (δ, τ) pairs. Given the limited time, it is not possible to do such an exhaustive search. We could at least say there is no association at some level, and this gives us the reason to separately look at the spatial marginal distribution and the temporal marginal distribution.

2.6 Spatial Model

Spatial marginal study means we do not look at the effect of the temporal dimension. Concretely, we 'average out' time. Since the data contains several types of crimes from 2013 to 2022, we aggregated them to their constituent census tract, and aggregate each year. By doing this, we obtain a 384×10 matrix $Tract\_Year$ where each row represents a tract, and each column represents a year. Entry $Tract\_Year_{s,t}$ is the count of crimes that happened in s-th tract and t-th year. From this matrix $Tract\_Year$, we construct the year average count of crimes vector $\{Tract\_Year\_Ave_s\}$ in each tract $s = 1, \cdots, 384$ by taking an arithmetic average over 10 years (each row).

Next, we merged this vector with demographic information in 'philadelphia_population_metrics.csv' for tracts. Now the data looks like this: each row represents a tract, and each column features including: (1) *[crime_type]_mean*, the yearly average crime counts across 10 years, (2) *geography_name*, the unique id for that tract, (3) *count_all_races_ethnicities*, (4) *percent_white_nh*, (5) *percent_black_nh*, (6) *percent_asian_nh*, (7) *percent_hispanic*, (8) *shape__area* and (9) *shape__length*.

Given that the dependent variable y_i is crime counts, We use (3) - (9) as independent variables x_i to estimate a Poisson regression model for each type of crime. This model helps to understand the average effects of demographic characteristics.

$$
\log y_i = \alpha + x_i^\intercal \theta + \varepsilon_i, \tag{2.1}
$$

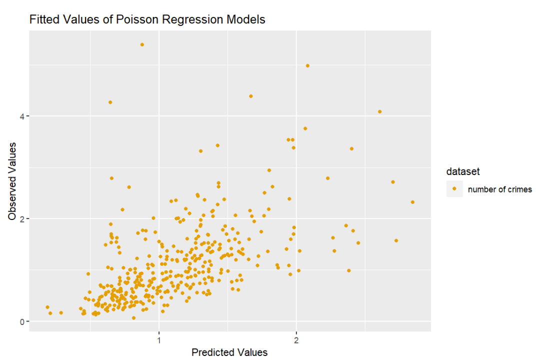
where $\alpha \in \mathbb{R}$ is the intercept, and $\theta = (\theta_1, \cdots, \theta_7)^\intercal \in \mathbb{R}^7$ is the marginal effect of each feature, with 7 being the number of features.

We estimate the parameter by using maximum likelihood estimation (MLE) since it is a parametric model. MLE has lots of nice properties such as consistent and asymptotic normal, while the latter allows us to use the normal distribution to construct confidence intervals and do hypothesis testing. Details of estimators can be seen in Appendix.

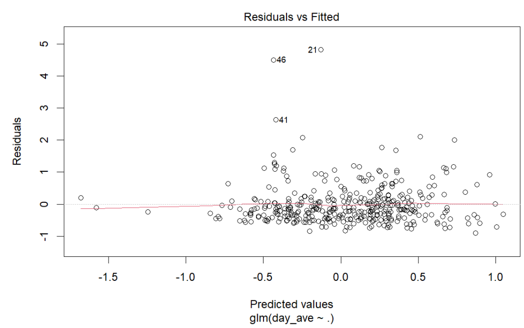
2.6.1 Results

Table 3: Summary Statistic

Statistic	N	Mean	St. Dev.	Min	Max
count_all_races_ethnicities	381	4,005.265	1,679.630	2	8,322
percent_white_nh	381	37.876	33.121	0.000	100.000
percent_black_nh	381	42.927	35.370	0.000	96.130
percent_asian_nh	381	5.926	7.741	0.000	63.160
percent_hispanic	381	10.884	16.235	0.000	91.226
shape__area	381	1,631,449.000	2,544,783.000	179,724.000	34,980,023.000
shape__length	381	5,533.390	4,071.063	1,726.925	40,164.960
day_ave	381	1.125	0.798	0.063	5.395

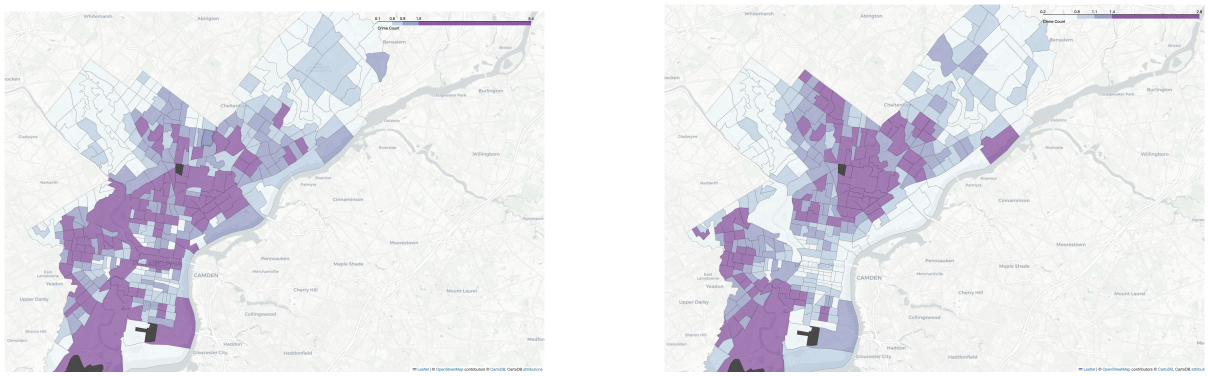


(a) Fitted Value of Spatial Marginal Poisson Regression



(b) Residuals and Fitted Value

Figure 9: Results from Spatial Marginal Model Fitting



(a) True Data (b) Fitted Value

Figure 10: Heatmaps of true (left) and fitted (right) average daily crime count

Table 4: Spatial Marginal Regression Results

	Dependent variable:
	day_ave
count_all_races_ethnicities	0.0001***
	(0.00003)
percent_white_nh	0.005
	(0.057)
percent_black_nh	0.013
	(0.057)
percent_asian_nh	0.011
	(0.060)
percent_hispanic	0.018
	(0.056)
shape__area	0.00000
	(0.00000)
shape__length	−0.00004
	(0.00004)
Constant	−1.348
	(5.592)
Observations	381
Log Likelihood	-Inf.000
Akaike Inf. Crit.	Inf.000
Note:	*p<0.1; **p<0.05; ***p<0.01

The first independent variable, "count_all_races_ethnicities," has a significant positive coefficient (0.0001) at the 1% significance level, which suggests that an increase in the count of all races and ethnicities (in other words, the population size) is associated with an increase in the dependent variable.

The other independent variables, "percent_white_nh," "percent_black_nh," "percent_asian_nh," "percent_hispanic," "shape__area," and "shape__length" do not have significant coefficients at the 10% or 5% significance levels.

This is to say that the population size could explain a proportion of the total crime count. However, we must note that statistical significance does not necessarily imply practical significance or causality. It only suggests that there is a statistically significant association between the independent and dependent variables. Therefore, additional analyses, such as causal inference methods or effect size calculations, may be needed to fully understand the practical implications of these results.

The important difference between the spatial model and the temporal model is that here we do not separate the dataset into the training set and the testing set. This is because we realise without considering the time-dependent nature of the data, it is hard to have high accuracy. It could be understood as a correlation analysis by controlling the other factor's effect.

2.7 Temporal Model

2.7.1 Model Specification

A Temporal Marginal Model posits that only past crime counts in a given area, referred to as the tract identity denoted as s in this study, impact the current count. A time series model is constructed to capture the crime count pattern for each tract. While the traditional time series models, such as the autoregressive integrated moving average model, are suitable for continuous random variables, these models are inappropriate for count time series, e.g. crime counts in our case. Thus, we narrow down model choices to count time series models, such as tscount model [5], which is a model applies generalized linear regression to time series, and Zero-Inflated Markov (ZIM) model[12].
To select the most suitable model, we further analyse the characteristic of crime counts using visualisations. Figure 11 displays the histogram of crime counts in a tract with the highest crime intensity, shown by Figure 5a. The histogram revealed a large proportion of time points when no crime was committed in the selected tract. The presence of excess zeros in the crime count data is referred to as zero inflation, and this characteristic has led us to consider the use of the ZIM model, which was specifically developed for handling this type of data.

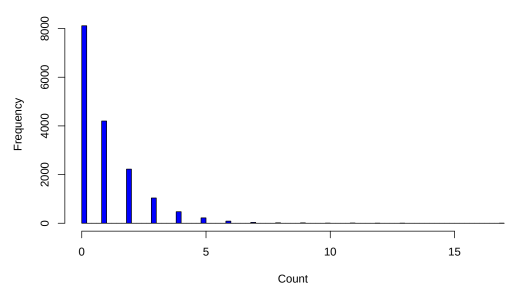


Figure 11: Histogram of crime counts in Tract 42101000500.

For the Tract s, the ZIM model treats the crime count at time t, $y_{s,t}$, as a random variable with probability mass function (pmf):

$$f(y_{s,t}|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \begin{cases} \mathbb{1} y_{s,t} = 0 & \text{with probability } \omega_{s,t}, \\ g(y_{s,t}|\mathcal{F}_{t-1};\boldsymbol{\beta}) & \text{with probability } 1-\omega_{s,t}. \end{cases} \tag{2.2}$$

This means the $y_{s,t}$ is assigned to zero with probability $\omega_{s,t}$, and otherwise, follows a distribution with pmf $g(y_{s,t}|\mathcal{F}_{t-1};\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ and parameters related to $\omega_{s,t}$ forms the unknown parameter set $\boldsymbol{\theta}$.

There are two choices for the pmf:

$$g(y_{s,t}|\mathcal{F}_{t-1};\boldsymbol{\beta}) = \exp -\lambda_{s,t}\lambda_{s,t}^{y_{s,t}}/y_{s,t}!, \tag{2.3}$$

$$g(y_{s,t}|\mathcal{F}_{t-1};\boldsymbol{\beta}) = \frac{\Gamma(y_{s,t}+\alpha_s^{-1})}{\Gamma(y_{s,t})\Gamma(\alpha_s^{-1})}\Big(\frac{\alpha_s^{-1}}{\alpha_s^{-1}+\lambda_{s,t}}\Big)^{\alpha_s^{-1}}\Big(\frac{\lambda_{s,t}}{\alpha_s^{-1}+\lambda_{s,t}}\Big)^{y_{s,t}}. \tag{2.4}$$

The corresponding distribution of Equation 2.3 is Poisson($\lambda_{s,t}$), and Equation 2.4 leads to a Negative Binomial distribution with mean $\lambda_{s,t}$ and variance $\lambda_{s,t} + \alpha\lambda_{s,t}^2$.
For both cases, $\lambda_{s,t}$ and $\omega_{s,t}$ are functions of past crime counts:

$$\lambda_{s,t} = \exp\left(\beta_0 + \beta_1 y_{s,t-1} + \cdots + \beta_{k_1} y_{s,t-k_1}\right) \tag{2.5}$$

$$\omega_{s,t} = \exp\left(\tilde{\beta}_0 + \tilde{\beta}_1 y_{s,t-1} + \cdots + \tilde{\beta}_{k_2} y_{s,t-k_2}\right) \tag{2.6}$$

When fitting the model, both cases will be considered, and we can set a hypothesis test to select which distribution to use. The null hypothesis is that the Poisson distribution is favoured, and similarly, the alternative hypothesis is that the model prefers Negative Binomial distributions. This hypothesis test can be simplified to whether the mean of a sample equals the variance because Poisson distribution has the assumption that means should be equal to the variance.

2.7.2 Implementation Details

For a tract s, we split the time series to training and testing data set with a ratio of 3:1 and fit the ZIM for tract s using the time series $(y_{s,1},\ldots,y_{s,T_{\text{train}}})$, where T_{train} is the last time index of the training data set.

To select the lag orders k_1 and k_2, we compute the Akaike information criterion (AIC) of the model with each combination of (k_1, k_2), then adopt the elbow method to find the optimal combination. Figure 12 displays the AICs of a total of 100 candidate models, with each one corresponding to one combination of (k_1, k_2), for k_1, $k_2 = 1, \ldots, 10$. The elbow method suggests choosing k_1 and k_2 to be 7.

Empirically, we reject the null hypothesis that the part of crime counts follows a time-varying Poisson distribution, so the following results are based on the ZIM models with Negative Binomial Distribution.

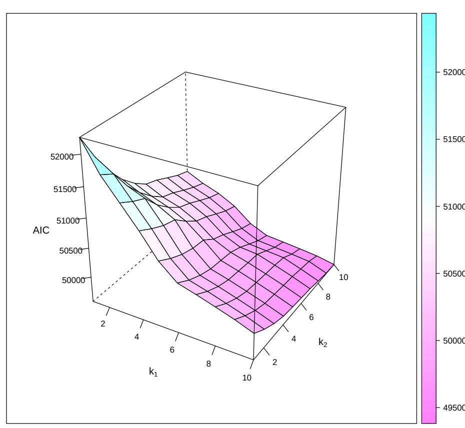


Figure 12: AIC values for the 100 candidate models fit to the training data, with k_1, k_2= $1,\ldots,10$.

2.7.3 Results

2.7.3.1 Model Performance

After fitting one zero-inflated model to each tract, we assessed the goodness of fit based on rooted mean squared error (RMSE) and adjusted root mean square error (Adjusted RMSE). The RMSE of a data set D has the following expression:

$$\text{RMSE}_D = \frac{1}{N_{\text{tract}}} \sum_{i=1}^{N_{\text{tract}}} \Big(\frac{1}{T_D} \sum_{t=1}^{T_D} (y_{i,t} - \hat{\mu}_{i,t})^2 \Big)^{1/2}, \tag{2.7}$$

where T_D is the length of the time sereis in data set D and N_{tract} is the number of tracts. The lower this metric is, the better a model performs in the data set.
The adjusted RMSE normalizes the RMSE by taking account of the fitted variance:

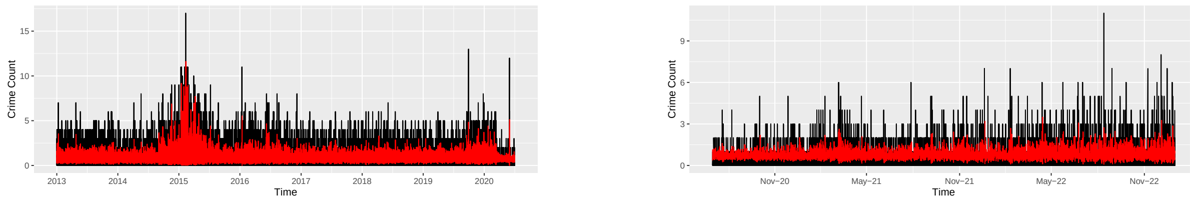
$$\text{Adjusted RMSE}_D = \frac{1}{N_{\text{tract}}} \sum_{i=1}^{N_{\text{tract}}} \Big(\frac{1}{T_D} \sum_{t=1}^{T_D} (\frac{y_{i,t} - \hat{\mu}_{i,t}}{\hat{\mu}_{i,t}\hat{\alpha}_{i,t} + \hat{\mu}^2})^2 \Big)^{1/2}. \tag{2.8}$$

Table 5 presents the performance of fitted models in training and testing data sets. RMSE shows that the model does not exhibit overfitting since the RMSE of the testing data set is lower than the one of the training data set. Adjusted RMSE gives opposite results, suggesting to use of a sliding window procedure to split the data set into training and testing parts, but we only focus on the current implementation due to time constraints.

RMSE	Training	0.5467
	Testing	0.435
Adjusted RMSE	Training	1.7086
	Testing	2.4084

Table 5: Performance of fitted models.

To provide more insights about model performance, Figure 13 focuses on the data in Tract 42101000500, the tract with the highest crime intensity, to compare time series plots of the true and fitted/predicted crime counts. Overall, the model is able to capture the time-varying pattern of crime counts in both data sets. For instance, the fitted crime counts in Figure 13a are consistent with the true ones by experiencing a peak around 2015. After November 2021, the predicted counts in Figure 13b are adjusted to higher values, which corresponds to the increase of the true counts. When crime counts are greater than zero, the zero inflation characteristic results in the model generating predictions that are lower than their actual counterparts. However, it is noteworthy that the variance of a crime count increases with the prediction, so a police officer should be more cautious when the model forecasts a relatively high expected crime count, because it may indicate more crimes than expected. Considering the performance and insights provided by the model, we recommend that the police department in the City of Philadelphia employ the ZIM model to predict the progression of crime counts and consider a relatively high prediction as a potential indicator of a corresponding high crime count.



(a) Training

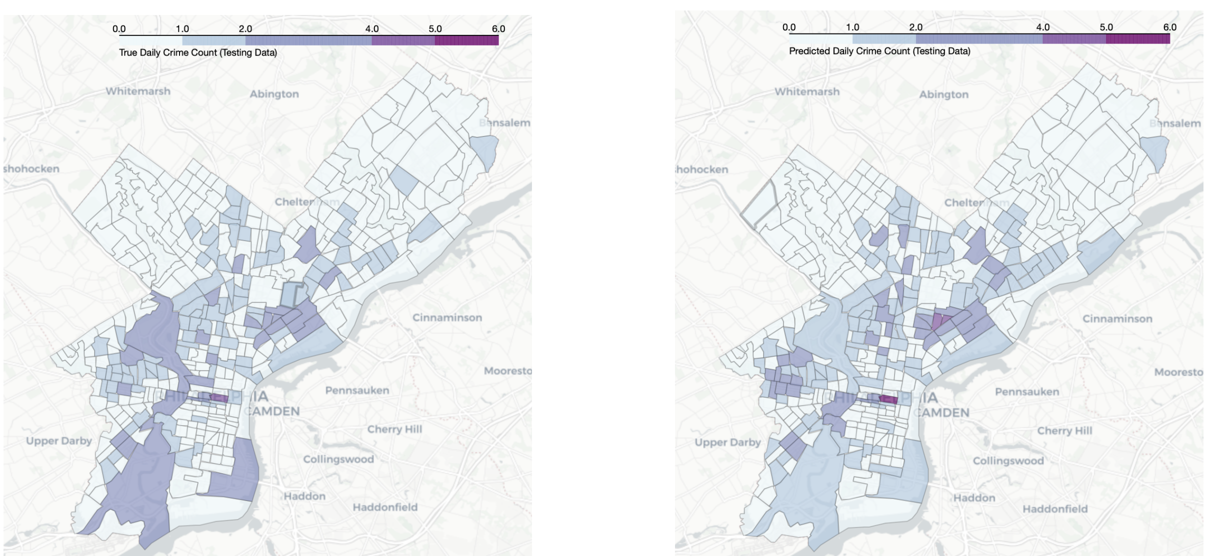
(b) Testing

Figure 13: Time series plots of true crime counts (black) and fitted/predicted crime counts (red) in Tract 42101000500.

2.7.3.2 Heat Map Comparison

In Section 2.7.3.1, we assess the merits of the fitted model with only one tract, this section demonstrates the utility of this time series model when considering all tracts.

Figure 14 compares the true daily crime count for each tract with its predicted counterpart. Both daily counts are averaged using the testing data set, and we set the same colour scale for both panels. Most coloured tracts in Figure 14a are also highlighted in Figure 14b. This indicates that the ZIM model can accurately predict which tracts are more dangerous than others based on the past information of tracts. For tracts coloured in both panels, predicted daily crime counts are close to the true ones since all of these tracts are classified to the same or a neighbouring colour. The darker colours in these misclassified tracts suggest the ZIM model overestimates the number of daily crimes in the testing data set. A possible reason is that these tracts used to high crime intensity before 2020-07-02 and their safety level in terms of daily crime count improved.

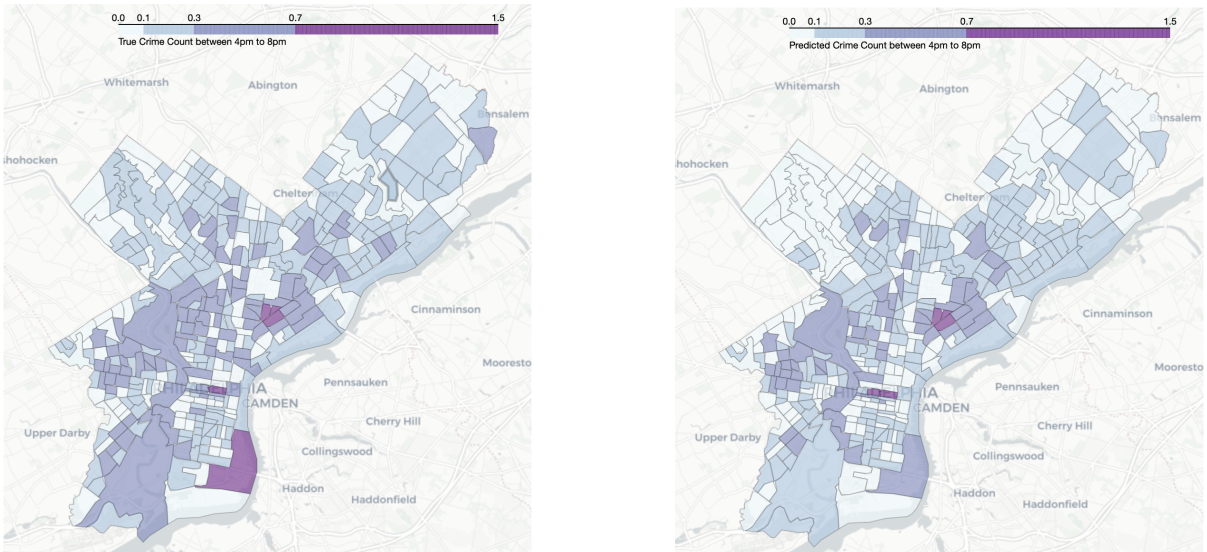


(a) True

(b) Predicted

Figure 14: Heatmaps of true (left) and predicted (right) daily crime count averaged from 2020-07-02 to 2022-12-31.

Instead of averaged daily crime counts, Figure 15 compares the averaged true and predicted crime counts from 4 pm to 8 pm, the time frame with the highest crime intensity according to Figure 6. Similar to Figure 14, sets of highlighted tracts in these two panels have a significant overlap, which reveals the predictability of the ZIM model. In contrast to Figure 14, there are lighter colours in the right panel, compared to the left one. This finding is consistent with Figure 13 in which the predicted crime counts are lower than the actual ones.



(a) True

(b) Predicted

Figure 15: Heatmaps of true (left) and predicted (right) crime counts between 4pm and 8pm, averaged from 2020-07-02 to 2022-12-31.

2.7.3.3 Coefficient Significance

This subsection studies the coefficient significance of fitted models for the top 5 dangerous tracts based on Figure 5a. Table 6 and Table 7 show estimated coefficients. A negative (positive) sign indicates that the corresponding intercept or lag has a negative (positive) association with related parameters ($\lambda_{s,t}$ or $\omega_{s,t}$). For $\lambda_{s,t}$, the estimated mean of crime count at time t in Tract s, intercepts are all negative and significant in Table 6, which means that the mean of a crime count will be less than 1 if the past 7 crime counts are all zero. The only lag that is significant for all 5 tracts is the first lag, but the third, fifth, sixth and seventh lags are significant in 4 out of 5 tracts. The significance of the coefficient of the sixth lag led to the finding that time series of crime counts have seasonality with a period of 1 day. Based on the negative signs of the coefficients of the third lag, we also find that the crime counts 12 hours ago are negatively associated with the current count.

In Table 7, we also detect significant and negative intercepts for all 5 tracts. This finding aligns with intuition because these tracts have high crime intensity which leads to a low probability that no crime will be committed in a 4-hour period. Only the coefficient corresponding to the sixth lag in each tract model is significant. It confirms again the seasonality of crime counts. A negative and significant coefficient of the sixth lag suggests the probability that a crime count following a distribution (either Poisson or Negative Binomial) is positively related to the crime count 24 hours ago.

Tract ID	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Tract 42101000500	**-0.129**	**0.099**	-0.021	**-0.137**	0.001	**0.081**	**0.114**	**0.056**
Tract 42101017702	**-0.191**	**0.090**	0.019	**-0.048**	-0.002	**0.059**	**0.104**	**0.059**
Tract 42101017701	**-0.395**	**0.093**	0.018	-0.028	-0.011	**0.088**	**0.132**	**0.047**
Tract 42101000402	**-0.418**	**0.127**	0.003	**-0.074**	-0.010	**0.095**	**0.130**	**0.045**
Tract 42101017601	**-0.349**	**0.096**	0.013	**-0.051**	0.034	0.101	0.094	0.029

Table 6: Estimated coefficients for parameter $\lambda_{s,t}$ of the top 5 dangerous tract. Bold values means the coefficient is associated with $\lambda_{s,t}$ at 0.0001 significance level.

Tract ID	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_5$	$\tilde{\beta}_6$	$\tilde{\beta}_7$
Tract 42101000500	**-0.501**	**-0.228**	0.118	**0.122**	**0.152**	**-0.300**	**-1.415**	**-0.356**
Tract 42101017702	**-0.804**	**-0.213**	**0.216**	**0.101**	0.057	**-0.248**	**-1.705**	-0.076
Tract 42101017701	**-1.097**	-0.260	**0.278**	**0.131**	0.002	-0.133	**-2.120**	-0.177
Tract 42101000402	**-0.901**	-0.298	0.118	0.041	0.065	-0.401	**-1.276**	-0.403
Tract 42101017601	**-0.920**	**-0.267**	**0.165**	0.094	0.092	0.032	**-2.489**	-0.188

Table 7: Estimated coefficients for parameter $\omega_{s,t}$ of the top 5 dangerous tract. Bold values means the coefficient is associated with $\omega_{s,t}$ at 0.0001 significance level.

3 Conclusion

This report presents a statistical approach to fight crimes in Philadelphia by predicting crime. Through exploratory data analysis, we identified high crime incidence areas in Philadelphia and patterns and trends in the crime data that guided the development of the time-series model and the spatial model. Then we build the model to explain the reason that crime incidence appears.

Crime event data is a typical spatial-temporal data. We first apply Knox test. The results of the Knox test indicate that there is no significant pattern or structure in the sequence of crime data over time, suggesting that the data appears to be randomly distributed. Build upon this hypothesis, only looking at spatial dimension or temporal dimension makes sense.

For the spatial model, it shows that population size is statistically significant in explaining the crime count in each census tract. The temporal model effectively identifies tracts with high daily and 4-hourly crime counts, and exhibits 24-hour seasonality and negative correlation with past crime counts for the five most dangerous tracts.

Overall, our statistical approach to fighting crimes in Philadelphia provides insights into the patterns and trends of crime data and highlights areas that require targeted policing efforts. It is hoped that this report can serve as a useful tool for law enforcement agencies in Philadelphia to allocate their resources more effectively and efficiently in combating crime, thus making the city safer for all its residents.

References

[1] Oscar Areliz. *Los Angeles Recorded 382 Murders in 2022*. Jan. 2023. URL: https://xtown.la/2023/01/09/los-angeles-382-murders-2022/.

[2] InTime. *3 Changes to Your Rotation Schedule that Can Save Lives*. https://intime.com/resources/blog/rotation-schedule/. [Online; accessed 10-April-2023].

[3] Shane D Johnson and Kate J Bowers. "The burglary as clue to the future: The beginnings of prospective hot-spotting". In: *European Journal of Criminology* 1.2 (2004), pp. 237–255.

[4] Ernest G Knox and Maurice S Bartlett. "The detection of space-time interactions". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 13.1 (1964), pp. 25–30.

[5] Tobias Liboschik, Konstantinos Fokianos, and Roland Fried. "tscount: An R package for analysis of count time series following generalized linear models". In: *Journal of Statistical Software* 82 (2017), pp. 1–51.

[6] Mariam Tafsiri Matthew Heeks Sasha Reed and Stuart Prince. *The Economic and Social Costs of Crime*. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/732110/the-economic-and-social-costs-of-crime-horr99.pdf.

[7] *Report Ranks America's 15 Safest (And Most Dangerous) Cities For 2023*. URL: https://www.forbes.com/sites/laurabegleybloom/2023/01/31/most-dangerous-cities-in-the-us-crime-in-america/?sh=2b41d1284b25.

[8] *Safest Philadelphia neighborhoods*. URL: https://www.neighborhoodscout.com/pa/philadelphia/crime#:~:text=With%5C%20a%5C%20crime%5C%20rate%5C%20of,here%5C%20is%5C%20one%5C%20in%5C%2029.

[9] Ralph B Taylor, Stephen D Gottfredson, and Sidney Brower. "Block crime and fear: Defensible space, local social ties, and territorial functioning". In: *Journal of Research in crime and delinquency* 21.4 (1984), pp. 303–331.

[10] Tribune Publishing Company LLC. *New York Daily News*. 2023. URL: https://www.tribpub.com/gdpr/nydailynews.com/.

[11] James C Wo. "Understanding the differential effects of land uses on crime: An examination across Philadelphia neighbourhoods". In: *The British Journal of Criminology* 59.6 (2019), pp. 1432–1454.

[12] Ming Yang, Gideon KD Zamba, and Joseph E Cavanaugh. "Markov regression models for count time series with excess zeros: A partial likelihood approach". In: *Statistical Methodology* 14 (2013), pp. 26–38.

4 Appendix

4.1 Theory in Poisson Regression

4.1.1 Data Generating Process

Poisson Regression is a type of generalised linear model, the Data Generating Process is:

For the i-th observation, given the latent variable x_i, each y_i is drawn from a Poisson distribution $y_i|x_i \sim Poisson(\mu_i)$. According to the property of Poisson distribution, $\mathbb{E}[y_i|x_i] = \mu_i$. The core assumption is that each parameter μ_i is linear in x_i: $\mu_i = \theta' x_i$.

In a word, we have 4.9, and this is equivalent to $\mathbb{E}[y_i|x_i] = e^{\theta' x_i}$.

4.1.2 Estimation

Since we have the conditional distribution of $y_i|x_i$, the maximum likelihood estimation could take its role. The probability density of observation i is:

$$f(y_i|x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!},$$

take the log, and note that $\mu_i = e^{\theta' x_i}$, we have:

$$\begin{aligned}\log f(y_i|x_i) &= -\mu_i + y_i \log \mu_i - \log(y_i!) \\ &= -e^{\theta' x_i} + y_i(\theta' x_i) - \log(y_i!).\end{aligned}$$

Given sample size n, the log-likelihood function is:

$$l(\theta; X, Y) = \sum_{i=1}^{n} \log f(y_i|x_i) = \sum_{i=1}^{n} \Big(- e^{\theta' x_i} + y_i(\theta' x_i) - \log(y_i!)\Big).$$

The maximum likelihood estimate is defined as:

$$\hat{\theta}^{\text{MLE}} = \arg\max_{\theta \in \mathbb{R}^p} l(\theta; X, Y).$$

However, in this case, we do not have a closed-form solution, but the value of $\hat{\theta}^{\text{MLE}}$ could be obtained numerically via algorithms such as Newton-Raphson or Adam, etc.

4.1.3 Inference

Then we have the fitted value of average count: $\hat{\mu}_i = e^{\hat{\theta}' x_i}$. Furthermore, in Poisson Regression, Deviance Residual for observation i is defined as:

$$\begin{aligned}\text{res}_i &= \Big[- \hat{\mu}_i + y_i \log \hat{\mu}_i - \log(y_i!)\Big] - \Big[- \mu_i + y_i \log \mu_i - \log(y_i!)\Big] \\ &= y_i - \hat{\mu}_i + y_i \log\Big(\frac{\hat{\mu}_i}{y_i}\Big).\end{aligned}$$

The sum of Deviance Residuals follows χ-square distribution if the model is correctly specified, thus the statistic could be used for the goodness-of-fit test. Under H_0 : The model is correctly specified, we have:

$$\sum_{i=1}^{n} \text{res}_i \sim \chi^2_{n-p},$$

where the degree of freedom is $n - p$, since the estimation of $\theta \in \mathbb{R}^p$ cost p number of degree of freedom.

The goodness of fit for Poisson regression is actually to check whether the sample variance is close to sample mean, which is a nature of Poisson distribution.

If we reject the null hypothesis, which means the issue of dispersion exists, this can be caused by several reasons: (1) omitted variables and functional forms. (2) Assuming that the model is correctly specified, the assumption that the conditional variance is equal to the conditional mean should be checked. There are several tests including the likelihood ratio test of over-dispersion parameter alpha by running the same model using negative binomial distribution. R package pscl (Political Science Computational Laboratory, Stanford University) provides many functions for binomial and count data including odTest for testing over-dispersion.

4.2 Negative Binomial Regression

Our approach is that: Use Exploration Data Analysis (EDA) to find potentially correct hypothesises. For example, H_0 : There is no seasonality in the counts of crimes monthly data in each tract, against H_1 : There is some. And then apply hypothesis testing to show that we reject the null hypothesis, thus, the seasonality is justified. Then we apply algorithms to eliminate the effect of season.

Then we could maybe assume the data is i.i.d. among all

Model 1. *Poisson Regression without considering the time dependence. As what has been done in the [11].*

$$\log \mathbb{E}[y_i|x_i] = \theta' x_i, \tag{4.9}$$

Note that the Poisson distribution has the same mean and variance. This property could be used to test whether the model is correctly specified.

Model 2. *Poisson regression with the lagged independent variable.*

Add how many orders of lags could be determined by AIC and BIC. Then do the goodness of fit test.

4.3 Cost of Crime Analysis

Cost of Crime Analysis was implemented [6], which aims to estimate the real cost of different types of crime. The cost of each type of crime is broken down into 3 components:

- cost in anticipation (e.g. cost of burglar alarms)
- cost in consequences (e.g. cost of stolen or damaged property)
- cost in response (e.g. cost of police and criminal justice system)

The sum of the 3 components represents the unit cost of each crime category. This was then multiplied by the number of crimes to provide an estimate for the total cost of crime.

Crimes	Number of crimes	Unit cost (£)	Total cost (£)
Homicide (criminal, justifiable, gross negligence)	3515	3217740	**11.3bn**
Assault (aggravated assault, no firearm)	318647	14050	**4.5bn**
Thefts (general theft, theft from vehicle, motor vehicle theft)	414343	4180	**1.7bn**
Robbery (no firearm, firearm)	59907	13160	**0.8bn**
Burglary (residential, non-residential)	72597	10695	**0.8bn**
Rape and other sex offenses	20238	22940	**0.5bn**
Criminal mischief (driving under the influence, vandalism, public drunkenness, disorderly conduct, vagrancy, loitering)	195727	1350	**0.3bn**

Figure 16: Cost of crime analysis

The amounts were converted from GBP to USD to remain consistent with the geographical location of the study (Philadelphia, PA).

4.4 Additional Graphs and Visualisations

1. **Tract map**(17): This map displays the boundaries of all tracts in Philadelphia, providing an overall view of the area under study. It helps us to understand the geography of the area and to gain a general understanding of the tract distribution and locations.

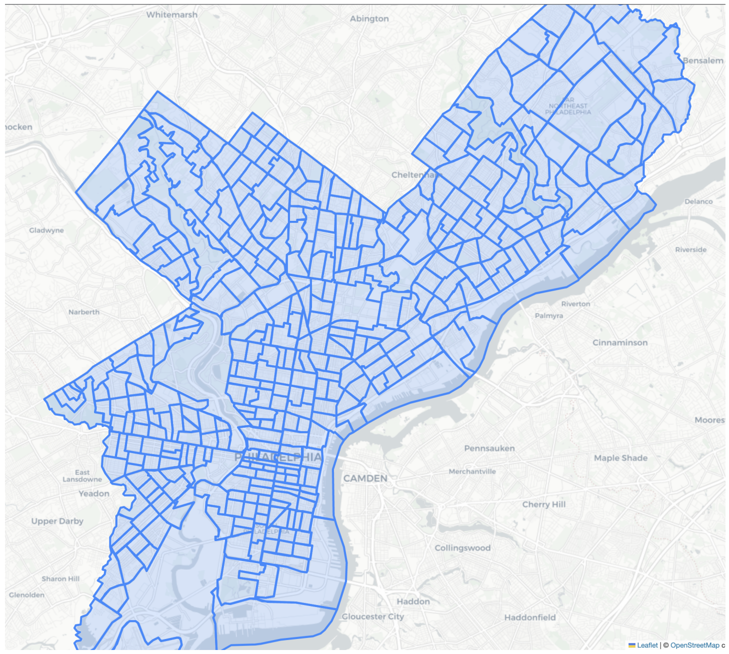


Figure 17: A map of the boundaries dividing different tracts in Philadelphia.

2. **Heat map of total crime incidence**(18): This heat map shows the overall crime incidence across the city of Philadelphia. It provides a visual representation of the hotspots of crime in the city and helps us to identify areas that require more police attention and care for public safety.

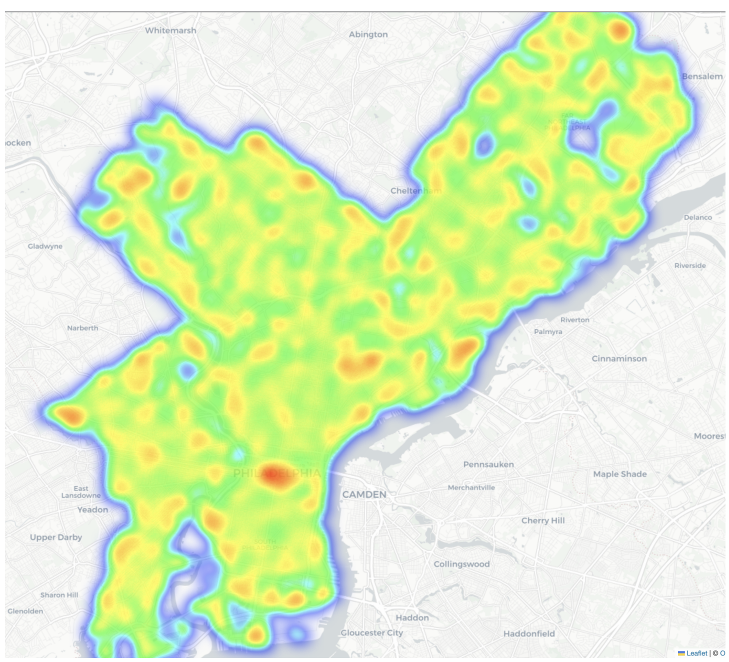


Figure 18: A heat map of the total crime incidence in philadelphia