CS506 Data Science Tools and Applications

Police Overtime Project

Final Report - Team F

Ashton Fox (ashfox@bu.edu, Class of 2024)
David White (dwhite24@bu.edu, Class of 2024)
Seunghwan Hyun (shhyun@bu.edu, Class of 2024)
Ahmad Sadiq (asadiq@bu.edu, Class of 2025)
Emily Opresnick (eopres@bu.edu, Class of 2024)

Introduction:

With the Boston Police Department having a budget of over $400 million dollars, we have focused on analyzing how this money is spent. Namely, we have looked at Police Overtime and the relationship between total time worked and overtime hours. However, Overtime pay is just one category of BPD spending. We have analyzed the earnings, overtime, operating budget, field activity, and crime datasets to determine how money is used within the Boston Police Department. Additionally, to further analyze these trends we looked at datasets for race and demographics, police watchlists, complaints, and discipline. All of these datasets can be found under the data folder in the repository.

We focused on answering three main questions: Identifying financial excess in BPD Spending, Characterizing Wasteful Overtime Practices, and Filling in Narratives around waste and Misconduct by Individual Police Officers. Through our analysis of these questions we decided on an extension question which focuses on the relationship between the Boston Police Departments spending and community safety.

Data Collection and Processing:

Crime Incident:

The Crime Incident data was not provided to us and was found on The City of Boston's website. We found this dataset useful when answering our Extension Question. In order to use this we had to convert the location (latitude, longitude) to a zip code format.

Demographic:

This dataset was provided to us, however, it contains limited information about the demographics of the BPD, split by rank into the intersections of race and gender. While there are entries for several years spanning back, until 2022, the sets provided are only screenshots that we can’t make models of, so we are limited to using 2022 and 2023 demographic data. Despite this, we used this data to help give us insights as we analyzed trends in other datasets while answering the key questions.

Earnings:

This data was provided to us and contained information about the Boston City Employees' Earnings which was used to help answer our first question. The datasets ranged from 2011-2022, and we analyzed trends in different periods of time: all of the data (2011-2022), the last decade (2012-2022), and recent years (2021-2022).

The datasets contain pay information for each employee, broken up into injury, overtime, regular, retro, Quinn, and detail pay. Additionally, the total pay for each employee is provided as well as their job title and department. To focus on the pay of the Boston Police Department we had to filter this data to only contain their employees. Additionally, we had to process the columns relating to pay to be floating point numbers so we could perform computations on the data.

Field Activity:

This dataset was provided to us and contains detailed information about specific cases of arrests/scenes, both for the suspect in question and the officer who responded. Along with the location of the arrest and supervising officer. We used this data to help answer our extension question. To preprocess the data set we had to account for the ways that the format of the tables changed. Between the first data from 2012 and now there have been 4 different schemas that had a variety of table names along with data. We also had to clean the data of any NaN values as well as incorrect data types.

Operating Budget:

This dataset was not provided to us and was found on the Boston.gov website. It provides information about the proposed budget for 2024 and the budgets for 2021-2023. This dataset can help us analyze how the budget has changed over time. We first had to replace any missing values with NaN. Additionally, the expense columns had to be formatted to floating point numbers so we could perform computations. Lastly, to look at the budget for the Boston Police Department, we filtered out other departments and only looked at the data relating to the BPD.

Overtime:

Within the Overtime folder you will find both court overtime data as well as Overtime pay data. The Court Overtime data has annual records of officers requesting overtime pay when they appear in court. We have analyzed the number of records for each year to get a better understanding of overtime court appearance for the last decade. This data has STARTTIME and ENDTIME which could be used to calculate hours worked by officers.

The overtime pay dataset ranges from 2014 to 2021. It provides records of officers requesting overtime pay and the address where the overtime took place. We have analyzed the number of records for each year to get a better understanding of overtime for the last decade and the discrepancy between ‘Hours Worked’ and ‘Hours Paid’.

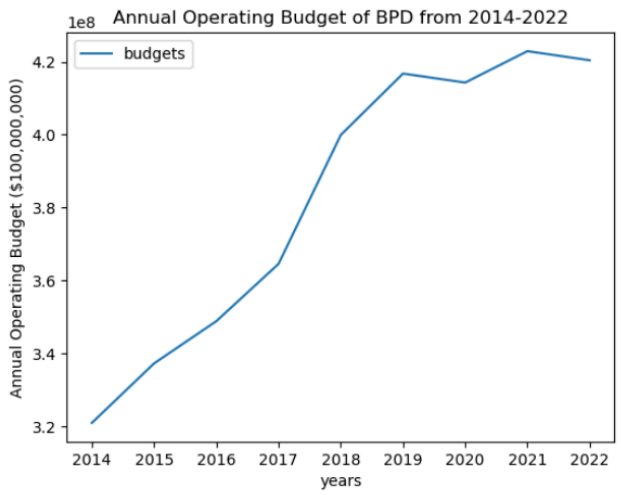
Watchlists and Complaints:

These datasets were not provided to us but required to answer the third question. They provided information about officers who are on the Suffolk County Watchlist, been disciplined for overtime abuse or misconduct, or have complaint records. To use these datasets we had to convert the format of the officer name in each dataset into the same format and remove duplicate values for complaints and salaries. Also calculating the number of years which an officer collected overtime during.

Base Analysis:

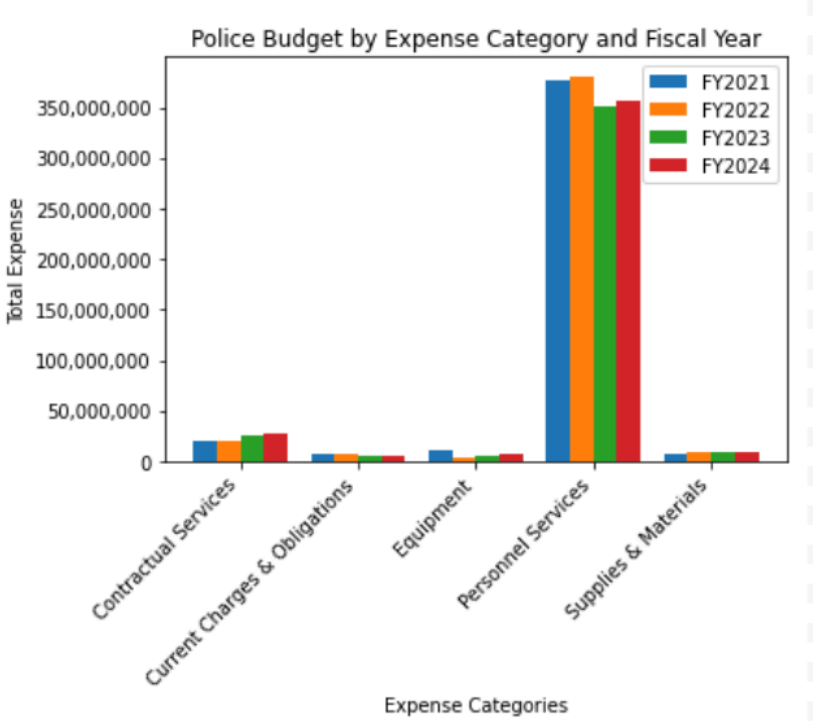
Identifying instances of financial excess in BPD spending

David, Ahmad, and Emily



Overall the BPD Operating budget has increased however recent years saw a 4.24% decrease

Our first key question was to identify financial excess in BPD Spending. We began answering this by looking at **how the BPD Operating Budget has changed since 2021**. We found that the overall budget had increased since 2014 but recent years saw a 4.24% decrease. Additionally, the budget is broken up between five categories: Contractual Services, Current Charges & Obligation, Equipment, Personnel Services, and Supplies and Materials. Personnel Services has remained the top category for budget in all four years. This category deals with employee pay including overtime and regular pay.

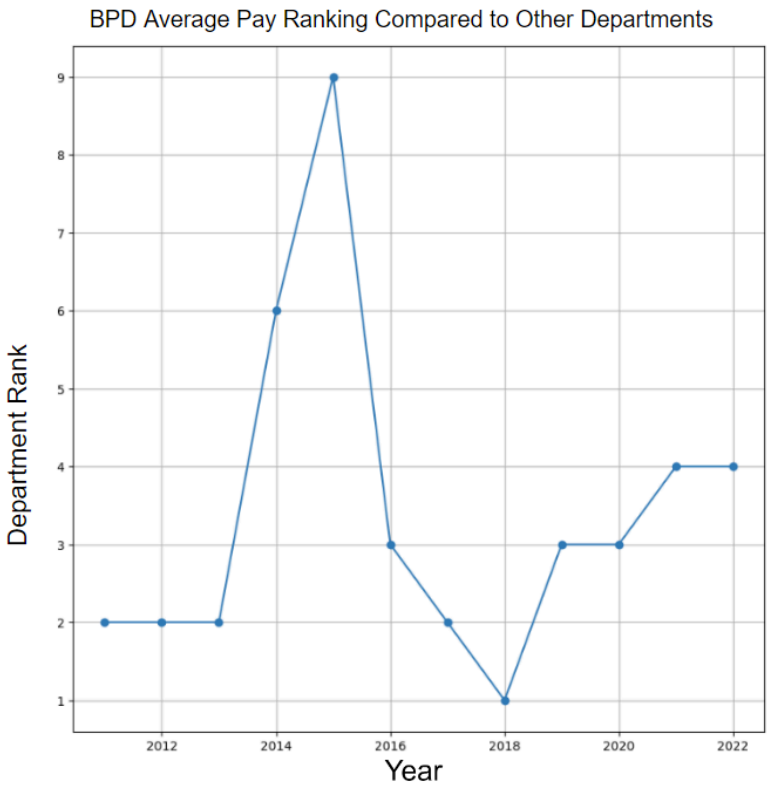


Personnel Services makes up a majority of the BPD Operating Budget and has remained the top category as it includes employee pay.

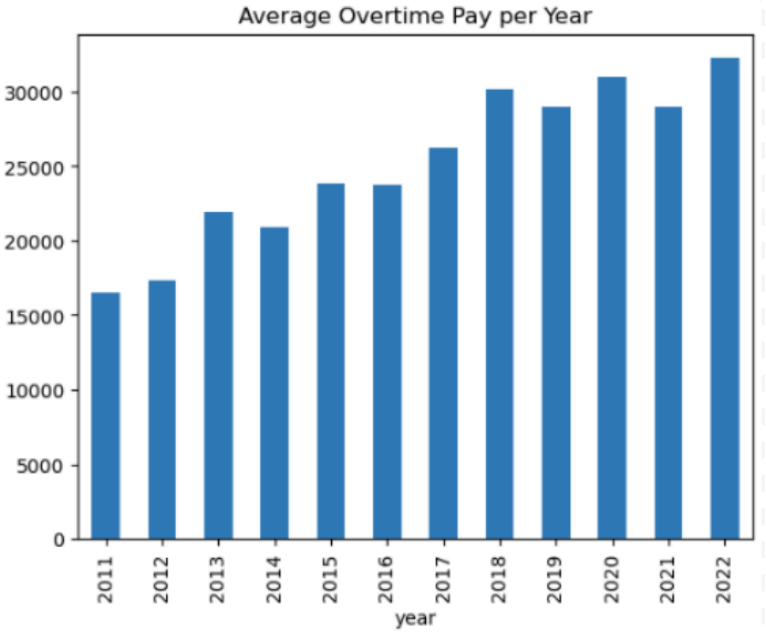
We then looked at the Earnings Data to further analyze how this budgeting money was being spent. From this we could look at **how funds have changed intra-departmentally**. We found that Officers are consistently the highest percentage of the combined pay for BPD employees, this makes sense given that there are more police officers than other roles. Surprisingly of the top 10 highest combined pay roles almost all had an uptick in combined pay in 2018 while many had a downtick in 2021.

Police Officers consistently had the highest percentage of combined pay for BPD Employees. Looking at their total earnings by year, there was a large increase in 2018 and has been decreasing since 2020.

Police Detective is somewhat of an exception among the top 10 highest combined pay roles as it reached a peak in 2018 and after falling in 2019 has remained consistent. To compare to other departments, we looked at the average pay of BPD employees versus other department employees and determined the top 10 earning departments. From that we could determine how frequent departments appeared. The BPD was frequently ranked one, meaning they had the highest average pay for many years. BPD on the other hand, frequently ranked 2,3,4 and in 2018 ranked first.



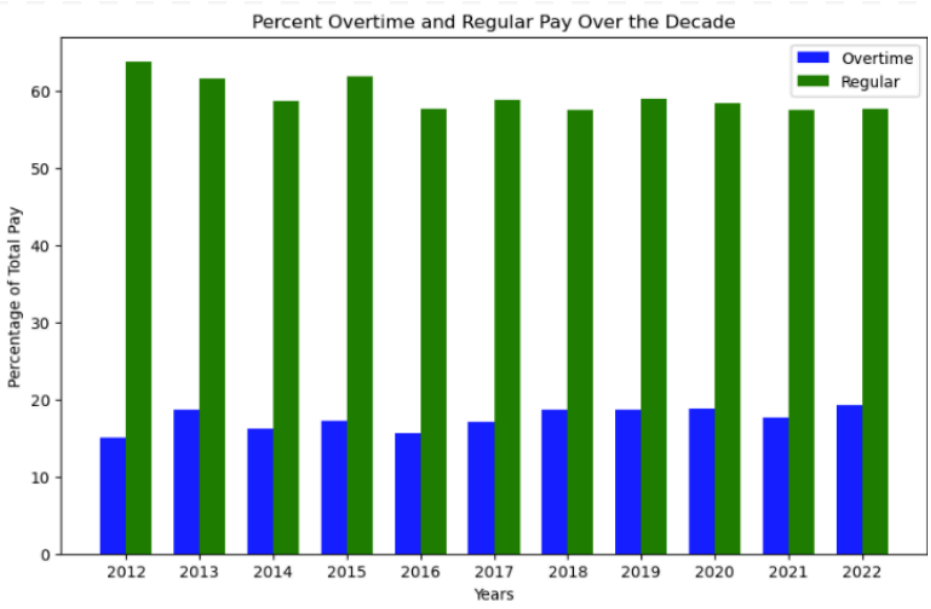
The BPD has remained one of the top 10 departments since 2011 when ranking by average pay. In 2018 they were the top department and since 2016 have remained in the top 5.



The Average amount of overtime pay has doubled from 15k in 2011 to 30k last year.

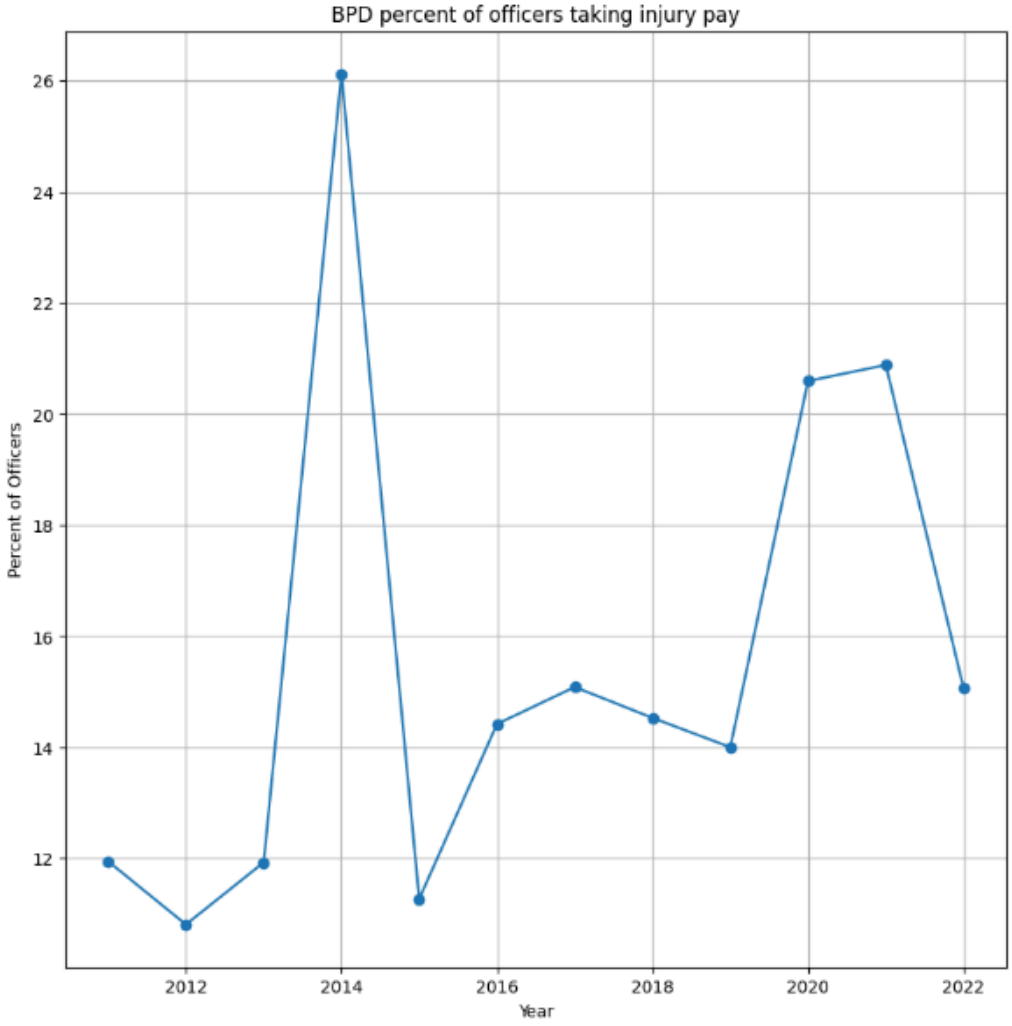
Additionally we analyzed how **BPD paychecks have changed throughout the years**. We found that the average pay, however, has gone up from a little over 60,000 in 2011 to upwards of 80,000 in the past 3 years. Regular pay has gone up slightly in the past 10 years, while overtime pay has gone up significantly – it doubled from 15k in 2011 to 30k last year.

We looked at the breakdown of this pay by analyzing changes in overtime, quinn, retro, injury, detail, other, and regular pay have changed from 2012-2022. To do this, we looked at what percent each category made up of the total pay for each year. From this data we found that quinn, other, and injury pay had increased whereas detail and retro have decreased. Overtime and regular pay have remained consistent.

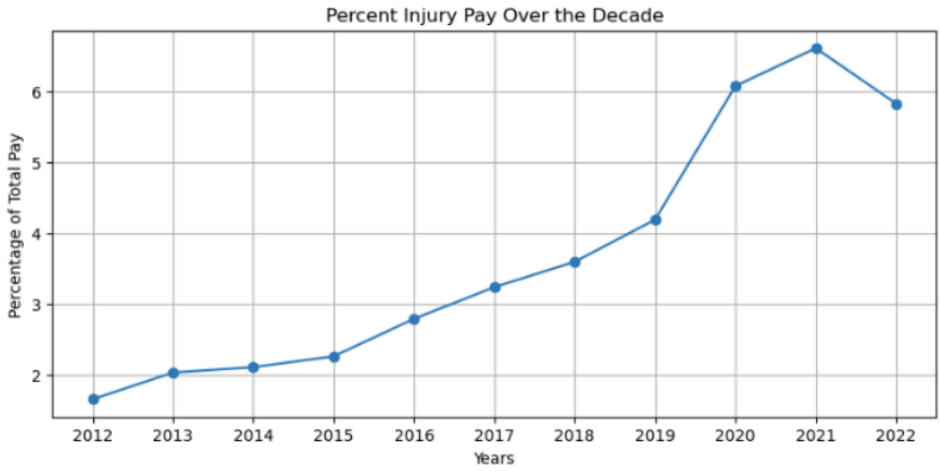


The percentage that Overtime and Regular Pay make up of the total pay have remained consistent

Lastly for this question, we looked at **how much BPD Officer pay came from injury pay and how many officers claimed injury pay**. Looking at BPD earnings from 2012-2022 we found that Injury pay has increased overall, with a slight decrease in 2022. Additionally we found that on average across the decade 15.9% of BPD employees claimed Injury pay.



On average about 15.% percent of BPD Employees took injury pay. The percentage of officers who took injury pay peaked in 2014 with 26%.

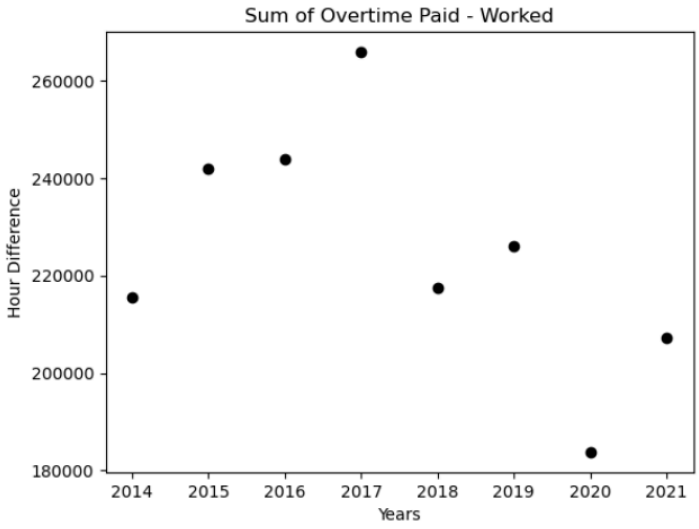


The percentage that injury pay makes up of the total pay has increased overall, however there was a slight decrease in 2022.

Characterizing wasteful BPD overtime practices

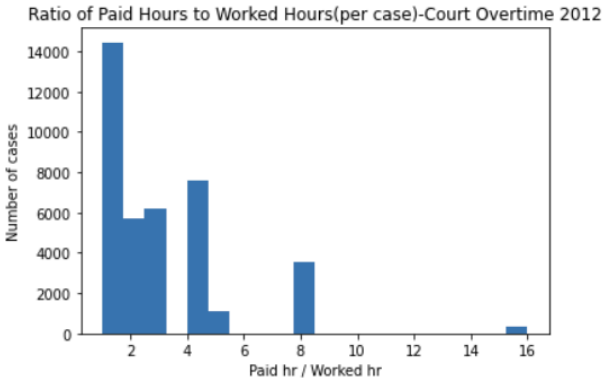
Jason and Ashton

The next question we had to answer was to characterize wasteful BPD Overtime practices. To find **wasteful overtime practices**, we focused on the discrepancy between overtime hours paid compared to overtime hours worked by analyzing overtime data from 2014 to 2021. Comparing the overtime hours paid and worked, the annual difference ranges from 180k hours to 260k hours but no clear pattern was found. The year of 2017 displayed maximum value and year of 2020 had minimum value.

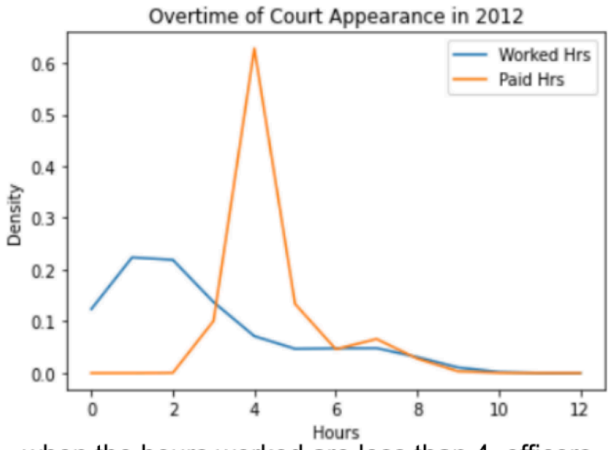


There was no clear pattern for the discrepancy of hours worked and overtime hours

To add **narratives to misconduct in overtime pay**, the distribution for worked hours and paid hours are very different for hours less than 6. We found that when hours worked are less than 4, most officers were paid for 4 hours. This trend was visible throughout all years



The distribution for ratio of paid hours to worked hours is skewed to the right. Outlier with value of 16 has 4 paid hours and 0.25 worked hours



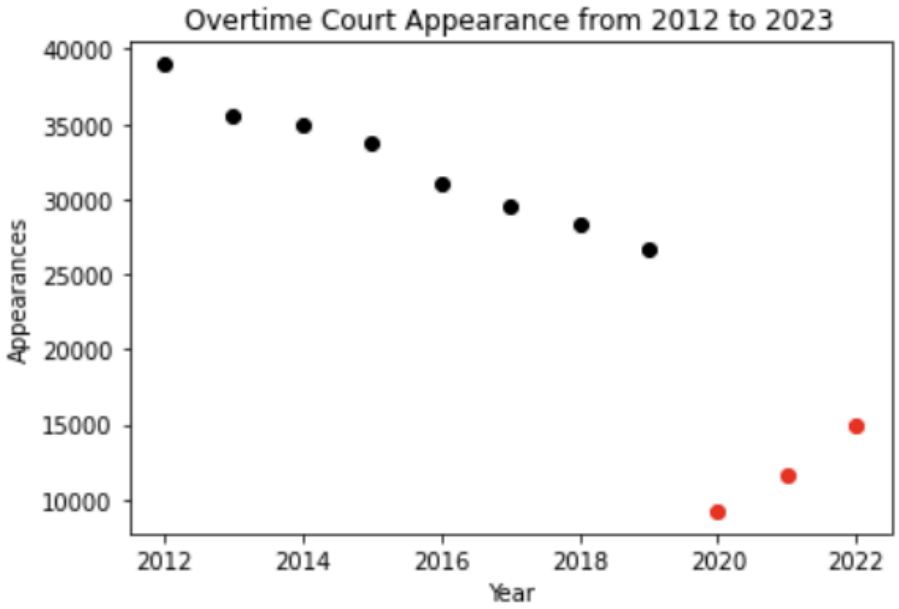
when the hours worked are less than 4, officers tend to request for 4 hours of overtime

BPD's overtime data did not have hourly pay for each officer and having the same rank can't guarantee equal pay, so we **could not calculate the financial discrepancy** due to lack of data.

Distribution of ratios of overtime worked vs. overtime paid is skewing to the right as time goes by. As the data gets skewed, the number of outliers is increasing. For 2012, the

greatest outlier was 16. This is because most discrepancy happens when the hours worked is less than 4. So an outlier with a value of 16 is an officer who worked 15 minutes billing 4 hours.

Overtime for court appearances have declined from 2012 to 2019 by 4,000 each year from 40,000 to 26,000. However, after COVID-19, appearances are increasing by 3,000 each year.

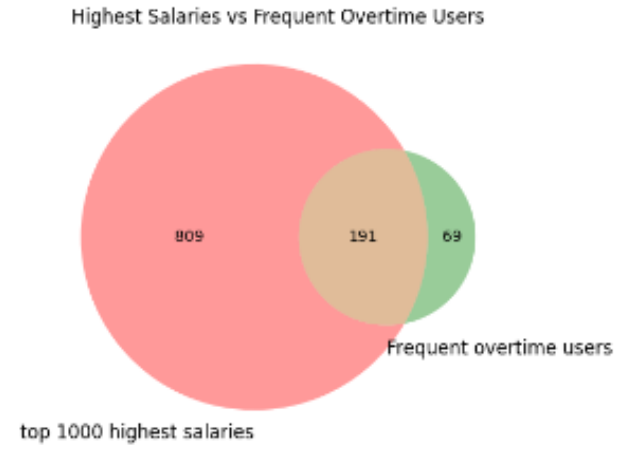


Overtime for court appearances have declined from 2012 to 2019

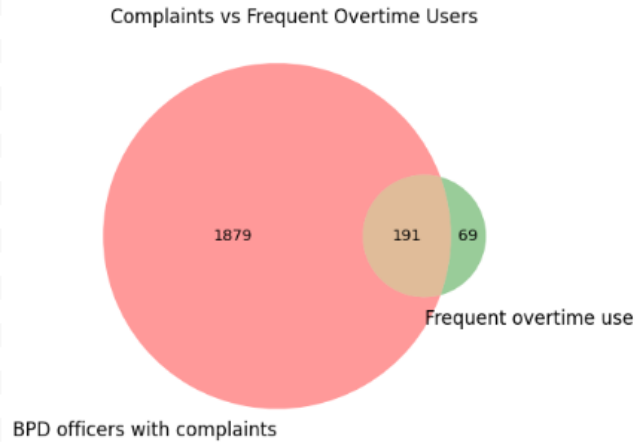
Using data to fill in narratives around waste & misconduct by individual BPD officers

David and Ashton

Our final question was to look at the **overlap between frequent overtime users and officers who have the highest salaries, are listed in Suffolk County police watch list, have been previously decided for overtime abuse, or have internal affairs complaints.** To do this we looked at the top 10% of officers who frequently claimed overtime (260 officers). We found that there was not a large overlap between frequent overtime users and the Suffolk County police watch list or officers who have a record of previous disciplinary action. We found that the vast majority of officers who were on the list had some form of internal complaints record, with 191 of the 260 checked.



A Venn Diagram of the overlap between highest paid BPD officers and the most Frequent Overtime Users Showing that 75% of the BPD's highest paid officers frequently use overtime.



A Venn diagram of the overlap between BPD officers with complaints and the most frequent Overtime users showing that roughly 70% of the most frequent overtime users have complaints against them.

We found that 73% of the officers who were frequently using overtime had a salary in the top 1000 salaries in the Boston police. Most notably among the 5 datasets we checked there was one officer who appeared on all of the lists. This officer has previously been disciplined for submitting false timesheets, and is currently under investigation for corruption. Overall there does appear to be some connection between the officers who took the largest amount of overtime per year and the officers with the highest pay, and with a record of internal and external complaints.

Extension Analysis:

The Impact of Boston Police Department spending on Community Safety

Jason, David, Ashton, and Emily

Background

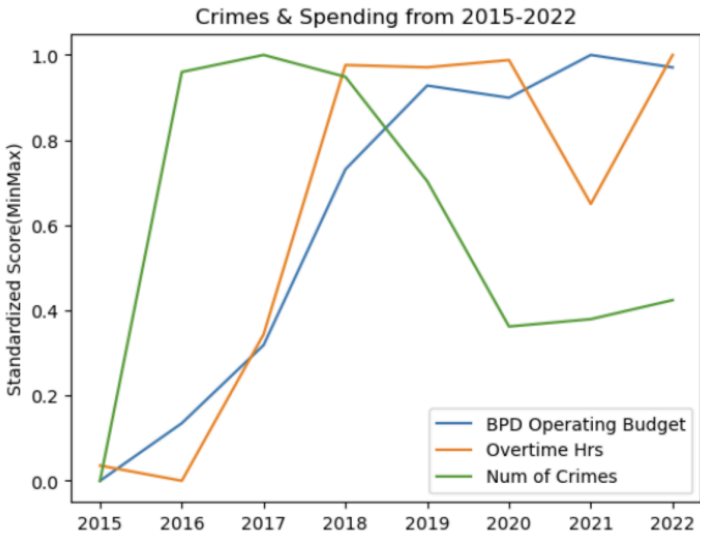
For the Extension project we focused on analyzing the impact of the Boston Police Department spending on Community Safety. As we were looking at the different datasets surrounding BPD spending, we wanted to explore how these numbers were impacting community safety. Given the significant budget allocation to the BPD, it's important to understand how police presence and activities can correlate with crime rates, community safety, and ultimately public trust in Boston. This analysis can reveal insights into the effectiveness of police spending and strategies, and whether they align with community needs.

In order to answer this, we created a few questions to guide our analysis. We focused on seeking out any correlations between overtime pay and crime rates as well as the BPD budget and crime rates. Additionally, we aimed to answer if the areas with higher amounts of field reports had higher or lower crime rates to determine if the number of on field police officers impacted crime rates.

To answer these questions we used crime datasets from the City of Boston, which are located in our GitHub repository data folder under Crime Incident. Additionally, to analyze the relationship between budget and crime, we needed more budget data from Boston as the csv file had data from 2021-2023. To accomplish this, we manually entered the budget data from 2014-2022. This information can be found on the City of Boston website, which has documents containing the operating budgets for current and past fiscal years.

Analysis

To check the relationship between overtime hours, BPD budget and number of crimes, we plotted each data's value from 2015 to 2022. Data was standardized using min-max normalization to scale the values from 0 to 1 to make relationships more visual.



We graphed the standardized values of Budget and Overtime to compare this to the standardized Number of Crimes

We predicted the **relationship between the number of crimes and overtime hours** to be positive. Contrary to our belief, our analysis using linear regression model gave us a R-squared of 0.000 and ‘overtime’ was not a significant variable to predict the ‘number of crimes’ as high p-value(0.972) demonstrates we fail to reject the null hypothesis. We did discover that the number of crimes could be a leading indicator for overtime hours by 2,3 years.

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.964e+04	6.75e+04	1.181	0.282	-8.54e+04	2.45e+05
overtime	3.443e-05	0.001	0.037	0.972	-0.002	0.002

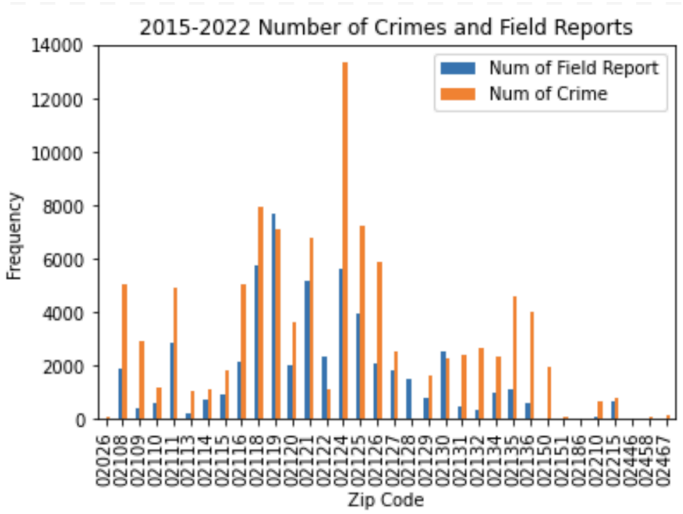
Using Linear Regression we found no relationship between the number of crimes and hours of overtime

Relationship between the number of crimes and BPD budgets was also analyzed using a linear regression model. R-squared was 0.000 and ‘budgets’ was not a significant variable to predict the ‘number of crimes’ with high p-value(0.884).

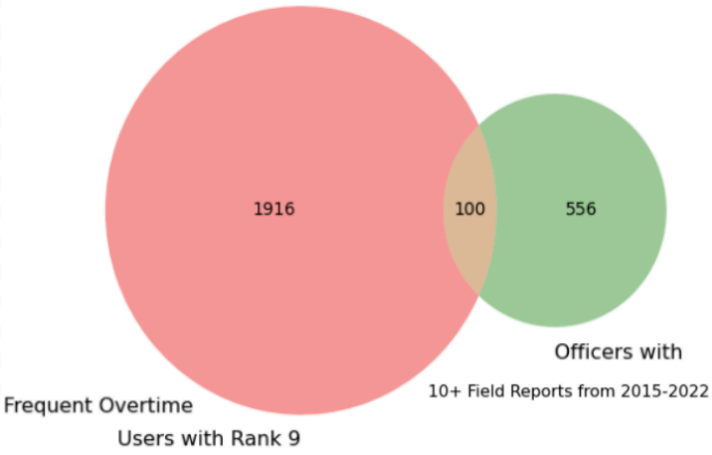
	coef	std err	t	P>|t|	[0.025	0.975]
const	9.411e+04	7.91e+04	1.189	0.279	-9.95e+04	2.88e+05
budgets	-3.071e-05	0.000	-0.152	0.884	-0.001	0.000

Using Linear Regression we found no relationship between the number of crimes and BPD Budget

We also analyzed how the **number of on field police officers and how they impact the crime rates** for different parts of Boston using the zip code. The number of police reports made per zip code was often considerably less than the amount of reported crime for that area between 2015-2022. We do not have an explanation for zip codes where there were more field reports than actual reported crime. From this we can conclude that police presence did not impact crime rates.



To see the relationship between crime and police activity we graped the number of field reports and crime reports from 2015-2022 by zip code.



There is not a large overlap between frequent overtime BPD Employees with rank 9 and officers with more than 10 field reports from 2015-2022

We looked into **if field officers with more overtime were in places with higher or lower crime rates** to determine if these extra hours worked had an impact on community safety. We found that there is not a large overlap between frequent overtime BPD Employees with rank 9 and officers with more than 10 field reports from 2015-2022.

Challenges And Limitations:

We have faced a few challenges throughout the project so far. The first challenge we faced was that the earnings data sets did not have a column for the employee identification number. This has been a challenge in our data analysis as there might be officers with the same names. Additionally, this has made it difficult to merge the earnings data with overtime data when finding overlap between the two. As a result, we have to assume that names are unique in the datasets.

We also faced a challenge with the budget data as the data only contained budget information from 2021-2024. As a result, we manually input the budget information for years 2014-2020 with data found on the City of Boston's Website.

When looking at demographic data, the information only goes back to 2022 which is limiting when looking at trends. We found that the demographic data only contains information about race and sex. There was no information about factors such as name, age, title, rank, tenure, and location which has also been very limiting as we cannot draw any conclusions about trends with overtime and demographics.

The overtime data's 'Hours Worked' columns and 'Hours Paid' were challenging to work with as the formats for expressing time were different. The 'Worked Hour' column had two formats: '%H' and '%H%M'. So we had to check if the value was less than 12 to see if it was in hour form. If not,, the last two digits were minutes and other digits were time. To unify the format, we created new columns 'WRKDHRS' and 'PAIDHRS' which were calculated by counting the amount of time in 'seconds' and dividing by 3600 so the units of columns were hours.

Additionally, The overtime dataset provides the officers rank, however, there was no key explaining the equivalent role. This was a challenge for us when answering our extension question, as we were looking at the overlap between frequent overtime users and field officers with 10 plus reports. We wanted to narrow down the frequent overtime users set to only include officers who would be conducting field reports to get an accurate representation of the overlap. However, without any knowledge of the officer's job title, this was challenging. To overcome this, we looked at the overlap for each of the ranks, and found that rank 9 officers had the greatest overlap with field officers. We were then able to draw conclusions after making this judgment. Also this is why we could not calculate the financial discrepancies for different years because we could not assume the hourly rate for overtime pay especially when we couldn't find what 'OTCODE' was and how overtime was calculated.

Lastly, when we delved deeper to analyze the regional relationship between number of crimes and number of field reports, we had to add a new column in the crime datasets which contained the zip code. We did this by using the longitude and latitude columns, however, it was very time consuming to convert these to zip code format. To ensure that the client would not have to run this timely process, we added the crime datasets with the zipcode to our data folder in the repository.

Conclusion:

Identifying instances of financial excess in BPD spending

The Annual funding of the Boston Police Department has increased most years since 2014, increasing from 320 million dollars per year in employee pay to 420 million dollars in 2022. In the next few years police funding is projected to go down slightly, but the vast majority will continue to be in employee pay. We found that the police department is consistently one of the highest paid departments by Boston, peaking at the most paid department in 2018. Despite the increase in overall pay the ratio of overtime to regular pay has remained the same, but injury pay has significantly increased over the past decade.

Characterizing wasteful BPD overtime practices

Annual sum of overtime hours (2014-2021) for the Boston Police Department displayed no particular pattern and remained within the 180k to 260k range. When it comes to discrepancy between paid hours and worked hours, we discovered that for cases when worked hours are less than 4, most officers will get paid for 4 hours and this is where the majority of discrepancy takes place. Because worked hours were small, the distribution of the ratio of paid hours to worked hours has an outlier of 16 and mean of 3 which makes the matter seem much more serious. Court appearances for overtime has been on a steady decline until COVID-19 from 40k to 26k but after COVID-19, it has been increasing for 3k every year from 10k in 2020.

Using data to fill in narratives around waste & misconduct by individual BPD officers

The Boston Police Department's most common users of overtime were often correlated with the groups that we expected, highly paid officers with a history of misconduct and complaints against them. There is a high correlation of 75% between overtime and these groups, and it seems likely that the more of the groups an officer belongs to the more likely they were to have abused overtime.

The Impact of Boston Police Department spending on Community Safety

Analyzing the annual Boston Police Department operating budget, total overtime in hours and number of crimes, we found that BPD's operating budget and total overtime in hours displayed a similar pattern. However, their pattern was different from the number of crimes in that while the number of crimes decreased from 2017, BPD's operating budget was steadily increasing every year. Using linear regression analysis to find the correlation, p-value was greater than 0.88 for overtime and BPD budgets. We delved deeper to check if areas with more crimes had more field police officers and discovered that the number of crimes had no relationship with the number of field reports and concluded that police presence has no relationship to the crime rates. Also, officers with more overtime had little relationship to being stationed in communities with higher crime rates.

Individual Contributions:

Ashton Fox: I worked on trying to answer a question we were given about racial/sex/rank demographics of BPD members who abuse the overtime mechanics, but the datasets that we were given were too shallow and only provided race and sex data indiscriminately across the BPD force, so the data analysis of this ended up leading nowhere. Before then, I looked at the campaign contribution and field activity data, but wasn't able to utilize it in any meaningful way until the extension proposal, where I was able to compile the density of where crime reports were filed by zip code, and pair it together with field report zip codes. This was helpful in the extension because it both helped to answer if, over the years, increased policing led to less crime, and if more policing translated to more overtime.

David White: For the first question I worked on processing data and determined examples of intradepartmental changes in BPD funding by title. As well as annual change in funding. I also worked on processing and visualizing the change in average pay per department in Boston to determine how the BPD was paid compared to other departments in the City of Boston. Then I worked on processing and analyzing the overtime, payroll, and complaint/s/discipline datasets. I also worked on preprocessing and doing a preliminary analysis of the field activity report dataset.

Seunghwan Hyun: I worked on a second question to find wasteful overtime practices. I analyzed the discrepancies between paid hours and worked hours by plotting annual differences, comparing the distributions of annual data and drawing the distribution of ratio of paid hours to worked hours. I also visualized annual overtime court appearances from 2012 to 2023 to find the pattern.For an extension question, I found the annual BPD operating budget on a government website and visualized the annual BPD operating budget, overtime hours and number of crimes and discovered the relationships among different variables. I used Min-Max standardization to smoothly compare the patterns. I also analyzed the number of crimes and number of field reports by zip code using the 'uszipcode' library to add a zip code column to crime data and visualized it with a side-by-side bar graph.

Ahmad Sadiq: I worked on identifying instances of financial excess in BPD spending, where I preprocessed the data and analyzed the budget and overtime datasets. I visualized the graphs in my own work. I compared the year-over-year changes in the BPD budget, pinpointing areas where funds had increased or decreased, both overall and within specific departments. By breaking down the data, I was able to discern how funding allocations shifted between departments over the two years. For the extension proposal question I proposed the idea.

Emily Opresnick: I focused on answering the first question about Identifying instances of financial excess in BPD spending. I did this by analyzing the earnings data from 2012 - 2022. Additionally, I looked at the BPD budget data to further analyze how the BPD is spending their budget. Lastly, I worked on answering the extension question by looking at budget, crime, overtime, and field report data. I helped identify if there was any relationship between the data through conducting linear regression as well as looking at overlaps between overtime officers and field officers.