

CS506 Project Meeting Notes

10/3/2023:

Meeting Topic: Project Description and Preliminary Questions

1. What is the project focus/overall goal?

To investigate how BPD budget, with a specific interest in how they register overtime with a focus on, but not limited to, the following topics

- Identifying instances of financial excess in BPD spending
- Characterizing wasteful BPD overtime practices
- Using data to fill in narratives around waste & misconduct by individual BPD officers

2. Why is this project important?

The city of Boston has a large spending budget of \$400 million thus it is important to analyze how this money is being spent. Audits are also important to moderate the power imbalance between the police and the community.

3. What type of data will you collect or be analyzing?

- Employee earning report
- Campaign contribution data
- BPD field activity data
- Overtime data

4. What are potential limitations of the project?

Recency of the data available, determining the root of causes for excess spending, Defining 'wasteful financial excess' with limited understanding of how police payroll operates

5. What are your next steps? Divide tasks amongst the team

Plan out data collection/cleaning and form the structure for a potential database, delegating directives to team members

6. Divide Tasks:

- (1) [Employee earnings data \(search police\)](#)
- (2) [Campaign contribution data](#)
- (3) [BPD field activity data](#)
- (4) [Overtime data from 2012-2022](#)

Dataset preprocessing and analysis

Name	Seunghwan	Ahmad	Ashton	Emily	David
Data Sets	1, 4	1,4	2,3	2,3	3, 4

Discussion Prompts:

Is it inherently political to fund the police more or less?

- Yes, whether people believe the police should be funded more or less inherently becomes a question about politics and society. The views on this topic vary based on political parties as well as the political climate of the world at the time. The following data was from a survey of American Adults in June of 2020:
 - A majority of Republicans and independents who lean to the Republican Party (61%) say spending on police should be increased, with 29% saying it should be increased a lot; 5% of Republicans say spending should be decreased, and 33% say it should stay about the same.
 - By contrast, 34% of Democrats and Democratic leaners say police funding should be increased, 25% say it should be decreased and 40% would like to see it stay about the same.
- It seems that the call for defunding the police was more prominent in 2020 than 2021. This could be because of the Black Lives Matter Movement which was very prominent in 2020. This movement called to defund the police in light of police brutality towards black americans.
 - The share of Black adults who say police spending in their area should be decreased has fallen 19 percentage points since last year (from 42% to 23%)
 - Democrats who say funding for local police should be decreased has fallen markedly – from 41% in 2020 to 25% today
 - Source for above data:
 - <https://www.pewresearch.org/short-reads/2021/10/26/growing-share-of-xx-americans-say-they-want-more-spending-on-police-in-their-area/#:~:text=>

[A%20majority%20of%20Republicans%20and.should%20stay%20about%20the%20same.](#)

What place does, or should, police union lobbying have in the discussion around police funding?

- The police unions should definitely have a voice in the conversations surrounding how the police department budget is delegated, but perhaps not as much sway in the final decisions - not as much as “lobbying” would imply there is.

10/10/2023:

Meeting Topic: Beginning to process data and learn Pandas functions

Findings from Datasets [1](#) and [4](#):

- Employee IDs are not given for some people, might have to sort by names
 - Will have to deal with duplicates (potentially two or more people having the same name) if this is the case
- Hours worked and hours paid do not align for some of the data
 - Will have to look into this discrepancy
- 1,507 titles but about half a dozen ranks
 - Need to figure out which ranks are higher
 - Can compare to salaries to see which ranks have higher salaries

Helpful Functions:

- Reading Data:
 - `pd.read_excel('file path')`
 - If data comes from excel sheet
 - `pd.read_csv('file path')`
 - If data from csv
- `df.columns`
 - Gets the columns of the dataset
- `df.isnull().sum()`
 - How much missing data there is
- `plt.hist(df['Column'])`
 - Example:
 - `overtime21['OVERTIME_FREQ'] = overtime21.groupby(['Emp. ID'])['Emp. ID'].count()`
 - `plt.hist(overtime21['OVERTIME_FREQ'])`

10/17/2023

Meeting Topic: Discussing Data Findings So Far

Campaign Contribution Data:

- OCPF Data (Office of Campaign and Political Finance)
 - Filtered by Employer
 - BPD and Boston Police Department
 - 478 rows
 - City Of Boston with Occupation as Police Officer
 - Resulted in 14 entries
 - Not sure if we include
 - Amount donated has negative values for some rows
 - Removed for data analysis
 - Total spent by BPD employees was \$113,444.56
 - The most someone donated was \$5,800
 - The least was \$10
 - 101 of the Donations went to **Martin Walsh**
 - He received \$28,917
- Active CC Contributions:
 - Filtered by Employer BPD and Boston Police Department
 - 113 Rows
 - Also might need to consider employer as city of boston with occupation as police officer
 - Amount donated has negative values
 - Removed for data analysis
 - Total amount: \$24,721.00
 - Annissa George was the most common recipient
 - 38 people donating to "George, Annissa"
 - Total: \$10,000
 - 28 people donating to "Essaibi George, Annissa"
 - Total: \$7,450
 - One person donated 11 times
 - Did not spend the most money though
 - Highest Contributor was Richard Driscoll
 - Totaled \$2,950
 - All rows in here are present in the OCPF data
 - The merge resulted in 113 rows which matches active CC contributions
- Active CC Summary
 - Has a lot of variations of BPD in the employers, need to find a way to get them all
 - This analysis is based off of a subset of them
 - No Negative values
 - Total Amount was \$26,421

- Annissa George was top recipient
 - Probably has overlap with active CC Contributions
- Active Police CC Contributions
 - Same issue as above with a lot of employers
 - Probably comes from the same data
 - When filtered by same employers there's the same amount of rows (119)
 - Same amount spent as above
- All Police Contributions
 - Need to filter out non Boston Police officers same issue as above
 - Total spent: \$113054.56
 - Walsh received the most money

Field Data:

- **FieldContact Table:**
 - Incident report from officer side
 - Important data: string description of report, officer id. Most other data will probably be irrelevant
 - Can link officer id with other data fields, and can view incident reports when needed to assess financial claims
- **FieldContact_Name Table:**
 - Descriptions of those taken into custody
 - "Was frisked" column, either boolean 0 or 1, sometimes set to NULL
 - Few entries where almost all columns are 0 or NULL
 - "Ethnicity", is either one of four: ["Hispanic Origin", "Not of Hispanic Origin", "Unknown", NULL]
 - "Build" is one of five: ["Thin", "Medium", "Heavy", "Muscular", NULL]

Overtime Data:

- Discrepancy between WRKDHRS & OTHOURS
 - ex) WRKDHRS=1 & OTHOURS=4
- Lunch & Dinner hours are included in WRKDHRS for some and not for others

Earnings Data(2021):

- 16392 out of 23204 rows have 'None' in 'OVERTIME'
 - Tracking NaN is tricky so I used following code to change NaN to 0

```
nan_dict = {np.nan: 0}
temp_earnings['OVERTIME'] =
temp_earnings['OVERTIME'].map(nan_dict)
```

To do:

- Figure out how to group officers by title / rank, possible cluster methods, because of the sheer quantity of officer titles make it hard to extrapolate data from
- Where to find (if possible) descriptions of data sets
- Keep doing as usual for next week
- Deliverable 1 Due Nov 1st (Wed)

Potential Extension Questions:

- Political Parties
 - Look at Campaign data and see who they donated to and which party they are affiliated with
- <https://www.wbur.org/news/2020/12/21/boston-police-medical-leave-overtime-costs>
 -

10/24/2023

Meeting Topic: Discussing Data Findings Continued

What Have We Worked On?

- Campaign Contribution Data
 - Finished Processing all of the datasets (with the exception of state campaigns)
 - Refactored how employer and occupation is filtered to include police officers who are employed by the city of boston
 - Discussed with the TA that it would be best to include this information
 - Some overlap with some of the datasets
 - For example the filtered CC Contributions and the Police CC Contributions as theoretically once filtered the CC Contributions should have the same data as the Police CC Contributions.
- Field Data
 - Finished combining/preprocessing data
 - Analysis of FIO reports by supervisor and district
- Overtime Pay
 - Analysis of Difference in Paid Hours and Worked Hours
 - Distribution by number of record & number of record per person
- Earnings
 - Analysis of Overtime pay
 - Distribution by number of record & number of record per person

What Are We Working on Next?

- Deliverable Due Next Week
 - *Collect and Pre Process Data*
 - Done
 - Preliminary Analysis of Data
 - Look into if there is overlap between individuals who have higher overtime pay and higher campaign contributions
 - Look into if there is overlap between individuals in field data who have over 14 standard deviations from the average amount of field data with individuals with higher overtime requests
 - Maybe those who have more field reports were working overtime
 - Merge datasets by names
 - Figure out relationship between rank and occupation and overtime
 - Answer 2 Key Questions
 - **Identifying instances of financial excess in BPD spending**
 - How has the BPD budget changed year-over-year?

- Where have funds grown or shrunk, overall and intra-departmentally?
 - How has funding shifted between departments?
- How have BPD paychecks changed year-over-year? Both the average amount, as compared with non-BPD Boston city employees, and the breakdown (regular pay v. overtime pay, etc.)?
- How much BPD officer pay came from injury pay? What percentage of officers took injury pay in a given year?
- **Characterizing wasteful BPD overtime practices**
 - Identifying instances of financial excess in BPD spending,
 - Characterizing wasteful BPD overtime practices
 - Using data to fill in narratives around waste & misconduct by individual BPD officers.
 - How do overtime hours paid compare to overtime hours worked? What does the discrepancy financially amount to, year after year?
 - How has overtime for court appearances changed year-over-year?
 - What is the distribution of ratios of overtime worked vs. overtime paid? Are there any outliers? (WRKDHS vs. OTHOURS in the court OT database).
 - Are certain officers (e.g., white, old, male, long tenure, high ranking title) more likely than others to have lower worked-to-paid ratios?
- To Do:
 - Share key IDs and the descriptions
 - ex) Top 5% of number of Overtime requests
 - ex) Top 1% of yearly Overtime pay
 - ex) Timothy M Horan ID 11610
 - 14 standard deviations away from the mean for number of fio reports supervised
 - Question 1
 - Emily
 - Ahmad Sadiq
 - David
 - Question 2
 - Jason
 - Ashton

Any Issues?

- None

Have We Talked With The Client Recently?

- Not Meeting with the client

10/31/2023

Meeting Topic: Discussing Data Findings and Key Questions

What Have We Worked On?

- Seunghwan Hyun
 - Merging overtime data, overtime earnings data, overtime court data from 2012 to 2022
 - Creating plot for merged data and yearly data to visualize the trend of overtime court appearance, discrepancy between paid hours to worked hours
 - Used pandas groupby to calculate overtime data by employee ID when the original data was in records basis
 - Transformed data to create new columns for data visualization
 - Calculated Hours Worked by merging data in int and HH:MM format
 - Calculated cut-off line for top 1%, 5%, 10% points which could be used as a cut-off for outliers later on.
- Emily Opresnick:
 - Looked at earnings from 2012-2022
 - Broke down the pay into 7 categories: Quinn, Other, Regular, Injury, Overtime, Retro, Detail
 - For each year, plotted the percent each of the categories made up from the total pay
 - Plotted the change in each category over the decade
 - Looked at the percentage of BPD employees who claimed each category and plotted this by decade to see trends
 - Looked at average earnings by department
 - Found the 10 departments with the highest average earnings each year and plotted them
 - Added Operating Budget dataset
 - Figure out how funding has changed over past few years
- Ashton Fox:
 - Created analyzed data table based on Sex/Race in the police department categorized by rank & across-the-board
 - Data records are limited so will have to scale past years based on 2022-2023
 - Something to note: Because rank won't carry over on a 1-to-1 scale with the rest of the data, will have to take the naive approach to assessing possible instances of worked-to-paid ratio discrepancies
 - eg. % of white male officers who receive more hours worked than actual work

- David White
 - Processed earnings data 2012-2022
 - Answering: **Identifying instances of financial excess in BPD spending**
 - Comparing Number of officers claiming injury pay year by year
 - Average total earnings
 - Funding breakdown by department
- Ahmad Sadiq:
 - Answered the following questions for the data of 2021 and 2022 for earnings report and the overtime 2021 2022:
 - **Identifying instances of financial excess in BPD spending**
 - How has the BPD budget changed year-over-year?
 - Where have funds grown or shrunk, overall and intra-departmentally?
 - How has funding shifted between departments?
 - How have BPD paychecks changed year-over-year? Both the average amount, as compared with non-BPD Boston city employees, and the breakdown (regular pay v. overtime pay, etc.)?
 - How much BPD officer pay came from injury pay? What percentage of officers took injury pay in a given year? ‘
 - Analyzed the data for the following questions and represented the data in graphs to visualize the results obtained to answer the questions.
 - Identified the different departments within BPD to see where the funding was going towards and where it shrank in the different types of department.
 - Showcased all the results with the group at the group meeting and discussed our individual findings.

What Are We Working on Next?

- Working on slides for Deliverable
 - Presenting on Thursday 11/2
- Checkpoint A
 - Collecting and Processing More Data
 - Operating budget data set
 - Shows how the budget has changed from 2021 - 2023 and what they have set for 2024
 - Meeting next week to discuss further findings

Any Issues?

- Some years in earnings data, the sum of all of the earnings categories does not align with the total sum of earnings
 - This might be because the way Pandas handles floats, with it being in scientific notation and then when fixing this it removes decimal points

Have We Talked With The Client Recently?

- Not meeting with client

