

DELIVERABLE 1 REPORT

Outline

- 1. Introduction to Problem Statement**
- 2. Data Collection and Cleaning Steps**
- 3. Exploratory Data Analysis (EDA)**
- 4. Possible Answers To Clients' Questions And Hypotheses**
- 5. Individual contributions of each team member**

1. Introduction to Problem Statement

Policing across the entire country of the USA has been a controversial issue, concerns about potential abuse of power and bias has been rising. In recent years, there has been an increased focus on the money spent on items such as military weapons and vehicles, and some of these expenditures, which are considered "wasteful," are being masked as overtime expenses. The Boston Police Department (BPD) has an operating budget over \$400 million. Therefore, it is important to analyze and understand how this money is used, particularly in overtime expenses.

2. Data Collection and Cleaning Steps

a. Data Collection:

- Data Source: Our main source of data is the datasets provided by the client.

(1) [Employee earnings data \(search police\)](#)

(2) [Campaign contribution data](#)

(3) [BPD field activity data](#)

(4) [Overtime data from 2012-2022](#)

- Data volume and size:
 - Data sets preprocessed:
 1. BDP Earning Dataset
 2. Non-BDP Earning Dataset
 3. Overtime Dataset
 4. Court Dataset
 5. Campaign Contribution Dataset
 6. Field Activity Dataset
 - Most of our datasets have a time frame from 2012-2020.

b. Data Cleaning:

- Data Type Conversion:
 - The provided data sets exhibit inconsistencies in data types within the same fields.

For instance, the “overtime pay” field in the employee earning dataset contains non-numeric characters such as ‘\$’, ‘,’, etc.
 - We created a function to convert fields facing this issue into a consistent numerical format
- Handling Missing Values:
 - For each datasets, we checked for missing values and used mean imputation for highly related data, and dropped the row if the missing values are scarce and randomly distributed
 - The time frames across datasets are not consistent (i.e: BDP Court overtime dataset only has data until October 2022 , not full year, BDP Field Contact dataset only has data until 2019, etc)

- **Standardization and Normalization:**
 - The field names (columns) within a data category are inconsistent over the course of the analysis time frame.
 - We normalized all column names for the sake of consistency and clarity
- **Dataset combination:**
 - Datasets are provided yearly, with each year in a single csv file. Thus, utilizing the dataset would be messy if we hard code it by year name. Thus, we created a list for each data category (Eg. Overtime, Injury etc.), and store all datasets in sequential order regarding the years. By doing so, we can call the dataset of one category by a single variable, with the index reflecting the year.

3. Exploratory Data Analysis (EDA)

3.1 Data Understanding

- **Employee Earnings Data**
 - The dataset provides the information of employees in the City of Boston from years 2011 - 2022.
 - Contains employee names, the department they work in, as well as their regular earnings, retro earnings, overtime earnings, injured earnings, paid detail earnings, education earnings, as well as other earnings such as bonuses, etc.
- **Campaign Contribution Data**
 - The dataset provides the information of Campaign and Political Finance (OCPF) in years of 2010 - 2020.

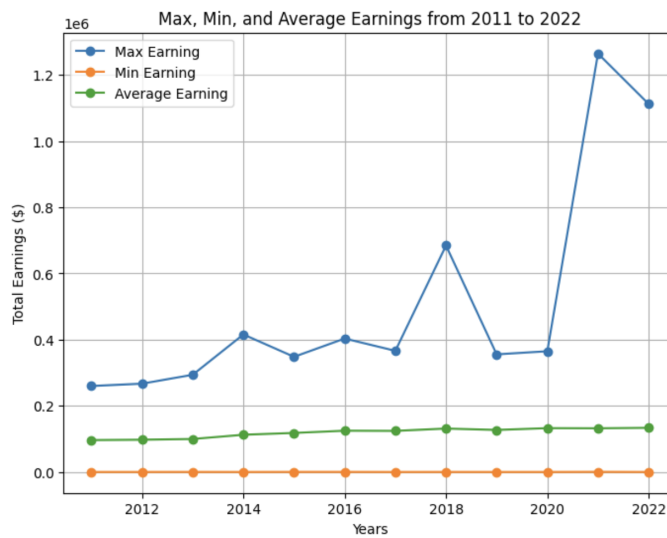
- Contains all the names of people who made the contribution, their occupation, employer, the amount they have contributed and the receiver of the contribution as well as other personal information such as the living address, zip code, CPF ID, etc.
- BPD Field Activity Data
 - The dataset provides the information of interaction of BPD with private individuals from 2011 - 2022 to keep their work transparent.
 - Contains the names of individuals who were in contact with the police officer, names of the police officers involved in the interactions as well as their supervisors name, the date and time of the interaction, place of the interaction, and the reason for the interaction, with other details such as the vehicle of the officer and the individual, id of the officer and the individual, etc.
- Overtime Data
 - The dataset provide the information of overtime working hours and payments of BPD employees in the years 2012 - 2022
 - Contains the personal information of the employees such as their names, working id, and rank. As well as the location of where they worked overtime, who their client was, the number of hours they worked overtime and how much they were paid per hour.

3.2 Plotting:

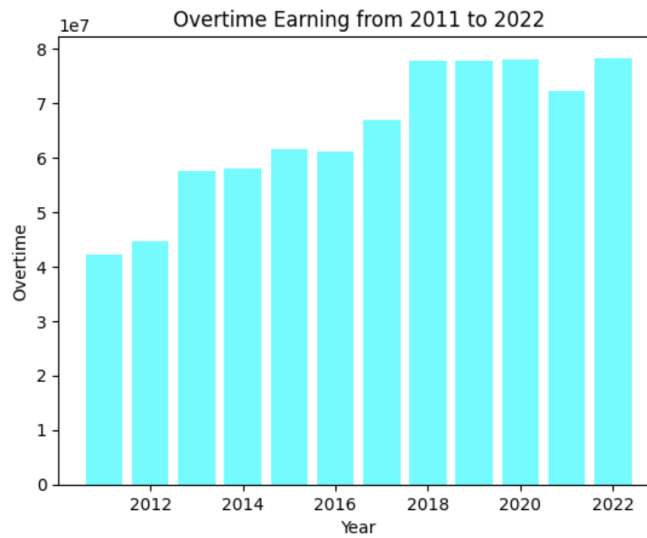
a. Total Earning:

Dataset: [BDP earning from 2011 to 2022](#)

- Min, Max and Average Total Earning of BDP officers from 2011 to 2022:



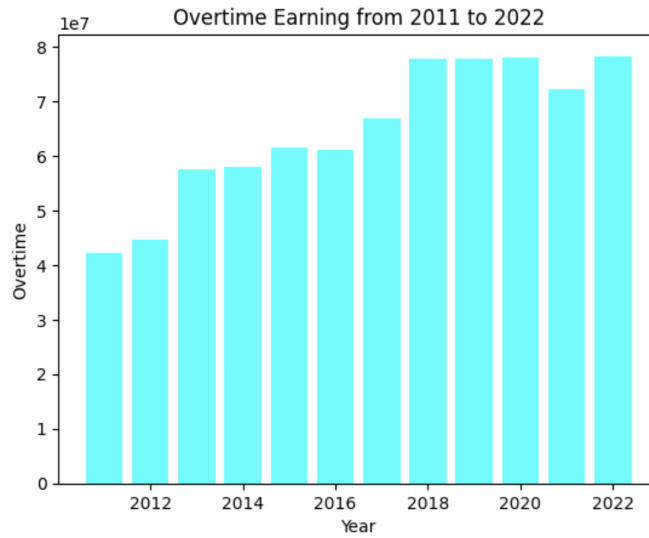
- Actual Total Earning of BDP officers from 2011 to 2022:



b. Overtime Earning:

Dataset: [BDP earning from 2011 to 2022](#)

- Actual Overtime Earning of BPD officers from 2011 to 2022:

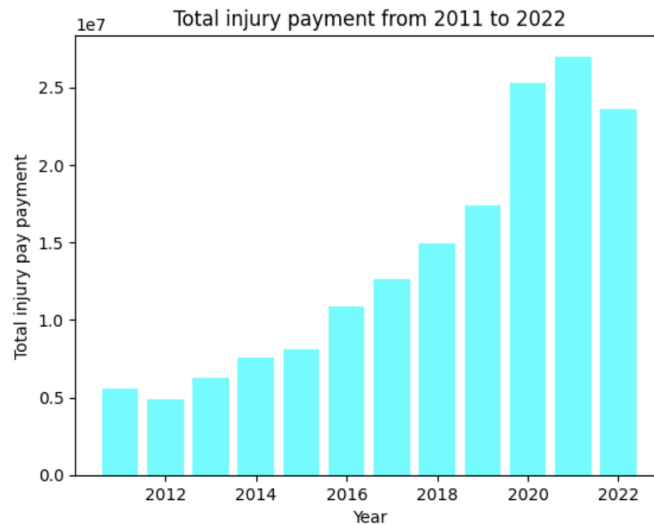


- Percentage Overtime Earning of BPD officers from 2011 to 2022:



c. Injury Pay

- Total Injury Payment from 2011 to 2022



4. Possible Answers To Clients' Questions And Hypotheses

Base Question 1: **How have BPD and paychecks changed year-over-year on average compared to non-BPD Boston city employees?**

- Answer

The data shows that salaries for Boston Police Department (BPD) employees have been going up over time. This suggests that both the normal pay and the extra pay they receive for working overtime have increased. The data indicates that the average salary for non-BPD city workers is approximately half that of BPD employees.

- Hypothesis:

We hypothesize that BDP paychecks have a linear relationship with officers 'Total Earning'.

Thus, to analyze BDP paychecks changed year-over-year, we analyze the trends in BDP 'Total Earning' from 2011 to 2022.

- Potential Flaws in Hypothesis:

- The hypothesis assumes a linear relationship between BDP paychecks and officers' total earnings, which needs further validation. The reason why we made this hypothesis is because we have no direct access to BDP paychecks (the actual checks that an officer received after deductions for taxes, benefits and other withholdings).

- Analysis Based on Hypothesis:

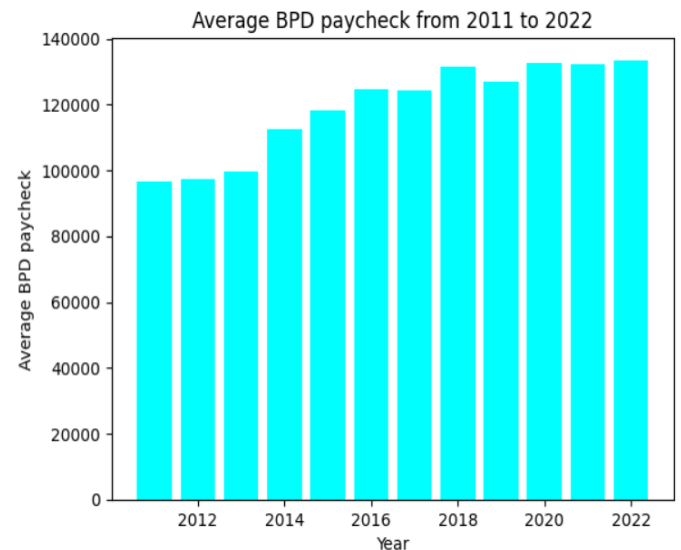
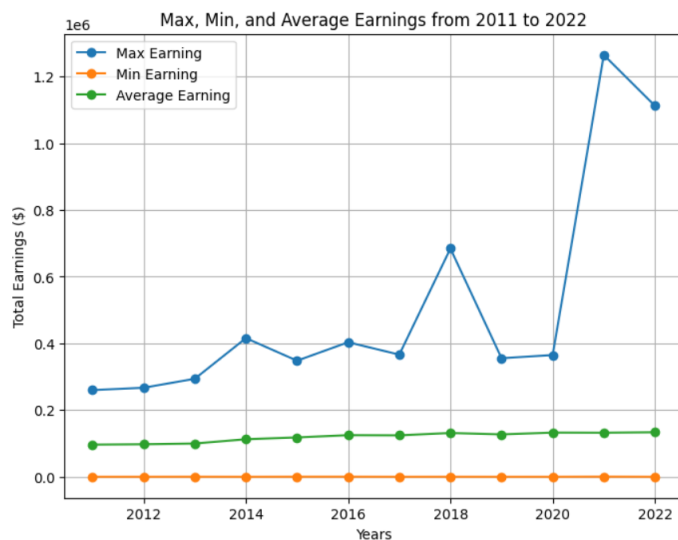
Analyzed Data:

- BDP

	Max Earning	Min Earning	Average Earning
2011	259914.04	11.70	96421.474132
2012	266971.82	58.52	97515.361269
2013	293892.24	187.69	99771.862159
2014	415709.53	9.36	112589.650642
2015	348096.80	223.02	118041.488626
2016	403408.61	238.85	124787.164775
2017	366232.65	3.50	124254.563280
2018	684410.90	105.90	131321.462320
2019	355538.70	2.50	127094.346316
2020	365001.16	25.00	132487.610436
2021	1264843.63	400.00	132114.566694
2022	1112348.25	23.68	133494.427569

- We can see that there is a great difference between Max-Earning and Min-Earning, with some cases where the minimum earning being only a few dollars. This can be indicative of potential outliers or atypical situations in our dataset.

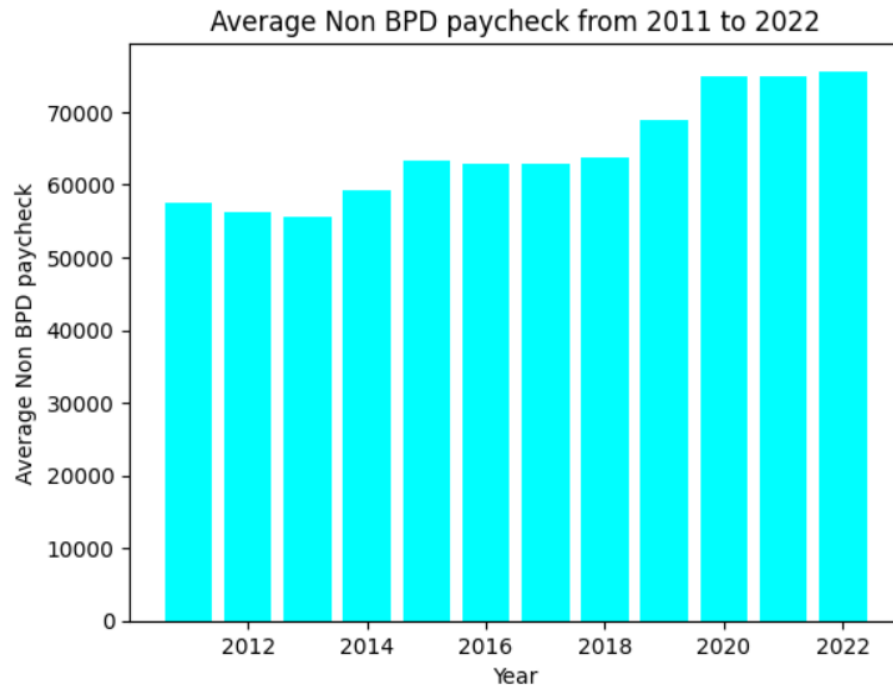
Line graph showing Min, Max and Average Total Earning changes from 2011 to 2022:



Analysis:

- The plot shows that, on average, BDP paychecks have increased over the years.
- However, the max total earnings per officer experienced a dramatic increase from 2020 to 2022. Specifically, the maximum earnings increased from \$259,914 in 2011 to \$1,112,348 in 2022.
- Notice an abnormal increase in police earnings between 2020-2022, we looked into it and found out that the officer was actually awarded \$2 million in a [gender discrimination lawsuit by the Federal Jury](#).

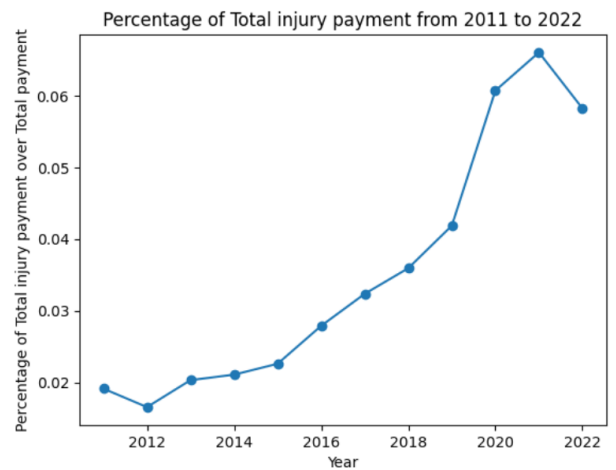
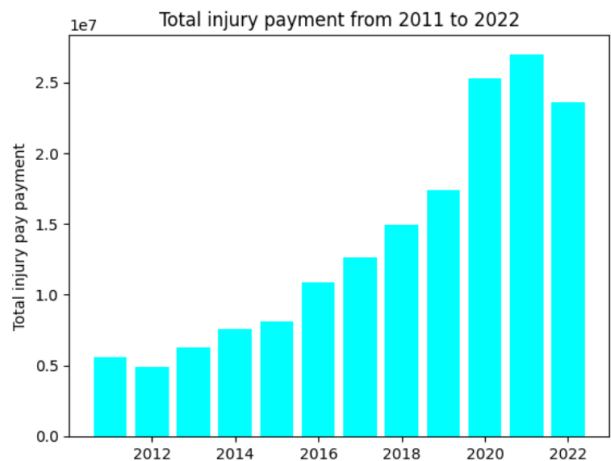
- Non-BDP Data:



- The plot show that in general, the average non-BDP paycheck increased from 2011 to 2022
- However, the average salary for non-BPD city workers was only approximately half that of BPD employees.

Base Question 2: How much BPD officer pay came from injury pay? What percentage of officers took injury pay in a given year?

- The injury payment of BPD officers ranges from 5% to 25% of their overall payment and it has an increase overtime.
- About 2% to 6% of BPD officers got injured. This also has an increase overtime. Which corresponds to the injury payment increase.



5. Individual contributions of each team member

[Nurassyl Medeu]

- Assisted with data processing and cleaning.
- Performed analysis on BPD and non-BPD paychecks 2011-2022:
 - Determined the early change of average paychecks of BPD employees and non-BPD employees
- Performed Statistical Analysis of Campaign Contribution data 2011-2022
 - Calculated and plot the average amount contributed by BPD employees and non-BPD employees

[Jiawei Sun]

- Helped with data processing and cleaning.
- Conducted analysis on Total Earning Data from 2011-2022:
 - Analyzed trends of total earning and overtime earning over years
 - Calculated the percentage breakdown of individual earnings.
 - Examined the percentage increment of individual earnings over time.

[Truc Duong]

- Helped with data preprocessing and cleaning
- Performed Statistical Analysis on BDP Earning Data from 2011-2022:
 - Determined average total earning of a police officer

- Analyzed changes in average total earnings from 2011 to 2022.
- Examined detailed statistics for police earnings in 2011 and 2022.
- Analyzed differences in overtime type in the overtime dataset
- Analyzed Overtime and Regular Earning changes over years

[Can Wang]

- Analyze data for injury payment and injury payment ratio 2011-2022.
- Analyze data for injury ratio of BPD officers per year 2011-2022
- Did research related to injury ratio and injury payment.
- Assesses if BPD officers have had better subsidies since 2011.
- Comparing the injury payment of BPD and non BPD 2011-2022.

[Al Mbaye]

- Performed Statistical Analysis on Total Spending by Category from 2011-2022
 - Determined the amount of total spending of each category
 - Determined the average amount of spending of each category
- Performed Statistical Analysis on Total Earning by Category from 2011-2022
 - Determined the amount of total earnings of each category
 - Determined the average amount of earnings of each category
 - Determined a ranking of total earnings
 - Determined a ranking of total earnings adjusted for inflation
- Helped with data preprocessing for court overtime