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:

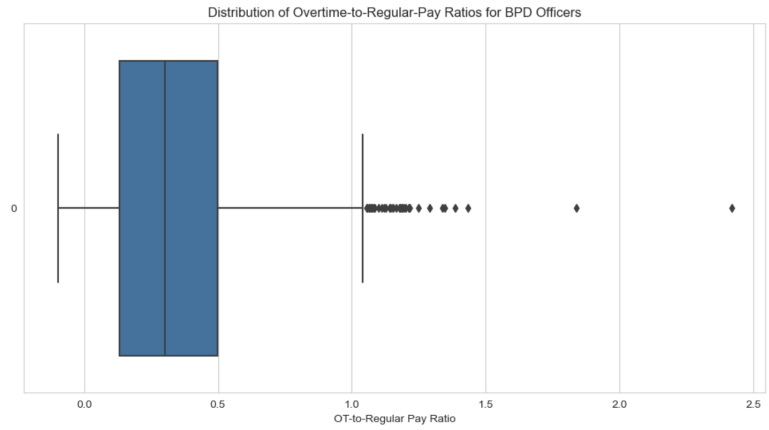
Performed analysis on the earning-data.csv dataset which includes general earnings data for city of Boston employees, and answered the following questions:

-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%

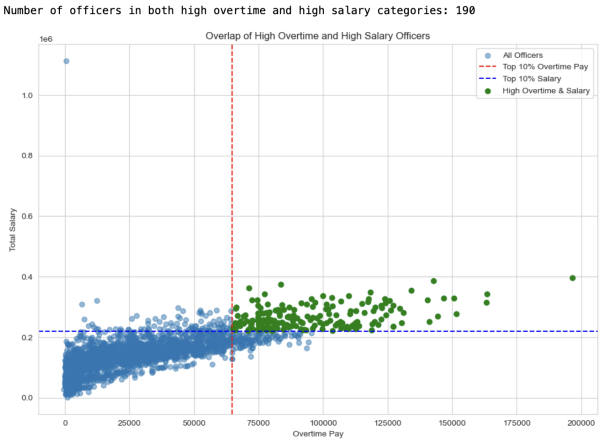
-Using data to fill in narratives around waste & misconduct by individual BPD officers
Top 10 officers who took the most overtime pay:

	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

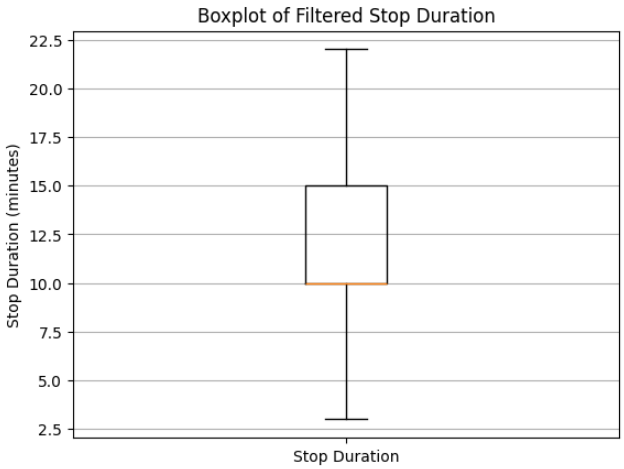
75% 15.00000

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:

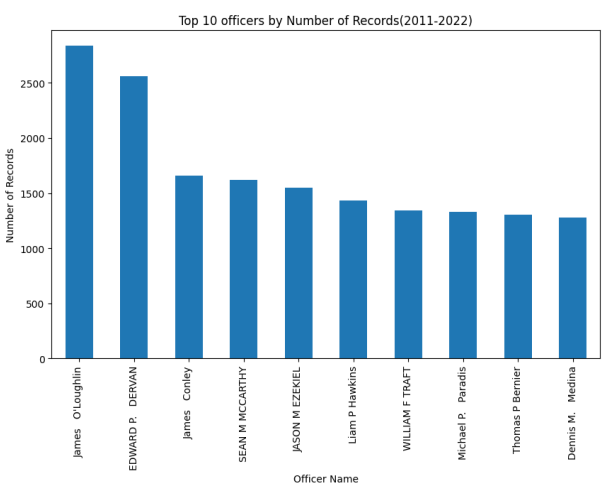
Explored the data to understand which officers and supervisors had the highest number of records.

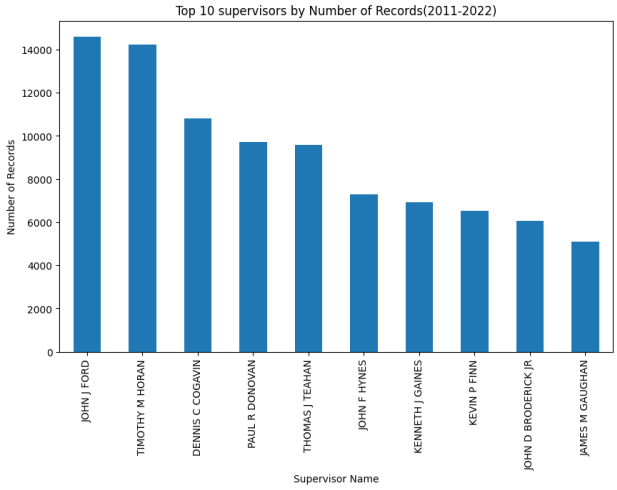
Initially utilized the unique IDs of officers and supervisors for this analysis.

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.





Overtime Data:

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 all of the base questions as mentioned in the project directives. 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. This displays instances of financial excess in BPD spending, characterizing wasteful BPD activities. It answers key questions from the base questions, "How do overtime hours paid compare to overtime hours worked? What does the discrepancy financially amount to, year after year?".

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

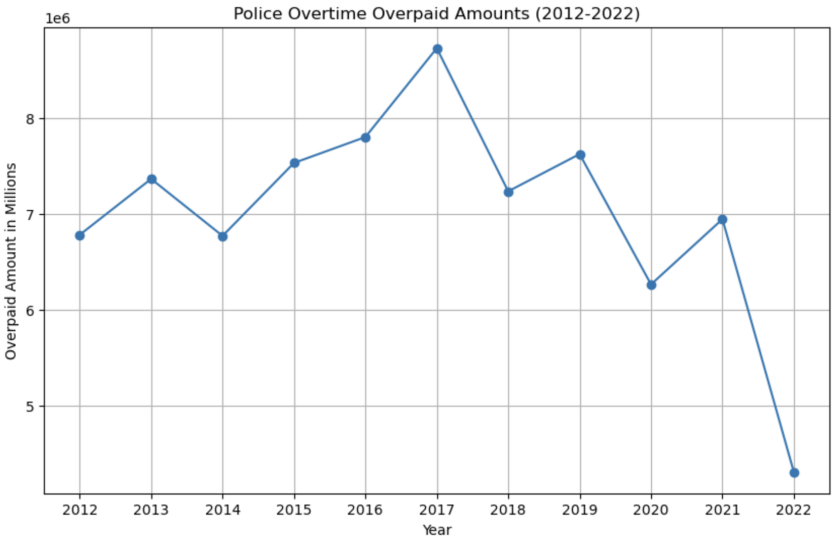


Fig: Plot showing the yearly trend of the amount the Police has overpaid in overtime

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, but most importantly "What is the distribution of ratios of overtime worked vs. overtime paid? Are there any outliers? (WRKDHRS vs. OTHOURS in the court OT database)". We do this 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. For finding the potential outliers we set the threshold to 0.1, which means a person actually works 1 hour for every 10 hours they get paid.

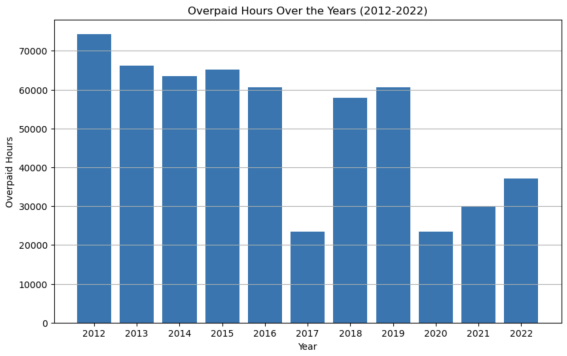


Fig: Bar chart showing a yearly trend of overpaid hours

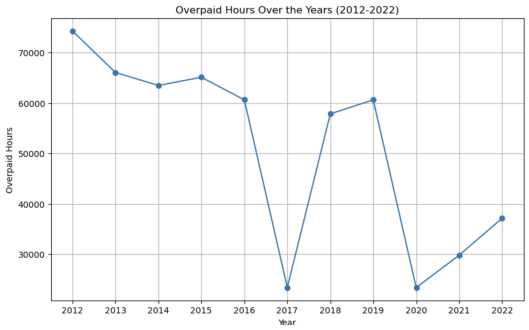


Fig: Line chart showing a yearly trend of overpaid hours

```
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
5035        Hyppolite,Jonathan A                0.062500
13884            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
```

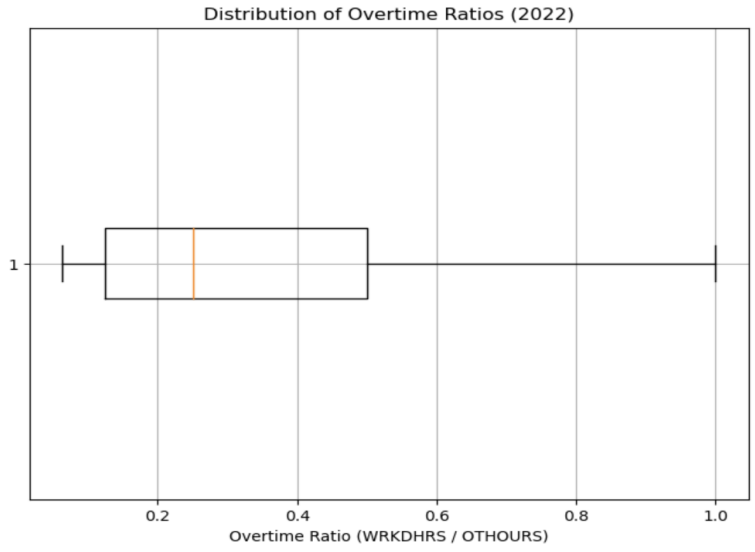
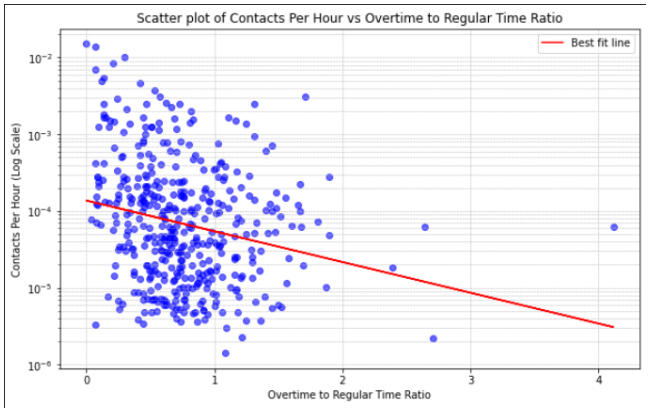


Fig: Distribution of overtime worked vs. overtime paid. Potential Outliers with names of the Officers

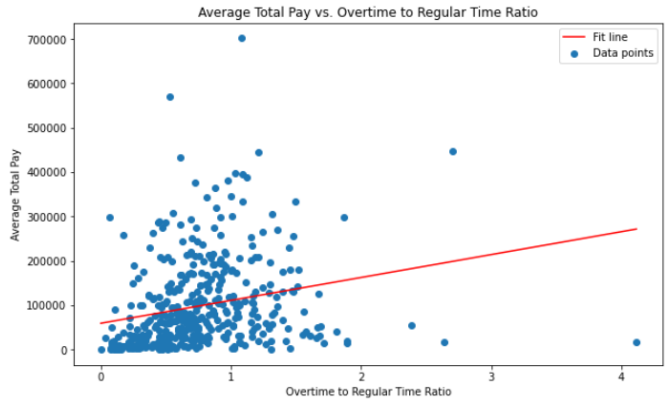
Overtime Hours and Field Contact:

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:



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.

To drive the economic non-viability of this system even farther, we also discovered a positive correlation relating the average total pay to the overtime/regular-time ratio. With these two graphs together, we can see that even though productivity is going down, the amount that the officers are getting paid still increases.



There several potential ways of solving this problem including creating quotas for work done during overtime hours, or simply paying officers a flat rate regardless of when the work is done.

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 + extension questions.
2. Chen Yang: Performed an analysis on the BPD field activity data and addressed a few of the challenges. This was also a major part of the extension project
3. Rithik Bhandary: Performed exploratory data analysis on the campaign contribution dataset and the overtime dataset. Further, extracted crucial insights from these datasets. Plotted trends and found valuable insights with figures. Plotted yearly trends that help us paint an overall picture. Answered many of the base questions visually.
4. Jakob Rundlett: Found correlation and produced graphs alluding to decrease in officer productivity during overtime hours + extension project on the same dataset.

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.

Extension Report:

Earnings Data:

Performed additional analysis on the earnings-data.csv dataset which includes remaining columns information outside of base questions
-Job title average regular pay, overtime pay and total gross pay
-Impact of the Quinn Education Incentive on salaries by job title. The Quinn Education Incentive is a government program that provides salary increases for police members who pursue and obtain a degree in a related field such as criminal justice.

These are specified in the tables below.

TITLE	REGULAR	OVERTIME	TOTAL_GROSS
Captain/Academy Instructor	160549.880000	96865.970000	315315.000000
Police Lieutenant/Hdq Dispatch	142843.866667	78730.383333	264348.566667
Police Sergeant/Mobile Oper	123595.810000	76441.426000	214360.321667
Police Lieut (Det)	146513.605000	72492.350000	274646.130000
Police Captain/DDC	155636.630714	70540.003571	275665.093125
Police Lieut/Acad Instr	140519.280000	68299.970000	260163.845000
Police Offc Breath $13.50	62929.815000	68073.040000	160543.115000
Police Lieutenant (Det)	124534.806667	68060.549565	281768.155000
Exec Asst (BPD)	134925.755000	61573.330000	174366.660000
Police Sergeant (Det)	118176.289474	60835.307117	225760.985470

TITLE	EMPLOYEE_COUNT
Captain/Academy Instructor	1
Police Lieutenant/Hdq Dispatch	3
Police Sergeant/Mobile Oper	6
Police Lieut (Det)	2
Police Captain/DDC	16
Police Lieut/Acad Instr	2
Police Offc Breath $13.50	2
Police Lieutenant (Det)	28
Exec Asst (BPD)	2
Police Sergeant (Det)	117

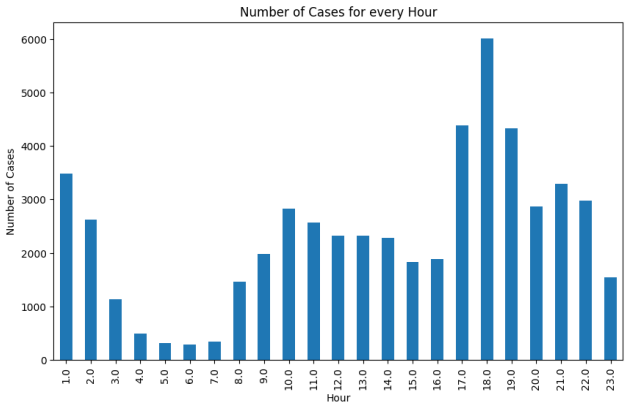
Fig: Regular Pay, Overtime Pay and Employee Count by job title

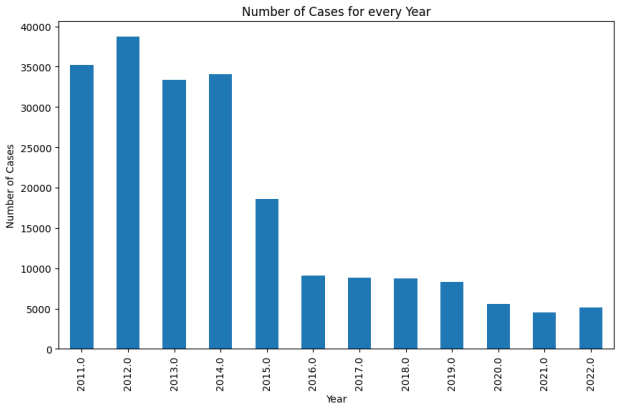
TITLE	Education_Impact	TOTAL_ GROSS	QUINN_EDUCATION	EMPLOYEE_COUNT
Lieut. Supv of Court Cases	19.576091	177923.010000	34830.370000	1
Police Offc Breath $13.50	18.113425	160543.115000	24770.100000	2
Police Captain/Hackney Invest	17.149412	234046.510000	40137.600000	1
Police Sergeant/Hdq Dispatcher	16.807861	182636.680000	30697.320000	1
Police Offc Ballistician 4$10	16.517460	101297.412500	21655.490000	4
Police Sergeant/Hackney Invest	16.368619	148381.360000	24287.980000	1
Police Lieutenant/Mobile Oper	15.375840	224829.150000	34569.370000	1
Dep Supn (Bpd)	15.337457	231725.855833	36206.223636	12
Police Lieut/Paid Detail Ser	15.098062	231038.520000	34882.340000	1
Police Sergeant/CHF RADIO DISP	14.831017	194075.001818	28158.626364	11

Fig: Impact of the Quinn Education Incentive by Job Title

BPD Field Activity Data:

In analyzing Boston Police field activity data, we observed that case frequency fluctuated across years and hours of the day.Annual analysis from 2011 to 2022 showed an initial high number of cases, a significant drop in 2015, and a continued decline thereafter. Hourly data for the same period shows a peak in cases at 17:00 hours, which may be related to police shift changes or higher crime rates at that time of day. These patterns are visualized in two bar charts: one showing the number of cases per year, and the other showing the distribution of cases at different times of the day. These charts illustrate the temporal distribution of activity in the field, providing a basis for assessing the effectiveness of the police presence and potentially informing funding allocation decisions.





Campaign Contribution Data:

In the directives that were provided to us, there was no mention of base questions for the campaign contribution dataset. Therefore, essentially whatever we have done is an extension. To maintain consistency, we created visualizations that are closely aligned to the directives given for the other datasets.

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. Additionally, we found the top 10 recipients based on how much money they have received, found the occupation of the top 10 contributors and top 20 occupations based on amount contributed.

```
Top 10 names who have received the most contributions:
 Matching Names
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
```

Fig: Names of the top 10 Active House Rep recipients

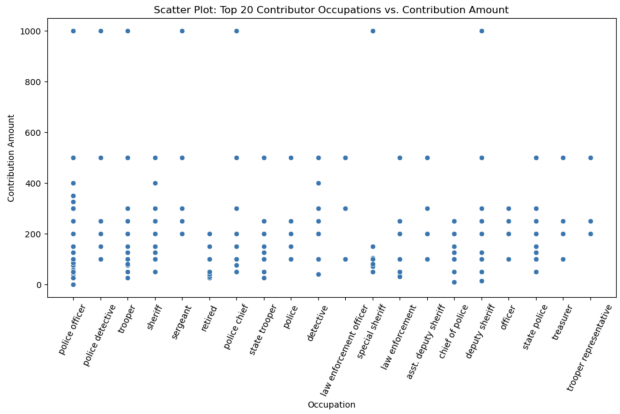


Fig: scatter plot for the top 20 contributions vs. contribution amount for Active Senator contributions

We also analyze the amount of contributions received each year by the House Reps and the Senators giving you a yearly trend, which can also provide insights into how much impact the incoming elections have on the amount of contributions.

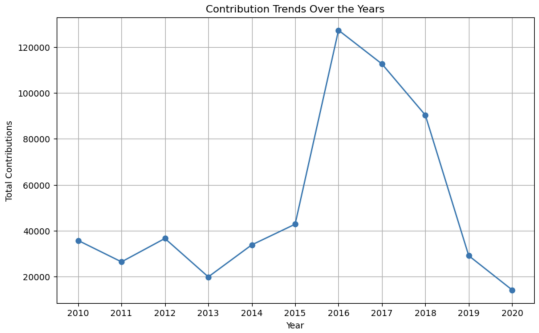


Fig: Yearly trend line chart for Active Senator Contributions

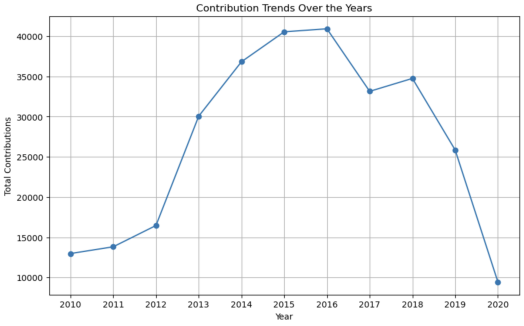


Fig: Yearly trend line chart for Active House Rep Contributions

Next, we dived into the active CC contributions dataset. We created visualizations based on the "Contributor" name to get a top 20 list of contributors and their contribution amount. On analyzing the amount received by each recipient, we 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, finding his top contributors and a year wise trend of how much he received. We also further restricted the findings to only people whose occupation is related to the police force.

	Contributor	Amount
143	Aggregated Unitemized Receipts	147763.85
12135	Murray, Georgia	25500.00
9572	Lee, Barbara	22350.00
1071	Benson, Deborah	21700.00
54	Aberly, Naomi	20550.00
5820	Flynn, Edward Michael	20000.00
1878	Burnes, Nonnie	12737.00
1462	Boston Teachers Union PAC	12500.00
7174	Hanley, Joseph	12050.00
1448	Boston Police Patrolmens Assoc. Pol Action Comm.	12000.00
9737	Leung, Edward	11250.00
10202	MA & Northern NE Laborers' District Council Po...	11250.00
10549	Maloney, Mark	11000.00
9753	Levenson, Norman	11000.00
530	Aulenback, Eric	9500.00
16510	Torres, Gregory	9500.00
2911	Choo, Arthur	9400.00
117	Adamson, Paul	9250.00
14460	Ross, Michael	9218.02
1347	Boland, Beth	9050.00

Fig: Top 20 contributors in the active CC contributions dataset

	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

Fig: Total amount received by each of the recipients sorted in descending order

Finally, we plotted a line chart for the yearly trend of contributions made to Flahert Jr., Michael F. by everyone and only police. There is also a contribution count which gives us the number of times a donation was made to Michael F. in each of the years sorted in descending order, this gives us crucial insights into his election campaign.

	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

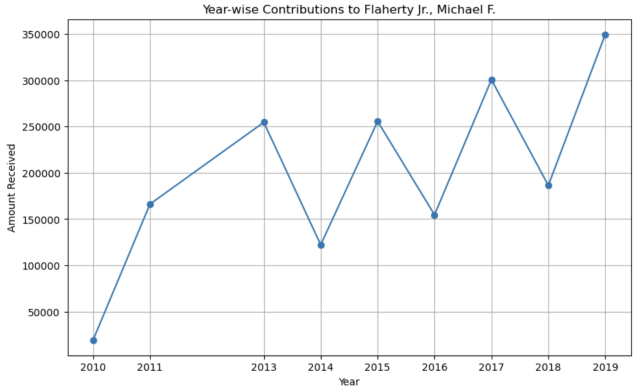


Fig: All contributions made towards Flaherty Jr., Michael F.

	Year	Contribution Count
0	2013	54
1	2017	48
2	2015	34
3	2014	23
4	2019	23
5	2016	21
6	2018	21
7	2011	10
8	2010	1

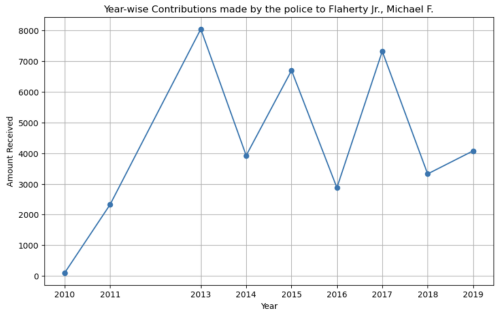


Fig: Contributions made by the Police towards Flaherty Jr., Michael F.

Overtime Data:

Although the overtime dataset served as an important dataset to help answer most of the base questions from the project directives, we also extracted further crucial insights that helped paint a better picture of the dataset. For the extension, in the 'year_details' data we found the the hours of the top 5 most overpaid officers, got insights regarding which areas call for the most amount of overtime services, also displaying it using a bar chart to show that some street names tend to call for an excessive amount of overtime. Probably, hiring more police officers in the police station near that area may help alleviate this problem.

```
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
```

Fig: Names of the top 5 most overpaid Officers

```
Top 5 street names with a count of the times an offcer has done overtime:
P.O. BOX 220801    15704
P.O. BOX 250       13870
SYLVAN ROAD         5462
ALLERTON ST.        5313
P.O. BOX 245        3662
Name: Street Name, dtype: int64
```

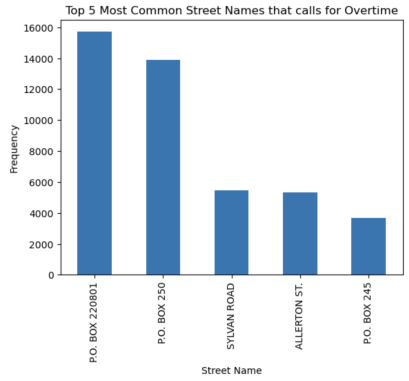


Fig: Insights on areas that call for surplus amount of Police overtime hours

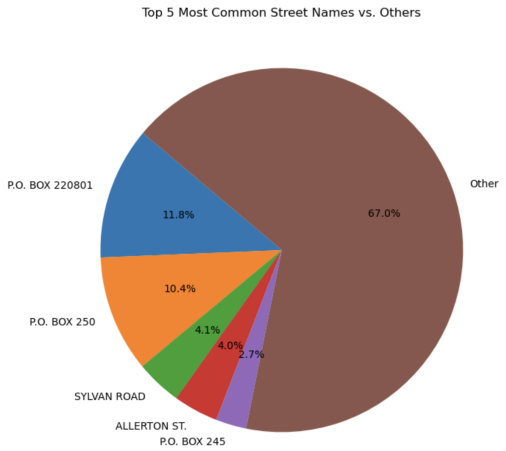


Fig: Pie chart on areas that call for surplus amount of Police overtime hours

Moving on to the 'court_overtime' data and its extension, we find the number of overpaid hours based on the ranks of the Officers, find the names of the Officers who are guilty of having a high number of overpaid hours along with their rank printed and finally, we also found the number of overpaid hours based on the 'Description' column of the data.

```
Rank-wise Overpaid Hours in Descending Order:
Rank Ptl: Total Overpaid Hours 23113.75
Rank Det: Total Overpaid Hours 9983.75
Rank Sergt: Total Overpaid Hours 2039.5
Rank SgtDet: Total Overpaid Hours 1744.25
Rank Lieut: Total Overpaid Hours 264.25
Rank Capt: Total Overpaid Hours 12.0
Rank LtDet: Total Overpaid Hours 8.0
Rank Civili: Total Overpaid Hours 0.0
```

Fig: Overpaid hours sorted by rank

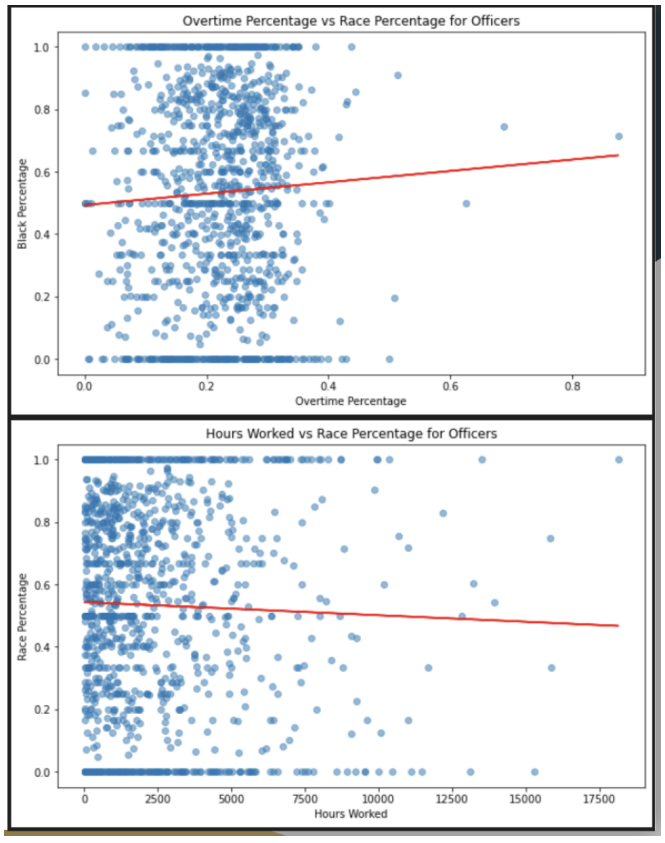
```
Top 20 Overpaid Employees:
Name: Lee,Waiman, Rank: Det, Overpaid Hours: 432.75
Name: Hernandez,Eddie, Rank: Det, Overpaid Hours: 283.75
Name: Flynn,Robert, Rank: Ptl, Overpaid Hours: 281.75
Name: Malcolm,David, Rank: Ptl, Overpaid Hours: 267.75
Name: Moy,Richard, Rank: Det, Overpaid Hours: 266.75
Name: Williams,David C., Rank: Det, Overpaid Hours: 245.5
Name: Edwards,Daran D, Rank: SgtDet, Overpaid Hours: 234.75
Name: Anderson,Lamont, Rank: Det, Overpaid Hours: 226.0
Name: Hegerich,Daniel J, Rank: Det, Overpaid Hours: 220.0
Name: Gallagher,William, Rank: SgtDet, Overpaid Hours: 208.75
Name: Wallace,Sean, Rank: Det, Overpaid Hours: 203.5
Name: Murray,Timothy D., Rank: Det, Overpaid Hours: 202.75
Name: Coakley,Kevin, Rank: Det, Overpaid Hours: 201.0
Name: Sanon,Serge A., Rank: Det, Overpaid Hours: 196.5
Name: Dineen,John J, Rank: SgtDet, Overpaid Hours: 194.75
Name: Kerrigan,Christopher, Rank: Ptl, Overpaid Hours: 189.5
Name: Juba,Dave E, Rank: Det, Overpaid Hours: 183.25
Name: McGovern,Omar, Rank: Ptl, Overpaid Hours: 182.5
Name: Monzon,Renzo A, Rank: Ptl, Overpaid Hours: 180.5
Name: Stoddard,James M, Rank: Det, Overpaid Hours: 179.0
```

Fig: Top 20 overpaid employees along with their rank and overpaid hours

```
Overpaid Hours by Description (Highest to lowest):
Description: COURT:TRIAL, Total Overpaid Hours: 15282.0
Description: COURT:CLERK MAG. HRG, Total Overpaid Hours: 8780.75
Description: COURT:MOTIONS HRG., Total Overpaid Hours: 5000.75
Description: COURT:CASE PREP., Total Overpaid Hours: 4858.25
Description: COURT:GRAND JURY TES, Total Overpaid Hours: 1550.75
Description: PROBATION HEARING, Total Overpaid Hours: 658.25
Description: ADMIN. HEARING, Total Overpaid Hours: 618.25
Description: COURT:PRETRIAL CONF., Total Overpaid Hours: 285.25
Description: COURT:JUD.TRAFFIC AP, Total Overpaid Hours: 77.0
Description: PAROLE HEARING, Total Overpaid Hours: 44.75
Description: SPECIAL CIRCUM.Frm26, Total Overpaid Hours: 9.5
Description: COURT, Total Overpaid Hours: 0.0
```

Fig: Overpaid hours sorted by description of the work done in court

Overtime Hours and Field Contact:



Used data from the BPD Field Interrogation and Observation (FIO) from the years 2015-2021 to get information on the percentage of black persons involved in police field contact

- Each point represents a police officer
- 1189 officers were excluded for not having contact with any individuals in the field (skewing the ratio)
- Surprisingly, total hours worked is negatively correlated, further discrediting overtime usage
- Increase in race ratio over overtime ratio is equal to 0.183 meaning an officer who has 5% more overtime compared to normal work time has contact with about 1% more black people.