Determine the More Efficient Implementation of Police Stations in Baltimore through Clustering

Bowen Deng James Guo Lisong He Jiachen Wang

Johns Hopkins University

December 19, 2024

Table of Contents

Crime Issues in Baltimore

- According to a new article in 2024, Baltimore has been ranked as the "deadliest large city in the nation." [1]
- The crime problem is a significant issue, which threatens the surrounding industries, [2] with robbery and violence crime. [1]

Source of Data

- We use data set from the **Open Baltimore website** on the recorded crime data. [3]
- This is the data set from the police department of Baltimore in terms of the criminal cases in the city.

RowID	CCNumber	CrimeDateTime	CrimeCode	Description	Weapon	Post	Gender	Age	Race	Location	Latitude	Longitude
138357	23L09461	12/31/2023 2:00:00 PM	6C	LARCENY		811				4600 EDMONDSON AVE	39.29341	-76.69569
⋮		*Further data are omitted for presentation purposes.*									⋮	

- We kept the most important rows. As we train the model for clustering, we use the following:

Column	**Reason for Keeping**
CrimeDateTime	We use the date as a criterion of the selection of data for training and validating.
CrimeCode	This code is a representation of the seriousness of the crime.
Description	We assign weight based on the seriousness of the description of the crime.
Latitude & Longitude	This represents the location of the crimes as a subset of the $\mathbb{R}^2$ space.

Goal of Project

Study about Police Response Correlating Crime Rate

According to the study, the police presence deters the happening of crime. [4]

Hence, **the better coverage of police through faster response time could help decrease the crime rate**. Our problem is now:

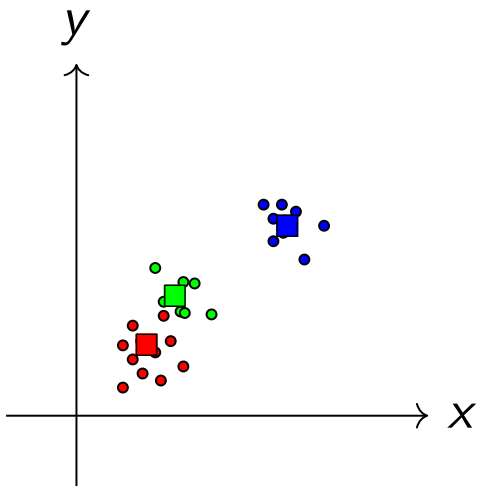
- **Efficiency**: We want to "span" the city up with less as possible police department to minimize the cost of operation.
- **Response time**: We want to minimize to total amount of response time for police to react to a crime.

Research Procedures

The problem naturally becomes a **clusterings problem**, as a type of **unsupervised learning**.

Setup in $\mathbb{R}^2$

- Consider the $(\text{latitude}, \text{longitude})$ as a subset of the $\mathbb{R}^2$ plane.
- Find the clusters of the crimes and centroids to minimize the distance from centroids to the crime locations.



Alternations in Methods

- The k-clustering algorithm can take different metric spaces (X, d). [5]
  - The typical metric space is $\mathbb{R}^n, \| \bullet \|_2$, or the Euclidean distance.
  - We will test on different metrics, such as ℓ^p metrics and an optimized metric we developed by ourselves.
- As the data includes more fields, we would also want to assign weights on our clustering.
  - We have thought about the **severity of crime**. In general, a **murder** is more sever than a **robbery**, and more sever than **larceny**.
  - There have been more robbery throughout the city and less murdering. [1]
  - Thus, it is important to distribute the police forces corresponding to the severity of cases.
- By introducing these heuristics, we want to compare the impact of these changes on the model.

K-Means Method

K-means clustering is a simple and effective method to associate points of a dataset with a class, *i.e.*, the clusters.[5]

- We start by randomly selecting a set of points as the its centroid.
- Then, we iteratively do:
  1. Assign each data point to a cluster based on the **distance metric** from the point to the centroids.
  2. Then, we calculate the new centroids of each cluster by taking the average of all the points that belong to the current cluster.
- The iteration terminates when:
  - When the number of iterations exceed the maximum number of iterations.
  - When the movement of each cluster centroid is within a level of tolerance.

Assigning Weights

- According to the dataset, we first calculate the probability of proportion of each type of crime in the variable `VIOLENT_CR`.

  Here are the relative frequencies.

proportion	VIOLENT_CR
LARCENY	0.363547
COMMON ASSAULT	0.248342
AGG. ASSAULT	0.147875
AUTO THEFT	0.122083
BURGLARY	0.079096
ROBBERY	0.026775
RAPE	0.008106
HOMICIDE	0.004176

Assigning Weights (Continued)

- Then we multiply all of the probability by 1000 and allocate numbers into each crime based on the **severity of crime**.

```
mapping =
      {"LARCENY":          8,  "COMMON ASSAULT":   79,
      "AGG. ASSAULT":    122,  "ROBBERY":         148,
      "BURGLARY":          27,  "AUTO THEFT":        4,
      "RAPE":             248,  "HOMICIDE":        364}
```

- Therefore, for our **K-means** algorithm, it is more inclined to put the weight to the more sever and serious crimes, such as homicide. The model would *theoretically* be locating the police departments to locations that more serious crimes tend to happen.

Metric Space and ℓ^p Space

Metric Space

Let (X, d) be a metric space, then X is a set, and $d : X \to \mathbb{R}_{\geq 0}$ is the metric that must satisfy the following properties for any $x_1, x_2, x_3 \in X$:

1. **Positivity**: $d(x_1, x_2) \geq 0$,
2. **Definiteness**: $d(x_1, x_1) = 0$ if and only if $x_1 = 0$,
3. **Symmetry**: $d(x_1, x_2) = d(x_2, x_1)$, and
4. **Triangular inequality**: $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$.

ℓ^p Space

Consider $\mathbb{R}^n$ as a n-dimensional vector space (technically, it could be infinite dimensional with some constraints), and suppose $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$. For $p \geq 1$, we define:

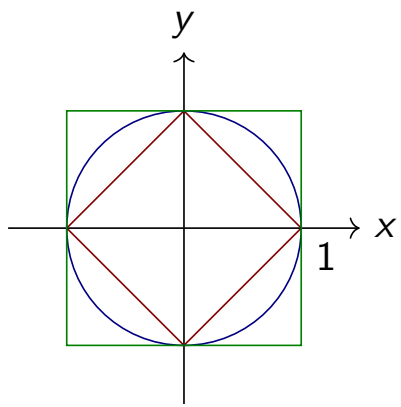
$$L_p\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) := \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}. \text{ [6]}$$

Use of ℓ^p Space

Here, we inspect the road map of Baltimore city:

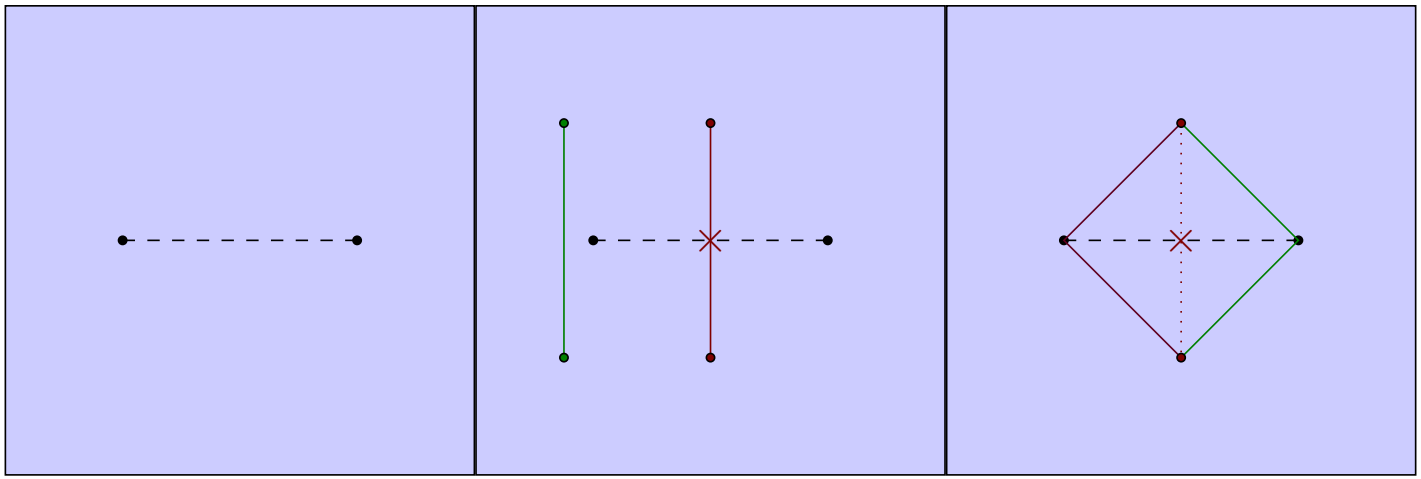
- Cannot consider it as strict grids (Manhattan distance) or straight line (Euclidean distance).
- We will be evaluating the ℓ^p metric with $p = 1$, 1.25, 1.5, 1.75, and 2.

Example: $\mathbb{R}^2$ Unit ball of L_1(red), L_2(blue), and L_∞(green).



Optimized ℓ^2 Metric

- Consider a **not total-connected** set, we can optimize the ℓ^2 metric as follows.



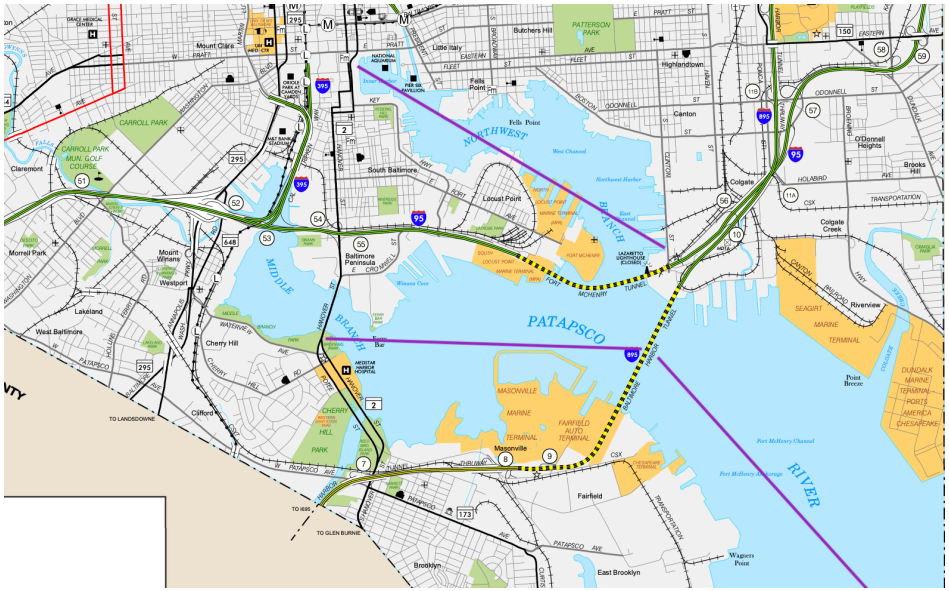
Optimized ℓ^2 Metric

Let $X \subset \mathbb{R}^n$ be a subset of a finite dimensional Euclidean space, and let $x_1, x_2 \in X$ be arbitrary. Consider $C : [0, 1] \to X$ as any path between x_1 and x_2 such that $C(0) = x_1$ and $C(1) = x_2$, we let the distance be:

$$d(x_1, x_2) := \inf_{C \text{ is a path in } X \text{ between } x_1 \text{ and } x_2} \text{len}(C).$$

Optimized ℓ^2 for Baltimore City

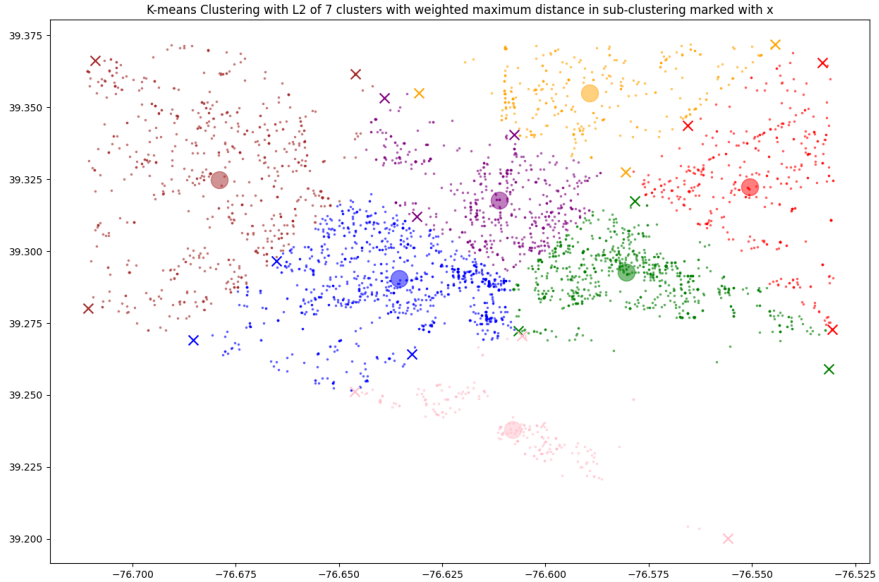
Thereby, implementation-wise, we will be creating a metric over a subset in $\mathbb{R}^2$ that excludes the watershed in Baltimore city.



When a path attempts to intersect one of the purple line segments in the figure, it has to either path through the left or the right endpoints of the purple line segments.

Evaluation Methods

- Within each cluster, *recluster* all points in it into 3 new clusters with the same distance metric but unweighted.
- For the new inner cluster, calculate the distance with the original centroid.
- Calculate the time needed to approach three furtherest points on each inner cluster from the original cluster using Google map.
- Record the longest two time of the three time from the last step within each new clusters and average them.



Evaluating Current Models

- We use the data from **October 1st, 2024** to **November 6, 2024** to train the model, and use excerpts of data from **November 7, 2024** to **December 6, 2024** to validate our model.
- The way that we compare our model with the Baltimore current setup (9 police departments) is to compare the response time between ours and theirs. [3]

Errors in Data Set

Prior to our results, we want to note some errors/missing components that are in the data set [3].
Here are some sample of data with error.

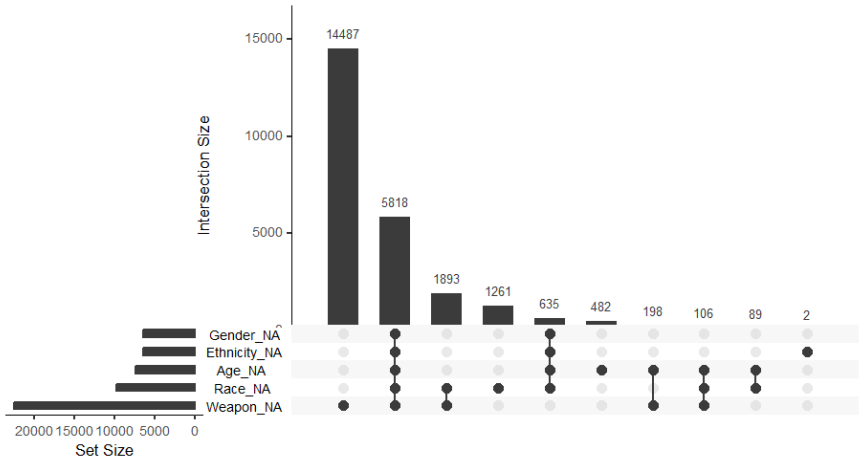
RowID	CCNumber	CrimeDateTime	CrimeCode	Description	Weapon	Post	Gender	Age	Race	Location	Latitude	Longitude
128782	23L09461	1523/09/08 17:30:02	7A	AUTO THEFT		421	F	0	WHITE	5100 HARFORD RD NORTHEAST	-76.565003	39.34795
128784	23C04992	1023/03/16 13:01:02	4B	AGG. ASSAULT	PER_WEA	822	F	30	BLACK	4600 PEN LUCY RD SOUTHWEST	39.285659	-76.693882
427984	18H02741	2018/8/7 16:30	6E	LARCENY	NA	822	M	-1	BLACK	3600 BROOKLYN AVE	39.23494022153	-76.59923441235
⋮				*Other error data are omitted for presentation purposes.*							⋮	

Examples of Errors

- `AUTO THEFT` in 1523, but the first car built after that $\Longrightarrow$ Incorrect date.
- Case happening at $(-76, 39)$, but Baltimore is at around $(39, -76)$, not Antarctica $\Longrightarrow$ Flipped coordinates.
- Criminals of ages -1, 0, and 1, while it is suspicious a child at 0 or 1 year old could using a knife or gun to hurt others, it is not possible for a person who was not born to rob others $\Longrightarrow$ Incorrect age.

Missing Data

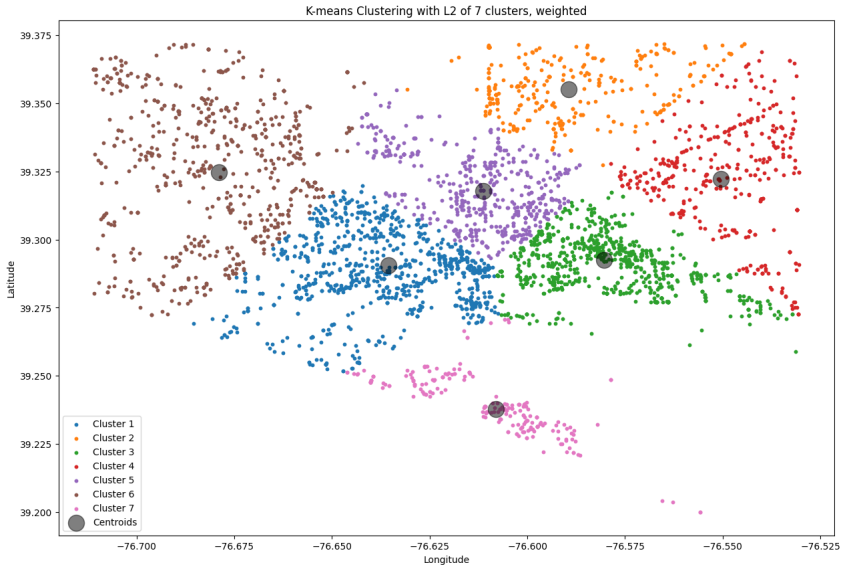
- There are missing age, race, address, or other information covered in the original dataset.
- Considering some of the officers are not familiar with the recording system, or they are tired during the recording period, these could be careless mistakes.

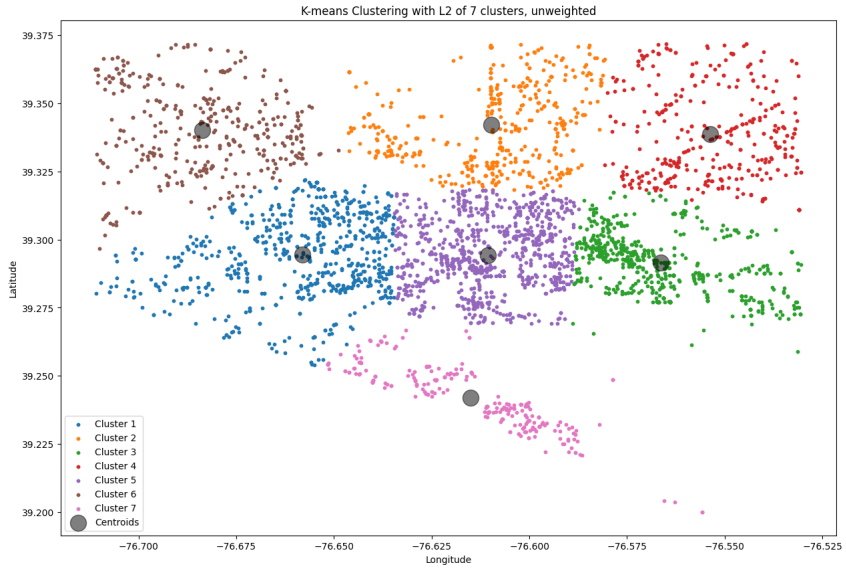


- However, there could also be cases when such information is unknown, and it might not necessarily be random. However, since we are ignoring those fields, these loss are fine.

Remarks on Weights

- We started the idea of training based on weights.
- It is not possible to quantitatively analyze the model with weights. Given that homicides are more severe than other crimes, they affect the position of centroids much more than other points in weighted cluster method.





Interpreting the ℓ^p Metrics

We compare the effectiveness between the different ℓ^p metrics for $p = 1$, 1.25, 1.5, 1.75, and 2 with $k = 7$. The results are as follows:

p	Minimum Travel Time (min)	
	Average	Standard Deviation
1	9.64	2.41
1.25	9.43	1.95
1.5	9.14	1.29
1.75	10.14	2.51
2	10.29	2.33

- $\ell^{1.5}$ norm performs the best among all with the smallest average and standard deviation of traveling time.

Compare with Current Models

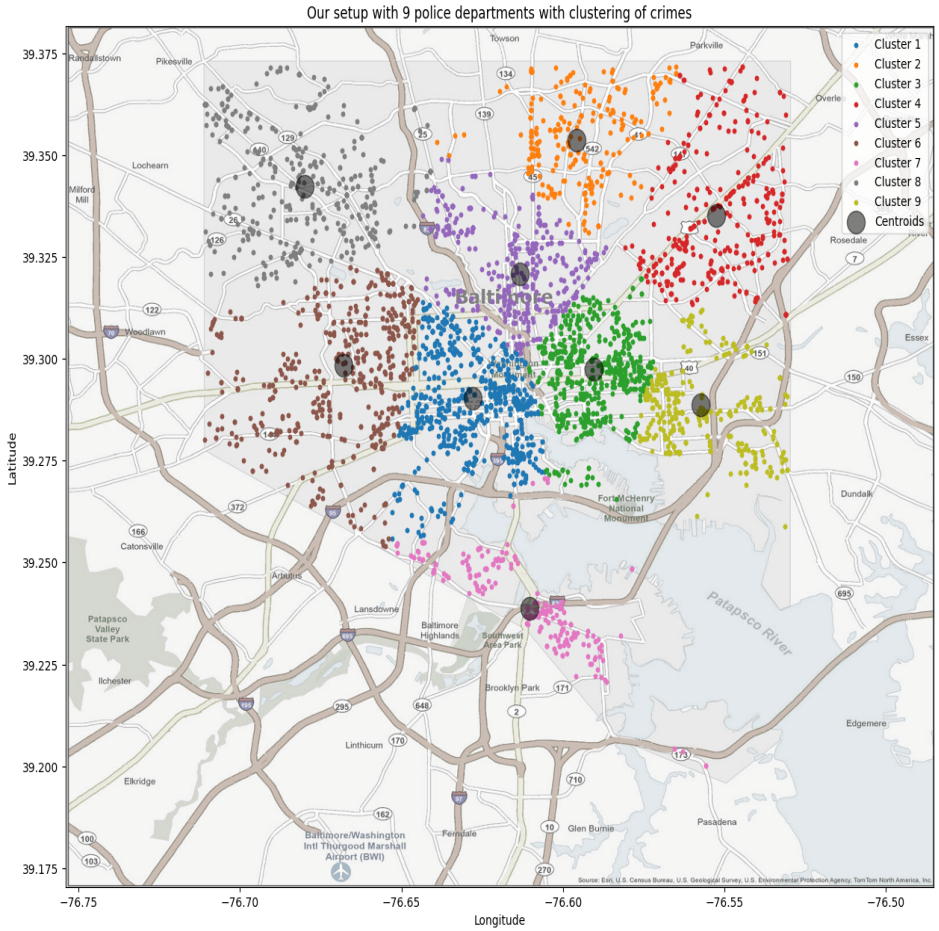
Here are the results of the models and comparisons.

Setup	Minimum Travel Time (min)	
	Average	Standard Deviation
Baltimore Current Setup	5.60	2.02
Our 7 Departments	5.51	1.91
Our 9 Departments	4.26	1.62

- Baltimore Current (9 departments) takes the longest time among all.
- For the 7 models, it slightly outperforms the current deployment while reducing the number of departments.
- Having the same number of departments as current significantly shortens the time by 2.5 minutes.

Our Conclusion of the Model

Using $p = 1.5$ weighted clustering, we have the following trained model.



Conclusions

- Implementing weighted k-means makes a big difference in terms of final clustering result. Assigning more weight on more serious crime is necessary.
- $\ell^{1.5}$ norm is selected as the best theoretical distance as it balances the trade-offs between Euclidean distance and Manhattan distance. The practice indeed proves this point.
- Reducing the number of police departments from 9 (current) to 7 retains the same intensity of deployment.
- Maintaining the same 9 police departments significantly time shortens time arrive on scene.

Suggestions and Further directions

- Leverage real-time crime trends to dynamically optimize resource allocation and patrol routes without relocating stations, like using clustering models to identify areas for temporary officer deployment or increased patrol frequency.
- Incorporate tools like Google Maps APIs for detailed and scalable evaluations.
- Apply predictive analytics to identify potential crime hotspots for preemptive measures.
- Test methods in cities with diverse crime patterns and geographies to validate adaptability and effectiveness.

References I

- V. Hill, "Baltimore named deadliest large city in the u.s., surpasses memphis and detroit," Fox News, August 2024. [Online]. Available: https://foxbaltimore.com/news/local/baltimore-named-deadliest-large-city-in-the-us-surpasses-memphis-and-detroit
- Campus security. Johns Hopkins University Public Safety. [Online]. Available: https://publicsafety.jhu.edu/community-safety/campus-security/
- (2023, July) Part 1 crime data. Open Baltimore. [Online]. Available: https://data.baltimorecity.gov/datasets/baltimore::part-1-crime-data/about
- S. Weisburd, "Police presence, rapid response rates, and crime prevention," *The Review of Economics and Statistics*, vol. 103, no. 2, pp. 280–293, 05 2021. [Online]. Available: https://doi.org/10.1162/rest_a_00889
- I. Shpitser, "Semi-supervised and unsupervised learning." [Online]. Available: https://courselore.org/files/2593442987984265654 5/lecture13.pdf

References II

- R. S. Elias M. Stein, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, November 2009.
- G. Geography. Baltimore map, maryland. [Online]. Available: https://gisgeography.com/baltimore-map-maryland/
- "General highway map: Baltimore city," Maryland Department of Transportation. [Online]. Available: https://www.roads.maryland.gov/Town_Gridmaps/100000_Baltimore%20City.pdf
- A. K. Elmagarmid, *Similarity Search The Metric Space Approach*. Springer Science, 2006.