
Beantown Ballistics: Crime Clustering in Boston

Nina Fiore, Kyle Mikami, and Brittany Regan

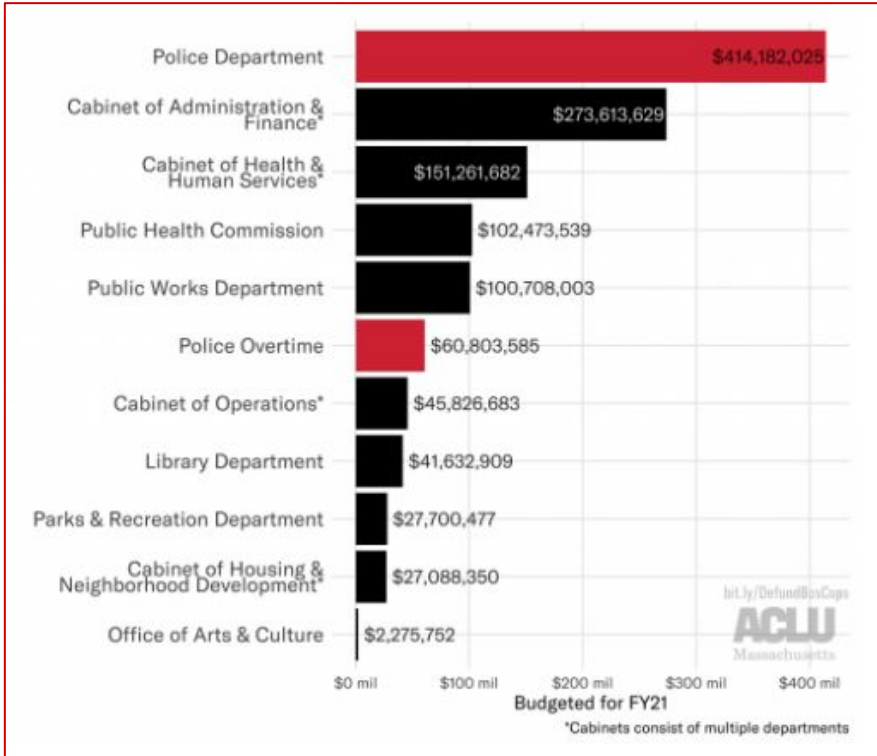


Agenda



- Project Background & Problem Description
- Data & Approach
- EDA
- Clustering Algorithms & Results
- Initial Findings & Next Steps
- References

Project Background & Context



- Crime is a prevalent social issue and has become a national discourse topic recently, driven by factors like an uptick in crime since 2019, excessive force/brutality, and the Defund the Police movement
- The Boston Police Department has an annual budget of over \$400M, the city's second largest budget item
- Understanding where crime occurs across Boston, which types of crimes are prevalent in certain areas, and how areas are related in terms of crime is critical given limited resources and a bright spotlight on splicing activities
- **Our analysis hence aims to cluster police districts/neighborhoods together based on types of crimes and trends**

Data & Approach

Data

Data was consolidated from two main sources:

1. BPD crime incident reports from 2015-2022, containing crime information along with time and location data
2. Boston neighborhood demographics from the US Census Tract, including population, racial makeup, etc.

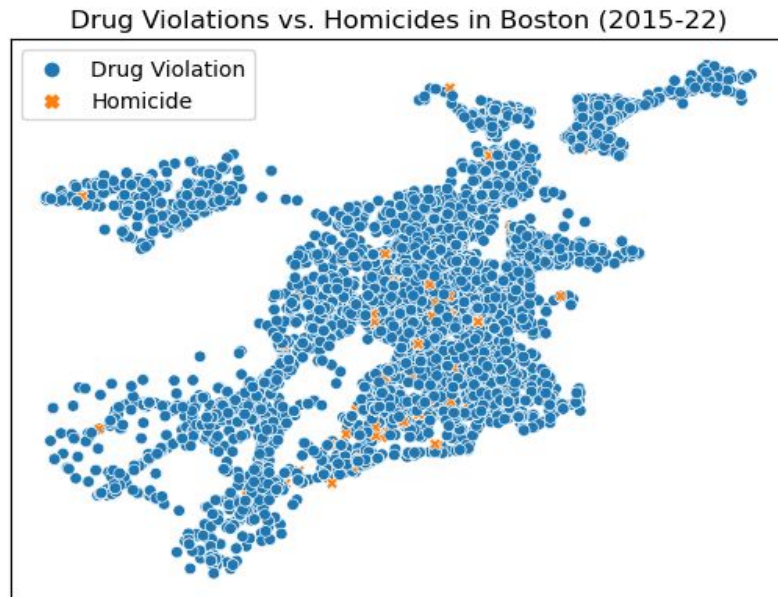
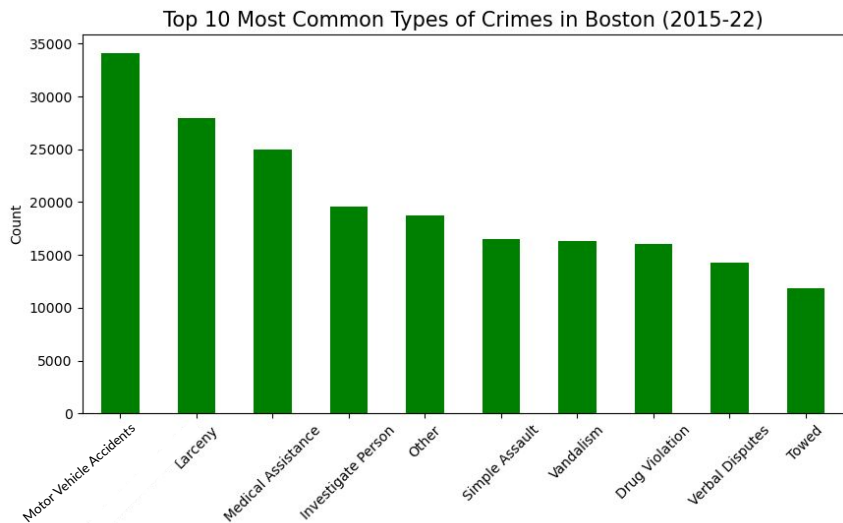


INCIDENT_ID	OFFENSE	OFFENSE	OFFENSE_DISTRICT	REPORTING_AGENCY	SHOOTING_OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOURLY	UCR_PART	STREET	Lat	Long	Location		
119207755	3115	Investigate	INVESTIG A B3	468	4/30/2018 9:00	2018	4	Monday	9	Part Three	HAZLETON	42.27997	-71.0955	(42.27997063, -71.09553354)		
119207733	619	Larceny	LARCENY E18	496	3/6/2018 8:00	2018	3	Tuesday	8	Part One	HYDE PARI	42.26922	-71.1209	(42.26922388, -71.12085347)		
119207666	2629	Harassmer	HARASSMI E5	662	10/31/2018 12:00	2018	10	Wednesda	12	Part Two	PRIMROSE	42.29077	-71.1302	(42.29076521, -71.13021098)		
119207538	2629	Harassmer	HARASSMI A1	96	4/9/2018 8:43	2018	4	Monday	8	Part Two	ATLANTIC	42.35526	-71.051	(42.35526402, -71.05098788)		
119207533	3208	Property L	PROPERTY D4	132	1/1/2018 0:00	2018	1	Monday	0	Part Three	COMMON	42.35352	-71.0728	(42.35352153, -71.07283786)		
119207501	2629	Harassmer	HARASSMI E5	668	10/1/2018 9:15	2018	10	Monday	9	Part Two	CEDRUS A'	42.2784	-71.1353	(42.27839822, -71.13527907)		
119207471	1107	Fraud	FRAUD - IN B3	458	11/18/2018 9:00	2018	11	Sunday	9	Part Two	BLUE HILL	42.30316	-71.0852	(42.30316091, -71.08515552)		
119207471	2629	Harassmer	HARASSMI B2	278	1/1/2018 0:00	2018	1	Monday	0	Part Two	HARRISON	42.33154	-71.0802	(42.33153805, -71.08015661)		

Analysis Approach

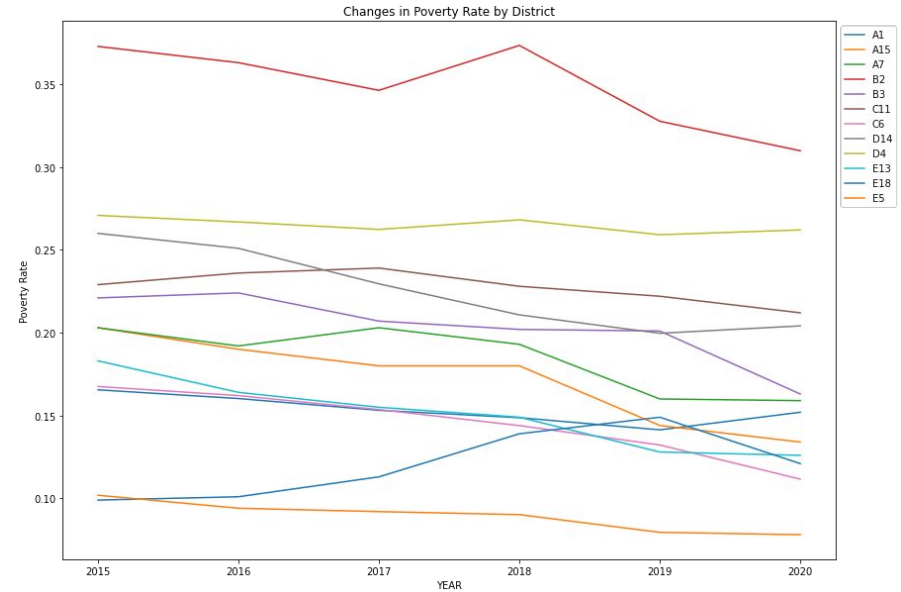
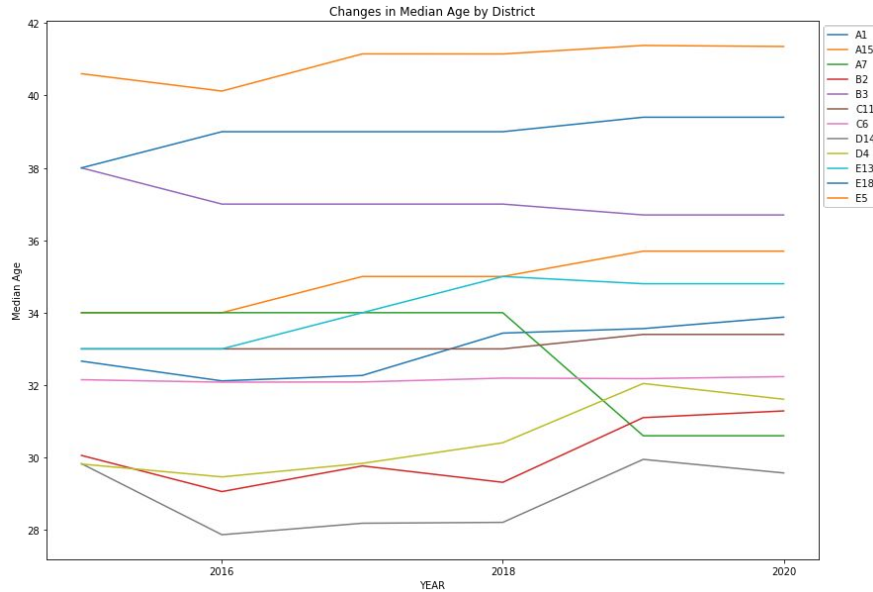
1. Gather/consolidate crime and demographic data
2. Significant data wrangling and tidying was required given variance between years, missing values, etc.
3. EDA
4. Analyze several clustering algorithms to group police districts together based on types of crime and prevalence
5. Analysis and interpretation of results

EDA of Crime Incident Datasets



- Boston Police Department incident datasets consist of over 644K recorded crimes from across Boston over the 8-year period from 2015-2022
- As expected, certain types of crimes such as car accidents and larcenies are common while highly violent crimes are relatively rare
- In addition, as seen above some crimes like drug violations occur all across Boston while crimes like homicides tend to only occur in certain areas

EDA of Census Tract Data



- Boston's neighborhoods have experienced some shifts over the past few years
- For instance, District A7 (East Boston) has seen a noticeable drop in median age and poverty rate given gentrification and younger people relocating there

Setup for Clustering Algorithms

- After tidying data and combining crime data with census tract data, we organized crime incident prevalence at a population per capita level by police district in order to run clustering algorithms on data
- Granular offense categories were grouped into broader categories to consolidate similar crimes together
- We also created versions of this dataset at an annual level in order to assess how clusters may change over time



DISTRICT	ARSON	ASSAULT	AUTO THEFT	BURGLARY	DRUG VIOLATION	FIREARM DISCOVERY / VIOLATIONS	HOMICIDE	LARCENY	MISSING PERSON REPORTED / LOCATED	MOTOR VEHICLE ACCIDENT RESPONSE
A1	0.052393575	18.34700024	2.386012636	3.010671871	9.929156415	1.035622066	0.031201515	35.15718	1.197290344	14.37218645
A15	0.035864777	5.6066256	1.49484288	1.464733429	3.954411111	0.537314915	0.017301746	11.8719	1.382745342	12.02038994
A7	0.018160741	5.467322264	1.415121861	1.766374449	4.048906672	0.436745581	0.043369423	6.644118	1.597021606	7.702074132
B2	0.074676722	13.92412781	3.355045128	3.438234207	7.695855732	1.849828604	0.183636791	14.60435	6.127904074	19.29854767
B3	0.111242536	29.77271948	7.235104129	7.387652773	14.36826252	3.784902503	0.526655972	22.59408	19.11930216	37.20353328
C11	0.015922515	6.339769999	1.886644441	1.911810774	3.937640381	0.726102639	0.066638435	7.798459	3.234462803	10.58100216
C6	0.028997239	10.16031132	3.369196428	3.003374487	9.119377122	0.838710092	0.046033459	18.84212	1.863379216	16.59276601

Clustering Results

District	Neighborhoods	KMeans++ Cluster	DBSCAN Cluster	Hierarchical Cluster
A1	Downtown, North End, West End, Chinatown, Beacon Hill	0	0	0
A15	Charlestown	1	1	1
A7	East Boston	1	1	2
B2	Roxbury, Longwood, Mission Hill	2	2	4
B3	Mattapan	3	3	3
C11	Dorchester	1	1	2
C6	South Boston, South Boston Waterfront	2	4	4
D14	Brighton, Allston	1	1	2
D4	South End, Back Bay, Fenway	5	5	0
E13	Jamaica Plain	2	1	1
E18	Hyde Park	2	1	1
E5	West Roxbury, Roslindale	1	1	2

Initial Findings & Next Steps



Initial Takeaways

- Both KMeans and DBSCAN algorithms grouped the same neighborhoods together in most cases
- The hierarchical algorithm had overlaps with both algorithms as well, but seemed to more closely align with DBSCAN's clusters
- The largest cluster encompasses a significant share of Boston's land area and includes neighborhoods thought of as distinct from each other, such as Dorchester, Brighton/Allston, and East Boston

Next Steps/Future Work

- Further optimize clustering algorithms and hyperparameter tuning
- Assess changes in district/neighborhood clusters over time by running algorithms on years individually
- Deeper dive into types of crimes that are prevalent in each district and in each identified cluster
- Explore clustering visualizations

References



1. McKinley Becker, Kaitlin. "Boston Police Sgt. Placed on Leave, DA Investigating Body Cam Videos That Show 'Troubling' Footage" [18-19 December 2020]. *NBC San Diego*. [Online] available: <https://www.nbcsandiego.com/news/national-international/boston-police-sgt-placed-on-leave-da-investigating-after-body-cam-videos-show-troubling-footage-from-protest/2476181/> [accessed 18 November 2022].
2. Vera, Amir. "Boston mayor fires city police commissioner after investigation into domestic violence allegations" [7 June 2021]. *CNN*. [Online] available: <https://www.cnn.com/2021/06/07/us/boston-police-commissioner-fired/index.html> [accessed 18 November 2022].
3. DV and Police Aggression
4. McCourt, Clara. "Boston police officer arrested following domestic violence incident" [30 October 2022]. *Boston.com*. [Online] available: <https://www.boston.com/news/local-news/2022/10/30/boston-police-officer-roselyn-lacroix-arrested-domestic-violence/> [accessed 18 November 2022].
5. McKinley Becker, Kaitlin, and Mike Pescaro. "Boston Police Officer, Out on Bail After Domestic Violence Charges" [30-31 October 2022]. *NECN*. [Online] available: <https://www.necn.com/news/local/boston-police-officer-arrested-following-domestic-violence-incident/2856784/> [accessed 18 November 2022].
6. Chambers, Lauren, with Kade Crockford, and Fatema Ahmad. "Unpacking the Boston Police Budget" [undated]. *ACLU Massachusetts*. [Online] available: <https://data.aclum.org/2020/06/05/unpacking-the-boston-police-budget/> [accessed 15 November 2022].
- 7.

References (continued)



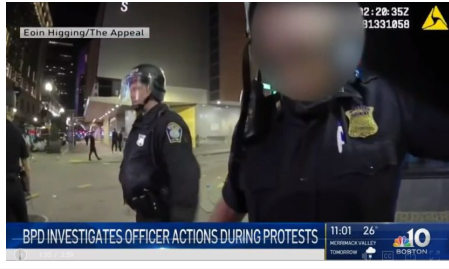
1. McKinley Becker, Kaitlin. "Boston Police Sgt. Placed on Leave, DA Investigating Body Cam Videos That Show 'Troubling' Footage" [18-19 December 2020]. *NBC San Diego*. [Online] available: <https://www.nbcsandiego.com/news/national-international/boston-police-sgt-placed-on-leave-da-investigating-after-body-cam-videos-show-troubling-footage-from-protest/2476181/> [accessed 18 November 2022].
2. Vera, Amir. "Boston mayor fires city police commissioner after investigation into domestic violence allegations" [7 June 2021]. *CNN*. [Online] available: <https://www.cnn.com/2021/06/07/us/boston-police-commissioner-fired/index.html> [accessed 18 November 2022].
3. DV and Police Aggression
4. McCourt, Clara. "Boston police officer arrested following domestic violence incident" [30 October 2022]. *Boston.com*. [Online] available: <https://www.boston.com/news/local-news/2022/10/30/boston-police-officer-roselyn-lacroix-arrested-domestic-violence/> [accessed 18 November 2022].
5. McKinley Becker, Kaitlin, and Mike Pescaro. "Boston Police Officer, Out on Bail After Domestic Violence Charges" [30-31 October 2022]. *NECN*. [Online] available: <https://www.necn.com/news/local/boston-police-officer-arrested-following-domestic-violence-incident/2856784/> [accessed 18 November 2022].
6. Chambers, Lauren, with Kade Crockford, and Fatema Ahmad. "Unpacking the Boston Police Budget" [undated]. *ACLU Massachusetts*. [Online] available: <https://data.aclum.org/2020/06/05/unpacking-the-boston-police-budget/> [accessed 15 November 2022].
7. "Crime in Boston" [2015-2022]. *Analyze Boston*, courtesy of Boston Police Department. <https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system/> [accessed 10 October 2022].
8. American Community Survey, U.S. Census Bureau. "Poverty Status In The Past 12 Months By Sex By Age" [2000-2021 available]. [Online]. Multiple datasets disaggregated by race/ethnicity available: <https://data.census.gov/cedsci/table?q=United%20States&t=Income%20and%20Poverty%3A/> [accessed 20 October 2022].

Appendices



The Background

BPD Sergeant placed on admin leave for bragging on bodycam about attacking demonstrators, discussing arrest quotas and the use of police cruisers as weapons, and relentlessly targeting protesters with pepper spray and batons.¹



May 31, 2020

Boston mayor Kim Janey terminates Police Commissioner Dennis White following an investigation into domestic violence allegations.² Domestic violence is prevalent among police and has been correlated to use of excessive force.³



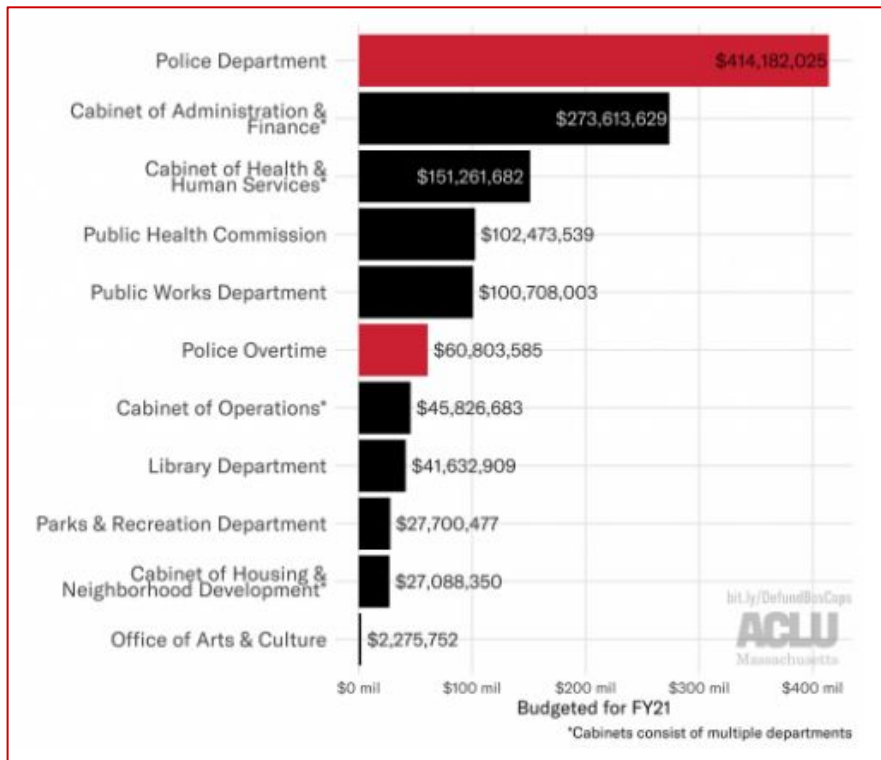
June 7, 2021

Roselyn LaCroix, 6-year police officer on paid administrative leave and \$500 bail after DV - destruction of tables, glass, and windows in the home, physical violence, and threatening his wife that he would light the home on fire with her inside.^{4,5}



October 30, 2022

The Background

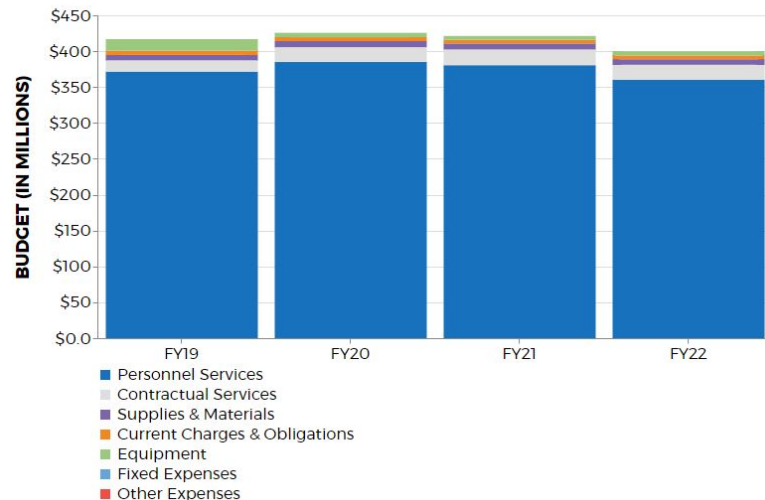


DEPARTMENTAL SPENDING

The FY19-FY20 amounts are actual expenditures. The FY21 amount is a projection and the FY22 amount is a budget.

SELECT A DEPARTMENT:

Police Department



The Data

ANALYZE BOSTON



[DATASETS](#) [NEWS](#) [TIPS](#) [LOG IN](#)

[Home](#) > [Organizations](#) > [Boston Police Department](#) > [Crime Incident Reports ...](#)



CRIME INCIDENT REPORTS (AUGUST 2015 - TO DATE) (SOURCE: NEW SYSTEM)

Dataset

Topics

Activity Stream

Showcases

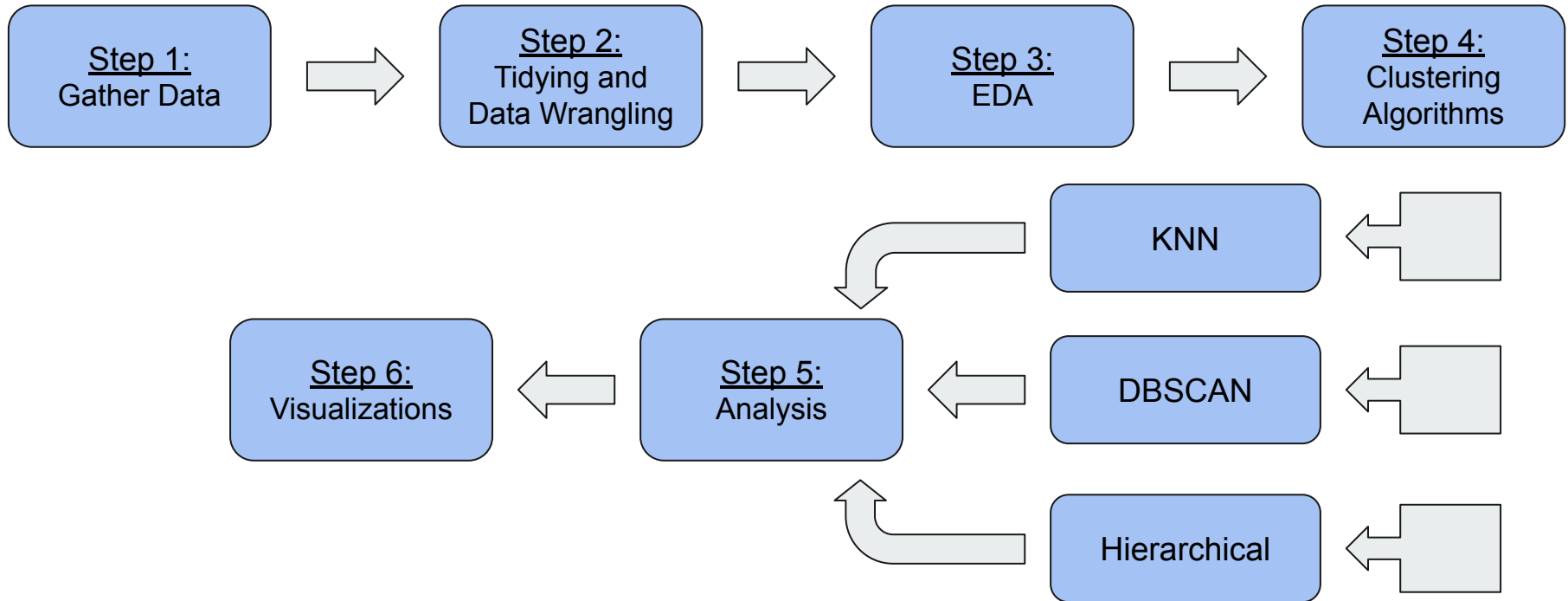
CRIME INCIDENT REPORTS (AUGUST 2015 - TO DATE) (SOURCE: NEW SYSTEM)

Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred. Records in the new system begin in June of 2015.

The Analyze Boston Data Exports posted now are the updated incident data from the Mark43 RMS Database which launched in September of 2019 and is complete through present with the exclusion of data that falls under MGL ch.41 s.98f. The 2019 data that was originally posted contained combined exports from the Intergraph RMS and the Mark43 RMS during 2019 but the Extract/Transfer/Load process was not updated during the transition. We are continuing work on producing the gap data from January 2019 to September 2019 from our legacy Intergraph RMS and will be publishing this as soon as the work is complete.

INCIDENT	OFFENSE	OFFENSE	OFFENSE	DISTRICT	REPORTIN	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_V	HOURL	UCR_PART	STREET	Lat	Long	Location		
I19207755	3115	Investigat	INVESTIGA	B3	468		4/30/2018 9:00	2018		4	Monday	9	Part Three	HAZLETON	42.27997	-71.0955	(42.27997063, -71.09553354)	
I19207733	619	Larceny	LARCENY	E18	496		3/6/2018 8:00	2018		3	Tuesday	8	Part One	HYDE PARI	42.26922	-71.1209	(42.26922388, -71.12085347)	
I19207666	2629	Harassmer	HARASSMI	E5	662		10/31/2018 12:00	2018		10	Wednesda	12	Part Two	PRIMROSE	42.29077	-71.1302	(42.29076521, -71.13021098)	
I19207538	2629	Harassmer	HARASSMI	A1	96		4/9/2018 8:43	2018		4	Monday	8	Part Two	ATLANTIC	42.35526	-71.051	(42.35526402, -71.05098788)	
I19207533	3208	Property L	PROPERTY	D4	132		1/1/2018 0:00	2018		1	Monday	0	Part Three	COMMON	42.35352	-71.0728	(42.35352153, -71.07283786)	
I19207501	2629	Harassmer	HARASSMI	E5	668		10/1/2018 9:15	2018		10	Monday	9	Part Two	CEDRUS A	42.2784	-71.1353	(42.27839822, -71.13527907)	
I19207471	1107	Fraud	FRAUD	- IN B3	458		11/18/2018 9:00	2018		11	Sunday	9	Part Two	BLUE HILL	42.30316	-71.0852	(42.30316091, -71.08515552)	
I19207471	2629	Harassmer	HARASSMI	B2	278		1/1/2018 0:00	2018		1	Monday	0	Part Two	HARRISON	42.33154	-71.0802	(42.33153805, -71.08015661)	

The Approach



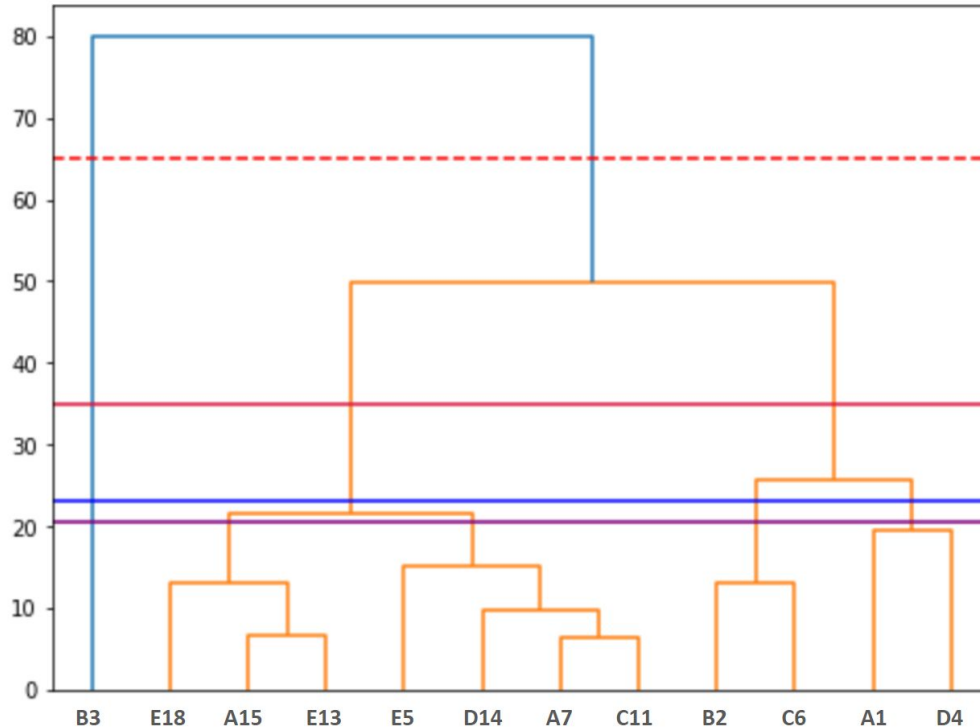
DBSCAN



District	Neighborhoods	Cluster
A1	Downtown, North End, West End, Chinatown, Beacon Hill	0
A15	Charlestown	1
A7	East Boston	1
B2	Roxbury, Longwood, Mission Hill	2
B3	Mattapan	3
C11	Dorchester	1
C6	South Boston, South Boston Waterfront	4
D14	Brighton, Allston	1
D4	South End, Back Bay, Fenway	5
E13	Jamaica Plain	1
E18	Hyde Park	1
E5	West Roxbury, Roslindale	1

- Using $\text{eps} = 10$ & $\text{min samples} = 1$ results in 6 clusters
- While most of the districts were their own cluster, it is interesting that the districts closest to downtown were clustered together

Agglomerative Hierarchical



- Most highly correlated crimes:
 - Vandalism, Verbal Disputes, and Violations
 - Restraining Order Violation and
 - Service, Vandalism, Verbal Disputes, and Violations
 - Search Warrant and:
 - Vandalism, Verbal Disputes, and Violations

