

Police Overtime - D

Problem Statement:

The problem at hand is to conduct an in-depth analysis of how the Boston Police Department (BPD) allocates and spends its operating budget, with a specific focus on the registration and utilization of overtime. The BPD operates with a budget exceeding \$400 million, and understanding how this budget is distributed is of paramount importance to ensure accountability, transparency, and to address potential issues of financial excess and wasteful spending. Additionally, there is room for creative data science work, including exploring sociological, environmental, and political dimensions related to policing and its impact.

Data Collection:

Primarily, the stakeholders have provided us access to the following 4 datasets on which we perform our exploratory data analysis-

1. [Employee earnings data \(search police\)](#) - This dataset contains information on the earnings of all employees of the City of Boston, including job title, name, department name, regular pay, gross pay, injury pay, "other pay" and overtime pay.
2. [Campaign contribution data](#) - This dataset has information on the contributions made to the house representatives, senator and CC contributions, with a focus on contributions made by the police force.
3. [BPD field activity data](#) - Spanning from 2011 to 2022, this dataset details interactions between the Boston Police Department (BPD) and the public via three record management systems: OLD RMS, NEW RMS, and MARK43. It sheds light on the frequency and duration of stops, highlighting specific officers and supervisors involved. The data aids in recognizing patterns, potential discrepancies, and mapping active officers and supervisors using key variables like 'stop_duration', 'contact_officer', and 'supervisor'. It offers a comprehensive view of BPD's field activities and trends over the years.
4. [Overtime data from 2012-2022](#) - This dataset provides valuable information to discover the discrepancies between the amount of authorized overtime hours and the amount of hours actually worked. Also, it provides insights on the overtime details for the duty performed by the police in the Court.

Analysis / Tasks Performed:

Earnings data:

Performed analysis on the earning-data.csv dataset which includes general earnings data for city of Boston employees, and answered the following questions:

-How much BPD officer pay came from injury pay? What percentage of officers took injury pay in a given year?

Percentage of BPD Officers Who Took Injury Pay: 15.07%

Total Injury Pay for BPD Officers: \$23,628,467.04

Percentage of Total Gross Pay from Injury Pay: 5.82%

-Identifying instances of financial excess in BPD spending,

-Characterizing wasteful BPD overtime practices

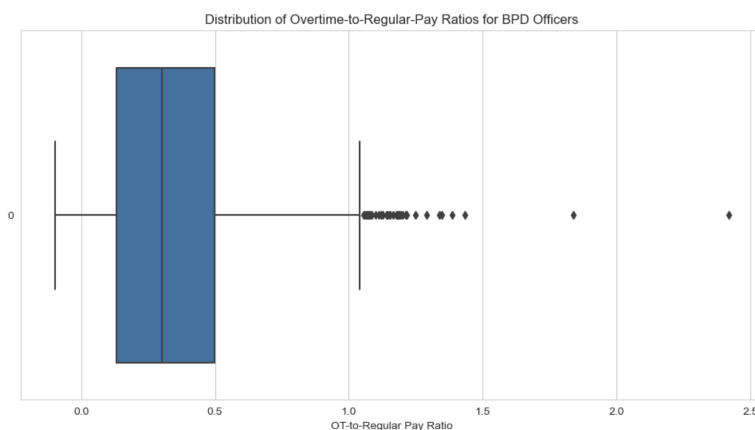
Performed analysis on the ratio of overtime to regular pay, and identified the top officers who took the most overtime pay, as well as created a boxplot to show the distribution of how officers take overtime pay:

Percentage of BPD Officers Who Took Overtime Pay: 79.77%

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

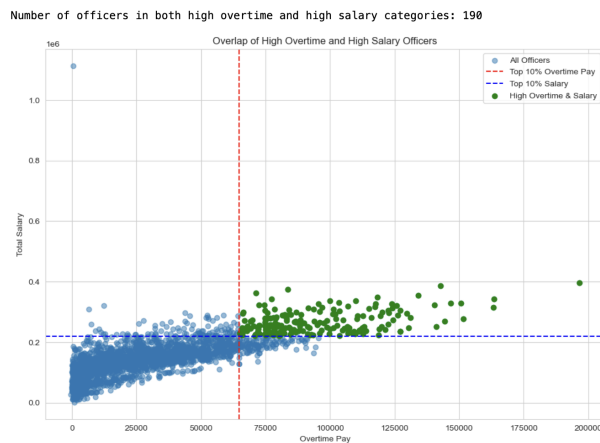
Top 10 officers who took the most overtime pay:

	NAME	REGULAR	OVERTIME	OT_to_Regular_Ratio
20438	Stewart,Greta E.	1977.92	4785.93	2.419678
11518	Woodley,Lorenzo I	11775.11	21638.24	1.837625
1846	Webster,Geneese	65294.61	93654.82	1.434342
111	Acosta,Jose L	109502.02	151608.89	1.384531
3	Demesmin,Stanley	145775.26	196515.25	1.348070
135	Jones,Craig D	108038.52	144457.87	1.337096
203	Christie,Albert C	109502.02	141222.99	1.289684
14	Barrett,Thomas E.	130930.12	163494.70	1.248717
387	Cornelius,Devon M	92637.86	112645.22	1.215974
260	Shikoluk,Michael	99922.46	121258.11	1.213522



It appears that there is a heavy skew in the amount of overtime pay taken by officers, with a large amount of outliers well above the average. In addition, the average ratio of overtime to regular pay is quite high, hovering around 30% of the regular pay being overtime.

-How much overlap is there between frequency overtime users and officers who have the highest salaries on the force?



BPD field activity data:

Upon investigating BPD field activity data. As the record system of BPD has changed twice from the past 10 years (old RMS, new RMS, Mark43), started with conducting a thorough analysis on the combined dataset, which encompasses detailed stop-and-frisk records for a specific region over the span of 2011 to 2022. I addressed the following research inquiries:

Understanding and Cleaning stop_duration Data:

Examined the 'stop_duration' variable, which initially contained a mix of string descriptors and numerical values.

Mapped string values to a median time, and converted everything to minutes for standardization.

Analyzed the distribution of stop durations and identified potential outliers or erroneous data entries.

Key Findings:

Total number of cases: 211,522

Cases with non-empty stop_duration: 35,971

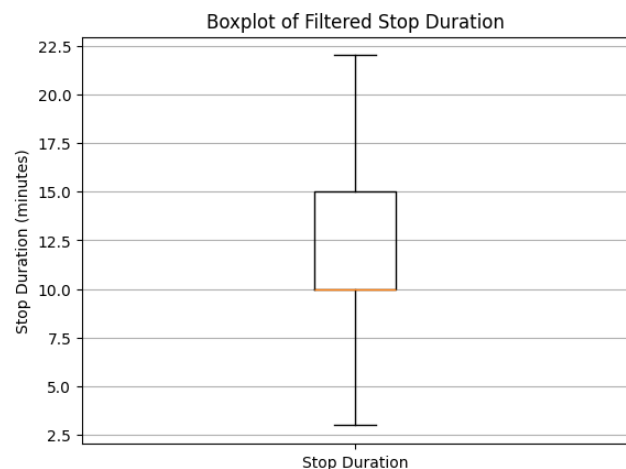
Average stop_duration: 49.47 minutes

```
count    37810.00000
mean      50.27818
std       2187.32763
min        1.00000
25%       10.00000
50%       10.00000
75%       15.00000
max      389700.00000
```

Name: stop_duration, dtype: float64

drop outliers by calculating the IQR (1.5)

```
count 33314.000000 mean 11.174971 std 4.489778 min 3.000000 25% 10.000000 50%
10.000000 75% 15.000000 max 22.000000 Name: stop_duration, dtype: float64
```



Analysis of Officer and Supervisor Activity:

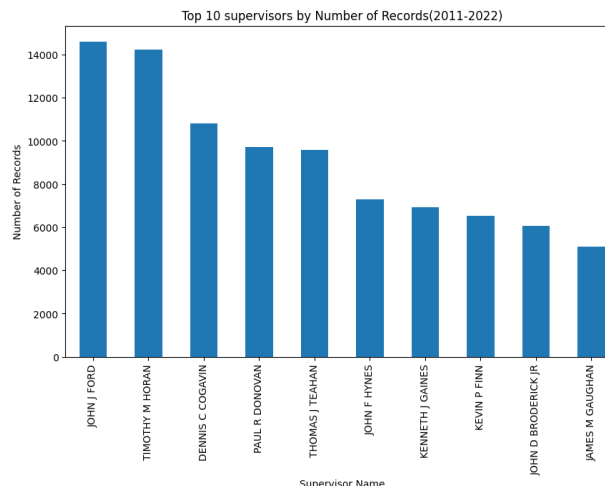
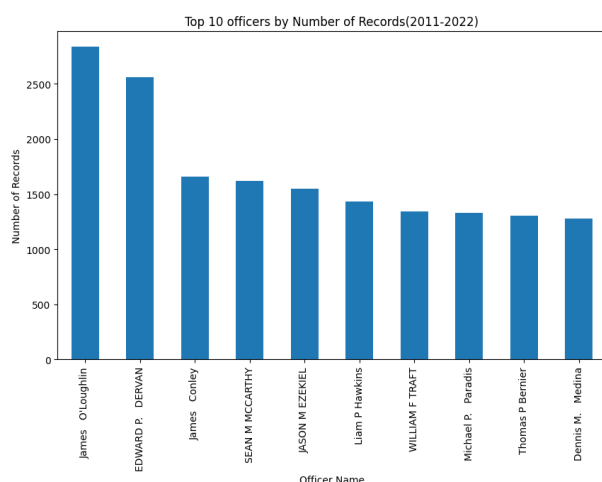
Explored the data to understand which officers and supervisors had the highest number of records.

Initially utilized the unique IDs of officers and supervisors for this analysis.

Later, enhanced the visual representation by mapping IDs to actual names to offer a clearer understanding of the most active personnel.

Challenges Addressed:

Dealt with potential naming inconsistencies by ensuring that a unique mapping between IDs and names was maintained. This helped in addressing potential issues arising from multiple naming conventions for the same individual.



Campaign Contribution Data & Overtime Data:

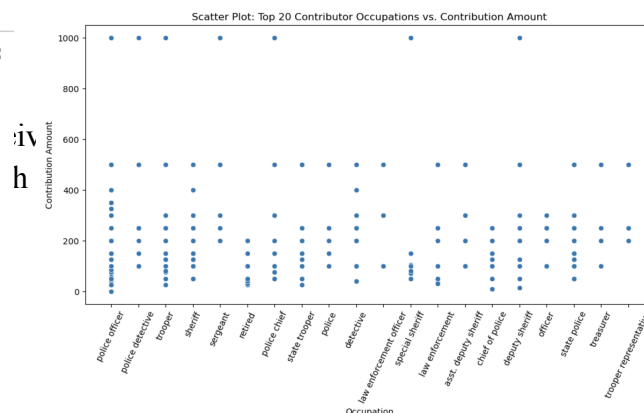
Exploratory data analysis was performed on the following two datasets as well, **Campaign contribution data** and **Overtime data from 2012-2022**, with which we were able to answer quite a few of the base questions.

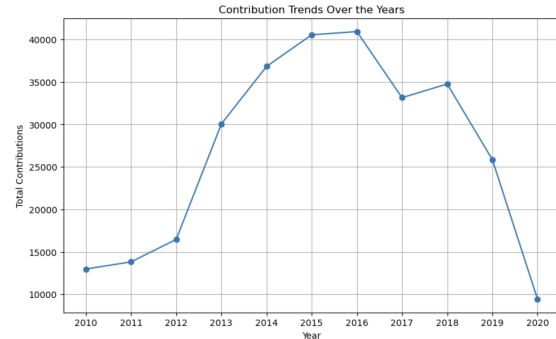
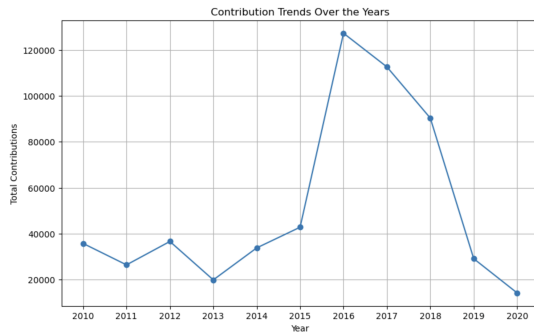
In the Campaign contribution data we first explored the data which provides us with details of the contributions made to the House Reps and Senators. Notably this dataset had 3 pieces of vital information, the contributors name, the recipient's name and the amount received by the recipient/house rep/Senator. We find the name and amount received by the top 10 recipients.

We create a scatter plot to observe contributor occupation vs. contribution amount for the house rep and Senator datasets respectively.

Top 10 names who have received the most contributions:

Matching Names	Contribution Amount
Michael Moore	163730.0
Stanley Rosenberg	100800.0
Karen Spilka	72800.0
Michael Rodrigues	46830.0
Walter Timilty	29925.0
Michael Rush	24920.0
Bruce Tarr	22750.0
Patrick O'Connor	17850.0
John Velis	8750.0
Marc Pacheco	8400.0

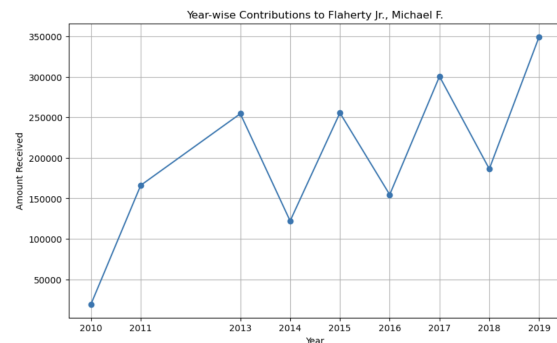




We also dived into the active CC contributions to discover that ‘Flaherty Jr., Michael F.’ is the recipient with highest contributions received (2010-2019) and hence we further analyzed the contributions made to him, finding the city from where he received most of his contributions and a year wise trend of how much he received.

	Recipient	Amount
7	Flaherty Jr., Michael F.	1808742.76
14	Wu, Michelle	1486452.16
4	Campbell, Andrea Joy	897720.68
13	O'Malley, Matthew J.	692661.37
9	Flynn, Edward Michael	686292.89
1	Baker, Frank	684131.47
5	Edwards, Lydia	554057.77
6	Essaibi George, Annissa	367026.41
10	George, Annissa	296705.87
11	Janey, Kim	249827.23
2	Bok, Priscilla MacKenzie	205397.47
12	Mejia, Julia M.	152378.26
0	Arroyo, Ricardo N.	138424.77
3	Breadon, Elizabeth A.	36725.10
8	Flaherty, Michael F.	10102.07

Year	Contribution Count
0 2013	1264
1 2015	1055
2 2019	994
3 2011	950
4 2017	938
5 2018	614
6 2014	612
7 2016	526
8 2010	168



In the overtime dataset we have two kinds of data. One, based on the yearly details ranging from 2012-2022 on how overtime has been used by the police officers and another based on the police officer's overtime based on his services in the Court. The main fields found upon exploring these datasets are Employee (Officer name), hours worked and hours paid.

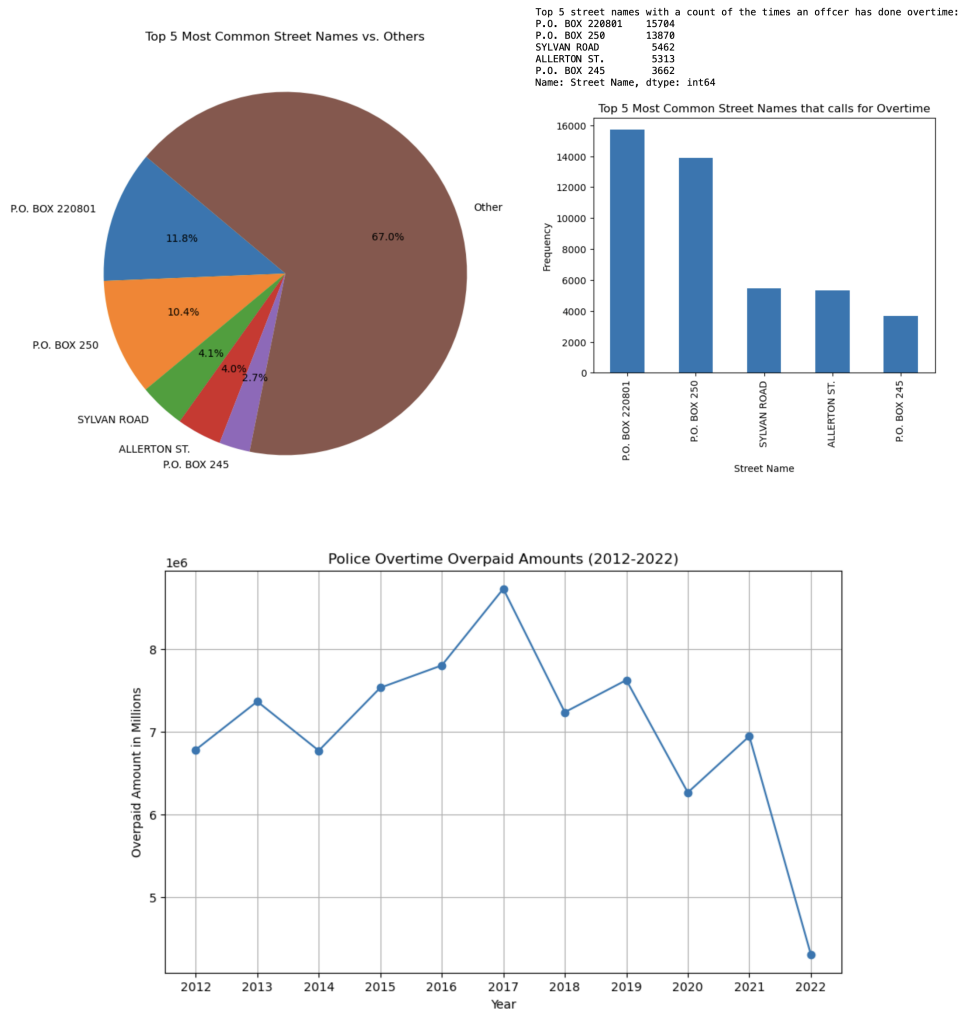
For the yearly dataset, we find the total amount of overpaid overtime hours and find the average rank of the officer who has applied for overtime in that year. On multiplying this we get answers to some of the base questions in the project. For instance, from the analysis of financial excess of BPD spending we can conclude about the amount of money the state would save in the year of 2022 if the police officers were paid according to their actual worked hours amounts to **6,781,003.98 \$**. We did this for every year and visualized a yearly trend via a line chart.

We also have extracted other crucial insights as shown by the images below for the year of 2012-

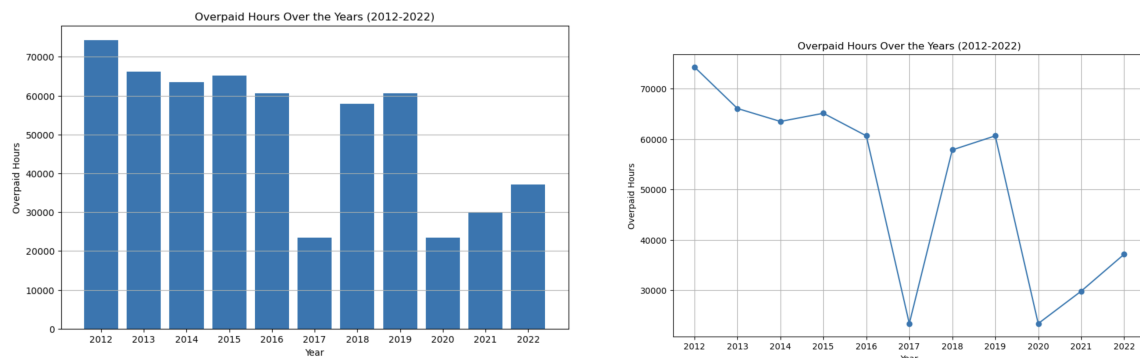
Matching Names	
Timothy Whelan	30750.00
Robert DeLeo	25275.00
Rob Consalvo	14500.00
Harold Naughton	11132.73
William Lantigua	10885.00
Daniel Hunt	10310.00
Brian Dempsey	9700.00
Jr. Puppolo	8500.00
Edward Coppinger	8350.00
Adrian Madaro	6950.00

The overpaid hours of the top 5 most overpaid officers:

Employee	
BURCH, CHARLES K.	826.77
FITZPATRICK, DENIS J	641.50
RUSSELL, PATRICK	632.37
CHEN, WILLIAM L	601.77
DORCH, KENNETH C	549.25

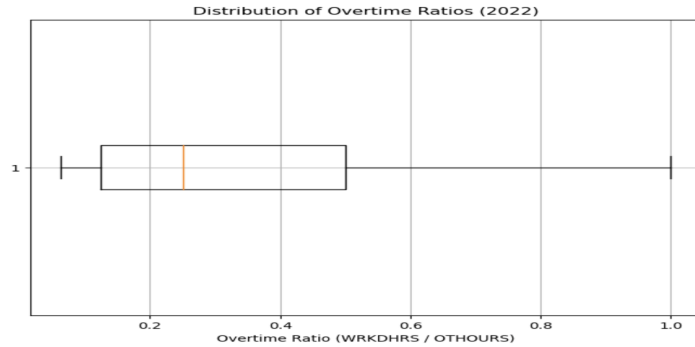


For the Court overtime dataset, we calculate the total overpaid hours due to discrepancies in the hours worked vs. hours paid of the Officers working overtime in the Court. Further, we answer a few more of the base questions of the project by visually comparing yearly overpaid hours as shown below and we also find the distribution of ratios of overtime worked vs. overtime paid to find potential outliers in the court OT database, which has been shown below for the year 2022.



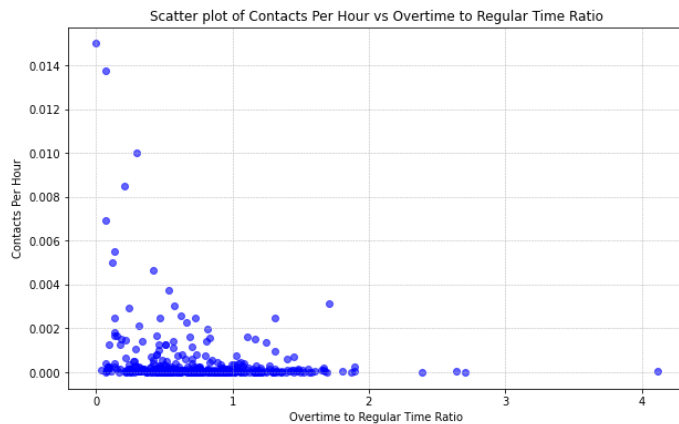
Mean Overtime Ratio: 0.38
Median Overtime Ratio: 0.25
Standard Deviation of Overtime Ratio: 0.29
Potential Outliers:

NAME	Average_Overtime_Ratio
14431 Agudelo-Echevarria, Marily	0.062500
8063 Becker, Matthew F.	0.093750
14740 Charles-Sampson, Azadi	0.093750
13188 Crabbe, David D.	0.093750
10355 Fantaroni, Evan M	0.062500
581 Foley, James M	0.093750
5635 Hyppolite, Jonathan A	0.062500
13894 Lawless, Timothy	0.093750
6697 Legacy, Mary Katherine	0.062500
14836 Lonergan, Ryan Charles	0.093750
6592 Mendes, Kevin	0.062500
14832 Milton, Christa A.	0.062500
8293 Murphy, Daniel Brian	0.062500
1745 Powell, Nadine	0.093750
4622 Ross, Allison D	0.062500
6696 Samuel-Lenehan, Angelique	0.062500
9943 Van Orman, Haley Margaret	0.062500
13950 West, Shawn L	0.062500
7061 Yanovitch, Scott	0.083333

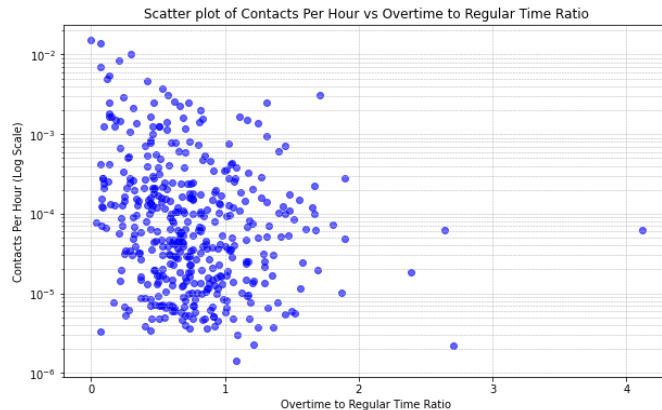


Overtime Hours and Field Contact:

In the interest of seeing the use of overtime hours, we decided to quickly see if a relationship exists between field contacts per hour, and the overtime/regular-time ratio. Since there was only a hunch there would be a correlation, we exclusively looked at the year 2022. Plotting the relationship we produced this plot:



Seen as the plot above did not expose too much detail due to extreme outliers, we produced another plot, this time with the contacts per hour graphed on a logarithmic scale:



As we can see, there is a clear negative correlation, potentially showing the decrease of officer productivity during overtime hours. Granted, the sample size is not large enough to draw such a concrete conclusion, but there is now more than enough reason to conduct more research into the potential relation.

Individual Contributions:

1. Aaron Zheng : Analyzed and performed analysis on the earning-data.csv dataset which includes general earnings data for city of Boston employees, and the relevant base questions.
2. Chen Yang: Performed an analysis on the BPD field activity data and addressed a few of the challenges.
3. Rithik Bhandary: Performed exploratory data analysis on the campaign contribution dataset and the overtime dataset which answer all of the base questions for the overtime dataset. Plotted trends and found valuable insights with figures.
4. Jakob Rundlett: Found correlation and produced graphs alluding to decrease in officer productivity during overtime hours.

Challenges and Limitations:

Certain base questions cannot be answered due to a lack of relevant data, including:

-Are certain officers (e.g., white, old, male, long tenure, high ranking title) more likely than others to have lower worked-to-paid ratios?

For this base question in the overtime dataset we have no mention about the age and ethnicity of the officers. Rank information is provided and we have visualized our findings based on rank.

-Have previously been disciplined for overtime abuse or other misconduct?

Yet again for this question we don't have a dataset that has information on the officers that have been disciplined for overtime abuse or misconduct.

Assumptions:

The dataset was not manually reviewed, so we assume that the data is accurate and correctly aligned.

Expected Next Steps:

Having Performed early analysis on the datasets provided to us, we aim to extend our work by finding more crucial insights and we will attempt to see if there's a bridge between the different datasets.

Project Completion Plan:

We propose to answer all of the remaining project base questions and to take inspiration from the other Police Overtime teams to help view our data from another perspective.

Extension Proposal:

In the overtime dataset, we have no mention about the age and ethnicity of the officers. If the stakeholders could provide us with a dataset that has this information, we could answer a few of the base questions and hope to draw out more insights. In the overtime dataset, we don't have a dataset that has information on the officers that have been disciplined for overtime abuse or misconduct. If the stakeholders could provide us with a dataset of these records we could attempt to extract more information. We already have visualized some important information from the campaign contributions dataset that are not part of the base questions. We could follow this up and find the big picture.