# CS506 Team E

* Police Overtime Team Weekly Scrum Report 3

1. **What we accomplished this week:**

* Find demographics data of BPD officers
* Gather field activity data
* Gather and preprocess BPD campaign contribution data
* How has overtime for court appearances changed year-over-year? (Truc)
* What is the distribution of ratios of overtime worked vs overtime paid? Are there any outliers? (WRKDHRS vs OTHOURS) (Al)
* Use gathered data to start answering 1-2 main questions

1. **Deliverable links:**

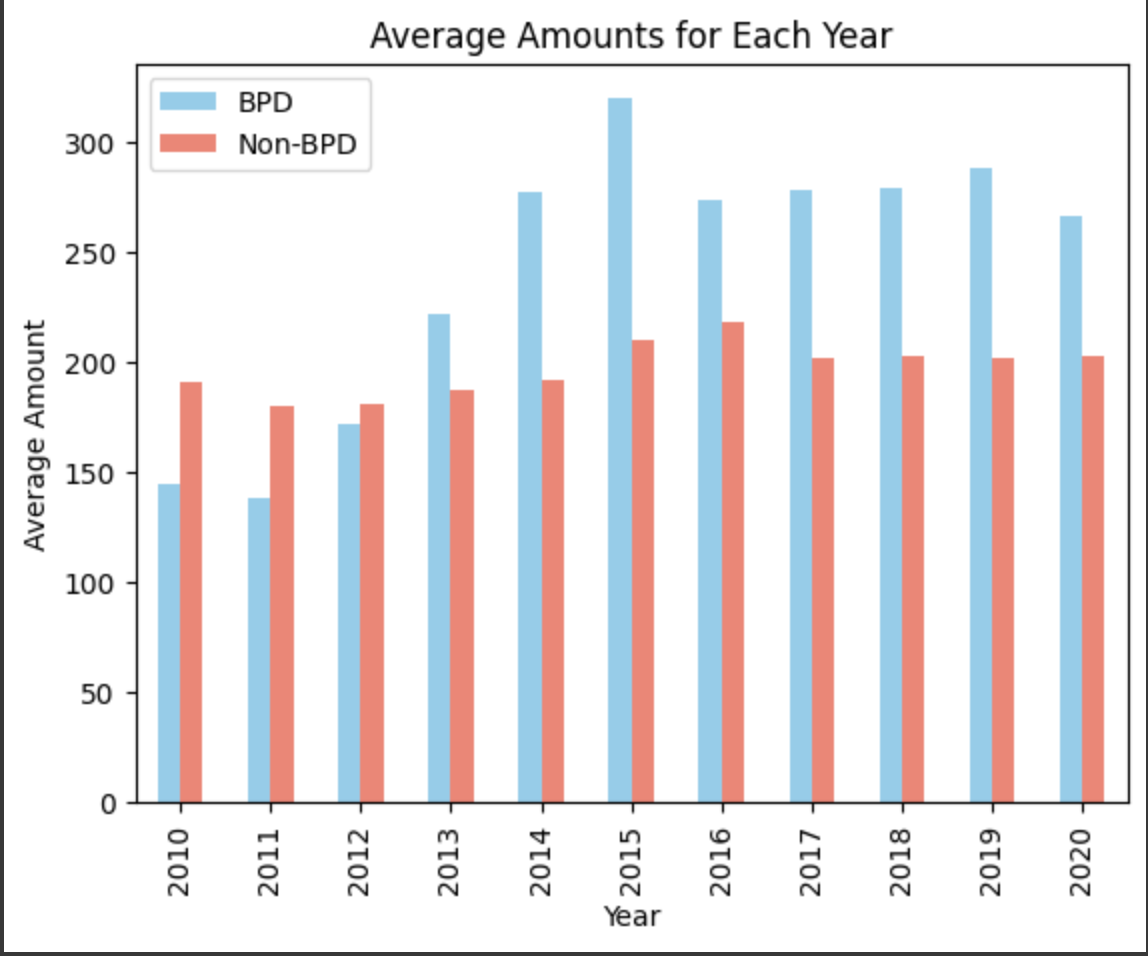
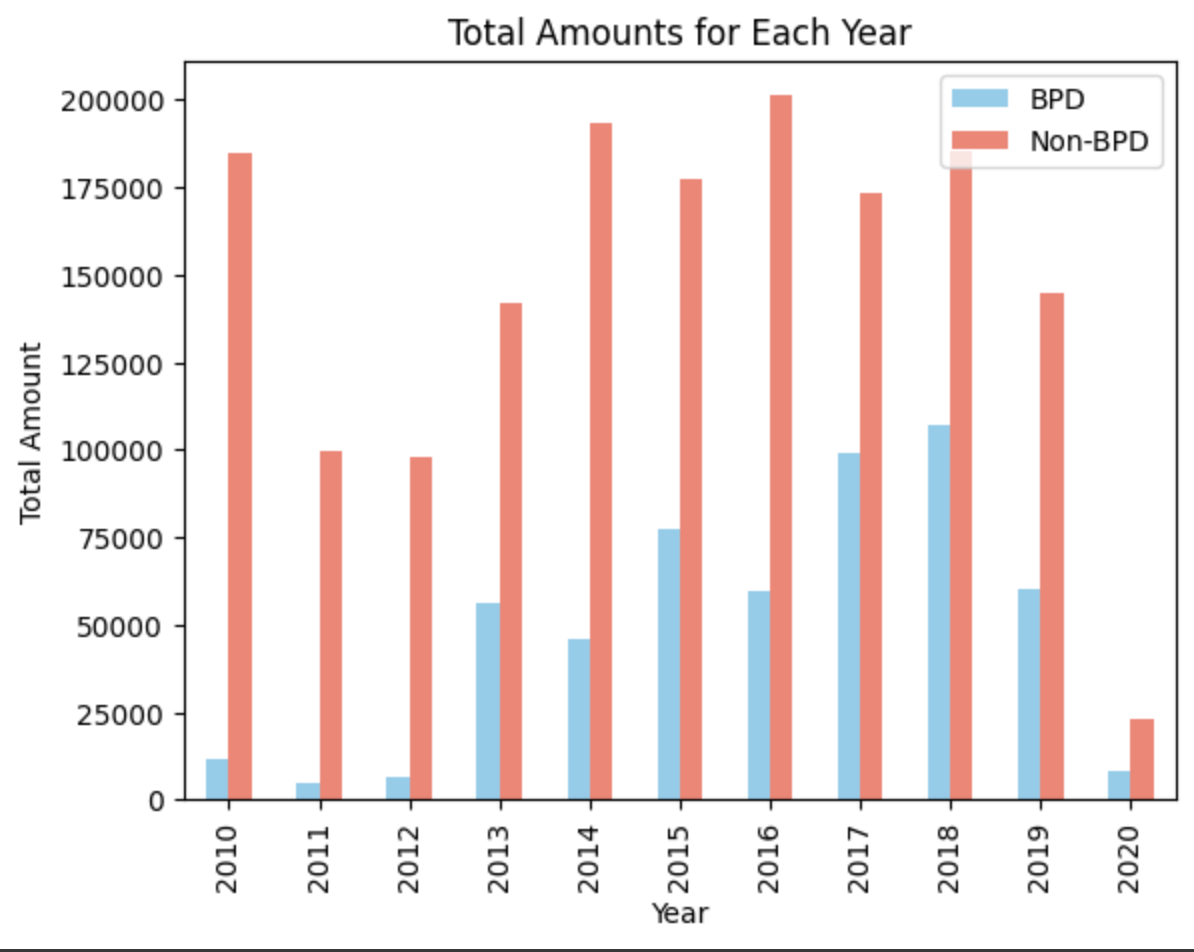
[insert links]

* <https://colab.research.google.com/drive/1gVfObsV1cbzK5XpTjjA9xc1m9cBou1LY#scrollTo=HNFvMTdVnkc6>

**2. Individual team member updates:**

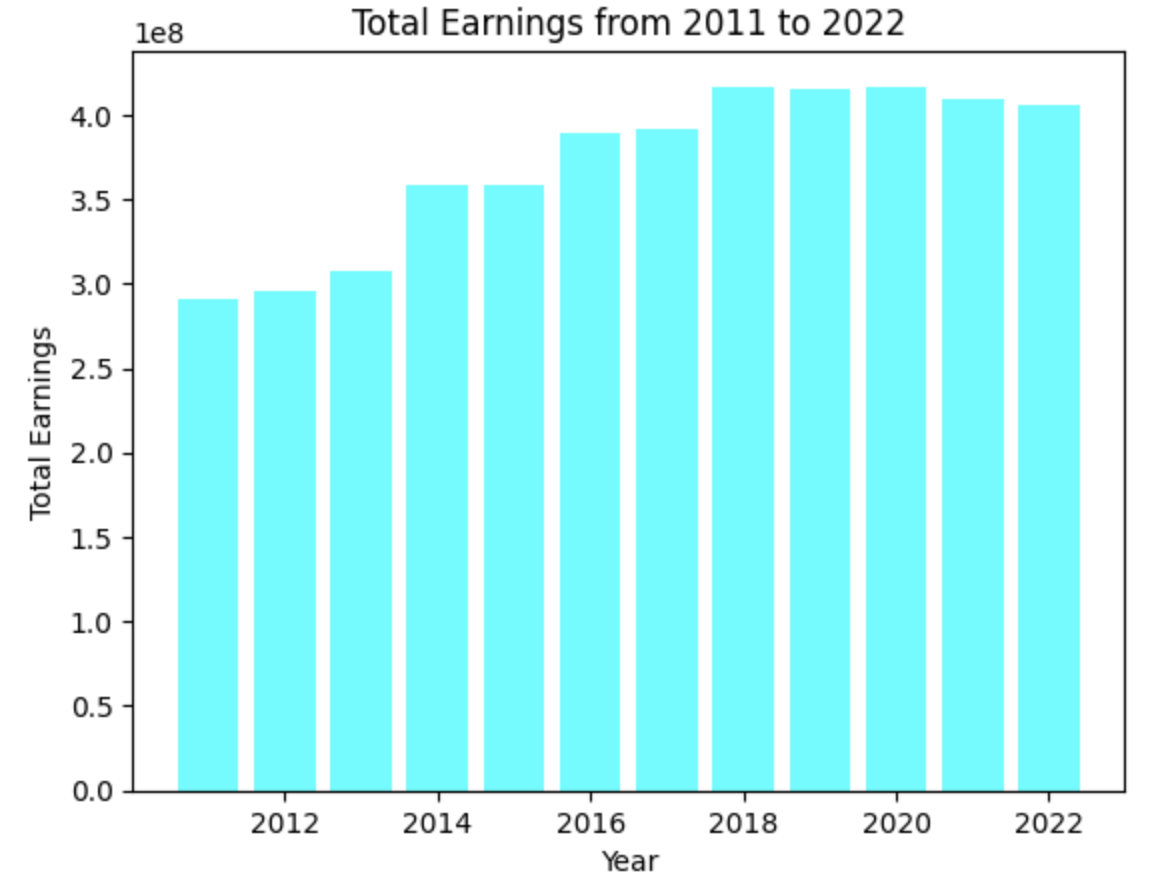
[Each team member should give a summary of their personal contributions]

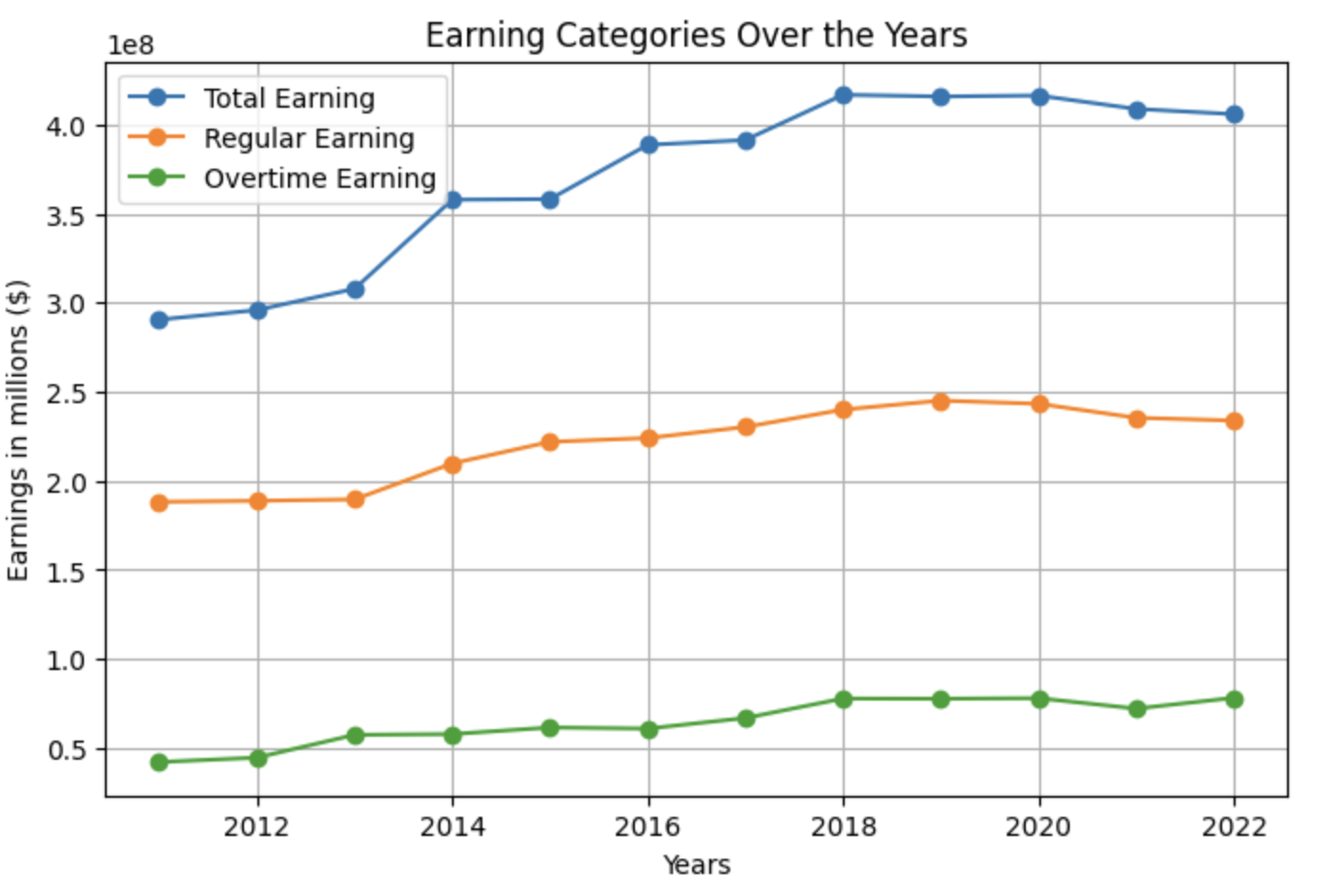
***[Nurassyl Medeu]***

* Added Field activity dataset to folder
* Prior analysis of Campaign Contribution data
  + Average contributions made by BPD employees versus non BPD Police employees
  + 
  + Total contributions made by BPD employees versus non BPD Police employees
  + 

***[Riva Sun]***

* Answering question from
* 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.)?
* The data shows that salaries for Boston Police Department (BPD) employees have been going up over time[1].
* Also, by the data plotted, extra pay and normal pay have both increased[2].
* The data indicates that the average salary for non-BPD city workers is approximately half that of BPD employees.

[1]: 

[2]: 

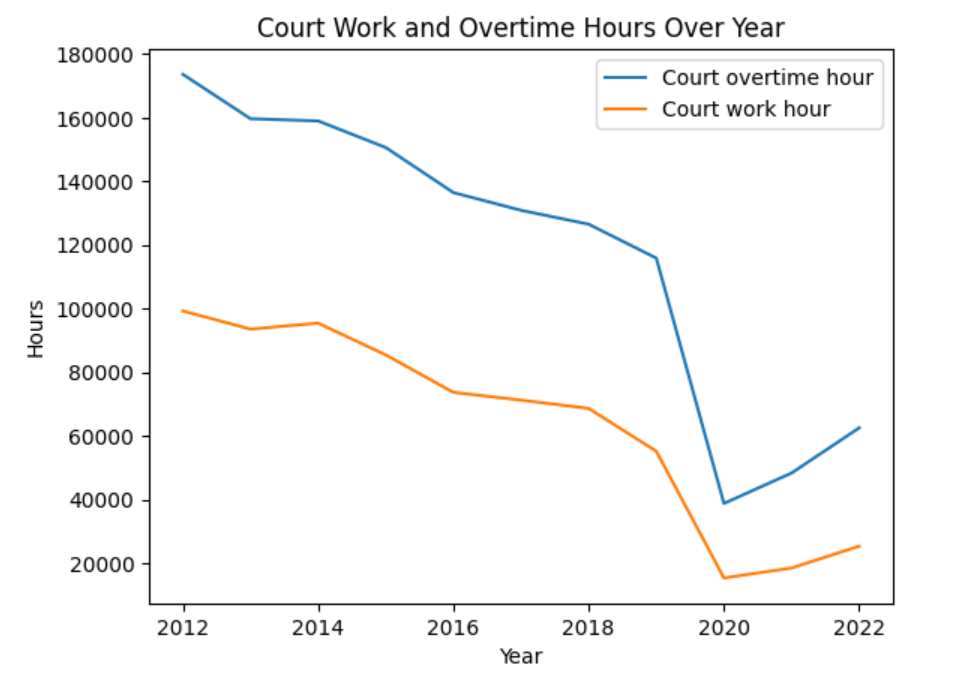
***[Truc Duong]***

* Find library/ tools to convert zip code -> longitude, latitude. Then we can use Folium and Selenium library to produce cluster map of frequency of overtime based on customer zipcode or department zipcode

-> Identify geographical characteristics of occurrences of overtime pay

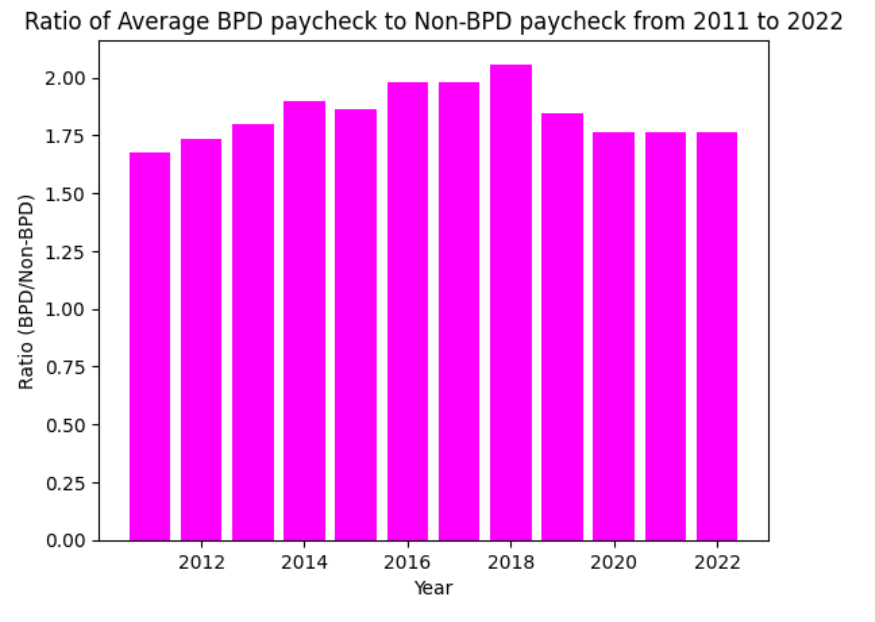
Some resource found:

* https://simplemaps.com/data/us-zips
* https://www.r-bloggers.com/2011/01/my-first-r-package-zipcode/
* Analyze Court Work hour and Overtime hour from 2012 - 2022



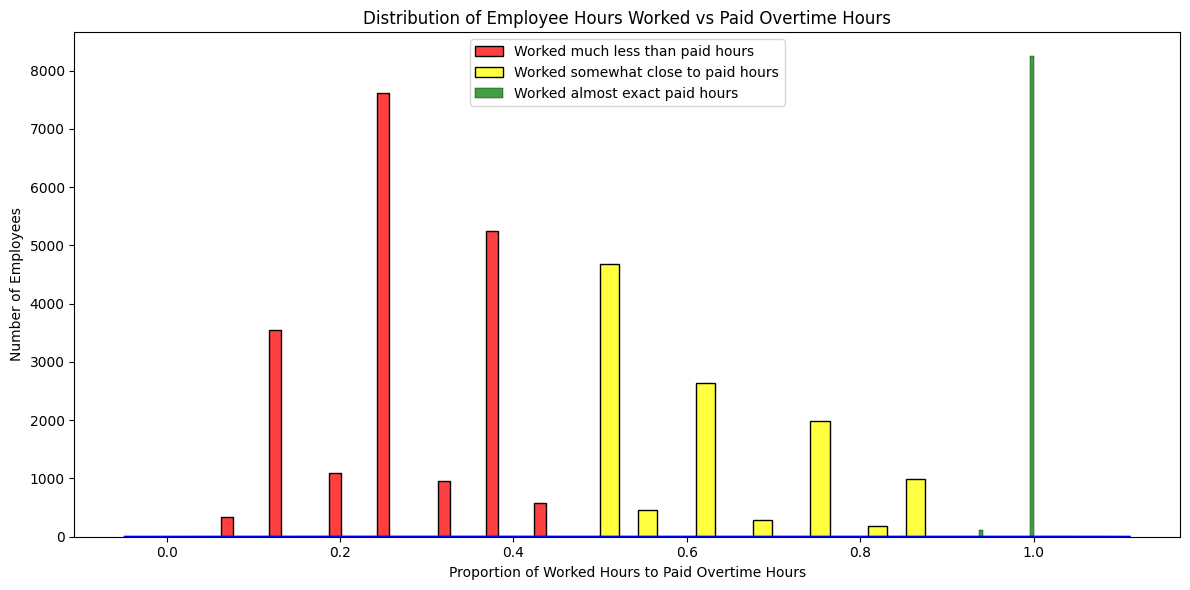
* We can see that there was a significant decrease in court work hours and overtime hours during 2020. A possible reason for this could be the covid pandemic during 2019-2020.
* Also, notice that during 2022, there is only data available until Oct 2022, not the entire year.
* We notice that overall, court overtime hour has always been higher than the court actual work hour

***[Can Wang]***

* Fixing some code that stops working after importing a new set of data. Due to variable name overlay.
* Analyze the earnings between BPD and non BPD. Found out that BPD usually has 1.8 times the payment that other departments may have. It is consistently higher.
* 

***[Al Mbaye]***

* Created and organized the court overtime datasets into structured csv files
* Loaded the court overtime data and conducted preprocessing to standardize the data
* Created a graph to display the distribution of ratios of overtime worked vs overtime paid



* The small bars at intermediate ratios appear to be outliers due to their low frequency

**3. Issues or blockers:**

* There are some inconsistencies in data across different years in the datasets, which make it difficult for us to do a general analysis.
* There are certain fields in the dataset that we don’t have definition or information about (i.e: officer ranks, overtime pay type, officer titles, etc). We’re reaching out to the PM to find approaches to this problem
* Difficult to find official demographics data of the police officers

**4. Plans for next week**

* Characterize wasteful BPD overtime practices by looking at how overtime hours paid compare with hours worked.
* Find out how much overlap is there between frequency of overtime users and officers who meet certain criteria
* Preprocess and start preliminary analysis on “campaign contribution data”
* Collect, summarize, clean up and answer the base questions required by the client